Université catholique de Louvain

Valibel - Centre de recherche Discours et Variation
ILC - Institut Langage et Communication

Thèse de doctorat

Evaluation contextuelle de la (dis)fluence en production et perception

Pratiques communicatives
et formes prosodico-syntaxiques en français

Auteure :
Iulia Grosman

Promotrices :
Prof. Anne Catherine Simon
Prof. Liesbeth Degand

Thèse soutenue en vue de l'obtention du grade de
Docteur en Langues et Lettres

Membres du jury :
Prof. Lambert Isebaert (Université catholique de Louvain, président)
Prof. Anne Catherine Simon (Université catholique de Louvain, promotrice)
Prof. Liesbeth Degand (Université catholique de Louvain, promotrice)
Prof. Maria Candea (Université Paris Sorbonne Nouvelle - Paris 3)
Prof. Philippe Boula de Mareüil (CNRS, LIMSI)
Prof. Laurence Meurant (Université de Namur)

Septembre 2018

Table des matières

Liste des figures i

Liste des tableaux vii

Liste des abréviations ix

Résumé (FR) xi

Résumé (EN) xiii

Remerciements xv

Avant-propos : De la « bafouille » … à la (dis)fluence xix

1 Introduction 1

1 Approches de la fluence . . . 2
1.1 Fluence cognitive, sociale ou linguistique ? . . . 2
1.2 Domaines de la linguistique ayant traité la notion de (dis)fluence . . . 7
1.2.1 Linguistique clinique : *dys*-fluence pathologique . . . 7
1.2.2 Linguistique des apprenants : natifs /vs/ non-natifs . . . 8
1.2.3 Linguistique computationnelle : erreurs de performance . . . 11
1.2.4 De la chaîne au schéma . . . 12
1.3 (Dis)fluences : signaux et indices à double facette . . . 15
2 Pluridimensionnalité du continuum fluence-disfluence . . . 18
2.1 Impression globale des locuteurs et/ou des auditeurs ? . . . 18
2.2 Achoppements syntaxique, prosodique, pragmatique (Gars-Fribourg) . . . 20
2.3 Disfluence et *monitoring* : de Levelt à Günther ? . . . 23
2.3.1 Langage : processus et résultat . . . 23
2.3.2 Disfluences, micro et macroplanifications . . . 25
2.3.3 Boucle perceptuelle : anticipation des erreurs . . . 26
2.3.4 Sources des différents types de *repairs* . . . 29
3 Définition synthétique de la prosodie . . . 32
3.1 Systèmes et fonctions . . . 32
3.2 Nature des unités et niveaux d'interactions . . . 34
3.3 Phonostyle et (dis)fluences . . . 36
4 Objectifs et canevas de la thèse . . . 39

2 Variations linguistiques des (dis)fluences **43**
1 Indices linguistiques au service de la (dis)fluence . . . 44
1.1 Approche acoustico-prosodique au sein de la zone de (dis)fluence . 44
1.1.1 Quelles zones de « turbulences » : *reparandum*, point d'interruption, *reparans* ? . . . 44
1.1.2 Corrélats acoustiques de durée des zones de (dis)fluence . 45
1.1.3 Corrélats acoustiques de l'intonation des zones de (dis)fluence . . . 48
1.1.4 Interactions des paramètres prosodiques et des zones de (dis)fluence . . . 49
1.2 Approche prosodique du contexte d'actualisation des (dis)fluences . 52
1.3 Définition syntaxique du contexte d'actualisation des (dis)fluences 57
1.4 Proposition d'évaluation contextuelle des (dis)fluences en production et en perception . . . 60
1.4.1 Limites de l'approche fréquentielle de la saillance . . . 61
1.4.2 Unification du concept : approche sociocognitiviste de la saillance . . . 63
1.4.3 Proposition de redéfinition et de formalisation du continuum fluence-disfluence . . . 65
2 Définitions des (dis)fluences . . . 70
2.1 Pauses silencieuses . . . 70
2.1.1 Hétérogénéité des définitions . . . 71
2.1.2 Eléments de fluence dialogique . . . 72
2.1.3 Structurante /vs/ d'hésitation . . . 73
2.1.4 Contextes syntaxiques en fonction des profils de durée . . . 74
2.2 Allongements vocaliques . . . 76
2.2.1 Grammatical /vs/ d'hésitation . . . 76
2.2.2 Détection et perception des allongements . . . 77
2.2.3 Contextes syntaxiques et prosodiques des allongements . 78
2.3 Pauses pleines . . . 80
2.3.1 Pause pleine /vs/ allongement vocalique . . . 81
2.3.2 D'appui /vs/ d'hésitation – continuité ou rupture . . . 82
2.3.3 Contextes syntaxique : autonomie /vs/ dépendance . . . 84
2.3.4 Contextes prosodiques : caractéristiques acoustiques . . . 86
2.4 Répétitions lexicales à l'identique . . . 88
2.4.1 Typologies des répétitions . . . 88
2.4.2 Structure interne de la répétition lexicale . . . 91
2.4.3 Caractéristiques acoustiques et prosodiques des répétitions 94
2.4.4 Rupture ou continuité : fluence ou disfluence . . . 97
3 Niveaux et types de ruptures : cooccurrences et contextualisation des marques . . . 98
3.1 Fréquences de cooccurrences . . . 99
3.2 Production : impact des cooccurrences sur la forme acoustique des marques . . . 102
3.3 Perception : impact des cooccurrences sur la perception et l'évaluation des marques . . . 103
4 Hypothèses : la disfluence comme outil de fluence ? . . . 105

3 Données et méthodologie **109**
1 Présentation des données de corpus . . . 109
1.1 Le *LOuvain Corpus of Annotated Speech - French* (LOCAS-F) . . 110
1.2 Le Corpus de parole des Humoristes français (C-Humour) 111
1.3 Description des paramètres situationnels des corpus 111
2 Principes d'annotation des corpus . . . 118
2.1 Annotations des (dis)fluences . . . 118
2.1.1 Annotations et validation au niveau lexical . . . 118
2.1.2 Annotations et validation au niveau syllabique . . . 122
2.2 Méthodologie de description des structures de (dis)fluences . . . 123
2.3 Annotations prosodiques et syntaxiques . . . 124
3 Distributions des phénomènes annotés dans le corpus . . . 131
3.1 Description quantitative du temps de parole . . . 131
3.2 Description quantitative des unités prosodiques et syntaxiques . . . 132
3.3 Description quantitative des (dis)fluences . . . 133
4 Principes d'une pragmatique expérimentale . . . 135
4.1 Complémentarité des études en production et en perception . . . 135
4.2 Méthodologie au service de la convergence des données : schéma récapitulatif . . . 136

4 Evaluation de la (dis)fluence des pauses silencieuses **139**
1 Méthodologie de catégorisation des pauses . . . 141
1.1 Théories de détection et de classification des (seuils de) pauses . . 141
1.2 Types de seuils, de mesures et de transformations . . . 144
1.3 Méthodologie de classification des pauses en *clustering* . . . 147
2 Objectifs et hypothèses . . . 150
3 Analyses . . . 152
3.1 Fréquence et durée des pauses en fonction des situations de parole 152
3.2 Fréquence et durée des pauses cooccurrentes à des (dis)fluences . . 156
3.3 Fréquence et durée des pauses aux frontières syntaxiques . . . 162
3.4 Fréquence et durée des pauses en fonction des contours intonatifs . 166
4 Conclusions autour de la rupture et de la continuité des pauses silencieuses . . . 174
4.1 Modélisation multifactorielle des durées pausales . . . 174
4.2 Synthèse et discussion . . . 184

5 Evaluation de la (dis)fluence des pauses pleines et des allongements **187**
1 Analyses des données de corpus . . . 188
1.1 Objectifs et hypothèses . . . 188
1.2 Fréquence et durée des pauses pleines et des allongements en fonction des situations de parole . . . 191
1.3 Fréquence et durée des pauses pleines et des allongements cooccurrents à des (dis)fluences . . . 198
1.4 Fréquence et durée des pauses pleines et des allongements aux frontières syntaxiques . . . 206
1.5 Fréquence et durée des pauses pleines et des allongements en fonction des contours aux frontières intonatives . . . 215
2 Evaluation perceptive des pauses pleines et des allongements . . . 219

2.1 Perception *offline* des pauses pleines et des allongements par des annotateurs experts . . . 220
2.1.1 Objectif et hypothèses . . . 220
2.1.2 Données et méthode . . . 221
2.1.3 Instructions . . . 223
2.1.4 Analyses et résultats . . . 224
2.1.5 Modélisation des allongements en fonction des résultats d'expérience . . . 226
2.1.6 Conclusions et discussion . . . 231
2.2 Comparaison des méthodologies *online* et *offline* de l'annotation des pauses pleines et des allongements d'hésitation . . . 232
2.2.1 Objectif et hypothèses . . . 232
2.2.2 Données et méthode . . . 233
2.2.3 Analyses et résultats . . . 236
2.2.4 Discussion sur la comparaison des méthodes . . . 240
3 Conclusions et discussion : la rupture et la continuité des pauses pleines et des allongements . . . 241

6 Evaluation de la (dis)fluence des répétitions à l'identique 249
1 Analyse des données de corpus . . . 250
1.1 Objectifs et hypothèses . . . 250
1.2 Fréquences et proportions des répétitions en fonction des situations de parole et traits situationnels . . . 252
1.3 Fréquences et structures des répétitions en fonction des cooccurrences . . . 256
1.4 Fréquences et positions des répétitions en fonction des unités syntaxiques et tours de parole . . . 266
1.5 Forme acoustique des répétitions à l'identique . . . 269
1.6 Conclusions sur l'évaluation des répétitions produites . . . 276
2 Perception des répétitions avec et sans marqueur d'édition . . . 279
2.1 Expérience 1 : fluence cognitive des répétitions . . . 280
2.1.1 Matériel et méthode . . . 281
2.1.2 Hypothèses . . . 287
2.1.3 Résultats . . . 288
2.1.4 Conclusions et discussion . . . 293
2.2 Expérience 2 : évaluation contextualisée des structures des répétitions . . . 294
2.2.1 Matériel et méthode . . . 294
2.2.2 Hypothèses . . . 300
2.2.3 Importance du stimulus, du participant et des seuils équidistants . . . 302
2.2.4 Résultats . . . 304
2.2.5 Conclusions et discussion . . . 309
3 Conclusions et discussions de l'évaluation des répétitions à l'identique sur les continuums fluence-disfluence . . . 311

7 Conclusion et discussion 315
1 Résumé des objectifs et des hypothèses . . . 316

2 Résumé des principaux résultats 318
3 Discussion et perspectives 325

Bibliographie **329**

Index des auteurs **354**

A Description de C-Humour : La parole des humoristes français **3**

B Description de LOCAS-F :
Louvain Corpus of Annotated Speech - French **9**

C Description et référence des annotations **19**

D Protocole ARC d'annotation des (dis)fluences (EN) **51**

E Données de corpus & matériel d'expérience **85**

F Structure des données de la thèse et accès aux données **101**

Liste des figures

1 Capture d'écran de l'émission « Quotidien » du 02 décembre 2016. xxi

1.1 Modèle de cognition située – inspiré de Malinin (2016:5). 3
1.2 Définition et interaction des types de contextes et de processus de formulations d'attentes – traduit de Schmid & Günther (2016:3). 5
1.3 Modèle contextualisé de tripartition de la fluence en langue non native – traduit de Segalowitz (2016:90). 8
1.4 Modèle non contextualisé de tripartition de la fluence en langue native – traduit de Götz (2011:42). 10
1.5 Schéma de structure interne d'une (dis)fluence – inspiré de Nakatani & Hirschberg (1993). 13
1.6 Modèle de production orale du langage Levelt (1989) – traduction adaptée de Mazur-Palandre (2009:63). 23
1.7 Modèle de traitement de la parole – inspiré de Postma (2000:99). 27
1.8 Association des types de *repairs* aux modules de *monitoring* et aux niveaux de syntaxe. 28
1.9 Modèle de compilateur prosodique au regard des systèmes de production et de compréhension de la parole de Di Cristo (2013:44). 36

2.1 Définition minimale du contexte syntagmatique local des (dis)fluences. . . 53
2.2 Amalgame des relations de saillance et de fréquences. 62
2.3 Quatre types de saillances issues de la comparaison d'indices linguistiques et des attentes – adapté de Schmid & Günther (2016). 64
2.4 Proposition de redéfinition du continuum fluence-disfluence par extension du cadre sociocognitif de la saillance. 66
2.5 Hypothèse de catégorisation des items sur le continuum fluence-disfluence. 67
2.6 Type de structure des répétitions et fonctions associées de Hieke (1985) à Shriberg (1995). 95

3.1 Distribution des temps d'enregistrement, nombre de mots et nombre de syllabes acoustiquement analysables par situation et corpus. 113
3.2 Distribution des enregistrements en fonction du degré d'élicitation. 115
3.3 Distribution des enregistrements en fonction du cadre d'interaction. 115
3.4 Distribution des enregistrements en fonction du nombre de locuteurs. . . . 116
3.5 Distribution des enregistrements en fonction du degré d'interactivité. . . . 116
3.6 Distribution des enregistrements en fonction du degré de préparation. . . . 116
3.7 Distribution des enregistrements en fonction du degré d'orientation médiatique. 116

3.8 Méthodologie de description d'une structure de (dis)fluence. 125
3.9 Terminologie utilisée dans les approches actuelles de la hiérarchie des domaines prosodiques – issu de Di Cristo (2011). 126
3.10 Schéma relationnel des annotations et des segmentations des corpus. . . . 130
3.11 Distribution des types de structures de (dis)fluence(s) par situation de parole. 134
3.12 Schéma des dynamiques sous-jacentes au cadre méthodologique. 138

4.1 Distribution (asymétrique) de la durée des pauses silencieuses du corpus Eurom en situation spontanée 142
4.2 Distribution de la durée des pauses silencieuses du corpus Eurom en situation spontanée après transformation logarithmique. 142
4.3 Diagramme en boîte à moustaches de la distribution des durées pausales (LOCAS-F). 145
4.4 Distribution de 4 mesures de distributions pausales. 148
4.5 Nombre d'occurrences et médiane de la durée pausale en fonction de la situation de parole dans LOCAS-F. 152
4.6 Différence de la durée des distributions pausales par situation de parole en logarithme et durée relative avec mise en valeur des variations individuelles. 153
4.7 Distribution des durées pour la situation du discours politique. 154
4.8 Distribution des durées pour la situation du journal parlé. 154
4.9 Distribution des durées pour la situation du récit conversationnel. 154
4.10 Distribution des durées pour la situation de l'homélie. 154
4.11 Graphique des tableaux croisés du degré de préparation et des chaînes de cooccurrences. 159
4.12 Durées pausales en fonction des cooccurrences. 160
4.13 Durées pausales en fonction des cooccurrences et du degré de préparation. 160
4.14 Durées pausales en fonction de la position syntaxique de la pause. 163
4.15 Durées pausales en fonction de la position syntaxique et du degré de préparation. 163
4.16 Diagramme en boîte à moustaches de la distribution des durées pausales en fonction de leurs cooccurrences et de leur localisation syntaxique et du contour intonatif précédant la pause. 165
4.17 Nombre d'occurrences et médiane de la durée pausale en fonction du contour intonatif la précédant. 167
4.18 Graphique des tableaux croisés des chaînes de cooccurrences et du contour intonatif précédant la pause. 169
4.19 Diagramme en boîte à moustaches de la distribution des durées pausales en fonction des chaînes de cooccurrence et du contour intonatif précédant la pause. 170
4.20 Schématisation de la durée des pauses en fonction des chaînes de cooccurrences et des contours intonatifs des syllabes précédentes. 173
4.21 Diagramme en boîte à moustaches de la distribution des durées pausales en fonction de la localisation syntaxique de la pause et du contour intonatif précédant la pause. 175

4.22 Diagramme en boîte à moustaches de la distribution des durées pausales en fonction des contours précédant la pause et de la localisation syntaxique de la pause, regroupé par chaîne de cooccurrences. 177
4.23 Distribution des enregistrements en fonction du degré d'élicitation. 178
4.24 Distribution des enregistrements en fonction du cadre d'interaction. 178
4.25 Contribution des variables (ajustement à la forêt d'arbres décisionnels). . 179
4.26 Arbre de décision des durées pausales en fonction de la position syntaxique, du degré de préparation et de la cooccurrence avec des (dis)fluences. 180
4.27 Arbre de décision des durées pausales en fonction de la position syntaxique et de la cooccurrence avec des (dis)fluences. 182
4.28 Arbre de décision des durées pausales en fonction de la position syntaxique, de la cooccurrence avec des (dis)fluences, du type de contour prosodique et du degré de préparation. 183

5.1 Diagramme en boîte à moustaches de la distribution des durées des pauses pleines et des allongements. 191
5.2 Distributions de deux mesures de durée des pauses pleines et des allongements. 192
5.3 Nombre d'occurrences et médianes des pauses pleines et des allongements en fonction de la situation de parole. 194
5.4 Distribution des durées des pauses pleines et des allongements des conférences scientifiques. 196
5.5 Distribution des durées des pauses pleines et des allongements des interviews sociolinguistiques. 196
5.6 Distribution des durées des pauses pleines et des allongements des performances radiophoniques. 197
5.7 Distribution des durées des pauses pleines et des allongements des récits conversationnels. 197
5.8 Graphiques des tableaux croisés du degré de préparation et des chaînes de cooccurrences des pauses pleines et des allongements. 202
5.9 Durées en fonction des cooccurrences des pauses pleines et des allongements. 204
5.10 Durées des pauses pleines en fonction des cooccurrences et du degré de préparation. 205
5.11 Durées des pauses pleines et des allongements en fonction de la position syntaxique. 207
5.12 Graphiques des tableaux croisés des pauses pleines et des allongements en fonction du degré de préparation et de la position syntaxique. 208
5.13 Durées des pauses pleines en fonction de la position syntaxique et du degré de préparation. 209
5.14 Diagramme en boîte à moustaches de la distribution des durées des pauses pleines en fonction de leurs cooccurrences et de leur localisation syntaxique. 210
5.15 Diagramme en boîte à moustaches de la distribution des durées des allongements en fonction de leurs cooccurrences et de leur localisation syntaxique. 211
5.16 Nombre d'occurrences et médianes de la durée des pauses pleines et des allongements en fonction du contour intonatif la précédant. 215

5.17 Représentation du mouvement intonatif intrasyllabique des allongements en fonction des cooccurrences. . . . 217
5.18 Représentation du mouvement intonatif intrasyllabique des pauses pleines en fonction des cooccurrences. . . . 218
5.19 TextGrids avant annotation. . . . 222
5.20 TextGrids après analyse prosodique. . . . 222
5.21 TextGrids après annotation. . . . 222
5.22 Diagrammes en barres des tendances d'annotation par annotateur et par phénomène. . . . 225
5.23 Diagramme en barres des tendances d'annotation par annotateur, par phénomène et par fichier. . . . 226
5.24 Matrices de confusions interannotateurs. . . . 227
5.25 Taille des effets fixes pour la prédiction de l'annotation des allongements dits d'hésitation. . . . 229
5.26 Fenêtre de dialogue du programme « Tapping Annotator ». . . . 234
5.27 Diagramme en boîte à moustaches des pauses pleines par type de détection. 236
5.28 Diagramme en boîte à moustaches des allongements par type de détection. 236
5.29 Distribution des pauses pleines par type de détection. . . . 237
5.30 Distribution des allongements par type de détection. . . . 237
5.31 Diagramme en boîte à moustaches du mouvement intrasyllabique des pauses pleines par type de détection. . . . 238
5.32 Diagramme en boîte à moustaches du mouvement intrasyllabique des allongements par type de détection. . . . 238
5.33 Diagramme en boîte à moustaches du mouvement intrasyllabique des pauses pleines par type de détection. . . . 238
5.34 Diagramme en boîte à moustaches du mouvement intrasyllabique des allongements par type de détection. . . . 238
5.35 Diagramme en boîte à moustaches de la durée des marqueurs détectés en fonction des pauses précédant/succédant et de la méthodologie de détection. . . . 239

6.1 Proportion de temps de répétitions sur le temps d'articulation. . . . 253
6.2 Association de répétitions par rapport aux autres (dis)fluences et segments non (dis)fluents par situations de parole. . . . 254
6.3 Association de répétitions par rapport aux autres (dis)fluences et segments non (dis)fluents par degré de préparation. . . . 255
6.4 Analyse en composantes multiples des structures de répétitions et des cooccurrences par variable. . . . 262
6.5 Analyse en composantes multiples des structures de répétitions et des cooccurrences par rapport à leur pertinence informationnelle. . . . 262
6.6 Analyse en composantes multiples par variable et par modalité avec représentation des ellipses de confiance. . . . 265
6.7 Répartition des profils de répétitions en fonction du rapport (durée du *reparandum* / durée du *reparans*). . . . 270
6.8 Répartition des profils de répétitions en fonction du rapport (durée du *reparandum* / durée du *reparans*), en fonction du terme d'édition et de la nature du répétable . . . 272

6.9 Hauteur moyenne de l'*onset* et de l'*offset* du *reparandum* et *reparans* des répétitions par sexe des locuteurs . . . 275
6.10 Exemple de six paires de stimuli. . . . 281
6.11 Instruction de l'expérience 1. . . . 284
6.12 Prosogramme de 4 stimuli correspondant au même énoncé. . . . 285
6.13 Temps de réaction par condition de fluence. . . . 289
6.14 Différences entre conditions par méthode des moindres carrés. . . . 292
6.15 Exemple de résultat (anonymisé) du test F-IRI envoyé à un participant. . 295
6.16 Image de contextualisation d'une situation de parole non médiatique. . . . 298
6.17 Image de contextualisation d'une situation de parole médiatique. . . . 298
6.18 Effets de la variance des stimuli et des participants sur la taille des effets aléatoires du modèle. . . . 302
6.19 Médianes de fluence perçue par affirmation et condition de fluence. . . . 303
6.20 Données brutes : évaluation des participants par degré de fluence phrastique de l'item et par dimension de fluence. . . . 305
6.21 Médianes de la fluence perçue par affirmation, en fonction des index F-IRI des participants. . . . 306

Liste des tableaux

3.1 Nombre et types de situations par corpus. . . . 112
3.2 Liste des (dis)fluences annotées. . . . 119
3.3 Classification de l'accord interannotateurs des (dis)fluences lexicales. . . . 120
3.4 Matrice de confusion de l'accord interannotateurs des (dis)fluences lexicales. 121
3.5 Accords interannotateurs des (dis)fluences au niveau syllabique. . . . 123
3.6 Types d'unités prosodiques et syntaxiques annotées dans les corpus. . . . 126
3.7 Catégorisation interactionnelle des segments temporels, syllabiques et lexicaux. . . . 131
3.8 Parole des locuteurs principaux : segments temporels, syllabiques et lexicaux. 132
3.9 Répartition des unités prosodiques et/ou syntaxiques annotées dans les corpus. . . . 132
3.10 Distribution des marques annotées dans les corpus. . . . 134

4.1 Distribution des durées des pauses en fonction des seuils inférieurs. . . . 146
4.2 Distribution des durées des pauses en fonction des seuils supérieurs. . . . 146
4.3 Fréquences et proportions des pauses en fonction des cooccurrences les plus fréquentes. . . . 158

5.1 Fréquences et proportions des pauses pleines et des allongements en fonction des cooccurrences. . . . 201
5.2 Position des pauses pleines par rapport aux tours de parole des dialogues et des multilogues. . . . 212
5.3 Position des allongements par rapport aux tours de parole des dialogues et des multilogues. . . . 213
5.4 Durées et proportions des enregistrements du sous-corpus. . . . 221
5.5 Protocole d'annotation des pauses pleines et des allongements dits d'hésitation. . . . 223
5.6 Tableau croisé d'accord interannotateurs sur toutes les marques. . . . 224
5.7 Accords interannotateurs par type de (dis)fluences au niveau syllabique. . . . 225
5.8 Tableau croisé d'accords interannotateurs des allongements. . . . 225
5.9 Tableau des effets fixes. . . . 229
5.10 Fréquence des phénomènes d'hésitation dans LOCAS-F en fonction des campagnes d'annotation. . . . 235
5.11 Résultat du croisement des campagnes d'annotation. . . . 235

6.1 Proportion de répétitions par rapport aux autres (dis)fluences et segments non (dis)fluents par situations de parole. . . . 254
6.2 Proportion de répétitions par rapport aux autres (dis)fluences et segments non (dis)fluents par degré de préparation. . . . 255

6.3 Proportion du nombre de répétés par corpus. 256
6.4 Fréquences des unités lexicales faisant l'objet d'une répétition par nombre de répétés et fréquences relatives par 1000 mots sur l'ensemble du corpus. 257
6.5 Longueur du répétable en nombre de mots. 258
6.6 Type de segments qui précèdent et suivent les répétitions. 261
6.7 Proportion relative d'actualisation des répétitions au sein des unités syntaxiques de LOCAS-F. 266
6.8 Fréquences absolues et relatives des répétitions en fonction du degré de frontière syntaxique dans LOCAS-F. 267
6.9 Position des répétitions par rapport aux tours de parole des dialogues et des multilogues. 268
6.10 Effets aléatoires de la modélisation des temps de réaction. 290
6.11 Effets fixes de la modélisation des temps de réaction. 291
6.12 Résumé des variables et niveaux de variables. 301
6.13 Effets fixes de la modélisation du degré de fluence en fonction du type de (dis)fluence. 304
6.14 Effets fixes de la modélisation du degré de fluence en fonction du type de média testé. 306

Liste des abréviations

	Signification (*équivalence anglophone)*
AA	Acoustiquement analysable
ARC	Action de recherche concertée
BDU	Unité de base du discours (*basic discourse unit)*
BIC	Critère d'information bayésien (*bayesian information criterion)*
CART	Arbre de classification et de régression (*classification and regression tree*)
DE	Délétion
md	Marqueur discursif (*discourse marker*)
ET	Terme d'édition (*explicit editing term*)
EN	Anglais
fàf	Face-à-face
F0	Fréquence fondamentale
FP	Pause pleine (*filled pause*)
FR	Français
FS	Faux départ (*false start*)
GI	Groupe intonatif
GMM	Mixte de modèle gaussien (*gaussian mixture model)*
ga0	Groupe accentuel accentuable
ga1	Groupe accentuel accentué
IL	Insertion lexicale
IP	Insertion parenthétique
IPU	Unité interpausale *(inter-pausal unit)*
L1	Langue première (langue maternelle)
L2	Langue seconde (langue apprise)
LG	Allongement (*lengthening*)
LLog	Rapport de Vraisemblance Logarithmique *log-likelihood ratio*
LOCAS-F	Corpus de parole Locas-F (*LOuvain Corpus of Annotated Speech - French*)
MCA	Analyse en correspondance multiple (*multiple correspondance analysis*)
MO	Mot outil
MP	Mot plein
POS	Partie grammaticale du discours (*part-of-speech*)
RI	Répétition à l'identique
RM	Répétition avec modification
SM	Substitution morphologique
SP	Substitution propositionnelle
TR	Troncation
UP	Pause silencieuse (*unfilled pause*)
UR	Unité de rection
WI	Intérieur (*within*)

UNIVERSITE CATHOLIQUE DE LOUVAIN

Résumé

Faculté de Philosophie, Arts et Lettres
ILC - Institut Langage et Communication

Docteur en Langues et Lettres

Evaluation contextuelle de la (dis)fluence en production et perception

par Iulia GROSMAN

L'objectif de cette thèse est d'étudier les effets des (dis)fluences sur la production et la perception de la parole par des locuteurs natifs francophones. La facilité avec laquelle nous produisons, percevons et interprétons la langue parlée dépend de divers facteurs linguistiques, situationnels et sociaux. Par conséquent, la façon dont on définit la « fluence » ou la « disfluence » dépend du type de contexte pris en compte et, plus précisément, du type d'attentes que l'on tire de ce contexte. Les marqueurs de fluence ou disfluence sont des dispositifs linguistiques tels que les pauses silencieuses et les pauses pleines, les allongements, les répétitions et les marqueurs discursifs (e.g. bon, ben, tu vois, quoi). L'hypothèse principale de cette recherche est que toutes les (dis)fluences ne sont pas égales et que certaines peuvent être vecteurs de fluence. Leur statut ambivalent leur permet de soutenir ou d'entraver la production et la perception de la parole en interaction. Les analyses ciblent spécifiquement la définition des attentes linguistiques, situationnelles et sociales des (dis)fluences et de la structure discursive. Elles révèlent notamment la variabilité et l'homogénéité des discours à travers différentes situations. Notre méthode combine l'analyse statistique et quantitative des corpus annotés et des expériences de perception. Chaque chapitre est consacré à un phénomène (pauses silencieuses, pauses pleines, allongements et répétitions). La *dis*fluence, linguistique ou contextuelle, y est déterminée par l'absence de congruence entre les attentes et le contenu produit ou perçu en contexte, potentielle créatrice d'une rupture discursive.

Mots-clés : *disfluence, fluence, prosodie, production, perception, parole*

UNIVERSITE CATHOLIQUE DE LOUVAIN

Résumé

Faculté de Philosophie, Arts et Lettres
ILC - Institut Langage et Communication

Docteur en Langues et Lettres

Evaluation contextuelle de la (dis)fluence en production et perception

par Iulia GROSMAN

(EN)

The objective of this thesis is to study the effects of (dis)fluencies on the production and perception of speech by French native speakers. The ease at which we produce, perceive and interpret spoken language depends on various linguistic, situational and social factors. Hence, the way one defines "fluency" or "disfluency" itself depends on which type of context is taken into account, and more precisely, what kind of expectations one draws from this context. (Dis)fluencies are linguistic devices such as silent and filled pauses, vocalic lengthening, lexical repetitions or discourse markers (e.g. well, you know, I mean). The main hypothesis of this research is that not all (dis)fluencies are equal and that some of them may participate in the evaluation of fluency. Hence, their ambivalent status can either support or hinder the production and perception of speech in interaction. These analysis specifically target the definition of linguistic, situational and social expectations of disfluencies and discourse structure, by observing the variability and homogeneity of discourses across various situations. The method combines the statistical and quantitative analyses of annotated corpora and perception experiments. Each chapter is devoted to one phenomenon (silent pauses, filled pauses, lengthening and repetitions). Linguistic or contextual disfluency is considered when there is a mismatch between expectations and the produced/perceived content in context, potentially creating discursive rupture.

Keywords : *disfluency, fluency, prosody, production, perception, speech*

Remerciements

Quel est le lien entre le monde de la recherche et le monde de l'acrobatie, les deux passions qui ont encadré cette thèse ? Le doctorant est un équilibriste à la recherche du dépassement de soi et d'un certain épanouissement intellectuel. L'intégration progressive du chercheur au monde universitaire le transforme ainsi en jongleur. Il est amené à concilier le jeu socio-professionnel de l'Academia avec celui de son évolution personnelle. Au début de chaque thèse et en fin de parcours associé, une chance inégalée est donnée : l'expression de la reconnaissance. Durant mes années de recherches, la gratitude a été un outil pratique quasi-quotidien m'accompagnant à travers ces deux mondes. Elle est devenue la condition nécessaire de l'actualisation de cette thèse dans une atmosphère de bien-être. J'espère que mes lecteurs ressentiront que la rédaction de ces remerciements ne répond pas à une simple pratique protocolaire mais consacre, au sens propre du terme, toutes les personnes qui ont accompagné les multiples facettes de mon développement.

Ma reconnaissance va avant tout au nombre de chercheurs, linguistes, sociologues, sémioticiens, acousticiens qui ont nourri le présent travail de thèse.

Malgré des années de pratique, il est deux personnes qu'il me semble impossible de remercier en toute équité, les Prof. Anne Catherine Simon et Prof. Liesbeth Degand qui ont dirigé ce travail. S'il est des modèles de femmes à suivre dans le monde universitaire, cela fait longtemps que je n'ai plus d'hésitation. Ces ressources critiques et éthiques tant au niveau intellectuel, professionnel qu'humain m'ont soutenue depuis maintenant bientôt 7 ans. Leur esprit aiguisé m'a convaincu de ma spécialisation dès mon arrivée à l'UCL. Leur rigueur a tant de fois remis dans le droit chemin mon esprit en ébullition, parfois aussi déterminé que dispersé, avouons-le. A elles deux, pour leur aiguillage attentif et leur bienveillance continue, en un mot comme en cent : Merci.

Dans le domaine de la linguistique, les femmes sont nombreuses. Je remercie aussi sincèrement mes collègues qui ont ramé dans la même galère, Ingrid, Amandine, Bénédicte, Anne-Sophie, Ludivine, Lucie. Merci de vous être épanouies avec moi dans vos thèses respectives.

Si le doctorat est souvent associé à un long voyage en solitaire, il n'en est pas moins le fruit d'une série de rencontres enrichissantes. En ce sens, je tiens à exprimer toute ma gratitude envers les chercheurs et chercheuses m'ayant encadré et encouragé avec beaucoup de bienveillance. Une pensée particulière revient à Catherine Bolly (Univ. zü Köln), Maria Candea (Univ. Paris III), George Christodoulides (Univ. catholique de Louvain), Laurence Mundschau (Univ. catholique de Louvain), Antoine Auchlin (Univ. de Genève), Mathieu Avanzi (Univ. catholique de Louvain), Amélie Cougnon (Univ. catholique de Louvain) et Roxane Bertrand (Univ. Aix-en-Provence).

La présente thèse s'inscrit dans le cadre d'une Action de Recherche Concertée (ARC) « Fluency - Disfluency : A multi-modal approach to Fluency and Disfluency Markers » financée par la Fédération Wallonie-Bruxelles (bourse nr. 12/17-044). Les équipes de VALIBEL, du CECL et du LSFB-LAB, présentes au sein de l'Université catholique de Louvain et de l'Université de Namur, ont mené pendant cinq années des recherches sur les liens entre fluence et disfluence en langue orale et langue des signes au travers de trois modalités : les discours en langues premières (français et anglais), en langue seconde (anglais) et en langue des signes (française de Belgique).

J'exprime ma gratitude la plus sincère aux chercheurs qui me font l'honneur d'accepter de faire partie de mon jury de thèse : Prof. Lambert Isebaert, Prof. Anne Catherine Simon et Prof. Liesbeth Degand (Univ. catholique de Louvain), ainsi que Prof. Maria Candea (Université Paris Sorbonne Nouvelle - Paris 3), Prof. Philippe Boula de Mareüil (CNRS, LIMSI) et Prof. Laurence Meurant (Université de Namur)

Je dois également beaucoup à mon entourage qui, parfois loin du monde de la recherche, n'a jamais douté que je mènerais à bien cette aventure. Si ces proches n'ont peut-être pas su arrêter de me dire « *et*, *euh* c'est quand que tu finis *déjà* ? », ils ont surtout compris quand et comment m'épauler dans les moments les plus difficiles. Merci à vous entre autres : Gui, Ilias, Caro, Lena, Marco, Avgi, Tom, Laura, Cha, Monia, Declan, David, Jean-François.

Enfin et surtout, à vous mes parents, un remerciement, mis en exergue, en fin de texte, vous met à l'honneur ; vous, académiques fous et libertaires qui m'ont soutenue à bien des niveaux. De loin, vous êtes tout de même parvenu à me tenir la main, à me mettre une petite tape dans le dos lorsque nécessaire. Pour la transmission de votre ouverture d'esprit et curiosité inégalée. Merci. Papa, un sens analytique s'est rarement mêlé à un esprit pragmatique aussi créatif : ils m'ont sauvée quand je ne voyais plus de solutions à mes problèmes. A toi, Maman, tes relectures incessantes et précieuses pendant 9 ans, ta patience, nos interactions matinales sur Skype motivant nos recherches respectives : elles ont bâti notre relation actuelle et m'ont permis de continuer à me construire en toute sérénité. Elles ont rendu possible ce manuscrit. Pour votre complémentarité, pour le sourire que vous m'apportez à chaque fois que je pense à vous, mon *muppet show* préféré, j'ai de la chance de vous avoir. MERCI.

« C'est au chercheur singulier de définir son corpus et ses méthodes, et surtout le sens de sa recherche, avec les disciplines, contre les disciplines, ou hors discipline.

Derrière cette défense d'une indisciplinarité, il y a évidemment une forme d'appel à l'autogestion de la recherche, à l'énonciation d'une voix personnelle qui, quoique traversée par les savoirs disciplinaires, ne répondrait pas à la question, déjà ancienne, du lieu d'où l'on parle, par l'affirmation 'de telle discipline'.

Et c'est paradoxalement à ce prix que les disciplines peuvent, probablement, ne pas perdre leur sens, le sens de la recherche. »

*(*Loty *2005:258)*

Avant-propos : De la « bafouille »... à la (dis)fluence

Un monologue écrit pour être lu. Un discours préparé. Une intervention publique et médiatique historique. Le 01 décembre 2016, pour la première fois dans l'histoire de la République française, un président ne brigue pas de second mandat :

> **Exemple (1)** [...] *aujourd'hui je suis conscient / des risques / que ferait courir une démarche la mienne qui* ***euh*** *ne rassemblerait pas* ***lage-largement*** *autour d'elle / aussi / j'****ai:*** *décidé de ne pas être candidat à l'élection présidentielle au renouvellement* ***donc*** *de mon mandat* [...] *dans les mois qui viennent* ***mon devoir mon seul devoir*** *sera de continuer à diriger le pays* [...] *en m'y consacrant pleinement /* ***et*** *dans le* ***dévouement*** [duvmã] *le plus total à la République / vive la République et vive la France* [...] ».[1]

En l'annonçant, François Hollande « *a bafouillé deux fois* » remarque l'équipe du Quotidien de Yann Barthès. Deux ? Certains linguistes annoteraient une pléthore de « bafouilles » ou du moins, d'interruptions du discours : des pauses silencieuses, des pauses pleines ou *euh*, des allongements, des reformulations, insertions, erreurs d'articulation, répétitions, etc. Ainsi, un court segment télévisuel nous fait prendre conscience de la différence entre les *bafouilles produites* et les *bafouilles perçues*.

Au journaliste ensuite de commenter « [...] et ce n'est pas un hasard si ces deux bafouilles entourent la phrase la plus importante du discours la phrase où il dit qu'il ne va pas se présenter [...] ». Les linguistes n'en ont pas fini de réfléchir : il y a aussi des *bafouilles interprétées* ! Ici, le locuteur-auditeur-commentateur s'accorde sur la logique de l'emplacement des bafouilles produites, et les associe à un événement non aléatoire chez le locuteur. Les régularités de ces événements semblent être de vraies clés de lecture de l'interprétable pour l'auditeur.

1. Transcription d'un segment extrait de l'émission « Quotidien » diffusée sur TMC – édition du 02 décembre 2016. La barre oblique marque la présence d'une pause tandis que les deux points indiquent les allongements perçus comme des hésitations. Le texte entre crochets symbolise la prononciation phonétique. Les deux bafouilles relevées par le commentateur Yann Barthès sont soulignées dans le texte. En gras, apparaissent les bafouilles repérées par le linguiste.

Que faire des autres indices volontaires ou signaux involontaires produits, généralement annotés par le linguiste mais ignorés de certains auditeurs-commentateurs ? Ne sont-ils pas perçus ou ne sont-ils simplement pas interprétés comme des achoppements critiques dans le discours du locuteur ? En effet, un auditeur ne perçoit pas nécessairement les mêmes bafouilles qu'un autre ; et lorsque ceux-ci les repèrent, ils ne leur accordent peut-être pas la même importance ni n'en font la même interprétation. Comment expliquer ces variations ?

Cet exemple rappelle à quel point la détection de ces phénomènes dépend du point de vue adopté. Ce sont ces bafouilles (*produites, perçues, interprétées*), et surtout, la manière dont ces trois facettes de bafouilles s'articulent et communiquent entre elles, qui font l'objet de cette thèse de doctorat. Ces différents objets sont souvent traités séparément bien que dans toute situation discursive les locuteurs les concilient : ils produisent, perçoivent et interprètent le flux discursif *de manière intégrée*. Intégrée, dans les deux acceptions du terme. Intégrée à un ensemble faisant partie d'un tout plus grand, car un individu et/ou une marque linguistique ne peuvent jamais totalement s'abstraire de leur temps ni de leur espace. Ils sont plus qu'un enchaînement de mots ou de séquences à un instant *t*. Intégrée également, vu la nature de la problématique qui articule les trois facettes. Celle-ci appelle à une approche *située*, *appropriée* vis-à-vis des dimensions théoriques et méthodologiques. Elle implique aussi des considérations épistémologiques inhérentes à la nature pluri- [2], inter- [3] voire trans- [4] disciplinaire du phénomène.

L'étude de ces bafouilles intéresse le linguiste en ce qu'elles sont reliées aux processus psycholinguistiques de production et de perception du langage, mais leur étude est aussi nécessaire lorsqu'il désire appliquer des procédures de traitement automatique du langage à du discours spontané. Ce travail de recherche tient à réconcilier production, perception et interprétation de nos bafouilles quotidiennes. Pour ce faire, il embrasse la dynamique d'allers-retours constants entre les résultats empiriques issus d'une part de l'observation des discours (*linguistique de corpus*) et d'autre part de leur perception (*linguistique expérimentale*).

2. Est-ce que ces multiples disciplines mobilisées par la recherche concourent à une réalisation commune, mais sans que chacune des disciplines n'ait à modifier sa propre vision des choses et ses propres méthodes ?

3. Ou notre recherche a-t-elle juste pour but d'élaborer un formalisme, « suffisamment général et précis pour permettre d'exprimer dans un langage unique les concepts, les préocupations et les contributions d'un nombre plus ou moins grand de disciplines » ?

4. Ou pourrions-nous parvenir à aller « au-delà de toute discipline » et à voir le locuteur et sa relation au concept de bafouille, de (dis)fluence, comme un objet qui traverse de manière cohérente toutes les disciplines possibles ? (Claverie 2010:22).

FIGURE 1 – Capture d'écran de l'émission « Quotidien » du 02 décembre 2016.

Chapitre 1

Introduction

Comme tout phénomène langagier, les « bafouilles » ont droit à leur néologisme : **les (dis)fluences** [1]. Celles-ci sont définies traditionnellement comme des « phénomènes d'interruption du flux discursif n'ajoutant pas de contenu prépositionnel à un énoncé » (Fox Tree 1995:709). Ces marqueurs s'actualiseraient tous les 6 à 20 mots sur 100 en fonction des discours et des locuteurs, des types de marque(s) pris en compte et dès lors de la manière dont le chercheur les définit. On y retrouve le plus souvent les pauses (silencieuses et pleines), les interruptions lexicales ou postlexicales, les autocorrections, les répétitions, les erreurs de prononciation, etc. Leur définition (*détection, identification/annotation* et *interprétation*) varie aussi en fonction des domaines scientifiques et de leur spécialisation. L'objectif de ce chapitre est de clarifier les acceptions du terme « disfluence » et de situer les concepts clés qui lui sont associés dans le cadre de cette recherche.

Les **écarts épistémologiques et terminologiques** peuvent s'appréhender en fonction du domaine d'étude envisagé (section 1). Si, le concept de fluence est fécond en sciences humaines, et plus particulièrement en psychologie cognitive et sociale (1.1), sa définition a également évolué différemment selon les domaines linguistiques envisagés. Les apports de ces domaines sont essentiels pour la présente recherche (1.2). Cette première partie révèle l'intérêt, voire la nécessité, d'une approche **pluridimensionnelle du continuum fluence-disfluence** (section 2) prenant en compte l'importance des définitions antérieures de la fluence et de leurs interactions (2.1). Elle montre la nécessité d'inclure dans notre recherche le point de vue du locuteur et de l'auditeur ainsi que leurs relations vis-à-vis de l'interprétation de l'interaction discursive (2.2). Il s'agit donc de définir notamment les conditions de production et de réception (perception et interprétation) du discours et leur impact sur le discours lui-même (2.3), entre autres sur sa forme lexicale mais

1. Appelées également *marqueurs de disfluence*, *phénomènes d'hésitations* ou *marques du travail de formulation*. A l'instar du terme « disfluence », celui d'« hésitation » désigne selon les auteurs soit les marques linguistiques elles-mêmes soit les processus cognitifs sous-jacents associés. Ces deux termes sont à la fois ambigus et trop précis. D'une part certains énoncés « bien formés » peuvent être hésitants (e.g. « et je ne sais pas quoi faire à ce propos ») et d'autre part il semble cavalier de réduire à priori toutes les marques linguistiques de « bafouilles » au seul processus cognitif « d'hésitation » d'un locuteur (Candea 2000a:12,20). Nous préférons ainsi le terme « (dis)fluence » qui désigne une marque linguistique pour laquelle nous ne présumons ni de l'impact communicatif ni de l'origine cognitive.

aussi *prosodique*[2] (section 3). Les objectifs centraux de cette thèse postulent ainsi que l'interprétation des (dis)fluences comme dispositifs fluents ou disfluents est **graduelle et conditionnée par l'analyse de leur usage en contexte**. La dernière partie dessine le plan de cette recherche en résumant les questions spécifiques à chacun de ses chapitres (section 4).

1 Approches de la fluence

> "As a rule, every cognitive task can be described along a continuum from effortless to highly effortful, which produces a corresponding metacognitive experience that ranges from fluent to disfluent, respectively. Thus, for example, watching a film at the cinema is more visually fluent than watching the same film on a small black and white television from the far end of a large room."
> (Alter & Oppenheimer 2009:209)

La richesse du concept de fluence dépasse les domaines d'étude de l'oral et de l'analyse de la parole de locuteurs natifs. Pour souligner son importance **hors du champ linguistique**, cette première section met en évidence les notions clés dans le cadre de la psychologie cognitive et sociale (1.1). Nous les ancrons dans un domaine plus particulier des sciences humaines, la linguistique qui a, elle aussi, développé de nombreuses acceptions de la (dis)fluence – *produite, perçue, interprétée* – en se nourrissant notamment de travaux de la **psycholinguistique**, de la **linguistique des apprenants** et du **traitement automatique du langage** (1.2). Nous verrons ainsi à travers les disciplines, en quoi ces marques peuvent se définir comme des signaux et des indices à **double facette** (1.3).

1.1 Fluence cognitive, sociale ou linguistique ?

Adoptons un instant une définition *intégrée* et extensive de la fluence comme « **une expérience subjective d'aisance avec laquelle l'on traite l'information** » (Alter & Oppenheimer 2007). Ce faisant, nous notons que la linguistique de l'oral n'a pas l'exclusivité des recherches sur la fluence : les sciences cognitives et sociales reconnaissent plusieurs significations de ce terme et de nombreux domaines d'applications.

Le concept de « fluence » a particulièrement nourri le domaine de la *psychologie cognitive*. Nombre d'études ont mis en relation l'expérience de fluence avec le fonctionnement de la mémoire (e.g. encodage et récupération d'informations, linguistiques et non linguistiques) ou de manière plus générale avec les processus cognitifs de haut niveau (i.e. déduction, raisonnement spatial, prédiction, attention et mémorisation). Le concept de fluence a notamment été mis à l'honneur par les **théories d'amorçage** (i.e. *conceptual priming*) reliées au domaine de la « *fluence conceptuelle* » : ces expériences soulignent par exemple

2. Brièvement, la prosodie est l'étude des phénomènes suprasegmentaux, se manifestant au-delà des segments phonétiques individuels, tels que le rythme et l'intonation (voir chapitre 1 section 3).

que la présence d'un *stimulus d'amorce* peut favoriser le traitement d'un autre *stimulus cible*, c'est-à-dire en quelque sorte sa fluence, ce non pas dans le seul cas de la parole, mais dans plusieurs modalités telles que sur la fluence perceptuelle de stimuli visuels linguistiques écrits et imagés (voir notamment Labroo, Dhar & Schwartz 2007).

L'impact de la fluence peut être décrit au niveau cognitif *individuel* : il s'agit d'évaluer la facilité de traitement de certains concepts par *un* individu. Les théories psychocognitives et sociales insistent sur le fait que tous les stimuli (conceptuels) sont **ancrés socialement** par ce même individu. C'est ainsi que le concept d'**incorporation** (*embodiment*) se développe au sein des théories de la **cognition incarnée** et donne naissance à la notion de *fluence incarnée*. Les psychologues observent par exemple l'impact positif de la posturo-mimo-gestualité et des réactions faciales lors du traitement de l'interaction par les locuteurs-auditeurs (Alter & Oppenheimer 2007). Petrova & Cialdini (2005) font le lien entre la *fluence incarnée* et la *fluence d'imagerie*. Ils soulignent l'interaction entre la fluence et son **ancrage contextuel et social** en observant que l'exposition prolongée à certaines situations entraîne chez l'individu une *fluence d'imagerie*, c'est-à-dire une capacité à évaluer une situation future comme plus plausible.

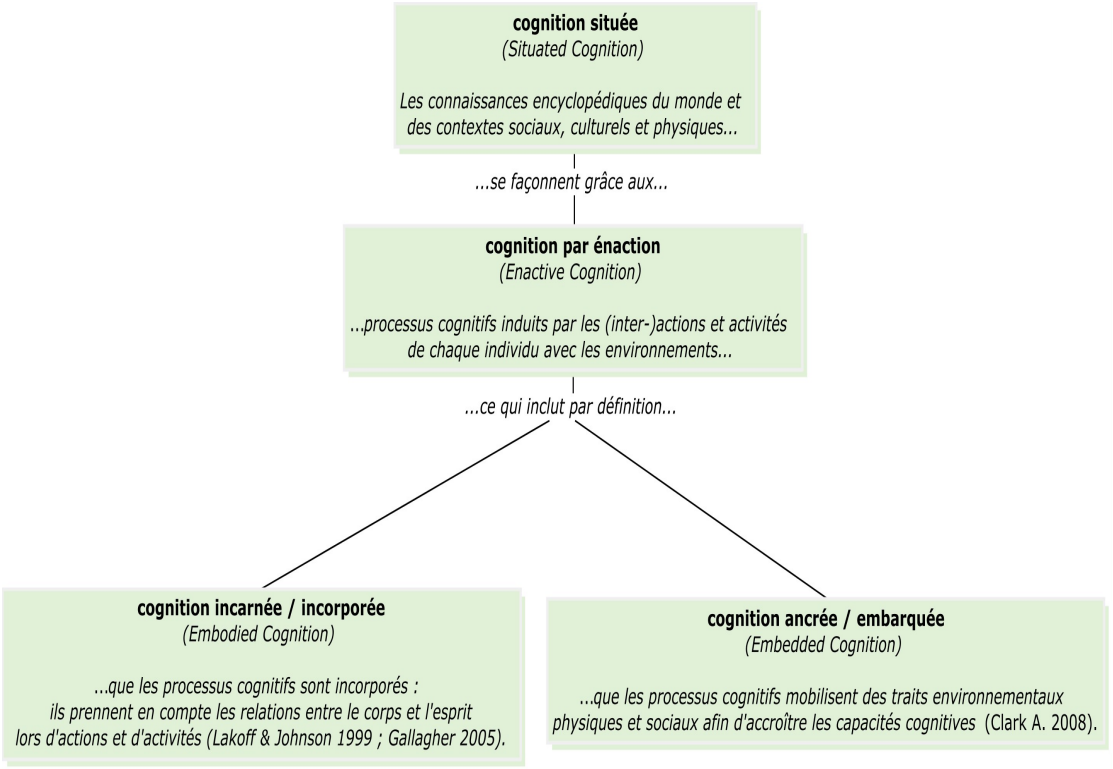


Figure 1.1 – Modèle de cognition située – inspiré de Malinin (2016:5).

Ces exemples ne constituent qu'un bref aperçu d'acceptions du mot *fluence*. Ils mettent en évidence que tout processus cognitif individuel est inclus dans un contexte social, culturel et physique. En d'autres termes, tout processus est **cognitivement situé** (Malinin 2016). La manière de percevoir et de traiter les informations « **en situation** » dépend de l'expérienciation de l'individu, autrement dit de son interaction avec l'environnement : on retrouve ici le concept de **cognition énactive**.

Le concept d'« **énaction** » peut se voir comme englobant deux systèmes d'interactions cognitives : (i) la **cognition incorporée**, tout processus cognitif naît du lien chez un individu entre son corps et son esprit et (ii) la **cognition ancrée** se référant au fait que les individus mobilisent leurs connaissances de la situation dans des environnements spécifiques (les « **traits situationnels** ») afin d'accroître leur capacité de traitements cognitifs. L'articulation de ces différents types d'interactions cognitives est illustrée en Figure 1.1.

Résumons. (1) Un processus cognitif, même envisagé au niveau de l'expérience individuelle, dépend du contexte social au sein duquel il s'actualise. (2) Il en va de même pour tout processus cognitif lié à la production ou à la perception (physique et interprétative) du langage. Langage, nécessairement *situé*. (3) Il en ressort des considérations théoriques, méthodologiques et épistémologiques : comment appréhender la contextualisation des « (dis)fluences », comment faire communiquer les disciplines, et peut-on accéder à un concept de « fluence » réellement interdisciplinaire, cohérent au sein de la linguistique et au travers des sciences humaines ?

Au centre de ces considérations, adopter une vision transdisciplinaire de la définition du contexte semble primordiale. Nous optons pour une définition des **contextes** en accord avec ce cadre théorique sociocognitiviste ébauché. Au niveau individuel, cette **cognition située** s'actualise et se modifie au fil des discours vécus (monologiques ou dialogiques) et ce de manière constante, i.e. pendant la production et la perception des discours par les locuteurs-auditeurs eux-mêmes. Cette actualisation et cette modification du contexte se font à travers ce que Schmid & Günther (2016) proposent d'appeler le « **Contexte Cognitif Immédiat** » (CCI), c'est-à-dire « l'état du système cognitif précédant immédiatement le traitement d'une information ».

C'est au sein de ce CCI que le locuteur-auditeur évalue une situation, un discours, un *input* linguistique. Les paramètres activés par ce CCI naissent de l'interaction entre quatre types d'information qui peuvent aussi être vus comme des contextes :

1. le **contexte linguistique** : ce qui a été dit avant ;
2. le **contexte situationnel** : participant(s), temps, endroits, cadre, objets ;
3. le **contexte social** : pratique communicative, rôles sociaux des participants ;
4. le **contexte cognitif général** : connaissances et routines générales et linguistiques encodées dans la mémoire à long terme.

Notons que le dernier paramètre, le contexte cognitif général, est fondamentalement différent des précédents. Il est basé sur la mémoire à long terme (« **contexte interne** ») et non sur le « **contexte externe** », c'est-à-dire la perception actuelle des évènements externes par les interlocuteurs. La Figure 1.2 schématise les différentes relations entre ces contextes.

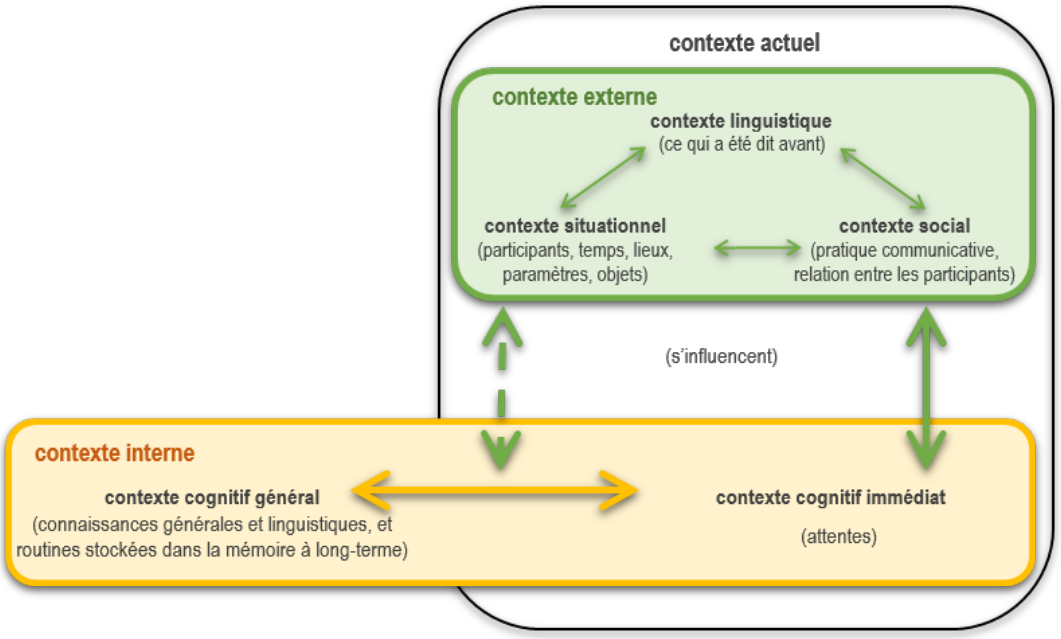


FIGURE 1.2 – Définition et interaction des types de contextes et de processus de formulations d'attentes – traduit de SCHMID & GÜNTHER (2016:3).

L'accès à cette cognition située passe par les interactions entre ces quatre types de contextes. Ce sont elles qui permettent aux locuteurs-auditeurs de formuler des **attentes**, au sein du **contexte cognitif immédiat** (SCHMID & GÜNTHER 2016:2). Les attentes naissent en effet de la comparaison du contexte cognitif immédiat et du contexte externe. Les diverses comparaisons de contextes appellent différentes sortes d'attentes. En les distinguant, on voit émerger trois différents types de fluences. Parmi elles :

- les **attentes linguistiques** liées à une évaluation de la **fluence phrastique interne** : *l'énoncé est-il bien constitué d'un tout cohérent et compréhensible, grammaticalement et syntaxiquement bien formé ?*
- les **attentes situationnelles** liées à l'évaluation de la **fluence linguistique située** : *la forme et le contenu du discours sont-ils compatibles, en adéquation, avec les connaissances liées au genre et les contraintes pragmatiques de la situation ?*
- les **attentes sociales**, liées à l'évaluation de **fluence sociale** ou **interpersonnelle** : *la forme et le contenu du discours sont-ils appropriés par rapport au rôle social attribuable aux locuteurs et/ou à leurs relations interpersonnelles ? Le locuteur prend-il en compte son interlocuteur et/ou parle-t-il avec une aisance relative ?*

Au chapitre 2 (section 1.4), nous repenserons aux relations entre le contexte et la formulation d'attentes d'une part, et aux concepts de saillance et de fluence d'autre part, naissant des interactions entre le contexte cognitif immédiat et les contextes internes (SCHMID & GÜNTHER 2016).

Revenons à l'objet de « fluence » dans un sens linguistique restreint cette fois, également présent en psychologique cognitive. En effet, à côté des fluences incarnée, imagée, conceptuelle examinées ci-avant, Alter & Oppenheimer (2009:225) énoncent deux acceptions supplémentaires de la fluence : la *fluence perceptuelle* et la *fluence linguistique* :

- La **fluence perceptuelle** a trait à la perception physique d'un stimulus (visuel ou auditif) et/ou à sa perception temporelle. Simplement dit, la fluence perceptuelle ***physique*** fait référence à la facilité que l'on a de percevoir un stimulus, sa « *saillance* » physique, tel qu'un contraste d'une forme par rapport à un fond, ou la perception plus aisée de certains lettrages et de tailles de caractères dans un journal (Reber & Schwartz 1999). La fluence perceptuelle ***temporelle*** renvoie elle à la durée ou au moment d'exposition à un stimulus : une durée plus longue ou la présentation de stimuli associés visuellement permet en général un traitement de l'information plus aisé (Winkielman & Cacioppo 2001).
- La **fluence linguistique** peut se scinder en de nombreux domaines selon les variables linguistiques envisagées. Les auteurs distinguent notamment la fluence phonologique (e.g. la facilité à traiter certaines suites phonétiques ou syllabiques), lexicale (e.g. l'aisance relative à traiter des mots simples /vs/ complexes) et syntaxique (e.g. le degré de complexité de traitement de suites de constructions grammaticales).

En résumé, quel que soit le domaine de recherche envisagé en psychologie cognitive et sociale, la fluence partage au moins deux propriétés principales : elle **facilite le traitement** d'un stimulus dans son environnement mais également l'**évaluation positive** que l'on en a. En effet, *in fine*, la psychologie cognitive a mis en évidence que l'évaluation fluente d'un stimulus est fortement associée par les sujets à sa **véracité**, le **degré de confiance** qu'on lui accorde ou l'**intelligence** qu'on lui prête (Alter & Oppenheimer 2007, 2009:229-231 ; Loy, Rohde & Corley 2016).

La notion de « disfluence » semble prolifique en ce qu'elle transcende de nombreux domaines des sciences humaines et sociales. Mais, comment ces différentes origines et effets de « fluence » ou de « disfluence » mis en valeur par la psychologie cognitive (*fluence conceptuelle, d'imagerie, physique, perceptuelle, linguistique, etc.*) sont-ils reliés à l'acception linguistique de la fluence ? Comment sont-ils envisagés en fonction des domaines linguistiques ? Se traduisent-ils également en différents types de « fluence » ou « disfluence » en fonction des domaines de la linguistique envisagés ?

1.2 Domaines de la linguistique ayant traité la notion de (dis)fluence

Si les concepts de fluence-disfluence se retrouvent dans plusieurs domaines des sciences humaines, celui de la linguistique ne manque pas à l'appel. Un élan théorique comme méthodologique à l'étude des « disfluences » de l'oral doit son impulsion à trois disciplines scientifiques distinctes qui adoptent parfois une acception différente de la (dis)fluence : *la (psycho)linguistique clinique*, *la linguistique des apprenants*, et *la linguistique computationnelle*.

1.2.1 Linguistique clinique : *dys*-fluence pathologique

La **psycholinguistique clinique** s'est attachée à décrire les hésitations dites « pathologiques » (e.g. bégaiements, aphasie, dyslexie, etc.). Depuis Wingate (1984) cité par Eklund (2004:158), on distingue au niveau terminologique les « *dys-fluences* », des hésitations « anormales » dues spécifiquement à l'état pathologique d'un sujet et les « *dis-fluences* » qui, par leur relation au concept abstrait de *fluence*, ne sont pas nécessairement motivées par des raisons pathologiques. Le discours pathologique a souvent été étudié au regard de celui tenu par des sujets non atteints de **troubles du langage**.

Ces études ont enrichi les connaissances sur les relations entre les **troubles cognitifs** et les troubles moteurs liés à la production et au traitement des (dis)fluences. Les (dis)fluences sont alors vues comme des dispositifs linguistiques utilisés pour la gestion temporelle du flux de parole, facilitant le traitement cognitif pour le locuteur comme pour l'auditeur (Bosshardt 2002 ; Byrd, Vallely, Anderson & Sussman 2012 ; Karniol 1995 ; Levelt 1989 ; Sasisekaran 2013). Point essentiel à notre développement : la (psycho)linguistique clinique reconnaît de ce fait la potentielle **fonction communicative fluente** des (dis)/(dys)fluences. Plus encore, à titre d'exemple dans le domaine, Johnson (1959) remarquait déjà que c'est bien la **réception négative** à une (dis)fluence qui entraîne son statut de **dispositif communicatif disfluent**. Par là même, l'auteur met en valeur le lien essentiel entre le contexte de perception et de production. Nous y reviendrons.

Ainsi, bien que la ***dysfluence*** (pathologique) ne soit pas l'objet premier de notre recherche, rien n'empêche de se référer ponctuellement à certains de ses apports précieux en matière de *dis*fluence. En réalité, la comparaison de productions orales par des sujets exempts et atteints de bégaiement a permis de comprendre les différences de fonctionnement des systèmes articulatoire et phonologique de ces deux populations et les éventuelles interférences attentionnelles de ces systèmes avec des processus du système exécutif central. Bosshardt (2002) le met notamment en évidence dans une expérience portant sur la répétition de mots et de non-mots.

1.2.2 Linguistique des apprenants : natifs /vs/ non-natifs

La **linguistique des apprenants** s'est, elle aussi, intéressée aux (dis)fluences, plus particulièrement dans le cadre de l'évaluation de la **compétence langagière** de locuteurs non natifs. A l'instar de la (psycho)linguistique clinique, cette évaluation a souvent été réalisée en comparaison au discours de « sujets normaux » : les locuteurs natifs (Bosker, Quené, Sanders & de Jong 2014 ; Denke 2005 ; Götz 2013b ; Osborne 2010). Sans insister sur le danger de poser le discours des locuteurs natifs comme toujours « fluent » (ou de marquer l'équivalence entre « fluence » et « absence de disfluence »), certaines de ces références majeures ont inspiré théoriquement comme méthodologiquement les recherches sur la fluence des locuteurs natifs.

Sans prétention d'exhaustivité, nous pensons spécialement à la contribution théorique de Segalowitz (2000, 2010, 2016) qui propose un cadre cognitif d'analyse de la fluence dans une approche tripartite synthétique. Segalowitz (2010:165) distingue la « ***fluence cognitive*** » qui s'attache à l'efficacité des énoncés ; la « ***fluence phrastique*** » qui s'intéresse aux propriétés de l'énoncé reflétant la fluence cognitive du locuteur et la « ***fluence perçue*** » qui tient à la formation d'inférences par l'auditeur à propos de la fluence cognitive du locuteur, basée sur leur perception de la fluence phrastique (Figure 1.3).

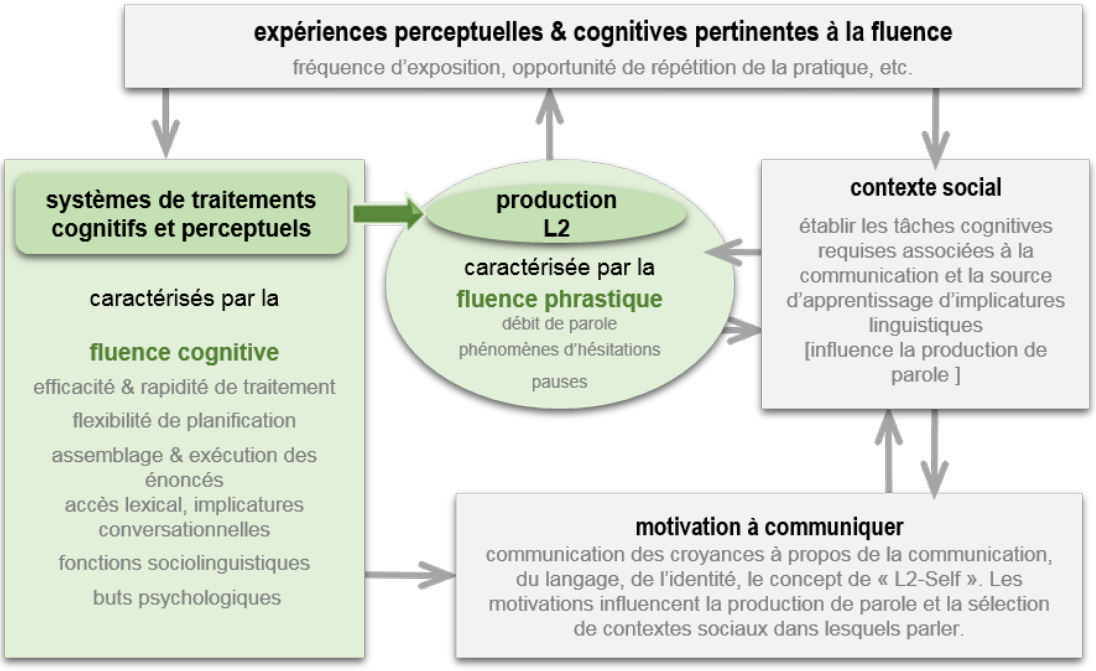


Figure 1.3 – Modèle contextualisé de tripartition de la fluence en langue non native – traduit de Segalowitz (2016:90).

Résumons. Quand un président « bafouille » lors de son allocution, le linguiste, annotateur acharné de tout segment discursif, s'attache traditionnellement à l'analyse de la *fluence phrastique* de l'énoncé. Le linguiste formule éventuellement des hypothèses sur la *fluence cognitive* que l'auditeur-évaluateur attribue à l'efficacité du discours. Mais, c'est ici *in fine* le journaliste-informateur qui fournit les indices les plus précieux de la *fluence*

perçue en partageant les inférences qu'il formule sur la fluence cognitive du locuteur. Dans notre démarche, cette **fluence perçue** permet de réévaluer de manière motivée et de pondérer l'importance toute relative de la *fluence phrastique*.

Par ailleurs, Segalowitz (2016:83-84) met en garde à l'égard de la distinction entre la **fluence cognitive *objectivable*** (mesurable objectivement par la fluence phrastique) et la **fluence cognitive *subjective*** correspondant à l'**évaluation de l'aisance** avec laquelle quelqu'un a réalisé une tâche (que ces fluences soient effectivement concordantes ou non, ce dernier cas s'actualise lors d'« illusions cognitives » : lorsque par exemple un stimulus semble familier, on se dit qu'il a dû être présenté récemment alors qu'en réalité il ne l'a pas nécessairement été) [3].

Dans le même ordre d'idées, Götz (2013b:9) souligne que l'évaluation de la **compétence orale** des apprenants germanophones de l'anglais [4] (*fluence phrastique*) n'est pas significativement associée à la perception de fluence par des locuteurs natifs anglophones (*fluence perçue*, mesurée par six variables : le caractère idiomatique de la parole, le registre, la diversité lexicale, la structure phrastique, l'accent et des traits pragmatiques).

Par ailleurs, le modèle théorique et la méthodologie développés par Götz (2011) dans le cadre de sa recherche doctorale offrent des propositions allant au-delà de la distinction de la fluence phrastique /vs/ perçue. Nous les présentons brièvement en mettant en valeur leurs apports notables pour le domaine de la recherche sur la (dis)fluence mais en pointant également les écueils éventuels.

L'auteure appelle « **fluencèmes** » (*fluencemes*) « des traits ou marques linguistiques abstraites de description de la fluence, permettant la construction d'une **distinction abstraite** et **idéalisée** des variables de fluences entre la fluence productive, perceptive et la fluence non verbale » [5]. A juste titre, l'auteure distingue deux types d'approche de la fluence : le degré de fluence établi sur base **du locuteur** (certaines marques permettent au locuteur d'augmenter son aisance à produire la parole) et le degré de fluence basé **sur l'auditeur** (certaines marques permettent d'augmenter pour l'auditeur la perception de fluence du discours du locuteur).

Götz (2011:42) oppose ainsi les **variables de production** de la fluence aux **variables de perception**. Elle associe par ailleurs un troisième élément représentant les **variables non verbales**. Ces types de fluence sont présentés comme s'actualisant par des variables différentes et **mutuellement exclusives** en Figure 1.4. Pourtant, l'auteure reconnaît

3. Cette démarche rappelle l'approche de Duez (1993) qui distingue la *pause subjective* de la *pause objective* (voir chapitre 2 section 2.1).

4. Götz (2013b) s'appuie sur la comparaison des productions de 50 locuteurs natifs /vs/ non natifs issus des corpus oraux LINDSEI-GE (*Louvain International Database of Spoken English Interlanguage*) et LOCNEC (*Louvain Corpus of Native English Conversation*).

5. Texte original : " The classification of features of speech into these three categories allows for an abstract description of the linguistic substance of fluency in speech. I label these eatures fluencemes, as they build the basis for the somewhat idealized and abstract distinction between productive, perceptive and non-verbal fluency variables, although it has become evident that many of the variables are interconnected and not easy to separate." (Götz 2011:41-42). Notre traduction. Nous surlignons.

elle-même qu'en réalité celles-ci se chevauchent – au moins partiellement. Bien que distinguer les deux types d'effets de (dis)fluence, en production et en perception, nous semble pertinent, nous désirons notamment remettre en cause le fait que les variables étudiées soient présentées, même au niveau théorique, comme mutuellement exclusives en fonction du domaine de fluence envisagé (e.g. une variable produite n'est-elle pas la plupart du temps perçue ?).

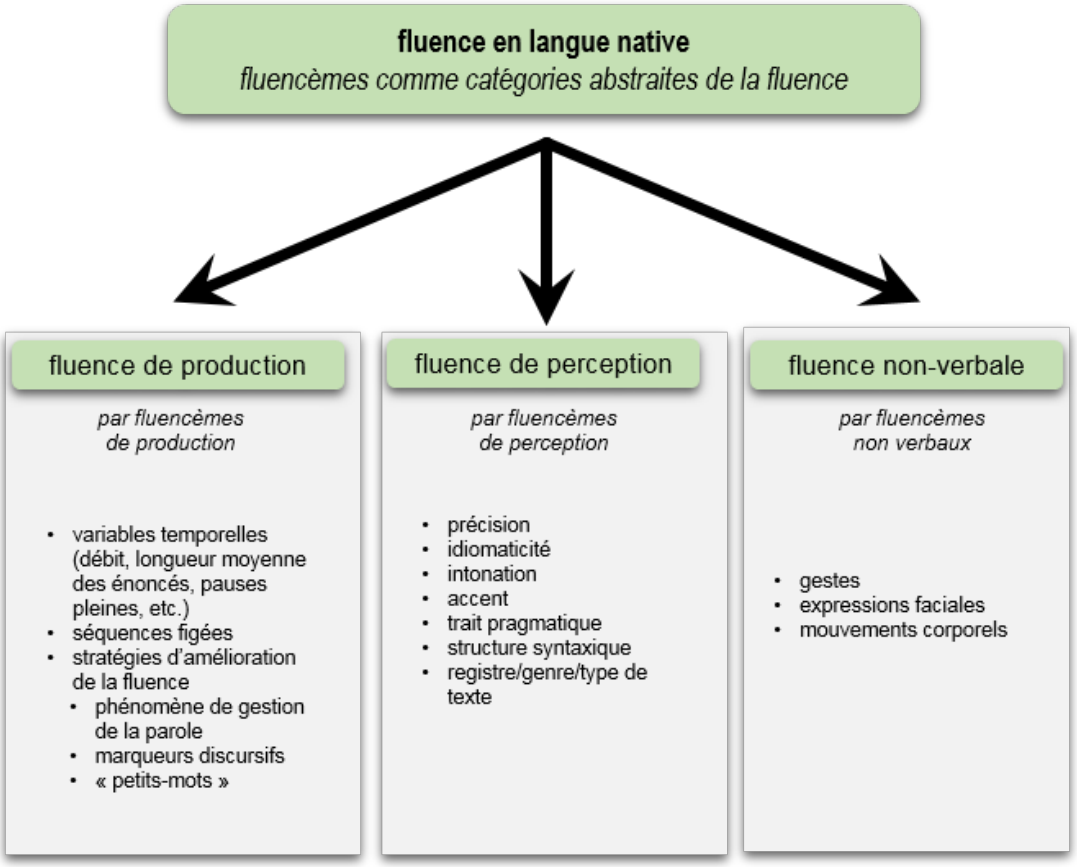


FIGURE 1.4 – Modèle non contextualisé de tripartition de la fluence en langue native – traduit de GÖTZ (2011:42).

Par exemple, selon ce modèle la « diversité lexicale » participerait à la perception de la fluence chez l'auditeur, mais ne serait pas présente en production comme un outil de fluence pour le locuteur. Pourtant, réduire la **charge cognitive** que demande une situation réduit également la complexité (mesurée ou perçue) de l'*output* discursif. En choisissant un *input* lexical moins diversifié, les locuteurs diminuent le nombre de « disfluences » produites (BOURAOUI 2008). A titre d'exemple additionnel, la production de ***marqueurs discursifs***[6] se retrouve comme une stratégie d'augmentation de la fluence en production, mais ne participerait pas, dans ce modèle, à l'évaluation d'une fluence

6. Les marqueurs discursifs sont des items lexicaux pragmaticalisés que l'on peut définir au sens large comme des indices métadiscursifs syntactiquement optionnels contraignant l'interprétation du discours et mettant en évidence les relations discursives, les structures de topicalisation ou les stratégies interpersonnelles des locuteurs (SCHIFFRIN 1987 ; SCHOURUP 1999).

accrue pour l'auditeur. Pourtant, la recherche témoigne à quel point les marqueurs discursifs peuvent **participer à la perception de fluence** par l'auditeur : ils segmentent le discours, facilitent la formulation d'inférences discursives, participent à la gestion interactionnelle, etc. (Crible 2017 ; Degand 2016 ; Degand & van Bergen 2016 ; Fischer 2000 ; Maschler & Schiffrin 2015 ; Zufferey & Degand 2013). Enfin, de manière plus triviale, les **signaux non verbaux** (i.e. gestes, expressions faciales, langage corporel, etc.) sont produits par les locuteurs *et* perçus par les auditeurs, ne devraient-ils pas alors s'insérer au sein des deux catégories précédentes ?

Ainsi, on reconnaît que ces deux dernières propositions, bien que rédigées au départ dans le cadre de recherches en linguistique des apprenants, ont des applications pour la recherche sur les productions et perceptions des **locuteurs natifs**. Les travaux de Götz (2011, 2013a) offrent une vision très pertinente de la littérature sur les (dis)fluences. Bien que les analyses proposées témoignent d'une richesse indéniable de par la quantité de variables envisagées, la distinction entre variables de production /vs/ de perception nous semble quant à elle trompeuse. La présence du néologisme *fluencème*, adopté comme une unité abstraite, en insistant plus sur l'aspect fluent que *dis*fluent, semble à première vue favoriser la mise en évidence des marqueurs étudiés comme des dispositifs communicatifs fluents. Mais, ce terme ne fait peut-être que **renverser le problème initial sans le résoudre**. En effet, si le terme « disfluence » insiste bien trop sur l'aspect négatif de certains marqueurs, le terme « fluencème » ne présenterait-il pas le désavantage réciproque en insistant a priori sur l'aspect fluent de ceux-ci ?

1.2.3 Linguistique computationnelle : erreurs de performance

La **linguistique computationnelle** sera historiquement le deuxième domaine de développement des (dis)fluences. Il s'est attaché à la détection des phénomènes dits « agrammaticaux », en dehors du système linguistique tels que les (dis)fluences. La *dis*fluence était alors définie comme une « **erreur de production** » (*performance error*), ou comme tout dérangement par rapport à la « prononciation idéale de la parole ». Pour atteindre cet idéal, les *dis*fluences représentent alors des cas où un « segment continu de matériel linguistique doit être supprimé afin de parvenir à la séquence initiale que le locuteur avait l'intention de produire, le plus probablement, prononcée sans répétition »[7] (Shriberg 1994:1). Pour ce domaine linguistique, comprendre leur fonctionnement favorise l'évaluation automatique de la **compétence langagière** (chez les locuteurs natifs comme non-natifs) (Bouraoui 2008 ; Cucchiarini, Strik & Boves 2000 ; Zechner & Bejar 2006), tandis que leur modélisation permet avant tout leur intégration aux systèmes de segmentation, d'annotation, de synthèse ou de reconnaissance de la parole.

7. Texte original : "cases in which a contiguous stretch of linguistic material must be deleted to arrive at the sequence the speaker 'intended', likely the one that would be uttered upon a request for repetition" (Shriberg 1994:1). Notre traduction. Nous surlignons.

Bien que l'intention de nombre des approches computationnelles ait été initialement de gommer et/ou remplacer ces parties à priori « disfluentes » du discours (chez Dowding et al. (1994) par exemple), et que ceci semble aller à l'encontre d'une approche reconnaissant l'**aspect bénéfique des (dis)fluences**, ces recherches ont l'avantage d'avoir **formalisé l'analyse** de ces marques et d'avoir facilité leur extraction automatique (Besser & Alexandersson 2007 ; Johnson & Charniak 2004 ; Nakatani & Hirschberg 1994 ; Shriberg 1994 ; Strassel 2004). Au niveau morphosyntaxique, on a identifié par exemple que les signes formels d'**achoppements du discours**[8] tels que les répétitions ont tendance à s'actualiser sur des **mots grammaticaux** monosyllabiques en frontière d'unité syntaxique (Duez 1991:106 ; Candea 2000a:30,99)[9]. Au niveau acoustique, les avancées de la linguistique computationnelle permettent d'identifier les régularités d'actualisations acoustiques (e.g. fréquence fondamentale, formants) des *euh* d'hésitations /vs/ des *euh* dits grammaticaux (Shriberg 1999:621 ; Shriberg, Stolcke & Baron 2001:165).

Toutefois, ce type de détection, basé uniquement sur des éléments lexicaux, est mis à mal lorsque la reconnaissance des mots est elle-même incertaine (Shriberg, Bates & Stolcke 1997:2384). Les variétés des méthodologies de détection et leur combinaison (e.g. approches par *parsing* syntaxique, par reconnaissance de patrons réguliers, par différentes techniques d'apprentissage automatique, etc.), ont favorisé l'émergence des **caractéristiques internes** aux (dis)fluences mais aussi à leur **contexte d'actualisation** syntaxique, prosodique et situationnel. Cette avancée technologique et méthodologique a permis une meilleure détection des phénomènes et de dépasser une vision des « disfluences comme des *erreurs de production* ». Exit l'objectif de gommage des (dis)fluences, lesquelles accèdent enfin au statut de **potentiels dispositifs communicatifs fluents** au sein des systèmes de synthèse et de reconnaissance de la parole, afin notamment de la rendre **plus naturelle** (Adell, Bonafonte & Escudero-Mancebo 2008).

1.2.4 De la chaîne au schéma

La linguistique computationnelle a révélé que si chaque (dis)fluence **s'accompagne de spécificités** acoustiques et distributionnelles[10], des **patrons récurrents** à la structure interne de toutes les (dis)fluences émergent également. A la suite des recherches de Levelt (1983, 1989), portant plus généralement sur la production et le traitement de la parole (section 2.3), Nakatani & Hirschberg (1993:47-48) sont les premières à proposer une **modélisation de la structure interne des (dis)fluences**. En respectant la terminologie établie dans le *Repair Interval Model* (RIM), elles identifient trois parties structurantes d'une zone de (dis)fluence (Figure 1.5). Le ***reparandum*** représente

8. Voir section 2.2 de ce chapitre.

9. Les mots grammaticaux, associés également aux « mots vides » ou « mots clitiques », s'opposent aux mots lexicaux (ou mots pleins, non-clitiques). Ils reprennent des catégories grammaticales en général non accentuables en français, telles que les prépositions et les déterminants. Voir notamment chapitre 2 section 2.2.3.

10. Voir chapitre 2 section 2.

l' « objet disfluent », la partie du segment qui va être répétée, éditée ou supprimée. L'astérisque « * » entre le *reparandum* et l'*interregnum* symbolise le point d'interruption (de la fluence). L'***interregnum*** est la phase entre le *reparandum* et le ***reparans*** ; elle inclut optionnellement une phase d'édition signalant la correction (composée de pauses, termes d'éditions et/ou autre(s) (dis)fluence(s)). La zone de retour à la fluence est appelée *reparans* (Levelt 1983:44-45) [11], de sorte que, en théorie, la suppression du *reparandum* et de l'*interregnum* donne lieu à une production « lexicalement fluente » (Shriberg, Stolcke & Baron 2001:161).

Un des prolongements directs de cette formalisation lie la structuration interne de ces marques avec les processus de traitement de la parole humaine. La littérature formule l'hypothèse selon laquelle les auditeurs bénéficient d'un **signal** « **acoustique** » qui leur permet de distinguer entre le contenu fluent et le contenu disfluent dans un discours. Certains auteurs ont cherché ce signal de manière ponctuelle autour du **point d'interruption** de la fluence. Il peut être actualisé par la présence d'une pause, d'amorces de mot(s), d'un signal d'édition au sein de l'*interregnum* ou par **des phénomènes non lexicaux**, acoustico-prosodiques tels que le ralentissement du débit, la perturbation de la fréquence fondamentale et le marquage mélodique (Hindle 1983:128 ; Bear, Dowding & Shriberg 1992 ; Nakatani & Hirschberg 1993, 1994 ; Shriberg, Bates & Stolcke 1997).

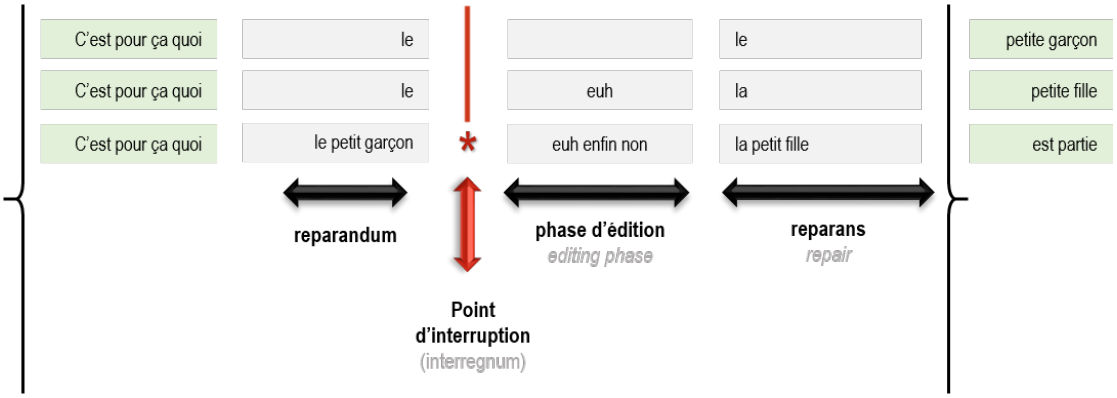


Figure 1.5 – Schéma de structure interne d'une (dis)fluence – inspiré de Nakatani & Hirschberg (1993).

Ces modifications acoustico-prosodiques seraient ainsi plus particulièrement marquées dans la **phase d'édition** notamment par la présence de pauses pleines et de termes explicites d'édition (e.g. « (je suis) (euh) (non pardon excusez-moi) (je veux dire) (je serai) [...] »). Ces recherches soulignent à quel point, au niveau des unités intonatives qui entourent la (dis)fluence, comme des segments internes à la structure d'une (dis)fluence, **la prosodie** est un indice précieux à la **détection**, la **contextualisation** et

11. Nous utilisons le terme de *reparans* au lieu de *repair* étant donné la confusion possible dans la littérature entre repair comme synonyme de *reparans* et de *repair* comme synonyme d'« (auto)-correction » équivalent de la « substitution/reformulation » en français (Eklund 2004:164).

l'**interprétation** de ces marques (Lickley & Bard 1998:205 ; Liu et al. 2006:1528 ; Nakatani & Hirschberg 1994:1620-1621 ; Shriberg, Bates & Stolcke 1997 ; Moniz, Trancoso & Mata 2009) [12].

Appelons « **chaîne de (dis)fluences** » toute combinaison de ces marques, sans limite minimale ou maximale de longueur. Toute chaîne de (dis)fluence(s) est un « schéma potentiel ». Le « **schéma potentiel** » se définit comme (1) une **chaîne de (dis)-fluence(s)** segmentable structurellement en fonction de plusieurs niveaux de granularité (e.g. [+/- terme d'édition], [+/- nombre de répété], [type de répété]), (2) une *combinaison de contextes* linguistiques et non linguistiques au sein desquels les chaînes de (dis)-fluence(s) s'actualisent et deviennent des *schémas* (e.g. régularités structurelles de la chaîne accompagnée de ses contextes prosodiques, syntaxiques et traits situationnels).

Pour qu'une unité de la langue soit considérée comme « potentiellement schématique », elle doit s'actualiser comme une **structure régulière** et récurrente lui permettant d'être représentée comme « une », ou au sein d'une catégorie abstraite singulière chez les locuteurs-auditeurs. Si ce processus d'association est dans un premier temps guidé par l'observation des productions de marques dans les corpus, la méthodologie soutient alors une approche cognitiviste-fréquentiste, où le corpus permet à lui seul de valider la présence des **représentations mentales prototypiques** ou **schémas** chez les locuteurs-auditeurs [13].

Un des résultats de cette méthode de regroupement permet d'augmenter le **degré d'abstraction** des chaînes : de regrouper des phénomènes en apparence différents mais aux propriétés internes schématiques communes. Shriberg (1994) regroupe ainsi 8 types différents de chaînes sous l'étiquette « REPétition », en fonction des éléments qu'elles intègrent. A titre d'exemple, des chaînes comme (i) *« la/ la/ euh l'annonce du président »* et (ii) « *ou/ euh ou/ oublie* » suivent toutes les deux la structure « répétition + pause pleine », elles comprennent une « **(dis)fluence composée** » et peuvent se décrire comme des « répétitions non isolées », dépendant du niveau d'abstraction reconnu et de la **granularité** nécessaire d'une analyse [14]. La notion de schéma favorise l'association des chaînes à priori différentes.

Cette même notion de « schéma » se retrouve dans des théories en production comme en perception. Le type de schéma défini dépend de la granularité de l'observation et du type de méthode utilisé pour extraire les interprétations. Ces interprétations sont dépendantes notamment de la **place donnée au contexte** et à la **relation locuteur-auditeur** : un schéma produit est-il un schéma perçu ? Les schémas sont alors potentiellement interprétables comme des signaux ou indices bénéfiques en contexte.

12. Voir chapitre 2 section 1.
13. Voir chapitre 2 section 1.4.1.
14. Les macro-labels et les regroupements abstraits des chaînes potentiellement reconnues comme des schémas sont présentés en détail chapitre 3 section 2.2.

1.3 (Dis)fluences : signaux et indices à double facette

En résumé, nous avons vu que les théories computationnelles classiques de reconnaissance et de synthèse automatique de la parole ont historiquement été les premières à s'intéresser à la reconnaissance et à l'annotation de (dis)fluences pour produire des systèmes exempts « d'erreurs ». Par la suite, la psycho- et la neurolinguistique y ont vu des **témoins aux processus sous-jacents** (et troubles) de production de la parole. Depuis une vingtaine d'années, un revirement de perspective a amené à considérer les **aspects bénéfiques** des (dis)fluences.

Ainsi, l'utilisation des (dis)fluences peut partiellement naître du besoin de **réajustement entre le produit et l'interprétable** ou, en d'autres termes entre le discours effectivement produit et les **attentes** du locuteur et/ou de l'auditeur de ce même discours (d'une situation de parole et des interlocuteurs) [15].

Les (dis)fluences sont susceptibles d'être des **indices à double facette**. Soulignons à ce stade que les locuteurs peuvent être conscients des intuitions et des inférences pragmatiques formulées par les auditeurs sur leurs discours (ces locuteurs sont aussi auditeurs !). Ces locuteurs peuvent par conséquent vouloir **contrôler** les (dis)fluences produites. Comme les nombreux exemples ci-avant et notre extrait présidentiel l'illustrent, si tout traitement de l'information peut en soi être placé sur une échelle de fluence-disfluence, il semble essentiel que pour mettre en relation les (dis)fluence(s) linguistique(s) *produites, perçues et interpretées*, ces échelles soient pluridimensionnelles, envisageant la mise en relation des différents **types de fluences** (e.g. fluence cognitive, incarnée, linguistique, perceptuelle etc.) (King, Loy & Corley 2017) et de leur(s) contexte(s) d'actualisation. Ainsi, si les (dis)fluences sont parfois des **symptômes** d'un problème de **linéarisation du discours** chez le locuteur, elles peuvent aussi constituer des **signaux** d'aide non négligeables dans la **résolution** de ces problèmes (Clark & Tree 2002:74-78 ; Auer 2009:11 ; Bouraoui 2008:26) [16].

Une grande partie de la littérature témoigne encore davantage de l'**aspect bénéfique** de certaines (dis)fluences linguistiques, vues tout à la fois comme des indices et des signaux. Quelques exemples plus spécifiques sont donnés ci-après.

Les (dis)fluences peuvent faciliter le traitement des *anaphores* et des *chaînes de coréférences* [17] (Arnold, Losongco, Wasow & Ginstrom 2000:51) ; elles déclenchent ainsi des *implicatures conversationnelles* [18] (e.g. telle que l'anticipation d'un contenu nouveau

15. Voir section 2.2 de ce chapitre.

16. Selon la terminologie de (Léon 1993:6), un **indice** est par définition involontaire et le plus souvent inconscient. A côté de ceux-ci nous percevons des **signaux**, volontaires et plus ou moins conscients. Les signaux sont utilisés dans la parole pour influencer les interlocuteurs.

17. Traditionnellement, on regroupe les expressions référentielles d'un texte selon la référence à une même entité du monde extralinguistique. Les éléments regroupés appartiennent à une *chaîne de coréférence*.

18. Brièvement, en pragmatique gricéenne, une *implicature conversationnelle* se réfère à ce qui est suggéré par un locuteur de manière implicite (et non explicitement par une propriété sémantique de l'énoncé). Une *implicature* se distingue d'une *implication*, qui est annulable sémantiquement si l'on

ou complexe (Arnold & Tanenhaus 2011:213 ; Watanabe, Hirose, Den & Minematsu 2008:91)) et partagent des caractéristiques avec les phénomènes de coordination lors de la création de listes (Guénot 2005:180 ; Auer & Pfänder 2007:60). Elles permettent également d'**attirer l'attention** d'un auditeur sur un mot cible, et favoriseraient dès lors sa reconnaissance rapide (Fox Tree 1995). Enfin, les (cooccurrences de) (dis)fluences faciliteraient l'identification de **frontières discursives** majeures pour l'auditeur (Liu et al. 2006 ; Rendle-Short 2004).

Les (dis)fluences sont aussi donc parfois **marqueurs de frontières**, de focus, pour le locuteur mais aussi pour les interlocuteurs. En effet, sur base d'un corpus d'interviews politiques, Boula de Mareüil et al. (2005) et Adda et al. (2007) ont également montré que les (dis)fluences participent à la **coordination de l'interaction** dans le cadre de l'alternance des tours, appelée également « **syntaxe dialogique** » (Clark 2002 ; Clark & Tree 2002 ; Moniz, Trancoso & Mata 2009). Ainsi, les (dis)fluences sont **des stratégies d'édition en ligne**, inhérentes au traitement de la parole en production et en perception auxquelles on associe de nombreuses **fonctions méta ou paralinguistiques**. L'existence de relatives régularités interlangues en perception et/ou de production (Clark & Tree 2002:92-93 ; Lai, Gorman, Yuan & Liberman 2007:2347) et leur participation à des processus sémantico-pragmatiques sont autant d'éléments en faveur de leur reconnaissance comme système intégrant de la langue, comme partie de sa « grammaire », du système linguistique conventionnalisé, concept qui n'a pas toujours recueilli un large assentiment.

Parfois cependant, certains auditeurs traitent négativement certaines informations en fonction de stimuli purement linguistiques. Sereno & Hawkins (1967) ont mis en évidence que la production de *dis*fluences (pauses pleines, répétitions et erreurs d'articulations) influence négativement le **jugement des locuteurs** sur l'évaluation de la *compétence des locuteurs* ainsi que de leur dynamisme, mais les auteurs n'ont pas trouvé d'association entre les (dis)fluences et l'évaluation du sujet discuté [19]. Toutefois, toujours par rapport à l'évaluation du locuteur et de son rapport au discours, les résultats des expériences de McCroskey & Mehrley (1969) indiquent que la présence de *dis*fluences conduit à une évaluation **moins convaincante** et **moins crédible** du discours des locuteurs.

Si toutes les productions de ces marqueurs peuvent être interprétées de manière univoque comme des signaux ou indices de la **malhonnêteté d'un propos** et/ou d'un individu en production et en perception (Akehurst, Kohnken, Vrij & Bull 1996 ; Arciuli, Mallard & Villar 2010 ; DePaulo et al. 2003 ; Vrij & Winkel 1991), il semble que les marques peuvent être justifiées et influencer les locuteurs à un **jugement favorable** (e.g. un humoriste prolonge sa fin de phrase car les rires de son public risquent de masquer la chute ou le début de la phrase suivante). Ainsi, la présence de marques **modifie les interprétations** des auditeurs et leur permet de déclencher plusieurs types d'**inférences**

ajoute une proposition, ou d'une *présupposition implicite* de nature sémantique. Enfin, une *implicature conversationnelle* dépend du contexte de la conversation, elle se distingue d'une *implicature conventionnelle* dépendante de l'énoncé lui-même.

19. Les participants étaient amenés à exprimer leurs attitudes par rapport aux musulmans.

pragmatiques (Corley 2017) et donc de nouvelles **attentes** de l'auditeur vis-à-vis de l'interlocuteur et/ou de la situation (Allopenna, Magnuson & Tanenhaus 1998 ; Loy, Rohde & Corley 2016).

En conclusion, ce bref aperçu des différents courants scientifiques s'intéressant à la fluence révèle les nombreuses acceptions du terme et des concepts clés associés. Nous considérerons ponctuellement les théories, les méthodologies et leurs conclusions. Ce travail de recherche doctorale se focalise sur **une définition de la (dis)fluence *produite, perçue* et *interprétée* de discours *non pathologiques* produits par des *locuteurs francophones natifs*.** En effet, lorsque la (dis)fluence n'est ni définie par un trouble pathologique, ni par un déficit langagier d'apprenant, comment peut-on la modéliser (détecter, définir, prédire et interpréter) au-delà d'une **expérience subjective d'aisance** avec laquelle le locuteur et/ou l'auditeur traite l'information ? Peut-on unifier la variété de définitions et d'actualisations des types de fluences-disfluences évoqués (sociales, linguistiques, cognitives – produites et perçues – de manière située) observées d'un point de vue pathologique, cognitif, psychologique, discursif ou computationnel, et si oui, comment ?

La (dis)fluence comme indice ou signal, pour le locuteur et/ou pour l'auditeur ? Dans toutes les situations et dans tous les contextes ? Qu'en est-il des divergences d'interprétations ? Avant de répondre à ces interrogations, plusieurs **questionnements méthodologiques** s'imposent : comment allier la **description formelle** de la fluence phrastique, de la fluence cognitive (objectivable ou subjective) et de la fluence perceptive ? Selon nous, en vertu de la palette d'interprétations possibles de ces marques, une catégorisation à priori binaire et univoque n'aboutira à aucun cadre théorique et méthodologique fidèle à une modélisation contextuelle, réaliste et réalisable, des (dis)fluences. Pour cela, il faut (1) se positionner au sein d'un type de **modèle de traitement de la parole** (orienté en production sur le locuteur et/ou en perception sur l'auditeur) et (2) prendre conscience des différents **processus cognitifs** (parallèles et/ou en séquentiels) que ce traitement implique lors de la production et de la réception de discours par les locuteurs-auditeurs, **dans une situation spécifique**. C'est ce que nous tâchons de réaliser dans la section ci-après.

2 Pluridimensionnalité du continuum fluence-disfluence

L'exploration des (dis)fluences en langue maternelle ou seconde s'étend donc sur un continuum allant de la psycholinguistique [20], qui voit en elles une réflexion des processus cognitifs sous-jacents au traitement de la parole (Chafe 1980 ; Goldman-Eisler 1972), à des études centrées sur la production de la parole, et plus particulièrement sur ses **dimensions temporelles** dont les (dis)fluences font partie (Goldman-Eisler 1968 ; Henderson, Goldman-Eisler & Skarbek 1966 ; O'Connell & Kowal 2008), en passant par des études de sociolinguistique appliquée, prenant en compte l'influence en production et en perception du **contexte communicatif** et de diverses variables (le participant, le sujet, la tâche, etc.) (Schegloff 1998 ; Schegloff, Jefferson & Sacks 1977). Ces recherches s'accompagnent de postulats sous-jacents aux modèles du traitement envisagé de la parole.

Comme nous l'avons vu, l'analyse de la (dis)fluence peut être orientée davantage sur la production du locuteur ou la perception des auditeurs. Ce choix dépend en partie du modèle de traitement envisagé de la parole (2.1). Les modèles accordent de ce fait une importance relative différente aux processus de **conscience**, d'**attention** et de ***monitoring*** auxquels les locuteurs s'adonnent lors de l'interaction discursive (2.3). Notre démarche propose une **approche mixte** : notre recherche se nourrit des études *phonostylistiques* [21] qui permettent de dégager des **régularités acoustiques** de production des (dis)fluences et d'éventuellement servir de point de départ à la formulation d'attentes et d'interprétations perceptives des locuteurs-auditeurs liées à ces régularités (3.3).

2.1 Impression globale des locuteurs et/ou des auditeurs ?

> "Is fluency, like beauty, in the eyes (and ears) of the beholder ?"
>
> (Freed 2000:243)

L'application méthodologique de la vision théorique tripartite de la fluence [22] n'est pas univoque. La fluence a d'ailleurs souvent été décrite comme une notion incluant une **série de concepts** clés plutôt qu'un tout formé de composantes bien définies.

Lorsque Lennon (2000:25) décrit la fluence comme une manière « rapide, lisse, précise, lucide et efficace de transmettre une pensée ou une **intention communicative** dans une langue sous contraintes temporelles du traitement en ligne du flux discursif » [23], il mêle

20. Les études fondatrices appartiennent au domaine de la psycholinguistique et portent sur l'anglais parlé. On les doit à Maclay & Osgood (1959) et à Goldman-Eisler (1958, 1961, 1968, 1972). Un compte-rendu détaillé de ces travaux est fait par Duez (1991).

21. Brièvement, les phonostyles « sont des styles sonores "tels qu'ils sont perçus en tant que caractéristiques d'un individu (jeune, vieux, homme, femme), d'un groupe social (prolétaire, bourgeois), ou d'une circonstance particulière (discours politique, sermon, etc.)." » (Léon 1993:3 ; cité par Simon, Auchlin, Avanzi & Goldman 2010).

22. Voir le modèle de Segalowitz (2016) présenté en section 1.2 de ce chapitre.

23. Texte original : "A rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal conditions of on-line processing". (Lennon 2000:25). Notre traduction. Nous surlignons.

de manière holistique des questions de fluences phrastique, cognitive et perceptive. En somme, la fluence fait ici référence à **la manière « effective » d'utiliser le langage** au regard de tous les modules de la « grammaire discursive », y compris du bon usage des (dis)fluences (Ginzburg, Fernández & Schlangen 2014:9).

Lennon (2000) fonde ses recherches sur un corpus de quatre étudiantes allemandes dans une université anglaise pendant 6 mois. Sur base de 2 enregistrements par étudiante (3 semaines /vs/ 6 mois après leur arrivée), l'auteur compare 12 **paramètres linguistiques** : le nombre de mots par minute, le nombre de mots par minute en excluant les mots répétés et les autocorrections, les pourcentages de répétitions, d'autocorrections, de pauses pleines par rapport aux unités syntaxiques minimales, le pourcentage de mots répétés et autocorrigés par rapport au nombre de mots prononcés, les pourcentages de pauses silencieuses /vs/ de pauses remplies par rapport au temps de parole, la longueur moyenne des unités interpausales, le pourcentage d'unités syntaxiques maximales suivies d'une pause (silencieuse ou remplie), ainsi que le pourcentage de temps total de pause et leur durée moyenne en frontière d'unité syntaxique maximale. Cette approche, quoique **relativement prescriptiviste** vis-à-vis de l'interprétation de la production et de la perception des marques du travail de formulation, a l'avantage d'étudier le **profil individuel** de chaque locuteur.

C'est d'ailleurs cette **vision holistique** de la (dis)fluence qui a caractérisé les premières recherches dans le domaine. Ces approches explorent les concepts centraux autour de l'**impression globale** de fluence davantage axée sur l'auditeur, avec peu de prétention d'investigations des composantes structurelles de la fluence. Bien que peu utilisées dans notre travail de recherche, elles ont permis d'associer l'impression globale de fluence avec les concepts d'**efficacité** et d'**automaticité**. La prise en compte des attributs des individus (locuteurs-auditeurs) et de la situation est devenue centrale au vu de cette définition. On retiendra également, sans prétention d'exhaustivité : l'approche de Crystal (3ème éd. 2010 ; 1988:421) qui décrit la **performance langagière** fluente du côté du locuteur comme un usage « sans effort »[24]. Cette dernière est fortement reliée à celle de Schmidt (1992:358-359) qui entend par fluence « l'automatisation des compétences langagières procédurales ».

Fillmore (1ère éd. 1979 ; 2000:51) était cependant parvenu à préciser le concept de fluence autour de quatre dimensions satellites. Il distingue la dimension **temporelle** (i.e. le flux continu de parole), la dimension **sémantique** et **syntaxique** (i.e. la cohérence et la logique des propositions), la dimension **sociopragmatique** (i.e. le caractère à propos d'une production en fonction du contexte de communication) et la dimension

24. De nombreux exemples dans la littérature scientifique adoptent cette description holistique de la fluence-disfluence. Par souci de concision, renvoyons le lecteur à « Overview : Varying Perspectives on Fluency », excellent résumé historique des approches holistiques de la fluence par Koponen & Riggenbach (2000).

créative (i.e. les usages métaphoriques du langage)[25]. Le locuteur « fluent », théorique dans le cadre de ces études non empiriques, maîtriserait parfaitement ces quatre dimensions.

Nombre de travaux en psycholinguistique se sont focalisés sur les deux premières dimensions (temporelles et sémantico-syntaxiques). S'ils apportent une rigueur méthodologique à la description de la (dis)fluence à partir de variables linguistiques quantifiables telles que le calcul du ratio type/token, la proportion de « phrases incomplètes », le nombre de mots ou morphèmes par minute (Fillmore 2000:58-59), ces études sont également confrontées à une limite. Elles omettent souvent des principes d'**interactions fortes** entre les **processus d'écoute et de parole**, et se révèlent alors incapables d'analyser **le rôle de la situation et du contexte**, comme **pratique communicative** agissante sur le discours.

Par là même, l'approche holistique omet souvent la nature pragmatique et sociolinguistique de la communication. Elle minimise l'importance dans la définition de la (dis)fluence de l'**acceptabilité communicative** de l'acte de parole produit en fonction des attentes de l'interlocuteur (Tomasello et al. 2005:276,279). Il a en effet été mis en évidence que les auditeurs-locuteurs prêtent en réalité moins attention à une « grammaire idéale » d'une phrase qu'aux **objectifs communicatifs** d'une situation (Debaisieux & Deulofeu 2001:71-72). Ainsi, une dimension de fluence pourrait ne pas être atteinte mais tout de même mener à une évaluation « fluente » du locuteur. Nous y reviendrons.

En conclusion, bien que la définition de Fillmore (1979) marque un premier pas vers une **description componentielle** de la fluence, ces approches demeurent souvent subjectives et difficilement formalisables (Hieke 1985:134-137). Avant tout qualitatives, elles ne permettent pas l'analyse du concept de fluence-disfluence en **variables discrètes**. Dans quelle mesure certaines marques langagières et certaines variables (extra)linguistiques doivent-elles être alors opérationnalisées, exclues, incluses et comment les positionner sur l'analyse du continuum dynamique de la fluence-disfluence[26] ?

2.2 Achoppements syntaxique, prosodique, pragmatique (Gars-Fribourg)

Au contraire des approches holistiques de Lennon (1990, 2000) ou de Fillmore (2000), les premières recherches en français sur la (dis)fluence s'attachent à une ébauche de description componentielle de la **fluence phrastique**. Elles découlent de la proposition de l'« **analyse en grille** », soumise à l'origine dans le but de faciliter la compréhension des transcriptions orales spontanées et de montrer une forme d'ordre derrière le « désordre apparent » (Blanche-Benveniste 1987, 1990a,b ; Blanche-Benveniste et al. 1979).

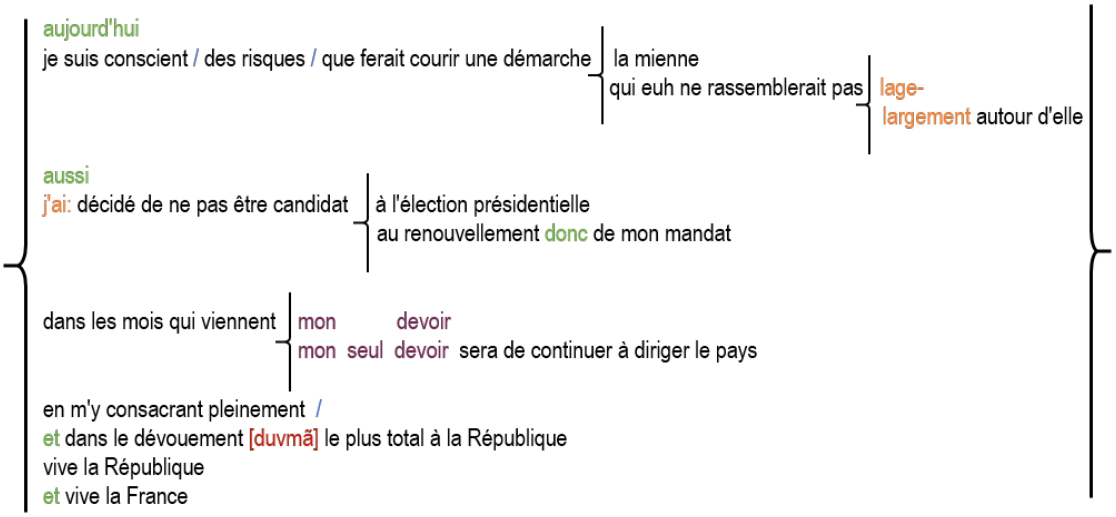
25. Texte original : "the ability to talk at length with few pauses,... the ability to talk in coherent, reasoned, and 'semantically dense' sentences, ... the ability to have appropriate things to say in a wide range of contexts, ... [and] the ability some people have to be creative and imaginative in their language use". (Fillmore 1979:26). Notre traduction. Nous surlignons.

26. Voir chapitres 2 et 3.

Basés sur la grammaire de dépendances, les questionnements de Blanche-Benveniste et du Groupe Aixois de Recherche Scientifique (G.A.R.S) sur une « grammaire de l'oral » ont révélé la difficulté de distinguer à l'oral le fruit d'**élaborations volontaires** /vs/ **involontaires du discours** ou, en d'autres termes, la fragile distinction entre les concepts de *coordination* et de *reformulation*.

De ces travaux résulte une des premières définitions d'un phénomène « disfluent » s'actualisant lorsque le déroulement de l'**énoncé syntagmatique**[27] **est brisé** et que cette occupation d'une même place syntaxique par plusieurs objets n'en modifie pas les caractéristiques sémantiques (Blanche-Benveniste 1987:137). Ainsi, par leurs descriptions d'une « syntaxe de l'oral », Blanche-benveniste et le G.A.R.S ont été les premiers à proposer une formalisation de l'« **achoppement dans le flux syntagmatique** » ou l' « **entassement paradigmatique** » et des diverses problématiques qu'il soulève – notamment l'**interprétation** (exemple 2).

> **Exemple (2)** – Proposition d'analyse et de représentation en grille de l'allocution présidentielle du 02 déc. 2016 (voir exemple 1).



La présentation de l'exemple d'avant-propos, formalisé ici en grille, met en évidence une **amorce de mot complétée**, « lage/ largement » correspond à un achoppement du flux syntagmatique, ainsi qu'à une **erreur de prononciation**, qui ne s'accompagne d'aucun signe de réparation ou d'édition de la part du président (« dévouement » prononcé [duvmã]), et donc d'aucune interruption du flux discursif. A fortiori, personne ne met en évidence la présence d'une « **reformulation** » (« mon devoir / mon seul devoir », une

27. L'approche de Blanche-Benveniste utilise les concepts d'axes syntagmatique et paradigmatique initialement développés par Saussure (1ère éd. 1916 ; 1989) dans son cours de linguistique générale. En résumé, l'*axe syntagmatique*, généralement représenté horizontalement, est celui sur lequel les différentes unités de la langue produites par un locuteur s'enchaînent séquentiellement lors de la production discursive. L'*axe paradigmatique*, illustré au contraire verticalement, représente un « ensemble de choix possibles » pour le locuteur, à un instant donné t lors de sa production discursive (e.g. « mon père est marié à [ma mère / une femme / quelqu'un] »).

répétition avec modification contenant une insertion). Celle-ci peut être présentée comme une réelle autocorrection, une **stagnation de l'axe syntagmatique**, mais elle ne semble *pas* être interprétée comme une *dis*fluence – comme un indice d'interruption ou de malaise discursif, comme une opération locale par laquelle « une séquence seconde éliminerait une séquence première » (Candea 2000a:32). Blanche-Benveniste (1990b) associe à des formes similaires en surface, un effet de précision et d'hyperonymie qui les rapproche parfois de « **retouches stylistiques** ».

Contrairement aux premiers modèles issus du traitement automatique du langage, l'objectif n'est pas une prise de décision entraînant le « gommage » d'un phénomène défini à priori comme *dis*fluent : on soulève justement **la multiplicité de degrés d'interprétation** entre rupture et continuité syntagmatique. Au linguiste de sélectionner les facteurs morphologiques, syntaxiques, acoustiques, prosodiques ou sociolinguistiques, permettant de désambiguïser ces « bafouilles », de départager « **l'erreur manifeste** » de la « **retouche intéressante** » (Blanche-Benveniste & Jeanjean 1986:161), c'est-à-dire de les positionner sur le continuum de la fluence-disfluence [28].

Dans le cadre de l'étude d'un phénomène linguistique, la nécessité de l'intégration des **éléments pragmatiques et sociaux** du langage appelle à une approche cognitive basée sur l'usage émergentiste (Barlow & Kemmer 2000). Avant de présenter ce cadre [29], nous revenons sur les premiers modèles cognitivistes de traitement de la parole par les locuteurs-auditeurs en mettant en valeur les aspects pragmatiques qui y sont impliqués, parfois en filigrane.

28. On retrouvera ces notions notamment dans les publications de Guénot (2005:180) et Dister (2007:123). Sur base des travaux de l'école du G.A.R.S, les études plus récentes réalisent la nécessité de l'intégration d'un modèle pragmatique pour la description contextualisée des (dis)fluences (Beliao & Lacheret 2013 ; Bolly 2010). Ces propositions distinguent notamment les niveaux de cohésions microsyntaxiques (i.e. régie par les unités de séquences fonctionnelles ou unités rectionnelles (Bilger & Campione 2002) et macrosyntaxiques (i.e. garantie par la dépendance illocutoire). L'inclusion de ces dernières constitue une avancée notoire vers une proposition robuste à la description des importants mécanismes paratactiques du français (voir chapitre 2 sections 1.2 et 1.3).

29. Voir chapitre 2 section 1.4.

2.3 Disfluence et *monitoring* : de Levelt à Günther ?

> "The complete speech act is a dynamic process, demanding the mobilization in proper sequence of a series of complex procedures and is the temporal integration of serial phenomena." (Goldman-Eisler 1968:6)

2.3.1 Langage : processus et résultat

Notre démarche adopte la prémisse communément admise que le langage oral s'offre à la fois comme un **processus** et comme un **résultat**. Cette approche est notamment développée par Levelt (1983:47, 1989:478) dans son modèle de production du discours. Selon ce modèle, la production du discours implique une combinaison de différents modules (Figure 1.6) :

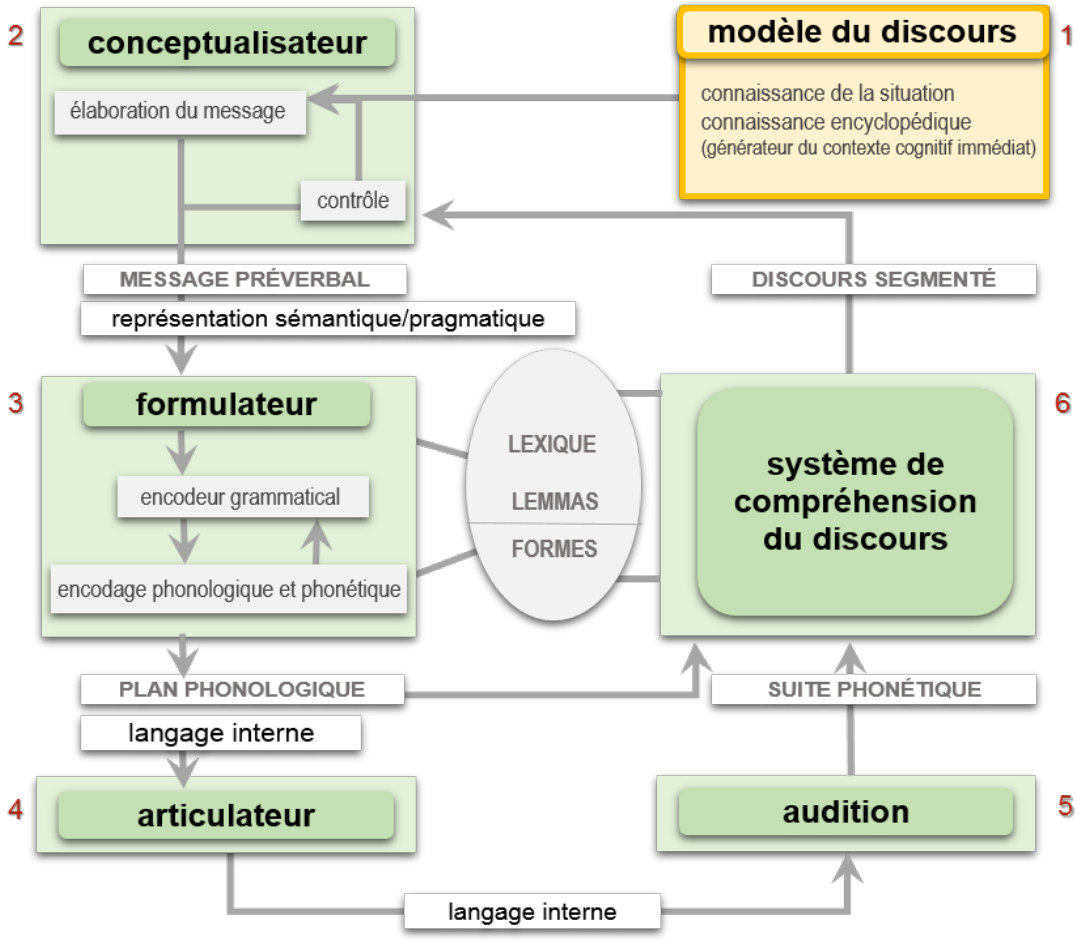


Figure 1.6 – Modèle de production orale du langage Levelt (1989) – traduction adaptée de Mazur-Palandre (2009:63).

1. Le *modèle du discours* dans Levelt (1989) englobe les connaissances de la situation et les connaissances encyclopédiques du locuteur. Bien que Schmid & Günther (2016) ne présentent pas un modèle de traitement de la parole en tant

que tel mais bien un modèle descriptif des types de saillances, on peut associer ce module discursif développé en psycholinguistique aux processus de génération du **contexte cognitif immédiat** développé dans le cadre sociocognitiviste [30].

2. Le *conceptualisateur* construit le message, ordonnance les idées, les connaissances de la situation et les connaissances encyclopédiques. Il permet de sélectionner et d'organiser les idées selon l'*intention communicative* du locuteur. C'est ici que s'organise la **représentation des contextes** communicatifs immédiats et du contexte externe.

Les représentations stockées sont liées à des **représentations sémantiques** (e.g. prédicat, agent, patient, instrument, expérient, etc.), ou **pragmatiques** (la modalité, i.e. topique, commentaire, assertion, interrogation, etc.) mais aussi à la **situation du discours** en tant que telle et sa relation au type de discours (dialogue, interactivité, préparation etc.). En résumé, le conceptualisateur planifie les connaissances (transformées et rapportées) et les rapporte sous forme de **messages pré-verbaux**. Il lie les connaissances *déclaratives* et *procédurales* (Levelt 1989).

Ces dernières sont plus ou moins spécialisées en fonction des **contraintes discursives**. En effet, selon le « type de texte », la structure du message est relativement conventionnelle, normée, évidente (Levelt 1995). Nous le verrons à la section suivante : c'est au sein du conceptualisateur que les processus de **micro et macroplanifications** régissent les messages prélinguistiques.

3. Le *formulateur* permet la production textuelle, l'accès lexical et phonétique. Il comprend l'*encodeur grammatical* et l'*encodeur phonologique*. Le premier sélectionne les items lexicaux et leurs caractéristiques syntaxiques tout en veillant à **leur cohésion** grammaticale. Aux significations et propriétés syntaxiques propres aux items lexicaux s'associent donc des caractéristiques morpho-phonologiques liées aux formes. Ces composants du lexique mental, les entrées lexicales, passent ensuite par l'encodeur phonologique.

4. L'*articulateur* permet l'*exécution motrice* des formes élaborées précédemment par l'encodeur phonologique, les chaînes ordonnées de lemmes.

5. L'*audition* permet de passer du langage interne au langage externe.

6. Le *système de compréhension du discours* prend en entrée la production langagière orale. Ce discours, structuré et segmenté, fait l'objet de **procédures de contrôle**. « Ce composant permet un retour du locuteur sur son message afin de corriger les erreurs éventuelles » (Mazur-Palandre 2009:66).

Dans notre exemple présidentiel, l'entrée « dévouement » est constituée (1) d'un sens, (2) d'informations syntaxiques (e.g. nom, nom concret), de propriétés morphologiques (+*s* au pluriel), et de caractéristiques phonologiques (e.g. unité polysyllabique, syllabe de

30. Voir section 1.1 de ce chapitre.

type CV, etc.). A ce profil, s'ajoutent les propriétés pragmatiques et discursives. Celles-ci influencent les micro et macroplanifications du discours effectuées dans le module du conceptualisateur (section 2.3.2) et du principe de *monitoring* adopté par les théories du *Perceptual Loop Model*. Nous le verrons, la présence de **différents types de moniteurs** donne lieu à différents types de *repairs*[31].

2.3.2 Disfluences, micro et macroplanifications

> « Le choix d'une forme se réalise par rapport à ce qui a été énoncé, ce qui est dit et ce qui va être dit et selon le but et le contexte communicationnel de la situation. Ainsi, le choix de l'item voiture dans un message, est en lien avec le message en cours. Selon que le référent de cette entrée lexicale est connu des interlocuteurs ou non, le locuteur fait des choix différents. Ainsi, le statut informationnel peut avoir un impact sur le déterminant ou encore sur la position syntaxique de l'item. » (Mazur-Palandre 2009:65)

Lors de la conceptualisation, les locuteurs s'engagent simultanément dans deux types de processus cognitifs :

- la ***macroplanification*** de leur intention communicative, c'est-à-dire la récupération des données pertinentes à la situation de communication (e.g. « Le plat manque d'épices. Dire à Marie que je voudrais du sel. »).

- la ***microplanification*** du contenu informationnel, impliquant la recherche de la forme propositionnelle et de la forme stylistique de l'énoncé et dépendante en grande partie de l'accessibilité des référents et de leur statut informationnel. (e.g. « Marie, pourrais-tu me passer le sel ? »).

Ces types de processus sont cognitivement demandeurs, et peuvent s'**alterner de manière séquentielle**. Les micro et macroplanifications incluent le concept de « monitoring ». Partiellement conscients, les *self-* et *other- monitoring* permettent aux locuteurs de comparer leur intention énonciative initiale aux énoncés effectivement produits et perçus. Pendant le traitement de la parole, les interlocuteurs évalueraient ainsi le ***degré d'appropriété*** et la ***pertinence*** des productions discursives au regard (de la situation et) des interprétations qu'elles permettraient d'inférer. En production, tout en poursuivant le processus mental de linéarisation du discours, le locuteur écoute son propre discours et en **réajuste** le contenu ou la forme si nécessaire, en fonction de son intention communicative. Ces réajustements sont notamment linguistiques : les « bafouilles » ou (dis)fluences sont le résultat du processus de *monitoring* (e.g. « Marie, pourrais-tu me passer le poi/ euh non le sel ? »).

31. Voir sections 2.3.3 et 2.3.4 de ce chapitre.

2.3.3 Boucle perceptuelle : anticipation des erreurs

Les (dis)fluences sont à la fois une manifestation du « **contrôle de qualité** » effectué par un locuteur (Hieke 1981:148) et une fenêtre sur les processus cognitifs sous-jacents en cours chez les locuteurs-auditeurs lors de leur pratique discursive en production comme en perception. Les modules de ***monitoring*** leur permettent d'effectuer ce « contrôle de qualité », **avant même qu'une erreur ne soit produite** (Doutrich 2000 ; Levelt 1983 ; Reber, Schwartz & Winkielman 2004 ; Schegloff, Jefferson & Sacks 1977). La Figure 1.7 est un prolongement de la Figure 1.6 : le modèle de production de la parole et ses associations avec les divers moniteurs. Plusieurs types de *monitoring* donnent ainsi lieu aux *types de repairs*[32].

De nombreuses expériences montrent que les locuteurs sont capables de détecter et d'anticiper leurs erreurs même masquées par du bruit (Hartsuiker 2014:418). Le locuteur peut effectuer un *feedback* sur son message avant et après l'entrée en action de l'articulateur. Ce principe d'**anticipation** du contenu et des erreurs à venir est aussi connu sous le nom de *Covert Repair Hypothesis* (Postma & Kolk 1993). On distingue l'***action externe*** – quand la correction suit la réalisation sonore du message – et l'***action interne*** – quand la correction précède la réalisation sonore du message. Ces deux types de *repairs* (*covert* /vs/ *overt*) sont en réalité distingués en fonction du moment où le *monitoring* est réalisé.

Les *overt-repairs*, (dis)fluences « latentes » ou « déguisées » sont des erreurs apparues durant le processus de planification mais corrigées avant d'être articulées (action interne). Elles ne sont ainsi jamais effectivement actualisées si ce n'est par la présence de pauses pleines, de termes d'édition ou de répétitions d'un ou plusieurs items lexicaux. Elles font référence à la **complétion d'un énoncé** qui est simplement **différée** par une (dis)fluence (Ginzburg, Fernández & Schlangen 2014:55). Elles émanent de *dis*fluences d'appropriété ou de *dis*fluences d'erreur (voir ci-après), et peuvent donc recouvrir des processus de traitement de la parole au stade de **conceptualisation** comme de **formulation**.

Le concept de **boucle perceptuelle**, introduit par les théories du *Perceptual Loop Model* – PLM) (Hartsuiker & Kolk 2001 ; Levelt 1989), offre un cadre cohérent à l'actualisation des processus de *monitoring*. Le contrôle se fait à plusieurs moments du déroulement de production de la parole par l'activation de différents moniteurs. On distingue trois phases dans le *self-monitoring* :

– La première phase concerne le processus d'hésitation qui apparaît suite à une erreur d'unité(s) linguistique et/ou informationnelle. Le locuteur s'interrompt.
  Cette phase de **détection** du trouble de la parole est réalisée grâce au processus de *monitoring post-articulatoire* (permettant d'écouter l'*overt-speech* ou contexte externe), et de *monitoring pré-articulatoire* (contexte interne ou *inner-speech*).

32. Dans la terminologie de Levelt (1983) le terme *repair* équivaut à celui de (dis)fluence. Voir section 2.3.4 de ce chapitre.

– La seconde phase a trait à l'apparition de **termes spécifiques** signalant un problème à résoudre dans le message. L'interruption est un signal d'arrêt marquant une compétition entre le processus de la pensée et de l'action. Elle signale une **erreur de performance** [33].
– Enfin, la troisième phase se définit comme la **reprise du message** engendrant la production de la correction de son propre message. Cette **reprise** du discours est réalisée avec ou sans *repair* selon que la complétion du contenu antérieur est nécessaire ou non.

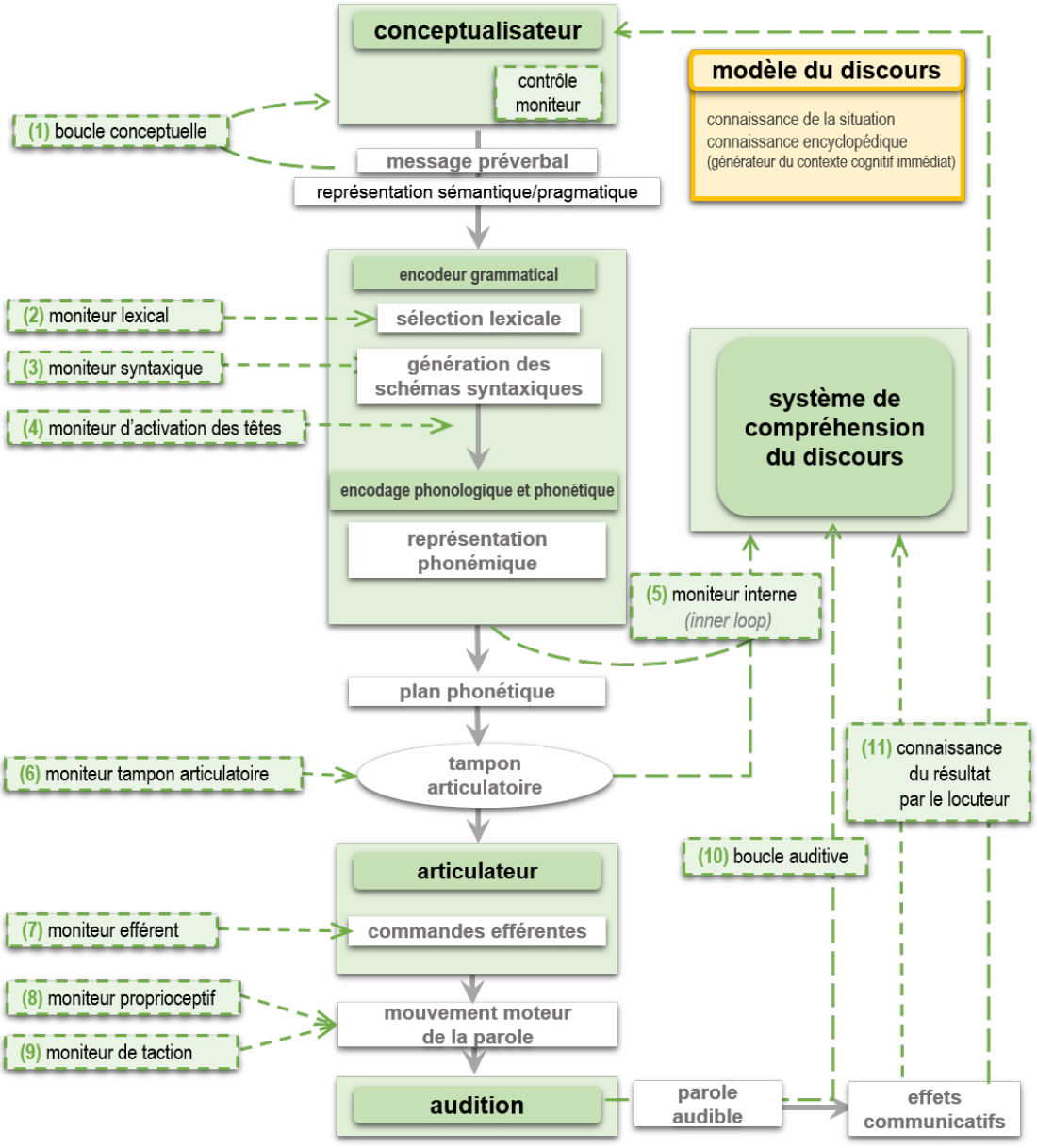


FIGURE 1.7 – Modèle de traitement de la parole incluant les boucles de *feedbacks* et les moniteurs, sous-jacent à la production et détection d'erreurs et à la production de (dis)fluences – inspiré de POSTMA (2000:99).

33. Nous revenons sur les prolongements de cette littérature au chapitre 2 section 1.1.1.

De leur côté Ginzburg, Fernández & Schlangen (2014:5,55) comptent également ces deux types de (dis)fluences :

- (Dis)fluence à **action externe** ou *covert-repair* qu'ils appellent les ***Backward Looking Disfluencies (BLDs)*** pour lesquelles le moment de l'interruption est suivi par une altération qui fait référence à un *reparandum* déjà prononcé.
- (Dis)fluence à **action interne** ou *overt-repair* qu'ils appellent les ***Forward Looking Disfluencies (FLDs)*** sont des moments où l'interruption est suivie par une complétion de l'énoncé temporairement retardé par une pause ou pause pleine ou une répétition d'une partie antérieure de l'énoncé. Une des différences principales est que ces (dis)fluences s'actualisent comme des *phases d'édition*, et révèlent donc l'existence de contenu à venir (*"soon-to-be-uttered word"*) (Ginzburg, Fernández & Schlangen 2014:50).

En conclusion, nombre d'auteurs font le lien entre le **type de moniteur** intervenant dans la (dis)fluence et le **type de perturbation** sur l'ordre syntaxique. La Figure 1.8 illustre (et simplifie) ces liens évoqués dans les sections antérieures. Dans le cadre de cette thèse, ces connexions sont intéressantes pour la formulation d'hypothèses liées aux effets phonétiques de durées entraînés par les (dis)fluences [34].

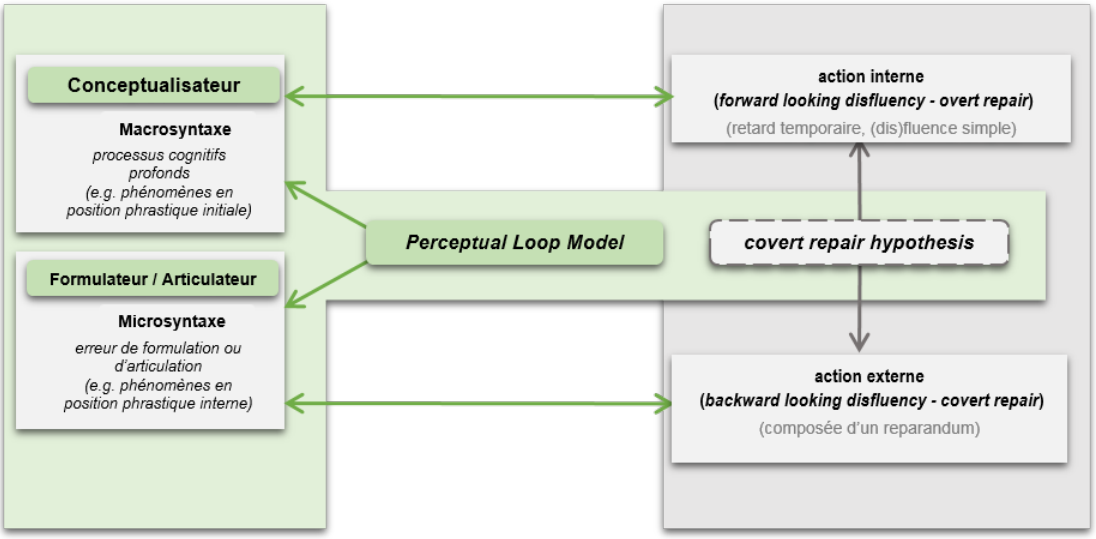


Figure 1.8 – Association des types de *repairs* aux modules de *monitoring* et aux niveaux de syntaxe.

34. Voir chapitre 2 section 1.1.2.

2.3.4 Sources des différents types de *repairs*

Levelt (1983:50-55) oppose les *covert-repair* aux *overt-repair*. Parmi ces *repairs*, on distingue au moins trois types de disfluences sous-tendant un dysfonctionnement au niveau du processus de **planification discursive**. Cette approche des (dis)fluences est fonctionnelle mais peut correspondre, de manière univoque ou non, à des (dis)fluences de surfaces spécifiques :

- **Les (dis)fluences d'appropriété (*appropriateness repairs*)** (Levelt 1983:51-53) représentent des cas où le locuteur revient sur son discours pour le **modifier** et/ou **spécifier** ce qu'il a énoncé. Elles regroupent des cas où le *reparandum* est grammaticalement correct mais a été jugé incorrect par le locuteur lui-même. Elles peuvent être dues à un problème d'*ambiguïté*, de *spécification*, de *précision contextuelle* (exemple 3a, b, c, respectivement).
  La plupart des **insertions** (ajout d'un élément au *reparandum*) reflètent des (dis)fluences d'appropriété. Les erreurs d'appropriété sont reliées aux processus de traitement de planification et de formulation de la parole, elles sont donc basées conceptuellement (Blackmer & Mitton 1991). Cette règle générale est à nuancer dans le cas des *substitutions* (remplacement d'un item lexical par un autre item synonymique d'un autre registre) qui suppose davantage une correction au niveau de la sélection lexicale.

  Exemple (3) – exemples forgés[35]
  a. on arrive à l'aéroport, au petit aéroport
  b. l'aéroport est à côté de la ville – juste après la grand route
  c. il faut tourner à droite, à ta droite

- **Les (dis)fluences d'erreur (*error repairs*)** sont associées au processus de formulation et classées comme des « disfluences de production » par Blackmer & Mitton (1991). Elles s'actualisent lorsque le locuteur a violé une règle grammaticale, produit une forme lexicale différente de celle qu'il avait l'intention de produire ou encore a mal prononcé un mot. Les substitutions sont typiques des *disfluences d'erreur*. Elles peuvent être *lexicales*, *syntaxiques* ou même *phonologiques* (erreur d'articulation) (exemples 4 a, b et c, respectivement).

  Exemple (4) – exemples forgés.
  a. il est à l'aérogare *euh* l'aéroport
  b. ils ont été bien reçus avec eux – par eux avec bras ouverts
  c. l'aréoport *euh* l'aéroport

- **Les (dis)fluences *D-Repair*** sont des cas où le locuteur décide d'entièrement réordonner son intention communicative. Le « D » venant de "Do I want to say this now ?". Elles sont, à l'instar des erreurs d'appropriété, basées conceptuellement. Généralement, celles-ci correspondent à des cas de *délétions* ou de *faux départs*.

35. Malgré cette classification, on ne peut s'empêcher de se questionner sur l'aspect « disfluent » des exemples 3a et 3b. L'insertion peut être vue comme une expansion de nom et non comme une *dis*fluence.

Exemple (5) – traduit de Levelt (1983:51).
a. We go straight on or [...] we come in via red, go then straight on to green
(On va tout droit ou [...] on va par le rouge et puis directement vers le vert)

Morel & Danon-Boileau (1998:75) font référence aux *D-repairs* en les caractérisant comme des marques liées à un « **malaise de la pensée** », c'est-à-dire au travail de conceptualisation, contrairement aux (dis)fluences d'erreur ou d'appropriété qui révéleraient des « malaises de la formulation ». Cette distinction s'associe également à celle déjà faite il y a plus de 30 ans par Chafe (1980:171) : parfois les locuteurs hésitent pendant qu'ils sont en train de décider de *ce dont ils vont parler*, et parfois il sont en train de décider *comment ils vont parler* de ce qu'ils ont choisi[36]. La locution **« (dis)fluence d'appropriété »,** souligne le fait que la définition de la (dis)fluence et les tâches de *monitoring* qui lui sont reliées, sont intrinsèquement dépendantes de la définition du « contexte ».

L'influence de ce contexte peut être différenciée en fonction du caractère **intrinsèque** /vs/ **extrinsèque** des paramètres de production et de perception de la parole[37]. Les niveaux intrinsèques sont directement dépendants des habilités cognitives du locuteur au moment de l'interaction c'est-à-dire des contraintes exercées sur la capacité de sa mémoire de travail (Chafe 1994:28-40). Les paragraphes ci-après donnent un bref aperçu des associations des contextes et des (dis)fluences.

1. Au **niveau intrinsèque conceptuel**, l'accroissement de la complexité cognitive, syntaxique ou prosodique est positivement corrélé à l'augmentation de (dis)fluences (Ferreira 1993 ; Goldman-Eisler 1968 ; Maclay & Osgood 1959 ; Wheeldon & Lahiri 1997). En d'autres termes, la probabilité d'apparition de (dis)fluences est généralement plus élevée lorsque vous essayez de vous remémorer et d'expliquer spontanément la recette d'un parfait au chocolat que celle d'un œuf à la coque. Dans le même ordre d'idées, la présence d'indécisions ou d'incertitudes quant à l'**intention communicative initiale**[38] aura également tendance à accroître la probabilité de *dis*fluences (Brennan & William 1995).

2. Au **niveau intrinsèque mais lexical**, l'accessibilité lexicale (qu'elle soit mesurée en terme de familiarité, de fréquence d'occurrence, période d'acquisition, etc.) et l'aisance articulatoire (e.g. type, nombre et combinaisons de phonèmes ou de syllabes) favorisent la production de discours exempts de (dis)fluences (Kapatsinski 2010:100-101).

3. Au **niveau extrinsèque** enfin, les locuteurs peuvent avoir tendance à produire des (dis)fluences en raison de distractions situationnelles « locales ». L'importance des processus de *monitoring* est notamment conditionnée par le **degré d'attention** que les locuteurs portent à leur discours. Comme le faisait remarquer Halliday (1987:67), la

36. Texte original : "Sometimes speakers hesitate while they are deciding what to talk about next, and sometimes they are deciding how to talk about what they have chosen." (Chafe 1980:171). Notre traduction. Nous *mettons en italique*.
37. Voir section 1.1 de ce chapitre.
38. Autrement dit leurs attentes linguistiques liées à la fluence phrastique interne.

production de (dis)fluences est caractéristique d'une **parole autoconsciente et autosurveillée** : dans une interaction spontanée en direct, peu de locuteurs ou auditeurs sont réellement conscients du flux de leur parole ; les locuteurs ont en général alors tendance à moins hésiter ou à s'arrêter dans leur flux discursif (Halliday 1987:69). Ainsi, les notions de *monitoring* et de fluence sont foncièrement dépendantes du **niveau d'attention**, **de conscience** ou **d'empathie** apportés à son propre discours et aux interlocuteurs. Par ailleurs, cette attention à son propre discours dépend partiellement de la sensibilité du locuteur vis-à-vis des besoins de l'auditeur, en d'autres termes, de la **synchronie interactionnelle** des interlocuteurs (Clark & Tree 2002:9-10).

Lorsque des **tâches cognitives lourdes** sont effectuées conjointement à l'acte de parole (facteurs extrinsèques), et en conséquence que la capacité de la mémoire de travail est réduite, on remarque une **augmentation de la production de (dis)fluences** (Christodoulides 2016). A l'inverse, lorsqu'un locuteur attache peu d'attention à la forme et/ou au contenu de son discours et/ou de son auditeur, l'**attention au processus de *monitoring*** se voit diminuée et par là même la production de (dis)fluences (Halliday 1987:69). Du point de vue de l'auditeur, si son attention et ses attentes par rapport au discours du locuteur sont réduites, il sera non seulement moins enclin à percevoir les éventuelles (dis)fluences, mais également à les juger comme **saillantes et/ou perturbatrices du flux discursif**.

En conclusion, le concept de *monitoring*, central aux théories de production et de perception des (dis)fluences, est un **processus cognitif non linéaire allant à l'encontre du résultat linéaire inhérent au flux oral de la parole**. Il implique une vérification constante de segments discursifs antérieurs ou à venir, venant d'être prononcés /vs/ planifiés par le locuteur et/ou attendus par l'auditeur. Cet enchaînement souligne à quel point la notion de fluence est liée à celle de monitoring et celle de *monitoring* à celle d'attention. Or, **le focus de l'attention dépend lui-même des attentes des interlocuteurs** (Chafe 1994:301 ; Zarcone, van Schijndel, Vogels & Demberg 2016 ; Schmid & Günther 2016). Nous émettons l'hypothèse que ces attentes sont liées en partie à **l'image sonore attendue**, présente dans notre **mémoire auditive**. Cette mémoire auditive, **partiellement partagée** par une communauté linguistique mais avant tout personnelle à chaque locuteur-auditeur, est plus ou moins homogène et précisément définie en fonction des **normes de production** de la parole. En linguistique, le domaine de la **prosodie** apporte de nombreux éclaircissements sur ces normes de production et de perception. ***Mais, qu'est-ce que la prosodie ?***

3 Définition synthétique de la prosodie

« Qu'est-ce que la prosodie du français ? », se demande-t-on plus précisément. Les réponses sont nombreuses et diverses. Elles varient selon la perspective du chercheur [39]. Plus d'un se sont essayés à l'élaboration d'une **définition synthétique** de celle-ci. Récemment, sur base de huit définitions de la prosodie, Di Cristo (2013:23) propose sa contribution :

> « La prosodie est une branche de la linguistique consacrée à l'analyse des propriétés formelles (phonologie prosodique) de la matérialité (phonétique prosodique) et de la fonctionnalité des éléments non verbaux de l'expression orale, non coextensifs aux phonèmes, tels que l'*accent, les tons, l'intonation, la quantité, le tempo* et les *pauses* [...]. (Leur) matérialité physique [...] se manifeste par les variations de la *fréquence fondamentale (F0), de la durée et de l'intensité* (paramètres prosodiques et acoustiques), ces variations étant perçues par l'auditeur comme des changements respectifs de *hauteur, de longueur* et de *volume sonore* (paramètres prosodiques auditifs). Les informations prosodiques ont la particularité d'être polysémiques et de véhiculer à la fois des informations d'ordre *linguistique, paralinguistique* et *extralinguistique*, qui se révèlent déterminantes pour l'interprétation des énoncés du discours et de la conversation. » (Di Cristo 2013:23)

Cette définition, quoique réunissant les éléments constitutifs essentiels de la prosodie, présente une **densité informationnelle** non négligeable pour le non-spécialiste. Sans prétention d'exhaustivité, nous répondrons à la question « qu'est-ce que la prosodie ? » en reprenant les mots-clés de cette définition, afin de donner nous l'espérons, au lecteur non prosodicien une vision claire d'un champ de recherche riche et relativement complexe [40].

3.1 Systèmes et fonctions

En section 1 de ce chapitre, nous avons ébauché une définition de la prosodie comme un champ d'étude d'un ensemble de **phénomènes suprasegmentaux** (i.e. se manifestant « au-delà » ou « au-dessus » des segments phonétiques individuels) formé par les **systèmes intonatif** et **rythmique** d'une langue orale.

1. Le **système intonatif** hiérarchise les unités de sens et en organise la cohérence (Léon 1993:45). Il fait référence au système des patrons mélodiques distinctifs qui ont pour domaine l'énoncé et ses constituants (Di Cristo 2013:4). Ces patrons peuvent être vus comme des formes globales indivisibles ou des unités hiérarchiques décomposables en unités primitives. Au niveau matériel, l'intonation correspond à

39. Cependant, certains ouvrages introductifs très compréhensifs sont à la disposition du lecteur. Nous conseillons particulièrement Lacheret-Dujour & Beaugendre (1999), Léon (1993) et Simon (2004, 2012) et plus récemment Di Cristo (2013).

40. L'importance de la prosodie dans l'analyse des (dis)fluences est traitée de manière plus précise au Chapitre 2.

des variations de la hauteur de la voix, exprimée acoustiquement par des changements de la fréquence fondamentale principalement. Par ailleurs, les variations intonatives traduisent au niveau phonostylistique des changements d'émotions ou d'attitudes ainsi que des manifestations des types de discours.

2. Le **système rythmique** se réfère à la récurrence de groupements fondés sur l'organisation hiérarchique de *niveaux d'accentuation* (ou de proéminences syllabiques). Dit simplement, le rythme se rapporte à l'alternance (plus ou moins régulière) de temps forts et de temps faibles. Par conséquent, le système rythmique est par définition intimement lié au **système accentuel** qui se rapporte, lui, au fonctionnement des accents dans une langue. L'accentuation est une contrainte linguistique permettant soit les distinctions lexicales, soit le découpage du flux sonore en unités discrètes décodables par l'auditeur (Léon 1993:45). L'*accent* se rapporte à la notion de proéminence, c'est-à-dire à une unité détachée de son environnement phonique (sur le plan physique et sur celui de la perception) (Di Cristo 2013:5). Précisons.

   Au niveau *hiérarchique*, en français l'*accent primaire* frappe la dernière syllabe pleine de mots dit « accentogènes ». Il se distingue de l'**accent secondaire** qui frappe en général la première syllabe des mots de cette même catégorie.

   Au niveau fonctionnel *lexical*, l'accent français a une **fonction démarcative**[41]. Il signale le début ou la fin des mots ou d'unités de rang supérieur tel que le syntagme. Au niveau fonctionnel *discursif* cette fois, l'accent assume aussi des fonctions rythmiques, syntaxiques, sémantiques et pragmatiques. Dit autrement, l'actualisation des accents est soumise à des contraintes relevant de l'organisation rythmique, de la syntaxe, de la sémantique et de la pragmatique.

   Enfin, au niveau *matériel*, c'est surtout la durée et la F0 qui marquent l'accentuation francophone, bien qu'en pratique une série de paramètres prosodiques peut participer au **marquage de l'accentuation** : fréquence fondamentale (impression de hauteur), amplitude (impression d'intensité), durée (impression de longueur) et spectre acoustique (qualité du timbre vocalique) (Delattre 1966).

On note que la prosodie et les concepts clés qui y sont rattachés définissent, par leur **nature formelle**, leur **matérialité physique** ou encore leur **fonctionnalité** (Di Cristo 2013:1). La nature des variables dépend du niveau envisagé.

41. L'accent a une fonction cumulative dans d'autres langues – il permet de dénombrer les mots d'un énoncé – ou une fonction distinctive – l'accent tonique sert par exemple à distinguer les mots lexicaux entre eux (Di Cristo 2013:6).

3.2 Nature des unités et niveaux d'interactions

> « Les étiquettes pour qualifier les styles sont aussi nombreuses que les paramètres prosodiques susceptibles de varier selon la situation. »
> (Simon, Auchlin, Avanzi & Goldman 2010:77)

Quels sont les niveaux de la prosodie et comment interagissent-ils ? Dutoit (2013:129-131) opère le même type de distinctions que Di Cristo (2013). Il identifie les variables selon la nature de l'observation. Les paramètres acoustiques (physiques) sont directement mesurables par observation du signal sonore. Le niveau perceptuel réfère aux effets perçus des paramètres acoustiques. Enfin, le niveau linguistique a trait aux fonctions des paramètres acoustiques perçus (Dutoit 2013:130-131).

Le **niveau perceptuel** met en relation les évènements prosodiques physiques et la manière dont les auditeurs les perçoivent (Mertens 2014). Les mesures acoustiques interagissent entre elles et ne forment pas une relation de parfaite équivalence entre les mesures acoustiques et les corrélats perceptuels. Définissons les niveaux acoustiques par rapport à leur corrélat perceptuel.

- La *fréquence fondamentale* a pour corrélat perceptuel l'**intonation** (*pitch*). En modélisation de la parole, on stylise la courbe de fréquence fondamentale afin de faire d'une multitude d'informations acoustiques discrètes un flux continu d'information : le signal auditif analysable – notamment par la transformation de Fourrier (Allen 1977). L'intonation est mesurée sur une échelle logarithmique en **semi-tons**, unité fidèle à la perception humaine.
- La *durée* a pour corrélat perceptuel la **longueur** (*lengthening*), elle s'incarne dans le temps, de manière absolue (en sec. ou millisec.) ou relative par rapport aux segments environnants.
- L'*intensité* a pour corrélat perceptuel le **volume sonore** (*loudness*), mesuré en décibels (dB).
- Les **formants** ont pour corrélat perceptuel la **qualité vocale**. Le spectre sonore est la somme de partiels sonores – les harmoniques.

Le **niveau linguistique** de la prosodie se définit par le fait qu'on associe à la prosodie des fonctions linguistiques. Les formes prosodiques révèlent elles aussi des fonctions communicatives à divers niveaux : linguistique, paralinguistique et extralinguistique (Mertens 2014).

- Au *niveau linguistique*, la prosodie est un **ponctuant** et **structurant** du discours tant du point de vue de sa production que de sa perception. Elle joue par exemple un rôle important dans la perception des frontières de mots mais également de **groupes prosodiques** (Cutler, Dahan & Van Donselaar 1997 ; Delais-Roussarie 2000 ; Lacheret-Dujour & Beaugendre 1999). Les groupes prosodiques sont en effet partiellement reliés au statut informationnel des **groupes de sens** – focus, topique,

parenthèse ainsi qu'à leur statut illocutoire. Comme le faisait déjà remarquer Delattre (1966), la prosodie permet notamment de distinguer différentes **modalités langagières** (e.g. questionnement, ordre, fait, citations, discours rapporté, etc.) et facilite par là même la gestion de l'interaction.

– Au *niveau paralinguistique*, certains patrons prosodiques ont été associés à l'ironie, au sarcasme, ou à différents autres **états émotionnels** (Attardo 2014). La prosodie joue alors un rôle dans les *implicatures conversationnelles* que les auditeurs peuvent formuler [42].

– Au *niveau extralinguistique* enfin, la prosodie témoigne de **variations sociolinguistiques** (sociale, géographique ou situationnelle). Ces variations s'incarnent notamment dans ce qu'on appelle des **phonostyles** ou **idiostyles** (Avanzi, Obin, Bardiaux & Bordal 2012 ; Boula de Mareüil 2014 ; Léon 1993 ; Simon, Auchlin, Avanzi & Goldman 2010). Nous y reviendrons.

Comment la prosodie s'intègre-t-elle au sein des systèmes de production et de perception de la parole, en situation ? Proposer l'intégration du système ou « compilateur » prosodique au sein d'un **système de production, de perception et de compréhension** de la parole présuppose qu'on s'accorde déjà sur une bonne définition de ce dernier. Au chapitre 1 section 2.3, nous avons, brièvement, présenté le modèle de production et de compréhension de la parole de Levelt (1983, 1989). Sur cette base, Di Cristo (2013:40-46) propose de modéliser le compilateur prosodique au regard des systèmes de production et de compréhension de la parole.

La Figure 1.9 souligne les liens directs et indirects de la prosodie avec les différents modules de production et de perception de la parole. En **production**, le lien principal se joue en amont de la conceptualisation, il a donc un impact important sur l'ensemble des processus de traitements subséquents. En **perception**, le compilateur prosodique intervient essentiellement comme un facilitateur de traitement (i.e. *processus inférentiels, segmentation et décodages linguistiques*, et de *décodage acoustico-prosodique*).

42. Rappel de la note de bas de page 18 : [...] une implicature conversationnelle se réfère à ce qui est suggéré par un locuteur de manière implicite (i.e. et non explicitement par une propriété sémantique de l'énoncé). Une *implicature* se distingue d'une *implication*, qui est annulable sémantiquement si l'on ajoute une proposition, ou d'une *présupposition implicite* de nature sémantique. Enfin, une implicature *conversationnelle* dépend du contexte de la conversation, elle se distingue d'une implicature *conventionnelle* dépendante de l'énoncé lui-même.

3.3 Phonostyle et (dis)fluences

> « La notion de choix, qui est à la base de la stylistique et de la phonostylistique, implique un effort plus ou moins conscient de la part du locuteur. Une fausse coupure, un silence prolongé, la voix qui reste en suspens, sont autant de marques d'un phonostyle oratoire. » (Léon 1993:9)

Synthétiquement, la fluence n'équivaut *pas* à un respect de **linéarité ininterrompue du flux discursif**, tout à fait irréaliste au regard de la capacité cognitive de certaines situations (spontanées mais formelles en particulier). Pour rappel, il est reconnu depuis Kasl & Mahl (1965) que la présence de (dis)fluences dans un discours dépend de la situation et plus précisément du canal de communication envisagé [43].

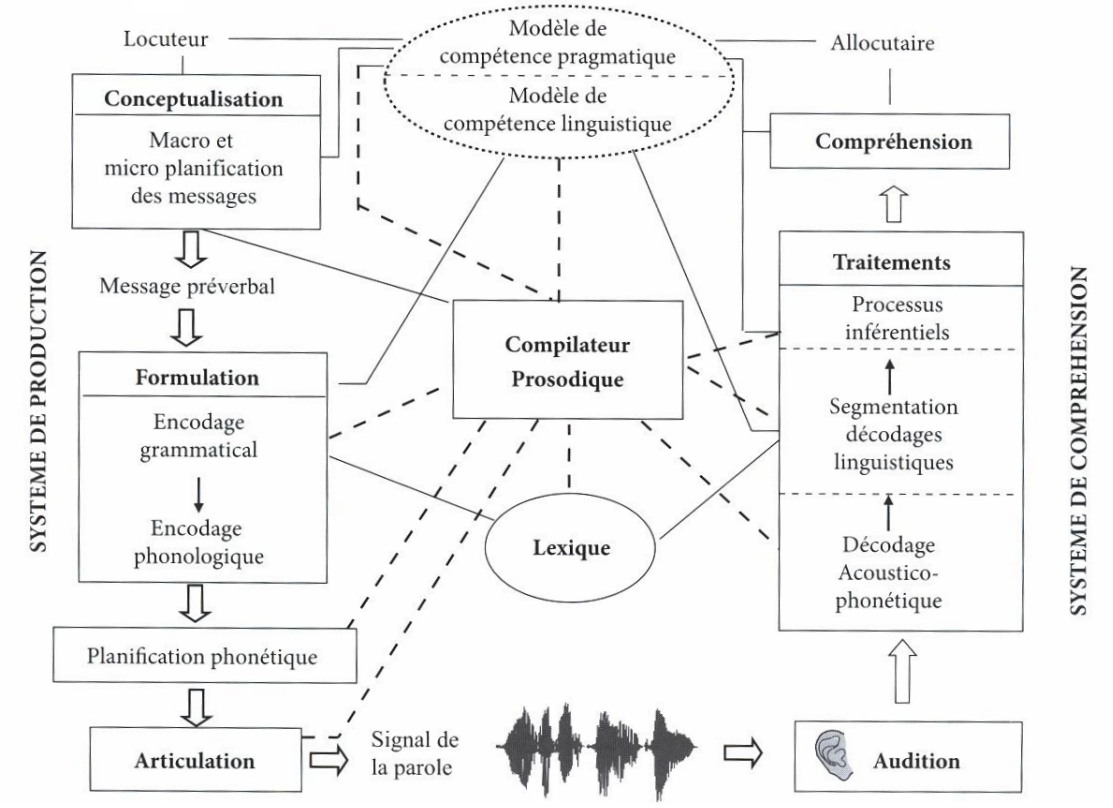


Figure 1.9 – Modèle de compilateur prosodique au regard des systèmes de production et de compréhension de la parole de Di Cristo (2013:44).

Reprenons l'exemple présidentiel d'avant-propos. Certaines « bafouilles » sont, **en contexte, plus acceptées que d'autres**. Hésiter dans une conversation spontanée avec vos pairs vous vaudra sans doute moins de jugement en faveur d'une (dis)fluence. Le contenu de votre discours n'étant pas programmé, il est accepté que la **charge cognitive** nécessaire à sa production soit davantage coûteuse et donc empreinte de

43. Les auteurs ont ainsi observé que les sujets sont plus enclins à produire des (dis)fluences en situation de télécommunication qu'en face-à-face. Ces résultats ont été confirmés par Bell, Eklund & Gustafson (2000).

petites marques du travail en cours de formulation (Berthold & Jameson 1999 ; Boomer 1965 ; Christodoulides 2016). Ce n'est pas le cas pour une allocution présidentielle préparée, qui répond à des normes et des **attentes socioculturelles** plus figées. Cette approche de l'évaluation de la « normativité » des attentes liées à une situation n'est encore qu'une autre manière de mettre en évidence le fait qu'une marque peut être interprétée différemment en fonction de son contexte d'énonciation. Le fondement même des analyses proposées dans cette thèse suit la prémisse largement acceptée selon laquelle, en production, la **pratique communicative**[44] influence la **charge cognitive** des locuteurs et affecte leur production discursive (Goldman et al. 2014 ; Simon, Auchlin, Avanzi & Goldman 2010).

L'étude des phonostyles éclaire ce qui peut se cacher derrière ces « **pratiques communicatives** ». Elle favorise le développement de la notion de (dis)fluence de manière contextuelle. Pour rappel, en utilisant la terminologie de Segalowitz (2016), la *fluence produite* est à la base des inférences que les auditeurs font à propos de la situation ou de l'état cognitif de locuteurs impliqués dans celle-ci (*fluence perceptive*). Dans cette approche, la notion de fluence perceptive dépend des **attentes spécifiques** des locuteurs quant à une situation à laquelle ils auraient attribué un ensemble de **normes sociales** (ou sociolinguistiques). L'impression de fluence ou de disfluence se module ainsi à travers l'**écart entre les normes et les attentes** d'un auditeur vis-à-vis d'une performance (i.e. un discours produit dans une situation de parole définie) et la performance telle qu'elle est actualisée (Alter 2013).

La **phonostylistique** s'intéresse aux variations lexicales, prosodiques et syntaxiques sous l'angle de la variation pragmatique, dans la mesure ou celles-ci ont un effet sur l'auditeur (Léon 1993). Cette discipline permet notamment d'établir un lien clair entre production et perception tout en contextualisant les facteurs à l'étude en ce qu'elle s'intéresse à la fois à l'**émetteur** et au **récepteur** du discours.

L'observation sur corpus de **régularités de production** (lexicales, prosodiques et syntaxiques) au sein d'une même situation de communication a permis d'opérationnaliser le concept au départ holistique de *phonostyle* développé par Léon (1993). Dans ses termes : « Certains types de discours, déclenchés par la mise en fonction (professionnelle), sont figés en **clichés mélodiques** reproduits toujours de la même manière. [...] Ce genre de stéréotype constitue ce que Fónagy appelle un 'cliché mélodique' » (Léon 1993:171).

L'application de **régularités prosodiques** ferait partie de la **mémoire auditive** des locuteurs et des auditeurs et refléterait la « **fonction identificatrice** » de la prosodie (Léon 1993:20-23). Comme mentionné ci-avant, outre le rôle structurant et les diverses fonctions linguistiques du rythme et de l'intonation, on reconnaît à la prosodie des fonctions dites « non linguistiques ». A côté de marqueurs prosodiques d'identification de style, on retrouve des **marqueurs prosodiques** d'émotions, d'accents régionaux, de sexe, etc. (Goldman et al. 2014:52). Un style est ainsi l'adaptation (*design*) de la production linguistique en fonction des attributs sociaux de l'auditoire (*audience*) (Bell

44. Et les paramètres situationnels y afférent tels que les degrés de préparation, d'interactivité, d'orientation médiatique, explicités de manière formelle au chapitre 3.

1984, 2001). Bien que le terme « disfluence » en tant que tel ne se retrouve que très rarement dans la littérature récente d'approches « conversationnelles » ou phonostylistiques, la mise en correspondance des données traitées (e.g. (dis)fluences en parole spontanée /vs/ en lecture oralisée) est indéniablement une source non négligeable de connaissances, notamment pour la présente recherche (Boula de Mareüil 2014).

Les phonostyles sont identifiés sur base de **corrélats acoustiques** et **situationnels**. Certaines situations de parole présentent des comportements prosodiques relativement similaires et homogènes. La parole radiophonique ou les lectures neutres à voix haute et le discours politique en sont d'illustres exemples. De ces régularités naît le *phonogenre,* défini comme « **une image acoustique typifiée** », donc plus ou moins conditionnée et homogène (Boula de Mareüil 2014 ; Goldman, Auchlin & Simon 2011). Dans le cadre de notre recherche, les phonogenres nous permettent d'appréhender les particularismes prosodiques définitoires d'un style de parole. On remarque là qu'il s'agit par ailleurs de discours hautement préparés et souvent exempts de marques formelles de (dis)fluence. Ces situations de communication sont « relativement normées et induisent des **comportements vocaux stéréotypés** » (Simon 2010:81).

Les phonostyles sont aussi et surtout une **réalité perceptive**. Dans le cadre d'une expérience perceptive, Goldman et al. (2014) ont demandé à des participants francophones et non francophones d'assigner à chaque échantillon de parole (filtré et non filtré) une pratique communicative spécifique. Les échantillons provenaient de chroniques radiophoniques didactiques, d'instructions d'itinéraires, de sermons religieux, de discours parlementaires, de commentaires sportifs en direct, de conversations informelles et de journaux parlés radiophoniques. Ces trois dernières situations sont celles les mieux identifiées par les participants francophones et non francophones (en condition filtrée et non filtrée). En somme, cette expérience perceptive met en valeur que les situations de communication démontrant le **moins de variation** acoustique (les phonostyles de production les plus homogènes), sont également les situations les plus **aisément reconnues** par les auditeurs. Dans le cadre de cette thèse, il s'agira notamment d'évaluer si l'homogénéité de production traduit une perception et une évaluation différentes des (dis)fluences à travers ces mêmes situations de parole.

4 Objectifs et canevas de la thèse

Le concept de fluence linguistique tel que parfois employé dans cette thèse ne doit pas effacer cette vision intégrée de la fluence : son origine cognitive ou son impact sociétal et par là même son lien éventuel avec les autres types de fluences définis ci-avant.

L'objectif de cette recherche doctorale est double. Nous aspirons d'abord à offrir une **description compréhensive** des dynamiques des marques locales de (dis)fluences au sein du flux discursif (formes acoustiques, cooccurrences, distributions prosodico-syntaxiques), et de leurs variations à travers les situations en français dans les productions des locuteurs francophones natifs. Nous proposons donc une **description contextualisée** de la production et de la perception des paramètres acoustiques internes aux (dis)fluences – en lien avec leur environnement **prosodique**, **syntaxique** et **situationnel** – pour les positionner sur le continuum de la fluence-disfluence. En d'autres termes, comment la fréquence d'usage et le contexte d'occurrences des (dis)fluences (paramètres prosodico-syntaxiques et sociopragmatiques) interagissent-ils en production et en perception ? Nous permettent-ils de catégoriser les (dis)fluences sur le *continuum* fluence-disfluence ?

L'**hypothèse principale** du projet ARC dans lequel s'inscrit cette thèse est que la fluence et la disfluence sont **deux faces du même phénomène** qui s'incarnent dans les (dis)fluences. Les formes et fonctions de ces marqueurs varient systématiquement en fonction de la langue, des locuteurs et des types de discours concernés. L'évaluation du degré de (dis)fluence dépendrait ainsi d'une **accumulation de marques** linguistiques (e.g. distribution et cooccurrences au sein des chaînes syntaxiques et prosodiques), diaphasiques (traits situationnels et situations de communication) voire diastratiques ou idiostylistiques (inhérentes au profil du locuteur). L'interprétation de ces (dis)fluences en fonction de ces indices linguistiques et contextuels permettrait de **désambiguïser leur rôle** en tant que partie d'un **dispositif communicatif fluent ou disfluent**.

Ensuite, nous désirons évaluer l'**impact** des (dis)fluences **au niveau perceptif** : nous postulons que c'est *in fine* le résultat des processus interprétatifs des locuteurs/auditeurs qui permet d'identifier le rôle de ces marques dans la dynamique de la (dis)fluence cognitive – individuelle comme sociale – et qui permettra d'accéder à leur désambiguïsation (ou tout du moins à une **évaluation graduelle** de leur caractère fluent /vs/ disfluent). Concrètement, l'hypothèse principale de cette thèse cherche à déterminer si l'évaluation fluente-disfluente d'une marque ou d'une production discursive par les auditeurs – estimées à partir de tests perceptifs – est associée ou non aux **fréquences objectives observées** dans les situations des corpus. L'intérêt majeur d'une issue positive serait, selon nous, notamment la validation ou l'invalidation d'une approche cognitive fréquentielle lors du traitement de données de corpus.

La première partie théorique et méthodologique de ce travail est développée des chapitres 1 à 3.

Le chapitre 1 prend en compte des **réalités sociolinguistiques et pragmatiques** sous-jacentes à la problématique de la fluence-disfluence, et défend une définition et une évaluation **intégrée et adaptative** de la fluence.

Le chapitre 2 présente les indices linguistiques mobilisés pour la définition de la (dis)-fluence. Il distingue entre autres les **indices linguistiques et extralinguistiques** émanant d'approches acoustiques, prosodiques, syntaxiques et situationnelles. Ces indices sont ensuite utilisés pour définir les (dis)fluences et leurs caractéristiques, à partir de la littérature francophone et/ou anglophone. Les marques considérées dans le cadre de cette recherche sont les pauses silencieuses, les pauses pleines et les allongements, et les répétitions à l'identique, que celles-ci apparaissent de manière isolée ou en cooccurrence. On y formalise les hypothèses générales de notre travail reposant sur une définition contextualisée des (dis)fluences. Cette contextualisation passe par la prise en compte premièrement de la relation entre les fréquences d'occurrence en production et la perception d'une continuité de l'axe syntagmatique et deuxièmement par la relation d'ajustement entre les attentes liées à la situation et la perception-évaluation d'une production discursive. Ce travail fait donc appel à un **questionnement méthodologique** plus large sur l'application de **méthodes mixtes** en sciences humaines.

Le chapitre 3 expose la structure de deux ensembles de **données orales empiriques** utilisés pour étayer notre propos. Ces deux *corpus* sont complémentaires, ils permettent l'analyse de la **variation intergenres et intralocuteurs**. On y décrit également le travail de segmentation, d'annotation et d'enrichissement des corpus. Une **description quantitative** de leur contenu donne au lecteur un premier aperçu des types de situations de parole à l'étude : leurs spécificités linguistiques et interactionnelles. Nul doute qu'un objectif méthodologique additionnel nous tenant particulièrement à cœur chapeaute cette problématique : l'exploration des liens possibles et de la complémentarité entre les pendants production et perception de la parole. L'originalité de nos études réside vraisemblablement dans leur enrichissement mutuel. Cette méthodologie au service de la **convergence des données** pose ainsi les principes d'une **pragmatique expérimentale**.

La seconde partie de la thèse expose les **résultats de nos recherches** : les chapitres 4 à 6 s'attachent respectivement à la description du fonctionnement des pauses silencieuses, des pauses pleines et des allongements et enfin des répétitions. Leur structure interne reflète le cheminement de notre **démarche méthodologique** : une analyse en production expose la variation situationnelle et individuelle observée pour chaque phénomène de nos corpus, tandis qu'une analyse en perception propose un design expérimental. Ce sont les résultats des analyses statistiques univariées ou multivariées, basés sur les **fréquences d'usages en production** qui motivent l'**expérimentation perceptive**. Les parties de discussions exposent enfin l'intérêt du retour aux données de production à partir des résultats de ces expériences.

Le chapitre 4 évalue la (dis)fluence des **pauses silencieuses**. On pose la question de **leur catégorisation** en fonction de critères acoustiques, distributionnels ainsi que de leurs environnements syntaxiques et prosodiques. A l'appui d'une analyse des différentes situations représentées dans les corpus, en raison d'une littérature abondante sur la pause silencieuse, on propose l'**application du degré de saillance et de rupture pour chaque pause du corpus**. La discussion évalue la validité de l'approche.

Le chapitre 5 développe l'évaluation de la (dis)fluence des **pauses pleines et des allongements**. Il s'interroge sur leur **détection, leur distribution et leur classification**. A l'instar du chapitre 4, on se questionne sur l'impact de leur environnement syntaxique et prosodique sur leur **saillance perceptive**. La mise en place d'expériences perceptives *online* et *offline* par des annotateurs experts et naïfs informe des paramètres acoustiques et linguistiques pertinents (i.e. mobilisés) lors des **tâches d'évaluation** des marqueurs sur le continuum de la fluence-disfluence.

Le chapitre 6 aborde une (dis)fluence structurée (composée d'un *reparandum*, d'un *interregnum* et d'un *reparans*) : les **répétitions à l'identique**. En intégrant les acquis des chapitres précédents, on examine les corrélats acoustiques et les distributions prosodiques des répétitions en fonction des **situations de parole** et des **types de structures** (nombre de composantes à la chaîne de (dis)fluence) envisagées. Enfin, deux expériences d'évaluation de la perception et du jugement des auditeurs sur les répétitions à l'identique (avec et sans termes d'édition, en contexte de production et hors contexte de production) permettent d'envisager un **retour vers les corpus**.

Le chapitre 7 revient sur les contributions théoriques et méthodologiques de cette thèse en en rappelant les objectifs et hypothèses. Ensuite, il résume et met en relation les résultats principaux de chaque chapitre. Enfin, en énonçant les bonnes pratiques à considérer lors de l'étude du continuum fluence-disfluence, il dessine les perspectives ultérieures de recherches.

Chapitre 2

Variations linguistiques des (dis)fluences

Le chapitre 2 parcourt la littérature descriptive et empirique des **indices linguistiques** au service de la modélisation des (dis)fluences (i.e. *description* et *évaluation* en *production* et en *perception*) (section 1). Son but est de réduire et de formaliser les paramètres des marques à observer pour leur catégorisation sur un continuum fluence-disfluence.

Nous commençons par présenter une approche empirique de l'observation acoustico-prosodique de la **forme interne** des (séquences de) (dis)fluences, telles que la durée et la fréquence fondamentale[1] (1.1). Puis, nous associons les observations de la forme interne à leurs contextes linguistiques : **(micro- et macro-) prosodiques** (1.2), **syntaxiques** (1.3), mais aussi **situationnels**. Nous formulons une proposition de cadre théorique d'**évaluation contextuelle** des différents types de fluences. A cette fin, nous retraçons la nature des relations entre les fréquences objectives d'occurrences de phénomènes, leur « saillance » en et hors contexte et les attentes formulées par le locuteur lors de la phase de conceptualisation du contexte cognitif immédiat (1.4)[2].

Dans le cadre de cette proposition également, nous (re-)définissons séparément chaque (dis)fluence à l'étude : *pause silencieuse* (2.1), *allongement* (2.2), *pause pleine* (2.3) et *répétition* (2.4). Cette description permet d'accorder à chaque marque son **set de propriétés**, de **tendances distributionnelles associé** aux contextes linguistiques et extralinguistiques définis antérieurement. Nous constaterons qu'une grande part des (dis)-fluences tend à apparaître **en combinaison** dans des contextes bien précis. Nous nous focaliserons ainsi sur la cooccurrence de ces séquences spécifiques de marqueurs (section 3). Enfin, après avoir décrit les ***schémas potentiels*** de chaîne de (dis)fluences sur

1. Comme le faisait déjà remarquer Guaïtella (1991), la mesure de l'intensité fait face à des réalités techniques et pragmatiques. Dans de nombreux enregistrements, on observe une modulation de l'intensité due à la distance entre les locuteurs et le microphone ou encore au formatage et à l'harmonisation effectuée en post-production. Ce dernier paramètre n'est donc pas abordé dans cette thèse.

2. Le croisement des différentes méthodologies d'application sur des données expérimentales et de corpus est développé au chapitre 3. Les méthodologies de désambiguïsation et de validation d'hypothèses spécifiques à chaque phénomène sont développées au chapitre qui leur est dédié.

base de la recherche empirique, expérimentale et sur corpus, nous reprenons les hypothèses validées et nous prolongeons la réflexion sur la **désambiguïsation** des marqueurs sur le continuum fluence-disfluence présenté. Ces hypothèses concernent notamment la forme prosodique interne et la relation aux contextes linguistique (cooccurrences, unités prosodico-syntaxiques), situationnel (pratique communicative, traits situationnels) et individuel (parole professionnelle, idiostyle, etc.) (section 4).

1 Indices linguistiques au service de la (dis)fluence

1.1 Approche acoustico-prosodique au sein de la zone de (dis)fluence

Au niveau acoustico-prosodique, la littérature sur les disfluences a émergé dès les années 80 des approches computationnelles. Avec pour objectif de rétablir la fluence verbale, les approches computationnelles caractérisent les indices acoustiques d'**interruptions du continuum langagier**. La rigueur de ces analyses a considérablement augmenté nos connaissances acoustiques comme prosodiques de ces phénomènes. Cette partie résume la littérature relative aux observations des formes acoustiques **internes à la structure des (dis)fluences** (1.1.1), leurs corrélats acoustiques sur les paramètres de durée et d'intonation séparément (1.1.2 et 1.1.3) ainsi que l'interaction même des paramètres prosodiques au sein d'une zone de (dis)fluence (1.1.4).

1.1.1 Quelles zones de "turbulences" : *reparandum*, point d'interruption, *reparans* ?

Les études distinguent les corrélats acoustiques selon la zone de (dis)fluence au sein de laquelle ils sont produits· : le *reparandum*, la *phase d'édition* (ou tout phénomène autour du point d'interruption) et le *reparans*. Hindle (1983) fut un des premiers à émettre l'hypothèse de la présence d'un **signal phonétique** d'édition explicite permettant aux systèmes informatiques (et parfois aux locuteurs) de détecter et d'identifier les *dis*fluences. Les avis semblent toutefois diverger sur la localisation de ce signal.

La majorité des systèmes automatiques s'accorde sur le fait que la partie de la (dis)fluence pour laquelle l'identification et la précision sont les plus élevées est le ***reparans*** (voir notamment Georgila, Wang & Gratch 2010). Pourtant, peu de corrélats acoustiques de la (dis)fluence sont observés dans le *reparans* par Shriberg, Stolcke & Baron (2001:166), à l'exception de certains types de *reparans* pour lesquels ils relèvent parfois un **contraste prosodique** avec le *reparandum*. Levelt & Cutler (1983) associent ce marquage à un **changement de contenu** entre les deux segments de la zone de (dis)fluence. Ils observent un marquage du *reparans* par rapport au *reparandum* dans 50% des cas pour les *dis*fluences entrainant des erreurs, contre seulement 20% pour les (dis)fluences impliquant des élaborations.

Reparandum ou *reparans* ? Certains chercheurs postulent que les indices acoustico-prosodiques plus pertinents à l'identification des (dis)fluences sont **autour du point d'interruption** et non au sein du *reparans* ou du *reparandum*. On observe ainsi un marquage par une pause interlexicale et plutôt à la fin du *reparandum* ou à l'attaque du *reparans*. Adoptant comme sources principales de détection le point d'interruption, les résultats de Nakatani & Hirschberg (1994) témoignent en effet de taux de précision et de détection automatique des (dis)fluences honorables de respectivement 91% et 86%. Shriberg (1999) et Shriberg, Stolcke & Baron (2001) ont également analysé les paramètres de durée, de fréquence fondamentale, de qualité vocale et d'énergie (amplitude) à ce même point de la zone de (dis)fluence. Les valeurs autour du point d'interruption ont été comparées au reste de la structure (début, milieu et fin du *reparandum*, phase d'édition, et *reparans*). Les auteurs rapportent des perturbations particulièrement **au point** et autour du **point d'interruption**, de manière plus significative (à la détection automatique) que dans le *reparandum* ou *reparans*. Ils confirment ainsi les observations de Nakatani & Hirschberg (1994).

Ces études constituent un atout méthodologique indéniable pour la présente recherche. Elles aident à sélectionner le nombre et les types de corrélats acoustiques à observer ainsi que les **lieux de mesure** les plus significatifs pour l'étude de la forme interne des (dis)fluences[3]. Mais quels sont exactement ces indices de « perturbations acoustiques » ?

Les auteurs observent des patrons inhabituels près du point d'interruption de la (dis)fluence. Ils notent souvent un **allongement** (vocalique, consonantique ou syllabique) et des phénomènes de **coarticulation** (i.e. la distorsion d'un phonème due à son environnement phonétique), de **glottalisation** ou de **laryngalisation**. Cette dernière est associée couramment à l'impression de « voix craquée » et est détectée par l'observation d'une courbe de F0 perturbée, résultat d'un resserrement des cartilages aryténoïdes dans le larynx. Nous proposons de nous attarder plus longuement sur les corrélats acoustiques de durée et d'intonation au sein des différentes zones de (dis)fluences.

1.1.2 Corrélats acoustiques de durée des zones de (dis)fluence

La **durée** est présentée comme le paramètre le plus déterminant dans la production (et la perception) des (dis)fluences, et ce quelle que soit la partie de (dis)fluence d'intérêt (*reparandum, reparans*, etc.). La durée est bien entendu avant tout un paramètre d'**organisation temporelle**, et ce à plusieurs niveaux.

3. Cependant, il ne faut pas perdre de vue que l'observation formelle du signal ne permet pas de savoir si ces marquages acoustiques sont utilisés par les locuteurs, et font varier ou non les interprétations en fonction des contextes d'actualisation (linguistiques et situationnels). A ce titre, il est important de noter que ces recherches se sont souvent focalisées sur des situations d'observations spécifiques ou peu écologiques (i.e. lectures, discours politiques, situations expérimentales) (Adda-Decker et al. 2008 ; Boula de Mareüil et al. 2013 ; Bouraoui 2008 ; Bouraoui & Vigouroux 2006 ; Broen & Siegel 1972 ; de Medeiros 2014 ; Vasilescu, Candea & Adda-Decker 2004).

En français, bien que ce paramètre ait avant tout une **fonction grammaticale** (lors de l'actualisation de l'accent démarcatif), il est également influencé par des facteurs cognitifs (para ou extralinguistiques). En effet, les chercheurs s'accordent sur le fait que l'allongement reflète un **temps d'encodage** plus long, associé à un ralentissement (ponctuel ou non) du rythme (Candea 2000a:20 ; Shriberg, Stolcke & Baron 2001:160). Duez (1991) et Carton (1999:40) considèrent par exemple que c'est la durée qui permet de différencier les *euh* « disfluents », des *euh* « brefs » ou des *euh d'appui*[4], et donc des phénomènes de structuration /vs/ de phonostylistique.

Quel type de blocage cognitif a été indiqué en surface par quelle (dis)fluence ? Comme évoqué, la durée est souvent utilisée comme fenêtre sur nos processus cognitifs. A ce titre, au lieu de tenter de reconstruire à l'exact le **cheminement cognitif** du locuteur, ce qui semble parfois impossible, Candea (2000a:88) propose de se baser essentiellement sur la durée des « marques du travail de formulation » pour savoir quelles marques interagissent avec des troubles de *conceptualisation* /vs/ de *programmation* lors du traitement de la parole.

L'auteure part de la prémisse selon laquelle « le travail de conceptualisation des unités se situe en amont du travail de programmation et d'agencement des unités linguistiques » (Candea 2000a:88). Elle formule alors l'hypothèse selon laquelle les **processus de conceptualisation** ont besoin « de plus de durée que le traitement des unités déjà sélectionnées » (i.e. de programmation). Elle associe aux problèmes de conceptualisation, un « **malaise dans la pensée** », et donc un temps d'écoulement plus long entre la production d'une (dis)fluence et le début d'une unité cible que pour un problème de programmation (i.e. « **malaise dans la formulation** ») (voir section 2.3.3)[5]. Quel que soit le fondement réel de la relation entre la « durée moyenne d'un phénomène » et le « temps de formulation », son analyse dans le cadre des (dis)fluences permet d'enchaîner sur le type de questionnements relié à la durée. Elle amène à s'interroger sur l'existence de différences significatives entre les quatre (dis)fluences étudiées (*pauses silencieuses, pleines, allongements,* et *répétitions*).

Toutes les (dis)fluences ne sont pas associées à des « hésitations *dis*fluentes » – parfois, ces marques reflètent justement la **« détection d'une erreur » par un locuteur**. Cette détection est rapide en elle-même (et se reflète dans l'actualisation de la marque). En conséquence de quoi, comme le font remarquer Shriberg, Stolcke & Baron (2001:162,164), toutes les marques de (dis)fluences ne sont pas obligatoirement associées à un allongement. On observe même souvent un **raccourcissement temporel** du *reparandum* et/ou du *reparans*, ainsi que parfois des pauses pleines proportionnellement courtes (i.e. < 0.2 sec.).

4. Voir section 2.3 de ce chapitre.

5. Bien entendu cette opposition à priori dichotomique est tout de suite réévaluée en terme de continuum de sorte que « certaines marques de TdF [travail de formulation] seraient ainsi nettement plus rapprochées d'un pôle que de l'autre et certaines autres se situeraient en revanche quelque part dans le continuum entre ces deux pôles, sans qu'il nous soit possible de déterminer avec précision leur(s) cause(s) » (Candea 2000a:88).

Au niveau acoustique, l'étude des pauses silencieuses isolées mobilise uniquement le paramètre de durée [6]. C'est donc davantage l'influence du contexte d'apparition des pauses silencieuses au regard de leur durée qui intéressera les linguistes, notamment au niveau de leur **positionnement syntaxique** et/ou de leur apparition au sein d'une **phase d'édition**. On notera ici que les paramètres de contextualisation mobilisés diffèrent en fonction de la marque envisagée.

L'utilisation par nombre d'auteurs du paramètre de la durée est également due au fait que la durée a l'avantage d'être à priori « **objectivement mesurable** » [7]. Les variables en relation avec la durée ont été formalisées par les études fondatrices de Grosjean & Deschamps (1972-1975) sur l'organisation temporelle du français [8]. Les mesures de durées comprennent notamment : le rapport temps d'articulation / temps de locution, le temps total de pause [9], la durée et le nombre de pauses silencieuses ainsi que des variables dites « secondaires » (i.e. certains incluent dans les mesures de durées les répétitions, les pauses « sonores » tels les *euh* et les *allongements* et les faux départs).

Dans le cas des « *(dis)fluences composées* » [10], le paramètre temporel le plus souvent utilisé est celui du rapport entre **la durée relative du *reparandum*** et **la durée relative du *reparans***. Comme déjà précisé, la durée est souvent (mais non nécessairement) marquée par un **allongement** de la rime ou syllabe précédant directement le point d'interruption. Si c'est le cas le plus observé par les chercheurs pour le *reparandum*, le *reparans* a en général des durées similaires aux caractéristiques des segments exempts de (dis)fluence (Duez 1993 ; Lickley 1994 ; Shriberg 1999 ; Candea 2000a:190-191 ; Shriberg, Stolcke & Baron 2001 ; Henry & Pallaud 2003).

En production, les (dis)fluences ont, comme vu ci-avant, des variations acoustiques de durée autour du point d'interruption et donc dans la ***phase d'édition***. Pour rappel, selon Levelt (1983), les termes d'édition sont le résultat du **besoin de planification** du *reparans* lors d'auto-interruption ; les phases d'édition sont alors associées avec **l'interruption de tous les niveaux de processus de production de la parole**. Lorsque Shriberg (1999) compare la longueur des pauses pleines et des voyelles au sein de contextes disfluents (phase d'édition) et en dehors, elle constate également leur

6. La manière de calculer cette durée est pourtant sujette à de nombreux débats (voir section 2.1 de ce chapitre et au chapitre 4 section 1).

7. Au niveau méthodologique, il existe cependant plusieurs façons de normaliser la durée des *reparandum* et des *reparans* (e.g. valeurs absolues /vs/ relatives).

8. Sans description de l'organisation suprasegmentale de la langue dans son ensemble. Les auteurs n'ont pas ici un objectif de modélisation de la langue, mais d'observation exploratoire des variables temporelles de la parole. De même, au sein de la littérature francophone, le modèle de Rossi (1999) ne fait pas mention des marques du travail de formulation – pour une critique de ce modèle nous renvoyons à Candea (2000a:64-68). Pour rappel, les études de Grosjean & Deschamps sont menées dans une démarche contrastive avec l'anglais, à partir d'un corpus de 30 interviews, non lues et radiophoniques, et de 20 enregistrements portant sur une tâche de description de dessins humoristiques (Grosjean & Deschamps 1972, 1973).

9. Chez Grosjean et Deschamps, ce temps total de pause n'inclut pas les pauses pleines, contrairement aux études de Duez (1991).

10. Une *(dis)fluence composée* contient un *reparandum*, un point d'interruption, une éventuelle phase d'édition et un *reparans*, signal de retour à la fluence (voir chapitre 3 section 2.1 et chapitre 1 section 1.2.4).

allongement relatif, surtout lorsque les pauses pleines **cooccurrent avec d'autres (dis)fluences**. Plus précisément, alors qu'en dehors de dispositif (dis)fluent ces marques durent entre 0.1 et 0.2 sec., elles peuvent aller jusqu'à 1 sec. en contexte (dis)fluent (Shriberg 1999:622).

En perception, la durée des marques ainsi que leur contexte d'occurrence prosodique et syntaxique influencent significativement sur la perception. Par exemple, selon Candea (2000a:26) les *euh* et les allongements vocaliques supérieurs au seuil de 500 ms. sont très majoritairement perçus (contrairement à ceux durant 200 à 500 ms.). Encore à titre d'exemple, les **pauses subséquentes** à des phénomènes de (dis)fluences seraient, elles, moins perçues que les **pauses isolées**. Mais on remarque qu'il s'agit là d'une confluence de paramètres prosodiques, syntaxiques et structuraux, sur lesquels nous revenons en détail en section 3 de ce chapitre.

Par ailleurs, de manière générale certaines variations temporelles émanent non seulement de l'impact des (dis)fluences mais aussi des **contraintes situationnelles**. Ainsi, les variables temporelles ont un rôle dans l'identification du **style de parole** : « [e]lles permettent de distinguer les messages produits en public des messages produits en privé, de définir l'objectif du message (informer, persuader...) et d'évaluer le statut social et la distance physique qui séparent les participants » (Duez 1997:283 ; citée par Bothuyne 2013:58). Nous le prenons en compte lors de l'analyse des durées relatives en fonction des situations.

1.1.3 Corrélats acoustiques de l'intonation des zones de (dis)fluence

La description prosodique intrinsèque de la (dis)fluence ne se résume pas à l'étude des paramètres de durée. Rares sont les auteurs intégrant les (dis)fluences dans leurs modèles de la parole et à fortiori, à décrire leur **comportement intonatif**.

Morel & Danon-Boileau (1998) dans leur modèle intonatif, prennent en compte la **valeur iconique** de base de chacun des paramètres prosodiques (i.e. F0, durée, intensité et présence de pauses) auxquels ils associent des aspects **linguistiquement conventionnalisés**[11]. Les auteurs suggèrent que les (dis)fluences s'insèrent dans des parties d'énoncé **en dehors de la co-énonciation** ; elles exhiberaient des caractéristiques prosodiques d'une **attitude « égocentrée »**. Morel & Danon-Boileau avancent que les disfluences sont en général situées en plage basse d'un registre intonatif (et intensité forte pour maintenir la parole malgré une rupture discursive ponctuelle). La ***hauteur relative de la F0*** serait pertinente, significative au niveau énonciatif : elle permettrait de détecter la *dis*fluence de par sa valeur iconique.

11. Nous ne revenons pas en détail sur le modèle énonciatif et intono-syntaxique de Morel & Danon-Boileau (1998). Brièvement la terminologie associée utilisée ci-dessus se définit comme suit : *le paragraphe* désigne une macro-unité de l'oral aux composants hiérarchisés en groupes prosodiques. Il se compose d'un enchaînement d'un ou plusieurs préambule(s) (à valeur thématique ou modale), rhème(s) (partie de l'énoncé présentée comme nouvelle par l'énonciateur) et *postrhème(s)* (constituant facultatif posé après le *rhème*) (Morel & Danon-Boileau 1998:20-30).

Guaïtella (1991) observe que les différences microprosodiques entre les segments plus ou moins disfluents se font par l'observation de leurs « contours mélodiques » (Guaïtella 1991:114 ; citée par Candea 2000a:93). Ces contours, appelés des **« intonations d'hésitation »**, suivent également une **ligne de déclinaison** purement physiologique : ils ne sont ainsi jamais montants, souvent **légèrement descendants** et à tendance plate bien que rarement parfaitement plats. De manière générale, la littérature s'accorde sur le fait que le **niveau relativement bas de la F0** rentre dans la définition même d'une *dis*fluence.

Shriberg (1999:621), Moniz, Trancoso & Mata (2009), Belz & Reichel (2015) corroborent ces résultats. En mesurant la fréquence fondamentale en début, centre et fin de pause pleine et l'amplitude du mouvement entre ces points, ils observent que leur intonation tend à reproduire une **courbe descendante**, linéaire et droite. Malgré ces tendances générales, les auteurs remarquent également que la valeur de déclinaison de la marque est influencée par son contexte d'apparition (notamment si elle est précédée ou non par une pause), i.e. l'intonation des marques est également **liée à la clause** à l'intérieur de laquelle celles-ci s'actualisent (O'Shaughnessy 1992 ; Shriberg 1999 ; Shriberg & Lickley 1993) [12]. Ces hypothèses (i) de registre intonatif relativement bas et (ii) de contours intonatifs plats ont été corroborées par d'autres études [13].

1.1.4 Interactions des paramètres prosodiques et des zones de (dis)fluence

Bien qu'il soit possible de présenter les corrélats acoustiques de la durée et de l'intonation sur les différentes zones de disfluences, il ne faut pas négliger (i) l'importance de l'**interaction des paramètres prosodiques** entre eux et (ii) la pertinence des relations prosodiques entre le *reparandum* et le *reparans* (Cole et al. 2005 ; Savova & Bachenko 2003 ; Shriberg 1995, 1999).

Au niveau des (dis)fluences simples comme composées, de nombreux auteurs ont posé l'hypothèse d'une interaction des paramètres entre eux. L'interaction permet d'identifier et de catégoriser les (dis)fluences. En français, Guaïtella (1991), citée par Candea (2000a:92), supposait déjà que la **durée** des (dis)fluences (ici particulièrement dans l'étude du *euh* et de l'allongement), est en corrélation avec la **forme du contour mélodique** réalisé.

L'étude de (dis)fluences composées d'un *reparandum* et d'un *reparans* (e.g. répétitions à l'identique, substitutions) peut être réalisée par la combinaison d'indices acoustico-prosodiques tels que la durée et la variabilité du pitch, et la présence de pause(s) entre les

12. Voir section 3 de ce chapitre.

13. C'est à ce titre qu'un parallèle entre les caractéristiques intonatives des marques du travail de formulation et celles de l'incise est souvent réalisé. L'incise est elle aussi souvent complètement plate et basse ; lorsque plus longue, elle peut être légèrement modulée et, si elle est enchaînée à un autre constituant du même énoncé, elle présente une petite montée finale qui rappelle très fortement l'amorce de fin de (dis)fluence (augmentation de la F0) relevée par Guaïtella (1991) et rappelée en section 2.3.1 de ce chapitre.

deux parties de la (dis)fluence (Levow 2004). Une méthodologie proposée par Levelt & Cutler (1983) invite à décrire les types de relations entre ces trois paramètres en trois questions :

- *Que fixe le reparans* ? Levelt & Cutler (1983) suggèrent ainsi que le *reparans* corrige la prononciation, la sémantique et/ou la syntaxe. Selon eux, les « erreurs » phonétiques et syntaxiques n'entraînent pas de marquage prosodique, contrairement aux erreurs sémantiques [14].

- *Y-a-t-il un fragment après le reparandum* ? Selon les auteurs, la présence d'un fragment après le *reparandum* (e.g.« *mam/ maman* » /vs/ « *maman/ euh non * maman* ») est symptomatique d'une réelle volonté de correction et induit par conséquent un positionnement conceptuel singulier [15].

- *Quel est le « statut » du reparandum par rapport au reparans* ? Selon Levelt & Cutler (1983), la forme prosodique relative du *reparandum* et du *reparans* permet déjà de distinguer différents degrés et types de *dis*fluences : erreur de prononciation, répétition, erreur sémantique et erreur syntaxique, etc.

Si la plupart des corrélats prosodiques mentionnés s'actualisent au sein du *reparandum* ou de la phase d'édition, Shriberg (1999) note globalement **une préservation des patrons intonatifs** et des relations locales de l'intonation entre le *reparandum* et le *reparans* malgré la présence d'allongements et quelques rares cas d'**emphase contrastive** dans l'élément du *reparans* (marqué intonativement) lors d'un changement de contenu.

Dans la même idée d'une approche acoustique pluriparamétrique de la (dis)fluence, Savova & Bachenko (2003) ont étudié empiriquement **la durée** et l'**intonation** de quatre types de (dis)fluence : *répétition à l'identique, substitution à l'identique, répétition avec modulation, répétition avec insertion* [16].

Plus précisément, les durées sont analysées pour le *reparandum*, la phase d'édition et le *reparans* [17]. L'étude se focalise donc sur la relation entre l'***onset*** et l'***offset* du *reparandum* et du *reparans***. Voici un résumé des riches résultats de leur étude [18].

14. En raison du possible marquage prosodique, on ne peut pas simplement enlever le *reparandum* et la phase d'édition pour laisser place à un « contenu fluent ».

15. Voir la distinction entre « *backward* » /vs/ « *forward* » *looking disfluency* au chapitre 1 section 2.3.3.

16. Un cinquième type de (dis)fluence avait été envisagé par les auteurs, la « répétition avec délétion » mais leur faible occurrence n'a pas permis de conclure leur analyse prosodique. En raison de la rigueur des analyses prosodiques et d'une classification étendue de la proposition de Levelt & Cutler (1983), seules les répétitions et les substitutions exactes ont été effectivement exploitées dans les analyses des auteurs. L'étude exclut également les cas de substitution avec délétion (e.g. « il y a cinq corrections sept en fait »).

17. Ce choix se base en partie sur une étude de Lickley (1994) ayant mis en lumière que la fin du premier mot correct après une (dis)fluence suffisait à elle seule aux humains pour identifier la *dis*fluence.

18. Il s'agit d'analyses effectuées par extraction automatique et correction manuelle pour les valeurs doublement plus hautes ou basses, avec exclusions des « *vocal fries* » (Olive, Greenwood & Coleman 1993) et élimination des sites avec des milieux de mot contenant des fricatives sourdes (valeurs de F0 trop hautes). Les contours intonatifs sont représentés en séquence (« *low* » /vs/ « *high* ») relativement à la F0 du mot.

– *L'hypothèse combinatoire*, fondée sur les travaux d'HINDLE (1983), est partiellement confirmée par les auteurs. Selon celle-ci, **l'intonation (et non la durée) permet de distinguer plusieurs (dis)fluences entre elles**, notamment dans les cas des substitutions /vs/ répétitions exactes. Les répétitions sont caractérisées par des patrons intonatifs **en parallèle** et des équivalences observées entre le *reparandum* et le *reparans*. Elles s'opposent aux substitutions avec modulation(s) pour lesquelles les auteurs n'observent qu'un patron intonatif **contrastif** avec un ton final abaissé. Par ailleurs, SAVOVA & BACHENKO (2003) remarquent que le *reparans* de la substitution a une étendue de F0 supérieure à celle du *reparandum* de la substitution ou du *reparans* de la répétition (ce qui peut, dans le cadre des théories PIERREHUMBERT (1980:1), s'interpréter comme une construction prosodiquement « marquée »).

– *L'hypothèse compensatoire* est partiellement confirmée. Celle-ci teste le fait que deux paramètres prosodiques ne sont jamais tous exploités à leur maximum, c'est-à-dire **conjointement** « **sur-emphatiques** », comme dans le cas d'une durée syllabique très importante combinée à une fréquence fondamentale agitée et dans les extrêmes. Selon les auteurs, cette mobilisation préférentielle de l'un ou l'autre paramètre permet d'éviter que l'auditeur ne détecte de manière erronée l'indice d'une **frontière discursive** lors de la production d'une (dis)fluence (i.e. un contour intonatif terminal ou conclusif est dans de nombreuses langues marqué par ces deux paramètres).

– *L'hypothèse « Règle du Reset élargi »*, inspirée de PIKE (1945), est partiellement validée. Selon celle-ci, l'*onset* de la F0 du *reparans* dépend de l'*onset* et de l'*offset* du *reparandum*. Les résultats montrent une corrélation forte entre ces deux points de la (dis)fluence (r = 0.64-0.99), mais bien plus faible lorsque les *reparans* sont entourés de silences[19]. Dans ce cas, ils observent que les éléments du *reparans* commencent à des **valeurs plus basses** que l'*onset* du *reparandum*. De même, ils confirment une forte corrélation entre l'*offset* du *reparandum* et les valeurs d'*onset* comme d'*offset* du *reparans*, indice qu'une dépendance existe au-delà de l'*onset* du *reparandum*[20].

Signalons que les auteurs désiraient tester *l'hypothèse du genre discursif* selon laquelle au travers des genres les (dis)fluences ne sont pas marquées par les mêmes traits prosodiques. Le corpus exploité ne leur a pas permis de confirmer la **variabilité intergenres**.

La structure des (dis)fluences pour laquelle l'identification et la précision sont les plus élevées est le ***reparans*** dans une majorité des systèmes. Cependant, nombre de travaux ont montré l'intérêt de **modélisations multifactorielles** (dites aussi *approches hybrides*). Celles-ci utilisent des modèles acoustiques prenant en compte les différentes parties de la (dis)fluence (paramètres prosodiques en début, au milieu et en fin des zones). Les modèles acoustico-prosodiques sont fiables, toutefois ils ne donnent qu'une **vision partielle** des **mécanismes de structuration du discours**. Les approches hybrides

19. Une des failles méthodologiques est que cette corrélation a besoin d'être comparée aux relations des *onsets* et *offsets* pour les segments fluents.

20. Cette dernière proposition fait par ailleurs écho à une observation de SHRIBERG (1995) selon laquelle lorsque les locuteurs modifient la durée du mot répété, ils préservent les patrons intonatifs et les relations intonatives locales des segments.

sont bien entendu dépendantes de la richesse des corpus et de leurs annotations. On associe en général des informations lexicales et morphosyntaxiques (e.g. *partie du discours, préfixe/suffixe, patrons graphiques, ponctuation éventuelle, mots outils /vs/ pleins*) ou syntaxiques (marqueurs discursifs lexicaux) et prosodiques (patrons intonatifs, phrases intonatives) plus élargies[21]. Ces modèles statistiques probabilistes favorisent les interactions entre variables voisines (Liu et al. 2006), entraînant des résultats bien supérieurs à ceux obtenus en traitant ces phénomènes de manière indépendante (Eklund 2004:146). Les sections suivantes s'attachent précisément à l'**explicitation des variables prosodiques et syntaxiques**.

1.2 Approche prosodique du contexte d'actualisation des (dis)fluences

L'évaluation de la (dis)fluence peut passer par la description et l'analyse acoustique ou microprosodique intrinsèque aux zones de la marque[22]. Toutefois, une évaluation contextuelle invite également à envisager l'influence de la marque sur la production et la perception plus globale du **rythme et de l'intonation**, c'est-à-dire le contexte prosodique plus large dans lequel ces (dis)fluences s'actualisent. Comment définir (identifier et catégoriser) le contexte prosodique d'actualisation d'une (dis)fluence ? Quand est-ce qu'une *rupture prosodique* entraînée par une marque équivaut à une **rupture discursive** ?[23]

Des questionnements subsidaires émergent. Premièrement, si plusieurs manières de **segmenter le flux de parole** en unités prosodiques existent, lesquelles adopter[24] ? Ensuite, étant donné l'impact des (dis)fluences sur la rupture ou la continuité des segmentations prosodiques elles-mêmes, n'est-il pas primordial de décrire l'**incidence de ces zones de (dis)fluences sur la matérialité prosodique** du discours afin de les prendre en compte dans la segmentation du flux en unités prosodiques ?

Duez (1991) définit le contexte prosodique d'actualisation d'une (dis)fluence selon un critère formel unique, en fonction du fait que la marque soit ou non directement **suivie** d'une *séquence sonore* (et optionnellement suivie d'une *pause silencieuse*[25]).

21. Les études de Baron, Shriberg & Stolcke (2002), Christodoulides, Avanzi & Goldman (2014), Dutrey et al. (2014), Georgila (2009), Heeman & Allen (1999), Moniz, Trancoso & Mata (2009), Shriberg (1995), Shriberg, Bates & Stolcke (1997), Shriberg, Stolcke & Baron (2001) et Stolcke et al. (1998) constituent des exemples d'approches hybrides.

22. Voir section 1.1 de ce chapitre.

23. Première question subsidiaire : cette rupture discursive issue de la rupture prosodique est-elle suffisante en soi pour catégoriser cette marque ou ce *schéma* de dispositif communicatif *dis*fluent en contexte ?

24. Di Cristo (2013) en différencie six en fonction de leurs contours prosodiques et des niveaux de frontières pris en compte (e.g. unité tonale, mot prosodique, unité intonative, syntagme intonatif, paratons mineur et majeur). Nous exposons les méthodes d'applications concrètes de quelques-unes au chapitre 3 section 3.2.

25. Duez (1991) emploie 10 auditeurs et un corpus d'environ 19 minutes produit par des locuteurs différents, Duez (1995) en emploie 25 et un corpus de 100 énoncés (68 hésitations au total), tandis que Candea (2000a) utilise un corpus de 30 auditeurs.

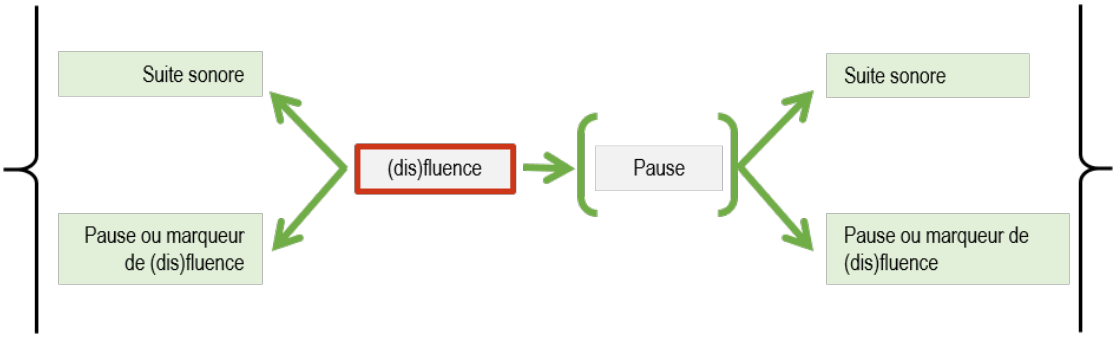


FIGURE 2.1 – Définition minimale du contexte syntagmatique local des (dis)fluences.

Cet exemple de **définition bivariée** du « contexte prosodique » d'actualisation d'une (dis)fluence illustre le point de départ de nombreuses approches de la littérature francophone, qui a par ailleurs produit maintes études fondatrices dans le domaine de la variation rythmique de la fluence verbale. Hormis certaines exceptions, la majorité de la littérature s'attache aux (dis)fluences **isolées**, éventuellement à la cooccurrence de pause(s) avec un ou deux autre(s) marqueur(s). Elle est souvent exempte de l'observation d'autres (dis)fluences composées ou d'**analyses avancées du système intonatif** – en raison des limitations techniques. Bien que ces limites réduisent la richesse des sources de désambiguïsation contextuelle de la (dis)fluence, ces recherches offrent une base riche propice à la reproduction de résultats sur des corpus plus étendus[26].

DUEZ (1991) apporte une distinction et une catégorisation des (dis)fluences (environnement direct *pause silencieuse* /vs/ *séquence sonore* autour du marqueur) qui, à défaut d'être la plus fine, est attrayante vu son apparente simplicité et sa transparence. Une « **séquence sonore** » est la séquence comprise entre deux pauses silencieuses, excluant les pauses remplies, les faux départs et les répétitions. Cependant, dans ces catégories peuvent pêle-mêle s'actualiser des éléments positionnés en début de constituant, d'autres (dis)fluences (substitutions, autocorrections), ou une frontière finale de constituant (e.g. marquée par une montée intonative qui se rapproche alors de la pause silencieuse marquant la présence d'une frontière plus que la continuité avec une « séquence sonore »).

Au sein des « pauses silencieuses », confondre par exemple la présence de ***pauses de structure*** marquant une frontière discursive (intermédiaire ou finale) et/ou ***pauses interactionnelles*** (tours de parole) avec des ***pauses de phase d'édition*** d'une (dis)fluence, semble peu pertinent dans le cadre de recherches plus étendues (CANDEA 2000a:282). Cette catégorisation pourrait ainsi amener à **distinguer** fortement des phénomènes relativement proches ou inversement à **confondre** une seule catégorisation pour des configurations fort différentes si on ne distingue que les quatre configurations

26. Citons notamment en exemple les études de PASDELOUP (1992, 2005), DUEZ (1982, 1993, 1998, 1999, 2001a,b,c) et CANDEA (CANDEA 2000a,b,c ; CANDEA, VASILESCU & ADDA-DECKER 2005 ; CLERC-RENAUD, VASILESCU, CANDEA & ADDA-DECKER 2004 ; VASILESCU, CANDEA, ADDA-DECKER et al. 2005).

suivantes, illustrées en exemple 6 : (1) pause silencieuse + *hésitation* + pause silencieuse, (2) séquence sonore + *hésitation* + séquence sonore, (3) séquence sonore + *hésitation* + pause silencieuse et (4) pause silencieuse + *hésitation* + séquence sonore.

Exemple (6) – issu de Candea (2000a:281-282)

a. essaie **de de:↗ /30/ de** le mettre avec une↗ toute↗ petite voix /550/ (Boucle, Prof, lignes 45-46)

b. ET: euh::: les fils deVAIENT la lancer:↗ **et:** LA où tomberait la plume↗ /40/ (Père fils, lignes 226-227)

c. la maman ours↗ qui dit la même chose↗ /80/ **et euh:** le petit BEbé:↗ il dit:↗ /35/ (Boucle, lignes 52-53)

/versus/

d. sui↗vre le petit: **euh::** le petit lutin↘ /75/ et: euh: le petit lutin↗ le soir↗ il fit un grand feu↗ et chanta↗ une chanson /25/ (Nain, lignes 366-367)

e. euh: après:↘ **euh:** /480/ chuchotements (rires) /300/ (Cordo, ligne 217)

Le modèle intonatif de Morel & Danon-Boileau (1998) est basé sur l'articulation du discours en *paragraphe, thème, rhème* et *postrhème*. Les auteurs font l'hypothèse qu'à chaque (dis)fluence correspondent des distributions et des **fonctions énonciatives** différentes. Cette proposition a l'avantage d'insérer les (dis)fluences dans un modèle de perception du rythme et de l'intonation en français.

Candea (2000a:282) enrichit les propositions de contextualisation(s) prosodique(s) des (dis)fluences. En appliquant le modèle de Morel & Danon-Boileau (1998) à l'analyse fonctionnelle des « marques du travail de formulation », l'auteure met en parallèle la distribution fonctionnelle des *marques de frontières intono-syntaxiques* avec l'actualisation de (dis)fluences. De plus, elle oppose en amont les occurrences simples **isolées** des **cooccurrences** de marqueurs. Elle remarque que les pauses pleines sont le plus souvent placées en **début de constituant** et **accompagnées d'un « ligateur »** (exemple 6). Elle associe cette cooccurrence à un *effet énonciatif* de **maintien** dans le plan discursif d'une **suite attendue**. Ce schéma marque en quelque sorte l'*absence de rupture*. Cette interprétation est appuyée par le fait que les configurations [pause pleine + ligateur] en position intraconstituant et en fin de constituant sont représentées mais respectivement peu fréquentes et extrêmement rares dans les corpus observés.

Heeman & Allen (1999) traitent de la modélisation des (dis)fluences en contexte, mais ici dans une approche computationnelle. Ils s'intéressent plus spécifiquement à la problématique de la reconnaissance des phrases intonatives dans la parole. Les phrases intonatives, les (dis)fluences et les « marqueurs de discours » partagent des **fonctions de structurations** linguistiques (marqueurs potentiels de frontières) et par là même de caractéristiques distributionnelles et prosodiques (e.g. tendance à être succédé par une pause, présence d'allongement syllabique en fin de segment, etc.).

Exemple (7) – issu et traduit de Heeman & Allen (1999:4-7)[27]

a. < you can carry them both on > | * | < two both on > the same engine
< tu peux tous les deux les porter sur > | * | < les deux sur > le même engin (d92a-1.2 utt40)

b. < we have the orange juice in two > | * oh | < how many did we need >
< on a du jus d'orange en deux > | * oh | < combien on en avait besoin > (d92a-1.3 utt75)

c. < the engine can take as many > | * um | < it can take > up to three loaded boxcars <la machine peut faire autant que > | * euh | < elle peut prendre > deux à trois cargaisons (d93-14.3 utt27)

Les auteurs proposent donc de **détecter conjointement** les **phrases intonatives**, **marqueurs de discours** et les **(dis)fluences** en raison précisément de leur entrelacement. Leurs résultats sont riches. Comme beaucoup, ils notent qu'en vertu de leurs similarités distributionnelles, les marqueurs de discours et les (dis)fluences tendent à être présents simultanément. Ils observent également qu'ils ont tendance à être positionnés en **début de segment**, i.e. attaque d'unité intonative majeure. La cooccurrence s'observe particulièrement dans le cas où la pause et le marqueur de discours apparaissent **au sein d'une phase d'édition**. Etant donné qu'ils signalent l'« erreur », ils laissent le temps de re-conceptualisation à l'auditeur et introduisent la reformulation pour l'auditeur (Hindle 1983 ; Levelt 1983). Les résultats du détecteur de frontière d'Heeman & Allen (1999) livraient sans conteste des chiffres robustes, avec une détection et une précision de 72% et 71% pour les frontières internes aux *tours de parole*, de 97% et 96% pour les « marqueurs de discours » et de 66%-74% pour les « disfluences »[28].

Simon & Christodoulides (2016) confirment l'interaction de la **perception accrue des frontières** lors de l'actualisation de pauses et de (dis)fluences par les auditeurs francophones natifs. Une expérience portant sur la perception de frontières (discursives, intonatives et/ou syntaxiques) par 125 locuteurs francophones natifs a livré de pertinents résultats. Lorsque les auditeurs se voient retirer le contenu syntactico-sémantique de la parole (condition filtrée), ils parviennent à mieux détecter la présence de frontières, surtout **lorsque la syllabe est allongée ou marquée par une pause pleine**, alors que ce traitement est inhibé en condition de parole non filtrée, naturelle.

27. La barre verticale symbolise une pause, l'astérisque l'*interregnum*, les segments entre crochets les *reparandum* et *reparans*.

28. Il est important de noter que nombre d'études portant sur les « marqueurs discursifs » entendent des « particules d'hésitation », parfois des interjections et rarement des marques plus diversifiées, contrairement à des recherches plus actuelles, ce qui tend à surévaluer le taux de précision des détecteurs et des catégoriseurs automatiques (voir notamment Christodoulides 2016 ; Crible 2017).

Exemple (8) – issu de Simon & Christodoulides (2016:99) (corpus LOCAS-F & C-PROM) [29]

a. < qui m'ont reproché de | euh d'insister trop euh sur la correction d'accents régionaux >

b. < un état qui protège ses habitants | et qui leur donne une patrie >

Ainsi, à l'exemple 8a. une pause silencieuse après « *de* » génère la perception d'une frontière, inhibée en condition maternelle en raison de l'**incomplétude syntaxique** de la construction. En 8b. la frontière a été perçue en condition non filtrée en raison de la **complétude syntaxique** (alors que les pauses silencieuses sont absentes), tandis qu'aucun participant n'a marqué de frontière en condition filtrée.

Ces résultats renforcent par ailleurs les **convictions méthodologiques** initiales de notre travail de recherche et précisent les variables à prendre en compte dans notre **modélisation et notre catégorisation pluridimensionnelle** des (dis)fluences. En plus des différences structurelles internes aux zones de disfluences (simple /vs/ composée, présence et type de phase d'édition, etc.), des traits acoustico-prosodiques (patrons prosodiques internes, proéminences, allongements, etc.), morphosyntaxiques (catégories grammaticales, Mots Pleins (MP) /vs/ (MO) Mots Outils, etc.), les contextes de frontières prosodiques sont des lieux propices d'actualisation à la production et à la perception d'une (dis)fluence.

Nous définissons les contextes prosodiques en termes de *groupe intonatif*, *contour intonatif*, *groupe accentuel accentuable* et *actualisé* [30]. Notre démarche est singulière en ce qu'elle cherche les marques de *dis*fluence (marque de frontière ou non) dans une **approche sociocognitive**. Son objectif n'est pas la détection de frontières purement intonatives. A notre connaissance, les recherches les plus actuelles adoptant des **critères intonatifs, syntaxiques et pragmatiques** sont celles du laboratoire MoDyCo (Beliao, Kahane & Lacheret 2013 ; Beliao & Lacheret 2013) (Paris Ouest Nanterre La Défense), bien que la question de recherche soit orientée dans une perspective résolument plus computationnelle, limitant l'analyse à la problématique de la coordination /vs/ subordination, c'est-à-dire de l' « élaboration » et du travail de formulation du discours dans sa connotation positive.

29. Dans l'annotation de Simon & Christodoulides (2016) la barre verticale « | » symbolise une pause.

30. Voir chapitre 3 section 2.3.

1.3 Définition syntaxique du contexte d'actualisation des (dis)fluences

> « Dans son acception cognitiviste-constructiviste, la macrosyntaxe permet [...] d'établir le lien entre le système linguistique, la situation de production langagière et l'ensemble des savoirs individuels et partagés des interlocuteurs. » (Bolly, Crible, Degand & Uygur-Distexhe 2015)

La définition du contexte syntaxique des (dis)fluences, comme celle du contexte prosodique résulte de choix méthodologiques et théoriques du chercheur. Dister (2007:224) constate que les critères de **définitions des positions syntaxiques** des marqueurs varient considérablement d'une étude à l'autre. Grosjean & Deschamps (1972) appliquent des critères en **syntaxe de dépendance** ; Campione & Véronis (2004) découpent les énoncés en « **segments discursifs** » (unités relativement brèves) ; Dister (2007) se restreint aux **constituants syntaxiques minimaux** équivalant au syntagme élémentaire ou *chunk* s'apparentant à ceux de ces derniers. En somme, si l'on cherche à savoir quels marqueurs entraînent davantage une **stagnation /vs/ une rupture syntagmatique** – ou si une rupture syntagmatique est toujours associée à une rupture discursive et/ou à une évaluation *dis*fluente – on se questionne avant tout sur le(s) type(s) de critère(s) de sélection des unités syntaxiques pertinentes à ce type d'analyse.

Pour rappel, au niveau **morphosyntaxique**, on observe de nettes tendances de certaines (dis)fluences à interagir avec la **catégorie grammaticale** des mots. On distingue ainsi traditionnellement les (dis)fluences s'actualisant sur des ***mots outils*** /vs/ ***mots pleins***. Ces termes sont empruntés à Tesnière (1988) et Morel & Danon-Boileau (1998). Ils correspondent peu ou prou à ce que certains appellent les *mots grammaticaux* (déterminants, articles, pronoms personnels, conjonctions de subordination, verbes auxiliaires, copules, prépositions) /vs/ *mots lexicaux* (verbes à l'indicatif, infinitif et participe, noms, adjectifs et adverbes)[31]. L'agencement des MO et MP construit le flux syntagmatique.

Par ailleurs, chaque actualisation de (dis)fluence entraîne un **degré de *rupture syntaxique* (ou cohésion) spécifique**. La stagnation sur l'axe paradigmatique peut être mineure, en terme de **nombre** ou **durée** d'éléments stagnants, et en fonction du **degré de violation de la liaison entre les constituants** (e.g. liaison du prédicat avec ses éléments constituants essentiels). Dans l'exemple 9a. la liaison syntaxique des constituants est mise en péril contrairement à 9b.

Exemple (9) – issu de Simon & Christodoulides (2016:101) (corpus LOCAS-F & C-PROM)

a. alors Arte Radio *euh* a choisi de | fabriquer des pastilles (B02)

b. elles voulaient quand même partir | *donc du coup* il a bien fallu qu'on trouve un endroit pour dormir

31. On dit par ailleurs les premiers à tendance atones et les seconds, toniques ou accentuables (chapitre 3 section 2.3).

Au-delà de la morphosyntaxe, reconnaissons la distinction de critères d'ordres ***micro*** /vs/ ***macrosyntaxiques***, opérée par le cadre stratificationnel du groupe de Fribourg. Ces deux modes d'analyse de l'**organisation des énoncés** se fondent sur le fonctionnement de la « rection » dans la grammaire de l'oral[32]. Les théories de Blanche-Benveniste (1990a, 2003) ont pour particularité de ne pas considérer les **degrés de dépendance** grammaticale de la syntaxe (Corminboeuf & Benzitoun 2014).

Pour rappel, dans le cadre des **unités microsyntaxiques**, on étudie donc le fonctionnement des unités dans la syntaxe de dépendance du verbe dont l'interprétation est restreinte à la proposition. Le contexte d'actualisation syntaxique sert à formuler des hypothèses cognitivistes ou psycholinguistiques du traitement des (dis)fluences. Les auteurs identifient les **ruptures syntaxiques mineures**, comme dans le cas des répétitions ou autocorrections immédiates et les **ruptures syntaxiques profondes** telles que les inachèvements et faux départs (Blanche-Benveniste 1987). Selon Morel & Danon-Boileau (1998:75), les **ruptures syntaxiques profondes** correspondent à des malaises de la **pensée** tandis que les **ruptures syntaxiques mineures** traduiraient simplement des malaises dans la **formulation**.

Candea (2000a:291) formule une proposition innovante, en faisant interagir les deux modes d'analyse lors de la description de la position syntaxique des (dis)fluences. Elle distingue au niveau microsyntaxique la (dis)fluence en position d'***intraconstituant*** /vs/ ***fin de constituant*** – propositionnel ou syntagmatique. Au niveau macrosyntaxique, elle met à part les (dis)fluence dites ***en fin d'énoncé***. L'énoncé est défini ici comme un « tour de parole ». Dans le cadre, les unités intonatives majeures, les (dis)fluences et marqueurs de discours affichent parfois des préférences pour une **position autonome**. L'auteure associe à cette position autonome (en fin d'énoncé) une portée énonciative plus large. Cette portée énonciative plus étendue induit une interaction plus importante avec les **conditions pragmatiques** de la situation (i.e. macrosyntaxiques).

L'association du degré de dépendance grammaticale (***ancrage syntaxique***) et de l'interprétation des fonctions de la (dis)fluence trouve également des résonances dans la littérature anglophone. Celles-ci concernent les moniteurs de conceptualisation et d'articulation au coeur de la théorie du ***perceptual loop model*** (Hartsuiker & Kolk 2001 ; Levelt 1989). Les auteurs lient l'idée de micro- et macroplanification avec la position d'actualisation des (dis)fluences (et les moniteurs qui leur sont associés)[33].

Shriberg (1994) en se basant sur Swerts (1998:489), met en relation la position syntaxique des (dis)fluences et le(s) **processus cognitif(s)** qui leur seraient sous-jacents. Au niveau « macrosyntaxique », les auteurs ont montré que les énoncés plus longs avaient (relativement) tendance à attirer plus de (dis)fluences, ainsi que des (dis)fluences plus complexes (Oviatt 1995). Shriberg (1994) énonce que les pauses pleines en **position phrastique initiale** reflètent un **processus cognitif plus profond** lors de la phase

32. Théorie développée et étoffée par le GARS – voir chapitre 1 section 2.2.
33. Voir chapitre 1 section 2.3.3.

(et du *monitoring*) de **conceptualisation**. Au contraire, les pauses pleines en **position phrastique interne** reflèteraient une recherche lexicale ou un problème local d'encodage phonologique et se passeraient donc lors de la phase (et du *monitoring*) de **formulation** ou d'**articulation** (Shriberg 1994). Ici, par contre, peu d'éléments de pragmatique sont pris en compte dans la description du fonctionnement des (dis)fluences.

Enfin, du côté de la **perception syntaxique**, la présence d'une frontière syntaxique, précédente ou subséquente, simultanément à l'actualisation d'une (dis)fluence favoriserait sa perception. Tandis que les pauses à l'intérieur de constituants seraient associées le plus souvent à des pauses stylistiques dans le corpus de Duez (1991), la **perception d'une frontière** induirait même la perception de ***pauses subjectives*** par les auditeurs. Les *pauses subjectives* désignent un endroit où un auditeur a perçu une pause silencieuse (absence de volume sonore) alors qu'aucune interruption du signal sonore n'est **acoustiquement observée** (Candea 2000a:131,450 ; Duez 1993). Les *pauses subjectives* semblent davantage perçues en frontière microsyntaxique (e.g. **intrasyntagmatique**) qu'en frontière macrosyntaxique (**proposition, rection, « énoncé »** , etc.)[34] et aussi entre leurs différents contextes d'actualisation linguistique (e.g. isolement /vs/ cooccurrence, début /vs/ fin d'unité prosodique, etc.).

La perception des pauses est également accrue par la **présence d'un allongement vocalique** ou *euh* (d'appui ou d'hésitation). Duez (1991:141-142) le note surtout en condition expérimentale de perception de parole inversée (i.e. dont le contenu sémantique est inintelligible). Par contre, la perception de frontière ne semble *pas* nécessairement être associée à des variations de F0 ou d'intensité (même élevées) sur la syllabe précédant la pause (à l'instar des pauses silencieuses). Les expériences de Duez (1991) ne semblent par contre pas montrer une différence du taux d'identification des pauses pleines lorsqu'elles font **varier l'intensité ou la F0** avant celles-ci.

En résumé, dans la plupart des cas, les pauses subjectives correspondent à un **allongement final**, qui concorde lui à la réalisation d'une frontière syntaxique. Nombre d'auteurs affirment que c'est davantage la position des marques par rapport aux *frontières syntaxiques* et le **débit de parole**, qui semblent avoir un effet important plus que la durée de la marque en elle-même. Moins de la moitié des auditeurs perçoivent des marques dépassant 500 ms. lorsque celles-ci sont à l'intérieur d'un syntagme (Candea 2000a:118). Hormis la présence ou l'absence d'une pause et d'une frontière, bien entendu : plus longue est la durée, plus le taux d'identification est élevé.

Nous le verrons dès le chapitre 4, nous postulons pour une méthode plus objective et contextuelle de la mesure de ces durées et de leur interactions avec les **frontières syntaxiques et prosodiques**. Les critères de définitions et les comportements syntaxiques

34. La section 2.3 du chapitre 3 revient sur les unités syntaxiques adoptées dans le cadre de nos analyses (unités interpausales, unités de rection, séquences fonctionnelles, groupes accentuables et unités de bases du discours) et ce de manière formelle.

décrits ci-avant, à l'instar des critères prosodiques, embrassent plus ou moins la **dimension pragmatique** du discours. Les choix de segmentation ont donc des implications théoriques parfois diamétralement opposées (bien que conciliables), offrant une vision très différente de la contextualisation syntaxique des (dis)fluences.

1.4 Proposition d'évaluation contextuelle des (dis)fluences en production et en perception

Ce travail s'attache à une description des (dis)fluences *en contexte*. Ce contexte se définit non seulement par l'environnement linguistique d'actualisation des marques (acoustique, prosodique, syntaxique) mais également par ses **caractéristiques situationnelles** d'actualisation des discours. Or, ces deux types de contexte communiquent.

Dans ce cadre, un des objectifs des descriptions phonostylistiques est d'accéder aux « tendances prosodiques centrales » communes à une ou plusieurs pratique(s) communicative(s)[35]. On entend par là, décrire en contexte, la réalité d'une certaine « norme discursive » produite par les locuteurs. L'**ancrage contextuel** de cette norme dépend des **représentations sociocognitives** des interlocuteurs. Celles-ci sont composées du *contexte cognitif immédiat*, produit des interactions des **contextes externe et interne**. Ces derniers regroupent d'une part des représentations de connaissances encyclopédiques (mémoire à long terme), et d'autre part, des représentations des contextes linguistique, social, et situationnel à un instant t[36].

L'observation fréquentielle des corpus offre des **normes statistiques** de production des discours. Ces normes statistiques observées présentent plus ou moins de régularités ou de variations au sein d'une même **pratique communicative**. Dans le cas de l'analyse phonostylistique, elles résument les **caractéristiques**, notamment **prosodiques**, des discours en situation. Ces attributs dessinent leur « **image acoustique typée** ». A titre d'exemple, on oppose souvent le *phonostyle* des discours issus de situations à orientation médiatique /vs/ exempts de dimension médiatique (vœux présidentiels, discours d'assemblée générale, revue de presse /vs/ parole spontanée, lecture, narration conversationnelle). Les premières sont caractérisées par une forte densité accentuelle et un registre intonatif plus étendu et dynamique (Grosman 2013b, 2016 ; Simon & Mertens 2013).

Le débat de la **convergence ou de la divergence** des résultats de **l'observation basée sur l'usage** (en linguistique de corpus) ou de l'observation fondée sur **la perception de cet usage** (en linguistique expérimentale) est loin d'être trivial. Les normes statistiques produites, basées sur l'observation de l'usage correspondent-elles à la perception de cet usage ? En d'autre termes, pouvons-nous confirmer une réelle convergence des normes et des fréquences ?

35. Voir chapitre 1 section 3.3.
36. Voir chapitre 1 section 1.1.

1.4.1 Limites de l'approche fréquentielle de la saillance

L'« effet de fréquence » considère que c'est dans la **répétition** d'une pratique communicative que se créent les « **patrons linguistiques** » (Bybee 2006)[37]. Avec une approche basée sur l'usage (l'observation des corpus dans le cadre des grammaires émergentes), ce sont ces expositions aux patrons linguistiques qui déterminent le **degré d'automatisation** et l'aisance de leur traitement en production et en perception chez les interlocuteurs.

Résumons. La « norme » observée sur un corpus (de production) offre au moins en apparence une *norme objective* tirée des statistiques de fréquences. Ces fréquences concrétisent des *usages*, représentatifs de par leur proportion dans une situation. Finalement, ces régularités aboutissent à la formulation des « **normes discursives** » descriptives d'une **pratique communicative**. Ces normes sont dites *discursives* lorsqu'elles prennent en compte l'interaction des différents types de contexte, généralement syntaxiques et pragmatiques (et donc situationnels). Des normes discursives phonostylistiques sont par exemple établies sur corpus dans les études de Degand & Simon (2009), lorsqu'elles mêlent analyses phonostylistiques et segmentations prosodico-syntaxiques du discours[38].

Une chaîne de (dis)fluences peut être plus ou moins fréquente dans un style de parole, et donc être reconnue au niveau fréquentiel comme une structure intégrée à des « **représentations mentales prototypiques** ». Des exemples de « schéma potentiel » sont révélés par les stéréotypes véhiculés en parole : « les Parisiens disent *euh* », « les bourgeoises disent bonjour-e 'comme ça' », « les jeunes disent 'ouais tu vois quoi' » ou même le simple principe que « le commentateur sportif parle vite ». Le **schéma actualisé** s'accorde ou non avec les normes discursives associées à un style de parole.

L'extrapolation est aisée : une **saillance cognitive** (issue de la notion de prototype) serait associée positivement à un « **degré de fluence** ». On entend ici par fluence, la **fluence cognitive**, mesurée en terme d'accessibilité et reliée à la fréquence ; et non pas la *fluence conceptuelle*, liée à la programmation et à la formulation, et aux **erreurs d'appropriété**[39]. Ce lien entre fréquence et fluence semble univoque. Une approche fréquentielle ou mono-méthodologique induit cependant une analyse *non dynamique, a-pragmatique et a-contextuelle* de la définition du continuum fluence-disfluence et de l'évaluation des marques en contexte.

37. Voir chapitre 1 section 2.1.

38. Le concept de ***norme fréquentielle*** rappelle la notion de **prototype**. La notion même de « prototypicalité » dans la littérature tend à devenir une notion « fourre-tout » fort controversée bien qu'on en reconnaisse les effets indiscutables sur les capacités cognitives (telle que la capacité de catégorisation). Nous renvoyons à Estes (1994) pour une revue de la littérature et au chapitre 1, section 1.2.4 pour une introduction sur le lien entre chaîne et schéma de (dis)fluence(s). Le prototype est défini en psychologie expérimentale comme un « **membre représentatif d'une catégorie** » (Rosch 1975) ; telle une matérialité acoustique ancrée dans le discours correspondant le plus à l'idée abstraite que se font les interlocuteurs du phonostyle et du phonogenre.

39. Justement, Candea (2000a:87) suggère qu'il ne faut pas minimiser l'importance du continuum entre ces deux « pôles de fluences ».

L'approche fréqentielle se retrouve dans le concept d'« **enracinement sociocognitif** » développé par Langacker (2000) et repris par Schmid (2010:38-41). La **fréquence d'occurrence** (de chaîne) de phénomènes fait varier l' « enracinement ». En d'autres termes, elle augmenterait la « **saillance cognitive** » et donc l'accessibilité du sens du stimulus – ici de la production et de la compréhension d'un schéma de (dis)fluence – par définition contextualisé (Janda 2010 ; Radden 1992). En résumé, selon ces théories, les combinaisons courantes de (dis)fluences constitueraient des enracinements. Ces théories ont pour avantage de mettre explicitement en avant la **relation entre fréquences et normes**.

Aisément, pour certains psycholinguistes cognitivistes le *prototype* fait référence à ce qui est *cognitivement le plus saillant* (Geeraerts 1989:606 ; Gilquin et al. 2008 ; Arppe et al. 2010:11), et nombre de linguistes n'ont pas manqué de faire l'**amalgame** avec « ce qui est *le plus fréquent* » (Gilquin 2006:159). Dans une vision assez simpliste du phénomène, il est possible de tomber sur un écueil de taille. Toutefois, plusieurs arguments, intuitifs, mettent directement en péril le **raccourci d'équivalence** illustré en Figure 2.2. Non seulement les résultats de ces approches ne convergent pas de manière systématique (Gilquin 2006 ; Gilquin et al. 2008 ; Glynn & Fischer 2010:10), mais surtout, l'équivalence des termes est méthodologiquement non applicable : leur opérationnalisation, les mesures qui leur sont associées ne peuvent être confondues.

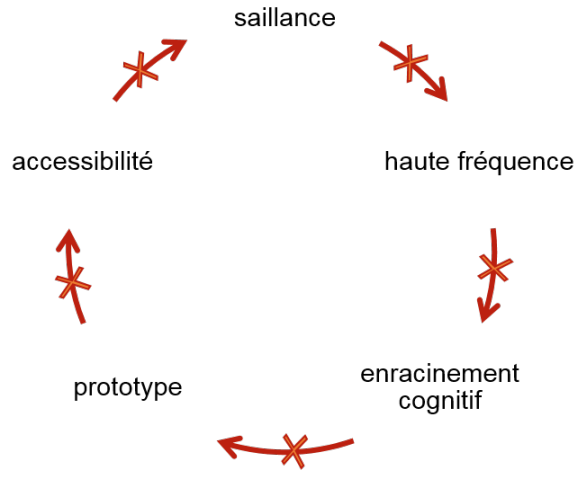


Figure 2.2 – Amalgame des relations de saillance et de fréquences.

En effet, alors que la *saillance* est mesurée par une relative **accessibilité** aux différents sens d'une structure, la *fréquence* est mesurée par **décompte sur des corpus**. La réponse à cette question de convergence entre fréquences et normes n'est donc pas si directe [40]. C'est précisément ce manque de **convergence entre saillance et fréquence** qui défie la capacité des corpus à servir de chemin d'accès à la cognition (Arppe et al.

40. De plus, comme énoncé précédemment, les propos d'Arppe et al. (2010) corroborent les expériences d'Alter qui a montré que les objets peu fréquents du monde social sont parfois particulièrement saillants (la forme se détache du fond – e.g. une tache d'encre ou une typographie sur une page de journal) et peuvent donc être synonymes de fluence dans le sens où ils favorisent l'attention d'une personne.

2010:11) bien que nombreux soient les auteurs argumentant en faveur de la compatibilité entre linguistique de corpus et théories cognitivistes (Chafe 1994 ; Gries 2011 ; Schönefeld 2011). Le plus fréquent ne veut pas toujours dire le plus saillant, loin de là. Comment **réconcilier les définitions** et les évaluations fréquentistes de la fluence-disfluence avec une approche linguistique de la saillance ancrée sociocognitivement ?

1.4.2 Unification du concept : approche sociocognitiviste de la saillance

Les théories de la cognition sociale prennent en compte que les « objets sociaux » ont un impact au niveau individuel. Le comportement social et discursif des individus n'est pas déterminé uniquement par la situation dans laquelle ils se trouvent mais plutôt par la **perception personnelle de cette situation**. C'est l'**activation d'un concept** par un élément contextuel qui permet aux sujets l'accessibilité à un ou une série de schémas. En d'autres termes, malgré les intérêts de l'approche fréquentielle, il semble impensable d'oublier, comme Blumer (1986:2) le formula si bien que les « humains agissent envers les choses sur base de ce que ces choses représentent pour eux » [41].

L'étude de la perception n'est pas l'étude de l'influence d'un stimulus dans l'absolu mais surtout de la **manière dont les individus forment des représentations mentales internes du monde extérieur** et de comment ces représentations guident leur comportement (Kihlstrom & Nathan 2013). En effet, la perception d'un stimulus implique chez l'individu un **changement du contenu en mémoire en laissant une trace de l'expérience perceptuelle même après disparition du stimulus**. Cette mémoire peut guider ou adapter notre comportement aux événements présents et, dans une certaine mesure, cette faculté de **se souvenir du passé** permet même d'**anticiper le futur**.

On ne peut s'intéresser aux mécanismes de l'évaluation de la *fluence* chez les locuteurs natifs sans définir le type de situation et de population (et donc de locuteurs-auditeurs [42]) adressés et leurs différentes expériences discursives (attitude face à la norme et aux jugements, etc.) (Blakemore, Winston & Frith 2004 ; Levinson 2012 ; Mildner 2008). A cet égard, on est enclin à supposer que (1) **les auditeurs formulent des attentes envers le discours de leurs interlocuteurs** et que (2) **ces attentes varient en fonction de leur « *norme représentative d'un discours* »** (Hambye & Simon 2004). En conséquence, cette norme représentative s'adapte en fonction des paramètres situationnels. La formation d'une **norme situationnelle** peut se développer au regard de plusieurs théories linguistiques. L'idée étant qu'en fonction du « **passé discursif** » des locuteurs, leurs attentes/normes discursives varieront et auront un impact sur ce qui est jugé saillant ou non et par extension, le degré de fluence ou disfluence d'un item linguistique.

41. Texte original : "Human Beings act toward things on the basis of the meaning that the things have for them." (Blumer 1986:2). Notre traduction. Nous surlignons.
42. Désormais « locuteurs ».

Comme le met en valeur le travail de synthèse de Schmid & Günther (2016:1), le concept de saillance est devenu ambigu, ce qui lui vaut souvent la confusion avec des interprétations ambivalentes en termes de *fréquence*. Selon ces auteurs, les différents types de saillance dépendent des **interactions de l'objet évalué** avec les **connaissances de la mémoire à long terme** des interlocuteurs tout comme de la **prise en compte du contexte** (Figure 2.3). Ainsi, un mot, ou tout autre item linguistique ou non linguistique, peut réveler quatre types de « saillance ».

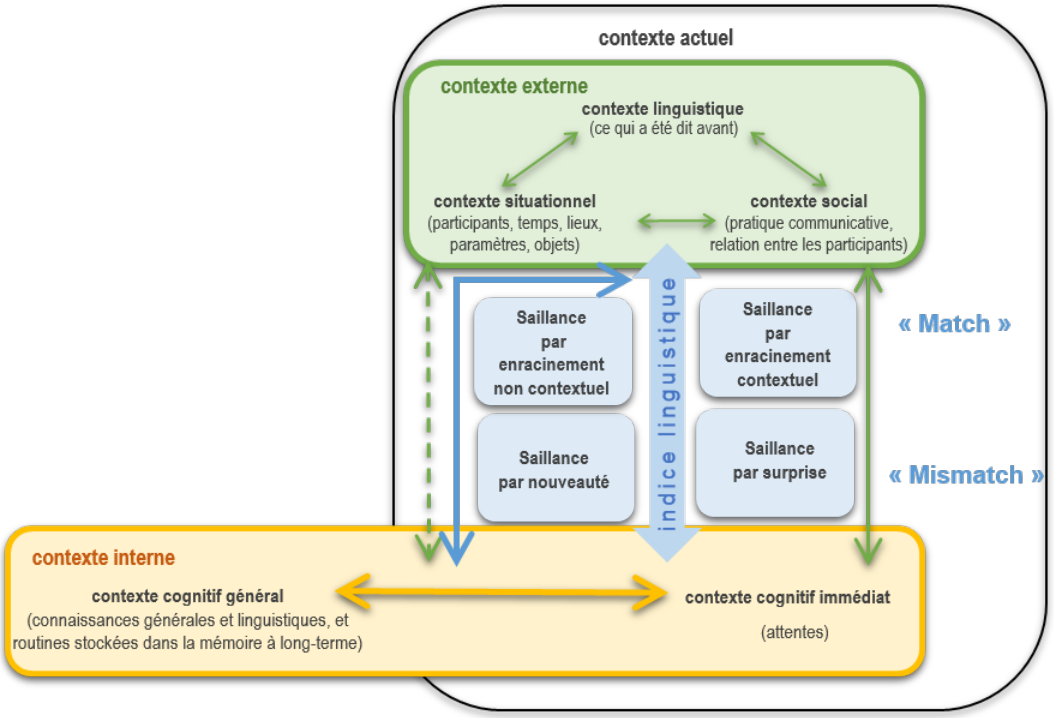


Figure 2.3 – Quatre types de saillances issues de la comparaison d'indices linguistiques et des attentes – adapté de Schmid & Günther (2016).

1. **Saillance par l'enracinement et la familiarité non contextuelle**. Un item est saillant car il s'agit du premier qui vient à l'esprit. C'est le scénario qu'on retrouve chez Langacker (2000) et Giora (2003).
2. **Saillance par enracinement et par probabilité contextuelle positive**. Un item est saillant car il s'agit du premier qui vient à l'esprit *dans ce contexte*. On retrouve l'approche de Geeraerts (2017) selon qui l'enracinement ne doit pas être vu uniquement comme une fréquence absolue ou en terme psychologique.
3. **Saillance par « surprise » ou probabilité contextuelle négative**. Un item est saillant car on ne s'attend *pas* à le rencontrer *dans ce contexte*. La saillance correspond ainsi à un effet de surprise, d'inattendu tel que chez Clark (2013) et Divjak (2016).
4. **Saillance par « nouveauté » ou non-familiarité totale**. Un item est saillant car il n'a jamais été vu/entendu auparavant. Il s'agit là d'une vision de la saillance comme « une information nouvelle pour la mémoire », plus qu'une surprise basée sur le contexe – question discutée par Barto, Mirolli & Baldassarre (2013).

Nous avons donné une définition du **contexte cognitif immédiat** (résultant de l'interaction, des contextes linguistique, situationnel et social ainsi que du contexte cognitif général) [43]. **Le traitement de la saillance résulte alors de la comparaison entre les contextes cognitifs et les attentes linguistiques**. En situation, les interlocuteurs partagent le contexte externe actuel mais pourraient avoir un contexte cognitif immédiat différent. Nous pensons que la résolution de ce conflit entre les différents types de saillance et leur relation à la fluence se situe dans la prise en compte de ces différentes relations.

1.4.3 Proposition de redéfinition et de formalisation du continuum fluence-disfluence chez les locuteurs francophones natifs

Dans un discours, les attentes font partie des représentations mentales activées par le **contexte communicatif immédiat**. Pour rappel, celui-ci est le fruit des interactions avec leurs représentations encyclopédiques (contexte externe) et contextuelles (contexte interne) (Schmid & Günther 2016) [44]. Ces **attentes** permettent l'évaluation d'un contenu par les interlocuteurs. L'évaluation de la fluence-disfluence peut s'envisager en fonction de la **convergence ou divergence** avec ces attentes (*confirmation* ou *déception*) ou de manière plus continue : *surconfirmation, confirmation, déception, surdéception* (*overmatch, match, mismatch, overmismatch*). L'évaluation de l'item linguistique comme la marque du travail de formulation naît du croisement des types d'attentes pris en compte dans l'évaluation.

Les **attentes linguistiques, situationnelles, sociales**, permettent respectivement l'évaluation des fluences linguistiques phrastique, situées (ou situationnelles) et socio-interpersonnelles. Ces attentes et fluences interagissent entre elles. Afin de prendre en compte ces interactions, nous proposons d'**étendre le cadre méthodologique** de Schmid & Günther (2016).

Définissons et opposons sur l'axe vertical les types de fluences relatifs au « **contexte externe** » des locuteurs (i.e. les connaissances encyclopédiques) à des évaluations dépendantes du « contexte interne » de production des discours. Ainsi, différents types de fluence émergent en fonction du **type d'attente**, mais également en fonction du **type de contexte** pris en compte. Sur l'axe horitontal, construisons un **continuum de convergence-divergence** de la situation (et des items linguistiques) avec les attentes du locuteurs (Figure 2.4). L'interaction de ces deux axes permet d'accorder des types (ou niveaux) de **saillance(s)** d'un stimulus linguistique produit ou perçu sur un continuum de fluence-disfluence. Ce niveau de saillance dépend des **attentes des interlocuteurs**.

43. Voir chapitre 1 section 1.1.
44. Voir chapitre 1 section 1.1.

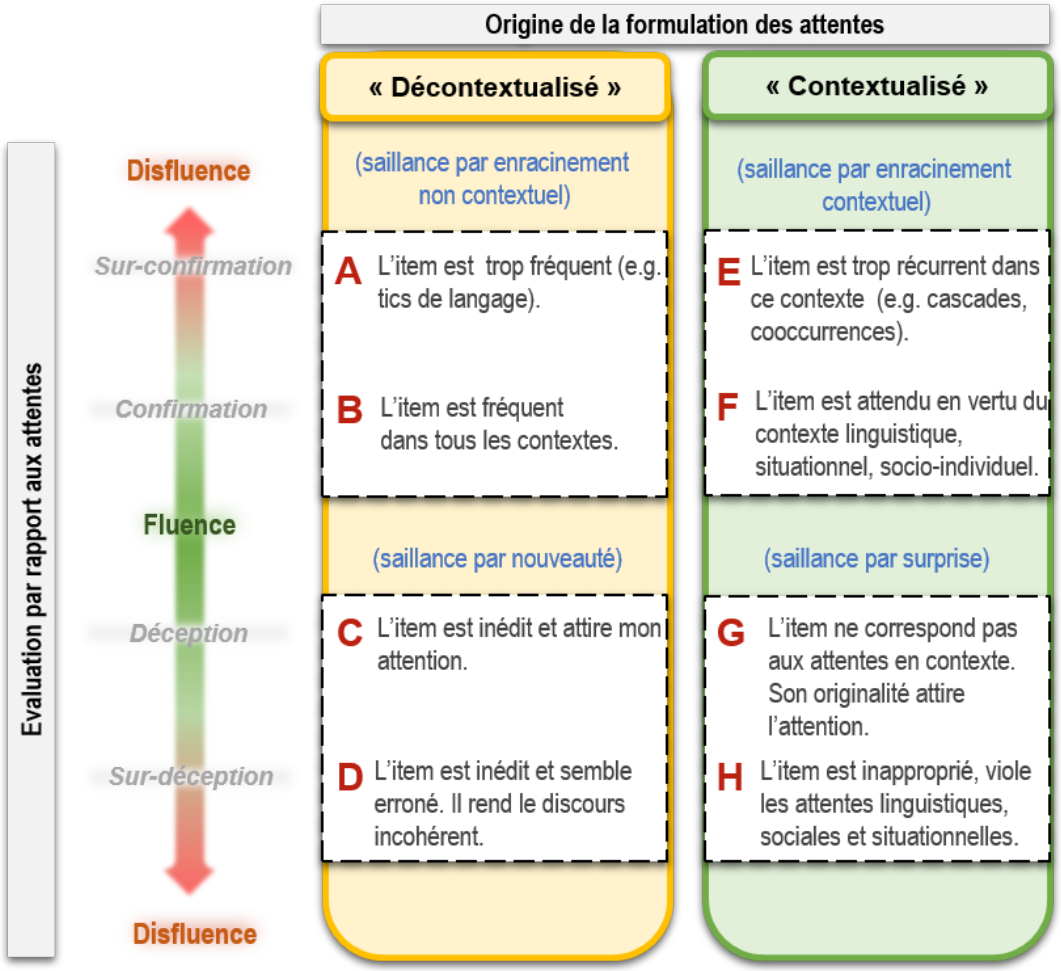


FIGURE 2.4 – Proposition de redéfinition du continuum fluence-disfluence par extension du cadre sociocognitif de la saillance.

Une réelle *dis*fluence correspond aux extrêmes des continuums : une **sur-saillance** marquée des attentes de l'auditeur vis-à-vis des **contextes externe et interne**, par rencontre ou déception du contenu linguistique, du registre et/ou de locuteur dans une situation. Ce cadre nous permet de distinguer des évaluations de la (dis)fluence en fonction de leur dépendance aux **contextes externe** ou **interne** au moment de production d'une (dis)fluence. On oppose ainsi une évaluation des représentations linguistiques contextuelles, situationnelles et sociales de production du discours à une appréciation **indépendante du contexte** – largement dépendante de **connaissances linguistiques à long terme**.

Nous pouvons placer ces types de fluence-disfluence sur un continuum allant de A à H, les cas de confirmation ou déception s'apparentant plus à de la fluence, tandis que les cas de **sur-confirmation** ou **sur-déception** des attentes s'apparenteraient plus à de la ***dis*fluence**. Ce continuum est schématisé en Figure 2.5.

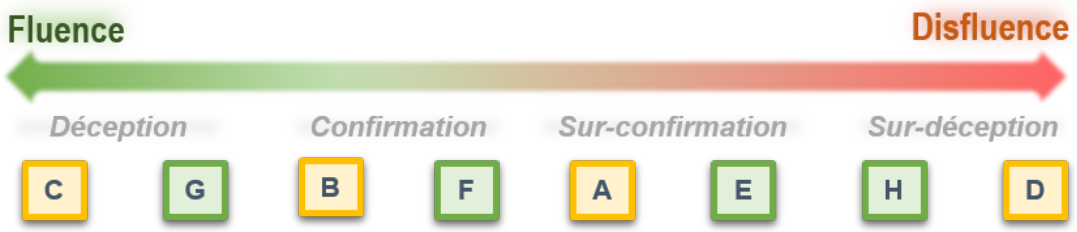


FIGURE 2.5 – Hypothèse de catégorisation des items sur le continuum fluence-disfluence.

A l'extrême gauche du continuum, les cas [C] et [G] correspondent à des phénomènes d'**écarts par rapport à une norme fréquentielle** et/ou aux attentes discursives des locuteurs. Bien qu'à l'encontre des régularités, ces cas, que certains catégoriseraient traditionnellement de « *dis*fluents », peuvent en réalité être porteurs d'une certaine fluence. Ces éléments **provenant d'un contexte normé** ressortent tout juste assez du fond pour **attirer et accroître l'attention** des interlocuteurs sur le contexte environnant. Les (dis)fluences ou segments prosodiques **relativement atypiques** en contexte **facilitent par conséquent l'attrait ou la mémorisation** de certains événements (GILQUIN 2006:168 ; GLYNN & FISCHER 2010:14 ; ALTER & OPPENHEIMER 2009:220).

> **Exemple (10)** – **Cas `C` et `G`** : Fluence par déception des attentes (décontextualisées et contextualisées) – issus des corpus LOCAS-F & C-Humour [45].
>
> a. Déception décontextualisée : saillance par nouveauté, l'item est inédit et attire l'attention.
>
> b. Déception contextualisée : saillance par surprise, l'item ne correspond pas aux attentes en contexte ; son originalité attire l'attention.

Cette approche permet aussi de mettre en évidence le fait que les **particularismes phonostylistiques ou idiolectaux** sont porteurs de fluence. Des stratégies compensatoires préférentielles de certains locuteurs, bien qu'en potentiel désaccord avec les normes fréquentistes, peuvent être tout-à-fait acceptables.

Dans les cas [B] et [F], l'item **correspond aux attentes** (fréquentielles et discursives) des locuteurs. Ces items ne sont donc pas marqués : ils sont relativement fréquents et faciles à traiter. Ils correspondent ainsi à la vision la plus classique de la fluence phrastique grammaticale (e.g. absence de marque et donc de marqueurs), en terme également de **fluidité**, de « parcours sans effort » (e.g. haute fréquence de la marque dans l'absolu en mémoire à long terme ou en contexte, isochronie rythmique, structuration syntaxique « efficace » etc.).

45. Les corpus à l'étude sont présentés en détail au chapitre 3.

Exemple (11) – **Cas `B` et `F`** : Fluence par confirmation des attentes (contextualisées et décontextualisées) – issus des corpus LOCAS-F & C-Humour [46].

a. Confirmation décontextualisée : saillance par enracinement non contextuel, l'item est fréquent dans tous les contextes.

b. Confirmation contextualisée : saillance par enracinement contextuel, l'item est attendu en vertu du contexte linguistique, situationnel ou socio-individuel.

Dans les cas [A] et [E], une évaluation de **disfluence mineure** peut surgir lorsqu'un item est bien attendu dans un contexte phrastique, situationnel, socio-individuel, mais est en fait trop présent ; il en devient *marqué* (e.g. cooccurrence ou accumulations de marque, tics de langage contextuels de type [*euh* + quoi] en fin de constituants, etc.). Nous sommes tous plus ou moins sensibles aux « **tic de langage** » de notre entourage. Ils se remarquent particulièrement quand une même marque produite par le locuteur est jugée par l'auditeur comme utilisée de manière excessive, que celle-ci le soit en raison de sa **présence objective** (approche fréquentiste) ou par une **proéminence perceptuelle** (induite par son utilisation en contexte).

Exemple (12) – **Cas `A` et `E`** : Disfluence par sur-confirmation des attentes (contextualisées et décontextualisées) – issus des corpus LOCAS-F & C-Humour.

a. Sur-confirmation décontextualisée : saillance par enracinement non contextuel. L'item est trop fréquent (e.g. tics de langage).

b. Sur-confirmation contextualisée : saillance par enracinement contextuel. L'item est trop récurrent dans ce contexte (cascades, cooccurrences).

A l'extrême droite du continuum, aux cas [H] et [D], une évaluation de **disfluence majeure** surgit lorsque l'unité linguistique **ne correspond pas aux attentes** (phrastiques, situationnelles, socio-individuelles) et/ou n'est pas **appropriée** voire **incompréhensible** par les locuteurs. Ici, les locuteurs s'attaquent en majeure partie à des (dis)fluences d'appropriété ou d'intelligibilité [47].

Exemple (13) – **Cas `H` et `D`** : Disfluence par sur-déception des attentes (contextualisées et décontextualisées) – issus des corpus LOCAS-F & C-Humour.

a. Sur-déception contextualisée : saillance par surprise. L'item est inapproprié, viole les attentes linguistiques, sociales et/ou situationnelles.

b. Sur-déception décontextualisée : l'item est inédit et semble erroné. Il rend le discours incohérent.

46. Les corpus à l'étude sont présentés en détail au chapitre 3.
47. Voir chapitre 1 section 2.3.3.

Ce cadre met en valeur l'intérêt d'une approche pluridimensionnelle du continuum fluence-disfluence – en production, en perception, contextualisée de manière linguistique mais aussi pragmatique. Nous avons revisité et remis en question les liens entre contextes - attentes - fréquences - saillances et (dis)fluences. Il est ainsi possible de confronter ces **normes statistiques produites** aux **normes perçues**. Cet état de l'art justifie que ce travail fasse appel à un questionnement méthodologique plus large sur l'application de **méthodes mixtes** en sciences humaines (Tashakkori & Teddlie 2003) (voir chapitre 3). Cette démarche était soutenue par l'hypothèse sous-jacente qu'une marque isolée n'est pas évaluée de la même manière selon sa distribution en contexte [48]. Elle peut être évaluée différemment en fonction de l'importance donnée aux contextes, des types de fluences considérés et de leurs interactions (e.g. continuum des contextes linguistiques, situationnels ou socio-individuels d'évaluation de la fluence-disfluence – voir les expériences de perception chapitre 6 section 2).

Dans cette introduction nous avons évoqué le principe de **regroupement** de (dis)fluences en chaîne [49]. Les deux sections suivantes s'attachent ainsi à la définition contextualisée des marques de manière isolée et en cooccurrence. Elles questionnent le **niveau de rupture** (discursive, syntaxique) que certaines marques peuvent engendrer, et la manière de lier ces **ruptures potentielles** avec notre modélisation théorique du continuum fluence-disfluence. Notre cadre d'analyse est résolument en accord avec les théories de pragmatique cognitive de Schmid (2012:303) qui proposent également d'identifier les liens entre pragmatique et cognition participant à la construction contextualisée du sens. L'interprétation du potentiel communicatif et/ou cognitif fluent ou disfluent des marques pour le locuteur-l'auditeur est établie sur base de leurs propriétés distributionnelles ainsi que de leur perception et leur évaluation par les auditeurs francophones natifs.

48. Nous entendons ici par distribution l'occurrence ou la cooccurrence positionnelle des éléments (avec possibilité, impossibilité d'insertion, de positionnement, de regroupement) (Mertens, Swiggers & Van den Eynde 1998).

49. Au niveau distributionnel, l'importance du regroupement en chaîne et la motivation vers une interprétation sociocognitive en terme de schéma sont appuyées par la littérature qui témoigne du fait que les (dis)fluences sont plus souvent actualisées en combinaison que de manière isolée (Candea 2000a ; Duez 1991 ; Grosjean & Deschamps 1975) (Voir également section 3 de ce chapitre).

2 Définitions des (dis)fluences

Cette section s'attache à définir et à décrire **de manière isolée** les (dis)fluence(s) étudiées dans cette thèse : les pauses silencieuses (2.1), les allongements (2.2), les pauses pleines (2.3) et les répétitions à l'identique (2.4).

Pour chacun de ces phénomènes, un bref état de l'art présente les paramètres essentiels à la définition, à l'évaluation et à la désambiguïsation de ces marques comme dispositif communicatif fluent ou disfluent, en fonction d'un faisceau de traits acoustiques, intonatifs, syntaxiques et situationnels, notamment eu égard à leur distribution et à leur impact sur la **continuité discursive**.

L'environnement lexical de ces trois phénomènes est envisagé à travers la combinaison de ces marqueurs ou la cooccurrence avec d'autres (dis)fluences additionnelles (e.g. répétitions avec modifications, amorces de mots, etc.) en section 3 de ce chapitre. Il permet de **formaliser les hypothèses générales** brièvement énoncées au chapitre 1 et communes à toutes les marques (section 4), tandis que des hypothèses spécifiques et une description plus fine de ces phénomènes sont données dans leur chapitre respectif.

2.1 Pauses silencieuses

Les pauses silencieuses – parfois appelées simplement *pauses*, pauses *vides* ou *non sonores* – sont une **nécessité physique** (*pauses respiratoires*) mais remplissent également des **fonctions pragmatiques**, **rhétoriques** (elles dramatisent le discours oral) et **linguistiques** (notamment syntaxique et intonative, de segmentation). Les pauses semblent en effet être des **dispositifs communicatifs** universaux régulés par des **contraintes spécifiques à la langue** ainsi que par le **contrôle du locuteur** (Hieke, Kowal & O'Connell 1983:214 ; Léon 1993:44-45).

La pause silencieuse ne correspond pas toujours à un **effort de formulation du locuteur** (Candea 2000a:106). Fox Tree (1995:709) rappelle en effet que « toutes les pauses ne sont pas des (dis)fluences ». Ce point de vue a déjà été mis en valeur par Riggenbach (1991:438) et Chambers (1997:539). Ces arguments sont soutenus par la mise en évidence de la **fonction expressive**, du rôle stylistique des pauses silencieuses – notamment dans la parole des hommes politiques français chez Duez (1985, 1991:28). Cette approche fonctionnelle et positive de la pause, ainsi que la **subjectivité inhérente à leur interprétation**, amène à se demander quelles pauses silencieuses doivent être prises en compte dans l'étude de la fluence-disfluence (Duez 1991:30). Les réponses varient en fonction des définitions de la « pause ».

2.1.1 Pause silencieuse : hétérogénéité des définitions

Il est difficile de comparer les études de pausologies, notamment en raison de leur **diversité terminologique**. L'état de l'art de Zellner (1994) le confirme. Cette disparité témoigne en réalité de l'hétérogénéité des phénomènes repris par les auteurs sous le terme de « pause » ou ***off-time*** . Pour illustrer ces recouvrements terminologiques hétéroclites : Maclay & Osgood (1959:24) y glissent les prolongations syllabiques, Goldman-Eisler (1961:167) y incluent les répétitions, Perfetti & Bertuccelli Papi (1985:345) y intègrent aussi toutes les « interruptions » (i.e. hésitations, corrections, paraphrases) et enfin Clark (1996:259) y inscrit toute suspension discursive même sans **interruption de la parole** (i.e (dis)fluences comme marqueurs discursifs).

Ces justifications d'inclusion ou d'exclusion évoquent soit la prétendue absence de **valeur informationnelle** de ces segments discursifs soit leurs **similitudes fonctionnelles et distributionnelles** (Guaïtella 1991:123 ; Candea 2000a:84,92 ; Campione & Véronis 2004:22). Si ces différences terminologiques et ces opérationnalisations méthodologiques se justifient souvent par des objectifs singuliers, elles entravent une **comparaison consistante** des résultats. C'est en partie pourquoi les pauses silencieuses sont parfois **exclues** de l'analyse des (dis)fluences malgré la fréquence importante du phénomène (Eklund 2004:160-161).

En français, les pauses silencieuses ont fait l'objet d'une analyse approfondie **dès les années 70-80**. Citons les études fondatrices de Grosjean & Deschamps (1972, 1973) portant sur un corpus d'interviews radiophoniques, mais aussi les travaux de Duez (1982, 1985, 1993, 2001c) s'appuyant surtout sur des discours et des interviews d'hommes politiques. On doit également beaucoup aux travaux de Blanche-Benveniste & Jeanjean (1986), Blanche-Benveniste (1987) et Blanche-Benveniste (1ère éd. 1997 ; 2010) qui se basent sur une grande variété d'enregistrements. Enfin, plus récemment, les analyses de Candea (2000a), de Campione & Véronis (2002, 2004, 2005) et de Henry, Campione & Véronis (2004) sur différents corpus ont également marqué et enrichi le paysage des études linguistiques francophones portant spécifiquement sur la pause silencieuse.

Nous adoptons une taxonomie analytique des (dis)fluences. Une pause silencieuse se définit avant tout comme une **« absence d'occupation du canal auditif »** (Duez 1993:25). Dans les termes de Zellner (1994:44), cette pause dite *objective* porte l'appellation de *pause physique*. Au reste, toute **pause physique** objective n'est pas nécessairement une **pause *perçue*** par les interlocuteurs. Par exemple, une pause de quelques millisec. n'est en général pas perçue comme une pause par les auditeurs (e.g. une pause « articulatoire » impliquée dans l'articulation d'une consonne occlusive)[50]. Il en va de même pour les pauses de longue durée au sein de l'interaction, qui ne bénéficient pas du même statut ni en production ni en perception (e.g. alternance de tour, silence, etc.).

50. Dans la littérature, de nombreux seuils minimum de durée ont été fixés, variant de 50 ms. à 350 ms. Un résumé des différentes propositions est fait par Candea (2000a:21-24). Nous exposons les challenges et les choix méthodologiques reliés à l'analyse des pauses silencieuses en détail au chapitre 4.

D'autre part, le locuteur peut également percevoir des ***pauses subjectives*** (Duez 1993:25), qui ne sont *pas* marquées par une absence d'occupation du canal auditif mais perçues comme des **interruptions du signal** par les auditeurs. Ce phénomène a été observé sur l'anglais au moins dès la fin des années 40 par Cowan & Bloch (1948:92) et plus récemment confirmé par Candea (2000a:123) en français – nous l'avons abordé en section 1.3 de ce chapitre.

En conclusion, de Goldman-Eisler (1968) à Goldman, Avanzi & Auchlin (2010), en passant par Duez (1993), la **non-congruence** entre les **pauses articulatoires, acoustiques** et **perceptuelles** est reconnue, tout comme le fait que certaines pauses sont mieux perçues par les locuteurs que d'autres, en particulier quand les interlocuteurs leur attribuent un rôle ou au moins une valeur fonctionnelle (e.g. interactionnelle, focalisante, structurante, stylistique, etc.).

2.1.2 Pause silencieuse : éléments de fluence dialogique

De nombreuses évidences empiriques existent notamment sur **le rôle structurant des pauses dans la parole**[51]. Prenons une seconde définition de la pause, reconnaissant son aspect structurant. Candea (2000a:23-24) désigne par pause silencieuse « un arrêt total de l'émission sonore à l'intérieur d'un énoncé », c'est-à-dire « **une interruption significative de toute émission sonore à l'intérieur d'une prise de parole** ». La définition de la pause silencieuse inclut donc ici **une dimension interactionnelle**.

Par conséquent, l'auteure exclut de l'analyse les pauses silencieuses présentes entre deux prises de parole de locuteurs différents[52]. En segmentant les unités de parole, certaines pauses agissent comme dispositifs de gestion de la **synchronie interactionnelle** des interlocuteurs (e.g. annonce d'un changement de sujet, retard ou problème de planification, intention de transfert ou maintien du tour de parole, révision ou abandon d'expression, etc.). Par extension, au niveau fonctionnel comme interactionnel, elles sont vues comme des **dispositifs de gestion cognitive** du flux discursif : elles permettent notamment au locuteur de produire un flux discursif **moins dense en « erreurs »** (Levelt 1989 ; Swerts 1998).

Nous adoptons cette **distinction interactionnelle** des pauses au sein de notre travail. A l'instar de Candea (2000a:24), nous ne fixons **pas de seuil maximum ou minimum** à priori des pauses pour la transcription des corpus. Toutes les pauses silencieuses (intratour

51. Renvoyons notamment à Candea (2000a,c) pour le français, et à O'Connell & Kowal (2004) pour une revue extensive de la littérature à ce sujet.

52. Ce positionnement interactionnel fait suite à l'article fondateur de Sacks, Schegloff & Jefferson (1974), et est reconnu par les propositions d'opérationnalisation de Edlund, Hirschberg & Heldner (2009) et Heldner & Edlund (2010) ainsi que de Fors (2011). Nous l'explicitons au chapitre 3 section 3.1.

de parole), même les plus longues, ont été prises en compte dans notre étude. Sauf explicitement mentionné, nous faisons de même au sein de ce travail : la seule exclusion sera liée à certains calculs statistiques afin d'éviter les **données aberrantes**[53].

2.1.3 Pause silencieuse : structurante /vs/ d'hésitation

Candea (2000a) adopte une **catégorisation fonctionnelle** des pauses, sur la base purement formelle du type d'environnement dans lequel elles s'insèrent. En reconnaissant que **les pauses silencieuses interagissent souvent avec d'autres (dis)fluences**[54], l'auteure est amenée à distinguer les pauses silencieuses *structurantes* /vs/ *non structurantes*.

La ***pause structurante*** se définit comme « **encadrée** par des séquences sonores appartenant au même locuteur et qui n'est pas immédiatement précédée par une marque de travail de formulation » (Candea 2000a:145). Au contraire, la ***pause non structurante*** est immédiatement **précédée** par un allongement ou un *euh* (« marque de travail de formulation ») ou insérée entre un répétable et un répété. Elle ne participerait pas à la hiérarchisation des constituants, et serait produite « **sans intention** » du locuteur, contrairement aux pauses silencieuses structurantes (Candea 2000a:178). Les pauses non structurantes constituent 17.5 % de son corpus contre 82% de pauses structurantes. Par ailleurs, Candea émet une réserve sur le fait que les pauses silencieuses **suivies** d'une marque du travail de formulation fassent en effet partie de **la même unité**. La marque du travail de formulation fait, selon elle, partie d'un **nouveau groupe** (et en occupe la première place sur l'axe syntagmatique).

Ensuite, au niveau microsyntaxique, l'auteure distingue les pauses en position **inter** /vs/ **intraconstituant** (Candea 2000a:166). Les pauses sont intraconstituantes lorsqu'elles s'actualisent entre un nom et son expansion. Ces dernières, si non structurantes (i.e. précédées par une marque) ne joueraient pas un rôle stylistique ou structurant mais seraient motivées par la **recherche de mots** : elles précéderaient un rajout, une expansion d'un contenu non programmé par le locuteur avant le début de la production de la pause. Selon Candea (2000a:270), **les pauses intraconstituantes** non structurantes constituent donc de réelles *dis*fluences.

Zellner (1994:42-43) différencie les types de pauses d'abord au **niveau syntaxique** : les pauses *intrasegmentales* des pauses *interlexicales*. Comme chez Candea, cette typologie a des **prolongements fonctionnels**. Certaines pauses sont ainsi dites *grammaticales*,

53. Il s'agit d'un choix méthodologique qui n'est pas sans implication théorique : dans le cas de certains chevauchements brefs de parole (dus à un *feedback* de l'interlocuteur de type « mh »), il est parfois malaisé de décider de l'inclusion ou l'exclusion d'une pause comme intratour ou intertour par rapport à la prise de parole d'un locuteur. Nous verrons au chapitre 4 que les rares cas d'exclusions à des fins statistiques sont empiriquement motivés sur base des mesures de durées des pauses du corpus (e.g. deux écarts types autours de la moyenne).

54. Renvoyons notamment à Clark & Tree (2002), Clark & Wasow (1998), Eklund (2004), Eklund & Shriberg (1998), Lickley (1994), Moniz, Mata & Viana (2007), Nakatani & Hirschberg (1994), Shriberg (1994) et Swerts (1998) pour une analyse prosodique des pauses et de leur insertion au sein de « séquences de (dis)fluences ». Ces éléments sont repris de manière détaillée au chapitre 4.

pauses de structure ou de jonction (Goldman-Eisler 1958 ; Barik 1968:157). Celles-ci s'insèrent entre les clauses et sont associées aux **traitements de l'information** tels que les planifications sémantique et structurelle du discours. Les pauses qualifiées de *non-grammaticales* sont souvent associées aux pauses d'*hésitation*.

Encore une fois, cette **catégorisation syntaxique** se traduit en **termes fonctionnels et informationnels** : les pauses en frontières d'unités syntaxiques sont décrites comme des lieux propices de transitions (changement de sujet ou de tour de parole) (*Transition Relevance Place*) (Sacks, Schegloff & Jefferson 1974 ; Morel & Danon-Boileau 1998:164-234). Leur actualisation au sein d'une clause suggèreraient une sélection plutôt précipitée d'un item lexical, ce qui les **rendrait perceptivement plus marquées** que les pauses dites *grammaticales* (Barik 1968:158 ; Morel & Danon-Boileau 1998:79 ; Reich 1980:380). Ces profils de (dis)fluences en fonction du contexte syntaxique d'occurrence de la pause semblent refléter des régularités prosodiques de durées.

2.1.4 Pause silencieuse : contextes syntaxiques en fonction des profils de durée

Cette catégorisation en fonction de critères syntaxiques rappelle les observations de Duez (1991) notant des différences significatives dans la durée des pauses silencieuses (entre 200 à 700 ms. de différence selon les sous-corpus envisagés) en fonction des positions possibles (intraconstituant, interconstituant, en fin de proposition ou en fin d'énoncé). Les observations de Candea (2000a:165,198) rapportent que (1) la durée des **pauses structurantes** est significativement **plus longue** que celle des pauses non structurantes, et que (2) la durée des pauses de **fin d'énoncé** est très largement supérieure à celle en fin de proposition ; les pauses en fin de proposition étant légèrement plus longues que les interconstituants. En somme, plus la **frontière syntactico-discursive** est importante, plus la **pause silencieuse est longue**. Cette tendance semble également se confirmer en anglais (Goldman-Eisler 1972).

L'**argument cognitif** ci-avant serait soutenu par des **manifestations prosodiques** : les noms et leurs expansions sont organisés comme deux constituants indépendants, le premier nom étant majoritairement marqué par un contour intonatif descendant (frontière finale) dans son corpus. Au contraire, il existe très peu d'occurrences de pauses silencieuses structurantes (non accompagnées de (dis)fluences) intraconstituantes. Selon Candea (2000a:173-174), cette observation favorise l'hypothèse selon laquelle toutes les **pauses silencieuses non précédées de « marques du travail de formulation »** seraient des **pauses structurantes**. Les rares **pauses structurantes intraconstituantes** ne sont en général pas accompagnées d'allongement ou de contours continuatif ou terminatif ; elles ne forment donc qu'un constituant sur les plans segmental et prosodique. L'auteur leur attribue donc davantage un **effet stylistique**. Ces distinctions à priori entre les types de pauses en fonction de leur environnement syntaxique et de leur durée auraient donc **des effets sur la perception des pauses**.

Les pauses ont par association une probabilité différente d'être perçues en fonction notamment de leur catégorie morphosyntaxique et de leur **localisation grammaticale**[55]. Candea (2000a:113,125) remarque que **la position en frontière grammaticale favorise la perception des pauses** (même des *pauses subjectives*), alors que 75% des auditeurs ont perçu les pauses distribuées en frontière syntaxique, seulement 39% des **pauses intraconstituantes** ont été détectées par ces mêmes auditeurs – cette observation doit cependant être considérée avec prudence, le protocole expérimental utilisant un support écrit – la transcription peut favoriser la perception des pauses silencieuses grammaticales démarcatives (Duez 1991:106). Toutefois, un découpage en unités énonciatives suivant le modèle de Morel & Danon-Boileau (1998) donne des résultats similaires : 69.5% de détection lorsqu'elles sont placées **en fin de *préambule* ou de *rhème*** contre 45% à l'intérieur du préambule ou du rhème.

Les **frontières syntaxiques** elles-mêmes (limites de propositions grammaticales, limites de syntagmes, etc.) **induisent la perception** d'une **pause subjective** ou favorisent la perception de la pause réelle située dans une telle configuration. Dans un protocole expérimental de Duez (1995), les pauses silencieuses ne sont jamais notées comme des hésitations par au moins 75% des sujets (contrairement aux allongements et aux répétitions pour lesquels 52% d'entre eux ont au moins été perçus par 75% des sujets comme des hésitations). Candea (2000a:129,178) interprète ce résultat comme suit : « Les auditeurs perçoivent l'**indice de recherche de formulation** produit par le locuteur et ne perçoivent pas la [pause silencieuse] subséquente en tant que pause, car elle **ne correspond pas aux représentations** qu'ils ont de la pause principalement comme indice démarcatif de constituants ».

Ces observations permettent de rappeler, à l'instar de Bell, Eklund & Gustafson (2000) que, le degré de fluence d'une pause varie de « **disfluence infaillible** » (e.g. lorsqu'elle vient interrompre un item lexical) à un « **outil conscient et intentionnel de structuration** » (e.g. lorsqu'elle s'insère entre des unités syntaxiques grammaticalement complètes). Selon Riggenbach (1991), même les pauses apparaissant au sein de la phase d'édition de (dis)fluence n'ont pas un rôle si univoque dans la perception de la *dis*fluence. Une catégorisation à priori sur l'échelle de la (dis)fluence sans une analyse du contexte discursif semble impossible, et même après, elle peut rester souvent douteuse. Et, si la pratique habituelle veut que les pauses « longues » et/ou intraclausales soient automatiquement labellisées « disfluentes », cela n'est pas sans poser problème : une **pause longue intraconstituante** peut, en contexte, avoir également un **effet rhétorique** ou **humoristique**.

En somme, les études antérieures ont permis d'éclaircir les différents types de pauses silencieuses présentes dans la parole, notamment en fonction de **critères interactionnels** et **syntaxiques**. Elles ont donc balisé la formalisation du contexte d'actualisation

55. surtout dans le cas de celles relativement courtes : 50-200 ms. chez Cowan & Bloch (1948) et 50-110 ms. chez Martin (1970). Nous dédions entièrement la section 1 du chapitre 4 à la question épineuse de la catégorisation des pauses en fonction de leur durée.

des pauses en production en leur associant des propriétés distributionnelles et fonctionnelles, sans pour autant qu'il y ait consensus : une pause non grammaticale n'est pas toujours plus courte et n'est pas toujours interprétable comme une pause d'hésitation. Une pause interconstituante n'est pas nécessairement plus longue et/ou signe de structuration. Nous proposons au chapitre 4 une nouvelle modélisation et catégorisation des pauses, notamment en fonction de leur durée relative dans leur environnement et d'évaluer le croisement de ces propriétés pour établir leur statut de dispositif communicatif plus ou moins fluent.

2.2 Allongements vocaliques

L'allongement vocalique comme marque d'hésitation est souvent exclu des études en raison de son **absence dans les transcriptions** mais aussi de la plus grande **variabilité interannotateurs** à laquelle le phénomène est soumis (Dister 2007:210-211). Nous avons la chance de disposer de riches corpus, permettant d'effectuer cette distinction entre pauses pleines et allongements vocaliques. Cette annotation permet d'évaluer le degré de *dis*fluence que peut engendrer l'actualisation d'un allongement. Cependant, elle pose aussi la question de la distinction prosodique et distributionnelle entre des **allongements vocaliques** dits *d'hésitation* /vs/ *grammaticaux*.

2.2.1 Allongement vocalique : grammatical /vs/ d'hésitation

L'allongement vocalique d'hésitation peut s'actualiser sur tout type de voyelles. Il serait particulièrement fréquent en français en raison de la proportion importante de syllabes ouvertes [CV] (Grosjean & Deschamps 1975). Une première définition des **allongements vocaliques d'hésitation** est reprise ci-après :

> « Tout allongement vocalique anormal en position finale de mot ou d'amorce de mot, présentant un contour plat et bas ou très légèrement descendant, représente un allongement marquant le travail de formulation en cours. Une voyelle commence à être anormalement allongée lorsque sa durée se situe entre 18 et 22 cs. selon les locuteurs. » (Candea 2000a:24-26)

Le terme « anormal », souligné par l'auteure dans le texte, tient à distinguer **les allongements vocaliques grammaticaux**, codés par le système phonologique de la langue française, des allongements vocaliques d'hésitation, moins prévisibles et, marqueurs d'une certaine *dis*fluence.

L'allongement grammatical est motivé par plusieurs raisons. L'accentuation, marquée par la durée, provoque un allongement des syllabes. Vaissière (1991) rapporte l'allongement relatif des syllabes accentuées surtout lorsqu'elles sont porteuses **d'un contour intonatif continuatif** – montant, parfois avec un crochet descendant – ou **d'un contour**

intonatif final – accent démarcatif avec une F0 descendante[56]. D'autres types d'accents s'actualisent aussi par un allongement, mais de manière moins marquée ; notamment, la tournure exclamative (avec une F0 maintenue dans un registre haut) et certaines emphases en finale de groupe ou de segment (Léon 1993:45).

2.2.2 Détection et perception des allongements

L'allongement vocalique et les allongements en général jouent un rôle essentiel pour la perception du rythme. En conviendront Fraisse (1974) et Di Cristo (1985) ou Duez (1991, 1993), pour lesquels l'allongement a un **rôle démarcatif** : il tend même à être **perçu comme une pause** ou comme une **marque de frontière** d'un groupe rythmique (Candea 2000a:133). Plusieurs méthodologies de détection ou d'identification des allongements d'hésitation ont été testées : d'une annotation complètement perceptive à une détection automatique en fonction d'un palier fixé par le chercheur, absolu ou dépendant de chaque locuteur.

Duez (1987:222), citée par Candea (2000a:263), soutient qu'un allongement final de 50% d'une voyelle la rend perceptible en tant que **marqueur de frontière**[57]. En l'absence de pause réelle, l'allongement sera plus probablement perçu comme une pause silencieuse (et donc une pause dite *subjective*)[58]. Léon (1993:95) cite Rossi (1999:2) et postule lui qu'un allongement est détecté à condition d'un accroissement minimum de la durée de l'ordre de 20% par rapport aux syllabes environnantes[59].

Candea (2000a:264) adopte un **seuil modulable** par locuteur. Un allongement est défini par la durée moyenne des noyaux syllabiques « longs », à laquelle elle ajoute un écart-type[60]. Cette méthode concilie **approche mathématique** et **perception humaine** : sur les 487 allongements vocaliques statistiquement significatifs, sur base d'une écoute et de ces seuils modulables, elle réduit à 284 les syllabes analysées comme porteuses d'un allongement d'hésitation significatif.

Par ailleurs, Candea (2000a:265) applique la démarche de Grosjean & Deschamps (1972-1975). Celle-ci consiste à ne permettre qu'aux **mots grammaticaux** d'être porteurs d'un allongement. Selon cette méthode, 88% des allongements vocaliques sont détectés automatiquement dans les corpus ; évitant une intervention humaine dans une

56. Cet allongement reste faible : il ne dépasse par exemple jamais le premier seuil de durée fixé par Candea (2000a:262) pour la sélection des allongements vocaliques, marques de travail de formulation.

57. L'auteure citant ne précise cependant pas, à notre connaissance, la manière dont cet allongement relatif est calculé, par rapport à une moyenne, aux noyaux vocaliques antécédents, etc. ?

58. Cela va de soi, surtout si la place d'une frontière est attendue par l'auditeur : celui-ci est actualisé sur un MP et non un MO, et plutôt en fin de constituant ou de syntagme (voir section 2.1 de ce chapitre).

59. L'auteur ne précise pas le nombre de syllabes environnantes prises en compte. Nous renvoyons au chapitre 3 section 2 pour la méthodologie d'annotation des allongements adoptée dans le cadre de cette recherche.

60. Le noyau syllabique « long » est défini pour chaque locuteur sur base de la durée moyenne du noyau + un écart type. Cependant, cette méthode est sujette à controverse : si la distribution de la durée des allongements ne suit pas une courbe normale, la prise en compte de la moyenne comme mesure statistique de base invite au débat.

tâche de compréhension. En ajoutant une contrainte prosodique telle que **l'interdiction de la présence d'un contour intonatif fortement modulé** (i.e. l'allongement vocalique est porteur d'un contour intonatif plat ou résolument descendant) – alors ce taux de réussite de catégorisation passe à 94.5% de détection. Les marges d'erreur restantes concernent **5% d'allongements d'hésitation s'actualisant sur des MP**[61], et quelques allongements stylistiques exceptionnels. Ainsi, coupler un critère intonatif et morphologique permet d'obtenir de très bons taux d'étiquetage.

Dans ce travail, nous adoptons une méthodologie d'annotation des allongements d'hésitation basée sur la **perception d'annotateurs** (uniques, multiples, « naïfs » ou « experts » en fonction des corpus). Nous décrivons les contextes prosodiques et syntaxiques de ceux-ci ainsi que leurs cooccurrences avec d'autres marques[62].

2.2.3 Contextes syntaxiques et prosodiques des allongements

Si pour beaucoup d'auteurs, seuls les mots grammaticaux sont porteurs d'allongements, alors par définition les allongements se situent **en début ou à l'intérieur de constituants**. Lorsque ceux-ci sont en finale de constituant, ils marqueraient un **inachèvement** (exemple 14). Pour Candea (2000a:283), c'est la position intraconstituant qui est privilégiée pour l'actualisation des *allongements vocaliques*.

Au niveau de l'intonation, la littérature rapporte que l'allongement vocalique d'**hésitation** se distingue de l'allongement vocalique phonologique **démarcatif** par ses contours intonatifs – ce dernier ayant un contour intonatif haut et souvent dynamique, modulé. L'allongement vocalique d'hésitation s'actualise dans **un registre de F0 relativement bas** et au **contour intonatif plat** (Vaissière 1991 ; Lacheret-Dujour & Beaugendre 1999 ; Candea 2000a:27,263). Par rapport à la distribution des *euh*, la position en début de constituant intonatif est privilégiée par les allongements vocaliques d'hésitation – ainsi les MO sont plus souvent allongés que suivis d'un *euh* (Candea 2000a:283).

Exemple (14) – issu de Candea (2000a:281)

a. essaie ↗ de de: /30/ de le mettre avec une↗ toute↗ petite voix /550/ (Boucle, Prof, lignes 45-46)

b. ET: euh::: les fils deVAIENT la lancer:↗ et: LA où tomberait la plume↗ /40/ (Père fils, lignes 226-227)

c. la maman ours↗ qui dit la même chose↗ /80/ et euh : le petit BEbé:↗ il dit:↗ /35/ (Boucle, lignes 52-53)

61. Pour rappel les mots grammaticaux, associés également aux « mots vides », « mots clitiques », s'opposent aux mots lexicaux (ou mots pleins, non-clitiques). Ils reprennent des catégories grammaticales en général non accentuables en français, telles que les prépositions et les déterminants. Voir notamment chapitre 2 section 2.2.3.

62. Voir chapitre 5.

d. sui↗vre le petit: euh:: le petit lutin↘ /75/ et: euh: le petit lutin↗ le soir↗ il fit un grand feu↗ et chanta↗ une chanson /25/ (Nain, lignes 366-367)

Pour rappel, au niveau de la durée, l'allongement vocalique d'hésitation serait plus long que celui marquant des frontières fluentes (Shriberg, Stolcke & Baron 2001:161). De même, selon Guaïtella (1991:114), les allongements d'hésitation sont un phénomène où sur une syllabe « se manifeste un allongement de durée, sans qu'il y ait modification du contour de F0 ». Si l'allongement s'accompagnait d'un **décrochage de la F0**, il serait assimilé à un accent. Seul le **contour plat** lui permet d'être associé à un allongement vocalique d'hésitation[63].

Au niveau fonctionnel, ces allongements vocaux sont associés par Guaïtella (1991b) à l'**activité de planification** et de **conceptualisation** des idées. Cependant, ils peuvent également servir au maintien du tour de parole par le locuteur[64].

L'allongement vocalique est un phénomène phonétique très souvent associé et lié à l'analyse de la pause pleine. S'ils apparaissent assez souvent concomitamment, leur **assimilation** sous une étiquette commune de « pauses remplies » semble parfois **faiblement justifiée** en théorie. La similitude de leurs propriétés de fonctionnement (acoustique) (Candea 2000a:88 ; Dister 2007:210 ; Campione & Véronis 2004) ne gomme pas entièrement leurs singularités distributionnelles, justifiant la **distinction énonciative** des deux phénomènes, notamment selon Morel & Danon-Boileau (1998)[65].

63. Cette hypothèse est vérifiée dans le cadre des différentes méthodologie d'annotations perceptives effectuées dans nos corpus (voir chapitre 3 section 2.1 et chapitre 5 section 3).

64. Notons que les conclusions de Duez (1991) accordent une importance plus grande à la chute de la fréquence fondamentale pour la prise d'un tour de parole qu'aux (dis)fluences. Cette hypothèse est donc selon Candea (2000a:85) à réévaluer dans le cas du français.

65. Voir section 2.3.1 de ce chapitre.

2.3 Pauses pleines

La *pause pleine* (Auchlin, Goldman, Simon & Avanzi 2018 ; Bachy et al. 2007 ; Bazillon et al. 2008 ; Bertrand et al. 2008 ; Simon & Christodoulides 2016), appelée parfois *pause non silencieuse* (Duez 1985, 1991 ; Lacheret-Dujour & Beaugendre 1999), mais aussi *pause remplie* (Blanche-Benveniste 1990a) ou *pause sonore* (Grosjean & Deschamps 1972, 1975), correspond à une marque explicite d'« hésitation » **quasi-lexicalisée** et **conventionnalisée**[66].

La pause pleine est une petite marque monosyllabique qui correspond à une :

> « [...] *voyelle prononcée* [ə, œ, ø voire ɛ̃ ou œm] pouvant être soit insérée en épenthèse en finale de mot et formant une syllabe supplémentaire, soit prononcée indépendamment, avant ou après un mot. Dans ce dernier cas elle peut être séparée du mot précédent et éventuellement du mot suivant par une interruption du signal sonore plus ou moins longue, ou bien, dotée d'une montée de l'intensité qui en fait une syllabe autonome par rapport à la séquence précédente. Cette voyelle représente l'indice le plus largement connu pour marquer conventionnellement la recherche de formulation en français oral. » (Candea 2000a:24)

Au niveau phonostylistique, les études de Fónagy (1989), Léon (1969) et Mosegaard-Hansen (1997) mettent en valeur le fait que le *euh* d'hésitation est favorisé par **un style « animé et peu formel »**. Il serait par ailleurs davantage prononcé par les jeunes et les femmes. De manière générale, Lacheret-Dujour & Beaugendre (1999:53) associent à ces « pauses sonores » un « **processus d'attente** et un moyen utilisé par les locuteurs pour augmenter le délai nécessaire à l'encodage de l'unité suivante ; indiquant l'inachèvement, [les pauses sonores] constituent la trace d'une activité de l'**élaboration verbale** ». Bien entendu, l'inachèvement n'est pas à confondre ici avec la *rupture* ou la *dis*fluence – on se le confirme quand on remarque que dans les corpus VALIBEL – ces *euh* sont rencontrés en moyenne une fois tous les 47 mots tous genres confondus (Dister 2007:233).

Après cette brève définition, nous désirons revenir sur la distinction *euh* d'hésitation et *euh* d'appui, ce qui éclaire déjà le lecteur sur les fonctions attribuables à cette marque (majoritairement structurantes, énonciatives, continuatives). Nous revenons ensuite sur les contextes syntaxiques et prosodiques dans lesquels cette marque peut être observée et sa relation avec l'**échelle de continuité /vs/ rupture** (acoustique, prosodique, syntaxique, discursive, énonciative), ainsi que sur les hypothèses phonostylistiques et sociocognitives reliées au **continuum fluence-disfluence** décrit en section 1.4.2 de ce chapitre.

66. Celle-ci possède en effet une entrée dans la pluspart des dictionnaires francophones sous *euh* ou *euh*. La locution « pause pleine » s'est également répandue à l'instar de celle de *pause remplie* par analogie avec la traduction anglophone les désignant sous le terme de *filled pauses* ou *fillers* (Brennan & William 1995 ; Shriberg & Lickley 1993 ; Swerts 1998).

2.3.1 Pause pleine /vs/ allongement vocalique

L'allongement et la pause pleine sont parfois regroupés sous la locution de *pauses remplies*. Comme mentionné ci-avant, les pauses pleines (*euh* d'appui inclus ?), allongements vocaliques (grammaticaux inclus ?) sont souvent confondus et/ou **traités ensemble** car ayant le même rôle et/ou les mêmes caractéristiques prosodiques. C'est le cas chez Guaïtella, mais également chez Duez (1991:71-78), le classement pauses pleines /vs/ allongements vocaliques est signalé en début d'ouvrage mais ensuite regroupé (Candea 2000a:84). Chez d'autres, la question de l'allongement est **ignorée** – plus en raison d'obstacles pratiques que de motivations théoriques (accès aux enregistrements, lourdeur d'annotation) (Dister 2007).

D'un point de vue théorique, chez Morel & Danon-Boileau, le *euh* et l'allongement ont des **caractéristiques prosodiques** communes et donc des **similarités iconiques** :

1. **durée supérieure** à celle d'une syllabe accentuée : signifiant un désir de continuation – que celle-ci soit motivée par un ajout ou une recherche de contenu (lexical et/ou syntaxique).
2. **plage de F0 basse** : associée à un repli sur soi, dû au fait que l'énonciateur se concentre sur la formulation de son propre discours.
3. **maintien de l'intensité** : signifiant à l'auditeur que le locuteur désire conserver la parole.

Cependant, Morel & Danon-Boileau (1998) font également l'hypothèse qu'à chaque (dis)fluence correspondent des distributions et des **fonctions énonciatives différentes**[67]. Résumons :

– L'allongement vocalique servirait ainsi à « **gérer la suite immédiate du discours** », il a une incidence uniquement locale sur la recherche de mot et a donc une **fonction prospective** (Morel & Danon-Boileau 1998:80). Il a donc une fonction d'opération modale, marquant la réévaluation du référant, une inflexion du point de vue. Au niveau distributionnel, il affecte surtout les éléments ligateurs et les verbes.

– Au contraire, le *euh* serait non pas une opération modale mais bien **référentielle**, il porterait donc plus sur des segments d'énoncé que sur des mots précis (Candea 2000a:91). Il ne marque pas le « manque de mot » mais « la difficulté à **organiser la relation** entre le segment déjà énoncé et le segment à venir », il a donc également une **fonction rétroactive** (Morel & Danon-Boileau 1998:82)[68]. Le mot précédant le *euh* déterminerait le statut et le rôle des segments d'énoncés à venir, le *euh* peut donc à lui seul jouer le rôle d'un « ligateur discursif ».

67. Voir section 1.3 de ce chapitre.
68. Il existerait des contextes où les locuteurs utilisent ces deux marques, dans le cas d'un « accompagnateur de ligateur » – bien que la nuance de fonction dans le cadre de distribution identique ne soit pas explicitée par Morel & Danon-Boileau (1998:83).

Candea (2000a:269) a alors comparé la distribution des *euh* et des allongements par rapport à la **catégorie morphosyntaxique** qui les précède directement. Nous retiendrons notamment :

1. A fréquence relative égale, les deux marques s'actualisent **davantage sur les MO** que sur les MP (i.e. six fois plus).
2. Les **marqueurs de discours** (ici des ponctuants-ligateurs) sont fréquemment porteurs d'une des marques voire des deux. Plus précisément, il n'existe pas de différence de pourcentage entre la probabilité d'avoir un ligateur (*donc*, *euh*) suivi d'un *euh* ou porteur d'un allongement d'hésitation.
3. La **nature de la voyelle** ne permet pas de prédire la forme d'une (dis)fluence (allongement ou *euh*). Cependant, la **structure syllabique** est pertinente. Candea (2000a:91) observe très peu d'allongements en syllabe fermée (de type [CVC]) qui privilégient les syllabes ouvertes (de type [CV]). Au contraire, les syllabes [CVC] sont davantage suivies de *euh*.

Malgré leurs prétendues similitudes, sur base de ces éléments distinctifs, nous faisons le choix dans nos analyses de distinguer à priori les pauses pleines des allongements vocaliques, et de justement reconnaître que ceux-ci pourraient avoir des fonctions similaires dans certains contextes mais d'être **regroupés** sous une autre catégorie dans d'autres – notamment lorsqu'ils se combinent sur un même item linguistique. En somme, nous partons du principe qu'il puisse exister un continuum allant du schwa épenthétique rythmique, à celui de *euh* de maintien et d'appui, en passant par le *euh* d'hésitation – et même, que leurs actualisations fonctionnelles et distributionnelles peuvent se combiner.

2.3.2 Pause pleine : d'appui /vs/ d'hésitation – continuité ou rupture ?

> « Il est classiquement admis que le *euh* 1) constitue une pause dans le flux de la parole et 2) marque une hésitation de la part du locuteur. Nous l'avons dit, ces deux « constats » se retrouvent fréquemment dans la littérature qui aborde le sujet, parfois sans aucune justification ni théorique, ni même empirique. » (Dister 2007:211-212)

Comme le fait remarquer Dister (2007), la production des *euh* dits d'hésitation, tout comme d'autres (dis)fluences, représente l'**activité métacognitive** qui accompagne l'activité de traitement de la parole. En termes de modélisation de la parole, le *euh* serait majoritairement lié à un malaise lors des **processus d'encodage** de la parole (Duez 1991:17 ; Brennan & William 1995). C'est une des raisons pour laquelle on lui associe souvent automatiquement l'étiquette *phénomène d'hésitation*. Quelques auteurs ont cependant remis en cause cette terminologie.

En effet, le lien des *euh* avec une activité métacognitive n'amoindrit pas leur **nécessité** en production. Au même titre que les pauses silencieuses, les pauses pleines participent à la **structuration discursive** du point de vue du locuteur comme de l'auditeur, voire à la **structuration interactionnelle** (Cutler, Dahan & Van Donselaar 1997 ;

Lacheret-Dujour & Beaugendre 1999 ; Swerts 1998). Au niveau **interactionnel**, Morel & Danon-Boileau (1998:82) notent qu' « un locuteur n'est jamais interrompu après la production d'un *euh* » – bien que, lors de l'analyse d'un corpus de récits Candea (2000a:42) relève des exceptions à cette règle. Elle en conclut que « un *euh* n'est jamais interprété par l'auditeur comme une marque de clôture et comme une invitation à prendre la parole » et que par conséquent un locuteur venant de produire un *euh* ne sera **pas plus interrompu** qu'un locuteur qui marque un léger arrêt au sein d'une structure encore inachevée.

Les pauses pleines ne servent donc pas uniquement à « marquer un temps dans la production d'un énoncé » mais aussi à « indiquer à l'auditeur que ce temps est clairement **destiné à poursuivre l'encodage** et non pas par exemple à céder la parole à l'auditeur » (Dister 2007:212). Le *euh* est donc avant tout un indice d'**intention de continuation** plus que celui de recherche de mots (Morel & Danon-Boileau 1998:82 ; Clark & Clark 1977).

Comme note Candea (2000a:24), il est parfois peu aisé de distinguer les *euh* d'hésitation /vs/ d'*appui*[69]. Cependant, en théorie, les deux *euh* se ditingueraient selon leur durée :

- Le *euh* d'hésitation est un *euh* autonome formant une seule **syllabe séparée** par une **interruption du signal sonore**. Ce *euh* va de 150 ms. à 1 sec. selon la littérature (Candea 2000a:42).
- Le *euh* d'appui, est un *euh* épenthétique, bref et surtout, **intégré phonétiquement** à la finale du mot. Il serait plus matérialisé par le son [@], et est relativement fréquent et typé du français de région parisienne.

Certains auteurs argumentent que ces deux types de marques se distinguent totalement (Mosegaard-Hansen 1997). Par exemple, Guaïtella ne voit dans le *euh* épenthétique qu'un processus de **mise en valeur lexicale** ou d'**équilibrage rythmique** (Guaïtella 1991:125). D'autres le posent sur un continuum de forme insegmentable : une épenthèse vocalique peut parfois être **aussi** un *euh d'hésitation* (Carton 1999:37-42 ; cité par Candea 2000a:25,505).

> **Exemple (15)** – issu de Candea (2000a:25) : « Dans les deux exemples suivants nous avons souligné les *euh* autonomes et nous avons symboliquement relié par un trait « _ » les *euh* associés à une finale de mot. »
>
> a. /130/ euh:: il était une fois↘ /50 (h expir)/ dans une petite maison↘ /100 (hinspir)/ euh:: dans les bois:↘ /50/ ts /60/ une famille ours euh /330 (toux)/ (Boucle, lignes 2-4)
>
> b. un soir↘ /20cs/ ils viennent_euh et tout ils se cachent derrière_euh: la porte de l'étal↘ (Cordonnier, lignes 201-202)

69. Ce problème ne se pose pas lorsque le *euh* est prononcé en début de tour de parole ou précédé par une pause silencieuse.

La question initiale est en effet également de savoir si ces *euh* se distinguent sur le plan de **leur forme**. Selon Carton (1999), rien si ce n'est l'intensité ne permettrait de distinguer ces deux types de *euh*. Les deux *euh* se distingueraient également en fonction de leur durée. Candea (2000a:207,209) applique donc à priori des critères rigoureux de distinction dans ses analyses de corpus – où les *euh* d'appui représentent 13% des marques contre 87% de *euh* d'hésitation :

> « Si la durée d'un tel *euh* [d'appui] dépasse le seuil d'une syllabe longue située en fin de groupe rythmique et dotée d'une **intonation continuative** (environ 20cs, en général) alors le *euh* en question sera étiqueté comme marque de TdF [travail de formulation] ; ou si l'intensité d'un tel *euh* est plus forte que celle de la voyelle accentuée qui le précède alors le *euh* en question sera étiqueté comme marque de TdF ; en revanche, si la durée d'un tel *euh* est inférieure à la durée moyenne d'une syllabe longue accentuée du même locuteur et que son intensité est inférieure à la syllabe en question sera étiqueté comme *euh* d'appui. » (Candea 2000a:207-208)

Les *euh* seraient distinguables non seulement par leur durée mais également **leur mélodie** : les *euh* d'appui s'amorcent par une déclinaison de la F0 (plus ou moins rapide) et continue, parfois vers un palier intonatif bas et totalement plat ; ils ne sont jamais plus hauts que la voyelle qu'ils précèdent, contrairement aux *euh* d'hésitation.

Dans notre démarche, nous ne distinguerons pas à priori ces marques fonctionnellement selon leur forme. Pour rappel, ceci renvoie à une autre problématique plus vaste : **une forme** ne pourrait-elle pas avoir **plusieurs fonctions** ou une fonction **plusieurs actualisations de surface** ? Nous n'opérons pas cette distinction également en raison du fait que la définition du *euh* épenthétique d'appui peut elle-même également être présentée comme une catégorie d'allongement (d'hésitation ou grammaticaux) [70].

2.3.3 Pause pleine : contextes syntaxiques – autonomie /vs/ dépendance

Candea (2000a:247-250) formule l'hypothèse que le *euh* est une marque autonome (i.e. pouvant s'actualiser dans n'importe quel contexte contrairement à l'allongement ou à la répétition) dont l'actualisation ne serait pas tant corrélée à la **structure syntaxique** qu'à la **structure énonciative**.

Nous nous intéressons aux différentes actualisations du *euh* : distributions situationnelle, syntaxique, prosodique et forme acoustique qui lui sont associées. Ci-après, nous nous attacherons au contexte d'actualisation des *euh* isolés. La section 3 de ce chapitre s'intéressera aux cooccurrences et à leur impact sur la rupture-continuité du discours. Gardons à l'esprit que le *euh* montre une **forte interaction avec d'autres (dis)fluences**. Une proportion de « 60 pourcent des pauses remplies sont précédées ou suivies par un autre type de pause » chez Grosjean & Deschamps (1972:51). De même, dans un sous-corpus de 414 occurrences, Dister (2007:231) remarque que 91% des *euh* se trouvent

70. Voir section 2.2 de ce chapitre.

au sein d'une chaîne de répétition ou d'autocorrections, et, parmi les 9,37% restant, 26 occurrences induisent une **rupture syntaxique** contre 13 une **continuité intra-syntagmatique**[71].

La fonction de rupture /vs/ continuité du *euh* semble à géométrie variable selon la littérature en raison des profils très différents que le *euh* peut adopter. Illustrons. Dans l'exemple 16, les *euh* [1] et [3] n'ont pas de répercussion sur la **linéarité de l'énoncé**. Les [5] et [6] sont **à l'intérieur** respectivement d'une répétition et d'une suite complexe. Enfin, les occurrences [2] et [4] indiquent une **rupture** respectivement d'un groupe syntaxiquement complexe et d'une rupture plus éloignée du *euh*. Nous allons considérer les différents contextes syntaxiques permettant d'interpréter les fonctions de ces *euh*.

Exemple (16) – issu de Dister (2007:234)

> oh c'était -| à c'était l'horreur hein attends on va chez les Espagnols il vient avec quoi enfin c'est l'horreur quoi tu vois tu te souviens |- de ça <blaND1> mm -| / enfin qu'est qu'est-ce qu'il fout là quoi / et et je vois les photos et je dis ah ah un intrus quoi enfin c'est vraiment et tout le temps quoi / les sorties c'était hors de question je m/ / combien de fois je me souviens **euh** [1] il y avait une soirée carnaval-là |- / et <blaND1> mm -| c'était c'était je vivais un peu enfin ce que mes parents **euh** [2] / enfin je sais |- pas si <blaND1> mm -| c'était c'est le le gros stéréotype de la fille qui reproduit mais **euh** [3] / enfin tu vois **euh** [4] les je me souviens d'une soirée carnaval où tu avais même été super déçue et / et moi j'étais super mal parce que / on on devait y aller puis Mick **euh** [5]/ Mick c'est du genre oui parce qu'il j/ **euh** [6] j/ je n'y allais pas sans lui c'était sûr je ne sais pas |- pourquoi je n'y allais pas <blaND1> mm -| sans lui (blaAT1 – Corpus VALIBEL)

La distribution du *euh* serait ainsi révélatrice « des positions dans la structure énonciative où les locuteurs éprouvent plus le besoin de marquer leur travail de formulation en cours » (Candea 2000a:248). Certains, tel Luzzati (1985), formulent l'hypothèse que ce *euh* a une fonction énonciative structurante, d'**appui du discours**. Ce qui justifierait sa **position finale de constituant**, tel un *hein, bon, quoi*. Morel & Danon-Boileau (1998) considèrent plutôt que le *euh* s'approcherait du fonctionnement de ligateurs énonciatifs, tels les marqueurs *ben, alors, donc*, et qu'il adopterait ainsi plus une **position intiale** de constituant de manière absolue ou en cooccurrence avec ces ligateurs.

Selon les observations de Candea (2000a:262), la moitié des *euh* sont observés « **en dehors d'un bloc syntaxique** » – entre deux pauses silencieuses, entre deux constituants juxtaposés, après un connecteur[72]. Selon Campione & Véronis (2004), les *euh* sont majoritairement (81%) positionnés au sein de segments discursifs (des unités « relativement brèves »). Effectivement, les *euh* s'actualisant au sein d'un bloc syntaxique (en cours de complétion) précèdent en général des « constituants fortement **prévisibles** à partir du contexte avant », tels que des **compléments essentiels**. Candea (2000a:237)

71. Nous y revenons en section 3 de ce chapitre.
72. Plus d'un ligateur sur 5 est accompagné d'un *euh* d'hésitation (Candea 2000a:237).

observe quelques *euh* précédant des constituants facultatifs (et non prévisibles). Selon elle, ceci montre que les contraintes syntaxiques entre un constituant et le *euh* qui lui succède sont moins fortes qu'entre un constituant et un autre constituant « réel ».

En somme, le *euh* est un outil pour le locuteur lui permettant de laisser ouvertes toutes les **possibilités syntaxiques** de continuation de son discours. Le *euh* signale qu'un constituant est susceptible de venir compléter une structure inachevée. Au niveau morphosyntaxique, ces *euh* seraient distribués plus fréquemment devant des mots lexicaux que grammaticaux (Maclay & Osgood 1959 ; Grosjean & Deschamps 1972:146). Ainsi, sur le corpus d'interviews radiophoniques, Grosjean & Deschamps (1972) observent 55,5% d'occurrences de *euh* « à l'intérieur de la phrase », dont 38,5% **en dehors de « points de ruptures grammaticales »**. Parmi ces marques, seulement 5.34% se trouvent entre une préposition et un syntagme nominal (SN), contre 25% à l'intérieur du SN ou du syntagme verbal (SV). De même, les MP sont plus fréquemment suivis d'un *euh* (d'appui ou d'hésitation) que les MO (1 MO sur 75 est suivi d'un *euh d'hésitation* contre 1 MP sur 27). Parmi ces MP, il semblerait que les noms et les verbes soient en moyenne un peu plus souvent suivis d'un *euh* que la totalité des MP (Candea 2000a:237).

2.3.4 Pause pleine : contextes prosodiques – caractéristiques acoustiques

Au niveau de leur **distribution intonative**, les *euh* sont le plus souvent placés en « début de constituant », et surtout après un ligateur. La position intraconstituant semble plus rare (bien que représentée) tandis que la position en fin de constituant est extrêmement rare. Chez Dister (2007:219), le *euh* est le plus souvent positionné en fin de tour de parole (et en chevauchement). Elle lui attribue également dans ce cas une fonction **continuative**. Candea (2000a:261) leur associe la fonction énonciative de « marquer l'**absence de rupture**, le maintien dans le même plan, la suite attendue ». Ainsi, selon Candea (2000a:224), les *euh d'hésitation* n'induisent pas une rupture de la suite de la construction syntaxique par rapport à ce qui est produit avant le *euh* (76% des cas). Au contraire, le *euh* a une capacité à « lier les constituants » (Candea 2000a:224). C'est ce qu'observent également au niveau prosodique les études portant sur le portugais européen, dans les corpus parlementaires exploités par Moniz, Batista, Trancoso & Mata (2012) : les pauses pleines ont majoritairement tendance à s'actualiser **à l'intérieur des « phrases prosodiques »** [73].

Le *euh* d'hésitation, à l'instar de la pause silencieuse, s'analyse également au regard de sa durée et de ses courbes intonatives. En moyenne, les pauses pleines dites d'hésitation sont portées par une durée significativement plus longue de la voyelle que des voyelles issues de syllabes comparables (Guaïtella 1991 ; Candea 2000a:228 ; Shriberg 2001:165). Au niveau de leurs **caractéristiques harmoniques**, les pauses pleines seraient marquées par une « déclinaison physiologique lente et [un contour mélodique] sans décrochage

73. Tandis que les pauses silencieuses suivent deux patrons distributionnels différents : elles apparaissent à l'intérieur des frontières syntaxiques dans le cas de la parole spontanée et en frontière ou sur des positions syntaxiques supérieures (unités de rection et séquences) dans les corpus de lecture Moniz, Batista, Trancoso & Mata (2012) et Moniz, Mata & Viana (2007).

de F0 », dans le registre bas du locuteur (Shriberg, Stolcke & Baron 2001:166). Guaïtella (1991b :121-122) relève **trois contours possibles** : (1) une déclinaison très douce, c'est la majorité des cas observés, (2) une déclinaison légèrement modulée par quelques petites montées de F0, ou enfin (3) une déclinaison se terminant par une amorce, c'est-à-dire, une montée relativement conséquente de la F0. Par contre, les contours plats seraient eux exceptionnels et les contours montants totalement inexistants. Nous proposons au chapitre 5 de tester la détection/catégorisation automatique de patrons des pauses pleines (courbes intonatives, distributions syntactico-prosodiques, marques cooccurrentes) et d'évaluer leur position sur le continuum fluence-disfluence.

2.4 Répétitions lexicales à l'identique

La répétition peut être prise dans un sens large (« dire la même chose avec des mots différents ») ou étroit (« répéter le même matériel [...] au sein d'une proposition, ou, à la rigueur, de deux propositions en succession immédiate») (Richard 2004:53). Dans cette acception étroite, la répétition est donc un bloc contenant au minimum deux éléments : le premier élément est un ***répétable***, suivi d'un second, identique lexicalement, un ***répété***. Toute unité de la langue est virtuellement *répétable*. C'est l'actualisation d'un répété qui permet à posteriori de détecter une « répétition ». Le répétable peut être répété plusieurs fois. Le terme de répétition dénote toujours l'intégralité du répétable et de tous ses répétés. Malgré cette apparente simplicité, **les répétitions** font l'objet d'une multiplicité de définitions dans la littérature francophone et anglophone.

Chez Candea (2000a), la répétition est la troisième marque la plus fréquente après les pauses silencieuses et les pauses pleines. Chez Dister (2007:131), la fréquence de chaînes de répétition(s) va de 0,10 à 10,18 répétitions tous les 100 mots. Mais, de quels types de répétitions parle-t-on ?

2.4.1 Typologies des répétitions

Langue /vs/ parole

En français, certaines répétitions dites « **faits de langue** » relèvent d'une exigence syntaxique due à l'homonymie (exemple 17a) ou contribuent à un effet oratoire d'intensification (exemple 17b-c), de gradation ou de modification du sens. Dans l'exemple 17d le premier *gentil* apporte une information tandis que le second est une confirmation de l'assertion. Enfin, on observe aussi des effets de mises en évidence par dislocation (exemple 17e). Ces répétitions ont pour particularité de pouvoir être présentes à l'écrit comme à l'oral. Elles font **progresser le discours** sur l'axe syntagmatique : « le sens se déduisant précisément de l'accumulation des sens du répétable et du (ou des) répété(s) » (Dister 2007:118 ; Bazzanella 2011:252).

Les répétitions « **faits de parole** » interrompent au contraire la linéarité du discours. Elles opèrent un **entassement** (ou « piétinement ») sur l'axe syntagmatique du discours (Blanche-Benveniste 1984). En effet, ces répétitions n'ont pas de fonction sémantique en ce qu'elles n'ajoutent pas de sens linguistique à un énoncé. Enfin, elles ne s'inscrivent pas dans le système grammatical ou syntaxique de la langue (exemple 17f-g).

Exemple (17) – exemples virtuels

a. **nous nous** retrouvons ici aujourd'hui

b. je suis **très très** heureuse de vous connaître

c. elle était **pas du tout pas du tout** mais alors pas du tout d'accord

d. mes amis sont **gentils // gentils** ouai on pourrait dire ça [74]

e. je suis certain que **vous / vous** monsieur / ça ne vous serait jamais arrivé

f. enfin ce que **je je** veux dire

g. non mais il a **pas pas** vraiment vu tout ça

Dans les faits, la distinction entre les **retouches intéressantes** et les **erreurs manifestes** n'est pas univoque [75] ; il en va de même pour les répétitions de langue /vs/ de parole. Dister (2007:122) soulève les cas de répétitions non ambigus (exemple 18a-c) dès qu'une analyse en dépendance est établie. D'autres cas cependant ne peuvent être désambiguïsés en se restreignant à la transcription (exemple 18d). Dans ces cas, **seule la prosodie semblerait capable de trancher**, d'indiquer par exemple la présence d'une hésitation /vs/ intensification (Henry & Pallaud 2002).

Exemple (18) – issu de Dister (2007:122-124)

a. et nous avons souvent des ennuis quand on se trompe -| et l'informatique **nous nous** conduit parfois à **nous nous** tromper [ilcDA1r]

b. plus // par exemple chez **nous // nous** étions à beaucoup // on parlait (silence) wallon [accPH1]

c. elle le fait d'une manière scientifique et elle se lance dans des explications à n'en plus finir / sur pourquoi dites-vous ceci **nous nous** disons cela il faudrait dire ceci voilà c'est évidemment une relation je dirais presque (dans un rire) professionnelle [iljPF1r]

d. on n'a pas suffisamment de recul et comme **nous nous** n'avons pas / vécu / étant trop jeunes nous n'avons pas vécu la transition [accCP1r]

Candea (2000a:28-30), Henry, Campione & Véronis (2004) et Shriberg (1994), excluent en amont les répétitions « faits de langue » de leurs analyses. Au contraire, dans le cadre de sa recherche portant sur la détection automatique des *dis*fluences dans les transcriptions, Dister (2007) inclut ces répétitions « faits de langue » afin de justement tenter de les distinguer automatiquement des répétitions « faits de parole ». Si l'objectif de recherche de Dister justifie à lui seul la démarche, il est aussi accompagné de **justifications épistémologiques et pragmatiques** telles que la difficulté de désambiguïsation, même parfois à l'écoute du contenu prosodique, et la volonté d'intégrer l'analyse de ces différentes répétitions dans une description unifiée (Dister 2007:130).

74. Ici encore, il est ambigu de déterminer, sans indices prosodiques si la répétition est une stagnation sur l'axe syntagmatique ou non, si la répétition relève plutôt d'une fluence ou d'une disfluence.

75. Comme nous le faisions remarquer au chapitre 1 section 2.2.

Monologique /vs/ dialogique

A l'instar des *self-* /vs/ *other-* repairs[76], on peut différencier les répétitions **monologiques**, qui sont le fait d'un unique locuteur, et les répétitions **dialogiques** lorsque la répétition concerne la reprise d'éléments énoncés par un locuteur et repris par un autre. Candea (2000a:247-251) enrichit la terminologie en parlant de *self-/auto-/same-speaker repetitions* /vs/ *allo-/second-speaker/etero- repetitions*. Par ailleurs, cette reprise peut être **lexicale, syntaxique, prosodique comme gestuelle**.

Ginzburg, Fernández & Schlangen (2007) proposent d'unifier les *self-* et *other-repairs* afin d'analyser les processus de *self-* et *other- monitoring* dans un cadre méthodologique unique. L'exemple 19 illustre un cas de répétition dialogique.

Exemple (19) – traduit de Ginzburg, Fernández & Schlangen (2007:59)

– Peter Miller ne vient pas
– Peter Miller ?
– Non, pardon Paul Miller

La répétition dialogique, appelée également « répétition en interaction » est aussi un signe d'adaptation, de coopération, d'accommodation ou d'alignement[77] entre les interlocuteurs. Richard (2015) parle de « **syntaxe collaborative** ». Cependant, attention à ne pas confondre la répétition dialogique avec la répétition due à la gestion interactionnelle. En effet, de nombreux cas de répétitions, ou de (dis)fluences de manière générale, apparaissent en frontière de tour de parole : un locuteur interrompu reprend la fin de son énoncé et le poursuit. Il s'agit bien là de répétitions monologiques mais **transtour** de parole.

Lexicale /vs/ prosodique

Tandis que la répétition *lexicale* consiste en la réitération d'un ou plusieurs items lexicaux – pouvant aller jusqu'à la répétition de syntagmes entiers – la répétition *prosodique* attrait à la **reprise d'un contour** ou d'un rythme. La répétition prosodique est courante lors de l'énonciation de listes (e.g. « je dois acheter du pain, du fromage, du lait, des tartines [...] »). Les *parallélismes prosodiques* par répétition de contours intonatifs ou par isochronie rythmique sont également reconnus comme des phénomènes de répétitions (Bertrand, Ferré & Guardiola 2013). Par ailleurs, une répétition lexicale peut être, mais n'est pas nécessairement, accompagnée d'une répétition prosodique. Nous y reviendrons.

Dans le cadre de cette recherche, nous nous intéressons essentiellement aux **répétitions lexicales monologiques intratour.** A l'instar de Candea (2000a:28), nous entendons la répétition de phénomènes « **en contiguïté** dans la chaîne sonore ». Cette définition

76. i.e. se corriger soi-même /vs/ corriger les autres.
77. Le choix terminologique dépend en partie du domaine linguistique considéré (e.g. psycholinguistique, phonétique, sociolinguistique) (Bertrand, Ferré & Guardiola 2013:143).

exclut donc les répétitions à l'identique appelées « partielles », c'est-à-dire avec modification ou insertion d'un nouveau contenu (e.g. « une maison, une belle maison », « je l'ai *euh* je l'avais vu ai ») et les phénomènes de reprises de mots [78].

2.4.2 Structure interne de la répétition lexicale

L'analyse des répétitions lexicales monologiques s'appuie traditionnellement sur la description de 4 dimensions relatives à la **structure interne** de la chaîne : le nombre de répétés, le type d'item répété, la longueur du répétable, le type de phase d'édition [79].

Nombre de répétés : répétition simple /vs/ multiple

La majorité des auteurs distinguent les répétitions en fonction du **nombre de répétés**, opposant ainsi les répétitions simples aux répétitions multiples (Morel & Danon-Boileau 1998:85). Dans les faits cependant, nombre d'entre eux ne traitent pas des répétitions multiples de manière distincte. Plusieurs raisons sont évoquées. Blanche-Benveniste (2010:19), citée par Dister (2007:134), justifie le fait de ne pas les distinguer sur base d'**arguments perceptifs**. Le nombre de répétés serait en effet indiscernable par les auditeurs, et pourrait aller jusqu'à sept sans que cela ne se remarque [80]. D'autres encore écartent simplement les répétitions multiples en raison du fait qu'elles sont **singulières au niveau du traitement prosodique**, ou peu fréquentes dans les corpus à l'étude. C'est le cas lors de la comparaison des durées et des intonations du *reparandum* par rapport au *reparans* chez Fox Tree (1995) et Shriberg (1995). Par ailleurs, les répétitions simples semblent avoir des caractéristiques acoustiques fort différentes de celles combinées à une ou plusieurs marque(s) [81].

78. Bien que les répétitions partielles ne fassent pas l'objet d'une analyse approfondie, nous les incluons dans nos annotation initiales sous la terminologie « Répétitions avec Modification » (RM). Candea (2000a:28) les appelle « autocorrections immédiates » (voir chapitre 3 section 2.2). En exemple de répétitions avec « reprise de mots » Candea (2000a:28) donne : « [...]ton chien n'obéit jamais il est vraiment pénible ton chien tu sais ». Ces répétitions, annotées dans nos corpus mais non reprises dans notre analyse sont appelées dans le cadre de notre protocole des « Répétitions Encadrantes ». Leur intérêt potentiel s'analyse au regard de leur fréquence importante en langue des signes francophone de Belgique (Notarrigo, Meurant, Van Herreweghe & Vermeerbergen 2016).

79. En section 3 de ce chapitre, nous envisageons également le(s) type(s) d'élément(s) précédant et succédant les structures.

80. Les expériences de Duez (1995) semblent pourtant indiquer que les auditeurs repèrent les répétitions et que, comme on peut s'y attendre, les chances de perception d'une répétition par les auditeurs sont positivement corrélées au nombre de répétés.

81. Voir section 3 de ce chapitre.

Au niveau distributionnel, les analyses sur corpus incitent à distinguer les répétitions **simples** /vs/ **doubles** /vs/ **multiples**. En effet, Dister (2007:131) observe 87% de répétitions simples (1 répété), 10% de répétitions doubles, et seulement 3% de répétitions de plus de deux répétés. Chez Candea (2000a:28), le nombre maximum de répétés s'élève à 5, contre 10 chez Dister (2007:131). Cette différence est en partie imputable au fait que cette dernière inclut les répétitions « faits de langues » [82].

Chez Candea (2000a:329-331) (contrairement à Morel & Danon-Boileau 1998), l'observation révèle que les répétitions multiples contiennent toutes **au moins une pause silencieuse**. En outre, l'auteure observe que lorsque les locuteurs se livrent à une recherche de mot infructueuse, ils ont tendance à se rabattre – en raison de la contrainte temporelle – vers la production d'un MP **générique** et/ou **fortement prédictible en contexte** (i.e. à haute fréquence ou déjà évoqué antérieurement dans le discours).

Catégorie de répétition : mots pleins /vs/ mots outils

La littérature francophone comme anglophone s'accorde sur la prédominance de répétitions impliquant **des MO** (monosyllabiques), en particulier des prépositions, déterminants, pronoms personnels clitiques sujet ou objets, conjonctions et ligateurs [83]. Les répétitions de MO sont quatre à cinq fois fois plus nombreuses que les répétitions de MP (Candea 2000a:316 ; Morel & Danon-Boileau 1998:84 ; Henry, Campione & Véronis 2004). Cette proportion augmente en fonction des méthodologies de décompte et des types de répétitions pris en compte, 70% chez Grosjean & Deschamps (1975:178-179) et allant jusqu'à 90% chez Blanche-Benveniste (1987:130-134). En incluant les répétitions « faits de langue », Dister (2007:159) n'observe ainsi que 23% d'occurrence appartenant à des mots lexicaux pour les répétables d'un seul mot, ceux-ci étant majoritairement des adverbes (19%).

Cette observation incite à affirmer que les répétitions de MO ont un **caractère systématique** tandis que les répétitions de MP auraient un **caractère relativement accidentel**. Selon Morel & Danon-Boileau (1998:86), le répétable d'une répétition portant sur des MO est **« automatique »** et **conditionné** par la structure syntaxique. D'ailleurs dans le cas des MO « simples » (sans cooccurrence), le plus souvent aucune marque du travail de formulation n'est relevée dans la suite du tour de parole, tandis que lors de l'occurrence en combinaison double, 58% seuls des occurrences de MO répétés sont suivies par d'autres (dis)fluences.

82. Le maximum de répétés atteint 10 dans les corpus de Dister (2007:131). Pour rappel, l'auteure prend en compte les répétitions faits de langues, or les répétitions $n \geq 7$ concernent uniquement les items adverbiaux *oui* et *non*. Ces exemples restent intéressants à comparer à d'autres répétitions $n \geq 5$ en ce qu'ils montrent que le nombre de répétés ne peut être pris en compte à lui seul comme preuve évidente de la disfluence ou fluence d'une chaîne.

83. Les proportions des catégories de mots grammaticaux varient d'un auteur à l'autre. Il semble cependant que de manière générale les marques ambiguës (pronoms /vs/ déterminants *le, la, les* sont plus propices à l'actualisation de répétitions – voir Dister (2007:159-161) pour une revue de la littérature à ce sujet et le chapitre 6 section 1 pour une comparaison avec nos corpus.

Les observations de Candea (2000a:325-327) confirment ces hypothèses. L'auteure postule ainsi que la répétition simple de MO **fonctionne de manière distincte** de celle en combinaison avec d'autres marques (e.g. allongement du répétable, pause silencieuse ou *euh*). La répétition simple de MO serait, elle, davantage due à un **simple retard articulatoire** de la production vocale (par rapport à la programmation d'unité) et s'approcherait d'un **cas léger de bégaiement** ou de gestion (à très court terme) de la production et du cheminement du travail d'élaboration syntaxique et non pas à **une recherche de formulation** chez le locuteur comme d'autres (dis)fluences (Morel & Danon-Boileau 1998:85-87).

Enfin, soulignons que cette proportion supérieure de mots grammaticaux devrait être **calculée de manière relative** à la proportion de mots grammaticaux présente dans chacun des textes. En effet, il n'est pas étonnant que cette catégorie soit sur-représentée comme lieu propice d'actualisation des (dis)fluences (répétitions et allongements) étant donné qu'elle l'est aussi de manière générale en langue (Morel & Danon-Boileau 1998:84 ; Dister 2007:161 ; Grosjean & Deschamps 1972:154).

Longueur du répétable

La longueur d'un répétable peut se calculer en (i) nombre de **mots graphiques**, (ii) nombre de **syllabes articulées** et (iii) en fonction de sa **durée** en absolue ou en durée relative du répétable (*ratio* durée du *reparandum* / durée du *reparan*)[84]. Les cas les plus fréquents sont ceux où le répétable est composé d'un seul mot – 80% des cas chez Dister (2007:136) – et/ou monosyllabique Candea (2000a) et Shriberg (1995). Par ailleurs, « le nombre d'occurrences des séquences répétées décroît selon que la longueur du répétable augmente ».

D'autre part, si dans le cas d'une mesure en **nombre de mots graphiques**, il n'y a pas à priori de différence entre la longueur du répétable et du répété, tel n'est pas toujours le cas lors de la mesure en nombre de syllabes articulées. En effet, il n'est pas rare d'observer des différences dues à l'**élision de voyelles** ou à l'inverse dues à l'**ajout d'un schwa** épenthétique (e.g. *cela* prononcé [sla] /vs/ [s2.la], *petite* prononcé [ptit], [p2.tit], [pti.t@] ou encore [p2.ti.t@]). Nous revenons sur les caractéristiques de longueur en fonction des caractéristiques prosodiques des répétitions ci-après.

Phase d'édition : répétition associée /vs/ directe

Le schéma formel des répétitions, comme toute marque composée d'un *reparandum* et d'un *reparans*, peut inclure une **phase d'édition** séparant le *reparandum* du *reparans*. On oppose alors les répétitions directes – dites aussi contiguës selon les auteurs — aux répétitions différées ou associées (Dister 2007:126 ; Henry & Pallaud 2002, 2004).

Trois positions de (dis)fluences sont possibles par rapport à une marque composée, avant le *reparandum*, après le *reparans* ou entre les deux segments. La phase d'édition peut comprendre des marques de différents types. Lorsque Candea (2000a:28-30)

84. Voir section 2.4.3 de ce chapitre.

inclut dans ses analyses les chaînes de répétitions porteuses de termes d'édition, elle remarque que ces derniers peuvent être un son, une syllabe, un mot, une amorce de mot, de plusieurs syllabes ou de plusieurs mots, tant que celui-ci ne porte **aucune valeur sémantique**. (e.g. « il est p/ p/ pas »). L'auteure observe qu'une pause silencieuse en phase d'édition apparaît en moyenne une fois sur deux (en particulier pour les répétitions de MO) (Candea 2000a:344).

Dister (2007:127), dans une vision moins étroite de la répétition, identifie plus de 21 items en phase d'édition. Parmi les plus courants, on retrouve des **interjections** (e.g. *ah, beh, boh, euh, ah, pff*, etc.) et des **marqueurs discursifs** (e.g. *disons, enfin, bon, ouais*, etc.) ou même des marques explicites d'édition (e.g. *je veux dire, comment dirais-je, par exemple*, etc.).

La position avant le répétable (MO ou MP) est plus rarement occupée par une pause silencieuse (14% des MO et un seul exemple pour les MP chez Candea (2000a), contre 49% des cas chez Grosjean & Deschamps (1975)). Enfin, Candea (2000a:333) décrit la dernière combinaison comme « pratiquement impossible ». La phase d'édition des répétitions à l'identique occupe une place privilégiée dans la littérature[85]. Par exemple, toutes les combinaisons doubles par lesquelles les répétitions sont concernées, le sont en vertu de la **présence d'un marqueur dans la phase d'édition** avec 38% d'allongements, 35% de pauses pleines, 27% de pauses silencieuses.

Les catégorisations, envisagées jusque-là de manière univariée interagissent. Une analyse multifactorielle révèle qu'elles se combinent de manière régulière, et que les patrons de combinaisons montrent des régularités acoustico-prosodiques. Nous verrons par exemple que la présence d'éventuelles **pauses avant et/ou après le répété** influence **la durée** des segments de la répétition et la **forme des contours** syllabiques (Candea 2000a ; Hieke 1985 ; Plauche & Shriberg 1999 ; Shriberg 1995).

2.4.3 Caractéristiques acoustiques et prosodiques des répétitions

La littérature catégorise les chaînes de répétition en fonction de la présence de pause(s) au sein et/ou succédant une répétition (Candea 2000a ; Hieke 1985 ; Shriberg 1995). On appelle **répétitions *prospectives***, celles suivies d'une pause – et contenant optionnellement une pause en phase d'édition. On appelle **répétitions *rétrospectives*** celles n'étant pas suivies d'une pause mais en contenant une entre le répétable et le(s) répété(s) (Figure 2.6).

85. Hormis l'étude de la structure interne des répétitions, Cole et al. (2005:55) notent parallèlement que les répétitions sont plus sujettes à contenir une phase d'édition que les substitutions (appelées ici *repair error corrections*) ainsi qu'à une actualisation sans troncation.

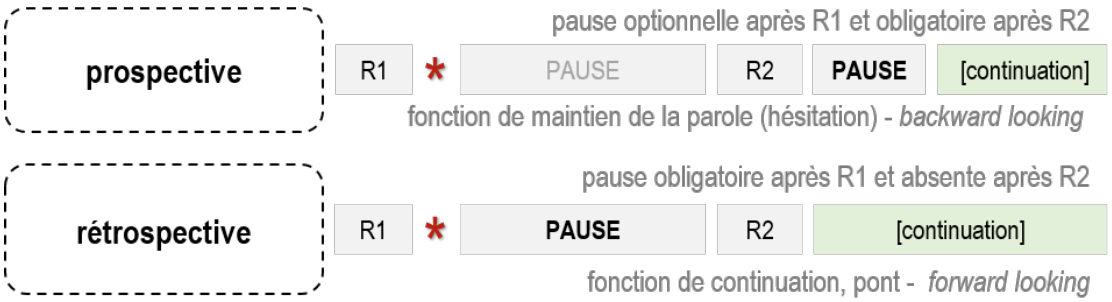


FIGURE 2.6 – Type de structure des répétitions et fonctions associées de HIEKE (1985) à SHRIBERG (1995).

Selon HIEKE (1985), la répétition prospective aurait alors une **fonction de *filler*** – la répétition est une simple hésitation servant au **maintien de la parole**. La répétition rétrospective servirait de pont vers le **recouvrement de la fluence** ; la connexion directe du *reparans* au reste de la chaîne discursive permettant ainsi une continuation fluide avec le matériel antérieur. Suite aux propositions de **relations formes-fonctions** de HIEKE (1985), SHRIBERG (1995) enquête sur la répartition de ces patrons et de leurs autres caractéristiques acoustiques (i.e. observation des mouvements de F0, hauteurs relatives et de la durée relative). Si les corpus montrent un nette tendance à l'actualisation de répétitions rétrospectives, leurs formes suggèrent que plusieurs hypothèses fonctionnelles sont envisageables.

Il n'y a que dans le cas des répétitions **rétrospectives** que la durée du *reparandum* est systématiquement **plus longue** que celle du *reparans* (*bridging hypothesis*). Dans le cas des répétitions prospectives, la variation relative de la durée des segments semble pouvoir aller dans les deux sens avec une tendance à l'allongement relatif du *reparans*, cohérente avec l'hypothèse de maintien, de suspens et d'hésitation.

CANDEA (2000a:190-191) distingue également **deux grandes catégories de répétitions**. La première englobe des chaînes où le répétable et le répété se suivent immédiatement ou sont séparés par une très brève pause silencieuse ([R1 + (pause) + R2]). La deuxième, comprend les chaînes où, non seulement le répétable **est allongé** de manière significative, mais est en plus suivi d'une hésitation *ou* d'une autre (dis)fluence (pauses, répétitions, autocorrections, etc.) ([R1-allongé + (*euh*/dis/pause) + R2]).

Les catégories de SHRIBERG (1995) et de CANDEA (2000a) ne se recouvrent pas entièrement. CANDEA (2000a) ne parle pas de la pause subséquente observée dans les répétitions prospectives. Cependant, elle met en valeur un trait important considéré par de nombreux auteurs : l'*allongement du répétable* – qui serait caractéristique de la répétition simple (MOREL & DANON-BOILEAU 1998:86). L'allongement du répétable est une tendance forte plus qu'une « règle » : 53% des MO répétables sont allongés dans les corpus de CANDEA (2000a:335). La tendance à l'allongement est plus marquée dans le cas des

configurations rétrospectives. Avec ou sans allongement attesté, dans une majorité des cas, le répétable est plus long que le répété et ce tout type de répétition à l'identique confondu (82%)[86].

A l'instar de Levelt & Cutler (1983), de nombreuses études se focalisent sur la relation entre *reparandum* et *reparans*[87]. Les profils intonatifs des répétitions sont révélés par au moins quatre points de mesure de **fréquence fondamentale** : le début et la fin du *reparandum* et du *reparans*.

Dans le cas des répétitions rétrospectives, les observations de Shriberg (1999) révèlent une tendance à (1) une **chute de la F0** au sein du *reparandum* (2) une amplitude et une pente de la F0 similaire entre le *reparandum* et le *reparans* (3) une équivalence de la hauteur des débuts de *reparandum* et de *reparans*. Ces observations corroborent exactement la **structure en parallélisme** observée vingt ans plus tard par Cole et al. (2005). En assimilant la courbe de F0 continuellement descendante des répétitions prospectives au profil intonatif des pauses pleines, les auteurs maintiennent que ces configurations ont bien une fonction de *filler*.

Enfin, certains font l'hypothèse que certaines répétitions seraient en réalité des « **substitutions cachées** » (*D-covert repairs*)[88] (Clark & Wasow 1998). Selon Shriberg (1995), ces types de répétition se distinguent par une F0 élevée du segment *reparans* – similaire à une valeur de F0 d'initialisation de segment. Elles seraient par ailleurs observées pour les répétitions prospectives et les non rétrospectives examinées ci-avant.

Cole et al. (2005) proposent une classification étendue des répétitions[89]. Quelques résultats globaux se dessinent. Proportionnellement, les répétitions **contiguës** sont plus fréquentes que celles associées à une pause pleine. Suivent les répétitions pour lesquelles le dernier mot dont le contour intonatif a été perçu comme marquant une « phrase intonative bien formée ». Moins courantes sont les répétitions avec **emphase prosodique** du *reparans* (durée accrue, F0 plus élevée) et les répétitions avec modification prosodique du *reparandum* par rapport au *reparans* (i.e. modification de la présence ou position d'un accent interprétées également comme une conséquence d'une **correction pragmatique** – plus que lexicale ou syntaxique[90]). Enfin, dans leur travail sur le *parallélisme prosodique*, Cole et al. (2005) émettent l'hypothèse selon laquelle l'éventuelle constance contrastive de certaines structures prosodiques du *reparandum* et du *reparans* des répétitions s'impose à l'auditeur comme un signal perceptif fort d'une (dis)fluence lui permettant d'éviter les erreurs de compréhension.

86. La tendance du répétable à être plus allongé que le répété est encore plus forte dans le cas de répétitions combinées à d'autres (dis)fluences (54% des cas pour les répétitions simples contre 86% pour les répétitions combinées) (Candea 2000a:336).

87. Voir section 1.1.2 de ce chapitre.

88. Voir chapitre 1 section 2.3.4.

89. Leur catégorisation permet de représenter 21,96% des (dis)fluences du corpus, fondée uniquement sur l'observation distributionnelle et prosodique, sans catégorisation à priori du type de (dis)fluence.

90. Pour chacun de ces types de répétitions, ils différencient les cas où la répétition s'actualise sur un fragment de mot ou sur un item lexical entier. Dans notre travail, comme mentionné au chapitre 3, la catégorie des répétitions n'inclut que celles des items lexicaux entiers – i.e. exclut les amorces.

2.4.4 Conclusion : rupture ou continuité – fluence ou disfluence ?

> « Le retour à l'identique permet également une meilleure **appréhension de l'élaboration du discours, du discours en train de se faire**. La répétition est alors un signal fort dans la progression et la hiérarchisation du discours. » (Richard 2015:6)

La catégorie des répétitions, particulièrement fréquente parmi les marques du travail de formulation, se révèle intéressante à plusieurs égards. Leur régularité est une porte d'entrée propice à l'analyse de phénomènes plus **complexes structurellement** (et donc prosodiquement) en ce qu'ils impliquent d'emblée davantage d'**analyses sémantique et syntaxique** (e.g. répétitions avec modifications, substitutions, autocorrections)

La tendance des répétables à s'actualiser sur « les petits mots du discours » témoigne de ce que la répétition relève, en général, d'une **réinitialisation syntagmatique** : un substantif est rarement répété seul (Blanche-Benveniste 2003). Ainsi, ce n'est pas sur le lexème seul que se manifeste le travail de formulation mais sur toute l'unité dont il est la tête (Dister 2007:163). Cette constatation favorise la reconnaissance de la répétition comme marque (ou moyen) de **continuité syntaxique** dans l'énoncé, puisque les constructions ouvertes par un mot outil sont au final achevées. D'ailleurs, dans un sous-corpus de Dister (2007), moins de 6% des répétitions (66 occcurrences) occasionnent une réelle **rupture syntaxique** de l'énoncé. L'auteure remarque par là-même qu'en réalité les ruptures de construction ne sont précédées d'aucune (dis)fluence lexicalisée précise.

Dans ce sens, Richard (2015:6) note elle aussi, que la répétition est révélatrice de l'**élaboration du discours** ; et que l'apparente stagnation sur l'axe syntagmatique qu'elle peut impliquer n'est pas toujours synonyme de rupture. Au contraire, en se répétant, le locuteur **assure le maintien** de la parole et vocalise son **intention de continuation et/ou de progression** discursive. Du côté de l'auditeur, nous avons vu que la répétition est rarement retranscrite par les naïfs (Voss 1979:132) et moins détectée comme signe de rupture que les faux départs ou autocorrections (Duez 1995).

Des répétitions « faits de langue » aux répétitions « faits de parole », plus ou moins disruptives dans l'énoncé, nous espérons avoir donné un aperçu des formes et de fonctions qu'elles peuvent accueillir. Malgré les apparentes régularités relevées ci-avant, les répétitions et les autres marqueurs tendent à **apparaître simultanément**, à s'actualiser ensemble. Nous analyserons leurs dynamiques dans nos corpus au chapitre 6. La question de leur relative fonction bénéfique – continuation de conceptualisation, de formulation ou coopération interactionnelle – se doit d'être soulevée d'une tout autre manière. La section suivante s'attache en particulier aux descriptions possibles de ces « **marques mixtes** ».

3 Niveaux et types de ruptures : cooccurrences et contextualisation des marques

Dans ce chapitre, on a décrit essentiellement les marqueurs s'actualisant de manière isolée, au sein d'un contexte prosodico-syntaxique, avec leurs régularités et leurs spécificités d'actualisations acoustiques. Nous avons toutefois volontairement omis la question complexe de leurs cooccurrences. Pourtant, un paramètre essentiel à la définition de la (dis)fluence en contexte relève de la combinaison des marqueurs entre eux. Ces « **accidents en cascades** » concernent 40 à 56% des phénomènes de (dis)fluences annotés dans les corpus [91].

En effet, chez Candea (2000b:134), la proportion de marques isolées (44%) est même inférieure à celles faisant partie d'une chaîne *double* ou *mixte* [92]. Ces observations incitent les auteurs à analyser de manière consistante, en amont, les occurrences simples /vs/ combinées des marques du travail de formulation. Dans ce travail de thèse, nous nous questionnons sur les possibles **régularités des différentes combinaisons**, leur impact sur une éventuelle rupture discursive (produite et/ou perçue) et sur l'acceptabilité ou l'appropriété en terme de fluence phrastique, situationnelle ou sociale. Cependant, une question subsidiaire semble pertinente : toutes les cooccurrences doivent-elles nécessairement être traitées comme faisant partie d'une **même chaîne** ?

> « Il est légitime de se demander si la pause silencieuse produite juste avant une marque de TdF [Travail de Formulation] pourrait en quelque sorte anticiper cette marque, tout comme la pause silencieuse peut prolonger l'indication de TdF en cours si elle est placée après une telle marque. Quel est le statut de l'élément subséquent à une pause silencieuse. » (Candea 2000a:155-156)

En effet, la complexité de ces combinaisons varie de par leur longueur et le type de marque actualisé. Différentes méthodologies de regroupement existent. Si certains auteurs écartent totalement les marques mixtes ou de plus de deux types (Shriberg 1994 ; Zellner 1994), d'autres s'attachent en particulier à ce qu'elles révèlent. Ainsi, Candea (2000a:155-156) soulève une interrogation pertinente dans la citation ci-avant. Selon Candea (2000a:159), la combinaison [pause silencieuse + marque] **ne forme pas une chaîne** en tant que telle, la marque qui succède la pause ouvrirait un nouveau groupe dont la programmation n'est pas directement disponible et qui nécessiterait de la part du locuteur de combler ce retard par une (dis)fluence à laquelle on associe principalement une **fonction colocutive**.

En résumé, Candea (2000a:155-156) propose que dans cette configuration la pause silencieuse endosse en réalité **deux fonctions amalgamées** : structurante et anticipatrice d'une (dis)fluence à venir. Ainsi, les allongements précédés d'une pause silencieuse, **sont**

91. Plus précisément, 44% chez Candea (2000a:134) et 40% chez Zellner (1994) pour la partie concernant les sujets non bègues de ces corpus. Cette dernière ne donne cependant pas plus de précisions sur les configurations observées.

92. A l'exception des pauses silencieuses, qui apparaissent justement plus souvent entre des chaînes sonores que combinées à une marque du travail de formulation.

en général devancés par ce que CANDEA appelle des *pauses structurantes* (marquant une frontière) et ne font donc pas partie de la même chaîne que l'allongement. Dans cette configuration, l'**allongement initialise** (et signale le retard de) l'unité suivante.

Certains allongements précédés d'une pause silencieuse sont pourtant précédés eux-mêmes d'un *euh* (la pause silencieuse étant ainsi non structurante par définition). Cette configuration semble donner une frontière relativement imperméable, l'allongement marque alors un **début de constituant** (et un début d'énoncé)[93]. L'auteure choisit donc de ne **pas intégrer les pauses subséquentes** aux chaînes de (dis)fluences car elles marqueraient l'ouverture d'un nouveau constituant, nous revenons sur ces questionnements au chapitre 4 lors de l'analyse du profil durationnel des pauses dans leurs contextes prosodiques et syntaxiques.

3.1 Fréquences de cooccurrences

Les cas de marques simples isolées sont donc relativement rares. Selon les recherches de CANDEA (2000a:442-443), l'ordre des fréquences des marques est **inversément proportionnel à leur tendance à marquer la rupture**. Le *euh*, marque la plus utilisée, est caractérisé le plus souvent par une continuité, sans rupture discursive. Près de la moitié des hésitations sonores (tous types confondus) sont précédées et/ou suivies par des pauses silencieuses (FAGYAL 1995 ; cité par CANDEA 2000a:139). De manière générale, GUAÏTELLA (1991b ; citée par CANDEA 2000:138) rapporte dans son corpus d'interviews que « 22,5% des hésitations vocales sont **précédées et suivies par des pauses silencieuses**, 50% sont suivies par une pause silencieuse et précédées par des paroles, 5% sont précédées par une pause et suivies par des paroles et 22,5% sont encadrées par des séquences sonores (précédées et suivies par des paroles) ».

La marque la plus fréquemment associée à d'autres est **la pause pleine** (FAYGAL 1999 ; CANDEA 2000a:139). Chez CANDEA (2000a:219), près de deux tiers des *euh d'appui (épenthétique)* sont des occurrences simples (pas de rupture dans la parole). Lorsqu'ils se combinent, c'est surtout avec des pauses silencieuses. Chez CANDEA (2000a:219) toujours, sur la totalité des *euh d'appui*, 24,3% sont de type prépausal, une différence importante avec HANSEN (1989 ; citée par CANDEA 2000:219) chez qui on atteignait 99.46%. L'auteure y voit un changement en cours. Le *euh* en combinaison à d'autres marques permettrait également d'**éviter la rupture discursive**, de favoriser le maintien d'une continuité.

Au niveau du nombre de constituants, pour les *euh* d'hésitation, CANDEA (2000a:239) rapporte que sur 469 occurrences, 46% seulement apparaissent de manière **isolée**, contre 36% **double** (dont les trois quarts sont accompagnés en réalité d'une pause silencieuse et n'induisent alors aucune rupture dans la construction syntaxique en cours), 11% triple et 7 de plus que trois associés. Ainsi, « non seulement les combinaisons *euh* + [pause

93. Voir CANDEA (2000a:295-300) pour plus de détails sur les configurations et les types d'enchaînements (e.g. énonciatif, syntaxique, discursif, etc.).

silencieuse] favorisent les durées les plus longues des pauses silencieuses [...], mais en outre il apparaît que la pause silencieuse est la marque avec laquelle le *euh* se combine le plus fréquemment : 73% des combinaisons doubles » (Candea 2000a:241).

Sur leur corpus d'interviews radiophoniques, Grosjean & Deschamps (1973) rapportent que 30% des *euh* sont suivis ou précédés d'une pause silencieuse (de manière non distinguée), que 6% sont encadrés (précédés et suivis) par une pause silencieuse, 24.5% sont précédés ou suivis d'une syllabe allongée et enfin 5.5% sont précédés ou suivis d'un faux départ ou d'une répétition. Ils notent ainsi que « le pourcentage de **pauses remplies précédées ou suivies d'une autre pause** est de 75,92% en anglais et de 59,45% en français [...]. Ces pourcentages élevés sont surtout dus au fait qu'un **grand nombre de ces pauses se produisent soit avant soit après une pause silencieuse**. [...] 47,26% du français, ce qui donne à penser que souvent la pause remplie renforce ou est renforcée par la pause silencieuse » (Grosjean & Deschamps 1975:176).

Après la pause, c'est l'**allongement** qui est le plus accompagné d'autres marques. Il maintient souvent la **continuité** également. Dans les corpus de Candea (2000a:291), l'allongement (portant sur des MO) tend à s'actualiser de manière isolée dans seulement 35% des cas. Au reste, il est associé à une autre marque dans 30% des cas (suivi par un ligateur ou une autre marque). Enfin, on le retrouve même dans des combinaisons triples et quadruples dans 21% et 14% des cas. Le pourcentage de combinaisons simple et double est ainsi inférieur au pourcentage pour les *euh* mais supérieur à celui obtenu pour les répétitions de MO (ou de ligateurs). Approximativement 30% des *euh* sont donc suivis par une pause silencieuse, alors que 19% seulement des allongements sont dans ce cas (Candea 2000a:185). Les MO allongés immédiatement suivis de *euh* sont plus fréquents que les MO (ou ligateurs) allongés immédiatement **précédés** de *euh*. Cette observation tend à conforter l'hypothèse selon laquelle « l'allongement vocalique est, comme le *euh*, essentiellement une marque continuative qui n'opère **pas de rupture** lorsqu'elle s'insère dans une chaîne sonore » (Candea 2000a:291).

La proprotion d'occurrences simples ou doubles est **très forte pour les allongements** (environ deux tiers dans le corpus de Candea (2000a)). Dans leur étude de 1972, Grosjean et Deschamps remarquent que dans 65% des cas, l'allongement se combine avec l'occurrence d'autres (dis)fluences. 46% d'entre eux seront précédés ou suivis d'une pause silencieuse, 10% sont suivis d'une répétition, 7% sont suivis d'un *euh*, 2% sont précédés d'un *euh* ou d'une « syllabe allongée d'hésitation » et sont souvent suivis par un *euh* (Lacheret-Dujour & Beaugendre 1999:50).

Les répétitions sont aussi sujettes à la cooccurrence de marques. Candea (2000a:320) observe ainsi que seules 12% des répétitions (de MO) sont produites de manière isolée (contre 29% en combinaison double, 24% en combinaison triple et 35% en combinaison de plus de trois marques). Cette même tendance s'observe pour les MP (Shriberg 1994, 1999 ; Watanabe, Hirose, Den & Minematsu 2008). La plupart des répétitions se combinent au moins avec l'allongement du répétable – dans 54% des cas chez Candea (2000a:325). L'importance de ce phénomène explique l'intérêt prononcé de la littérature

pour l'étude de la répétition autre que contiguë. La pause après la répétition serait, elle, relativement rare ; elle ne s'observe que dans les cas où la répétition et la **pause silencieuse sont encore suivies** d'une autre (dis)fluence (Candea 2000a:322).

La faible proportion relative de **pauses silencieuses** en cooccurrence avec d'autres marques s'explique notamment par le fait que les pauses silencieuses recouvrent un **caractère obligatoire**, relativement prédictible. Elles sont la trace d'une structuration plus maîtrisée. Au contraire, les (dis)fluences sont une trace de structuration « **temporairement non maîtrisée** ». Elles ont un caractère proportionnellement plus **facultatif** mais aussi **imprévisible**. La réinitialisation et/ou la réorganisation hiérarchique des énoncés favorisent la cooccurrence.

Enfin, il existe aussi des combinaisons dites **hybrides**, dans le cas des « marques triples » ou plus. Dans le cas de combinaison triple, c'est la répétition qui semble être la plus courante, accompagnée de pause(s) silencieuse(s) ou pleine(s). Rares sont les auteurs s'aventurant dans leur analyse fine étant donné que plus le nombre de marques combinées est élevé, plus les cas attestés sont rares (Shriberg 1994 ; Candea 2000a:145,476). Au sein des combinaisons triples, toutes les possibilités semblent également présentes [94].

L'ordre des fréquences des marques est **inversément proportionnel** à leur tendance à marquer la **rupture** [95] (Candea 2000a:442-443). La cooccurrence de marques représente une **alternative à l'allongement** d'une même marque, tout en allouant au locuteur un temps élevé pour programmer une unité subséquente. Dans les corpus de Duez (1991), la cooccurrence des ***euh* d'hésitation** et des **syllabes allongées d'hésitation** concerne jusqu'à 57% des phénomènes annotés au sein d'un syntagme pour les interviews politiques contre 35% dans les interviews amicales. Ces nombres montent de 57% à 64% dans le cas des phénomènes annotés en frontières syntaxiques. La cooccurrence est plus fréquente en frontières syntaxiques qu'au sein de syntagmes, afin notamment de **maintenir la continuité discursive**. Au reste, l'accumulation de marques est moins présente (voire totalement absente) dans les discours politiques, en raison du style particulièrement contrôlé des locuteurs (Duez 1991:74 ; citée par Candea 2000a:137).

Le cooccurrence met en exergue la relation entre le niveau de proéminence d'une marque et le degré de rupture qu'elle peut entraîner. Cette distinction fond-forme peut nous donner des indices sur la **saillance des chaînes** au sein des contextes discursifs. Enfin, comme nous l'avons vu, le degré de rupture discursive est lié aux marques elles-mêmes, mais aussi à leur **forme acoustique en contexte**. L'accumulation de marqueurs ne résulte pas uniquement de phénomènes linguistiques ; elle est aussi influencée par les **contraintes situationnelles** exercées sur le discours. Nous proposons de distinguer l'observation en production et les effets en perception de ces cooccurrences.

94. Si *P* une pause, *D* une (dis)fluence, on observe des configurations de type (D + **D** + P), (D + P + **D**) + (D + **D** + **D**). Si l'on considère la pause comme pouvant ouvrir une chaîne de (dis)fluences on ajoutera les chaînes (P + D + P) et (P + D + **D**).

95. Voir section 2 de ce chapitre.

3.2 Production : impact des cooccurrences sur la forme acoustique des marques

> « La présence ou l'absence d'une pause silencieuse précédant ou suivant immédiatement une hésitation vocale a une influence sur sa durée. »
> (Candea 2000a:95)

La cooccurrence a un **impact sur la forme prosodique** des marques. Les interactions des cooccurrences et de leur durée peuvent être décrites de manière régulière. On remarque que la hiérarchie des fréquences des pauses pleines, allongements d'hésitation et des répétitions à l'identique, est directement corrélée avec la **durée des pauses silencieuses subséquentes**. En résumé, les *euh* introduiraient les pauses plus longues mais garantissent la continuité. Les allongements introduiraient des durées plus brèves mais garantissent moins la **continuité sans rupture**. Les répétitions, apparaîtraient en même temps que des pauses silencieuses subséquentes de durées plus brèves et introduiraient davantage de rupture sur le plan syntagmatique. Les répétitions impliquent toujours une rupture minime, tandis que l'autocorrection introduit toujours des ruptures plus importantes.

Quelques généralités prosodiques ont été observées sur les marques encadrées de séquences sonores /vs/ d'autres marques (précédant, succédant, ou encadrant). Guaïtella (1991b :125), citée par Candea (2000a:138), note que ce sont les marques **encadrées par des pauses silencieuses** qui ont les durées les plus longues. Par ailleurs, les marques encadrées par des séquences sonores présentent une **déclinaison de F0 plus faible** (voire nulle) et une durée plus brève que celles précédées, succédées ou encadrées par des « marques d'hésitation ». Ce sont les marques **précédées par une pause silencieuse** qui ont la déclinaison la plus forte de F0 et une durée assez longue – que celles-ci soient suivies par une séquence sonore ou une autre pause silencieuse.

Par ailleurs, au niveau de leur forme, Candea (2000a:135) suggère qu'il existe un effet global du type de marque (*allongement, pause pleine, répétition ou autocorrection*) sur la durée de **la pause subséquente**. Par exemple, la pause silencieuse est significativement plus longue (supérieure à 500 ms.) dans la configuration [pause pleine + pause silencieuse] qu'en configuration isolée (Candea 2000a:184). De même chez Guïtella (1991b) les hésitations vocales *euh* ont en moyenne une durée plus longue (et une décliaison mélodique plus accentuée) lorsqu'elles sont précédées par une pause silencieuse. Cette observation est en partie confirmée par Swerts (1998) en néérlandais.

Les recherches de Candea (2000a:158) complètent ces observations, en confirmant que rien n'indique que les pauses silencieuses *structurantes* suivies de marques de travail de formulation soient plus longues que celles suivies d'une séquence sonore. Candea (2000a) compare la durée des pauses silencieuses précédées d'un allongement à celles des pauses silencieuses structurantes d'une part et des non structurantes d'autres part. Elle note que, les pauses silencieuses **précédées d'un allongement** sont en moyenne légèrement plus brèves que les pauses non structurantes (à l'opposé des observations des

pauses silencieuses précédées par un *euh*). Enfin, par rapport aux pauses silencieuses structurantes, celles précédées d'un allongement sont significativement plus brèves, alors que cette différence de durée ne serait que marginalement significative pour les *euh*.

En résumé, il existerait une différence entre la **durée des pauses silencieuses subséquentes** à une pause pleine /vs/ subséquentes à un allongement. Les premières sont en moyenne plus longues que les dernières. Selon Candea (2000a:187), l'écart serait soit dû à la différence du rôle énonciatif du *euh* et de l'allongement, soit aux particularités distributionnelles morphosyntaxiques des marques (e.g. présence d'un allongement dû à la fin d'un constituant). Par ailleurs, cette pause silencieuse serait plus courte entre des **MO répétés** qu'entre des **MP répétés**, bien que cette mesure de la durée doive être associée au débit relatif en plus de la présence d'un *euh* ou d'un allongement avant la pause silencieuse, qui semble avoir une influence sur la durée de cette dernière (Candea 2000a:334). En effet, les pauses silencieuses produites en phase d'édition de répétition, lorsque le répétable a été allongé ou suivi d'un *euh*, ont une durée significativement plus longue que toutes les autres pauses silencieuses (Candea 2000a:190). Cette durée plus longue peut être associée au **besoin d'encodage** plus important afin que le locuteur puisse continuer son énoncé (Clark 1994).

Candea (2000a:230) n'a pas pu mettre en évidence de différence de durée moyenne entre les pauses pleines en position **initiale** /vs/ **médiane de constituant**. Cependant, à l'instar de Guaïtella, l'auteur constate bien une différence significative entre la durée des pauses pleines en position médiane de constituant et celles produites entre deux pauses silencieuses (isolées des constituants syntaxiques), les secondes étant plus longues que les premières (36,8 cs. /vs/ 45,2 cs. en moyenne).

3.3 Perception : impact des cooccurrences sur la perception et l'évaluation des marques

La combinaison des (dis)fluences semble non seulement impacter leur forme (e.g. modification des courbes intonatives, allongement/raccourcissement des durées, etc.) en production, mais ces actualisations ont bien entendu des **effets sur la perception et l'évaluation qu'en ont les auditeurs**. De manière générale, les auteurs observent que très peu de (dis)fluences sont perçues par une majorité d'auditeurs-évaluateurs. Seulement 4.5% des marques d'hésitation sont perçues par au moins 90% des auditeurs contre 18.4% des pauses silencieuses. Quelle qu'en soit la nature, les **occurrences simples** passent plus souvent inaperçues par rapport aux combinaisons de marques, surtout dans le cas des pauses isolées dans l'expérience de Candea (2000a:134).

Au niveau de l'**évaluation**, les pauses silencieuses situées *après* une (dis)fluence seraient moins perçues que les autres marques et surtout, ne seraient pas perçues par les auditeurs comme des indices de rupture. En effet, avant tout, la marque **inhibe la perception** et l'évaluation sévère de cette pause silencieuse (Candea 2000a:132,148). Dans leur corpus, Grosjean & Deschamps (1973) notent que 7.66% des pauses silencieuses sont précédées

d'un faux départ ou une répétition. La présence d'une pause silencieuse à proximité d'une marque semble par contre rendre la marque en elle-même **plus proéminente** et favoriserait son évaluation comme une hésitation (Candea 2000a:125,131).

Toujours chez Candea (2000a:284), la moitié des allongements (portant sur des MP) se combinent avec d'autres (dis)fluences. La combinaison [*allongement* + pause pleine] est la plus fréquente[96]. Par contre, l'auteur n'atteste d'aucune combinaison [allongement + pause silencieuse]. L'auteur relie ces observations avec le fait que cette dernière configuration pourrait être perçue par les auditeurs comme **la marque d'une frontière** (Candea 2000a:287). Ce type d'observations incite à penser que le fonctionnement de la pause silencieuse isolée et celui de la pause combinée ou subséquente sont tellement différents qu'ils révèlent en réalité de **phénomènes complètement distincts**.

Candea (2000a) a remarqué sur les corpus étudiés que la durée des pauses silencieuses qui interviennent entre le répétable (lorsque celui-ci ne porte aucune marque de travail de formulation tel que l'allongement) et le répété ne dépasse pratiquement jamais 40 cs. Or, au niveau perceptif cette durée représenterait le **seuil minimum nécessaire** pour obtenir une réaction verbale de l'auditeur. Elle constitue la durée minimum du **décodage sensoriel** suivi du **décodage verbal** (env. 25-30 cs.) à laquelle s'ajoute le temps minimum nécessaire à l'encodage d'une simple marque d'écoute. Autrement dit, aucune réaction de la part de l'auditeur ne semble être possible (ni attendue) en deçà de ce seuil de 40 cs. Cela nous incite à penser que la répétition simple ou combinée avec la pause silencieuse (qui porte essentiellement sur les MO) est une marque **purement egocentrée**, sans visée colocutive ; elle n'est absolument pas destinée à être entendue par l'auditeur. Il se peut qu'elle témoigne de l'intention du locuteur de **pouvoir enchaîner** très vite sur la suite.

La brièveté de la durée des pauses silencieuses combinées uniquement avec la répétition peut compenser la « dérogation » au principe de l'ordre naturel de la perception. En effet, si l'on respecte ce principe, on entend d'abord le répétable étiqueté non pas comme répétable mais comme **unité laissée en suspens** (à la fin d'une séquence qui ne comporte pas de marque de travail de formulation) suivie d'une pause silencieuse qui devrait être logiquement étiquetée comme « pause silencieuse structurante » . Bien que la plupart des répétables soient des MO et que la présence d'une pause silencieuse structurante après un MO soit assez étonnante, l'auditeur-écouteur n'a concrètement **aucun indice certain qui lui permette de prévoir la répétition** après la pause silencieuse de l'unité laissée en suspens. Seulement, compte tenu des durées de ces pauses silencieuses, on s'aperçoit que la suite (à savoir le répété) est produite par le locuteur après un laps de temps qui permet à peine à l'auditeur de décoder ce qu'il vient d'entendre ; le locuteur enchaîne son énoncé après une rupture minime qui **semble passer quasiment inaperçue**. Nous testons la perception des répétitions et de marques associées au chapitre 6 section 2.

96. Suivie des combinaisons [*allongements* + répétitions] et [*allongements* + autocorrections immédiates].

4 Hypothèses : la disfluence comme outil de fluence ?

Avant d'exposer les **données et la méthodologie** à la base de la suite de nos recherches, nous récapitulons les **prémisses** et les **questionnements** révélés au fil des deux chapitres antérieurs, afin de formaliser les **hypothèses de recherche**.

Un enchaînement de **prémisses principales** chapeaute cette thèse étant donné les questionnements de la recherche :

- La **pratique communicative** influence la charge cognitive et affecte la **production discursive**.
- Les (dis)fluences sont des indices à **double facette** pouvant être **bénéfiques ou entraver** la production et la perception de la parole en interaction.
- La fluence ne se limite *pas* au respect d'une **linéarité ininterrompue** du flux discursif.
- La désambiguïsation des marques linguistiques de (dis)fluence se fait par l'**analyse du contexte**.

Pour rappel, les **trois questionnements** de ce travail sont les suivants :

- Q1. « Qu'est-ce que la fluence/disfluence pour les locuteurs natifs ; comment la définir et l'analyser au regard des situations ? » (chapitre 1 section 1).
- Q2. « Quelles sont les relations entre les **marques de disfluences**, les **ruptures prosodiques et syntaxiques**, et la **rupture discursive** ? » (chapitre 2 section 1).
- Q3. « Quelles sont les configurations qui entraînent une rupture suffisant à produire une **évaluation disfluente** en situation ? » (chapitre 2 sections 2 et 3).

Traditionnellement, le linguiste formule des hypothèses sur la **fluence cognitive** que l'auditeur attribue au discours. Ces hypothèses nous donnent une idée de la **fluence perçue** (c'est-à-dire des inférences qu'un auditeur formule sur la fluence cognitive du locuteur). Cette démarche permet au linguiste de réévaluer l'importance relative de la **fluence phrastique** (Voir chapite 1 section 1.2.2).

Pour tester ce cadre théorique de la fluence, notre recherche prend **un tournant théorique, méthodologique et appliqué**.

Notre proposition d'**hypothèse principale** porte sur **la définition même de la fluence** et du **continuum fluence-disfluence** chez les locuteurs francophones natifs en situation de parole.

H1 : L'évaluation de la fluence produite et perçue résulte de la **convergence** entre les **attentes liées à la situation** et le **discours effectivement produit et perçu**. Cet ajustement découle de nos **attentes linguistiques, situationnelles ou interpersonnelles**.

Nous avons à ce propos ébauché un **cadre théorique** au chapitre 2 section 1.4.3. On y distingue un continuum fluence-disfluence en fonction de la **convergence ou divergence avec ces attentes**. Dans cette proposition, une réelle (dis)fluence correspond aux **extrêmes du continuum** : une sur-saillance (rencontre ou déception) marquée des attentes de l'auditeur vis-à-vis des contextes externes et internes, linguistiques, situationnels ou interpersonnels. Autrement dit, une sur-déception ou une sur-confirmation des attentes.

Cette hypothèse théorique est accompagnée d'une **hypothèse méthodologique** testant la validité de l'**approche cognitive fréquentiste** (chapitre 2 section 1.4.1). En s'interrogeant sur une analyse des (dis)fluences qui pourraient être porteuses de **fluence produite**, **perçue** et **interprétée**, nous confrontons les normes statistiques produites aux normes perçues. On formule cette hypothèse non directionnelle :

H2 : L'évaluation d'une marque discursive par les auditeurs est **associée aux fréquences objectives observées dans les corpus**. Les **attentes discursives** sont liées aux **normes fréquentielles** de situations.

Une marque et ses cooccurrences pourraient donc être évaluées différemment en fonction de l'importance donnée aux différents types de contextes internes. De nombreuses **hypothèses spécifiques**, relevées dans la littérature du chapitre 2 section 2, font référence à chaque marque de (dis)fluence en elle-même. Concrètement, des questionnements différents s'appliquent à une pause silencieuse, à une pause pleine, à un allongement ou à une répétition. Chaque chapitre d'analyse sera basé sur des hypothèses précises sur leur forme, leur structure cooccurrente, leur impact sur les unités prosodiques et syntaxiques ainsi que leur distribution situationnelle.

Au **niveau appliqué**, pour évaluer quelle configuration entraîne une évaluation disfluente, on considère ainsi qu'il existe plusieurs **types de divergences par rapport aux attentes** et donc de types d'évaluation de la fluence-disfluence :

H3 : Plusieurs types d'évaluation peuvent être distingués en fonction des types d'attentes considérés : la disfluence **linguistique phrastique**, la disfluence **situationnelle** et la disfluence **sociale/interpersonnelle**.

Cette distinction tripartite amène à des sous-hypothèses relatives à chaque type d'évaluation.

– L'**évaluation disfluente linguistique** phrastique correspond à l'accumulation de marques. Elle est observée dans chaque section de chapitre portant sur **la cooccurrence**.

  De manière générale, on formule l'hypothèse selon laquelle les **structures très fréquentes** sont **traitées facilement** et jugées **moins disfluentes** au sein des situations.
  Plus précisément, l'**accumulation de marques** provoquerait davantage de **ruptures prosodico-syntaxiques et discursives**. Les marques accumulées sont **plus facilement détectées** et évaluées comme disfluentes.

– L'**évaluation disfluente situationnelle** est estimée en fonction **des variants et des invariants de chaque pratique communicative** et donc en fonction des **normes sociales** qui y sont inhérentes. Ces normes sont présentes dans le contexte cognitif immédiat lors de la production discursive et de son évaluation.

  Globalement, les marques linguistiques (dis)fluentes produites dans des situations de parole aux paramètres prosodiques et syntaxiques homogènes, les **pratiques phonostylistiques normées**, sont plus sujettes à être **évaluées comme disfluentes** que les mêmes marques produites dans des situations soumises à peu de contraintes.
  Plus précisément, certaines situations laissent place à l'**expression d'un idiostyle** : chaque locuteur pourrait posséder son propre profil de (dis)fluence en fonction de la situation de parole. Certains locuteurs s'adaptent davantage que d'autres tout comme certaines situations demandent plus d'adaptation que d'autres.

– L'**évaluation disfluente sociale / interpersonnelle** dépend **des capacités interpersonnelles et des besoins** des locuteurs.

  La production et la perception d'un discours comme disfluent est caractéristique d'une **parole (auto)-consciente et surveillée**. Une attention élevée augmente la perception de la disfluence tandis qu'une capacité de projection empathique diminue la sévérité de l'évaluation (chapitre 1 section 2.3.4).

Chapitre 3

Données et méthodologie

La présente approche analytique de la production et de la perception des (dis)fluences se fonde sur l'**observation empirique de l'usage** (Barlow & Kemmer 2000). Elle fait appel à des données authentiques afin de tester les hypothèses pragmatiques et cognitives exposées au chapitre 2.

Ce chapitre présente les **données de corpus utilisées** et les **principes méthodologiques** présidant leur exploitation. On détaille la structure des données en termes de situations de parole et de traits situationnels (section 1). On développe les principes d'annotations des corpus [1], dont la validité aux niveaux syllabiques, lexicaux, prosodiques et syntaxiques est confirmée par des expériences d'interannotation (section 2). Ensuite, on présente une description quantifiée de la distribution des données brutes et annotées (section 3). Enfin, on propose une réflexion sur la complémentarité des résultats produits par les deux méthodologies linguistiques mobilisées – données de corpus en production et données expérimentales en perception – et les applications qui en ressortent (section 4).

1 Présentation des données de corpus

Cette recherche s'appuie sur deux corpus totalisant 5.52 heures de parole. Les échantillons de parole proviennent du ***LO**uvain **C**orpus of **A**nnotated **S**peech – **F**rench* (LOCAS-F ; Martin, Degand & Simon 2014) et du **C**orpus francophone de parole d'**Hum**oristes (C-Humour ; Grosman 2016) [2]. Les sections 1.1 et 1.2 résument le contenu respectif de ces corpus. Au total, ces données incluent 74 enregistrements (26 issus de C-Humour et 48 de LOCAS-F). La section 1.3 résume les durées d'enregistrement par corpus (par situation de parole et par trait situationnel).

1. Les manuels des corpus LOCAS-F (Annexe B) et C-Humour (Annexe A) font un résumé du contenu de chaque corpus, tandis que l'Annexe C reprend les détails techniques spécifiques à chaque niveau d'annotation.

2. Ponctuellement, cette recherche exploite d'autres corpus. Nous en faisons explicitement mention à chaque occasion.

1.1 Le *LOuvain Corpus of Annotated Speech - French* (LOCAS-F)

Le corpus *LOuvain Corpus of Annotated Speech - French* (LOCAS-F ; Martin, Degand & Simon 2014) compte 48 échantillons sonores produits entre 2005 et 2015 et représentant **14 pratiques communicatives différentes** : la conférence scientifique, le débat politique, le discours académique, le discours politique, l'homélie, l'interaction formelle (asymétrique), l'interaction informelle (symétrique), l'interview radiophonique, l'interview sociolinguistique, l'interview télévisée, le journal parlé radiophonique, la lecture radiophonique, la performance radiophonique à micro-ouvert et le récit conversationnel en face-à-face.

D'une durée totale de **4 heures**, les 48 enregistrements[3] du corpus regroupent 43000 mots simples, prononcés par 16 locuteurs en situation de monologue, 41 locuteurs en situation de dialogue et 19 locuteurs en situation de multilogue. Les 64 locuteurs uniques sont des francophones natifs d'origines géographiques variées. On identifie 37 enregistrements de locuteurs belges, contre 35 de locuteurs français et 5 de locuteurs suisses. Enfin, le corpus est relativement équilibré au niveau du sexe : 43% des locuteurs sont féminins et 57% masculins.

Les **travaux antérieurs** sur ce corpus (Degand, Eychenne, Simon & Van Damme 2011 ; Degand & Simon 2005 ; Martin, Degand & Simon 2014 ; Simon & Degand 2011) ont assuré le développement de plusieurs annotations mobilisées par la présente recherche. Celles-ci incluent notamment l'alignement de la transcription sur le son et sa segmentation phonétique, syllabique et lexicale via EasyAlign (Goldman 2011). Dans le cadre du projet FRFC[4], les échantillons ont été segmentés et **annotés syntaxiquement** en séquences fonctionnelles (Bilger & Campione 2002) et en unités de rection (Lefeuvre 1999). Le niveau lexical a également été enrichi d'annotations en **unités prosodiques** (unités majeures et intermédiaires, et contours intonatifs) (Mertens & Simon 2013). Un niveau supplémentaire découle du croisement de ces annotations : la segmentation en unités de base du discours (*Basic Discourse Unit* – BDU) (Degand & Simon 2009).

Dans le cadre de cette thèse, nous y avons ajouté un nombre d'annotations et de segmentations dont le détail est donné en Annexe C. A titre indicatif, on mentionne le développement des plus importantes : les **annotations des (dis)fluences** (aux niveaux lexical et syllabique), l'annotation automatique en **partie du discours** (*part of speech*, désormais *POS*) et sa correction manuelle, la segmentation en **groupes accentuels accentuables et accentués**, ainsi qu'en unités séparées par des pauses[5].

3. Pour des raisons de traitement, on retrouve autant de paires fichiers son et de Textgrids que d'interlocuteurs dans l'enregistrement. Ainsi, au niveau technique les 48 enregistrements correspondent à 77 paires de fichiers.

4. Projet FRFC n°2.4535.02, « Établissement d'une procédure de segmentation du discours oral en unités minimales (MDU) sur la base de l'interaction de critères syntaxiques et prosodiques et étude de la contribution des différents types de MDU à la compréhension », 2007 - 2008 (promotrices : L. Degand, A. C. Simon ; chercheuse impliquée : A. Dister).

5. Voir section 2 de ce chapitre.

1.2 Le Corpus de parole des Humoristes français (C-Humour)

Le corpus francophone de paroles d'humoristes (C-Humour ; Grosman 2016) comprend 26 échantillons de parole de 8 humoristes français, récoltés entre 2009 et 2011. Sa durée totale est de **1h56 minutes**. Il compte 6 monologues (les chroniques radiophoniques) et 20 dialogues[6]. Ces derniers incluent 4 interviews radiophoniques, 8 dialogues conversationnels informels et 8 sketchs sur scène.

L'échantillonnage de ce corpus permet de comparer le profil linguistique d'un même locuteur au travers de **trois situations**. Il s'intègre ainsi dans la problématique d'une définition adaptative, flexible et situationnelle de la (dis)fluence. La structure équilibrée du corpus encourage une étude contrastive selon plusieurs axes : le sexe de l'humoriste (4H / 4F), son degré d'expérience (+10 années / -10 années sur scène) et la situation de parole. Il permet d'étudier la **variation phonostylistique** de la parole de chaque humoriste au sein des différentes situations de parole (Grosman 2013b). A l'instar de LOCAS-F, le corpus a été aligné automatiquement au moyen d'EasyAlign (Goldman 2011). La description de son contenu est reprise en Annexe A. Il fait l'objet de 18 annotations dont le détail est présenté en Annexe C.

1.3 Description des paramètres situationnels des corpus

> « La notion de genre est une notion biface qui fait correspondre une face interne (les fonctionnements linguistiques) avec une face externe (les pratiques socialement signifiantes) [...]. L'instabilité de la relation entre formes et comportements sociaux institutionnalisés est une difficulté centrale pour toute définition à priori des genres. » (Branca-Rosoff 1999:116)

Cette recherche **examine l'aspect fluent ou disfluent des (dis)fluences, notamment au travers de la variation de la parole au regard des situations**. La description de la variation situationnelle s'accommode de deux approches complémentaires. Une approche holistique représente de manière globale les situations de parole en fonction des spécificités liées aux tâches communicatives associées. Une approche componentielle distingue les situations en fonction d'un set de traits contextuels. Le Tableau 3.1 résume le nombre de situations de parole et d'échantillons d'enregistrement par corpus. On distingue ainsi 18 types de situations de parole et 103 échantillons uniques de production de locuteurs.

6. Parmi les huit humoristes, quatre prononcent des dialogues dans le cadre de la chronique radiophonique contre 4 en interviews. Deux échantillons supplémentaires ont été ajoutés pour deux des humoristes en raison de la brièveté de leurs chroniques.

	LOCAS-F	C-HUMOUR
Types de situations de parole (n.)	14	4
Locuteurs uniques (n.)	62	8
Enregistrements (n.)	48	26
Fichiers de transcription	77	26

TABLEAU 3.1 – Nombre et types de situations par corpus.

Le panel A de la Figure 3.1 illustre la distribution des enregistrements des 18 situations de parole des corpus LOCAS-F et C-Humour. Le temps total d'enregistrement varie de 10.4 à 44 minutes ($\mu = 19.48$, $\sigma = 9.8$) par situation. Cet écart se justifie par des *taux d'articulation*[7] variables entre les situations.

La parole issue de sketchs sur scène compte un nombre important de segments de rires du public (réduisant le temps de parole du locuteur principal) ; contrairement à la parole issue de chronique radiophonique qui a un **débit articulatoire élevé**. Pour illustrer autrement la répartition des données analysables du corpus, les panels B et C de la Figure 3.1 représentent les échantillons en nombre de mots et en nombre de syllabes acoustiquement analysables (AA)[8].

Cette **approche holistique par situation** unifie la représentation des enregistrements par situation. Par là même elle peut parfois réduire l'hétérogénéité des enregistrements entre eux. D'une part, deux enregistrements d'une même situation peuvent se rattacher à des réalités différentes (e.g. une interview conversationnelle informelle peut être médiatisée ou réalisée dans un cadre privé). D'autre part, deux enregistrements de situations distinctes peuvent en réalité partager nombre de caractéristiques situationnelles (e.g. un discours politique et une homélie incarnent tous deux des discours monologaux et hautement préparés).

Les **hypothèses variationnistes**[9] nécessitent une description situationnelle plus fine des données de corpus. Depuis plus d'une vingtaine d'années, la linguistique a adopté plusieurs modélisations componentielles des contextes de parole (ESKENAZI 1993 ; KOCH & OSTERREICHER 2001 ; LLISTERI 1992 ; SIMON 2010 ; SIMON, AUCHLIN, AVANZI & GOLDMAN 2010). Celles-ci sont en général scalaires et formalisent leur description du contexte suivant une combinaison de **traits descriptifs des situations**, lesquels sont virtuellement infinis.

KOCH & OSTERREICHER (2001) proposent d'analyser la **dimension conceptionnelle** des productions discursives en les distinguant selon la manière dont est conçu et reçu un discours en fonction des contraintes du contexte. À ce niveau, on suppose que la

7. Le taux d'articulation est déterminé par le temps total de parole articulée (exempt de silences) divisé par le temps total d'enregistrement.

8. Les segments acoustiquement analysables (AA) excluent de la parole du locuteur principal (« L1 ») certains éléments empêchant l'analyse acoustique des échantillons tels que : (1) les chevauchements, (2) les phénomènes paralinguistiques (rires, toux, bruits) et (3) les silences interlocuteurs (*gaps*).

9. Présentées au chapitre 2 section 4.

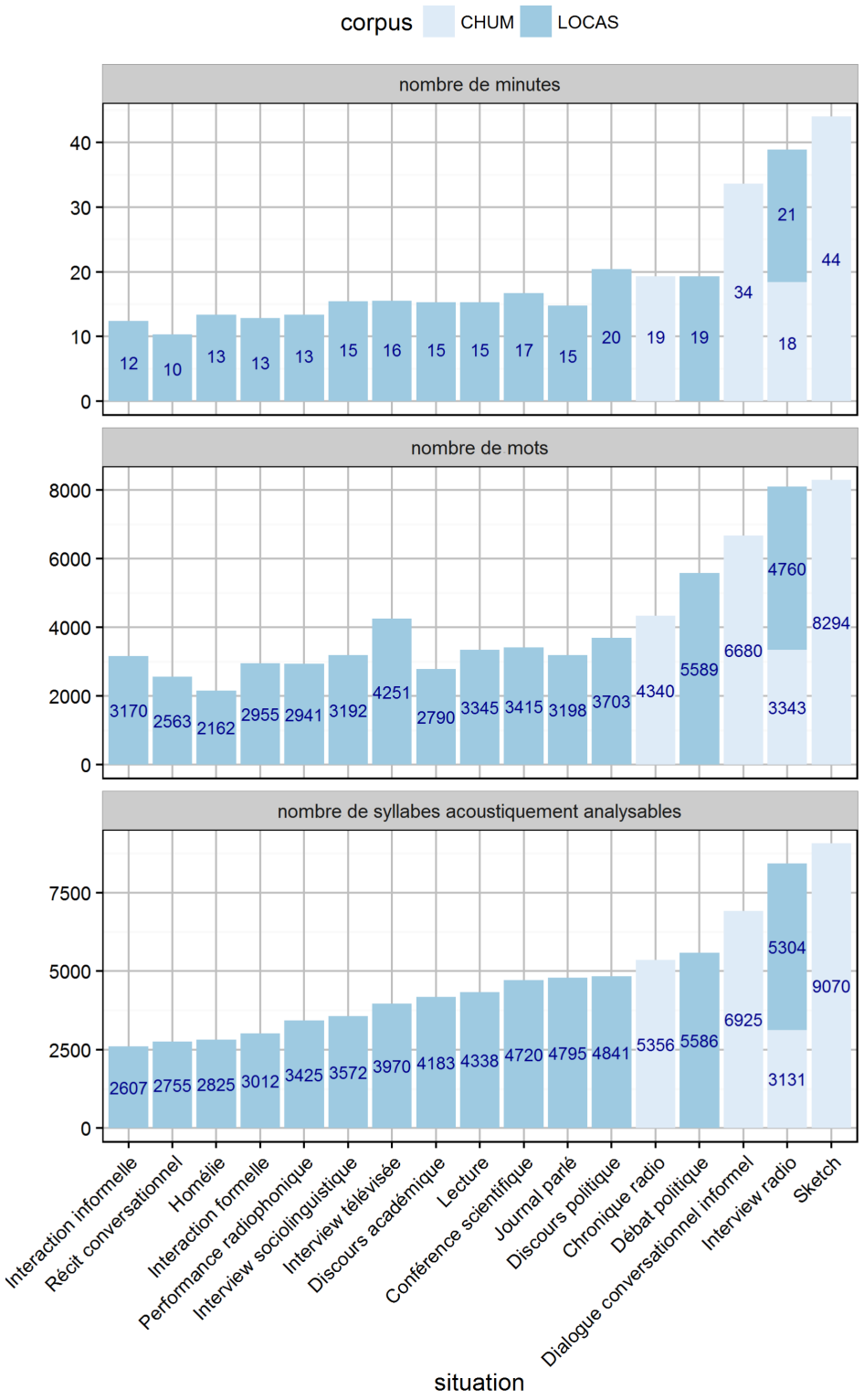


FIGURE 3.1 – Distribution des temps d'enregistrement, nombre de mots et nombre de syllabes acoustiquement analysables par situation et corpus.

conception du discours est influencée par un certain nombre d'éléments dans la situation : (1) le caractère privé /vs/ public d'une communication, (2) le degré d'intimité des locuteurs, (3) le degré d'émotivité, (4) l'ancrage actionnel et situationnel, (5) l'ancrage référentiel dans la situation, (6) la coprésence spatio-temporelle, (7) le degré de coopération communicative des interlocuteurs, (8) le caractère monologal /vs/ dialogal d'une situation, (9) le degré de spontanéité de la communication et (10) le degré de liberté ou fixation thématique. Les deux extrêmes du continuum permettent selon les auteurs de distinguer les discours incarnant la « **langue de la distance** » de ceux de la « **langue de la proximité** », équivalent respectivement à des styles **formels et informels**.

Afin de permettre une meilleure **comparabilité** entre les situations, en reformalisant les travaux de Koch et Osterreicher (*ibid.*), et faisant suite aux propositions de Simon (2010), Crible, Dumont, Grosman & Notarrigo (2016) ont adapté la définition de chaque trait situationnel aux **différentes modalités** concernées et différents types d'apprenants. **Six traits** ont été retenus avec, pour chacun d'entre eux, deux ou trois valeurs (/vs/ 10 dans la proposition initiale de Koch et Osterreicher) : (1) le degré d'élicitation, (2) le nombre de locuteurs en présence, (3) le degré de préparation, (4) le degré d'interactivité, (5) le degré d'orientation médiatique et (6) le degré de professionnalisation.

Cette méthodologie de catégorisation en traits situationnels a fait l'objet d'une **évaluation par 4 juges** experts sur 37 situations de parole distinctes (dont celles dans nos corpus)[10]. Cette évaluation portait sur la représentation des juges par rapport aux étiquettes des situations présentes dans les corpus – aucune tâche d'écoute n'était inclue. Ceux-ci devaient catégoriser les situations par les six traits situationnels. Le score Kappa de Fleiss (Fleiss 1971)[11] sur les six traits révèle un **accord interjuges global élevé** ($K = 0.88, n = 222, raters = 4, z = 113, p < 0.001$). Dans le détail, chaque score par trait situationnel reflète un accord presque parfait : situé entre [0.9-1] pour les degrés de médiatisation et d'élicitation, entre [0.75-0.9] pour le nombre d'interlocuteurs, le degré de préparation et le caractère professionnel d'une situation, et strictement inférieur à 0.75 pour le degré d'interactivité (ce qui reste un accord substantiel). Les évaluateurs tendent donc à catégoriser de manière similaire les situations par traits situationnels. La suite de cette section présente les données en fonction de cette classification, en minutes d'enregistrement.

- Le **degré d'élicitation** fait référence à la présence et au poids du protocole expérimental agissant comme une contrainte sur l'interaction. Les corpus à l'étude correspondent en majorité à des situations naturelles authentiques, sans aucun protocole expérimental, non suscitées par une action de recherche scientifique (Figure 3.2). Vingt-cinq minutes du corpus sont dites *semi-dirigées*. Ce sont des **productions naturelles mais supervisées** par un protocole expérimental souple dans le cadre d'une recherche scientifique

10. Les 37 situations de parole sont issues de 6 corpus : LOCAS-F (Martin, Degand & Simon 2014), C-Humour (Grosman 2016), la compilation de corpus DisFren (Crible 2017), LINDSEI (Gilquin, De Cock & Granger 2010), LOCNEC (De Cock et al. 2004) et le corpus LSFB (Meurant 2015).

11. L'utilisation d'un Kappa pondéré (Fleiss, Cohen & Everitt 1969) n'a pas été encouragée en raison de la délicatesse de son implémentation au-delà de deux juges.

guidant le choix du thème mais ne contraignant pas le locuteur à adopter une formulation préétablie. L'énonciateur y sélectionne librement ses mots. Ce trait situationnel caractérise l'interview sociolinguistique (`intfor`) et le récit conversationnel (`rcv`).

Les situations sont naturelles. Toutefois nombre d'entre elles sont suscitées par la **profession d'un des interlocuteurs** – ce qui réduit en général l'interactivité (Figure 3.3). Ainsi, 274.8 min. du corpus (83% du temps d'enregistrement) sont à caractère professionnel. Les situations privées, non professionnelles, de LOCAS-F et de C-Humour sont les interactions et les dialogues conversationnels informels (`conv-i`).

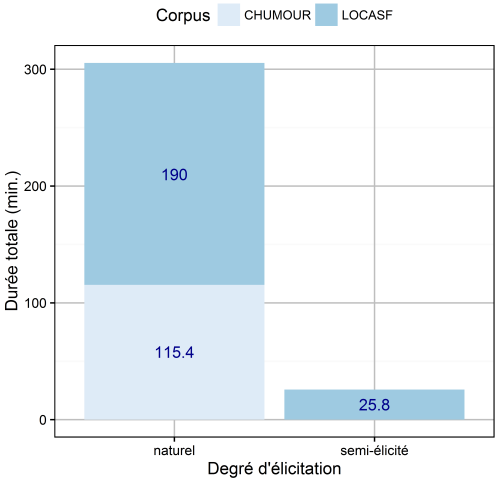


FIGURE 3.2 – Distribution des enregistrements en fonction du degré d'élicitation.

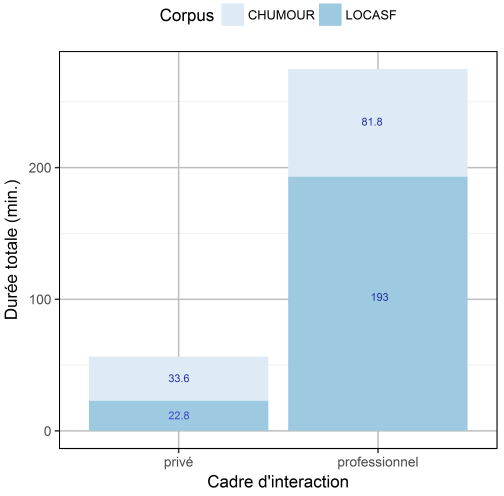


FIGURE 3.3 – Distribution des enregistrements en fonction du cadre d'interaction.

- Un troisième trait situationnel considère le **nombre de locuteur(s) actifs** dans l'interaction. La distribution des monologues, des dialogues et des multilogues est illustrée en Figure 3.4. Les données représentent en majorité des monologues et des dialogues avec une faible proportion de multilogues. La nécessité de sélectionner des enregistrements **prosodiquement analysables** (i.e. impliquant peu de chevauchement de parole), ainsi que l'influence de la gestion interactionnelle sur l'actualisation des (dis)fluences, motivent cette faible proportion [12].

- La Figure 3.5 représente la distribution des enregistrements en fonction du **degré d'interactivité**. Ce quatrième trait définit la capacité d'adaptation des interlocuteurs au regard des attentes sur le statut des locuteurs dans l'interaction. Ainsi, on entend par « situation interactive » une **relation symétrique** entre les locuteurs, permettant à chacun d'eux de prendre la parole. Logiquement, une situation non interactive est asymétrique : un des locuteurs prend et garde de manière quasi-continue la parole, sans possibilité d'intervention des interlocuteurs. Enfin, une situation semi-interactive se caractérise par le fait qu'un locuteur occupe plus la parole qu'un autre. Elle n'exclut pas pour autant des interventions ponctuelles des locuteurs secondaires.

12. Voir chapitre 1 section 1.3.

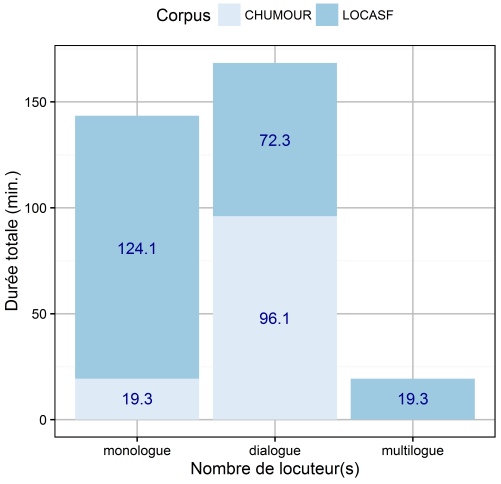


FIGURE 3.4 – Distribution des enregistrements en fonction du nombre de locuteurs.

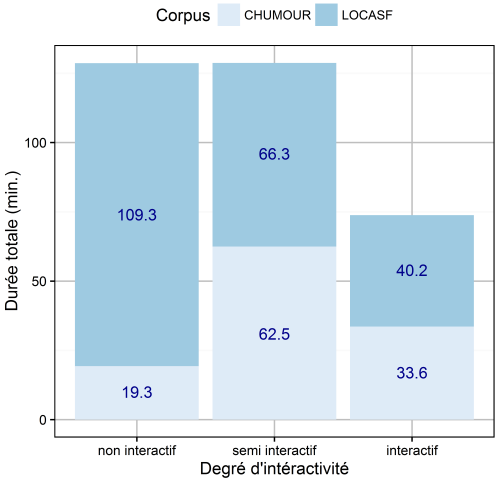


FIGURE 3.5 – Distribution des enregistrements en fonction du degré d'interactivité.

- Le trait suivant définit **le degré de préparation** orale et/ou écrite du discours du locuteur principal. Dans les situations préparées, le locuteur a entièrement élaboré, avant l'interaction, le contenu et la forme de son discours (i.e. le discours a été scripté et est formulé avec ou sans support textuel). Dans les situations spontanées, le locuteur improvise son discours au fur et à mesure de l'interaction et ce dernier n'a donc pas fait l'objet d'une préparation rigoureuse (BLANCHE-BENVENISTE 1997:17). Enfin, on définit un discours comme semi-préparé lorsque le locuteur a préparé la teneur globale de son discours sans l'avoir scripté dans sa totalité – l'interaction peut inclure un support textuel partiel. Les données sont relativement équilibrées au niveau du degré de préparation des discours avec 116 minutes de discours spontanés et semi-préparés et une proportion de discours préparés légèrement inférieure (98,5 min.) (Figure 3.6).

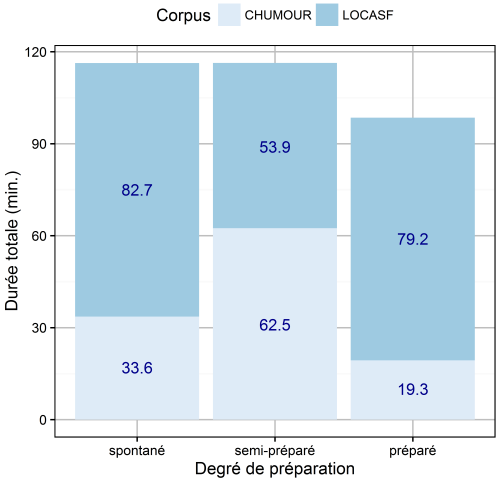


FIGURE 3.6 – Distribution des enregistrements en fonction du degré de préparation.

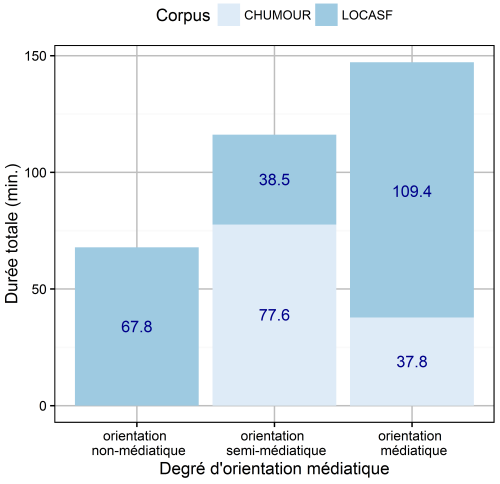


FIGURE 3.7 – Distribution des enregistrements en fonction du degré d'orientation médiatique.

- Le sixième trait situationnel est **le degré d'orientation médiatique** – soit l'importance de la diffusion dans les grands médias comme objectif principal de l'interaction. Les situations à caractère semi-médiatique représentent des interactions qui auraient eu lieu même si elles n'étaient pas diffusées par un média (e.g. homélie, discours politique, etc.) (Figure 3.7). Ici encore, les corpus offrent les trois types d'interactions, une part importante étant laissée à la parole médiatique (147 min.).

L'analyse en traits situationnels révèle un corpus relativement équilibré mais également une **colinéarité de certains traits** (notamment le degré de médiatisation et le degré de préparation ou le degré d'interactivité avec le nombre de locuteurs). L'utilisation de la description en traits présente un avantage méthodologique et analytique puissant en vertu de sa capacité à lier et à distinguer les situations entre elles. Ces traits sont utilisés en combinaison avec les étiquettes de situations de parole lors des chapitres d'analyses.

2 Principes d'annotation des corpus

> "Even when the experts all agree, they may well be mistaken."
> (Bertrand 1996:281)

Les corpus contiennent chacun plus d'une vingtaine de types d'annotation. Comme déjà mentionné, l'Annexe C reprend chaque annotation dans leurs détails méthodologique et technique. Cette section explique les annotations principales mobilisées dans le cadre de l'analyse des (dis)fluences en s'attachant particulièrement à leur validation par expériences de multiples annotateurs (2.1) et aux relations possibles entre les différents niveaux d'annotation (2.2).

2.1 Annotations des (dis)fluences

> "In connected speech, syllable boundaries often differ from a word's or morphemes canonical syllabification. The domain of syllabification is the phonological word, which can be smaller or larger than the lexical word due to morpho-phonological processes like inflection or cliticisation [...]."
> (Cholin & Levelt 2009:666)

Si les théories d'encodage et de planification de la parole s'accordent sur la réalité de l'**unité syllabique**, certaines divergences subsistent quant à l'interaction des différents mécanismes de planification (syllabiques et lexicaux) pouvant interférer pendant l'encodage phonétique (Cholin & Levelt 2009:683). Sans se positionner sur ce débat qui dépasse le cadre de cette recherche, nous bénéficions d'une **double annotation** des (dis)fluences : lexicale (2.1.1) et syllabique (2.1.2). Enfin, l'organisation syntagmatique des (dis)fluences répond à des structures régulières[13]. La méthodologie de description de ces structures varie en fonction de la granularité selon le sujet d'étude[14].

2.1.1 Annotations et validation au niveau lexical

Une analyse de la littérature a permis de comparer plusieurs protocoles d'annotation des (dis)fluences. En accord avec la littérature récente (Avanzi 2010 ; Candea 2000a ; Eklund 2004 ; Moniz 2013 ; Pallaud, Rauzy & Blache 2013 ; Shriberg 1994 ; Strassel 2004), une **approche componentielle** des structures de (dis)fluence a été adoptée. Sans en faire le détail, les différences entre les protocoles d'annotations résident principalement dans (1) le nombre de phénomènes annotés, (2) la présence d'un parti pris théorique d'annotation de phénomènes cognitivement et pragmatiquement fluents et/ou disfluents, (3) le niveau d'annotation (lexical ou syllabique) (4) la reconnaissance de la structure interne d'une (dis)fluence – *reparandum*, *interregnum* (point d'interruption et termes d'édition) et *reparans*.

13. Voir chapitre 1 section 1.2.4.
14. Voir chapitre 3 section 2.2.

Par exemple, l'intention originelle de Shriberg (1994:1) était une intégration d'un maximum de phénomènes par l'adoption d'une **théorie relativement neutre**. Elle y intègre les répétitions, substitutions, insertions, délétions, pauses pleines, termes explicites d'édition, certains usages des marqueurs discursifs mais également les conjonctions de coordinations, les amorces des mots, les erreurs d'articulations et les incomplétudes syntaxiques. Cependant, elle exclut les pauses silencieuses isolées et certains marqueurs discursifs en raison de la **prétendue intentionnalité** du locuteur de les produire. Ainsi, si son intention originelle était une théorie neutre de la (dis)fluence, elle restreint cependant l'annotation à des **(dis)fluences « agrammaticales »**.

En suivant la terminologie de Levelt (1989) présente dans son *Repair Interval Model* de son modèle de production de la parole, l'annotation inclut les ***overt* et *covert-repairs***[15]. Dans les termes de Ginzburg, Fernández & Schlangen (2014:4-6), elle reprend donc les (dis)fluences de *backward-looking*, se référant à un *reparandum* déjà énoncé, et les (dis)fluences de *forward-looking*, se référant à la simple complétion d'un énoncé retardé par la présence d'une pause ou d'une répétition.

Etiquettes	**Fluencèmes (FR)**	**Fluencèmes (EN)**
UP	Pause silencieuse	*Unfilled Pause*
FP	Pause pleine	*Filled Pause*
DM	Marqueur de discours	*Discourse Marker*
ET	Terme d'édition	*Editing Term*
FS	Faux-départ	*False Start*
TR	Troncation	*Truncation*
RI	Répétition à l'identique	*Identical Repetition*
RM	Répétition avec modification	*Modified Repetition*
SP	Substitution propositionnelle	*Propositional Substitution*
SM	Substitution morphologique	*Morphological Substitution*
IL	Insertion lexicale	*Lexical Insertion*
IP	Insertion parenthétique	*Parenthetical Insertion*

Tableau 3.2 – Liste des (dis)fluences annotées dans la tire principale – inspiré de Crible et al. (2016).

Dans le cadre de l'ARC *Fluency & Disfluency Markers*, un schéma d'annotation adapté aux **données multimodales** (langue des signes /vs/ langue orale), non spécifique à une langue (anglais /vs/ français) ou à un type de locuteurs (natifs /vs/ non natifs) était requis (Crible et al. *in prep.*). Le protocole formulé respecte le parti pris de ne pas annoter à priori le caractère fluent ou disfluent des (dis)fluences. Enfin, il opte pour une annotation basée sur la **forme de surface**, au niveau du mot simple. Les

15. i.e. les (dis)fluences impliquant un changement, une addition ou délétion de segment (*overt-repairs*) et celles principalement constituées autour d'un point d'interruption (*covert- repairs*).

détails théoriques et méthodologiques du protocole sont disponibles en Annexe D. Le Tableau 3.2 reprend la liste des (dis)fluences annotées dans les corpus C-Humour et LOCAS-F.

La qualité d'une étude se confirme notamment par la **reproductibilité** des analyses et des annotation(s) mobilisée(s). A cet égard, une mesure d'accord interannotateurs peut servir d'indice de fiabilité et de qualité d'une annotation. Nous avons comparé l'annotation lexicale des (dis)fluences par nous-même et une autre annotatrice experte sur 8 enregistrements issus de la parole radiophonique (4 interviews de C-Humour et 4 interviews de LOCAS-F). Après avoir présenté les résultats de la convergence des données, nous évoquons les éventuels **défis d'ambiguïtés ou d'interprétations** révélés par l'accord interannotateurs.

L'évaluation repose sur 25.7 minutes de parole articulée [16] (1542.6 sec.), soit 6892 lexèmes à annoter selon les catégories présentées dans le Tableau 3.2. L'annotation de référence porte l'acronyme IG et l'annotation d'hypothèse LC. En vertu de la combinaison possible des étiquettes (e.g. un marqueur discursif répété à l'identique porte l'annotation < DM-RI >, car on combine l'annotation DM avec l'annotation RI), le set de données initial confrontait 32 étiquettes. L'accord interannotateurs obtenu sur l'ensemble des données s'élève à $K = 0.67$ ($raters = 2, z = 98.60, p < 0.001, n = 6892$), ce qui correspond à un **accord fort**. Le Tableau 3.3 détaille ce score.

		valeurs prédites (IG)	
		n | %	**n | %**
valeurs observées (LC)		1413| 20.5	162 | 2.4
		(True Positive)	*(False Negative)*
		645 | 9.4	4672 | 67.8
		(False Positive)	*(True Negative)*

TABLEAU 3.3 – Classification de l'accord interannotateurs des (dis)fluences lexicales.

Tandis que 20.5% et 67.8% des segments ont respectivement été annotés ou exemptés d'annotation par les deux juges, 9.4% des données correspondent à une erreur de type-I (*False Positive*) où IG a détecté un segment non annoté par la juge LC. Au contraire, seuls 2.4% de segments correspondent à une erreur de type-II, i.e. qui sont annotés par LC mais ne l'ont pas été par IG (*False Negative*). Une annotatrice a donc eu tendance à annoter plus qu'une autre.

16. Cette étude exclut de l'accord interannotateurs l'évaluation des pause silencieuses (<UP>), leur segmentation par les deux experts ayant fait l'objet d'un traitement différent et relevant davantage de critère de segmentation et de détection que d'une annotation humaine.

Une seconde évaluation exclut les étiquettes comptabilisant strictement moins de 10 occurrences dans les corpus de LC et d'ID (règle `n-sup-10`). Les étiquettes concernées par la règle `n-sup-10` correspondent toutes à des cas de suites de (dis)fluences complexes – simples ou composées (e.g. < RMSP0 >, répétition avec modification combinée à une substitution propositionnelle) [17]. Cette manipulation facilite la visualisation de la **matrice d'évaluation**, réduite à 13 étiquettes pour chaque annotateur. Le score Kappa global n'en est pas modifié. Au Tableau 3.4, la diagonale bleue représente les cas d'accord parfait entre les annotatrices. Le gras souligne les désaccords de détection. La première ligne horizontale rassemble les cas non annotés par LC mais annotés par IG.

		IG													
		_	**DM**	**DMRI**	**ET**	**FP**	**FS**	**IL**	**IP**	**RI**	**RM**	**SM**	**SP**	**TR**	**TOT**
LC	**_**	4715	185	5	7	6	10	19	34	131	98	17	106	14	5347
	DM	52	410	9	**15**	0	0	0	1	1	0	0	1	0	489
	DMRI	0	0	13	0	0	0	0	0	8	0	0	0	0	21
	ET	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	FP	1	1	0	0	139	0	0	0	0	0	0	0	0	141
	FS	9	0	0	0	1	10	0	0	1	2	7	4	1	35
	IL	10	2	0	0	0	0	25	0	1	0	0	4	0	42
	IP	0	2	0	0	0	0	1	19	0	0	0	0	0	22
	RI	15	3	13	0	0	1	0	0	393	2	6	3	1	442
	RM	26	0	0	0	0	0	0	0	**58**	103	8	2	4	202
	SM	2	0	1	0	0	0	0	0	1	0	36	2	0	43
	SP	11	0	0	0	0	0	0	0	1	0	**21**	17	1	52
	TR	6	0	0	0	1	2	0	0	2	1	0	0	52	64
	TOT	4847	603	41	22	147	23	45	54	597	206	95	139	73	6892

TABLEAU 3.4 – Matrice de confusion de l'accord interannotateurs des (dis)fluences lexicales.

Remarquons que, de manière générale, IG a tendance à annoter davantage de marqueurs que LC, en particulier dans le cas des marqueurs discursifs (DM = 185), répétitions à l'identique (RI = 131) et substitutions propositionnelles (SP = 106). A l'inverse, seuls 52 DM annotés par LC n'ont pas été détectés par IG, tandis que 15 d'entre eux sont catégorisés comme des ET (*Editing Term*, termes d'éditions). Pour finir, une partie substantielle des désaccords provient en réalité d'une **confusion**, d'une part entre RM (*répétition avec modification*) et RI (*répétition à l'identique*) (tendance de IG à annoter plus de RI que de RM par rapport à LC), et entre SM (*substitution morphologique*) et SP (*substitution propositionnelle*) d'autre part (tendance de IG à annoter plus de SM que de SP par rapport à LC). Ces deux dernières observations laissent entrevoir une faille intéressante du protocole quant à l'annotation des RM et SP. Ces deux (dis)fluences composées font en réalité intervenir des **notions de morphosyntaxe** (catégorie lexicale /vs/ grammaticale) et de syntaxe (séquences fonctionnelles). Des analyses divergentes de ces niveaux

17. N'étant plus prises en compte dans cette seconde évaluation, ces suites de chaînes de (dis)fluences complexes ont automatiquement été remplacées par l'étiquette « _ » symbolisant l'absence d'annotation.

pourraient être à l'origine des **désaccords de catégorisation** des (dis)fluences. L'affichage de ces niveaux d'annotations pendant l'expérience interannotateurs permettrait sans doute d'accroître le score.

Enfin, lorsque le score Kappa est calculé sur l'ensemble des segments annotés comme porteurs d'au moins une (dis)fluence par les deux annotateurs, le score d'accord interannotateurs est significativement plus élevé et reflète ainsi un excellent accord interannotateurs ($K = 0.82, raters = 2, n = 1413, = 64.83, p < 0.0001$). En d'autres termes, lorsque les deux annotateurs détectent une (dis)fluence, **dans 82 cas sur 100**, ils sont d'accord sur la catégorisation de cette marque.

Peu d'études existent sur l'accord interannotateurs des (dis)fluences annotées par des experts sur un nombre de marqueurs si important. Besser & Alexandersson (2007:187) rapportent un score Kappa de 0.93[18] ($n = 792$[19] extraits du AMI meeting corpus, $raters = 4$). Ils relèvent également plus de **difficultés à la détection** qu'à la catégorisation et plus de difficultés à l'annotation des répétitions avec modification et des DMs que pour les catégories plus classiques telles que les pauses pleines et les substitutions morphosyntaxiques. A titre de comparaison uniquement, bien que Hough, de Ruiter, Betz & Schlangen (2015:2-3) n'annotent pas la catégorie de (dis)fluence mais plutôt la partie de la structure de la disfluence à laquelle le mot appartient (*reparandum, repair, pause longue* et *rires*), ils effectuent 4 calculs d'accords interannotateurs pour chacune de ces catégories (e.g. le mot fait partie ou non du *reparandum*) et observent des valeurs de Kappa allant de 0.8 à 0.9 sur 20000 lexèmes des données issues du DAP Kousidis, Pfeiffer & Schlangen (Dream Apartment Corpus ; 2013). Ce score relativement plus élevé s'explique notamment par la différence de **granularité de l'annotation**.

2.1.2 Annotations et validation au niveau syllabique

Traditionnellement, la recherche sur les (dis)fluences s'est focalisée sur leurs propriétés lexicale (e.g. partie du discours, chaînes de cooccurrences, etc.) et fonctionnelle (e.g. interactive, structurante, etc.). Nombre d'études proposent désormais une analyse des **propriétés phonétiques des (dis)fluences**, notamment au point ou autour du point d'interruption de la disfluence (Shriberg 1999:621 ; Hindle 1983:124). Une analyse cantonnée au niveau lexical ne rend pas compte des propriétés phonétiques : l'**allongement des durées de la rime ou de la syllabe**, la conservation des **contours intonatifs** entre le *reparandum* et le *reparans*, la laryngalisation, les phénomènes de coarticulation, la qualité vocalique, etc. ne sont analysables qu'au niveau syllabique. Le marquage prosodique permettrait entre autres de distinguer les *reparandum* impliquant une « erreur » de ceux impliquant une « élaboration » (Levelt & Cutler 1983).

18. Pourtant, les scores interannotateurs détaillés sont tous sous la barre des 0.93.
19. A noter que les auteurs analysent 1206 disfluences dont 791 « segments », sans précisions supplémentaires.

Ainsi, une annotation syllabique, nommée ***delivery***, des phénomènes de (dis)fluences et des phénomènes paralinguistiques a été réalisée sur les corpus. Une partie de cette annotation a fait l'objet d'une **évaluation par trois experts** francophones natifs. Cette évaluation porte sur 12 min. de parole équivalant à 3266 syllabes issues de 3 enregistrements de LOCAS-F et de 3 enregistrements du corpus de thèse de Rousier-Vercruyssen (Fossard et al. 2018). **Quatre catégories** ont été évaluées : (1) syllabe non disfluente, (2) syllabe porteuse d'un allongement d'hésitation, (3) syllabe représentant une pause pleine (*euh* isolé, mais également *hum, mh* et schwa post-tonique), et, de manière confondue, (4) le *reparandum* de quatre autres (dis)fluences (i.e. *répétitions, amorces, autocorrections, insertions*). Les résultats sont détaillés au Tableau 3.5 [20].

Phénomène	Kappa	Z	Sig.
Tout segment	0.72	75.53	***
Segment non annoté	0.76	74.78	***
Pause pleine	0.86	85.55	***
Allongement d'hésitation	0.64	63.11	***
Reparandum d'autres (dis)fluences	0.44	43.19	***

Tableau 3.5 – Accords interannotateurs (global et par facteur) des (dis)fluences au niveau syllabique.

Le résultat global indique un **accord interannotateurs élevé** ($K = 0.72, n = 3266, raters = 3, p < 0.001$) suggérant une détection substantielle et une bonne identification des (dis)fluences au niveau syllabique, particulièrement pour les pauses pleines et les allongements d'hésitation. Une étude approfondie de ces désaccords fait l'objet d'une étude séparée au chapitre 5 section 2.1. Il est notamment question d'évaluer l'**influence des paramètres** expérimentaux (annotation *online* /vs/ *offline*), acoustiques (allongement objectif de la syllabe, courbes intonatives), syntagmatiques (cooccurrence des (dis)fluences), morphosyntaxiques (types de lexème), syntaxiques et prosodiques (position du phénomène dans une unité) **sur la détection et la perception** des (dis)fluences syllabiques (Grosman, Rousier-Vercruyssen & Duvivier 2016).

2.2 Méthodologie de description des structures de (dis)fluences

L'organisation syntagmatique des (dis)fluences répond à des structures régulières [21]. La méthodologie de détection et de description de ces structures varie en fonction de la granularité désirée par l'étude. Ce niveau de granularité fait intervenir au moins deux niveaux d'analyse, illustrés en Figure 3.8.

20. *** indique une significativité de $p < 0.001$.
21. Notamment celles de *reparandum*, *interregnum*, *reparans*. Voir chapitre 1 section 1.2.4.

1. **Le type de structure :** au premier niveau, on distingue les (dis)fluences en fonction des composants essentiels de leur structure : les (dis)fluences à composante **simple** (`DS`) (*pauses silencieuses, pauses pleines, troncations inachevées, faux départs*) et (dis)fluences **composées** (`DC`), lesquelles sont structurées autour d'un *reparandum*, d'un *reparans* et d'un *interregnum* optionnel.
2. **La complexité de la structure interne :** à un second niveau, les (dis)fluences se distinguent par leur environnement syntagmatique direct. Dans le cas des `DS`, on distingue les (dis)fluences **isolées** (i.e. absence de cooccurrence) (`DS.A`) de celles **en cooccurrence** avec d'autres marques (`DS.CO`) (e.g. une pause silencieuse seule /vs/ une pause silencieuse suivie d'un marqueur discursif). Dans le cas des `DC`, les (dis)fluences se différencient notamment par la présence et le **type d'*interregnum*** en leur sein : *interregnum* **absent** (`DC.IA`) (e.g. répétition directe), *interregnum* rempli par une **(dis)fluence simple** (`DC.IS`) (e.g. répétition avec une pause en phase d'édition), une **(dis)fluence composée** (`DC.IC`) (e.g. répétition combinée à une troncation) ou une **combinaison des deux** (`DC.ISC`) (e.g. répétition combinée à une troncation et une pause).

Certaines analyses, notamment dans le cadre de la sélection de stimuli pour les expériences de perception, font intervenir des niveaux supplémentaires de description des structures internes (e.g. nombre d'éléments dans le *reparans*, types de contours intonatifs) et des chaînes de (dis)fluences (e.g. présence et position d'éléments succédant ou précédant) [22]. Lors de la désambiguïsation d'une (dis)fluence comme vecteur de fluence /vs/ disfluence, ces différentes structures sont mises en relation avec leur forme acoustique, leur contexte prosodique et leur contexte syntaxique.

2.3 Annotations prosodiques et syntaxiques

Les corpus comportent des niveaux d'annotations permettant des mises en relation croisées. Ainsi, par exemple, une même annotation peut englober plusieurs lexèmes. Certains niveaux d'annotations sont **hiérarchiquement dépendants** d'autres. L'ensemble des annotations permet la contextualisation prosodique et syntaxique des (dis)fluences [23]. Cette section présente leur définition sommaire et leurs principales caractéristiques.

Le corpus LOCAS-F est davantage annoté aux niveaux syntaxique et prosodique que le corpus C-Humour. Certaines analyses de contextualisation des (dis)fluences sont donc effectuées uniquement sur LOCAS-F. Le Tableau 3.6 reprend les types d'unités syntaxiques et prosodiques annotées dans les corpus [24].

22. Les schémas détaillés et plus complexes de ces structures sont repris en Annexe D. La répartition des structures de (dis)fluences dans nos corpus est donnée en section 3.3 de ce chapitre.
23. Le détail technique de ces annotations est repris dans les Annexes C.
24. Etant donné que les distributions de la durée et longueur (en mot ou syllabe) des unités présentées ne suivent pas une courbe normale, la présentation des moyennes et écarts-types n'est pas pertinente.

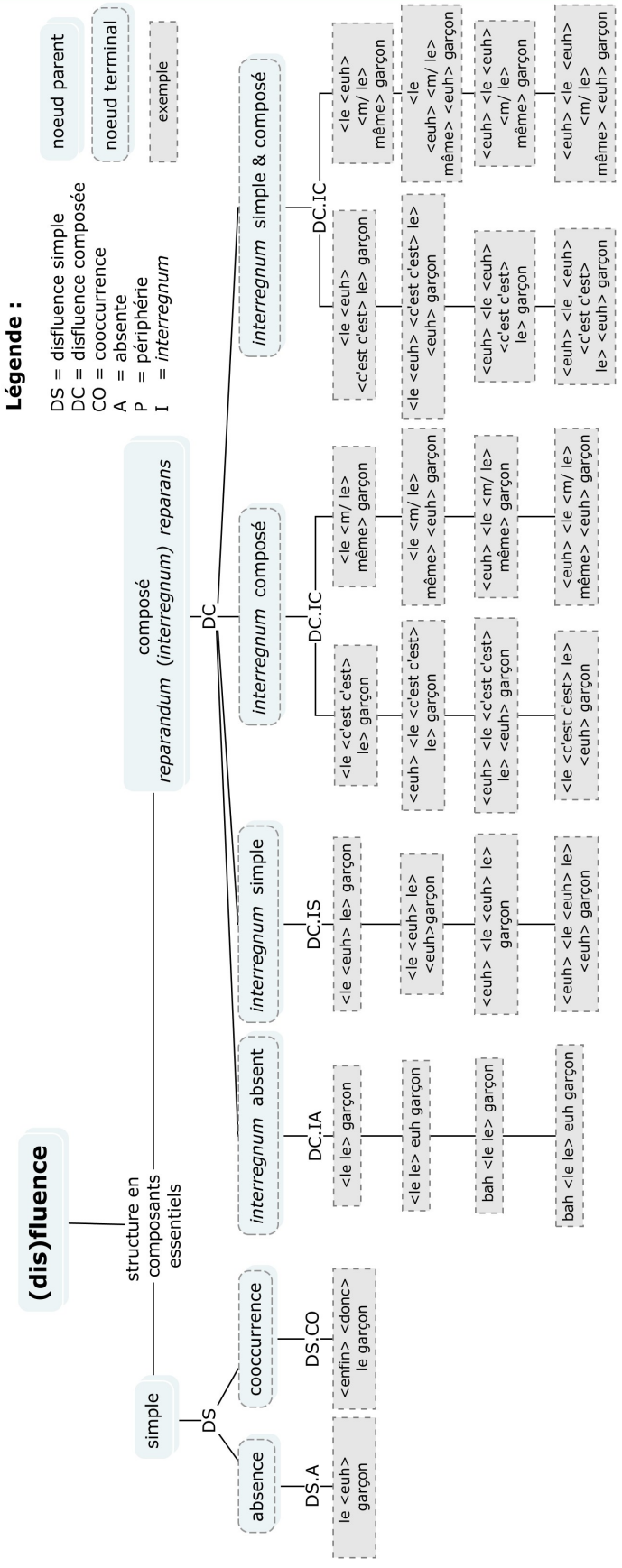


FIGURE 3.8 – Méthodologie de description d'une structure de (dis)fluence.

Abréviation	Unité suprasegmentale
BDU	*Unité de base du discours*
IPU	*Unité Inter Pausales*
UR	*Unité de rection*
SEQ	*Séquence fonctionnelle*
ga-0	*Groupe accentuel accentuable*
ga-1	*Groupe accentuel accentué*

TABLEAU 3.6 – Types d'unités prosodiques et syntaxiques annotées dans les corpus.

Le **Groupe Accentuel accentuable** (`ga-0`) ou *accentème* désigne une représentation abstraite déterminée par la présence d'un **accent lexical virtuel** (MERTENS 1993:40 ; LACHERET-DUJOUR & BEAUGENDRE 1999:42). En français, le groupe accentuel est l'unité minimale porteuse d'un accent final primaire. Sa formation part du principe que les mots lexicaux sont susceptibles de recevoir un **accent primaire sur leur syllabe finale**. Au contraire, les mots fonctionnels sont dits *atones* et ne s'intègrent dans un groupe accentuel qu'en dépendance à une tête. La frontière droite d'un groupe accentuel accentuable s'actualise pour toute **syllabe pleine de mot**. Il y a, dans la structure profonde, autant de groupes accentuels que de mots lexicaux (MERTENS 2008:101).

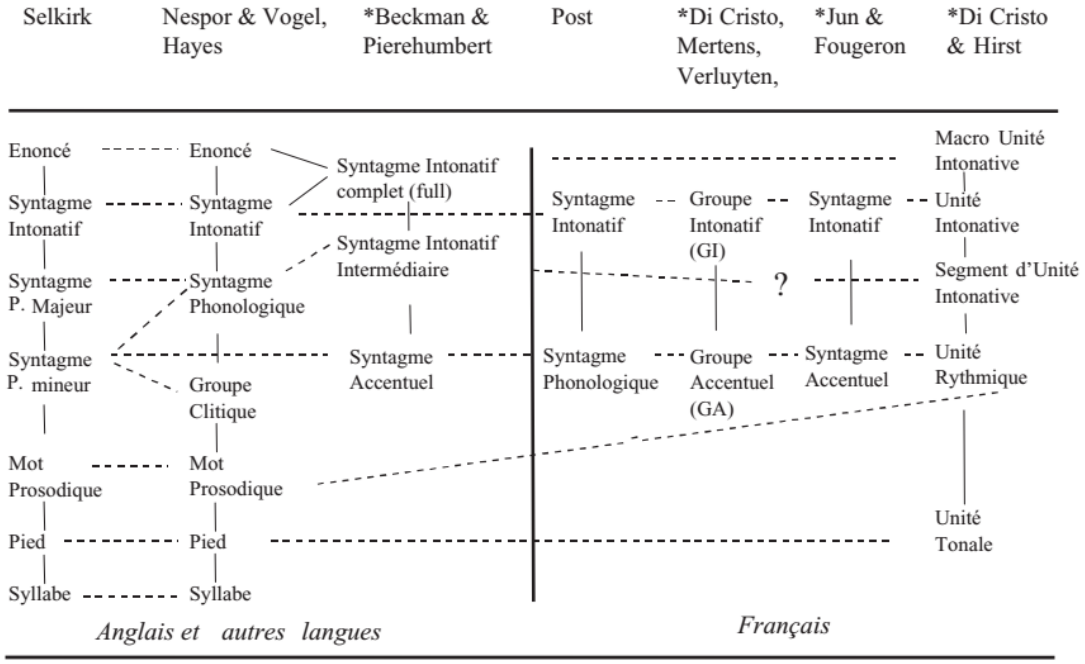


FIGURE 3.9 – Terminologie utilisée dans les approches actuelles de la hiérarchie des ´ domaines prosodiques. L'astérisque devant les noms d'auteurs signale les approches basées uniquement sur des critères prosodiques – issu de DI CRISTO (2011).

Le **Groupe Accentuel accentué** (`ga-1`) est une unité de surface basée sur le groupe accentuel accentuable et reconnaissant les **mécanismes de désaccentuation et de transfert accentuel**. Ces mécanismes respectent des principes rythmiques (e.g. deux

accents finaux contigus sont évités) et syntaxiques (e.g. regroupement syntaxique d'un seul objet informationnel) reflétant la structuration intonative du message (Mertens 2008:100-102). Le ga-1 correspond à une suite de syllabes dont la dernière est marquée par un accent primaire (Lacheret-Dujour & Beaugendre 1999:45)[25]. L'accent primaire est lui-même défini par une syllabe finale d'un mot lexical, **porteuse d'une proéminence sonore** (e.g. marquée par une augmentation de l'intensité, un détachement ou glissando intonatif, ou une durée syllabique relativement supérieure aux syllabes environnantes) (Mertens 1993:21).

La syntaxe est annotée à deux niveaux : celui de l'unité de rection et de ses séquences fonctionnelles. L'**Unité de Rection** (`UR`) est organisée autour d'un noyau (i.e. élément recteur). Il s'agit généralement d'un verbe tensé, bien que le noyau de l'unité puisse également être un nom, pronom, adverbe ou groupe prépositionnel prédicatif (Lefeuvre 1999). La **séquence** (`SEQ`) est une unité fonctionnelle segmentant hiérarchiquement les unités rectionnelles en (i) séquence sujet (SS), (ii) séquence verbale (SV), (iii) séquence objet (SO) et (iv) séquence régie (SR) (Tanguy 2012).

L'**Unité Inter-Pausale** (`IPU`) se définit comme une vocalisation par un locuteur unique bornée par des pauses silencieuses au sein d'une bribe de parole émise par un locuteur (« *talk-spurt* »). Elle est précédée et suivie par la parole d'un autre locuteur et perçue par celui-ci comme une bribe de parole (Jaffe & Feldstein 1970 ; Wlodarczak & Wagner 2013:1434).

L'**Unité de Base du Discours** (`BDU`) est une segmentation générée à partir de la correspondance entre les unités de rection et les unités prosodiques majeures. Nous distinguons ainsi cinq types de BDU. Les **congruentes (BDU-c)** où une unité de rection correspond à une unité prosodique majeure. Les **régulatives (BDU-r)**, constituées d'un élément syntaxiquement non régi (i.e. connecteur, adverbe de phrase ou marqueur discursif) et isolées par une frontière prosodique. Celles **groupées par l'intonation (BDU-i)** sont des cas où une unité prosodique regroupe plusieurs unités de rection. A l'inverse, celles **groupées par la syntaxe (BDU-s)** sont des cas où une unité de rection est segmentée en plusieurs unités prosodiques majeures. Enfin les **mixtes (BDU-x)** regroupent plusieurs unités syntaxiques et plusieurs unités prosodiques majeures qui ne s'achèvent pas de manière synchronisée.

La Figure 3.10 reprend schématiquement les **liens entre les différentes annotations**. On distingue (1) les données brutes (*en gris*), (2) les niveaux de segmentation (*en mauve*) et (3) les niveaux d'annotation dépendants d'un autre niveau de segmentation (*en bleu*). Par une lecture de gauche à droite, nous résumons ci-après les liens entre les annotations. Nous renvoyons le lecteur à l'Annexe C pour une formalisation détaillée.

25. Le concept de « groupe accentuel » désigne des objets différents selon la posture épistémologique des théories. Ici, la reconnaissance de la hiérarchie accentuelle (accents secondaires /vs/ primaires) valorise les unités linguistiques au détriment des unités abstraites (e.g. pieds accentuels).

Dans un premier temps, à partir des paramètres acoustiques, la transcription orthographique est phonétisée (tire `phones`) et regroupée en syllabes (tire `syll`) dont la combinaison offre une segmentation au niveau du mot simple (tires `words`/`tok-min`) annoté en partie du discours (`POS-treetag` et `POS-dismo` représentent deux sets d'annotation à granularité variable). De cette annotation grammaticale découle la tire `lex` qui distingue les mots grammaticaux considérés comme clitiques des mots lexicaux non clitiques (i.e. accentuables). Indépendamment, l'annotation des (dis)fluences et des phénomènes contextuels liés telle que prévue par le protocole-ARC (Crible, Dumont, Grosman & Notarrigo 2016) est alignée sur cette segmentation lexicale (tires `disfluency-arc` et `diacritics`). Ensuite, la tire `token` donne elle-même lieu à plusieurs niveaux de segmentation : Au niveau inférieur, les mots simples sont segmentés automatiquement en phones et regroupés en syllabes dans la tire `syll`. Au niveau supérieur, des dépendances existent :

1. En tours de parole (i.e. en fonction de(s) (l') allocutaire(s) de la parole) sur la tire `speaker` (e.g. locuteur principal, locuteur secondaire, chevauchements ou aucun allocutaire – *gap*).

   Par ailleurs, les segments « silence » présents dans la parole du locuteur principal sur la tire d'annotation `speaker` permettent de calculer un seuil moyen de la durée des pauses par locuteur et de construire la tire (de segmentation) `IPU.file` et la tire (de segmentation) `IPU.genre` (*Inter-Pausal Unit* ou unités séparées par des pauses)[26].
2. En séquences fonctionnelles (`sequence`), elles-mêmes regroupées en unités de rection (`rection`).

Enfin, de nombreuses tires d'annotations mettent en relation les niveaux lexical et syllabique. Par exemple :

1. Les chevauchements de la tire `speaker` et toute autre perturbation du signal sonore sont signalés au niveau syllabique dans la tire d'annotation `delivery`.

   Delivery est ainsi une tire d'exclusion des phénomènes paralinguistiques qui pourraient interférer avec l'annotation des proéminences syllabiques (tire `promauto`).
2. Les annotations de la tire `if` (initial-final) indiquant la position d'une syllabe dans un mot et dans un groupe accentuel accentuable sont issus des informations syllabique, lexicale et syntaxique.
3. Le « *chunking* » en *groupe accentuel accentuable* (tire `ga-0`) est fondé sur `lex`. Ces deux annotations sont couplées à la tire `promauto` : les proéminences finales s'actualisant sur une syllabe finale d'un mot lexical (non marquée par un schwa post-tonique) segmentent le discours en groupes accentuels accentués (`ga-1`).

En outre, au niveau syllabique, les **données acoustiques extraites automatiquement** (e.g. *amplitude, durée relative, F0 relative, mouvement de F0, toux/rires,* etc.) ont permis de segmenter automatiquement les enregistrements en **frontières mineures**

26. Voir chapitre 4 section 1.

et majeures (tire `boundary`) et d'assigner notamment à chaque syllabe précédant une frontière un **contour intonatif** (tire *contours*). Enfin, la détection des coïncidences entre les frontières prosodiques (tires `boundary`), les contours intonatifs (tire `contours`) et les unités syntaxiques (tire `rection`) donne naissance à une segmentation en unités de base du discours (tire `BDU`).

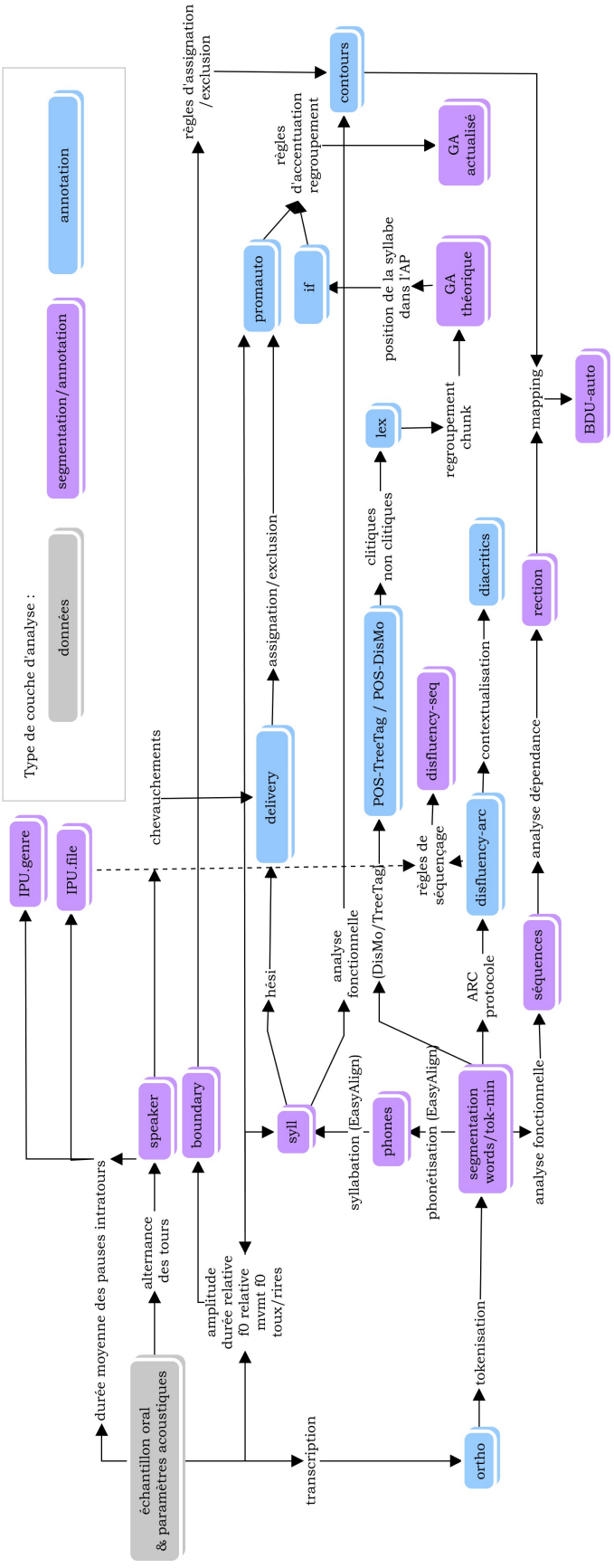


FIGURE 3.10 – Schéma relationnel des annotations et des segmentations des corpus.

3 Distributions des phénomènes annotés dans le corpus

> "Multi-Dimensional studies of register variation have uncovered both surprising similarities and notable differences in the underlying dimensions of variation." (Friginal & Hardy 2014:241)

Cette section décrit quantitativement les données des corpus. La répartition du temps de parole met en évidence la parole prosodiquement analysable et le nombre d'unités prosodiques (3.1) et syntaxiques (3.2) contenues dans les corpus. La fréquence des phénomènes de (dis)fluences et leur répartition selon les traits situationnels donnent un premier aperçu des possibilités d'analyses qu'offrent ces annotations (3.3).

3.1 Description quantitative du temps de parole

Suite à l'article fondateur de Sacks, Schegloff & Jefferson (1974), aux propositions d'opérationnalisation de Edlund, Hirschberg & Heldner (2009) et Heldner & Edlund (2010) ainsi que de Fors (2011), on distingue au sein du temps de parole d'un locuteur, les **pauses interlocuteurs** des **pauses intralocuteurs** (Tableau 3.7). Les premières, dites aussi *gaps*, apparaissent entre deux tours de parole, lorsqu'il y a un changement de locuteur[27].

On distingue ainsi pour chaque enregistrement (1) les pauses interlocuteurs (*gaps*), (2) la parole du ou des seconds locuteurs (« L_2 »), (3) la parole du locuteur principal (« L_1 ») et (4) la parole produite en chevauchement par deux locuteurs ou plus (« L_n+L_n »).

	LOCAS-F			C-HUMOUR		
	(min)	(syll)	(lex)	(min)	(syll)	(lex)
Pauses interlocuteurs (*gap*)	11.02	1278	1278	2.85	365	365
Parole de L_2 (uniquement)	117.27	1891	1887	27.14	694	694
Parole de L_1 (uniquement)	206.59	58338	45234	81.58	25704	21665
Parole chevauchée (L_n+L_n)	15.67	3632	3275	3.86	365	992
Total	**350.5**	**651369**	**51674**	**115.4**	**27897**	**23716**

Tableau 3.7 – Catégorisation interactionnelle des segments temporels, syllabiques et lexicaux.

Cette première catégorisation permet d'isoler les segments où le locuteur principal est propriétaire de la parole (chevauchée ou non chevauchée). Parmi ceux-ci, on distingue (Tableau 3.8) la parole **Acoustiquement Analysable** (AA), exempte de chevauchement d'un ou de plusieurs second(s) locuteur(s), de bruits paralinguistiques (e.g. toux, rires) ou de bruits extralinguistiques (e.g. microphone, porte qui claque).

27. Voir Annexe C, tire `speaker`.

		LOCAS-F			C-HUMOUR		
		(min)	(syll)	(lex)	(min)	(syll)	(lex)
Parole de L1 (AA)	pause	40.35	4988	4988	10.08	1471	1471
	articulé	160.02	52287		68.75	23686	
	Total	200.97	57275		78.83	25157	
Parole de L1	AA	200.97	57275		78.83	25157	
	non AA	16.80	4219		6.61	1681	
	Total	217.77	61494	48034	85.44	26838	22657

TABLEAU 3.8 – Parole des locuteurs principaux : segments temporels, syllabiques et lexicaux.

Ainsi, la durée totale du temps de parole des locuteurs principaux sur les deux corpus est de 303,21 minutes (5h), ce qui représente 70691 mots (88332 syllabes). L'exclusion des segments non-AA réduit le temps de parole à 280 minutes (4.66h). La parole non articulée de L1 indique le temps de pauses intralocuteurs (50.43 min soit 6459 pauses), ce qui augmente la parole articulée et acoustiquement analysable à 75973 syllabes. Seule la parole acoustiquement analysable sert de base à l'analyse acoustique et prosodique des (dis)fluences.

3.2 Description quantitative des unités prosodiques et syntaxiques

L'organisation prosodique des (dis)fluences est analysée au regard de l'organisation prosodique et syntaxique du discours[28]. Le Tableau 3.9 présente le **nombre d'unités prosodiques et/ou syntaxiques** annotées dans les corpus, ordonnées de la macro-unité la plus englobante (l'unité de base du discours) à la micro-unité la plus réduite (le groupe accentuel accentuable).

Abréviation	**LOCAS-F**	**C-Humour**
	(n)	**(n)**
BDU	2912	
IPU	4102	1469
UR	12097	
SEQ	16966	
ga-1	20181	10847
ga-0	20594	20626

TABLEAU 3.9 – Répartition des unités prosodiques et/ou syntaxiques annotées dans les corpus.

On décompte plus de 40000 groupes accentuels accentuables dans les corpus, rassemblés en 31000 groupes accentuels accentués. Les séquences fonctionnelles de LOCAS-F (près de 17000) sont réunies au sein de 12000 unités de rection. Enfin, le corpus rassemble près

28. Pour une définition des unités prosodiques et syntaxiques, voir section 2.3 de ce chapitre.

de 3000 BDU, soit un peu moins que le nombre d'unités séparées par des pauses (41000). Cette multitude de segmentations prosodiques et syntaxiques croisée à la quantité d'unités annotées dans les corpus est bénéfique à l'analyse des (dis)fluences. Elle permet d'une part d'envisager des **paramètres de contextualisation différents** des (dis)fluences (et de dresser un profil de chaque (dis)fluence sous plusieurs angles d'approches) et d'autre part, de procéder à l'analyse statistique sans que la taille d'échantillon ne fasse défaut.

3.3 Description quantitative des (dis)fluences

Cette sous-section présente quantitativement les données extraites des corpus à deux niveaux : **(dis)fluences lexicales** et **structures de (dis)fluences**[29]. Les données fournissent un aperçu global de la distribution de ces phénomènes au sein des corpus ainsi que de l'impact de l'unité envisagée sur la description des phénomènes de (dis)fluences.

Premièrement, un décompte des phénomènes annotés au niveau lexical sur l'ensemble de la parole des locuteurs principaux des phénomènes (AA et non AA) est donné par le Tableau 3.10[30]. Excepté la pause silencieuse, la (dis)fluence la plus courante au niveau lexical est le DM, suivi de près par les RI, FP et RM. Les SP, TR, IP, SM, DEL et ET, représentent dans les deux corpus moins de 5% des mots annotés. Au niveau méthodologique, différentes mesures du taux de (dis)fluences sont envisagées. Si nous avons exposé ci-avant la proportion de mots porteurs d'une (dis)fluence par corpus, nous envisageons dans certaines sections la description de ces phénomènes en fonction de leur environnement syntagmatique, notamment la **complexité de la structure de (dis)fluence(s)**[31]. La Figure 3.11 donne un aperçu de la répartition des (dis)fluences cooccurrentes au sein des situations de parole dans les corpus LOCAS-F et C-Humour. L'annotation lexicale regroupée par structures de (dis)fluences livre un premier aperçu des possibilités offertes par nos annotations. Ces observations justifient également que l'**unité d'analyse** de la (dis)fluence varie en fonction de l'objet d'étude envisagé. Chaque chapitre se focalisant sur une (dis)fluence et ses patrons de cooccurrences, nous ne nous attarderons pas davantage sur les détails de cette distribution, bien qu'on puisse déjà constater que la répartition des types de (dis)fluences au sein des situations ne semble pas arbitraire (Arnold & Tanenhaus 2011 ; Bortfeld et al. 2001).

29. Les principes sous-jacents à l'annotation des (dis)fluences à ces différents niveaux sont décrits en section 2 de ce chapitre.

30. Nous rappelons qu'un item lexical peut être porteur de plusieurs (dis)fluences (e.g. un mot porteur d'une troncation <TR> peut être également le point d'interruption d'une construction <FS>). Le décompte des (dis)fluences peut donc, en théorie, être supérieur au décompte de mots du corpus. Le nombre de segments annotés portant au moins deux tags est de 434 dans LOCAS-F et de 261 dans C-HUMOUR.

31. Il est ainsi possible d'envisager une analyse par mille mots, par minute, par unité prosodique, par unité syntaxique (en indice, nombre ou en durée) ou en fonction de leur cooccurrence avec d'autres (dis)fluences (e.g. « répétition avec substitution » fait émerger une analyse en collostructions, le sens des constructions en contextes et la manière dont celles-ci interagissent. Voir notamment Stefanowitsch & Gries (2003) et Gilquin (2013)). Voir chapitre 3 section 2.2.

	(Dis)fluence	LOCAS-F		C-HUMOUR	
		n	%	n	%
UP	Pause silencieuse	5078	10.78	1487	6.56
FP	Pause pleine	1071	2.23	355	1.57
DM	Marqueur de discours	3296	6.86	1972	8.7
DEL	Délétion	52	0.11	7	0.03
ET	Terme d'édition	43	0.09	33	0.15
FS	Faux-départ	146	0.3	98	0.43
TR	Troncation	519	1.08	168	0.74
RI	Répétition à l'identique	2804	4.84	1455	6.42
RM	Répétition avec modification	1027	2.14	546	2.41
SP	Substitution propositionnelle	612	1.27	367	1.62
SM	Substitution morphologique	362	0.75	160	0.71
IL	Insertion lexicale	132	0.27	81	0.36
IP	Insertion parenthétique	402	0.84	106	0.47
	Total annoté « (dis)fluent »	**15644**	**32.57**	**6835**	**30.18**
	Total non annoté « (dis)fluent »	**32772**	**68.23**	**16068**	**70.92**
	Total	**48416**	**101**	**22903**	**101**

TABLEAU 3.10 – Distribution des marques annotées dans les corpus.

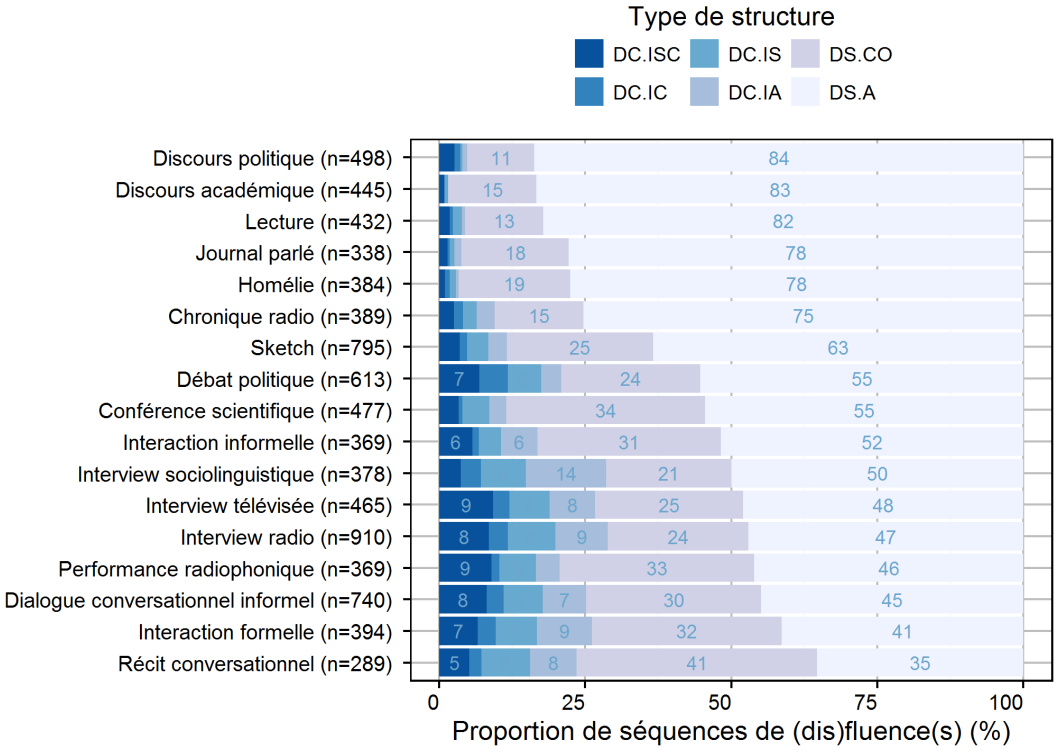


FIGURE 3.11 – Distribution des types de structures de (dis)fluence(s) par situation de parole.

4 Principes d'une pragmatique expérimentale

La structure de cette thèse reflète le cadre méthodologique de notre recherche. Chaque chapitre se focalise sur une (dis)fluence, combinant analyse des données de production (observation des corpus) et des analyses des données de perception (expériences perceptives). Chaque expérience perceptive fait appel à une méthodologie propre à la tâche de perception envisagée. Les sections suivantes présentent globalement la méthodologie de recherche. Elles explicitent les **implications théoriques et techniques** de l'association entre les données issues de **différentes sources** (4.1), et permettent de synthétiser les **liens entre théories** linguistiques, choix méthodologiques, données en production et données en perception (4.2). Cette section méthodologique répond ainsi aux questionnements théoriques du chapitre 2 section 1.4.1.

4.1 Complémentarité des études en production (sur corpus) et en perception (expérimentales)

> « A l'intérêt d'une indisciplinarité des objets d'étude s'ajoute celui d'une indisciplinarité des démarches employées. » (Loty 2005:252)

La problématique de notre travail de recherche mobilise des données de production et de perception. Méthodologiquement, nous utilisons les conclusions des analyses de corpus pour énoncer nos hypothèses concernant la **formulation d'attentes** perceptives d'un locuteur vis-à-vis d'une situation de parole.

En **linguistique cognitive**, une des hypothèses sous-jacente centrale postule que la structure de la langue **émerge de son usage** et que la compétence grammaticale est non modulaire, notamment en raison de la reconnaissance d'un continuum lexico-syntaxique (Janda 2010:15). Cependant, l'étude de la cognition à travers les données linguistiques *offline* reste indirecte et souffre souvent de l'application d'une méthodologie unique. Arppe et al. (2010:11) parlent de « **monisme méthodologique linguistique** ». Or, depuis une dizaine d'années, la linguistique cognitive sait que pour atteindre une connaissance compréhensive du langage, celle-ci a tout à gagner de la diversité des types de données, étant donné que chacune d'elles comporte ses limitations ou apporte ses types spécifiques de réponses (Schönefeld 2011:15-21).

Brièvement, certains auteurs reconnaissent la **complémentarité** de l'approche sur corpus et de l'approche expérimentale. Gilquin & Gries (2009:6) pointent ainsi premièrement l'**absence de dichotomie stricte** entre données de corpus et d'expérience. En soi, les deux constituent des corpus linguistiques dont la production et collecte est plus ou moins naturelle. A ce titre, les conclusions des analyses de corpus peuvent **converger** avec les conclusions issues de l'expérience, l'un servant ainsi de **validateur** à l'autre. D'autres fois, le corpus sert souvent simplement de base de données à la **composition des stimuli**. C'est le cas pour Duez (1985), qui teste expérimentalement les hypothèses

de catégorisation (motivées par l'observation de corpus) des *euh structurant* /vs/ *d'hésitation* en utilisant des stimuli issus de trois styles de parole (i.e. ici l'expérience sert de validateur des observations sur corpus, le corpus sert de base à la composition des stimuli).

Toutefois, assurer la **comparabilité** et l'**opérationnalisation** des hypothèses et des questions de recherche demeure l'enjeu capital d'une approche multiméthodologique (Arppe et al. 2010:14). On opposera ainsi souvent les deux techniques par leur nature *online* /vs/ *offline* (Grondelaers, Brysbaert, Speelman & Geeraerts 2002:19-20 ; Gilquin & Gries 2009:19 ; Arppe et al. 2010:2) et par conséquent par la **nature des phénomènes** qu'elles sont le plus à même de mettre en valeur. La question de la **convergence des preuves scientifiques** est fondamentale. Les données de corpus autorisent par exemple des généralisations basées sur la ***fréquence relative***. Celles sur des données expérimentales élicitées permettent par contre certaines généralisations sur la ***saillance relative*** (Arppe et al. 2010:17)[32]. Les résultats quantitatifs sur corpus sont une méthode objective (par rapport aux analyses qualitatives sur corpus) mais livrent peu de résultats « en finesse ». La qualité première des résultats élicités basés sur l'expérimentation est leur caractère ciblé. Ils dépendent toutefois foncièrement d'hypothèses sous-jacentes formulées à propos de la **relation entre un comportement observé et l'activité mentale**. C'est excellent pour récolter l'information sur le traitement de la parole et la saillance conceptuelle. Ainsi, avant de parler de résultats convergents ou divergents, il est important de **s'accorder sur ce que ces résultats représentent** (i.e. *une mesure de latence mesure-t-elle vraiment le niveau d'attention du sujet ? La dilatation des pupilles mesure-t-elle le niveau de concentration ou simplement le niveau d'activité d'un sujet ?*, etc.).

L'**approche multiméthodologique** permet d'approfondir pertinemment la réflexion sur la manière dont travaux empiriques et théoriques peuvent se nourrir, et de mieux comprendre la relation entre ces deux types de données.

4.2 Méthodologie au service de la convergence des données : schéma récapitulatif

La structure de cette thèse met en valeur de manière empirique l'intérêt de la complémentarité des approches basées sur l'usage, en production et en perception. Les annotations des (dis)fluences sur corpus servent notamment tour à tour de base à la formulation d'hypothèses cognitives *online* et à la sélection de stimuli d'expérience. Les résultats d'expériences peuvent quant à eux servir à générer de nouvelles annotations. Ils peuvent aussi servir à la création de nouveaux corpus d'évaluation de la perception d'une production comme fluente ou disfluente. Ainsi, tout en évitant un **syncrétisme méthodologique**, les deux approches se nourrissent perpétuellement dans ce travail. La Figure 3.12 schématise les différentes dynamiques sous-jacentes au cadre méthodologique. On y distingue :

32. Voir chapitre 2 section 1.4.

1. Les **étapes théoriques** en amont et en aval se basent sur/ou enrichissent le savoir scientifique (*en mauve*). En amont, ces théories sont plus ou moins justifiées empiriquement. En aval, elles s'enrichissent notamment par la convergence des résultats issus des données en production et en perception.
2. Les **étapes méthodologiques** (*en jaune*) reflètent un travail conceptuel comme technique d'annotation ou d'extraction des annotations et des informations linguistiques.
3. Les étapes issues de l'observation des **données de production** (*en bleu*), représentent majoritairement des étapes d'extraction de données ou d'extraction d'hypothèses suite à l'observation des corpus de parole.
4. Les étapes issues de l'observation des **données de perception** (*en vert*), peuvent intervenir en amont lors de l'annotation perceptive des données (*online* et *offline*), ou lors de l'examen perceptif des données sur des sujets experts et non experts.

En résumé, le point de départ de ce travail est le corpus annoté. Celui-ci résulte de l'application du protocole d'annotation de l'ARC Fluence-Disfluence, basé sur des théories linguistiques énoncées au chapitre 1, appliqué sur les **données orales brutes alignées**.

Après l'évaluation des protocoles (par accord interannotateurs) issue d'expériences d'annotations *offline* et *online*, les informations syntagmatiques, acoustiques, syntaxiques et prosodiques en production sont reliées aux scores interannotateurs en perception. On peut ainsi formaliser des **tendances d'annotation et de perception** et extraire les premières relations entre fréquences d'occurrences et tendances perceptives *offline* (dans l'ensemble, par situations, par trait situationnel, etc.). On peut également formuler les premières **hypothèses reliant ces deux types de données**.

En vertu des hypothèses sur ce que constitue une (dis)fluence plus ou moins fluente, l'étape suivante est de **sélectionner les stimuli d'expérience à partir du corpus** et de les tester par des expériences perceptives *online* et/ou *offline*.

Le résultat des études de perception permet d'obtenir un score de (dis)fluence par segment et de réévaluer l'annotation des marqueurs et des théories linguistiques (e.g. *Qu'est-ce qui est fluent ? En vertu de quels facteurs ?*). Ces scores laissent également envisager d'autres applications telles que la visualisation du profil de *dis*fluence par échantillon et un retour au corpus d'origine afin de créer une **nouvelle annotation de pondération** de chaque chaîne de (dis)fluence(s) présente au sein du corpus[33].

33. Nous y revenons au chapitre 7.

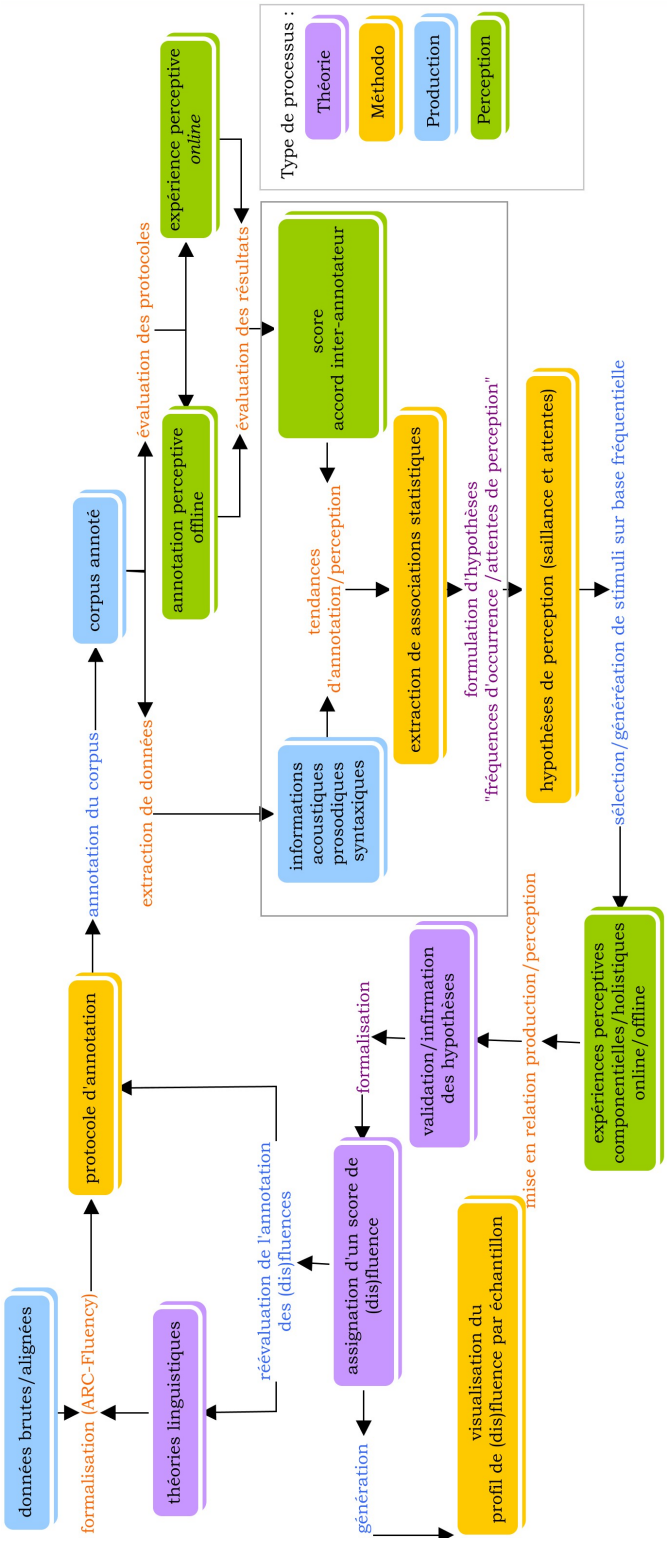


FIGURE 3.12 – Schéma des dynamiques sous-jacentes au cadre méthodologique.

Chapitre 4

Evaluation de la (dis)fluence des pauses silencieuses

> "Rest in the pause between breaths.
> Pause in the rest between thoughts.
> Bask in the space between words.
> Stop in the stillness of a calm lake and listen."
>
> (Rappaport 2004:57)

L'analyse rythmique de la parole peut envisager la segmentation du signal acoustique en **composantes mesurables** faites de l'alternance de séquences d'**intervalles articulés** et de **pauses**. Les pauses s'acquittent de multiples fonctions, des plus élémentaires (e.g. respiration, tenue liée à l'articulation de plosives) à celles davantage liées au flux discursif (e.g. segmentation syntaxique, prosodique, sémantique, stylistique et leurs interactions) (Duez 1982 ; Simon, Auchlin, Goldman & Christodoulides 2013). Nombre de ces catégorisations fonctionnelles prétendent que la durée des pauses varie selon leur fonction [1].

Nous avons montré au chapitre 2 que les études antérieures ont formalisé le **contexte d'actualisation** des pauses en leur associant des **propriétés distributionnelles** et **fonctionnelles** sans pour autant qu'il y ait consensus : une « pause non grammaticale » n'est pas toujours plus courte qu'une « pause grammaticale ». De plus, une pause non grammaticale n'est pas non plus systématiquement interprétable comme une « pause d'hésitation ». En somme, on ne peut prétendre à partir d'une analyse monofactorielle associer à une pause une seule et unique fonction. Dans l'objectif de positionner les pauses sur un continuum de fluence-disfluence, nous tentons cependant de repérer des (ir)régularités distributionnelles, à partir d'une diversité de paramètres linguistiques et situationnels congruents ou divergents.

1. Voir chapitre 2 section 2.1.

Ce chapitre propose une **modélisation multifactorielle** de « catégorisation » des pauses, notamment en fonction de leur durée et de leur contexte d'actualisation. Après être revenue sur les choix méthodologiques de l'analyse des pauses en contexte (section 1), nous formulons les objectifs et formalisons les hypothèses de ce chapitre (section 2). Enfin, nous évaluons le croisement de (1) la variation situationnelle des distributions des durées pausales (section 3.1), (2) des chaînes de cooccurrences de (dis)fluence(s) (section 3.2), (3) de la localisation syntaxique des pauses (section 3.3) et (4) des contours intonatifs précédant la pause (section 3.4). Nous posons ainsi les prémisses d'un manuel de « variance et invariance » des durées pausales, afin d'établir leur(s) position(s) possible(s) sur le **continuum fluence-disfluence**.

1 Méthodologie de catégorisation des pauses

Les méthodologies de mesures et de transformations (notamment logarithmiques) des durées conditionnent nombre d'analyses *post hoc* (e.g. catégorisation des pauses, analyse interactionnelle en tours de parole, définition d'unités intonatives, segmentation en unités séparées par des pauses, etc.) (Campione & Véronis 2002:199 ; Peshkov, Prévot & Bertrand 2013:95). Ces choix méthodologiques ne sont donc en rien anodins ou a-théoriques : ils doivent être envisagés de manière critique et avec précaution.

Nous revenons brièvement sur les théories antérieures de **détection** et de **classification** des pauses silencieuses (section 1.1) en offrant notamment une synthèse des **types de seuils, de mesures et de transformations** mathématiques applicables (section 1.2). Nous présentons les particularités méthodologiques de la **classification hiérarchique** (*clustering*) appliquée ici aux durées pausales des corpus LOCAS-F et C-Humour (section 1.3).

Nos résultats mettent en avant l'**impact des transformations** et des seuils sur la classification des données, mais aussi l'intérêt de questionner la **granularité des niveaux d'analyses** en fonction de la question de recherche, notamment à travers la mise en évidence de la variation inter et intrasituationnelle et des **spécificités inhérentes à chaque locuteur et/ou situation**. Ainsi, si cette réflexion sur la méthodologie descriptive contextuelle des durées pausales et sur le **rejet de seuils** de durées minimaux et/ou maximaux s'illustre dans l'étude intrinsèque et contextuelle des phénomènes de (dis)fluences, nous pensons que ses prolongements n'y sont pas restreints.

1.1 Théories de détection et de classification des (seuils de) pauses

L'analyse statistique des pauses (pleines et silencieuses) présente un nombre de défis méthodologiques dus au **caractère typiquement asymétrique** de leur distribution. Cette particularité distributionnelle entrave l'application de méthodes de catégorisations reposant sur l'hypothèse de normalité (Oehmen, Kirsner & Fay 2010). Concrètement, dans le cas d'une **distribution non normale**, la moyenne de la durée des pauses ne peut servir de point de départ valide à la segmentation du discours. La Figure 4.1 issue de Campione & Véronis (2002:200) offre un exemple de la distribution des durées pausales dans le cadre de la parole spontanée.

Certains chercheurs dépassent (ou contournent) cet écueil en appliquant une transformation logarithmique des données et/ou en appliquant des seuils minimum et maximum de durée. Les études divergent cependant quant à la **(log)normalité** de leur distribution. Si les données de de Jong & Bosker (2013:19) respectent tout à fait les conditions de lognormalité et n'attestent donc pas de **distributions bigaussiennes**, la majorité des distributions observées dans la littérature reflètent une combinaison variant de **deux à trois courbes normales** (Figure 4.2). Il semblerait que la distribution des pauses soit

en effet majoritairement **(log)multimodale** (Campione & Véronis 2002:200 ; Demol, Verhelst & Verhoeve 2007:990 ; Goldman, François, Roekhaut & Simon 2010:162 ; Rosen 2005:411-412) [2].

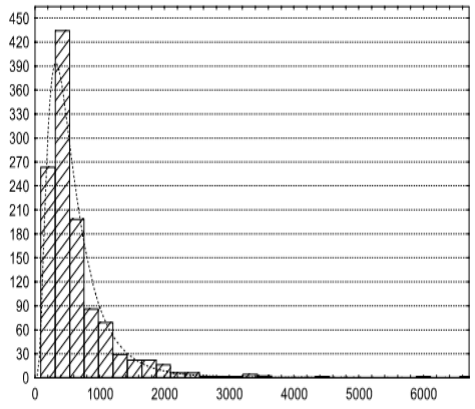


Figure 4.1 – Distribution (asymétrique) de la durée des pauses silencieuses du corpus Eurom en situation spontanée

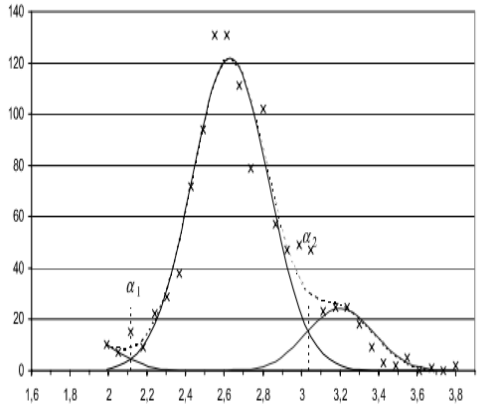


Figure 4.2 – Distribution de la durée des pauses silencieuses du corpus Eurom en situation spontanée après transformation logarithmique.

Certains chercheurs postulent que chaque composante normale permettrait de distinguer des « pauses brèves » (et/ou « moyennes ») et des « pauses longues » (Demol, Verhelst & Verhoeve 2007:992-999 ; Kirsner, Dunn & Hird 2003:14 ; Little et al. 2013:135). Kirsner, Dunn & Hird (2005) observent par exemple une **distribution bimodale** de durées de pauses silencieuses sur leur corpus anglophone. Par extension, les auteurs attribuent à la première composante l'appellation de « pauses courtes », qui correspondrait à des pauses liées aux **processus d'articulation**. La seconde composante correspondrait davantage à des « pauses moyennes » ou « pauses longues ». On associe ces dernières à des processus cognitifs « plus profonds », telle que la **segmentation discursive**.

En analysant un corpus de 5.5 heures de parole lue et spontanée dans cinq langues, Campione & Véronis (2002) rapportent une **distribution (log)bimodale** mais aussi **(log)trimodale** des pauses. Les auteurs distinguent alors par des seuils précis les « pauses brèves » (inférieures à 200 ms.), les « pauses moyennes » (de 200 à 1000 ms.) et les « pauses longues » (supérieures à 1000 ms.). Ces dernières n'ont été observées que dans le cas de la parole spontanée.

Ces observations rejoignent les conclusions statistiques de Demol, Verhelst & Verhoeve (2007) portant sur un corpus de 4 heures de parole dans trois pratiques communicatives et six langues. Ils soutiennent eux aussi que la durée logarithmique des pauses peut être évaluée comme une distribution bigaussienne, et ce dans le cas de **débit de parole** rapide comme lent, à travers toutes les langues étudiées (i.e. néerlandais, anglais, français, italien, roumain et espagnol).

2. Les 6000 pauses silencieuses représentées proviennent du corpus multilingue Eurom (5.5 heures de parole en anglais, français, allemand, italien, roumain, espagnol).

Goldman, François, Roekhaut & Simon (2010) rapportent également qu'une distribution multimodale peut être modélisée et décomposée en un **mixte de courbes lognormales**. Les auteurs avancent que l'aspect unimodal, bimodal voire trimodal des distributions pausales **dépend des situations de parole** observées. S'appuyant sur un corpus francophone de 40 minutes rassemblant quatre pratiques communicatives (lecture, narration conversationnelle, journal parlé télévisé, et conférence universitaire), les auteurs observent, en groupant les communications par pratiques communicatives, que la distribution des durées pausales est **majoritairement bimodale**, à l'exception de la situation du journal parlé télévisé qui, trimodale, serait la seule à témoigner de « pauses longues » (Goldman, François, Roekhaut & Simon 2010:164)[3].

Que les distributions soient bi- ou trimodales, les auteurs s'accordent sur le fait que ce type de distribution invalide les analyses se basant sur une unique **mesure de tendance centrale**[4] (e.g. moyenne de la durée pausale). Par contre, chaque (sous)catégorie de pauses pourrait être analysée comme une distribution lognormale, et inviterait donc à utiliser la **durée moyenne** comme unité d'analyse des pauses « courtes » et des pauses « longues ». Selon nous, cette catégorisation binaire ou ternaire des pauses résultant d'une transformation logarithmique doit être **motivée statistiquement** mais aussi **linguistiquement** et pragmatiquement. Or plusieurs objections surgissent :

- **Méthodologiquement**, l'observation et la réunion du nombre de catégories pausales par situation de parole, basées sur la **transformation logarithmique des durées absolues** pourraient être le résultat d'un simple artefact méthodologique gommant la variabilité interindividuelle. La définition en termes de situation de parole est-elle pertinente au vu de la particularité de chaque corpus, résiste-t-elle à la variation intrasituationnelle (i.e. interlocuteurs) ?

- Au niveau de la **production**, la transformation logarithmique ne serait-elle pas également créatrice d'un **artefact** de bi- ou trimodalité ? En d'autres termes, retrouve-t-on cette observation dans le cas d'**autres transformations** mathématiques ?

- Au niveau de la **perception**, une pause absolue définie « statistiquement » comme « courte » est-elle **interprétée** comme telle par un auditeur en contexte ? En effet, étant donné l'importance des phénomènes phonémiques, prosodiques (contours intonatifs) et syntaxiques sur la perception du débit de parole, et donc de la durée pausale[5], il semble peu écologique qu'une durée absolue, définie comme « courte » grâce à une transformation logarithmique, résonne de manière indifférente comme « courte » en tout contexte et pour tous les auditeurs.

3. Soulignons donc que chez ces derniers c'est une situation préparée qui laisse s'actualiser des « pauses longues », alors qu'il s'agissait d'une situation spontanée chez Campione & Véronis (2002).

4. La multimodalité des distributions conditionne l'application de tests statistiques paramétriques /vs/ non-paramétriques.

5. Voir notamment Zellner (1994) et plus récemment Obin, Rodet & Lacheret (2008) pour une revue des méthodologies de mesures de débits relatifs ainsi que notre proposition section 1.2.

Après être revenue brièvement sur les types de seuils et de mesures de transformations, nous appuyons ces objections par l'observation empirique des données de corpus. Nous proposons ensuite une méthodologie de classification des pauses en *clustering* **prenant en compte le débit relatif et le locuteur**. Cette proposition de **normalisation** constitue la base des analyses d'**évaluation du degré de (dis)fluence des pauses silencieuses** dans nos corpus.

1.2 Types de seuils, de mesures et de transformations

Faut-il exclure certaines pauses parce qu'elles sont « **trop brèves** » ou « **trop longues** » ? Grosjean & Deschamps (1975) intègrent les pauses silencieuses dépassant le seuil minimum de 250 ms. ; Duez (1985) établit un seuil adapté au débit de chaque locuteur à partir de la durée moyenne des occlusives intervocaliques, et ce seuil varie entre 180 et 250 ms. ; Candea (2000a) retient toutes les pauses silencieuses à partir de 200 ms. sans limite supérieure ; enfin, Lacheret & Victorri (2002) utilisent eux un seuil minimum de 300 ms. pour annoter les pauses silencieuses dans un corpus radiophonique, etc.

Dans leur revue de la littérature, Campione & Véronis (2002:199) révèlent la **diversité des seuils** de durées possibles : seuils minimum et maximum, seuils absolus /vs/ relatifs – par locuteur ou par situation – prenant en compte ou non un modèle de durées phonémiques et/ou syllabiques intrinsèques, dépendant lui-même d'un éventuel modèle prosodique (d'allongement syllabique), etc. Or, la détection et la classification catégorielle des pauses en fonction de leur durée (e.g. brève, moyenne, longue) dépend des seuils critiques fixés par l'analyste ainsi que des types de mesures appliqués. Enfin, ces méthodologies de normalisation influencent elles-mêmes la prise en compte de la potentielle **variabilité inter et intralocuteurs** des distributions pausales.

La valeur du seuil minimum varie de 180 ms. à 300 ms. selon les études et Campione & Véronis (2002:202) montrent de manière intéressante que la modification du seuil minimum de durée porte à conséquence sur les analyses subséquentes. Ainsi, ces auteurs révèlent que la durée moyenne des pauses est **plus brève en parole spontanée qu'en parole lue** si l'on n'utilise aucun seuil de durée ; que la proportion s'équilibre avec un seuil minimum de 200 ms. et un seuil maximum de 2000 ms. ; et que la durée moyenne des pauses devient **plus brève en parole lue qu'en parole spontanée** si l'on ne conserve que le seuil maximum de 2000 ms. En d'autres mots, l'utilisation de seuils de durée peut modifier sensiblement un jeu de données (Hieke, Kowal & O'Connell 1983:202 ; Campione & Véronis 2002:201-202) et rendre difficile les comparaisons avec d'autres études.

La définition de seuils limites peut potentiellement **exclure des informations** importantes d'un jeu de données (Hieke, Kowal & O'Connell 1983:202 ; Campione & Véronis 2002:201-202). Cette exclusion non motivée met non seulement en péril la

comparabilité des résultats de nombre d'études – mais éventuellement leur validité – les catégories de pauses découlant des **seuils de durée minimum et maximum fixés par l'analyste**. Les seuils minimum de détection de pauses varient de 100 à 300 ms. selon les études, le plus commun siégeant à 200 ms. (Heldner & Edlund 2010:557) [6]. Certains auteurs argumenteraient pour un seuil minimum basé sur le seuil minimum de perception (e.g. 50-150 ms.). Encore une fois, il s'agit alors d'être assuré des résultats d'expériences antérieures sur la perception de la pause – tous contextes confondus.

En outre, ces seuils ne rendent pas nécessairement compte de la **manière dont les pauses sont perçues**. Duez, qui établit un **seuil relatif** au débit de chaque locuteur, a ainsi prouvé que la perception d'une pause peut même survenir en l'absence de toute interruption de la phonation (cf. la notion de « **pause subjective** » dans Duez (1993)).

En effet, le **contexte d'actualisation** est essentiel à la perception d'une pause. Les recherches en perception et en psycholinguistique ont établi que le seuil de perception des pauses ne correspond pas nécessairement à la matérialité physique d'une pause. Le système sensoriel humain montre que le **seuil de perception** d'une pause est plus élevé que celui de la pause physique et qu'il est également modulé en fonction des **segments la précédant** (Zellner 1994:43). L'adoption d'un seuil minimum en durée absolue fixé à priori, bien que séduisante, ne semble donc pas totalement motivée surtout dans le cadre d'études perceptives [7].

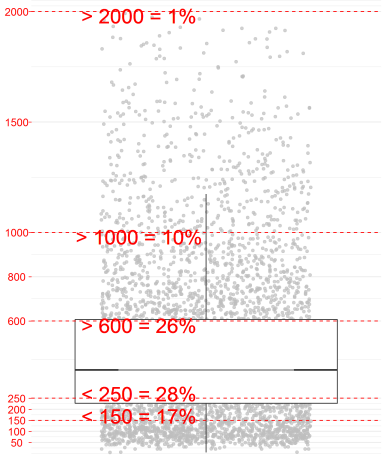


Figure 4.3 – Diagramme en boîte à moustaches de la distribution des durées pausales (LOCAS-F), avec indications de la proportion des données concernées en fonction des seuils communément utilisés (ms).

Nombre d'auteurs adoptent alors un **seuil adaptatif** – minimum, maximum, et/ou catégoriel. Celui-ci est en général non seulement davantage motivé, mais surtout plus **sensible à la variabilité** inter et intralocuteurs. Chez de Jong & Bosker (2013), le seuil est motivé en fonction du **débit d'articulation** de chaque locuteur. Chez Duez (1985), il dépend de la durée moyenne et de la déviation standard des **stops intervocaliques**.

6. Cette variabilité des seuils s'explique en partie par la prise en compte ou non de certaines pauses en fonction de leur position dans l'interaction (voir chapitre 3 section 3.1).

7. Voir section 1.3 pour la solution adoptée dans cette étude.

Certains auteurs choisissent également d'appliquer des **seuils individuels** de pauses pour établir leur classification par locuteur (de Jong & Bosker 2013:18-19). Enfin, tandis que d'autres défendent une catégorisation avertie en raison de cette bimodalité variable (Goldman, François, Roekhaut & Simon 2010:164), certains se refusent avec précaution encore à l'application de tout seuil (Little et al. 2013:195).

Nous retenons pour notre analyse les pauses **internes aux productions verbales** des locuteurs, en excluant les pauses occasionnant un changement de locuteur (entre deux tours de parole ou à l'occasion d'un chevauchement de parole). Sur un total de 8759 pauses silencieuses réunies dans le corpus LOCAS-F, on décompte **4969 pauses silencieuses** internes aux tours de parole.

Seuil inférieur (ms)	Fréquence (n)	Proportion cumulée (%)
50	116	2,24
100	522	10,07
150	899	17,36
200	1163	22,46
250	1467	28,33
Total	4845	100

Tableau 4.1 – Distribution des durées des pauses en fonction des seuils inférieurs dans LOCAS-F.

Seuil supérieur (ms)	Fréquence (n)	Proportion cumulée (%)
600	1353	26,12
800	816	15,76
1000	517	9,98
2000	55	1,062
3000	17	0,33
Total	4845	100

Tableau 4.2 – Distribution des durées des pauses en fonction des seuils supérieurs dans LOCAS-F.

On observe dans le diagramme en boîte à moustaches présenté à la Figure 4.3 la distribution des durées des pauses, en indiquant les seuils couramment utilisés. Ainsi, **50% des pauses ont une durée comprise entre 250 et 600 ms.** (rectangle intérieur du diagramme). Si l'on exclut les pauses d'une durée supérieure à 1000 ms., cela concerne **9,98% des pauses** ; si l'on exclut celles supérieures à 2000 ms., cela concerne 1,06% des pauses. À l'autre extrémité du diagramme, on observe que **28,33%** des pauses ont une durée **inférieure à 250 ms.** et que 17,36% ont une durée inférieure à 150 ms. Ces données sont exprimées de manière plus détaillée dans les Tableaux 4.1 et 4.2 présentant les fréquences et les proportions cumulées des pauses les plus brèves et les plus longues.

Sur la base de ces données, nous avons décidé d'**exclure un minimum de pauses** (moins de 6% des données). Les pauses concernées correspondent aux critères purement mathématiques présentés dans l'encadré, sans seuils définis par des à priori linguistiques. Ainsi, le nombre de pauses analysées pour LOCAS-F s'élève à **4845 pauses silencieuses issues de LOCAS-F** [8].

8. Le même procédé à été utilisé pour les pauses de C-Humour. Pour rappel, ce dernier n'est pas annoté en contours intonatifs ou frontière syntaxique, il n'est donc pas mobilisé dans les sections 3.3 et 3.4.

Exemple de critère d'exclusion sur LOCAS-F :

1. **Durée absolue** : 7 pauses (0,4% des données) supérieures à 5 sec. ou inférieures à 20 ms. ont été exclues car elles peuvent être attribuées à des erreurs d'annotation.
2. **Logarithme de base 10** : 117 pauses ont été exclues car leur valeur était 1.5 fois inférieure ou supérieure à l'écart interquartile. Ces pauses représentent respectivement 1.79% et 3.4% des données (voir Tukey 1977).

1.3 Méthodologie de classification des pauses en *clustering*

La présente sous-partie porte sur les **pauses intratour** issues des corpus LOCAS-F et C-Humour. Une étude respectant la même méthodologie a été réalisée sur un corpus plus important et a été présentée dans le cadre de la collaboration Christodoulides, Grosman, Degand & Simon (2017)[9].

Si nous avons refusé d'appliquer des seuils définis à priori (e.g. pause supérieure à 200 ms. et inférieure à 1500 ms.), la modélisation de la distribution pausale nécessite le **nettoyage des données** considérées comme mathématiquement aberrantes. Sur les 7599 pauses intratour originales, en durées absolues 9 pauses **inférieures à 20 ms** et **71 supérieures à 5 sec.** ont été écartées. Sur les 7519 pauses restantes, 184 pauses ont été écartées en raison de leur dispersion par rapport à l'**écart interquartiles**[10]. En somme, les données portent sur **7335 pauses silencieuses** (68% issues de LOCAS-F, 32% de C-HUMOUR).

Sur base des questionnements méthodologiques antérieurs, nous envisageons une modélisation des pauses normalisant les durées de chaque pause silencieuse en fonction du **débit de parole local normalisé**. La mesure est fondée sur l'intégration des syllabes environnantes dans une fenêtre d'observation à court terme. La durée relative est ici définie comme **la longueur de la pause divisée par la moyenne arithmétique de la longueur des segments adjacents**[11], au sein d'une fenêtre de plus et de moins 5 segments (incluant les syllabes articulées comme les pauses). En analysant les pauses dans leur contexte immédiat, on propose une analyse de la distribution de la durée relative des pauses, qu'on suppose plus proches de la perception humaine que les durées absolues hors contexte.

9. Les résultats sont présentés dans G. Christodoulides, I. Grosman, L. Degand & A. C. Simon. 2017. « Methodologies for modelling silent pause length. Insights on individual and situational variation from a large-scale corpus study ». Louvain-la-Neuve, Belgique. Les analyses se basent sur un « metacorpus » d'une durée totale de 17.4 heures (186 895 mots simples). Celui-ci compile 276 enregistrements dans 31 pratiques communicatives. Il reprend des enregistrements de LOCAS-F (Martin, Degand & Simon 2014), C-Humour (Grosman 2016), le Driving Simulator Cognitive Load Corpus (Christodoulides 2016), C-Phonogenre (Pršir, Goldman & Auchlin 2014) et le corpus Rhapsodie (Lacheret et al. 2014).

10. Plus précisément, les donneés respectivement inférieures et supérieures de 1.5 fois à l'écart interquartile ($1.5 * 0.454 = 0.805$) de sorte que $\sigma_{neg} = 1.69$ et $\sigma_{pos} = 3.501$.

11. Le matériau analysé a été extrait en utilisant le logiciel de gestion d'analyse et de gestion de corpus Praaline (Christodoulides 2014).

Cette méthodologie rappelle au niveau théorique celle adoptée par Obin, Rodet & Lacheret (2008) : la durée des pauses et les durées relatives des syllabes sont toutes considérées **sans égard aux propriétés intrinsèques des syllabes environnantes**. Contrairement à ces derniers, nous rejetons l'hypothèse sous-jacente selon laquelle « le débit local est réinitialisé à chaque pause » (Obin, Rodet & Lacheret 2008:3), sous peine de raisonnement circulaire. Le **débit local** a été calculé avec une fenêtre de Hanning de 650 ms. (Pfitinger 1996). Ce modèle présente notamment l'avantage de réduire la dispersion des données.

Les panels de la Figure 4.4 montrent la distribution des durées pausales selon **quatre mesures** : la durée absolue en ms. (A), la durée absolue transformée en logarithme de base 10 (B), la durée relative de chaque pause silencieuse (C), et la transformation logarithmique de base 10 pour chaque durée relative de pauses silencieuses (D).

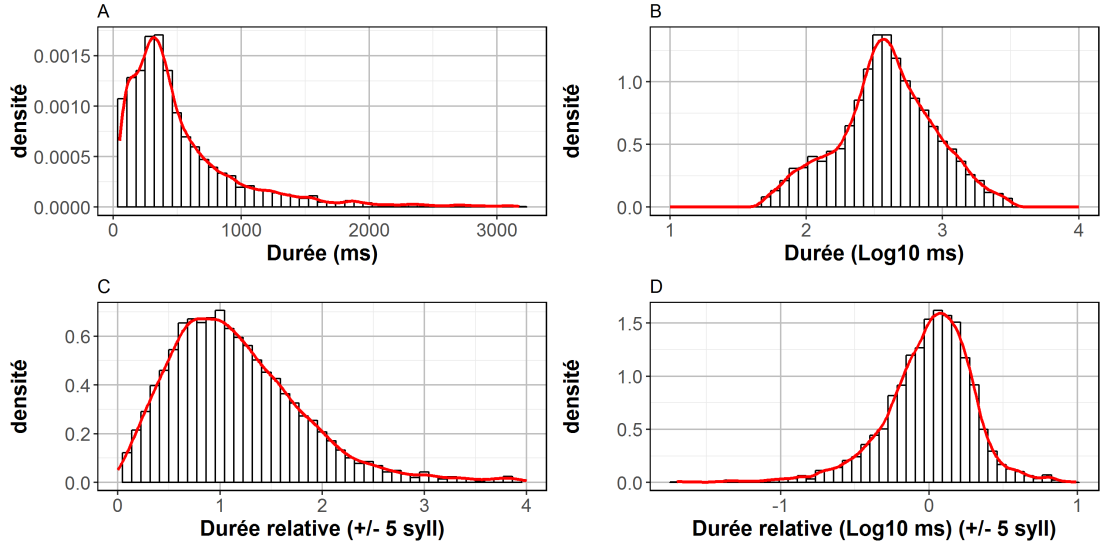


Figure 4.4 – Distribution de 4 mesures de distributions pausales : durée absolue, logarithme de base 10 des durées absolues, durée relative et logarithme de base 10 de la durée relative.

Le panel (A) confirme le **caractère asymétrique** de la distribution des pauses en durée absolue. Le panel (B) remet également en cause la lognormalité de la distribution de ces pauses **observée sur l'ensemble du corpus** [12] : la présence de deux pics laisse supposer une distribution effectivement multimodale actualisée comme un mixte de courbes gaussiennes. Cette distribution logarithmique semble bimodale, c'est-à-dire qu'elle pourrait être formée de deux composantes : un groupe réduit de pauses « brèves » et un groupe plus important de pauses « longues ». Ces résultats contrastent avec ceux de Rosen (2005) pour qui la distribution pausale s'approche suffisamment d'une distribution lognormale. Ils correspondent cependant à ceux de Goldman, François, Roekhaut & Simon (2010:162) et Demol, Verhelst & Verhoeve (2007:991). On remarque que les panels (C) et (D) **effacent cette prétendue bimodalité**, qui pourrait donc ne résulter

12. Un résultat similaire s'observe par type de situation ou par traits situationnels, voir section 3.1 de ce chapitre.

que d'un **artefact mathématique**. En d'autres termes, une mesure relative, prétendûment plus proche de la perception humaine, gomme la possible présence de deux groupes distincts de pauses (« courtes » /vs/ « longues ») – lorsqu'observée sur l'ensemble du corpus.

Nous avons cherché à établir quelles mesures sont les plus appropriées pour calculer la durée des pauses et s'il est possible d'établir empiriquement des **seuils de durée** pour distinguer différentes catégories de pauses (brèves, moyennes, longues) [13]. Pour cela, nous évaluons si ces données sont générales ou **spécifiques à la situation de communication ou au locuteur**. En effet, nous le verrons ci-après, ce type de distribution bimodale ne se retrouve pas systématiquement si l'on considère les durées des pauses produites par chaque locuteur individuellement ou dans les échantillons de parole correspondant à une situation communicative spécifique.

13. Nous avons développé un script permettant de définir pour chaque locuteur et/ou genre les seuils minimum et maximum des groupes de pauses. Ce script a par exemple servi à la segmentation en Unités Séparées par des Pauses (IPU, *Inter Pausal Units*). Par défaut de place, et de relation directe avec la problématique ci-présente, nous ne présentons pas ces résultats détaillés dans cette thèse. Nous nous focalisons davantage sur la nécessité d'appliquer de tels seuils de manière individuelle ou situationnelle en fonction de la variation interlocuteurs.

2 Objectifs et hypothèses

Dans les sections suivantes, nous présentons les analyses de fréquences et de durées des pauses silencieuses par situation de parole et par locuteur (section 3.1), selon leur localisation au sein d'une chaîne de (dis)fluences cooccurrentes à une pause (section 3.2), selon la localisation syntaxique de la pause (section 3.3), et enfin selon les contours intonatifs de la syllabe la précédant (section 3.4). Nous prenons ensuite en compte tous ces facteurs simultanément afin de modéliser le comportement des durées pausales dans leur ensemble (section 4.1). L'objectif est de faire émerger les **normes distributionnelles** et de comprendre comment se comporte le flux prosodico-syntaxique au regard des pauses silencieuses et des cooccurrences de (dis)fluences. Avant cela, nous tentons de répondre aux sous-questions suivantes :

1. Les profils de pauses doivent-ils être analysés en tenant compte des **caractéristiques de la situation de parole** (degré de préparation, degré d'orientation médiatique, etc.) ou **du locuteur** ?
2. La durée d'une pause silencieuse est-elle influencée par la **chaîne de (dis)fluence(s)** cooccurtente dans laquelle elle s'intègre ?
3. La durée d'une pause silencieuse est-elle influencée par le **contexte syntaxique** (e.g. unité de rection, séquences fonctionnelles, groupes accentuels) ?
4. La durée d'une pause silencieuse est-elle influencée par le **contour intonatif des segments** la précédant ?

L'état de l'art a permis de montrer que de nombreuses études empiriques existent sur la fréquence et la durée des pauses silencieuses en français. Souvent, les études basées sur les corpus les plus grands (Campione & Véronis 2002 ; Christodoulides, Grosman, Degand & Simon 2017) ne peuvent pas traiter la question de la localisation des pauses car les **annotations fines** font défaut. Au contraire, les corpus richement annotés sont en général de taille plus réduite, et illustrent souvent **une seule situation** de parole (interview radiophonique chez Grosjean & Deschamps (1975) ; parole en classe chez Candea (2000a) ; discours politique chez Duez (1991)) ou un nombre réduit de celles-ci (parole lue /vs/ spontanée chez Campione & Véronis (2002) ; quatre situations de parole chez Goldman, François, Roekhaut & Simon (2010)).

Le corpus LOCAS-F contient des échantillons représentant 14 situations de parole et a fait notamment l'objet d'une annotation des disfluences, des unités syntaxiques et des contours intonatifs. Ce corpus nous permet donc de tester une série d'hypothèses générales (que nous affinons dans chaque sous-partie de ce chapitre) :

- les pauses aux frontières syntaxiques sont **plus nombreuses** que les pauses dans l'environnement de marques d'hésitation ;
- la durée des pauses aux frontières syntaxiques est **supérieure** à la durée des pauses environnant les marques d'hésitation ;
- la durée d'une pause **augmente proportionnellement** au degré de la frontière syntaxique où elle est localisée ;

– les **contours intonatifs terminaux favorisent l'allongement des pauses subséquentes**, tandis que les contours continuatifs sont accompagnés de pauses de plus courte durée.

En outre :
– ces tendances pourront être nuancées ou précisées **en fonction des situations de parole** ;
– on cherchera à voir **quels facteurs** expliquent la variation de la durée des pauses environnant les (dis)fluences ;
– finalement, on proposera une modélisation multifactorielle de la durée pausale intégrant les critères situationnels, syntaxiques, intonatifs et liés aux (dis)fluences.

Ces analyses nous permettront en conclusion de revenir sur la distinction généralement admise entre pauses silencieuses **structurantes** (à fonction de segmentation) et **non structurantes** (à fonction notamment d'hésitation), afin de voir si cette distinction résiste à une analyse multifactorielle (section 4).

3 Analyses

3.1 Fréquence et durée des pauses en fonction des situations de parole

La durée médiane des pauses de LOCAS-F fluctue de **289 ms. à 518 ms.** à travers les 14 situations de parole. L'observation de la Figure 4.5 indique que la durée pausale varie en fonction des situations de parole, mais également **à l'intérieur** de celles-ci. Une analyse de la variance (ANOVA) sur les durées en logarithme de base 10 montre une variation significative des conditions ($F(13, 4831) = 34.05, p < .001$).

Un test post-hoc de Tukey DSH [14] montre que **7 groupes de pauses** peuvent être distingués **en fonction des situations de parole**. Il n'existe donc pas de différences significatives de la durée des pauses au sein de ces sept groupes [15]. Les pauses les plus longues se retrouvent au sein du groupe (1) homélie, discours politique, discours académique, tandis que les plus courtes sont observées au sein du groupe (7) conférence scientifique, débat politique et journal parlé.

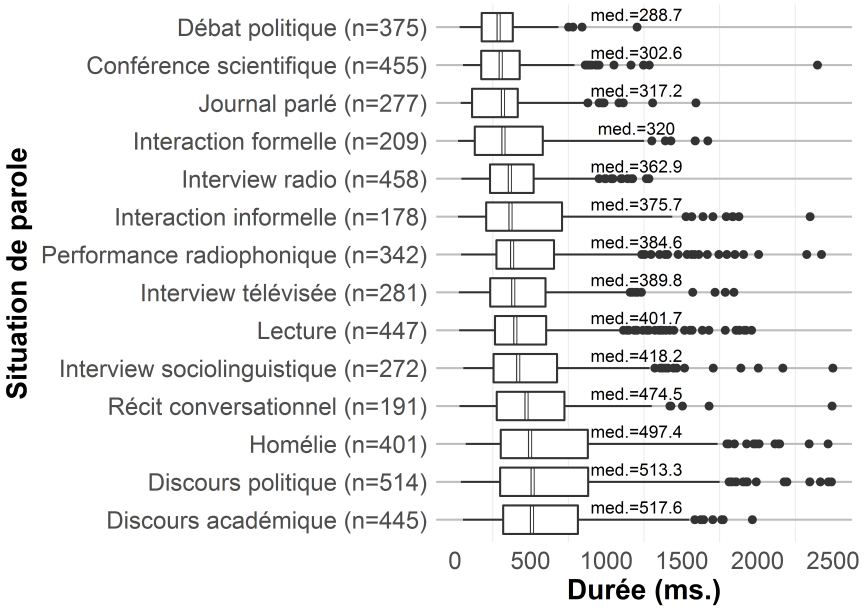


FIGURE 4.5 – Nombre d'occurrences et médiane de la durée pausale en fonction de la situation de parole dans LOCAS-F.

14. Le test statistique de Tukey DSH (différence significative honnête) effectue plusieurs comparaisons multiples en une seule étape. Il compare toutes les paires de moyennes possibles en se basant sur une distribution des rangs *studentisée*, similaire à la distribution selon la loi de Student utilisée dans le *test-t*.

15. Dans l'ordre des pauses les plus longues aux plus courtes nous avons donc, les parenthèses symbolisent les regroupements : [1] (Homélie > Discours politique > Discours académique) > [2] (Récit conversationnel > Performance radiophonique) > [3] (Interview sociolinguistique) > [4] (Lecture > Interaction informelle > Interview télévisée) > [5] (Interview radio) > [6] (Interaction formelle) > [7] (Conférence scientifique > Débat politique > Journal parlé).

Ces résultats nous incitent à appliquer un **regroupement hiérarchique** (*clustering*[16]), non pas au niveau du corpus entier (section 1.3) mais en fonction des situations de parole et des locuteurs. Nous tentons donc de répondre au premier questionnement : « Les profils de pauses doivent-ils être analysés en tenant compte des caractéristiques de la situation de parole (degré de préparation, degré d'orientation médiatique, etc.) et/ou du locuteur ? ».

La Figure 4.6 représente les distributions pausales de quatres situations de parole (durées logarithmiques transformées et durées relatives). En apparence, certaines situations illustrent une distribution unimodale, bimodale et d'autres trimodales. A première vue, ceci incite à penser que l'uni-, bi-, trimodalité serait **une question de « phonostyle »**.

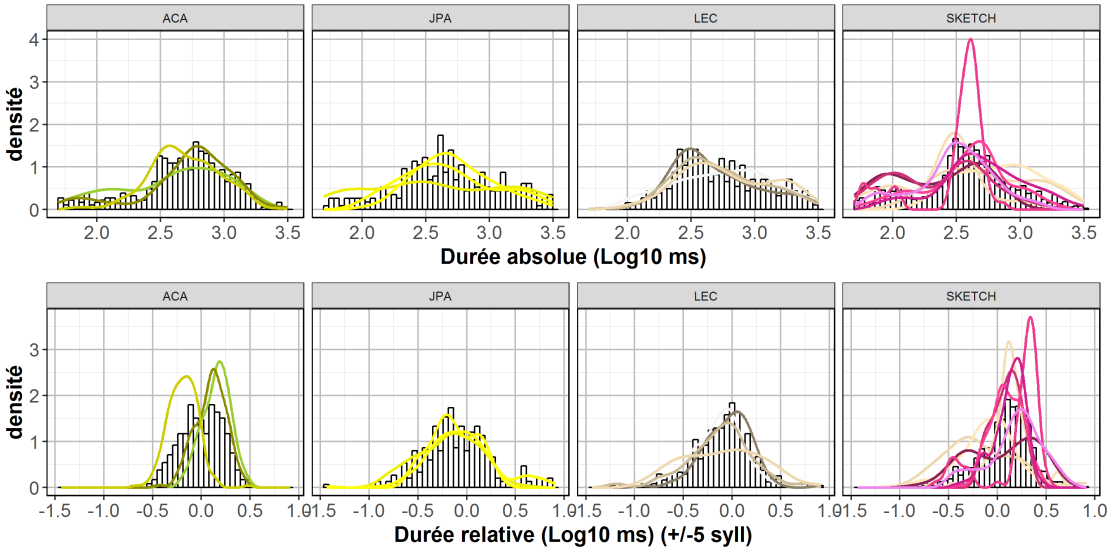


FIGURE 4.6 – Différence de la durée des distributions pausales par situation de parole en log_{10sec} (panel haut) et durée relative de log_{10sec}, (panel bas) avec mise en valeur des variations individuelles.

Notre étude corrobore ainsi en partie les résultats de GOLDMAN, FRANÇOIS, ROEKHAUT & SIMON (2010:162), DEMOL, VERHELST & VERHOEVE (2007) et LITTLE et al. (2013), tout en caractérisant un nombre plus important de données et en envisageant l'influence des traits situationnels. Cependant, comme le montrent les Figures 4.7 à 4.10, des **variations individuelles importantes** peuvent apparaître au sein d'une situation de parole (variations interlocuteurs figurées par les lignes). Nous présentons les distributions pour quatre situations de parole pertinentes. Dans la majorité des cas, la **bimodalité tend à disparaître** au niveau de la distribution pausale de chaque locuteur.

Les échantillons de discours politiques (Figure 4.7) montrent une distribution bimodale chez 3 locuteurs sur 5, et laissent entrevoir que les **durées moyennes des pauses varient d'un locuteur à l'autre**. Dans les journaux parlés radiophoniques (Figure 4.8), les distributions de chaque locuteur sont **plus homogènes** et également bimodales : ce genre parait donc relativement **standardisé**. Par contre, les récits conversationnels sont

16. Nous utilisons le package `mclust` (version 5.2.3) (FRALEY 2008) sous R (*R Core Team* 2016).

caractérisés par une distribution **plutôt unimodale** et relativement homogène entre les locuteurs (Figure 4.9). Enfin, les homélies sont intéressantes car les pauses de l'ensemble des locuteurs se décrivent plutôt selon une distribution unimodale (Figure 4.10).

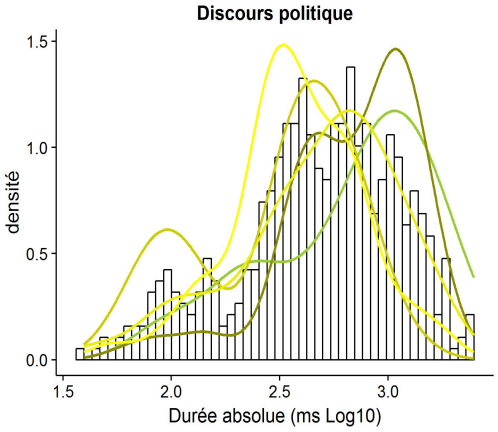


FIGURE 4.7 – Distribution des durées sur 375 pauses et 4 locuteurs pour la situation du discours politique (LOCAS-F).

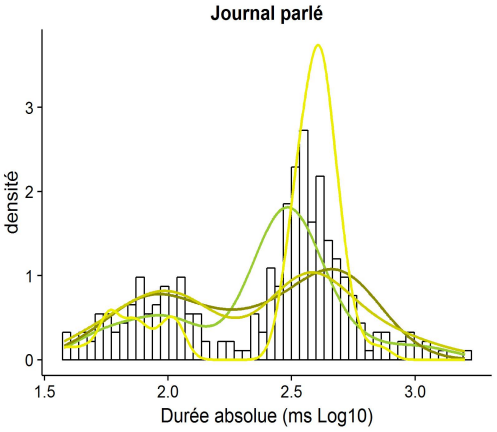


FIGURE 4.8 – Distribution des durées sur 277 pauses et 4 locuteurs pour la situation du journal parlé (LOCAS-F).

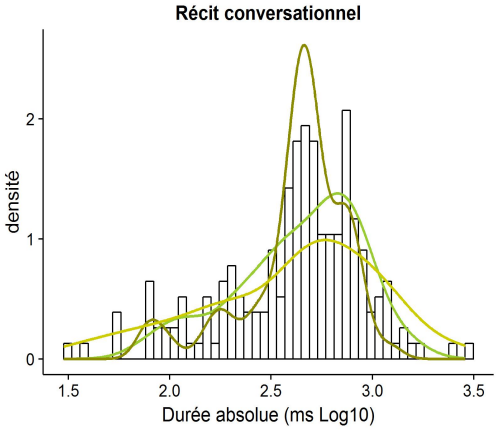


FIGURE 4.9 – Distribution des durées sur 191 pauses et 3 locuteurs pour la situation du récit conversationnel (LOCAS-F).

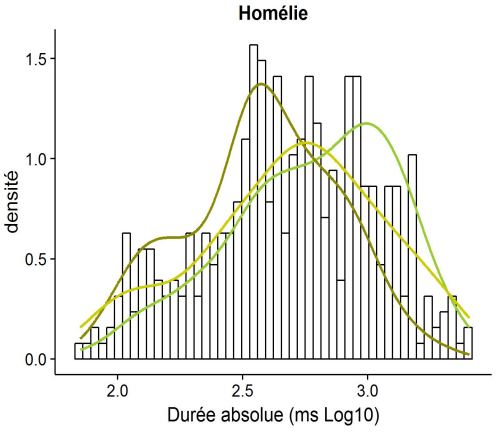


FIGURE 4.10 – Distribution des durées sur 401 pauses et 3 locuteurs pour la situation de l'homélie (LOCAS-F).

La variation entre l'aspect uni- et multimodal des distributions pausales est ainsi similaire à ce qu'observent AVANZI et. al (2014). Leurs résultats distinguent d'une part des discours politiques et contes de fées aux distributions pausales trimodales, et d'autre part des lectures de nouvelles et de dictées aux tendances bimodales. Certains auteurs formulent ainsi l'hypothèse que la composante de distribution « longue » est due au **caractère écrit** des pauses, et que la composante longue des discours politiques est connectée à un **aspect rhétorique** (CHRISTODOULIDES, GROSMAN, DEGAND & SIMON 2017:151-153). Contrairement à ces derniers, nous n'observons pas une distributon trimodale des discours politiques, que ce soit au niveau de la situation de parole ou du locuteur.

La **variation individuelle** est plus importante dans certaines situations que dans d'autres. De plus, la durée relative (Log10) suit presque systématiquement une distribution unimodale. Cette nouvelle perspective questionne la pertinence et l'adéquation de modéliser la distribution des pauses silencieuses comme un mixte de courbes gaussiennes. Elle suggère également que le **débit de parole du contexte d'apparition** de la pause joue un rôle non négligeable dans la distribution – bien entendu ces hypothèses devraient être testées par des expériences de perceptions ciblées.

Nous concluons de ces observations que le caractère unimodal, bimodal voire multimodal d'une distribution devrait être établi **au minimum par situation de parole, et parfois par locuteur**. Nous proposons méthodologiquement que les analyses précédentes servent à l' **évaluation du degré de variation intrasituationnelle** et soient utilisées comme **point de départ** à l'évaluation de la **nécessité d'utiliser des seuils de pauses adaptatifs ou non**.

Nous analysons ci-après la distribution et la durée des pauses silencieuses relativement à trois types de phénomènes : les chaînes de cooccurrences autour des pauses, les frontières d'unités syntaxiques et les contours intonatifs précédant la pause. Ce faisant nous poursuivons un double objectif, méthodologique et de modélisation.

3.2 Fréquence et durée des pauses cooccurrentes à des (dis)fluences

Nous présentons dans cette section une **analyse de la localisation séquentielle des pauses**, c'est-à-dire de leur position relativement aux unités linguistiques de différents niveaux. Grosjean & Deschamps (1975) remettent en cause la possibilité d'assigner une fonction eu égard à sa localisation contextuelle :

> « Tout silence de plus de 0,25 sec [...] est considéré comme une pause silencieuse, quel que soit son rôle dans l'activité langagière, à savoir pause d'hésitation, pause de respiration, pause grammaticale (dans laquelle nous incluons les pauses utilisées pour l'emphase et celles permettant de lever certaines ambiguïtés). En effet, ces types de pauses se confondent très souvent : un sujet profitera, par exemple, d'un arrêt grammatical pour respirer et pour hésiter, ou encore, se trouvant obligé d'hésiter à l'intérieur d'une phrase, en profitera pour respirer, etc. Essayer de déterminer rigoureusement la nature exacte de la pause silencieuse nous semble extrêmement délicat [...]. Nous avons donc préféré étudier toutes les pauses silencieuses comme un tout. »
>
> (Grosjean & Deschamps, 1975:148)

Malgré cette mise en garde, il nous a semblé pertinent de tester **si la fréquence et la durée des pauses sont corrélées à leur position dans la chaîne parlée**. Pour nous, il ne s'agit pas d'associer à une pause une unique fonction mais bien de voir si des régularités émergent. Par ailleurs, une distinction intéressante a été établie par Candea (2000a) entre les pauses structurantes et les pauses non structurantes. Une **pause structurante** désigne « toute pause silencieuse qui est encadrée par des séquences sonores appartenant au même locuteur et qui n'est pas immédiatement précédée par une marque de TdF [Travail de Formulation] » (2000a:145). Lorsqu'une pause silencieuse est immédiatement précédée d'une marque du travail de formulation (*euh*, allongement d'hésitation, répétition, faux départ, autocorrection immédiate), elle est considérée comme **non structurante** (Candea 2000a:178). Nous tenons compte de cette distinction pour présenter l'état de l'art sur la variation de la durée des pauses selon leur cooccurrence avec des marques d'hésitation (section 3.2) mais aussi leur **localisation syntaxique** (section 3.3) et leur(s) interaction(s) avec les **contours intonatifs** (section 3.4).

Selon Candea (2000a), les **pauses silencieuses non structurantes**, qui suivent immédiatement une marque du travail, ont « un fonctionnement très différent des pauses structurantes » (p. 146). Premièrement, elles **n'occasionnent presque jamais une alternance de tour de parole** (p. 147) et elles « ne sont **pas perçues** par les auditeurs comme des **indices de coupe** » (p. 148). Deuxièmement, ces pauses silencieuses non structurantes ne semblent **pas du tout perçues par les auditeurs**, que ce soit dans l'étude de perception mise en place par Candea (voir chapitre 3 2000a) ou dans une étude similaire de Duez (1995, citée par Candea 2000a:178). Par conséquent il semble légitime de considérer les pauses non structurantes comme une **catégorie distincte**.

Globalement, les **pauses non structurantes** (ou d'hésitation) sont **moins fréquentes** que les pauses structurantes (ou syntaxiques). Dans l'étude de Candea (2000a:179), elles constituent 17,5% du nombre total de pauses. Il est difficile de trouver des chiffres à comparer à ceux-là dans la littérature car peu d'études ont distingué ces deux catégories de pauses, soit parce que les auteurs estimaient qu'une telle distinction était difficile à appliquer (voir la citation de Grosjean & Deschamps *supra*), soit parce qu'une telle distinction requiert un corpus finement annoté.

Concernant la durée, Candea (2000a:161,180) rapporte que la durée moyenne des pauses non structurantes (566 ms.) est significativement **plus brève** que celle des pauses structurantes (677 ms.). Cette différence s'expliquerait par le fait que **la pause silencieuse qui suit une marque d'hésitation en serait un prolongement**, et que l'ensemble « marque d'hésitation + pause non structurante » serait perçu comme **une seule unité**.

Nous nous tournons maintenant vers la fréquence des pauses **en fonction de leur cooccurrence avec différentes (dis)fluences**. L'annotation des disfluences[17] a été exploitée de manière à extraire des chaînes des éléments qui précèdent et suivent immédiatement les pauses silencieuses.

Pour rappel, chaque pause intratour est catégorisée en fonction de son environnement immédiat. Les pauses dites « **seules** » sont précédées et suivies de tokens qui n'ont pas été annotés comme des (dis)fluences potentielles (exemple 20a) : les pauses peuvent être **précédées et/ou suivies d'un marqueur de discours** (exemple 20b), d'un *euh* (exemple 20c) ou d'une **autre disfluence** (comme une troncation ou un faux départ) (exemple 20d). Enfin, les **pauses en *interregnum*** sont inclues entre un élément à corriger et sa correction (exemple 20e).

Exemple (20) – Exemples de chaînes de pause(s). Issues de LOCAS-F.

a. des millions et des millions de personnes viennent ici **<pause>** à la grotte de Massabielle (hom_2)
b. **et <pause>** selon le New-York Times (jpa_2)
c. on est tous montés dans les taxis londoniens ***euh* <pause>** quatre par quatre (rcv_3)
d. avant de **pr/ <pause> et** en prenant *euh* dans chaque maison une personne (rcv_3)
e. je ne rencontre pas **une <pause> un** individu à gauche qui ne me parle pas de cette question (pol_4)
f. **voilà <pause> *euh* donc** on est parti au mois de juillet (rcv_3)

Les pauses peuvent également s'insérer dans des chaînes de **trois (ou plus) (dis)fluences** (exemple 20f). Ces chaînes peuvent contenir plusieurs pauses (e.g. pause + marqueur + pause + marqueur). Nous avons fait le choix de les regrouper sous **l'étiquette « mixte » ($n \geq 3$)**. Le Tableau 4.3 présente la fréquence de chaque type de chaîne. Au total, 3518 chaînes ont été extraites (certaines contiennent plus d'une pause).

17. Voir chapitre 3 section 2.1.

Chaîne de cooccurrence	Fréquence (n)	Prop (%)
pause seule	2593	53,5
pause + md	249	5,1
md + pause	684	14,1
pause + *euh*	146	3,0
euh + pause	211	4,4
pause + dis	115	2,4
dis + pause	104	2,1
pause *interregnum*	368	7,6
chaînes $n \geq 3$ (pause + md + *euh*, pause + *euh* + pause, etc.)	375	7,7
Total	4845	100

TABLEAU 4.3 – Fréquences et proportions des pauses en fonction des cooccurrences les plus fréquentes. *pause = pause silencieuse, md = marqueur de discours, euh = pause pleine.*

Les pauses intervenant entre de simples tokens sont les plus fréquentes (53,5%), suivies par les **pauses après un marqueur de discours** (14,1%). Si l'on y ajoute les pauses précédant un marqueur de discours (5,1%), un *euh* (3%) ou une (dis)fluence (2,4%), on obtient un total de **78% de pauses silencieuses qui ne suivent pas une (dis)fluence** et pourraient être considérées comme **pauses structurantes ou grammaticales**. Ce total confirme les observations de CANDEA (chapitre 3 2000a) qui compte dans ses données 82,5% de pauses structurantes. Viennent ensuite les pauses internes à des chaînes de (dis)fluences (7,6% et 7,7%) et celles qui suivent un *euh* (4,4%) ou une autre (dis)fluence (2,1%), soit un total de 22% de pauses d'hésitation.

L'analyse de la distribution des chaînes de pauses en fonction du **degré de préparation** de la parole montre qu'il existe des **associations fortes** entre certains types de chaînes et les situations de parole préparée, semi-préparée ou spontanée (Figure 4.11). En ordonnée se trouvent les types de chaînes incluant des pauses silencieuses et en abscisse se trouvent les situations de parole regroupées selon le trait « degré de préparation » (spontané, semi-préparé ou préparé). La taille des boîtes est proportionnelle aux nombres d'occurrences et la **taille de l'effet** (corrélation positive ou négative) est indiquée par la **mesure de résidu** (à droite).

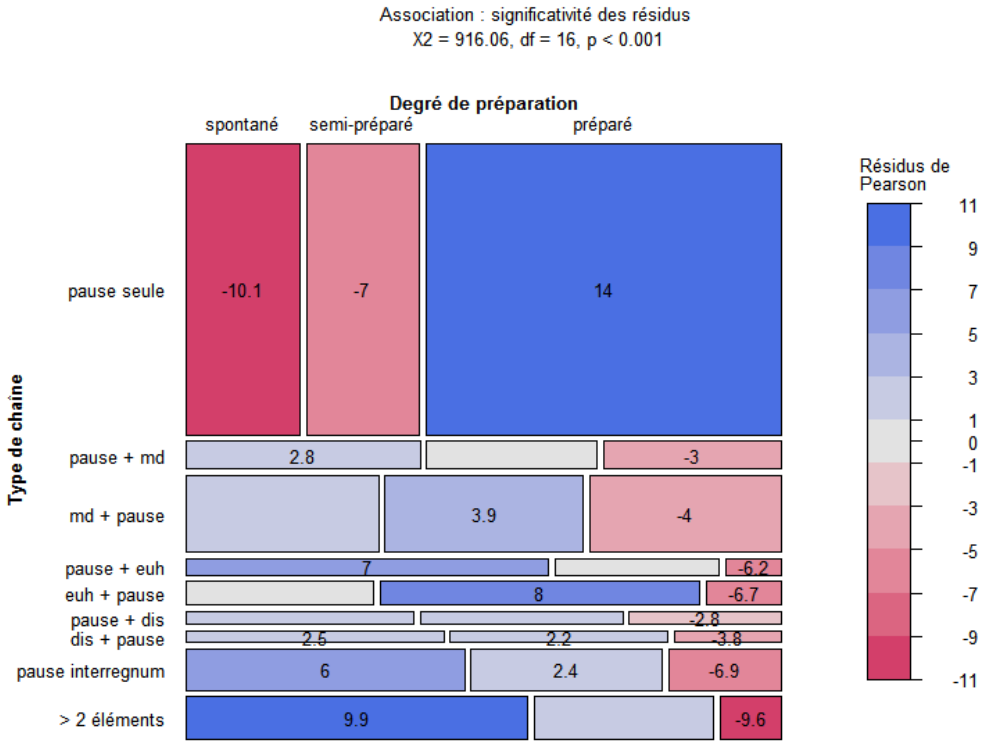


FIGURE 4.11 – Graphique des tableaux croisés du degré de préparation et des chaînes de cooccurrences.

L'observation la plus évidente est que les **pauses isolées** sont prédominantes dans les **situations de parole préparée** (journaux parlés, homélies, discours politiques, etc.) et plus rares dans les styles non préparés (conversation informelle, narration conversationnelle, etc.). Les configurations les plus typiques des **situations spontanées** sont les **séquences mixtes**, puis les pauses en *interregnum*, les *euh* suivis d'une pause et les pauses précédant un *euh*. En termes de fréquence, on n'observe pas de différence entre (*euh* + pause) et (pause + *euh*). Enfin, les pauses qui suivent ou précèdent un marqueur de discours sont moins fréquentes dans les situations préparées, mais de manière moins saillante.

La Figure 4.12 met en évidence les **différences de durées en fonction des marques en cooccurrence avec la pause**. De manière générale, les **pauses précédées d'un *euh* semblent être les plus courtes**. La tendance au raccourcissement relatif de la pause suivant une marque ne se retrouve pas dans la configuration « md + pause » ni dans « dis + pause », qui semblent au contraire plus longues.

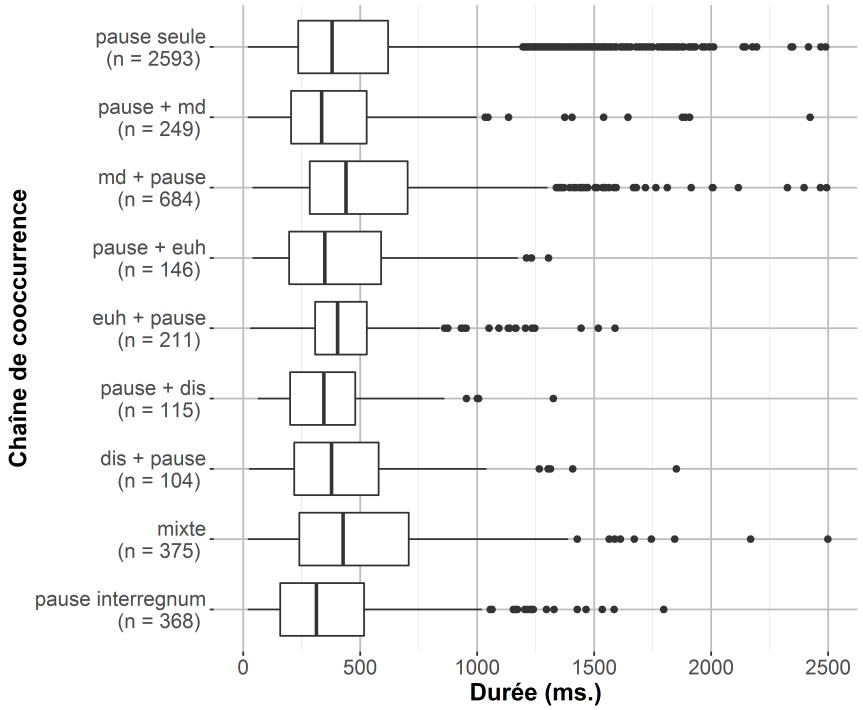


FIGURE 4.12 – Durées pausales en fonction des cooccurrences.

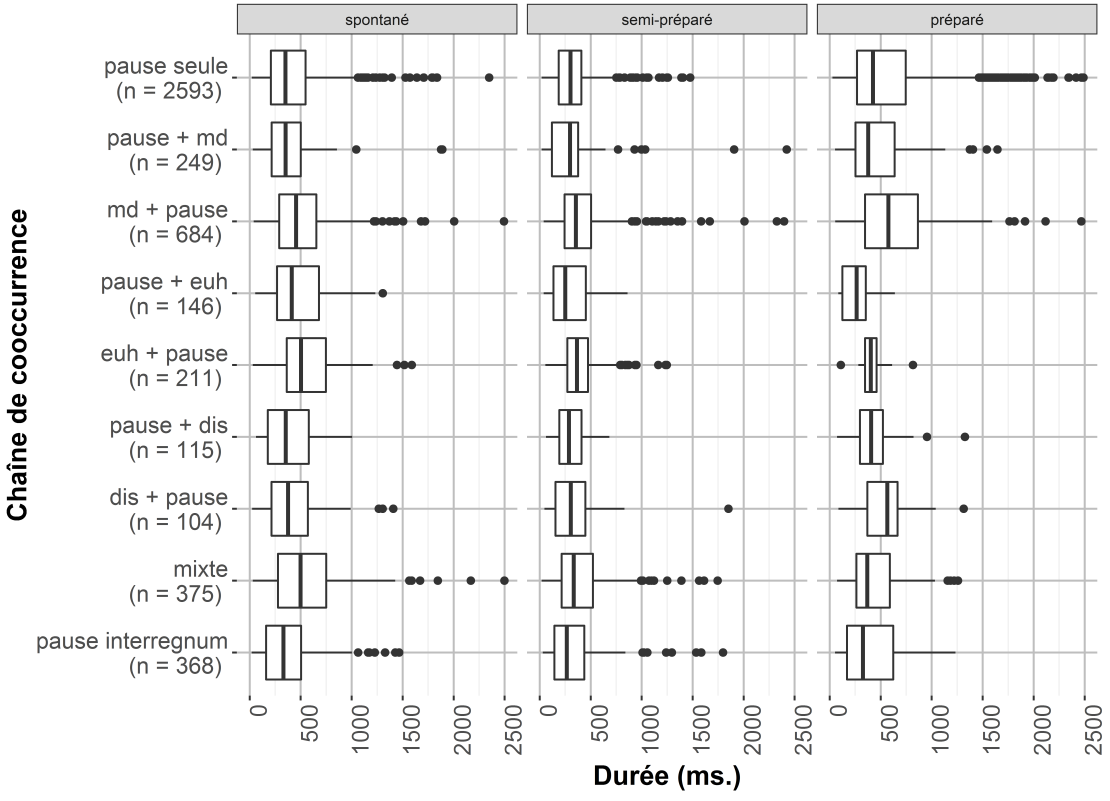


FIGURE 4.13 – Durées pausales en fonction des cooccurrences et du degré de préparation.

Une analyse de la variance (ANOVA) sur la durée des pauses en fonction des cooccurrences donne lieu à une variation significative entre les conditions ($F(8, 4836) = 8.58, p < .001$). Un test post-hoc de Tukey DSH montre que les durées pausales en fonction des cooccurrences peuvent être réunies en **4 groupes**. Les **pauses en combinaison mixte et précédées d'une marque (md ou *euh*) sont les plus longues**. Au contraire, la **pause au sein d'un *interregnum* est la plus courte**. Enfin, la chaîne (dis + pause) fait partie du même groupe que les chaînes cooccurrentes de pauses suivies d'une marque. Les pauses isolées constituent un **groupe isolé également et de durée moyenne** par rapport aux autres cooccurrences [18].

Peu de différences émergent en fonction du degré de préparation du discours (Figure 4.13). L'ordre des durées pausales est préservé entre les différents degrés de préparation, bien que ces durées soient globalement **plus longues en situation préparée** ($F(2, 4842) = 77.7, p < .001$). Dans la section suivante, nous rendons compte du **rôle des facteurs syntaxiques** pour expliquer la distribution des pauses silencieuses. Cela permettra notamment de centrer l'attention sur les pauses seules (non combinées avec des disfluences).

18. On observe donc quatre groupes distincts de durées pausales, des plus longues aux plus courtes : [1] (md + pause > *euh* + pause > mixte) > [2] (pause seule) > [3] (dis + pause > pause + *euh* > pause + md > pause + dis) > [4] (pause *interregnum*).

3.3 Fréquence et durée des pauses aux frontières syntaxiques

Plusieurs études soulignent que la fréquence et la durée des pauses sont positivement **corrélées avec le niveau de la frontière syntaxique** où elles occurrent, établissant une **fonction segmentante** et **hiérarchisante** de la pause silencieuse. Ainsi, Grosjean & Deschamps (1975:167) montrent que 60,46% des pauses dans leur corpus d'interviews radiophoniques se distribuent en « **fin de phrase** », contre 39,54% à l'intérieur d'une phrase (où elles se situent majoritairement dans le syntagme verbal, ou entre le syntagme verbal et un autre constituant). Dans leurs données, une différence de durée existe mais n'est pas significative (la médiane de la durée des pauses en fin de phrase s'élève à 500 ms. contre 400 ms. à l'intérieur des phrases).

À partir des effectifs rapportés par Candea (2000a:154), on observe que 77,66% des **pauses silencieuses structurantes sont positionnées en fin d'énoncé** ou de proposition, contre 19,05% en fin de constituant et seulement 3,30% à l'intérieur d'un constituant. Dans ces données, la durée moyenne des pauses est significativement **plus longue en fin d'énoncé** (881 ms.) ou de proposition (608 ms.) qu'en fin de constituant (486 ms.).

Ces tendances sont confirmées par Goldman, François, Roekhaut & Simon (2010:163) qui remarquent que la variable syntaxique[19] la mieux corrélée avec la durée d'une pause est la position **en fin d'unité de rection** – l'unité syntaxique maximale dans leur modèle, qui équivaut globalement à la fin d'énoncé ou de proposition chez Candea (2000a). Utilisant une procédure expérimentale de production contrôlée, Krivokapic (2006) pointe également, outre le niveau de frontière syntaxique, **l'impact de la longueur des constituants qui précèdent et suivent la pause** sur la longueur de celle-ci.

La Figure 4.14 confirme l'hypothèse que la **durée moyenne d'une pause augmente en fonction du degré de frontière syntaxique**, comme cela a été montré précédemment (Candea 2000a ; Goldman, François, Roekhaut & Simon 2010 ; Grosjean & Deschamps 1975). Les pauses se distribuent dans 44,1% des cas ($n = 2138$) entre les constituants syntaxiques minimaux (groupes accentuels accentués) ; dans 35,8% des cas ($n = 1735$) à la fin des unités syntaxiques maximales (unités de rection) ; dans 10,79% des cas ($n = 518$) entre les séquences syntaxiques (unités syntaxiques intermédiaires) ; et enfin dans 9,4% des cas (454) à l'intérieur des groupes accentuels accentués. On constate cependant que l'**écart entre les pauses inter-séquences et inter-rections** est nettement **moins marqué** qu'entre pauses intra-ga, inter-ga et inter-séquences.

19. L'annotation utilisée dans Goldman, François, Roekhaut & Simon (2010) est similaire à celle utilisée sur l'entièreté de LOCAS-F. Celle-ci est détaillée au chapitre 3 section 2.3 ainsi que dans l'Annexe C sections 11, 12, 16 et 17.

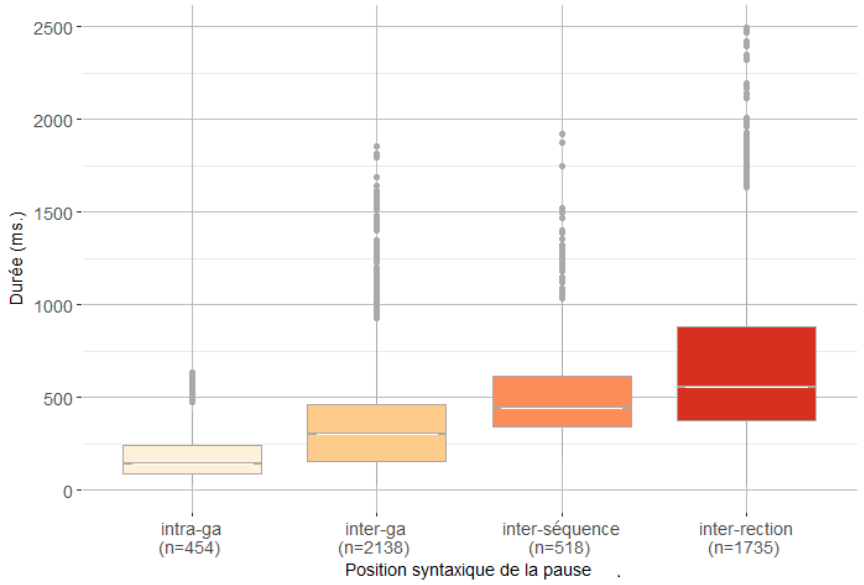


FIGURE 4.14 – Durées pausales en fonction de la position syntaxique de la pause (Kruskal-Wallis chi-squared = 1457.2, $dl = 3$, $p < 0.001$).

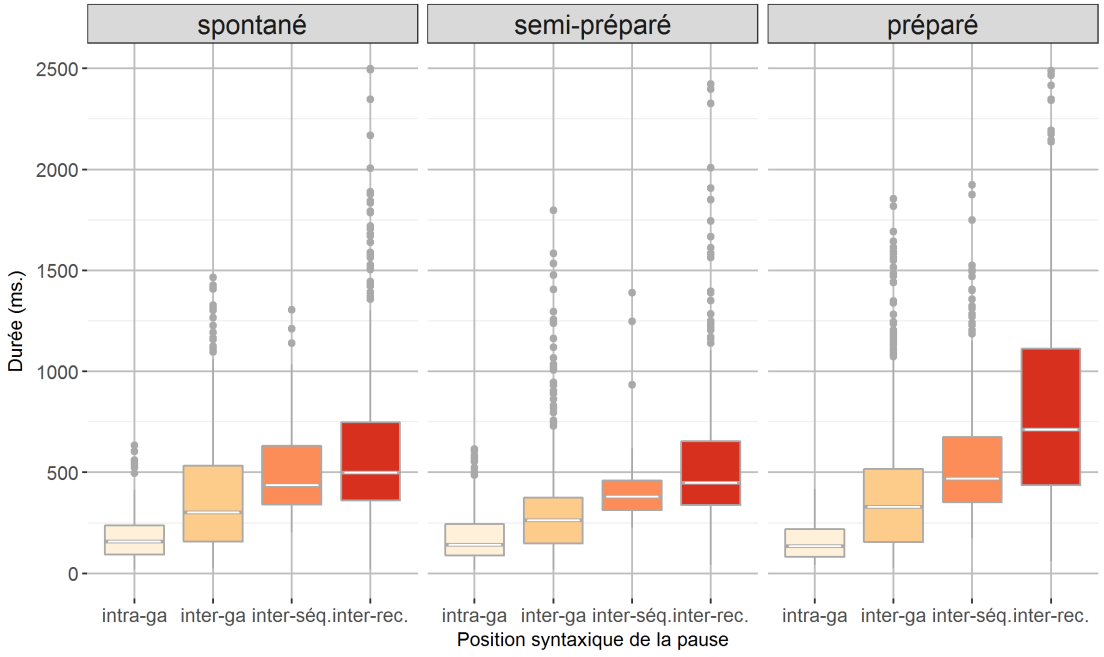


FIGURE 4.15 – Durées pausales en fonction de la position syntaxique et du degré de préparation (K, spontané = 402.5 : K, semi-préparé = 394.8, K, préparé = 643 : $dl = 3, p < 0.001$).

Cette distribution est influencée par le degré de préparation (Figure 4.15). Comme on pouvait s'y attendre, les pauses **internes aux constituants syntaxiques minimaux**, qui **interrompent une unité cohésive**, sont **moins fréquentes** en **parole préparée** qu'en parole semi-préparée ou spontanée. Les pauses qui distinguent le mieux la parole préparée sont les pauses de niveau intermédiaire, entre les séquences syntaxiques, ce qui peut s'expliquer par le fait que les **séquences syntaxiques sont plus longues et plus complexes en parole préparée** qu'en parole spontanée [20].

Les pauses **entre les unités syntaxiques maximales** (de rection) sont les **plus fréquentes en parole spontanée** – ce qui peut s'expliquer par la brièveté des unités maximales dans ce style – et les moins fréquentes en parole semi-spontanée – cette dernière tendance étant plus difficile à interpréter. La durée des **pauses inter-rections** est significativement **plus longue** dans le style préparé que semi-préparé ou spontané (Figures 4.14 et 4.15), différence qui ne se marque pas aux niveaux inférieurs.

Si l'on considère que les **pauses internes aux groupes accentuables**, qui interviennent dans un constituant syntaxique minimal, sont « **disfluentes** », ce schéma confirme que **les pauses (dis)fluentes sont plus brèves que les pauses structurantes** (Candea 2000a:270).

Enfin, nous analysons l'interaction entre la position syntaxique et des chaînes de disfluences sur la durée pausale (Figure 4.16), les deux annotations ayant été faites indépendamment.

Dans presque tous les types de chaînes, **la durée de la pause augmente avec le degré de frontière syntaxique**. C'est le cas particulièrement pour les **pauses seules ou en cooccurrence avec des marqueurs de discours**, que nous avons qualifiées de **pauses structurantes ou grammaticales**. En effet, les trois premières lignes de la Figure 4.16 montrent que les pauses sont les plus longues en position syntaxique inter-rections (frontière syntaxique maximale), et **diminuent progressivement avec l'affaiblissement de la force de frontière syntaxique** : les pauses sont un peu moins longues aux frontières inter-séquences, encore moins longues en position inter-ga, et plus courtes en position intra-ga.

20. Le corpus LOCAS-F compte 26607 groupes accentuables, 16979 séquences syntaxiques et 12099 unités syntaxiques maximales de rection. Ces unités présentent des durées plus courtes en parole spontanée et plus longues en parole préparée (pour chaque unité, on donne la durée moyenne et la médiane) :
- groupe accentuable en spontané, moyenne = 841 ms. et médiane = 450 ms. ; en situation préparée, moyenne = 623 ms. et médiane = 560 ms. ;
- séquences en spontané : moyenne = 1110 ms. et médiane = 465 ; en préparé : moyenne = 1333 ms. et médiane = 717 ms. ;
- unités maximales de rection en spontané : moyenne = 1422 ms. et médiane = 508 ms. ; en préparé : moyenne = 2506 ms. et médiane = 1063 ms.

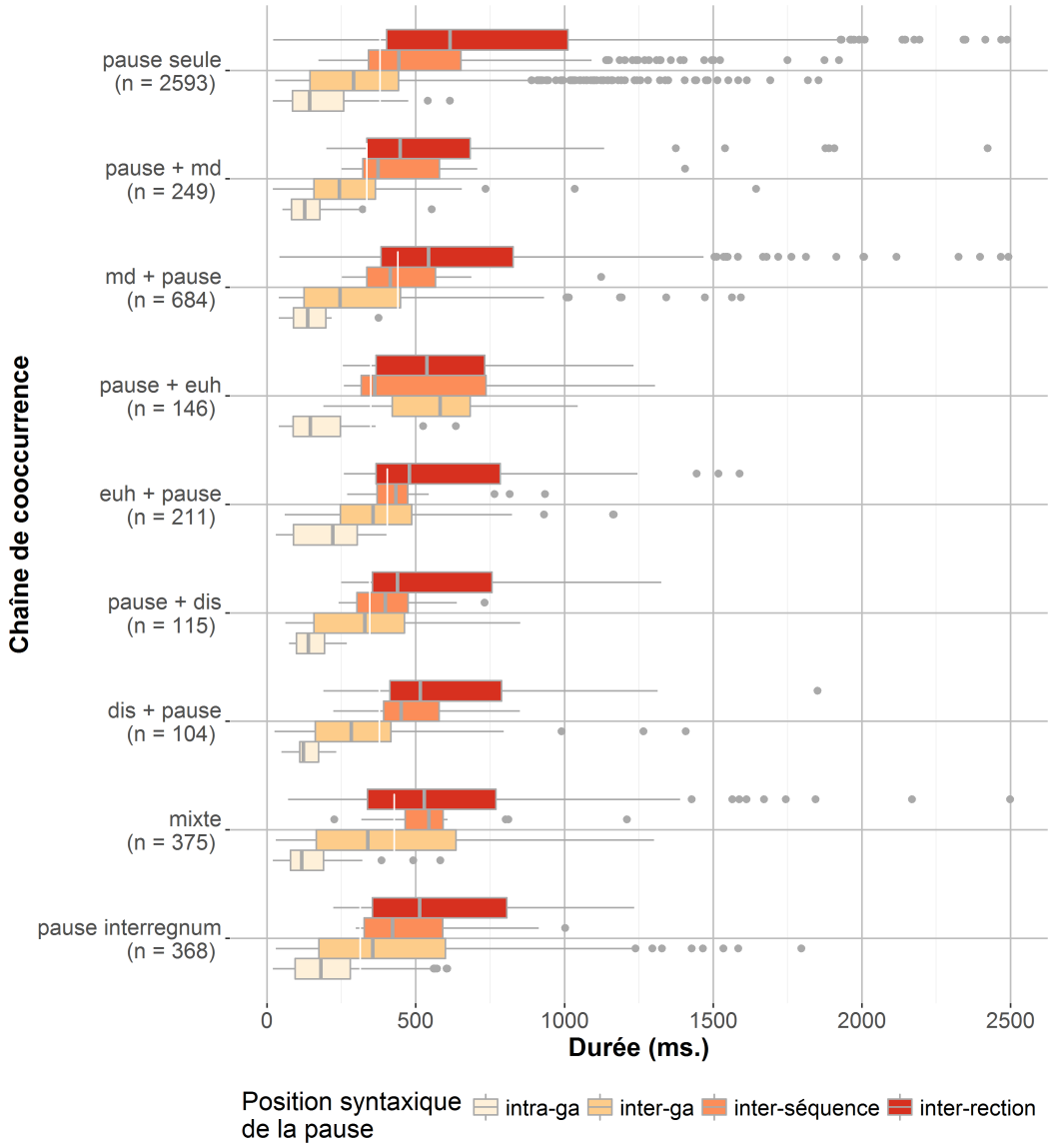


FIGURE 4.16 – Diagramme en boîte à moustaches de la distribution des durées pausales en fonction de leurs cooccurrences et de leur localisation syntaxique et du contour intonatif précédant la pause.

Deux cas font exception. Premièrement, la durée des (**pauses silencieuses** + ***euh***) (n = 152) semble influencée par le degré de frontière syntaxique, **sauf dans les cas intra-ga** où la pause est beaucoup plus brève. C'est aussi le cas, dans une moindre mesure, pour (*euh* + pause) et les pauses dans les chaînes de plus de 3 éléments (la dernière ligne de la figure). On pourrait en conclure que **(pause + *euh*) s'apparente davantage à un environnement disfluent que (*euh* + pause)**.

Dans la section suivante, nous analysons l'interaction des contours intonatifs sur la durée des pauses. Ces analyses permettent de préciser la fonction des pauses isolées et des chaînes de pauses au regard de l'intonation, et donc de leur éventuel effet de rupture /vs/ continuité intonative et/ou discursive.

3.4 Fréquence et durée des pauses en fonction des contours intonatifs

Cette section examine les interactions entre les distributions des pauses silencieuses et les contours intonatifs des segments précédant la pause. Nous posons les questions suivantes. textbfLe type de contour intonatif est-il associé à la durée de la pause subséquente ? Cette relation varie-t-elle selon la situation ? Enfin, si l'on veut pouvoir associer les pauses structurantes et non structurantes à certains contours intonatifs, **comment interagissent ces trois facteurs** (cooccurrences de (dis)fluences, contours intonatifs, position syntaxique de la pause) ?

Méthodologiquement, l'annotation des contours intonatifs a été effectuée manuellement [21] : à chaque **frontière prosodique majeure** et **intermédiaire** a été assigné un contour intonatif. Les frontières majeures sont notées « /// », et les frontières intermédiaires « // » [22].

Chaque dernière syllabe d'une unité prosodique intermédiaire ou majeure, a reçu une annotation du contour intonatif. En fonction du contour mélodique, cette syllabe se voit attribuer une lettre. Pour rappel, le **contour terminal (T)** est mélodiquement **descendant**. Le **contour continuatif-montant (C)** comporte soit une **montée intrasyllabique** soit intersyllabique (un intervalle positif) par rapport à la syllabe précédente. Le **contour de suspens (S)** est **plat**, sans écart mélodique par rapport à la syllabe qui précède. Enfin, le **contour de focalisation (F)** comporte un **mouvement montant-descendant** partant d'une cible haute et avec une descente raide. Les syllabes marquées (NA) regroupent les segments précédants la pause auxquels aucun contour intonatif n'a été assigné. Nous nous réferons à ces syllabes dans la suite du chapitre comme des « syllabes non marquées » [23].

Nos analyses visent à vérifier les hypothèses suivantes :

a. Les pauses silencieuses ont tendance à être **plus courtes après un contour continuatif** (qu'un contour terminatif ou de suspens). Dans une logique de fluence cognitive, le contour continuatif appelle à la poursuite du discours et donc à un temps de pause relativement bref (sous peine d'alternance de tours).

b. **Au niveau de la cooccurrence**, on s'attend à ce que les chaînes de cooccurrences de (dis)fluences soient entourées de pauses relativement longues lorsqu'elles sont précédées d'un contour continuatif. La continuation intonative associée aux (dis)fluences relevant d'une **incohérence discursive**, la pause révèlerait une réelle hésitation du locuteur.

21. L'annotation des contours se base sur la détection de frontières selon leur hauteur mélodique, durée relative et la **durée de la pause subséquente**. Elle a été effectuée manuellement, mais les annotatrices avaient à leur disposition une visualisation de la détection de la fréquence fondamentale et une annotation automatique fine, qui permettaient de valider la perception auditive à la base de l'annotation manuelle. Pour un détail du protocole d'annotation, consulter Degand, Martin & Simon (2011). Nous n'avons pas distingué le niveau de frontière intermédiaire du niveau majeur dans les présentes analyses.

22. L'annotation des frontières prosodiques et des contours intonatifs est détaillée en Annexe C sections 18 et 19.

23. Cependant, il est possible que ces segments marquent tout de même des frontières qui n'auraient pas été perçues par les annotateurs humains.

c. Un contour continuatif **accompagné de (dis)fluences et de pauses longues** signalerait éventuellement une **discontinuité intonative** ; celle-ci s'accompagnerait d'une **discontinuité syntaxique** particulièrement lorsqu'il s'actualise dans des positions syntaxiques intermédiaires (intra-ga, inter-ga, intra-séquence).

d. En outre, **au niveau de la cooccurrence et des frontières syntaxiques**, on s'attend à ce que les pauses silencieuses **structurantes intraconstituantes** [24] ne soient pas accompagnées de marquage continuatif ou terminatif. En effet, les pauses structurantes intraconstituantes ne forment qu'un seul constituant sur le plan segmental et prosodique et ne sont donc pas suivies de frontières.

Nous testons avant tout si la **durée d'une pause est corrélée au contour intonatif la précédant**, toute situation confondue. L'observation de la Figure 4.17 indique que la durée médiane des pauses de LOCAS-F fluctue de 325.2 ms. à 499.7 ms. en fonction du contour intonatif précédant la pause silencieuse [25].

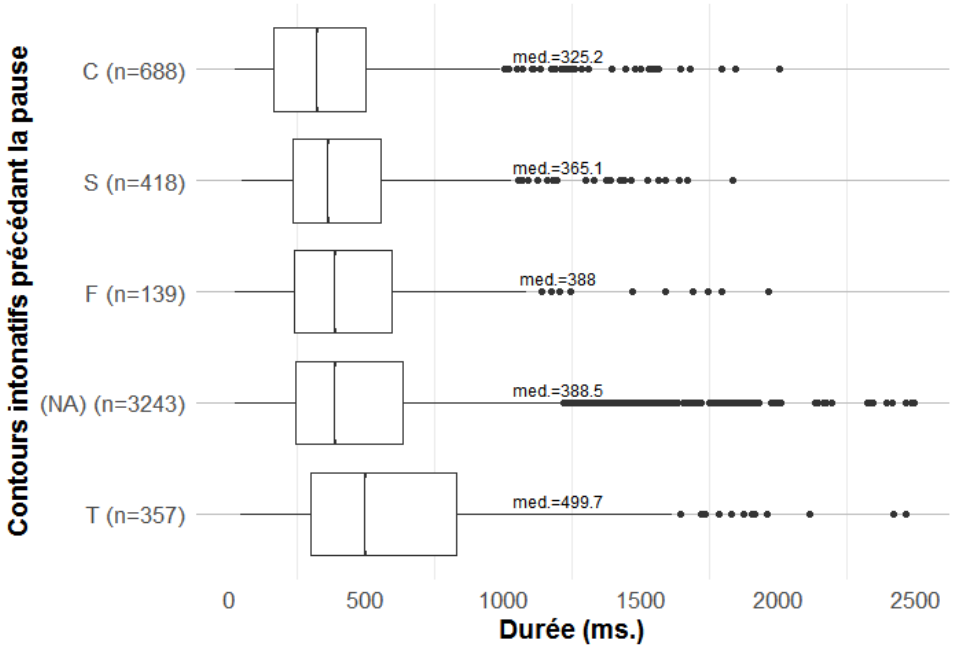


FIGURE 4.17 – Nombre d'occurrences et médiane de la durée pausale en fonction du contour intonatif la précédant.

24. Les pauses structurantes peuvent être intraconstituantes dans la terminologie de CANDEA (2000a:165,198). L'auteure observe que la configuration pause isolées (non accompagnées de marqueur de (dis)fluence) en position intra-ga ou intra-séquence se retrouve dans trois cas : comme effet stylistique, en intra-séquence comme « rajout » après un constituant syntaxiquement complet ou comme « hésitation », et en intra-ga entre un mot outil et un mot plein. Bien que relativement rares dans le corpus de Candea, ces configurations sont fréquentes dans le corpus politique de DUEZ (1991).

25. Dans cette section, les durées des pauses sont représentées dans les graphiques en millisec. Cependant, nos statistiques sont effectuées sur les données transformées en logarithme de base 10. Nous les reportons donc comme tel. Lorsque nécessaire, nous retransformons les résultats à des fins de lisibilité en chiffres naturels.

Une analyse de la variance (ANOVA) sur ces résultats indique une variation significative entre les conditions ($F(4, 4840) = 23.734, p < .001$). Un test post-hoc de Tukey DSH permet de distinguer **3 groupes de contours** en fonction de leur association avec les durées pausales. Il montre que les pauses précédées d'un contour terminal (T) ou d'un contour continuatif (C) appartiennent à des groupes de durées différents (à $p < .05$). Ainsi, les pauses **précédées d'un contour continuatif** sont **plus courtes** que celles précédées par d'autres contours. A l'inverse, les pauses **précédées par un contour terminal** sont généralement **plus longues**. Les contours de suspens (S), de focalisation (F) et les syllabes non marquées (NA), n'entrainent pas de changement significatif des durées pausales leur succédant.

Notre second questionnement a trait à l'association entre le **contexte immédiat des pauses et les types de contours intonatifs les précédant**. Nous avons vu en section 3.2 que les pauses en combinaison de plus de deux éléments, ainsi que les chaînes (md + pause) et (*euh* + pause) ont tendance à être plus longues. Cet allongement est-il aussi associé à la présence d'un **contour intonatif terminal** sur la syllabe précédente ? En d'autres termes, y a-t-il une **addition des effets** sur la durée des pauses (contexte de cooccurrence et contours) ? *In fine*, peut-on associer les pauses structurantes et non structurantes à certains contours intonatifs ?

Premièrement, les analyses confirment une **forte association** entre les **contours et les types de cooccurrences** associées aux pauses ($\chi^2 = 359.6, df = 32, p < 0.001$). En abscisse de la Figure 4.18 se trouvent les types de chaînes incluant des pauses silencieuses et en ordonnée les contours des syllabes précédant la cooccurrence. La taille des boîtes est proportionnelle au nombre d'occurrences et la **taille de l'effet** (corrélation positive ou négative) est indiquée par la **mesure de résidu** (à droite).

La première ligne souligne le fait que les **pauses seules** ont tendance à être précédées par des **syllabes non marquées**, exemptes de contours, notées (NA). Les résidus négatifs indiquent que les pauses seules sont également moins souvent précédées d'un contour continuatif ou de suspens que les autres pauses.

Au contraire, les **contours continuatifs** et de **suspens** sont fortement associés à des chaînes de cooccurrences telles que **(pause + md) ou (pause + dis)**, mais aussi à des chaînes de cooccurrences de **plus de 2 éléments** et de **pauses en *interregnum***. Nous notons également que les chaînes **(md + pause) et (pause + *euh*)** sont rarement précédées par un contour de suspens mais plutôt par un **contour terminal**. Enfin, la configuration (md + pause) s'associe par ailleurs au contour continuatif, tandis que la configuration **(pause + *euh*) a tendance à le rejeter en faveur d'un contour terminal**.

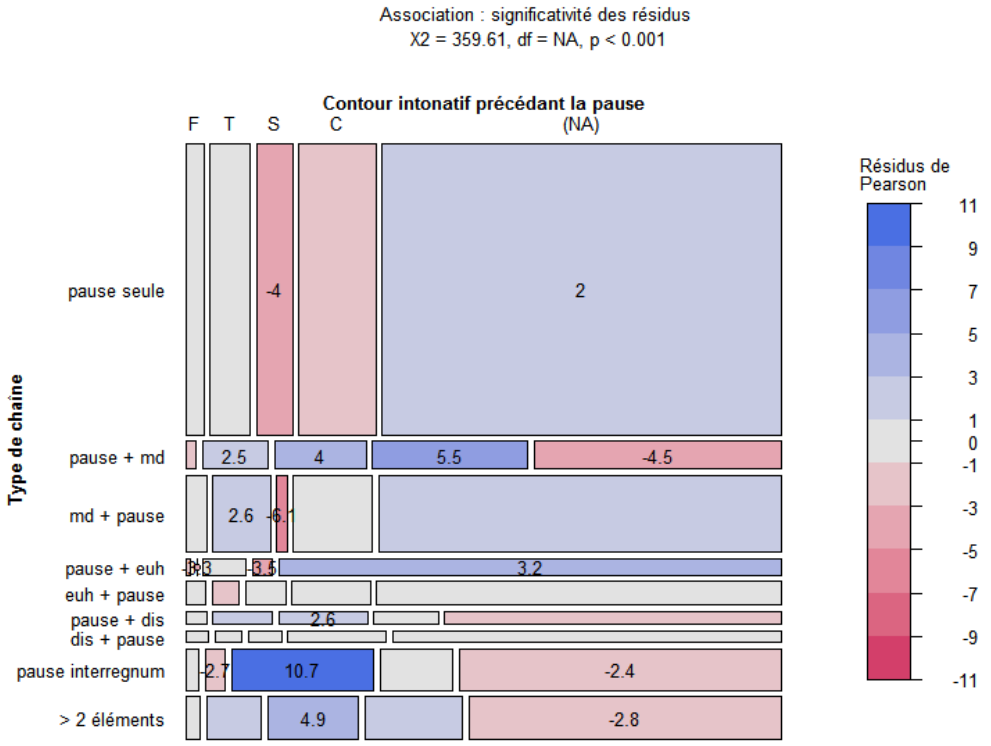


FIGURE 4.18 – Graphique des tableaux croisés des chaînes de cooccurrences et du contour intonatif précédant la pause.

L'association entre contour précédant la pause et la cooccurrence de pause est donc fortement marquée. Avant d'interpréter ces résultats au regard du continuum fluence-disfluence, ou de l'éventuel effet de rupture /vs/ continuité (intonative) de chaque chaîne, nous proposons de regarder si ces associations avec les contours ont un impact sur la **durée de la pause** dans la chaîne de cooccurrence. Peut-on par exemple observer **plusieurs patrons de durée pour un même type de chaîne de cooccurrences** en fonction du contour tonal ? La Figure 4.19 met donc en rapport ces deux facteurs au regard de la durée de la pause. Par ailleurs, nous analysons celle-ci en tenant compte des résultats précédents (notamment les Figures 4.12, 4.17 et 4.18).

Le groupe de chaînes de cooccurrences aux pauses les plus longues réunit les **(md + pause), (*euh* + pause) et les chaînes de plus de deux éléments**.

Les chaînes (md + pause) sont caractéristiques des situations de parole spontanée et semi-préparée. Elles **rejettent le contour de suspens** et sont plutôt liées à la **présence d'un contour terminal** ou l'absence de contour. Le contour de focalisation est également marginalement observé en antéposition de ces chaînes. La durée des pauses dans les chaînes (md + pause) est **significativement différente selon les types de contours qui les précèdent** ($F(4, 679) = 2.803, p < 0.05$). Les comparaisons post-hoc, effectuées en utilisant le test de Tukey DSH, confirment que les pauses sont **plus allongées en contour terminal** (479 ms. ; $\mu = 2.68, \sigma = 0.37$) et en **l'absence de contour** (428.35 ms. ; $\mu = 2.63, \sigma = 0.32$). Dans les rares cas d'un

contour de focalisation ou de suspens, les durées sont intermédiaires – respectivement F($\mu = 2.59, \sigma = 0.38$) et S($\mu = 2.55, \sigma = 0.26$). Enfin, la durée de la pause est particulièrement courte lorsque le md est marqué par un contour continuatif (333.74 ms. ; $\mu = 2.52, \sigma = 0.38$).

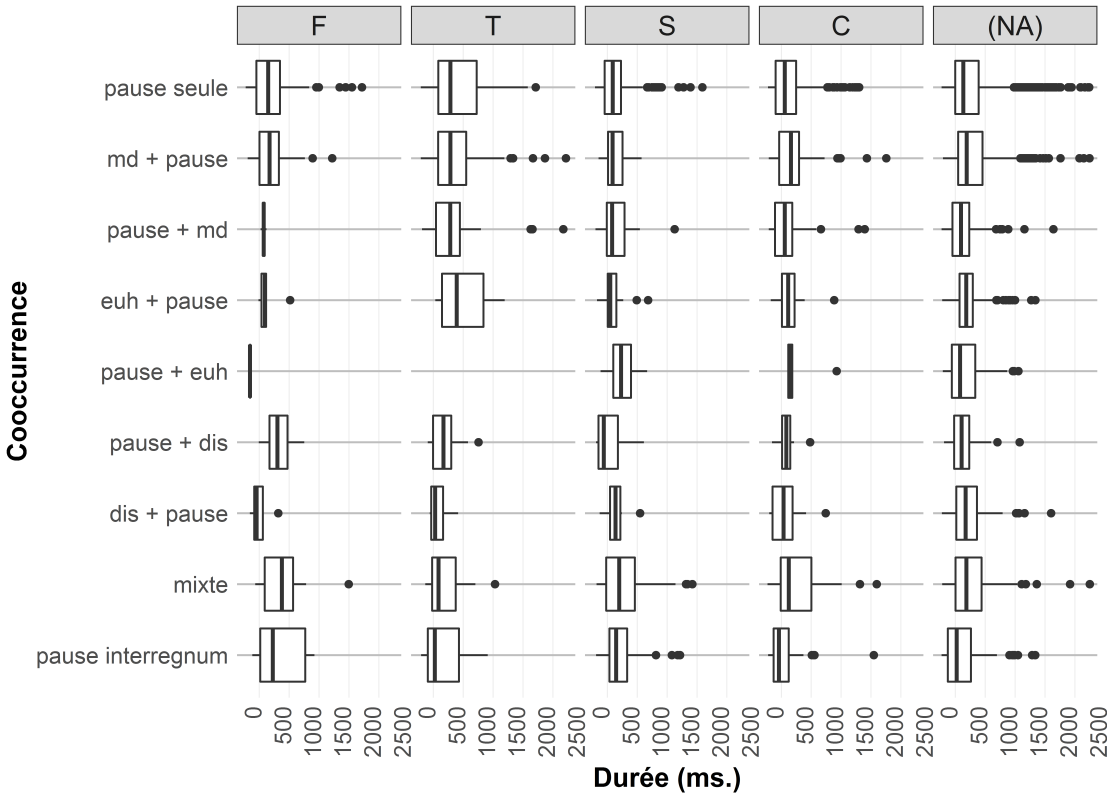


FIGURE 4.19 – Diagramme en boîte à moustaches de la distribution des durées pausales en fonction des chaînes de cooccurrence et du contour intonatif précédant la pause.

Les chaînes (*euh* + pause) sont, contrairement aux (md + pause), négativement associées au contour terminal. Elles sont par contre liées au **contour de focalisation**, relativement rare par ailleurs. Des différences de durée des pauses existent également en fonction du contour les précédant ($F(4, 206) = 2.90, p < 0.05$). Lorsqu'elles sont associées à un **contour terminal**, la durée des pauses suivant le *euh* reste significativement plus **élevée** ($\mu = 2.80, \sigma = 0.26$). Ces chaînes sont davantage antéposées par un **contour de focalisation**. La pause y est alors d'une **durée moyenne** ($\mu = 2.54, \sigma = 0.29$), similaire à celle des contours précédents non marqués, mais significativement plus longue que lorsque précédée de contours continuatif ou de suspens (respectivement $\mu = 2.5, \sigma = 0.27$ et $\mu = 2.47, \sigma = 0.29$).

Les chaînes de plus de deux éléments comprenant une pause, caractéristiques du spontané, sont fortement associées aux **contours continuatif et de suspens**, et dans une moindre mesure terminal. Elles sont très rarement précédées d'une syllabe exempte d'un contour. Contrairement aux chaînes antérieures, nous n'observons **pas de différences significatives de durée** des pauses en fonction des contours. Leur durée moyenne siège à $\mu = 2.57$ (375.5 ms.).

Les **pauses seules** sont particulièrement représentatives des situations de parole préparée. Elles sont fortement associées à l'**absence de contour** ou à un **contour terminal** antécédent (et dans une moindre mesure, à des contours de focalisation). Au contraire, elles sont rarement précédées d'un contour de suspens ou continuatif. La durée des pauses varie fortement en fonction du contour ($F(4, 25 = 17.79, p < 0.001$). Les pauses sont marquées par une moyenne et une étendue significativement **plus élevées en contour terminal** ($\mu = 2.71, \sigma = 0.35$). Viennent ensuite dans l'ordre décroissant les syllabes antéposées à la pause non marquée. Il n'existe pas de différence de durée entre le contour de focalisation et de suspens lorsque la pause est seule (respectivement, $F(\mu = 2.55, \sigma = 0.40)$ et $S(\mu = 2.50, \sigma = 0.34)$. Enfin, la durée de la pause précédée par un contour continuatif est significativement plus courte que lorsque précédée d'autres contours ($\mu = 2.45, \sigma = 0.36$).

Un troisième groupe de chaînes de cooccurrences rassemblait, sans différences significatives des durées de pauses entre ce groupe, les chaînes **(dis + pause), (pause + *euh*), (pause + md), (pause + dis)**. Cependant, des différences d'associations avec les contours ainsi que de durées apparaissent au sein de ce groupe.

Les chaînes (dis + pause) sont relativement rares en fréquence mais typiques du discours spontané ou semi-préparé. Les pauses qui s'y actualisent sont marginalement associées à un contour continuatif sur la syllabe la précédant (i.e. sur la (dis)fluence). Les pauses précédées d'une chaîne de (dis)fluence(s) **ne sont pas de durée significativement différente en fonction du contour qui les précède** ($F(2, 99) = 1.883, p = 0.119$). Leur moyenne siège à $\mu = 2.53$ (343.5 ms.).

Les chaînes (pause + *euh*) sont particulièrement fréquentes en situation de parole spontanée. La syllabe précédant ce type de chaîne est en général **non marquée**. Quelques syllabes marquées par un contour **de suspens** sont attestées mais de manière marginale. Une forte association négative existe avec les contours continuatif et terminal. Par conséquent, la configuration (pause + *euh*) peut être vue comme un **signe de rupture intonative, voire discursive**. Il n'existe par ailleurs pas de différence significative de la durée de la pause en fonction du contour intonatif la précédant ($F(3, 142) = 2.049, p = 0.11$).

Les chaînes (pause + md) sont plus courantes en discours, particulièrement spontané. Celles-ci sont fortement associées à des syllabes antécédentes portant un contour **continuatif**, **de suspens** et **terminal**. Au contraire, les chaînes (pause + md) rejettent entièrement l'absence de contour et le contour de focalisation. Des différences significatives de durée existent entre les durées de pauses en fonction des contours qui les précèdent ($F(4, 244) = 3.154, p < 0.05$). Le contour terminal (marginal) a la durée la plus élevée, les contours de suspens et de focalisation ont des durées similaires relativement inférieures (2.505 et 2.504 respectivement) tandis que le **contour continuatif et l'absence de contours** sont suivis par des pauses significativement **plus courtes** (e.g. la moyenne du contour continuatif siège à 260.53 ms. ($\mu = 2.42, \sigma = 0.35$)).

Les chaînes (pause + dis), à l'instar des chaînes (dis + pause) sont typiques du discours spontané. Les pauses précédant une (dis)fluence sont les plus courtes du groupe (de manière non significative). Elles sont associées aux syllabes antéposées porteuses d'un **contour terminal ou de suspens**. La différence entre les durées des pauses en fonction du contour qui les précède est marginale ($F(4, 110) = 2.616, p = 0.039$). Ainsi, bien que l'ANOVA soit en-deçà du seuil de chance, le test post-hoc de Tukey DSH ne permet pas d'établir des différences significatives entre les groupes.

Enfin, les plus courtes durées ont été attestées lorsque **la pause s'actualise au sein d'un *interregnum***. Ces durées sont particulièrement associées à un **contour de suspens** les précédant et négativement associées à une absence de marquage ou à un contour terminal. Il existe des différences de durée des pauses au sein d'un *interregnum* au regard des contours intonatifs les précédant ($F(4, 363) = 4.69, p < 0.001$). Deux groupes se distinguent : **de manière atypique**, les durées des pauses précédées d'un contour de **focalisation** ou **de suspens** sont significativement **plus élevées** – respectivement 446.88 ms. ($\mu = 2.65, \sigma = 0.35$) et 378.61 ms. ($\mu = 2.58, \sigma = 0.29$), que les pauses précédées par un contour terminal (286.37 ms. ; $\mu = 2.46, \sigma = 0.41$) ou par une syllabe non marquée ou porteuse d'un contour continuatif (222.83 ms. ; $\mu = 2.35, \sigma = 0.35$).

La Figure 4.20 résume la densité d'informations présentée ci-avant. Elle reprend l'amorce de croisement entre les différentes variables : durées des pauses, contours intonatifs, variation situationnelle, et surtout, combinaison des pauses au sein des chaînes de (dis)-fluences.

Le coin supérieur gauche symbolise les groupes de chaînes de cooccurrences aux durées pausales les plus grandes. Il s'oppose au coin inférieur droit, regroupant les chaînes de cooccurrences de durée inférieure. Les carrés gris regroupant les chaînes permettent d'opposer **quatres groupes aux durées significativement différentes**. En leur sein, il n'existe pas de différence de durée significative entre les pauses. Chaque chaîne se voit attribuer des contours, eux-mêmes **positionnés en fonction du degré d'allongement** provoqué sur la pause leur succédant. En périphérie droite, on retrouve les **contours associés positivement** à une chaîne. En périphérie gauche les contours associés négativement à celle-ci. Notons enfin que seuls les contours dont les résidus dépassent un seuil de +2/-2 ont été repris dans ce schéma, et qu'un double signe positif ou négatif « ++/- - » rappelle un **résidu relativement important**.

Par exemple, la **pause seule** est fortement associée à la présence de **syllabes non marquées** (++NA) et de **contours terminaux** (+T) sur la syllabe antéposée. Le contour terminal entraîne une pause significativement **plus longue** que la syllabe non marquée. Elle est négativement associée aux contours de suspens et continuatif. Ces derniers sont liés à une durée de pause inférieure à celle des syllabes non marquées, mais non significativement différents entre eux, d'où leur alignement sur l'axe horizontal.

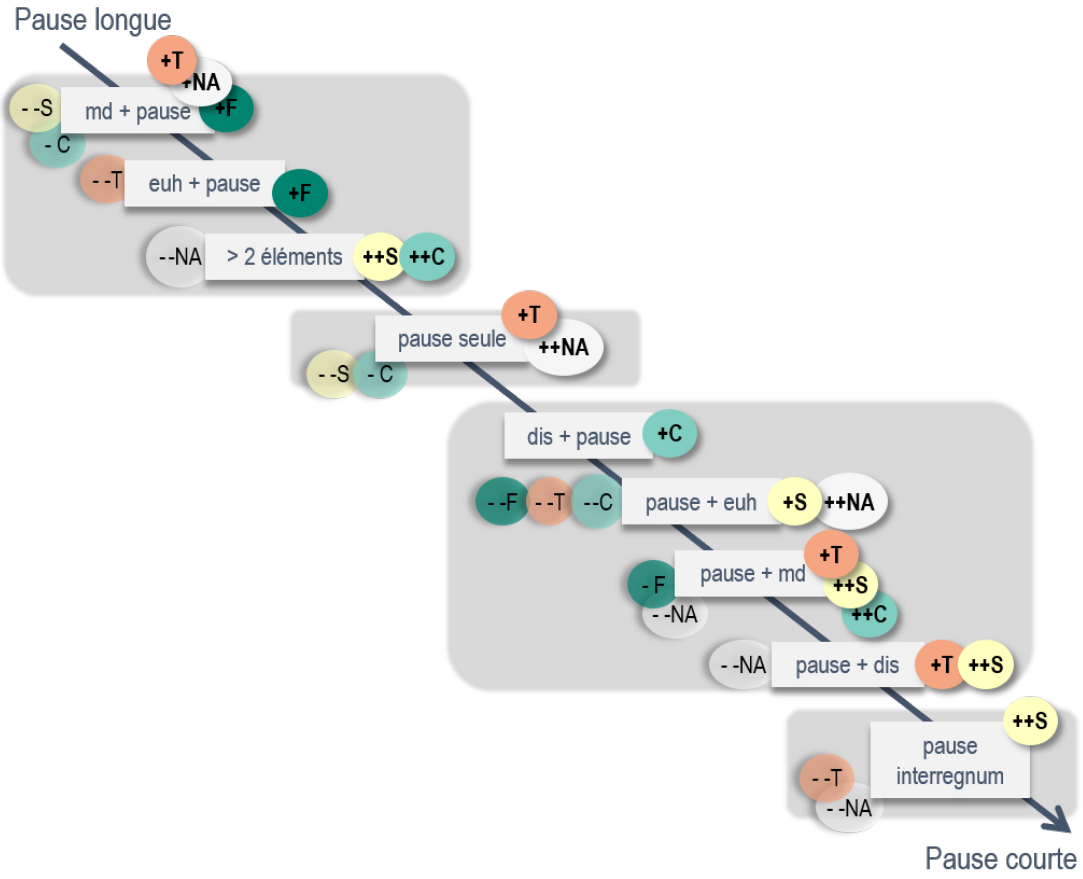


FIGURE 4.20 – Schématisation de la durée des pauses en fonction des chaînes de cooccurrences et des contours intonatifs des syllabes précédentes.

Nous aurions pu y ajouter des éléments des sections antérieures. D'aucun peut supposer qu'en **périphérie gauche** du schéma nous aurions tendance à trouver davantage de **pauses inter-rectionnelles** et non intra-ga, et des **contours terminaux** plutôt que continuatifs. Cependant, les interactions entre les différentes configurations sont nombreuses et méritent une analyse plus fine. Nous proposons donc dans la dernière partie de ce chapitre une méthode de **modélisation multifactorielle reprenant l'ensemble des paramètres**.

4 Conclusions autour de la rupture et de la continuité des pauses silencieuses

4.1 Modélisation multifactorielle des durées pausales

Cette section propose une modélisation multifactorielle de la durée pausale. Avant d'appliquer un algorithme de forêt d'arbres décisionnels de classification et de régression, nous nous arrêtons sur deux croisements des durées pausales permettant de synthétiser, dans une approche descriptive, les sections antérieures.

Les sections 3.3 et 3.4 s'attardaient sur l'interaction de la durée pausale avec les unités syntaxiques et les contours intonatifs. Chaque fois, la durée pausale a aussi été envisagée en fonction de la chaîne de (dis)fluence(s) présente dans le contexte immédiat de la pause silencieuse. Nos analyses ont ainsi confirmé d'une part que **les durées des pauses augmentent avec le degré de frontière syntaxique**, et d'autre part que **la durée des pauses aux contours terminaux est significativement plus importante que celles situées après les contours continuatifs**. Mais peut-on également démontrer que tous les contours intonatifs se comportent de manière similaire au regard du degré de frontière syntaxique ? En d'autres termes, **certains contours sont-ils plus sensibles au marquage de frontière(s) syntaxique(s) que d'autres** ?

Ainsi, dans une approche purement prosodico-syntaxique des pauses, sans considérer leur environnement au sein de (dis)fluence, la Figure 4.21 illustre les durées pausales en fonction de leur localisation syntaxique et du contour intonatif précédant la pause.

En apparence, tous les contours voient la **durée de la pause subséquente augmenter avec le degré de frontière syntaxique**. C'est le cas pour le contour de suspens, pour lequel on observe des durées de pauses significativement différentes en fonction de la localisation de la pause ($F(3, 414) = 39.16, p < 0.001$). Celles-ci suivent l'ordre habituel avec la durée la plus longue en position inter-rectionnelle (538 ms. ; $\mu = 2.73, \sigma = 0.24$) et la plus courte lorsque la pause est en position intra-ga (201 ms. ; $\mu = 2.30, \sigma = 0.28$).

Toutefois, les **contours continuatif, terminatif** et l'**absence de contour**, bien que présentant des différences significatives au sein de leurs conditions, ne permettent de distinguer que **trois groupes de durée**. En effet, pour ces trois contours, il n'existe **pas de différences significatives de durée de la pause en position inter-rectionelle ou inter-séquentielle**.

Si la **position syntaxique de la pause** présente est bien significativement **plus brève** lorsqu'elle est précédée d'un **contour de focalisation** ($F(3, 135) = 18.04, p < 0.001$), il n'existe pas de différence entre tous les niveaux de facteurs. Un test post-hoc de Tukey DSH indique que la différence de durée entre les pauses dans le contexte d'un contour de focalisation ne permet de distinguer que **deux groupes de durée**. Les **pauses courtes** (intra-ga et inter-ga) – respectivement 128 ms. ($\mu = 2.11, \sigma = 0.13$) et 256 ms. ($\mu =$

$2.41, \sigma = 0.34$), s'opposent aux **pauses longues** (inter-séquences et inter-rections) – respectivement 457 ms. ($\mu = 2.66, \sigma = 0.24$) et 575 ms. ($\mu = 2.76, \sigma = 0.31$), sans différence significative au sein même de ces deux groupes où la pause est précédée d'un contour de focalisation.

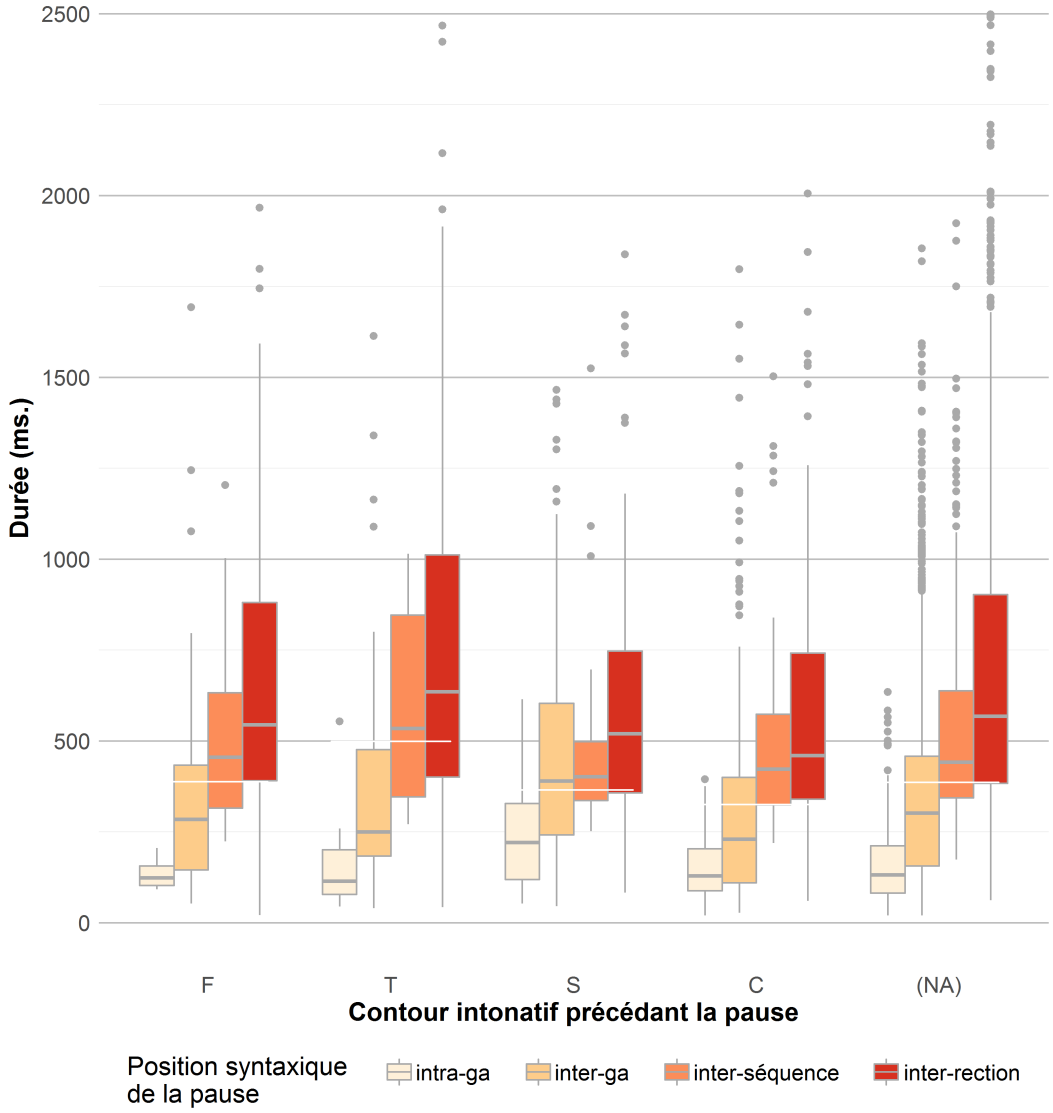


FIGURE 4.21 – Diagramme en boîte à moustaches de la distribution des durées pausales en fonction de la localisation syntaxique de la pause et du contour intonatif précédant la pause.

Enfin, la Figure 4.22 reprend le croisement de tous les paramètres présentés. Nous prenons le temps de souligner quelques points révélateurs de ces dernières statistiques descriptives[26] avant de nous focaliser sur la modélisation inférentielle.

26. Afin de ne pas perturber la lisibilité du texte, les annexes E section 1 reprennent les tableaux de statistiques descriptives liés à la Figure 4.22.

Les chaînes de pause *interregnum* ($n = 368$) sont majoritairement **inter- et intra-ga** (respectivement 4%+ et 30.7%). Elles s'actualisent en priorité après des **syllabes non marquées** (56.5%) ou des **contours de suspens** (25%). Le contour de focalisation accompagne en apparence des pauses de durée plus élevée en position inter-rectionnelle, celles-ci ne sont en réalité qu'au nombre de trois. Par contre, la durée moyenne des pauses **augmente effectivement en position inter-ga lorsque précédées par un contour de suspens**. La moyenne de la durée pausale croît de 210.01 ms. en position intra-ga à 638 ms. en position inter-rectionnelle.

Les chaînes > à 2 éléments ($n = 365$) sont au contraire majoritairement **inter-rectionnelles** (52%) mais sont précédées cette fois d'un **contour continuatif** (17%) ou d'une **absence de marquage** (17%). Ainsi, les chaînes de cooccurrences à deux éléments ont tendance à s'actualiser dans une **rupture syntaxique mais sans frontière intonative**. Le locuteur s'arrête au niveau syntaxique mais ne l'avait pas anticipé au niveau intonatif.

Les chaînes (dis + pause) ($n = 104$) sont moins nombreuses. La moitié d'entre elles s'actualisent en **position inter-ga** (49%) et avec un contour intonatif **non marqué** (68% des cas) ou continuatif (17.31%), à l'instar des chaînes de 2 élements. Ainsi, il semblerait que cette chaîne marque moins de rupture syntaxique, et soit encore moins marquée intonativement. La durée moyenne globale de ces pauses est également plus faible, ce qui entraîne davantage de continuité dans le flux discursif.

Les chaînes (pause + dis) ($n = 115$) suivent relativement le même schéma que les chaînes (dis + pause) : elles sont majoritairement en **position inter-ga** (48%) et précédées d'une syllabe **non marquée** ou d'un contour continuatif (respectivement 59% et 11%). Cependant, elles sont plus associées à un **contour de suspens** : le locuteur laisse en suspens l'unité intonative, l'interrompt syntaxiquement et effectue sa reprise par une pause (relativement courte) et une (dis)fluence.

Les chaînes (pause + *euh*) ($n = 146$) sont caractéristiques de la **position intra-ga**. Leur comportement se rapproche en cela, des chaînes où la pause prend la position d'*interregnum*. Cependant, elles peuvent aussi être associées à une profonde rupture syntaxique (25% en **position inter-rectionnelle**), sans doute plus abrupte : les syllabes des pauses précédées suivies d'un *euh* ne sont **jamais marquées par un contour terminal** et sont dans 88% des cas précédées d'une **absence de frontière intonative** sur la syllabe précédant la pause.

Les chaînes (*euh* + pause) ($n = 211$) sont au contraire rarement attestées en position intra-ga (5.21%). On les retrouve, à l'instar des (md + pause) (avec lesquelles elles partagent les mêmes durées pausales), majoritairement en **position inter-ga** (48%) ou **inter-rectionnelle** (33.6%). Intonativement, leur frontière gauche est soit non marquée, soit marquée par un contour continuatif (respectivement 71% et 14%). Il semblerait donc que ces deux configurations (*euh* + pause) et (md + pause) soient davantage un **signe de continuité** que de rupture.

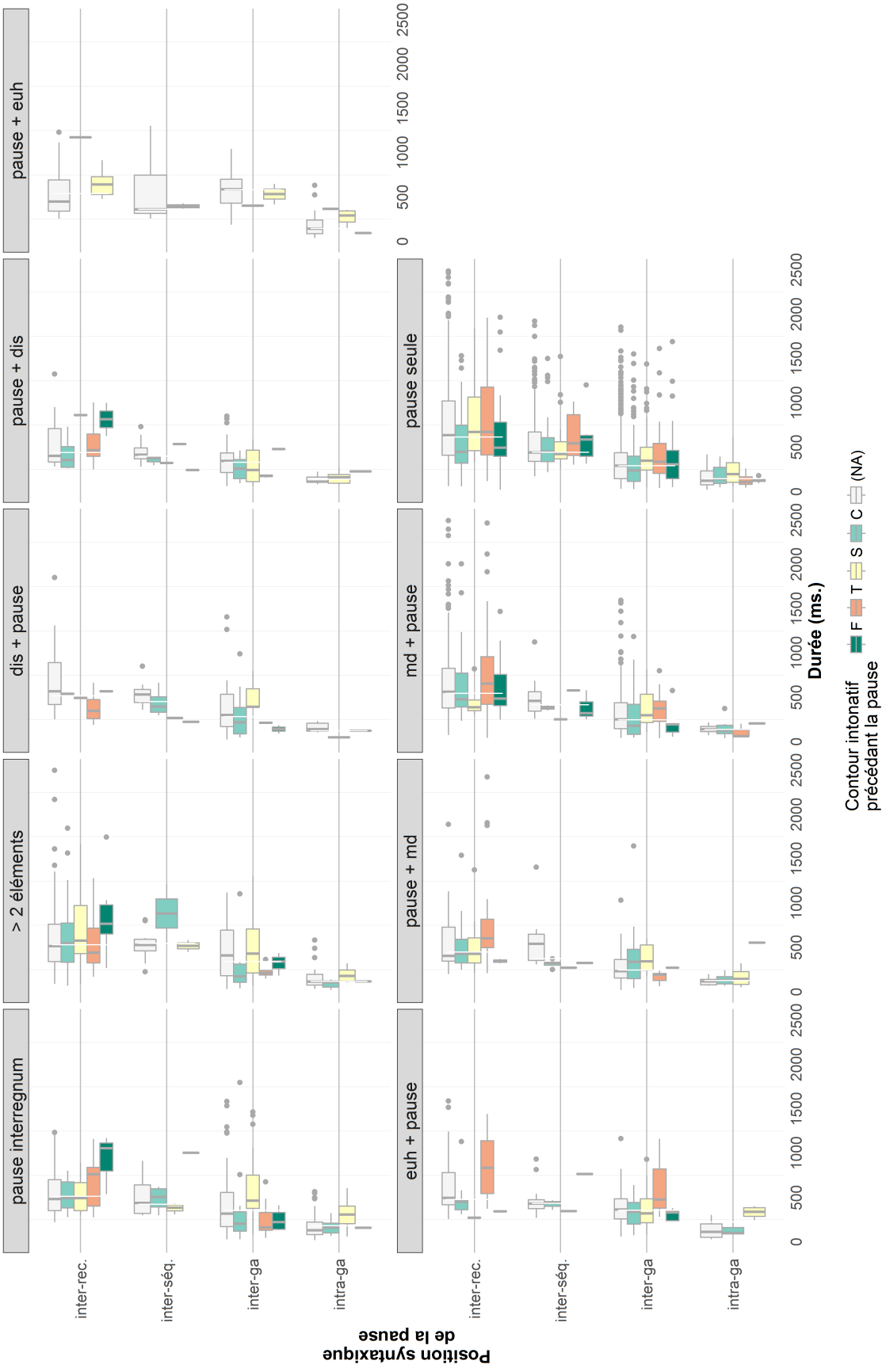


FIGURE 4.22 – Diagramme en boîte à moustaches de la distribution des durées pausales en fonction des contours précédant la pause et de la localisation syntaxique de la pause, regroupé par chaîne de cooccurrences.

Les chaînes (pause + md) ($n = 249$) ont des profils d'actualisation diversifiés. On les retrouve majoritairement en **inter-rection** ou inter-ga (45% et 30%) mais dans des proportions non négligeable également en **intra-ga** (16.5%) et inter-séquence (10%). En d'autres termes, la chaîne (pause + md) **n'est incompatible avec aucune position syntaxique**. Si moins de la moitié des occurrences de pauses sont précédées de syllabes non marquées (43%) ou continuatives (27%), une partie des actualisations sont associées au contour **de suspens** (16%) et terminal (11%).

Les chaînes (md + pause) ($n = 684$) sont majoritairement en **position inter-rectionnelle** (62%) ou inter-ga (31%) à l'instar de la configuration (*euh* + pause). Elles peuvent être interprétées comme des **outils de continuation intonative** (14% de C et 71% de NA). La configuration accepte cependant la présence d'un contour terminal sur le md lui-même (10%)

Les pauses seules ($n = 2593$) s'actualisent le plus nettement en **contexte inter-ga** (51%). Seules 30% d'entre elles se trouvent en position inter-rectionnelle. On peut présumer que cette dernière localisation syntaxique **invite à davantage de md ou d'alternance de tour**.

Par ailleurs, si nous avions en première section de ce chapitre mis en avant la variation situationnelle, nous nous sommes ensuite uniquement focalisés sur les croisements de la durée des pauses avec les facteurs intono-syntaxiques autour des pauses et des (dis)-fluences (sections 1.3 et 3.1). Nous ré-intégrons désormais les facteurs « situations de parole » dans une approche descriptive, afin de sélectionner les variables situationnelles pertinentes à la modélisation (Figures 4.23 et 4.24).

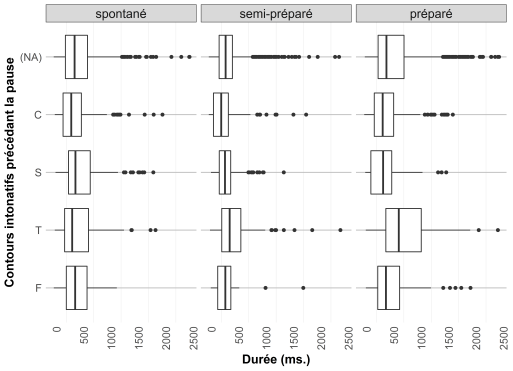


FIGURE 4.23 – Distribution des enregistrements en fonction du degré d'élicitation.

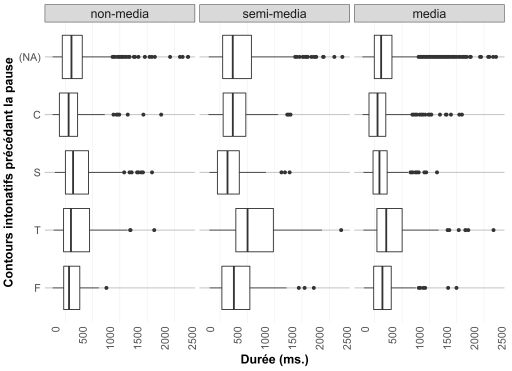


FIGURE 4.24 – Distribution des enregistrements en fonction du cadre d'interaction.

Dans une approche inférentielle, nous entreprenons premièrement une modélisation des durées pausales à partir des traits situationnels et des paramètres linguistiques caractérisant l'environnement des pauses : **cooccurrence avec des disfluences, localisation de la pause par rapport aux frontières des unités syntaxiques, contour intonatif précédant la pause**. L'application d'un algorithme de **forêt d'arbres décisionnels** de classification et de régression permet de sélectionner les variables à partir de leur

contribution informationnelle. Une modélisation à partir des variables « situation de parole », « cooccurrence avec une disfluence », « position syntaxique », « contour prosodique précédant » permet d'expliquer un taux honorable de 38,82% de la variance des durées pausales. Néanmoins, la permutation de la variable « situation de parole » avec celle des traits situationnels ne diminue le pouvoir prédictif du modèle que de 2% (36% de variation prédite). La Figure 4.25 présente la contribution de chaque variable à une forêt d'arbres décisionnels (*nombre d'arbres* = 1000). Nous pouvons conclure que c'est la syntaxe qui prédit le mieux la durée d'une pause silencieuse, suivie de la cooccurrence avec les (dis)fluences, du trait situationnel de degré d'orientation médiatique, du contour intonatif précédant la pause et du degré de préparation.

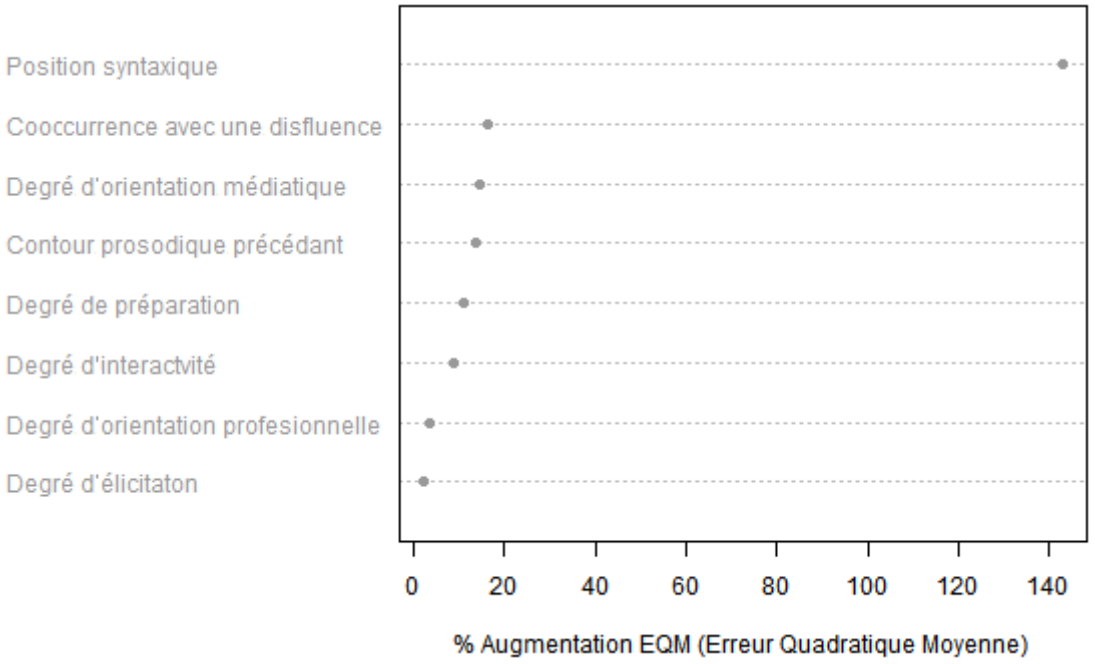


FIGURE 4.25 – Contribution des variables (ajustement à la forêt d'arbres décisionnels).

De manière plus détaillée, l'arbre de décision illustré par la Figure 4.26 montre que les **pauses les plus longues** sont situées entre **les unités de rection dans la situation de parole préparée**. Cet effet de la syntaxe est moins marqué en parole semi-préparée et spontanée. En conséquence, les pauses aux frontières de rection en parole semi-préparée ou spontanée ont la même durée que les pauses aux frontières de séquence en parole préparée. D'autre part, les **pauses les plus brèves** se situent à l'intérieur des **unités syntaxiques minimales** (ga), quel que soit le degré de préparation de la situation de parole. Entre ces deux extrêmes, se trouvent les pauses qui cooccurrent avec des (dis)fluences. Les nœuds 11 et 12 séparent ainsi respectivement le groupe [1] (*euh*+ pause), (pause + *euh*), (pause *interregnum*) et (chaîne $n > 2$) du groupe [2] (pause + dis), (dis + pause), (pause + md), (md + pause), (pause seule), ce dernier ensemble caractérisant des pauses plus brèves.

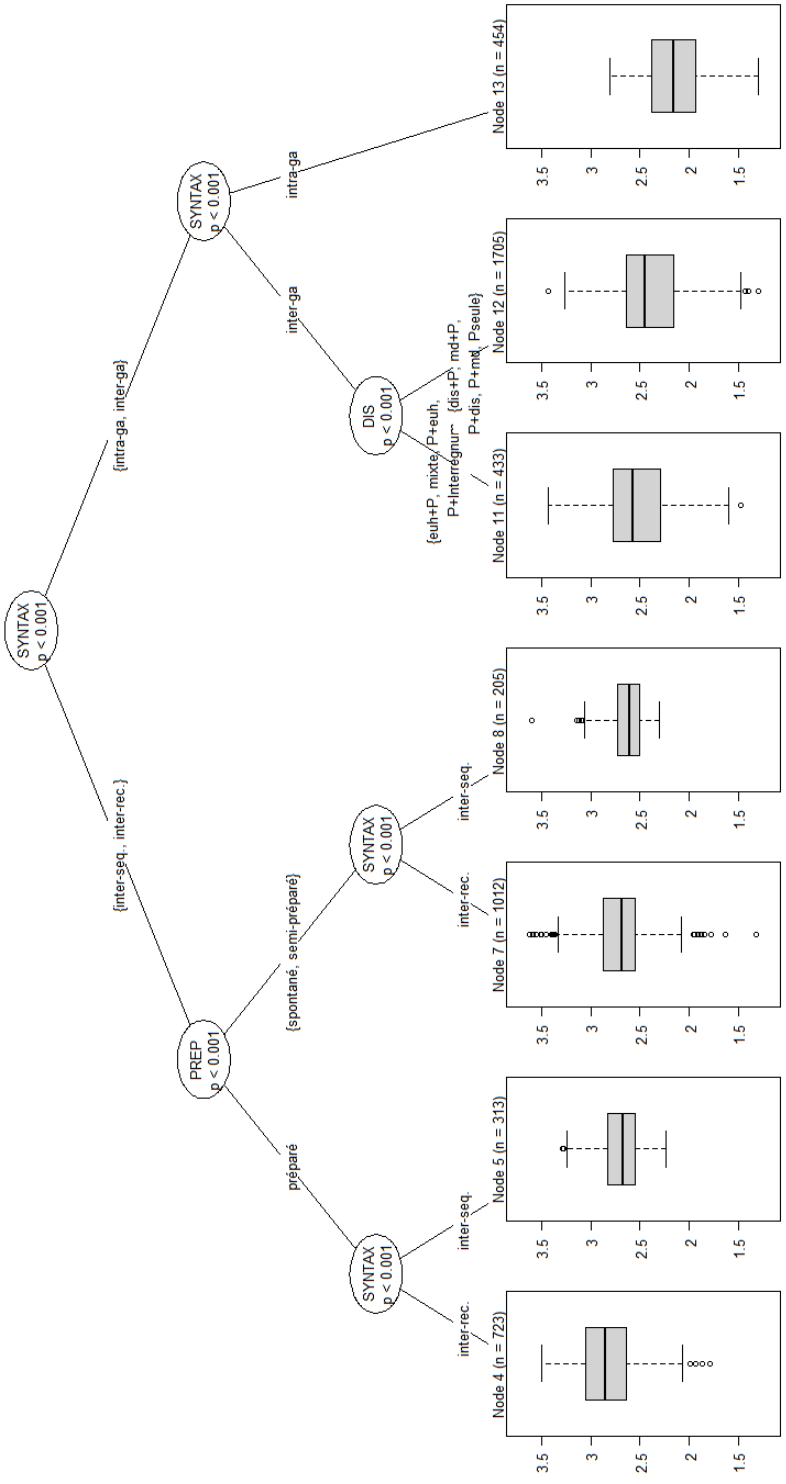


FIGURE 4.26 – Arbre de décision des durées pausales (log10 ms.) en fonction de la position syntaxique, du degré de préparation et de la cooccurrence avec des (dis)fluences.

Les pauses seules se trouvent aux deux extrémités de l'arbre (nœuds 4 et 13). Les pauses seules les plus longues sont situées entre les unités syntaxiques maximales (inter-rections), tandis que les plus courtes le sont entre ou dans les unités syntaxiques minimales (inter-ga et intra-ga). Sans perdre en précision, nous avons décidé de représenter ces données selon deux variables uniquement (position syntaxique et cooccurrence avec des disfluences) (Figure 4.27). Cette dernière option permet de montrer que, **à position syntaxique égale, les pauses avant ou après une disfluence sont plus brèves que les pauses seules**. Cela pourrait s'expliquer par le fait que la disfluence en elle-même offre déjà un **temps de planification**. Cependant, cette figure ne met pas systématiquement en évidence un comportement différent des pauses avant ou après disfluence (e.g. (*euh* + pause) /vs/ (pause + *euh*)), mais montre plutôt que les pauses dans un environnement de *euh* ou en *interregnum* tendent à être un peu plus longues que les pauses dans l'environnement d'un md ou d'une disfluence (dans la position inter-ga).

Enfin, la Figure 4.28 prend en compte également les contours intonatifs précédant la pause. Celle-ci permet de mettre en évidence que les contours intonatifs sont surtout utiles lorsqu'il s'agit de **distinguer les durées pausales entre les groupes syntaxiques minimaux** (inter-ga et intra-ga). Dans le cas des inter-ga (nœuds 14 à 16), les pauses précédées d'un contour continuatif sont toujours plus longues que celles précédées d'autres contours ou non marquées par un contour. Dans le cas des groupes intra-ga (nœuds 18 et 19), les syllabes suivies de pauses non marquées par les contours sont de durées plus réduites. Cette classification et régression s'expliquent en partie par le fait qu'a**ssez peu de pauses et (dis)fluences sont attestées en position intra-ga**.

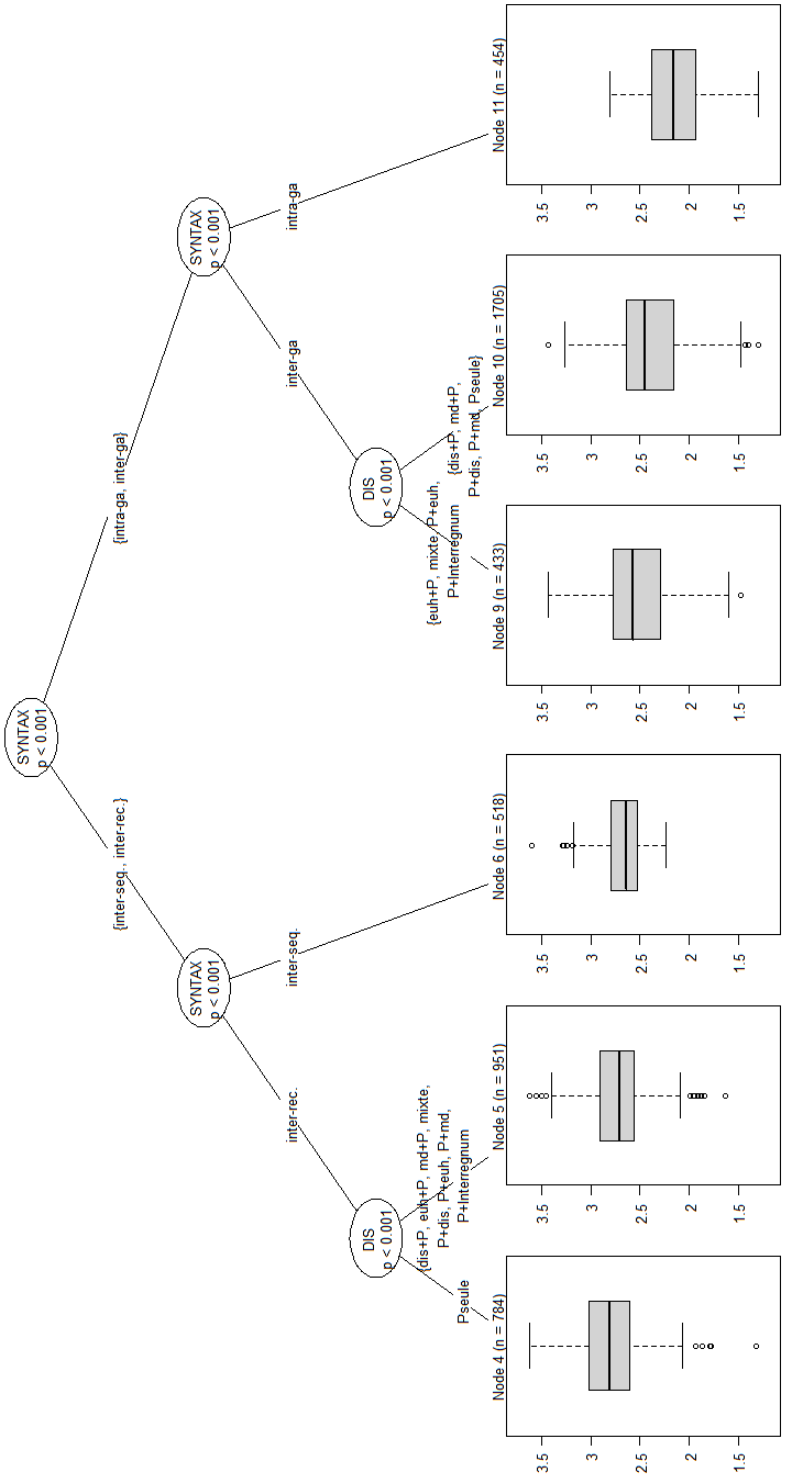


FIGURE 4.27 – Arbre de décision des durées pausales (log10 ms.) en fonction de la position syntaxique et de la cooccurrence avec des (dis)fluences.

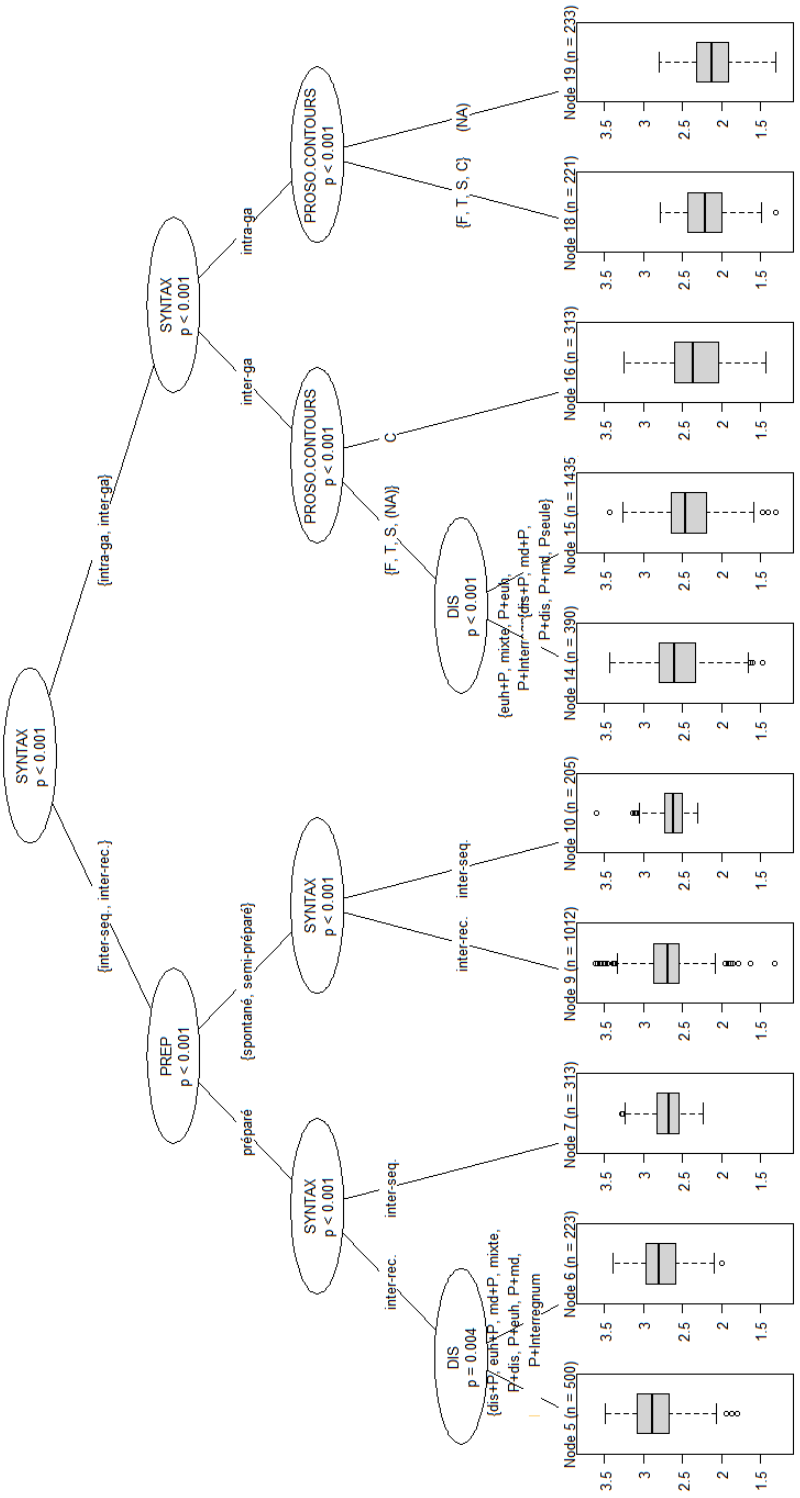


FIGURE 4.28 – Arbre de décision des durées pausales (log10 ms.) en fonction de la position syntaxique, de la cooccurrence avec des (dis)fluences, du type de contour prosodique et du degré de préparation.

4.2 Synthèse et discussion

Ce chapitre contient une étude à la fois **méthodologique et de modélisation** de la distribution et de la durée des pauses silencieuses aux frontières d'unités syntaxiques et/ou intonatives, et en cooccurrence avec des (dis)fluences. Les données analysées concernent la distribution des pauses silencieuses intratour dans 14 situations de parole différentes.

Ce chapitre a mis l'accent sur l'importance des **questionnements méthodologiques** ayant trait à la description des distributions des pauses et à leurs catégorisations fonctionnelles. A cet égard, nous avons **refusé les seuils définis** par un à priori linguistique et nous avons mis en évidence l'**impact des types de transformations** sur la forme des distributions, notamment via une méthodologie de classification des pauses en *clustering*.

En effet, en raison de la **variabilité des débits et des durées pausales**, nous ne préconisons pas d'appliquer des seuils fixes de durée pour rejeter certaines données (e.g. comme on retrouve souvent dans la littérature, un seuil inférieur à 200 ms. ou supérieur à 500 ms.). Les seules pauses exclues dans nos analyses le sont sur la base d'un **raisonnement statistique** (i.e. exclusions des données aberrantes liées au traitement des corpus). Cette méthodologie permet également d'**exclure un nombre réduit de données** (moins de 4% des pauses).

Un des objectifs principaux étant de positionner les pauses sur le continuum fluence-disfluence, nous avons opté pour une **granularité fine de l'observation des pauses en contexte**. Dans une approche **monofactorielle**, nous avons tour à tour enquêté sur la variation des durées en fonction des situations et des traits situationnels (section 3.1), des cooccurrences à des (dis)fluences (section 3.2), des frontière syntaxiques (section 3.3) et des contours intonatifs précédant la pause (section 3.4).

Ces analyses ont mené à une **modélisation multifactorielle des durées pausales**. Cette méthode a l'avantage de faire ressortir les grandes tendances de durée pausale sans faire précéder à priori un facteur plutôt qu'un autre.

Nous proposions aux chapitres 1 à 3 d'envisager le statut d'un phénomène sur le continuum fluence-disfluence en **fonction de deux axes** : (1) son influence sur la rupture ou continuité discursive (prosodique, syntaxique) et (2) son adéquation aux tendances générales, aux « normes discursives ».

Les résultats de la section 3.1 de ce chapitre montrent que la distribution de la durée pausale varie **tant entre les situations de parole qu'en leur sein**. Nous en concluons que le caractère unimodal, bimodal voire multimodal d'une distribution devrait être établi **au minimum par situation de parole, et parfois par locuteur**. Nous avons effectivement mis en en évidence que certains genres sont **plus standardisés** que d'autres au niveau de leur distribution pausale (e.g. le journal parlé ou l'homélie). La décision de l'analyste doit être avisée. Si la variabilité intragenre est faible, l'on pourra s'accorder

sur une représentation uni-, bi- ou trimodale par genre. Si la **variabilité intragenre** entre les locuteurs est élevée, il faudra considérer une distribution par locuteur et/ou l'**approche par modèles mixtes**[27].

En conséquence, nous posons que les phénomènes pausaux « hors norme » dans ce type de discours sont plus difficilement classifiables comme « disfluents » étant donné que la « **norme idéale** » y est plus floue. Le **hors-norme** peut cependant toujours être porteur de rupture : on remarque ainsi que dans le discours politique une tendance à la rupture est portée par les pauses et le contour continuatif : de manière potentiellement surprenante, le locuteur continue alors qu'on se serait attendu à une rupture. Ce comportement rend les pauses et les chaînes de cooccurrence de (dis)fluence(s) **particulièrement saillantes**. Par exemple, les pauses intraconstituants (c'est à dire intra-ga + intra-séquence) non structurantes marqueraient la recherche de mots et seraient donc **perceptivement plus marquées**. Nous émettons donc l'hypothèse qu'elles sont jugées plus **disfluentes** si, dans le discours, il y a **besoin de marquage**.

Ensuite, l'analyse en section 3.2 de la distribution des pauses silencieuses cooccurrentes avec des contextes dits fluents ou disfluents fait apparaitre que **78% des pauses** peuvent être qualifiées de « **structurantes** » ou « **grammaticales** », alors que **22% des pauses** sont dites « **d'hésitation** » se situant à l'intérieur de chaînes de disfluences ou immédiatement après un *euh* ou une autre disfluence. Plus spécifiquement, il existe des associations fortes entre certains types de chaînes et les situations de parole préparée, semi-préparée ou spontanée. Ainsi, les pauses structurantes (pauses isolées) sont prédominantes dans les situations de parole préparée (journaux parlés, homélies, discours politiques, etc.) et plus rares dans les styles non préparés (conversations informelles, narrations conversationnelles, etc.). Ces derniers se caractérisent plutôt par des configurations de pauses dans des **séquences mixtes** (combinant différentes (dis)fluences) ou dans une **séquence réparatrice**.

Par ailleurs, le degré de préparation de la situation de parole influence également la **localisation syntaxique des pauses** silencieuses (section 3.3). Plus précisément, les pauses internes aux constituants syntaxiques minimaux, qui **interrompent une unité cohésive**, sont moins fréquentes en parole préparée qu'en parole semi-préparée ou spontanée. Les pauses vraiment spécifiques à la parole préparée sont celles situées entre les unités syntaxiques intermédiaires (séquences), alors que les pauses entre unités maximales (de rection) sont les plus fréquentes en parole spontanée. Cette différence est à imputer au fait que **les unités syntaxiques sont plus longues et plus complexes en parole préparée** et plus brèves dans les situations spontanées (*infra.* note de bas de

27. Nous adoptons cette approche dans une majorité de nos modélisations perceptives. Voir par exemple chapitre 6 section 2.

page 20). Il pourrait donc s'agir d'un besoin d'insérer une pause à **intervalle temporel régulier**[28]. Enfin, la durée moyenne d'une pause augmente en fonction du degré de frontière syntaxique, quelle que soit la situation de parole ou la séquence fonctionnelle environnante.

La section 3.4 envisage l'interaction entre la durée des pauses et les **contours intonatifs des segments précédant la pause**. Elle montre que les pauses silencieuses ont tendance à être **plus courtes après un contour continuatif** (qu'un contour terminatif ou de suspens). Dans une logique de **fluence cognitive**, le contour continuatif appelle à la poursuite du discours et donc à un temps de pause relativement bas (sous peine d'alternance de tours). Au niveau de la cooccurrence, additionnellement, si la pause suivie d'un contour continuatif est longue, elle a également tendance à être accompagnée d'une **chaîne de cooccurrences de (dis)fluence(s)**. Enfin, au niveau de la cooccurrence et des frontières syntaxiques, les **pauses silencieuses structurantes intraconstituantes** ne sont pas accompagnées de marquage continuatif ou terminatif. En effet, les pauses structurantes intraconstituantes ne forment qu'**un seul constituant sur le plan segmental et prosodique** et ne sont donc pas en périphérie de frontières.

L'**analyse multifactorielle** à partir des trois types de paramètres retenus – situationnels, syntaxiques et liés aux (dis)fluences – nous permet de conclure que c'est la **syntaxe qui prédit le mieux la durée d'une pause silencieuse**, suivie de la cooccurrence avec les (dis)fluences, du degré de préparation médiatique, du contour prosodique précédant la pause et du degré de préparation.

Certaines de ces observations empiriques ont déjà été confirmées par des **études de perception**. Par exemple, Simon & Christodoulides (2016) ont montré que les pauses silencieuses, même longues, ne sont pas perçues comme des indices de frontières lorsqu'elles sont situées à l'intérieur d'un groupe syntaxique, à cause de l'incomplétude syntaxique qui inhibe la perception de la pause. Il serait intéressant de tester plus spécifiquement la **perception des pauses environnant les (dis)fluences** (Duez 1995), pour comprendre si elles augmentent ou diminuent la perception de ces marques et interrompent le traitement du flux discursif. Nous le testons en partie au chapitre 6 section 2.

Dans le chapitre suivant, nous nous attardons avec une méthodologie et une structure similaire à ce premier chapitre sur l'analyse des pauses pleines et des allongements. Nous y ajoutons l'observation des contours internes aux *euh* et aux allongements. En outre, nous proposons d'affiner un regard critique sur la détection même de ces phénomènes en fonction de leur contexte ainsi que de leur durée en évaluant l'apport de différents types d'annotation : automatique, *online*, et *offline*.

28. Nous ne nous focalisons que très rarement sur le *tempo* dans cette thèse. Cependant, nous faisons intervenir la distance des (dis)fluences entre elles pour évaluer l'impact sur le continuum fluence-disfluence au sein d'un discours.

Chapitre 5

Evaluation de la (dis)fluence des pauses pleines et des allongements

Ce chapitre décrit les régularités distributionnelles **acoustiques, prosodiques, syntaxiques** et **situationnelles** des **pauses pleines** et des **allongements**. Ces marques constituent en moyenne 5% du flux verbal. Elles sont reconnues comme des caractéristiques inhérentes des productions discursives (Eklund 2004 ; Fox Tree 1995). Elles peuvent présenter des distributions et des **patrons de cooccurrences** réguliers. En 1972 déjà, pour Grosjean et Deschamps (1972:51), 60% des pauses pleines et des allongements sont **suivis ou précédés de pauses silencieuses**. Les (cooccurrences de) marques seraient tellement intégrées, que les interlocuteurs jugeraient **plus disfluents** les segments de parole contenant des hésitations sans pause silencieuse subséquente qu'avec (Moniz, Trancoso & Mata 2009). Toutefois, peu d'études ont examiné si la perception même des hésitations est **univoque**, et si celle-ci est influencée par la **présence d'hésitations environnantes**. Nous traitons la production et la perception des pauses pleines et des allongements de manière conjointe vu leurs potentielles **similitudes acoustiques** et leurs **distributions complémentaires**.

La première partie du chapitre est consacrée à l'**observation des corpus** sur base de nos annotations (section 1). La seconde regroupe plusieurs études sur la **confrontation de méthodologies** d'annotation des pauses pleines et des allongements ; elle étaie les éléments influençant leur annotation et leur perception (section 2). Enfin, nous proposons une discussion (1) sur les fonctions et les formes des pauses pleines et des allongements, (2) la continuité ou la rupture discursive à laquelle ils peuvent être associés et (3) leur relation avec les normes situationnelles, et les attentes nécessaires à l'évaluation du continuum fluence-disfluence (section 3).

1 Analyses des données de corpus

Cette section compare les **durées** et les **distributions** des pauses pleines et des allongements ainsi que leur(s) interaction(s). Elle permet d'évaluer le **degré de rupture** ou de **continuité discursive** que leur actualisation entraîne. Comment les durées et les courbes des pauses pleines et les allongements *dits d'hésitation* se distribuent-ils au sein du corpus ? *In fine*, confirmons-nous des **distributions complémentaires** de ces deux marques et de leurs cooccurrences [1] ?

Nous reconnaissons qu'**il existe un continuum** entre le schwa post-tonique allongé et le *euh* intonativement détaché (e.g. *une petite* {*fill**e:*** / *fill•**e*** / *fille* ***euh***}). Les données analysées sont tributaires de la **perception** et des **choix des linguistes**, des annotateurs de corpus qui ont assigné l'un ou l'autre **extrême de ce continuum** (i.e. allongement /vs/ pause pleine) [2]. Ainsi, si nous étudions les pauses pleines et les allongements comme deux **catégories séparées**, nous ne distinguons pas les différents **types de pauses pleines** entre elles (e.g. allongement du schwa *euh* /vs/ allongement d'une voyelle et d'une nasale *euhm*) [3].

Une fois précisé les objectifs et les hypothèses de notre étude (1.1), nous ré-appliquons dans la première section les étapes d'analyse du chapitre 4. Nous sommes guidée par la série de questions suivantes : peut-on distinguer plusieurs distributions de pauses pleines et d'allongements en fonction de **leurs durées** au regard des situations (1.2) ? Comment ces marques s'associent-elles pour former des **chaînes de cooccurrences** (1.3) ? Quelles associations observe-t-on au regard de la **segmentation syntaxique** et des **tours de parole** (1.4) ? Enfin, quelles sont les **formes intonatives** et les associations fonctionnelles des pauses pleines et des allongements (1.5) [4] ?

1.1 Objectifs et hypothèses

Nombre d'études sur les pauses (pleines) incluent les allongements comme objet d'étude comparable étant donné leurs **similarités acoustiques et fonctionnelles**. Eklund (2004) traite simultanément les pauses pleines et les allongements car ils ne dépendent tous les deux que des **paramètres de fréquence fondamentale et de durée**. Selon Clark & Tree (2002) cependant, leur **distribution contextuelle** et leur(s) fonction(s) diffèrent. Schématiquement, les allongements reflèteraient une **mise en attente discursive** actualisée (au niveau du processus phonologique), tandis que les pauses pleines

1. A l'instar de Moniz, Mata & Viana (2007) mais à l'opposé de Clark & Tree (2002).
2. La deuxième partie de ce chapitre abordera pertinemment cette question en confrontant plusieurs méthodologies d'annotation (section 2).
3. Nous renvoyons le lecteur au chapitre 2 section 2.3.1 pour une discussion théorique et au chapitre 3 section 2.1 pour une présentation de la méthodologie d'annotation des pauses pleines et des allongements.
4. Peut-on, par leur forme acoustique, distinguer les pauses pleines et les allongements intégrés dans une chaîne de (dis)fluences de ceux s'actualisant de manière isolée ?

seraient le signal d'un **retardement à venir** (au niveau lexical). Le type de retardement varierait notamment en fonction de la **forme de la pause pleine** (Candea 2000a:90-95).

Au niveau discursif, la littérature oppose **les fonctions référentielle et rétroactive** des pauses pleines aux **fonctions modale et prospective** des allongements. Les *euh* signaleraient la difficulté à organiser la relation entre « ce qui est dit » et « ce qui va être dit ». Les *euh* sont alors des **ligateurs discursifs**, témoins du processus d'encodage. Ils peuvent indiquer l'inachèvement, mais ils n'induisent pas nécessairement une rupture discursive et/ou la perception d'une *dis*fluence. Ils marquent une **intention de continuation** plus que de clôture. Les allongements opéreraient une distinction énonciative avec les *euh*. Leur fonction modale et prospective s'incarne dans le fait qu'ils s'attacheraient à **la gestion immédiate** du discours, inhérente à un **travail de planification** et de conceptualisation (e.g. ré-évaluation du référent) ou de **maintien de la parole** (Lacheret-Dujour & Beaugendre 1999:53 ; Dister & Lacheret-Dujour 2011:17 ; Morel & Danon-Boileau 1998:82 ; Candea 2000a:91) [5].

Sur base de la littérature, nous n'émettons pas l'hypothèse d'une distribution bimodale des durées des pauses pleines et/ou des allongements. Au niveau discursif, rien ne laisse présager une détection et une classification de pauses pleines « **courtes** » /vs/ « **longues** ». Cependant, ces durées seraient influencées en fonction de la **cooccurrence avec d'autres (dis)fluences** :

1. Les différences de durées de pauses pleines et d'allongements en fonction de leur cooccurrence avec d'autres (dis)fluences seraient statistiquement significatives.
2. Les pauses pleines et les allongements ont en majorité une **durée supérieure lorsqu'ils sont en cooccurrence avec d'autres (dis)fluences** (à l'instar des observations de Shriberg (1999)).
3. Plus précisément, les durées des pauses pleines en chaînes (*euh* + pause) seraient plus remarquables que celles des chaînes (pause + *euh*), étant donné que ces dernières seraient davantage **initiatrices d'un nouveau constituant**.

Au niveau de la forme de la courbe intonative interne des pauses pleines et des allongements, celle-ci tendrait majoritairement à reproduire une **courbe descendante linéaire droite** (Duez 2001b ; O'Shaughnessy 1992 ; Shriberg 1999). Elle peut cependant aussi s'actualiser en **une courbe ascendante ou plate** (Duez 2001b:39). Trois types de pauses pleines devraient être prépondérantes : (1) chute progressive, (2) chute

5. Ces éléments ont été abordés plus en détails au chapitre 2 section 2.3.1.

irrégulière, (3) chute suivie d'une montée soudaine (Guaïtella 1991)[6]. Pour des raisons techniques, nous n'aborderons toutefois pas la question de leur hauteur intonative par rapport au registre du locuteur[7].

Concernant l'interaction des **formes de contours prosodiques avec les groupes intonatifs**, nous ne postulons pas l'existence d'une distribution significativement différente des contours de pauses en fonction des groupes intonatifs (majeurs ou mineurs) (Duez 2001b:39). Néanmoins, les pauses pleines et les allongements seraient **associés à différents niveaux de regroupements intonatifs** :

- les allongements seraient majoritairement **à l'intérieur des unités prosodiques** ;
- au contraire, les pauses pleines seraient **aux frontières**, notamment en vertu de leur **interaction avec l'unité intonative** qu'elles ont tendance à **scinder en deux**.

Nous pouvons formuler des hypothèses similaires pour le positionnement des marqueurs **au regard des frontières syntaxiques** : on peut s'attendre à ce que les allongements dits d'hésitation s'actualisent davantage en **médiane d'unité** qu'en frontière. Au contraire, les pauses pleines, faisant le lien entre « ce qui a été dit » et « ce qui va être dit », pourraient s'actualiser en **frontière d'unités syntaxiques** qui appellent à être coordonnées (ga ou rection) ; ces endroits étant aussi ceux attestés par des études expérimentales (Roberts & Kirsner 2000) et de corpus (Schneider 2014).

Nous examinons également la relation entre la **forme prosodique et les durées** des marques. On émet notamment l'hypothèse d'une **indépendance totale** de la durée de la pause et des différences de fréquence fondamentale et des contours, sauf pour les allongements et les schwas épenthétiques porteurs d'un contour intonatif (Duez 2001a:41). Enfin, au niveau de la **forme prosodique et de la localisation syntaxique**, on ne postule pas de différences à priori de la forme de la courbe intonative des pauses pleines en fonction de leur localisation syntaxique ou d'une chaîne de (dis)fluence(s) (O'Shaughnessy 1992).

Dans ce chapitre, nous observons avant tout la distribution des pauses pleines et des allongements par **clustering hiérarchique** ; d'abord sur l'entièreté du corpus et ensuite par situation (1.2), en passant par un rappel des **transformations** possibles et une évaluation des **données aberrantes**. Nous observerons ensuite les **contextes de (co)occurrences** des pauses pleines et des allongements (1.3). Nous regardons leur interaction avec les **unités syntaxiques** et les **tours de parole** (1.4). Enfin, nous décrivons notamment les **paramètres intonatifs** autour des pauses pleines et des allongements, et surtout les **mouvements intonatifs** en leur sein (1.5). Le croisement de ces quatre paramètres est ensuite analysé au regard des données de perception.

6. Cependant, cette pente de F0 n'est marquée que par 20% de déclinaison entre le début et la fin de la pause (Duez 2001a ; Guaïtella 1991 ; O'Shaughnessy 1992).

7. **Au niveau de leur hauteur intonative**, leur fréquence fondamentale serait spécifique au locuteur quelle que soit leur localisation dans l'énoncé (Léon 1969 ; Shriberg & Lickley 1993). Ainsi, la pause pleine s'actualise au niveau du registre intonatif prononcé entre le **dernier pic intonatif** et le **registre tonal moyen du locuteur** (Moniz, Trancoso & Mata 2009:1721). En d'autres termes, il existerait une « grande similarité entre la valeur de F0 du début de la pause pleine et de l'attaque non marquée du syntagme » (Duez 2001b:31).

1.2 Fréquence et durée des pauses pleines et des allongements en fonction des situations de parole

Les panels de la Figure 5.1 reprennent la distribution de 1425 pauses pleines (*à gauche*) et de 1215 allongements (*à droite*) de LOCAS-F et de C-Humour [8]. Les pauses pleines du corpus sont relativement **plus courtes que les silencieuses**. **Leur dispersion est plus étroite** : 46% des pauses pleines durent moins de 250 ms. (contre seulement 28% des pauses silencieuses), et moins de 2% d'entre elles dépassent 1 seconde (contre 10% des pauses silencieuses) [9].

FIGURE 5.1 – Diagramme en boîte à moustaches de la distribution des durées des pauses pleines (*gauche*) et des allongements (*droite*).

En comparant les pauses pleines et les allongements, nous observons que : (1) la **durée des allongements** est globalement **supérieure** à celle des pauses pleines ; (2) la distribution des allongements est **plus étroite** que celle des pauses pleines. Transformée en secondes, la moyenne des pauses pleines est estimée à 268.4 ms. contre 347.5 ms. pour celle des allongements. Ces observations sont statistiquement confirmées. Un test de Wilcoxon-Mann-Whitney sur les durées absolues des pauses pleines et des allongements confirme qu'il s'agit de **deux populations différentes** ($W = 598380, p < 0.001$). L'application d'un test-t d'échantillons indépendants de Welch sur les durées transformées en logarithme [10] confirme également ces constats ($t(2393, 4) = -12.68, p < 0.001$).

8. Pour les pauses pleines, 75% des données proviennent de LOCAS-F contre 25% de C-Humour. Pour les allongements, 81% des données sont issues de LOCAS-F contre 19% de C-Humour. Pour les deux corpus, nous avons adopté la même méthodologie de seuils qu'au chapitre 4. Moins de 0.4% ($n = 6$) des pauses pleines ont été exclues et considérées comme données statistiquement aberrantes. Cette exclusion concerne 39 points pour les allongements soit 3.2% des données. Ainsi, nous travaillons au final sur 1419 pauses pleines et sur 1476 allongements.

9. Voir chapitre 4, Figure 4.3.

10. Le test-t avec une correction de Welch a l'avantage de ne pas avoir pour prémisse l'égalité des variances entre les deux échantillons indépendants. Cependant, il conserve la prémisse de normalité des résidus, nous l'appliquons donc aux données transformées en logarithme.

En d'autres termes, 95% de l'intervalle de confiance de la différence entre la moyenne de la durée des pauses pleines et des allongements est entre -0.09 et -0.12 logdur (soit 813 et 760 ms.).

Les panels de la Figure 5.2 montrent la distribution des durées de ces deux marques selon deux mesures : durée absolue (panels A, *gauche*), et logarithme de base 10 (panels B, *droite*). Contrairement à la distribution des pauses silencieuses dont les distributions multimodales s'actualisent comme mixte de courbes gaussiennes, les distributions des pauses pleines et des allongements sont unimodales.

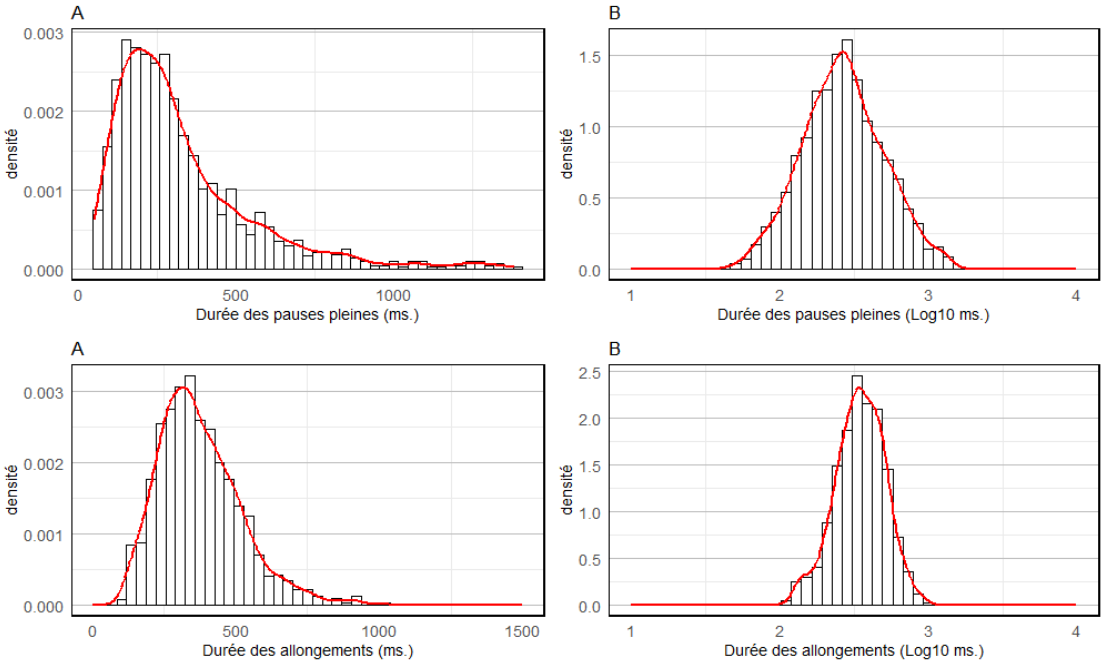


FIGURE 5.2 – Distributions de deux mesures de distributions des pauses pleines (*en haut*) et des allongements (*en bas*) : durée absolue et logarithme de base 10 pour LOCAS-F et pour C-Humour.

Les panels (A) confirment le caractère asymétrique de la distribution des pauses en durée absolue. Les panels (B) **confirment la lognormalité – l'unimododalité** des distributions de la durée des pauses pleines et des allongements, observées sur l'ensemble du corpus. Ainsi, si les schémas soulignent certaines différences entre les distributions de durées des deux marques, ils appuient la lognormalité, l'unimodalité et l'asymétrie des deux ensembles distincts de données (les pauses pleines *en haut* et les allongements *en bas*).

A travers les situations de parole, la **durée médiane des pauses pleines** de LOCAS-F et de C-HUMOUR s'étend de **107 ms. à 342 ms.**. L'observation du panel supérieur de la Figure 5.3 confirme que la durée pausale varie non seulement en fonction des **situations de parole**, mais également **en leur sein**. Une analyse de la variance (ANOVA) sur les durées en logarithme de base 10 montre une fluctuation significative des conditions ($F(14, 1404) = 22.292, p < 0.001$).

Un test post-hoc de Tukey DSH[11] révèle que **7 groupes de pauses pleines** peuvent être distingués par leur durée **en fonction des situations de parole**. Il n'existe donc pas de différences significatives de la durée des pauses au sein de ces sept groupes[12]. Les pauses les plus longues se retrouvent dans le groupe (1) interaction formelle et dialogue conversationnel informel, tandis que les plus courtes se relèvent dans le groupe (7) journal parlé et discours politique.

L'examen du panel inférieur de la Figure 5.3 confirme que la durée des allongements se module en fonction des situations de parole, et **à l'intérieur** de celles-ci. La **durée médiane des allongements** par situation de LOCAS-F et de C-HUMOUR va de **202 ms. à 476 ms**. Une analyse de la variance (ANOVA) sur les durées en logarithme témoigne d'une fluctuation significative au sein des conditions ($F(16, 1159) = 6.5, p < 0.001$).

Les résultats du test de Tukey DSH révèlent **5 groupes d'allongements en fonction des situations de parole**. Il n'existe donc pas de différences significatives de la durée des allongements au sein de ces cinq groupes[13]. Les allongements les plus longs sont observés pour (1) l'interaction formelle et l'homélie. Les plus courts le sont pour (5) la chronique radiophonique et la lecture.

Notons cependant que les durées en fonction des situations de parole peuvent doubler d'une situation à l'autre. Il peut dont être intéressant de prendre en compte cette variation intersituationnelle lors de l'évaluation de la (dis)fluence : les normes et les attentes peuvent varier d'une situation à une autre, et donc aussi l'évaluation que les interlocuteurs font du discours en situation.

11. Le test statistique de Tukey DSH (différence significative honnête) effectue des comparaisons multiples en une étape. Il compare toutes les paires de moyennes possibles en se basant sur une distribution des rangs *studentisée*, similaire à la distribution selon la loi de Student utilisée dans le *test-t*.

12. Dans l'ordre des pauses les plus longues aux plus courtes nous avons donc : [1] (Interaction formelle > Dialogue conversationnel informel) > [2] (Sketch > Interview sociolinguistique, Récit conversationnel > Interaction informelle >Conférence scientifique) > [3] (Interview radio) > [4] (Homélie) > [5] (Chronique radio > Interview télévisée) > [6] (Performance radiophonique >Débat politique) > [7] (Journal parlé > Discours politique). Les parenthèses symbolisent les regroupements.

13. Dans l'ordre des allongements les plus longs aux plus courts nous avons donc : [1] (Interaction formelle) > [2] (Homélie > Interview sociolinguistique > Conférence scientifique > Récit conversationnel > Interview radio > Performance radiophonique > Dialogue conversationnel informel) > [3] (Discours politique) >[4] (Interaction informelle > Journal parlé > Sketch > Débat politique) > [5] (Interview télévisée > Discours académique > Lecture > Chronique radio). Les parenthèses symbolisent les regroupements.

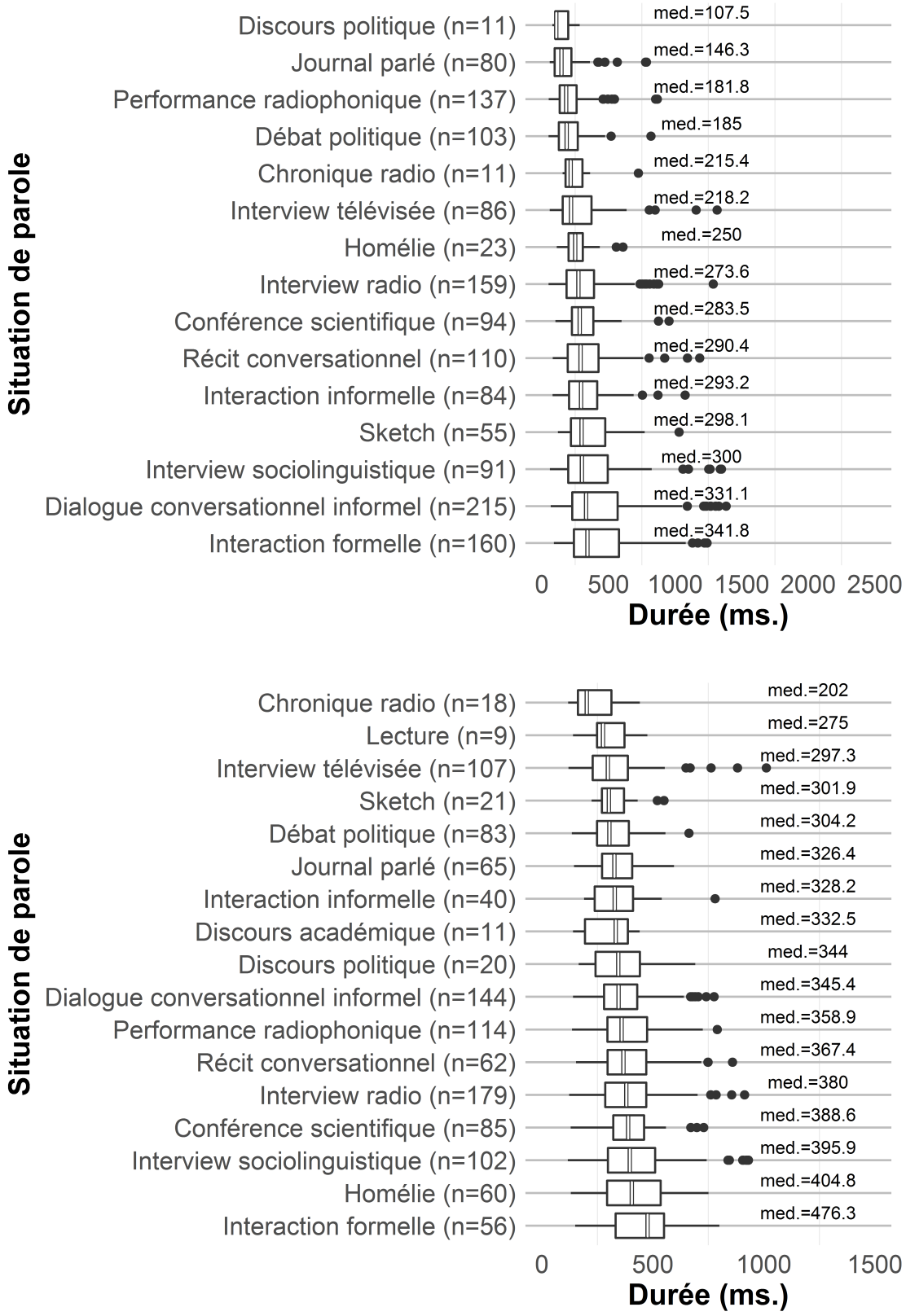


FIGURE 5.3 – Nombre d'occurrences et médianes des pauses pleines (*en haut*) et des allongements (*en bas*) en fonction de la situation de parole pour LOCAS-F et pour C-HUMOUR.

Les profils de pauses pleines et d'allongements doivent-ils être analysés en tenant compte des **caractéristiques de la situation de parole** (i.e.degré de préparation, degré d'orientation médiatique, etc.) et/ou **du locuteur** ? Ces variations intersituationnelles sont-elles consistantes ? Afin d'observer la **variabilité interlocuteurs et intrasituationnelle**, nous appliquons également un **regroupement hiérarchique** (*clustering*) par situation de parole et par locuteur [14]. Certaines situations montrent des **variations individuelles importantes** (Figures 5.4 à 5.7). Nous présentons les distributions pour quatre situations de parole représentatives (i.e. conférence scientifique, interview sociolinguistique, performance radiophonique, et récit conversationnel).**Au niveau des pauses pleines**, les échantillons de conférences scientifiques (Figure 5.4, *à gauche*) montrent une distribution unimodale chez les 4 locuteurs et laissent entrevoir que les **durées moyennes des pauses pleines varient peu** d'un locuteur à l'autre. Dans les interviews sociolinguistiques (Figure 5.5, *à gauche*), les distributions de chaque locuteur sont **tout aussi homogènes** et également unimodales : cette situation parait donc relativement **standardisée** malgré sa spontanéité. Les performances radiophoniques et les récits conversationnels (Figures 5.6 et 5.7, *à gauche*) sont caractérisés par une distribution **plutôt unimodale** et relativement homogène entre les locuteurs, à l'exception d'un locuteur dans la situation performance radiophonique mais pour lequel la présence de pauses pleines « courtes » (première gaussienne) reste marginale. **Au niveau des allongements**, des conclusions similaires peuvent être tirées de l'observation des panels de droite. On remarque peu de variation intrasituationnelle et des distributions majoritairement unimodales.

En conclusion, la variation des **durées interlocuteurs et intrasituationnelles** est donc moins marquée pour les pauses pleines et les allongements que pour les pauses silencieuses, malgré des différences de durée intrasituationnelles relevées ci-avant. Nous ne relevons que très rarement des distributions bimodales clairement distinctes. Au sein même des situations, bien que la **variation individuelle** soit plus notable dans certaines situations que dans d'autres, la durée (Log10) suit systématiquement une distribution unimodale. Il ne paraît donc pas pertinent de modéliser la distribution des pauses pleines comme un mixte de courbes gaussiennes ou de différencier le type de traitement des données en fonction des situations.

Nous analysons aux sections suivantes la distribution et la durée de 1419 pauses pleines et de 1716 allongements relativement à quatre types de phénomènes : les chaînes de cooccurrences, les frontières d'unités syntaxiques, les tours de parole et les groupes intonatifs. Ce faisant nous poursuivons un double objectif, méthodologique et de modélisation.

14. Nous utilisons le package `mclust` (version 5.2.3) (FRALEY 2008) sous R (*R Core Team* 2016).

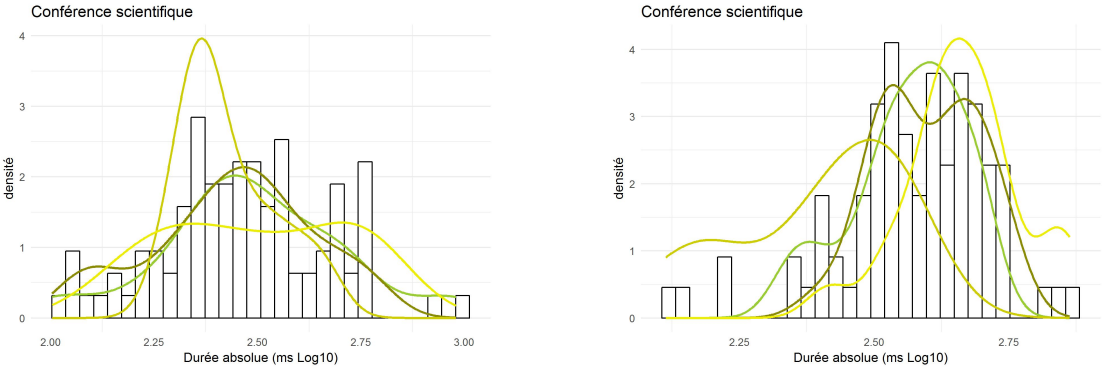


FIGURE 5.4 – Distribution des durées sur 94 pauses pleines (*panel gauche*) et sur 150 allongements (*panel droit*), pour 4 locuteurs dans la situation « conférence scientifique » (LOCAS-F).

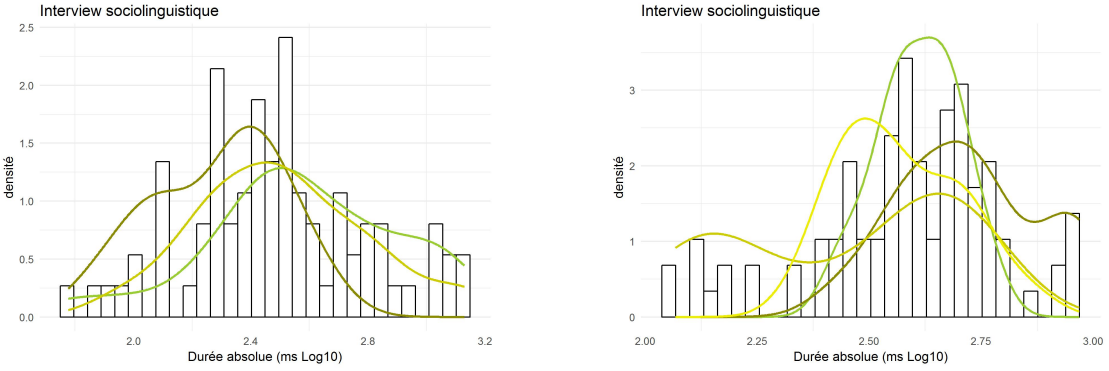


FIGURE 5.5 – Distribution des durées sur 91 pauses pleines (*panel gauche*) et 195 allongements (*panel droit*), pour 3 et 5 locuteurs dans la situation d'« interview sociolinguistique » (LOCAS-F).

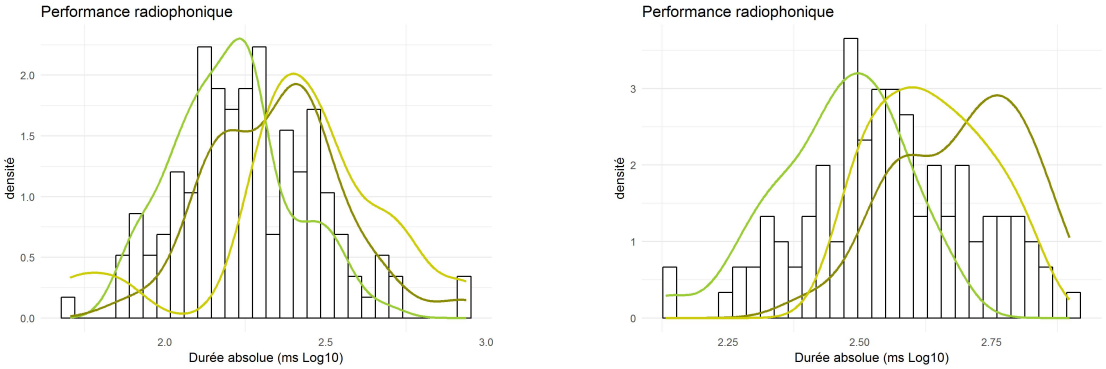


FIGURE 5.6 – Distribution des durées sur 137 pauses pleines (*panel gauche*) et sur 210 allongements (*panel droit*), pour 3 locuteurs dans la situation de la « performance radiophonique » (C-HUMOUR).

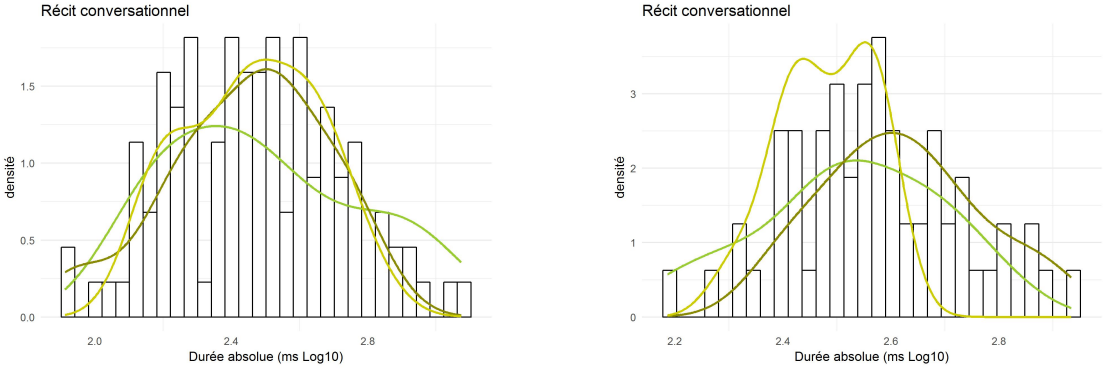


FIGURE 5.7 – Distribution des durées sur 110 pauses pleines (*panel gauche*) et sur 134 allongements (*panel droit*), pour 3 locuteurs dans la situation du « récit conversationnel » (LOCAS-F).

1.3 Fréquence et durée des pauses pleines et des allongements cooccurrents à des (dis)fluences

A l'instar des choix faits au chapitre 4, chaque pause pleine et chaque allongement sont catégorisés en fonction de son **environnement immédiat**. Les marques dites « seules » sont précédées et suivies de tokens qui n'ont pas été annotés comme des (dis)fluences potentielles (exemples 21a et 21b). Les marques peuvent **être précédées et/ou suivies** d'un marqueur de discours (exemple 21c), ou d'**une autre (dis)fluence** (exemples 21d et 21e). Les marques en *interregnum* sont incluses entre un élément à corriger et sa correction au sein d'une phase d'édition (exemples 21f et 21g). Enfin, les marques peuvent être aussi dans des chaînes plus longues et/ou plus complexes, supérieures à deux éléments (exemples 21h et 21i). Ces chaînes peuvent contenir plusieurs pauses (e.g. [pause + marqueur + pause + marqueur]). Nous avons choisi de les regrouper sous l'étiquette « > à 2 éléments ».

Exemple (21) – Exemples de chaînes de pauses pleines et d'allongements issus de `chmSG2` (C-Humour) :

a. vous savez je suis tout à fait ***euh*** civilisé

b. le président de la république **nomme::** les responsables de l'audiovisuel

c. il a fait un très beau travail ***euh* allez** à Libération

d. j'avais **jamais (0.700) jamais *euh*** fait de de presse écrite

e. et je pense **que:: c'est c'est** maintenant qu'il faut faire savoir

f. il est très intéressant de voir comment le pouvoir a essayé **de:: de de** faire taire certaines personnes

g. vous voulez que **si *euh* (0.543) si** la gauche passe

h. ***euh* voilà en plus** c'est une (0.100) pour moi c'est vraiment une expérience toute nouvelle

i. parfois **sous la:: sous la menace *euh* sous sur écoute** parfois

Au total, 1419 pauses pleines et 1176 allongements ont été analysés. Le Tableau 5.1 présente la fréquence de chaque type de chaîne. A l'instar de Moniz, Mata, Trancoso & Viana (2008), qui observent que plus de 20% de pauses pleines sont isolées, nous relevons également **un quart des marques** en isolation : 23% des pauses pleines et 27% des allongements.

Ensuite, les corpus révèlent que respectivement 42,8% et 44.5% des pauses pleines et des allongements sont **en cooccurrence avec une autre marque**[15]. Comme Grosjean & Deschamps (1975), qui notent que 47.5% des pauses pleines sont combinées avec des pauses silencieuses (avant et/ou après), nos corpus témoignent de **44.9% de pauses pleines en cooccurrence directe avec une pause silencieuse**. Par ailleurs, chez

15. Ces pourcentages sont obtenus en cumulant les résultats des deuxième et des troisième groupes de lignes du Tableau 5.1.

Moniz, Mata, Trancoso & Viana (2008), dans 80% des cas, les pauses pleines et les allongements sont **suivis de pauses silencieuses**. Selon les auteures, cette tendance soutient l'idée que leur présence est effectivement le signe d'un **retardement de la production discursive**. L'absence de la pause est par ailleurs sévèrement pénalisée en portugais (Moniz, Trancoso & Mata 2009:1719). Dans nos corpus, cette tendance est moins nette mais néanmoins présente : respectivement **25% et 12.4% des pauses pleines et des allongements sont suivis d'une pause silencieuse**.

Ainsi, une première distinction entre les marques est attestée : les cooccurrences dans des **chaînes de deux éléments** suivent des **configurations différentes** pour les pauses pleines et pour les allongements, ce qui **appuie l'hypothèse d'une distribution complémentaire de ces marques**.

L'observation des second et des troisième groupes de lignes du Tableau 5.1 révèle que les **pauses pleines ont tendance à être *suivies* de marque(s) de (dis)fluence(s)** tandis que les **allongements ont tendance à être *précédés* de (dis)fluence(s)**. Par exemple, nous notons que pour les **pauses pleines**, les **configurations (*euh* + pause silencieuse) et (*euh* + md)** sont favorisées (respectivement 14.9% et 9.2% des contextes du *euh*). Pour les **allongements**, ce sont les configurations **(pause silencieuse + allongement) et (*euh* + allongement)** qui sont privilégiées (respectivement 12.8% et 13.4% des contextes).

En suivant la même méthodologie que celle adoptée pour la définition des « **pauses (silencieuses) structurantes** » pour définir de potentielles « pauses (pleines) structurantes » : on relève que **50% des *euh* et 39% des allongements ne suivent *pas* une autre (dis)fluence**. Au chapitre précédent, nous remarquions que 78% des pauses silencieuses correspondaient à cette configuration. Il y a donc **moins de *euh* ou d'allongements potentiellement structurants** que de pauses silencieuses potentiellement structurantes. Par contre, à l'instar des pauses silencieuses, la position de **marque interne à des chaînes de (dis)fluences** favorise l'actualisation des pauses pleines et des allongements : 6.3% des pauses pleines et 7.8% des allongements sont en **phase d'édition**.

Dans un corpus d'interviews radiophoniques, Boula de Mareüil et al. (2005:5) observent que les cooccurrences les plus fréquentes de pauses pleines sont, dans l'ordre : les (*euh* + dis), les (md + *euh*), les (*euh* + répétitions) et les (md + répétitions). Les auteurs remarquent de manière générale que la **chaîne (md + *euh*) est deux fois plus fréquente que celle (*euh* + md)** [16]. Cette tendance des *euh* à s'actualiser à la frontière droite est également attestée par Degand & Gilquin (2013). Nos données **confirment ces observations** à travers une variété de situations. En incluant les chaînes > à 2

16. De plus, cette étude présente l'avantage d'une définition large des marqueurs de discours, comparable à la nôtre, intégrant certaines conjonctions, ainsi qu'une annotation inclusive des répétitions (e.g. *répétitions à l'identique, répétitions emphatiques, répétitions encadrant un parenthétique*).

éléments, **11% des *euh* succèdent à un md**, tandis que **20% des *euh* précèdent un md** [17]. Ces résultats sont aussi comparables à ceux de Crible, Degand & Gilquin (2017:84-85).

Dans une moindre mesure, une **similarité distributionnelle** peut être observée avec les allongements : 3.7% succèdent à un md, tandis que **le double**, 8.7%, précèdent un md [18]. Il sera intéressant d'étudier ces actualisations au regard des **frontières syntaxiques** et/ou **tours de parole** afin de confirmer si ces configurations sont synonymes de rupture ou continuation discursive (voir section 1.4).

Nous avons donc relevé quelques différences distributionnelles des chaînes de pauses et d'allongements (e.g. différentes tendances de cooccurrences), mais quelques similarités également (e.g. tendance à précéder un md). On note qu'il existe des **associations fortes** entre certains types de chaînes et les situations de parole préparée, semi-préparée ou spontanée.

Les tableaux d'association de la Figure 5.8 regroupent en ordonnée les types de chaînes incluant des pauses pleines (*en haut*) et des allongements (*en bas*). En abscisse, on retrouve les situations de parole regroupées selon le trait « degré de préparation » (i.e. *spontané, semi-préparé* ou *préparé*). La taille des boîtes est proportionnelle au nombre d'occurrences et la **taille de l'effet** (corrélation positive ou négative) est indiquée par la **mesure de résidu** (à droite). De manière générale, la force de l'association entre le degré de préparation et le type de chaînes – la **taille d'effet** – est **supérieure pour la pause pleine** ($\chi^2 = 152.8$, $df = 16$, $p < 0.001$) **que pour l'allongement** ($\chi^2 = 69. df = 20, p < 0.001$).

Pour les deux marques, la position de **marque isolée est favorisée en situation préparée** (+5.7 et +4.9 respectivement pour la pause pleine seule et l'allongement seul). A l'inverse, les chaînes supérieures à deux éléments y sont **atypiques**. Nous remarquons par ailleurs que :

- les chaînes (*euh* + pause silencieuse) et (allongement + pause silencieuse) sont caractéristiques des discours préparés et semi-préparés. Elles sont sous-représentées en parole spontanée.
- les chaînes (pause silencieuse + *euh*) et (pause silencieuse + allongement) sont caractéristiques des discours spontanés. Elles sont sous-représentées en paroles préparée et semi-préparée [19].

17. La chaîne (md + marque) est extraite en cumulant les données du premier groupe de lignes et le septième groupe de lignes du Tableau 5.1.

18. Les auteurs observent également que de manière plus générale les chaînes (md + dis) sont deux fois plus présentes que (dis + md). Une explication avancée par certains auteurs est que le md est souvent utilisé pour initier un message, et qu'une fois qu'une (dis)fluence est initiée, le marqueur de discours serait moins employé dans le *reparans*.

19. Dans une moindre mesure car de manière non significative pour les allongements.

Chaîne de cooccurrences	*Pause pleine*		*Allongement*	
	(n)	(%)	(n)	(%)
`marque` seule	327	23	317	27
`marque` + p. sil	211	14.9	56	4.8
`marque` + *euh*	10	0.7	22	1.9
`marque` + (dis)fluence	26	1.8	37	3.1
`marque` + md	131	9.2	29	2.5
p. sil + `marque`	139	9.8	150	12.8
euh + `marque`	8	0.6	157	13.4
(dis)fluence + `marque`	46	3.2	43	3.7
md + `marque`	37	2.6	26	2.2
p. sil + `marque` + p. sil	52	3.7	34	2.9
p. sil + `marque` + *euh*	2	0.1	5	0.4
p. sil + `marque` + (dis)fluence	63	4.4	29	2.5
p. sil + `marque` + md	76	5.4	31	2.6
euh + `marque` + *euh*	1	0.1	4	0.3
euh + `marque` + p. sil	6	0.4	34	2.9
euh + `marque` + (dis)fluence	5	0.4	38	3.2
euh + `marque` + md	3	0.2	26	2.2
(dis)fluence + `marque` + p. sil	49	3.5	14	1.2
(dis)fluence + `marque` + *euh*	3	0.2	6	0.5
(dis)fluence + `marque` + (dis)fluence	89	6.3	92	7.8
(dis)fluence + `marque` + md	16	1.1	8	0.7
md + `marque` + p. sil	44	3.1	7	0.6
md + `marque` + *euh*	7	0.5	1	0.1
md + `marque` + (dis)fluence	16	1.1	5	0.4
md + `marque` + md	52	3.7	5	0.4
Total	**1419**	**100**	**1176**	**100**

TABLEAU 5.1 – Fréquences et proportions des pauses pleines et des allongements en fonction des cooccurrences. *p. sil = pause silencieuse, md = marqueur de discours, euh = pause pleine.*

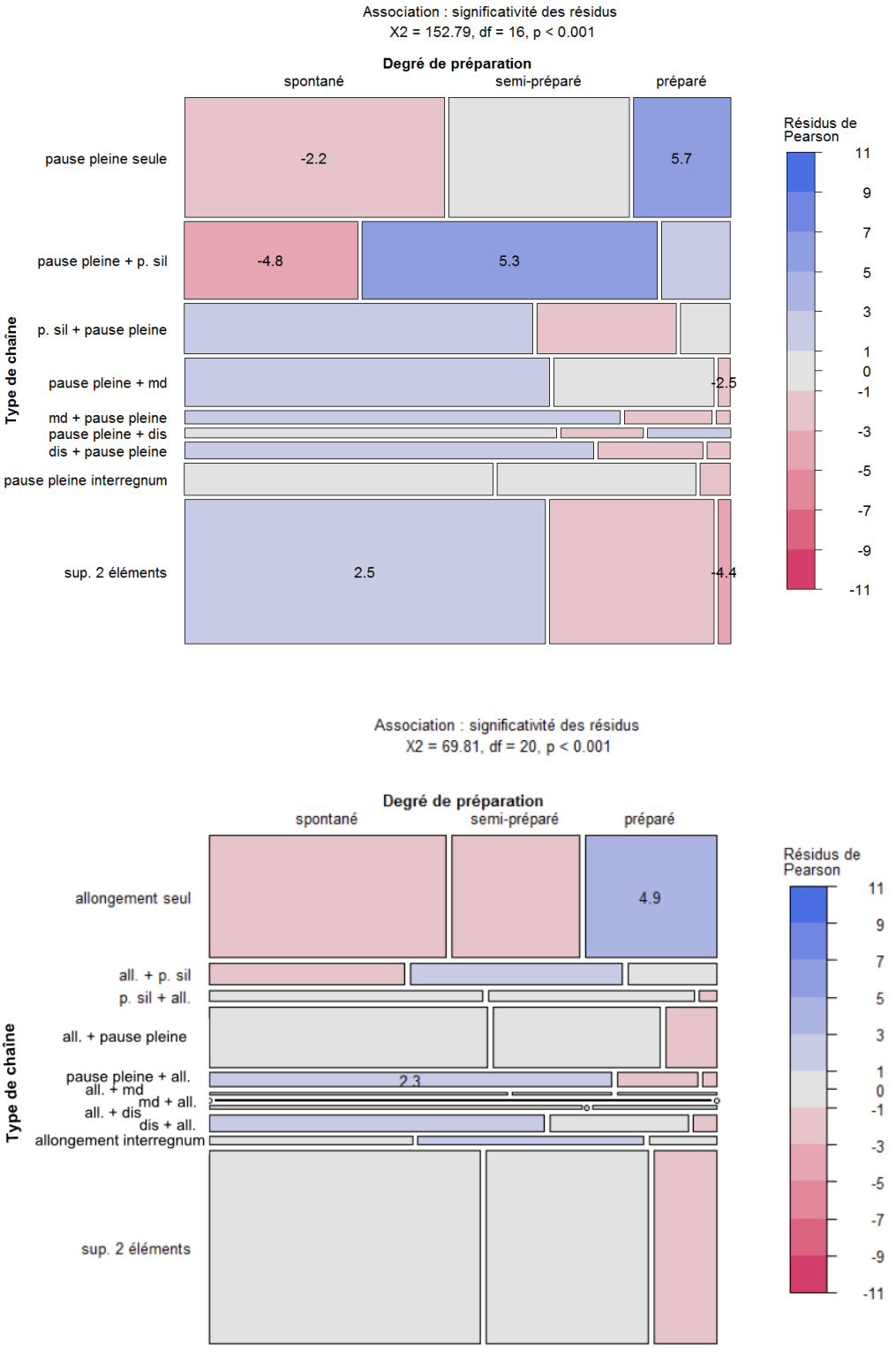


FIGURE 5.8 – Graphiques des tableaux croisés du degré de préparation et des chaînes de cooccurrences des pauses pleines (*en haut*) et des allongements (*en bas*).

Nous désirons observer ces comportements au regard de **la durée des pauses pleines et des allongements dans ces configurations**. Pour rappel, en début de chapitre nous avons émis l'hypothèse que la durée moyenne des configurations suivies d'une pause silencieuse serait supérieure à celles précédées d'une pause silencieuse. Les panels de la Figure 5.9 mettent en évidence les **différences de durées en fonction des marques en cooccurrence avec les pauses pleines** (en haut) et les **allongements** (en bas).

De manière générale, dans les **configurations initiées par une pause pleine** (e.g. pause pleine + pause silencieuse, pause pleine + md, pause pleine + dis), la **durée de la pause pleine est inférieure** aux configurations où celle-ci succède à une autre marque (e.g. pause silencieuse + pause pleine, md + pause pleine, dis + pause pleine). Cette conclusion va donc **à l'encontre de notre hypothèse initiale**.

Les durées pausales en fonction des cooccurrences peuvent s'organiser en **3 groupes**. Les durées les plus élevées de pauses pleines se notent (effectivement) pour les combinaisons suivies d'une pause pleine (*euh* + pause pleine, md + pause pleine, pause silencieuse + pause pleine). Les durées les plus longues de pauses pleines sont également relevées pour celles insérées dans des chaînes $>$ à 2 éléments. Ces durées les plus substantielles s'opposent aux **plus courtes** qu'on retrouve pour les **pauses pleines en position d'*interregnum***, pour les **pauses pleines suivies d'une (dis)fluence, et pour les pauses pleines seules**. Enfin, un groupe intermédiaire compte les autres configurations. On y retrouve, du plus court au plus long : les pauses pleines suivies d'une pause silencieuse, les pauses pleines précédées d'une (dis)fluence, les pauses pleines suivies d'un md et enfin les pauses pleines suivies d'un *euh*. Une analyse de la variance (ANOVA) sur la durée des pauses pleines en fonction des cooccurrences révèle une variation significative entre les conditions, ($F(10, 1408) = 8.9, p < 0.001$). Un test post-hoc de Tukey DSH confirme les observations précédentes.

Au contraire, les résultats de la variation de la durée des allongements en fonction des cooccurrences sont non significatifs dans leur ensemble ($F(10, 1165) = 1.83, p = 0.0506$). Un test post-hoc de Tukey DSH révèle que seules deux configurations se distinguent comme plus longues que les autres : (allongement + md) et (dis+ allongement) ont une moyenne de groupe de 433 ms. Le reste des configurations s'étend de 230 à 363 ms. en moyenne, mais sans différence significative au sein du groupe.

Enfin, nous relevons des différences de durées eu égard au type de cooccurrence mais aussi du degré de préparation (Figure 5.10). L'ordre des durées des pauses pleines est préservé entre les différents degrés de préparation, bien que ces durées soient globalement **plus longues en situation préparée** ($F(2, 1062) = 70.7, p < 0.001$). Pour les allongements, on observe également des différences en fonction du degré de préparation ($F(2, 1173) = 3.5, p < 0.003$), les allongements en **situations préparées** ($M = 324.2$ ms.) sont, sans surprise, plus courts que ceux en situations spontanées ou semi-préparées ($M = 352$ ms.). Dans la section suivante, nous rendons compte du **rôle**

des facteurs syntaxiques et des tours de parole pour expliquer la distribution des pauses pleines et des allongements. Cela permettra notamment de centrer l'attention sur les **pauses pleines et les allongements isolés** (non combinés avec des (dis)fluences).

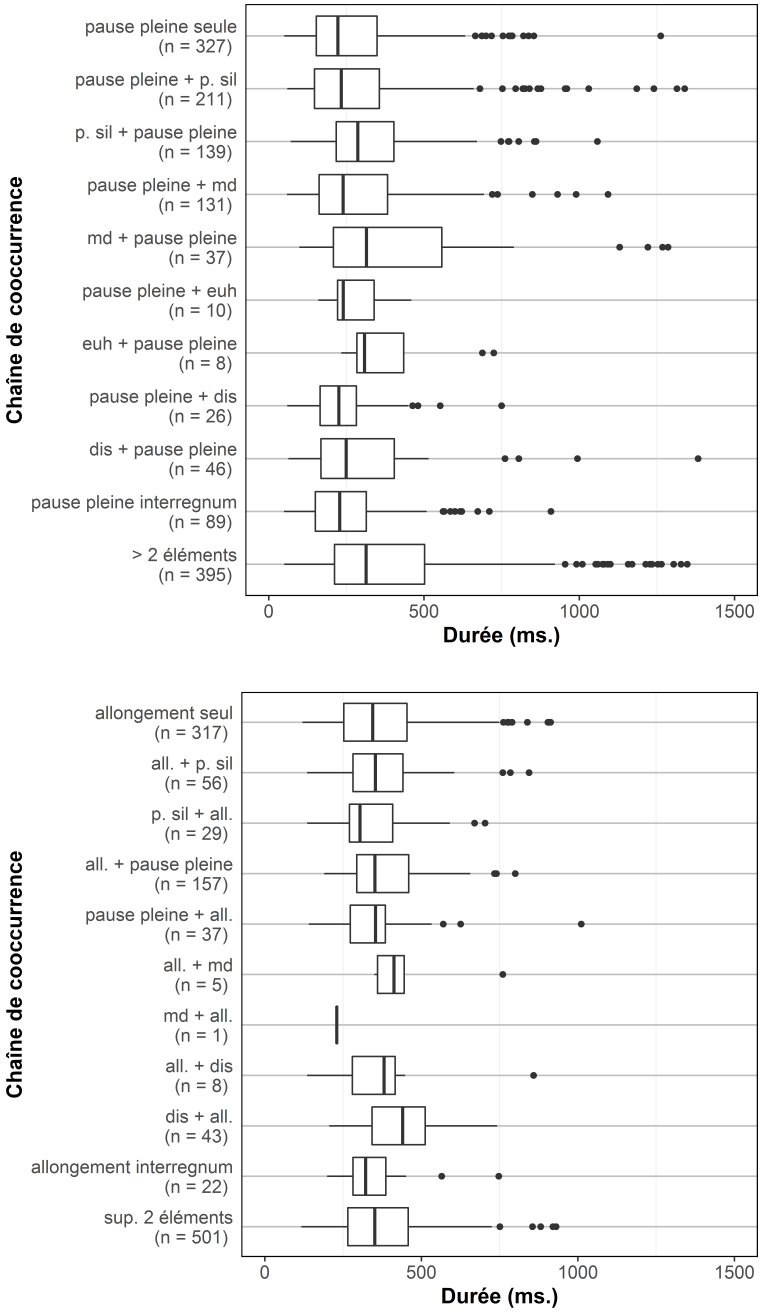


FIGURE 5.9 – Durées en fonction des cooccurrences des pauses pleines (*en haut*) et des allongements (*en bas*).

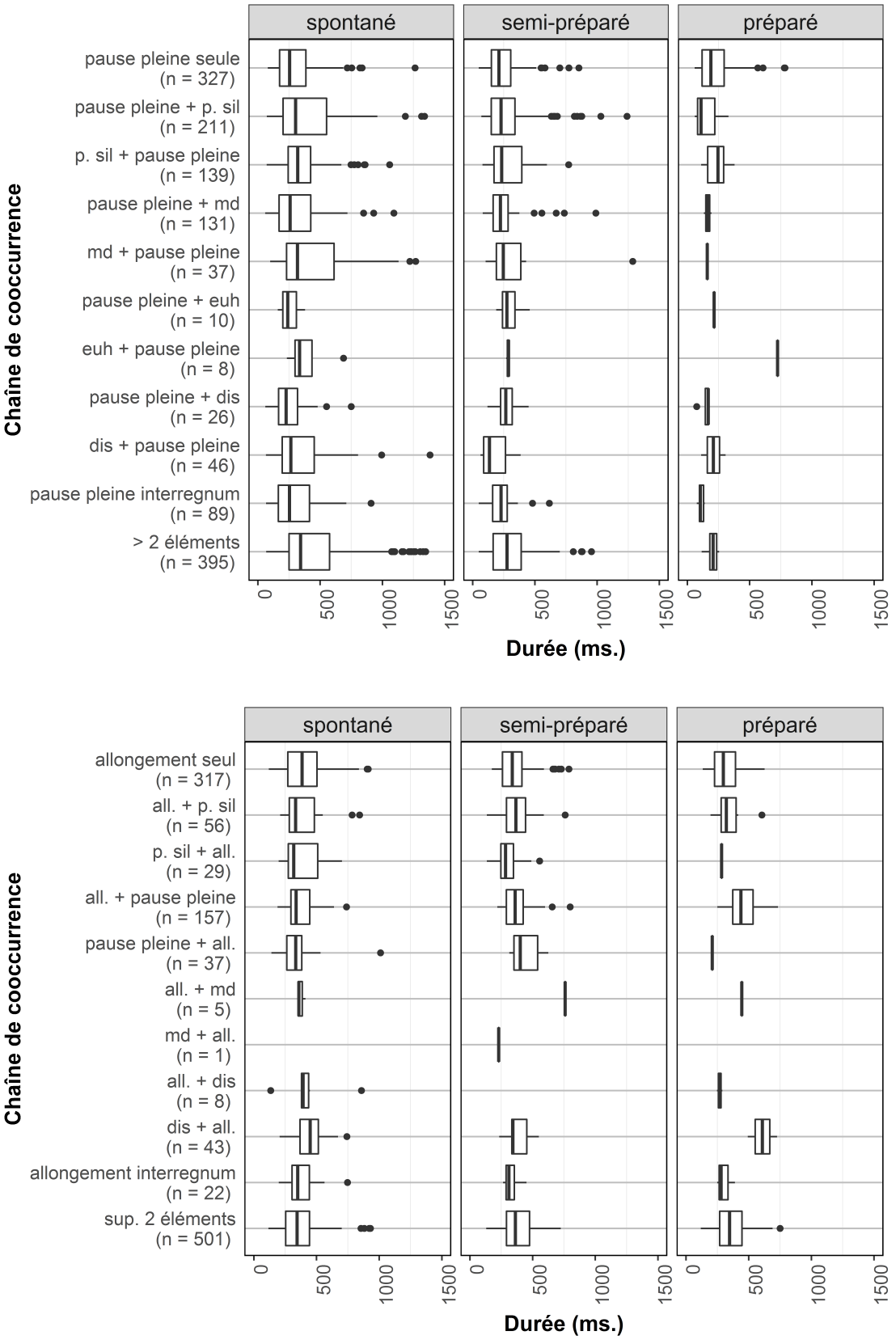


FIGURE 5.10 – Durées pausales en fonction des cooccurrences des pauses pleines (*en haut*) et des allongements (*en bas*) et du degré de préparation.

1.4 Fréquence et durée des pauses pleines et des allongements aux frontières syntaxiques

En début de chapitre, nous n'avons pas formulé d'hypothèse sur **la durée ou la fréquence des pauses pleines et des allongements en fonction de leur position syntaxique**. Dister (2007:233) soutient que « le *euh* n'a pas d'influence sur le déroulement syntaxique de l'énoncé : il le suspend momentanément, que ce soit à l'intérieur d'un syntagme ou à la frontière de deux syntagmes ». Cette réflexion suit les observations de Boula de Mareüil et al. (2005:4), ses données l'attestent : les pauses pleines peuvent s'actualiser **quasi en tout lieu** (35% se trouvent en inter-ga ou intra-séquence, contre 14% en inter-rection). Cependant, les pauses pleines restent **peu présentes au sein de syntagmes nominaux** (i.e. entre un déterminant et un nom). Selon l'auteur, cette position serait davantage favorable à l'**actualisation d'allongements** ou d'autres (dis)fluences.

Nos analyses ne confirment que partiellement les observations antérieures. **La durée des pauses pleines varie en fontion de leur position syntaxique** ($\chi^2 = 148.74$, $df = 3, p < 0.001$). Cependant, plus d'un tiers des pauses pleines sont en **position inter-rectionnelle** (37%, $n = 394$), ce qui est plus que les 14% notés par Boula de Mareüil et al. (2005:4). Ensuite, 13.5% ($n = 144$) sont en position inter-séquences et 30% ($n = 328$) inter-ga. Enfin, une partie non négligeable des pauses pleines s'actualise également au sein même du ga, entre le déterminant et le nom, jusqu'à 18.7% ($n = 199$). Il est intéressant de voir si les durées des allongements intra-ga sont en effet plus élevées et/ou davantage en cooccurrence avec d'autres marques[20].

Ensuite, **la distribution des allongements varie également en fonction des frontières syntaxiques** ($\chi^2 = 681.5, df = 3, p < 0.001$). Leur actualisation majoritaire en **position intra-ga** est davantage marquée que pour les pauses pleines (61.1%, $n = 581$). Les catégories inter-ga, inter-séquences et inter-rection sont moins représentées. Elles avoisinent respectivement 18.6%, 8.3% et 11.9%. L'actualisation d'autres (dis)fluences est favorisée entre des unités syntaxiques plus grandes. L'allongement d'hésitation est peu synonyme d'interruption du discours. Les données soulignent la fonction **modale et prospective** des allongements qui incarne une gestion immédiate du discours au niveau micro-syntaxique.

Le panel supérieur de la Figure 5.11 confirme l'hypothèse que la **durée moyenne d'une pause pleine augmente en fonction du degré de frontière syntaxique** ($\chi^2 = 44.1, df = 3, p < 0.001$). Plus précisément, les pauses intra-ga et inter-ga s'opposent et sont plus courtes que les pauses inter-séquence et inter-rection. En effet un test pairé de Wilcoxon permet de distinguer **deux groupes** : il n'existe pas de différence significative entre les pauses pleines en intra-ga et en inter-ga ($p = 0.89$), ni entre inter-séquences et inter-rections ($p = 0.26$). Par contre, les différences sont bien significatives

20. Nous y revenons à l'observation des données de la Figure 5.14.

entre ces deux groupes ($p < 0.005$). Au contraire, le panel inférieur de la Figure 5.11 révèle que la durée de l'allongement ne varie pas en fonction de son positionnement syntaxique ($\chi^2 = 2.24, df = 3, p < 0.5$).

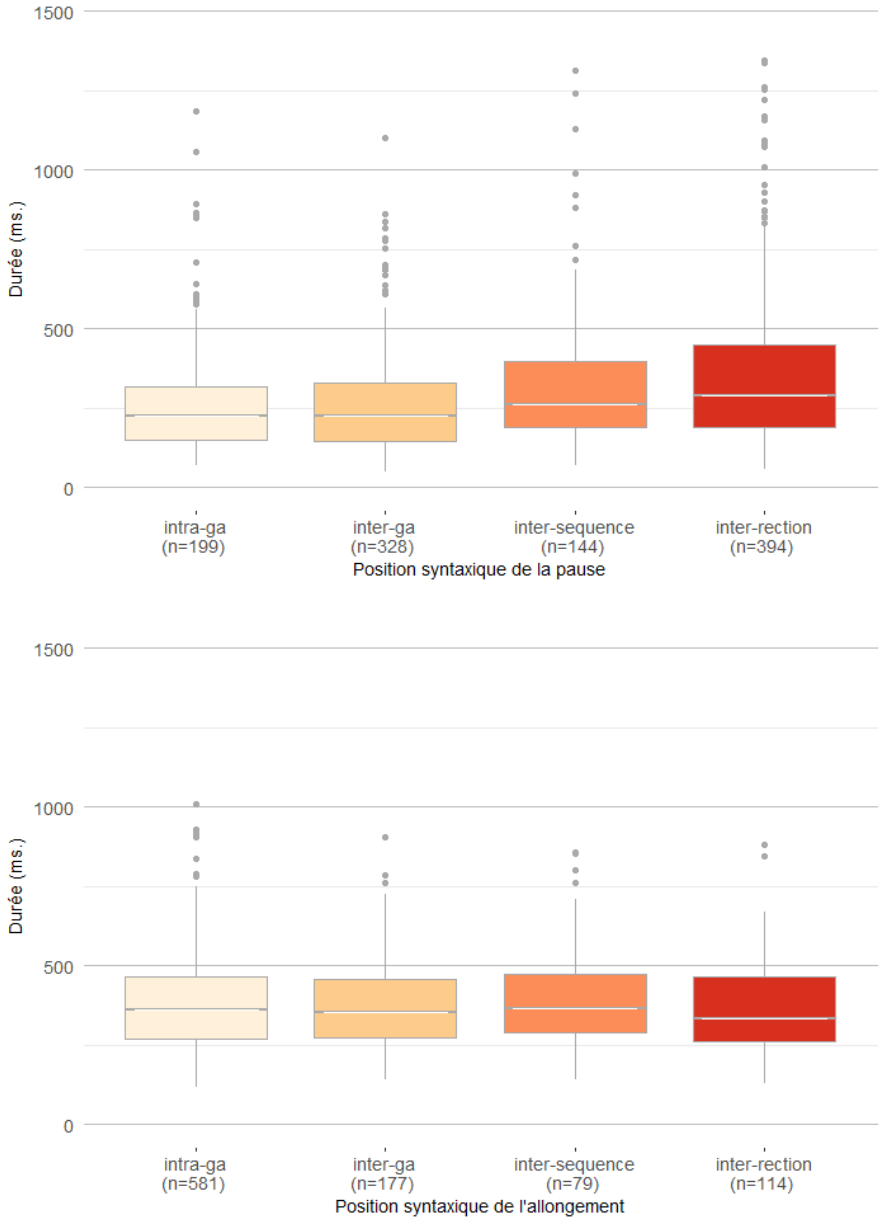
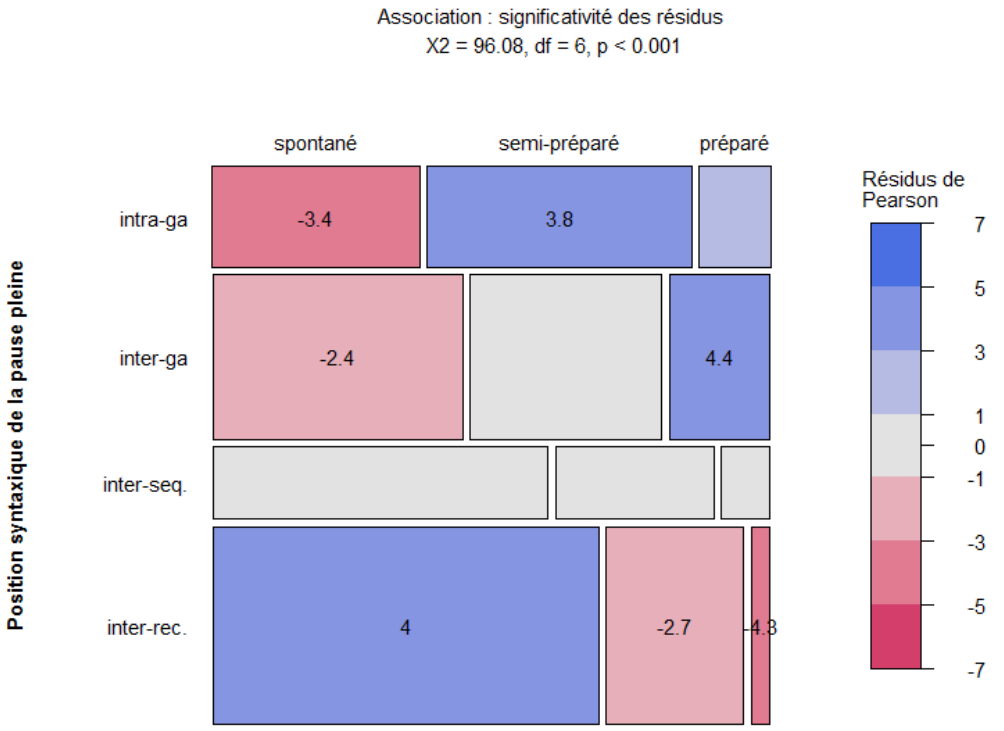


FIGURE 5.11 – Durées des pauses pleines (*en haut*) et des allongements (*en bas*) en fonction de la position syntaxique.



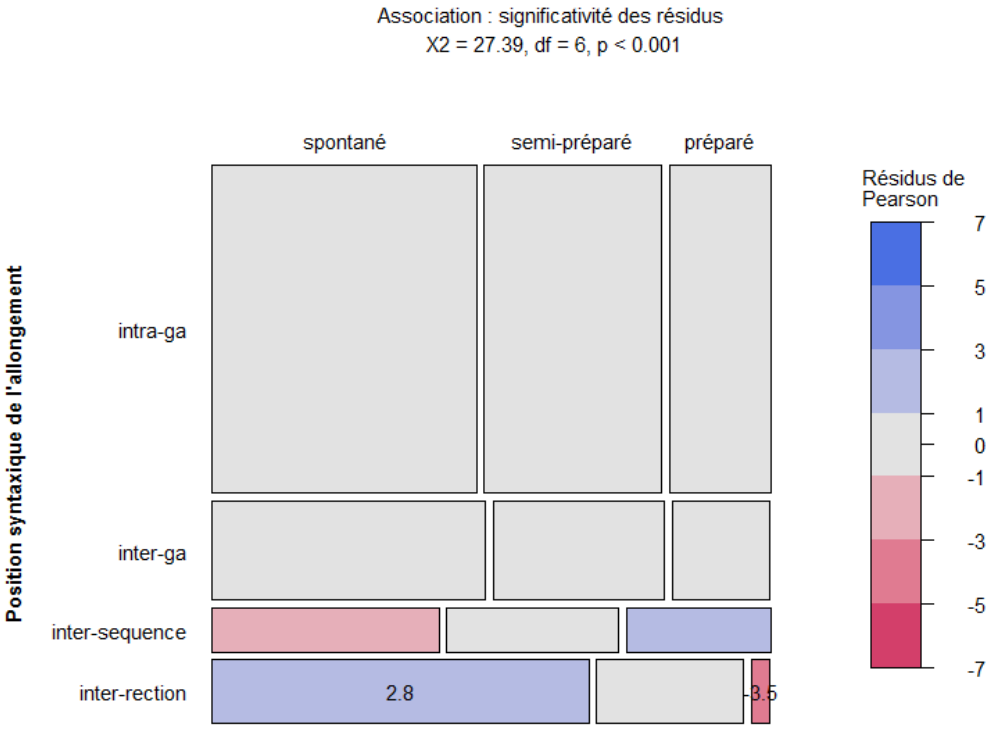


FIGURE 5.12 – Graphiques des tableaux croisés des pauses pleines (*en haut*) et des allongements (*en bas*) en fonction du degré de préparation et de la position syntaxique.

Nous avons ensuite analysé si les préférences distributionnelles interagissent avec le **type de situation**. Il existe en réalité une relation forte entre le degré de préparation et la distribution syntaxique des marques au sein de groupes syntaxiques (Figure 5.12). Cette association est **davantage prononcée dans le cas des pauses pleines** ($\chi^2 = 96.1, df = 6, p < 0.001$) que des allongements ($\chi^2 = 27.4, df = 6, p < 0.001$).

En spontané, la distribution des pauses pleines est significativement différente à travers les groupes syntaxiques ($\chi^2 = 11.12, df = 3, p < 0.05$). La position **inter-rectionnelle** est clairement plus exploitée (résidu +4). A l'inverse, en situation semi-préparée ($\chi^2 = 44.1, df = 3, p < 0.001$). Ce sont les intra-ga ou inter-ga qui sont favorisées pour l'occurrence de la pause pleine – de même, marginalement et de manière non significative en situation préparée ($\chi^2 = 6.6, df = 3, p < 0.09$). Enfin, pour **les allongements**, on note que si leur majorité se trouve en **intra-ga** ($n = 581$), la position inter-rectionnelle est davantage représentée en spontané qu'en préparé (résidu +2.8). Au contraire, la position inter-séquence (**entre les chaînes fonctionnelles**) serait plus typique de la situation préparée et sous-représentée en spontané.

Nous étudions également si la durée des pauses pleines et des allongements varie en fonction de ces deux paramètres. Comme déjà énoncé, nous n'avons pas observé de variations significatives au niveau de la durée des allongements en fonction du degré de préparation et/ou du niveau de frontière syntaxique. Par contre, nous notons **des différences de durées des pauses pleines au sein des situations spontanées** ($\chi^2 = 11.12, df = 3, p < 0.015$) et semi-préparées ($\chi^2 = 13.09, df = 3, p < 0.005$). Dans ces deux cas, les pauses pleines en position inter-rectionnelle sont plus longues que celles en position inter-ga et intra-ga (pour $p < 0.05$).

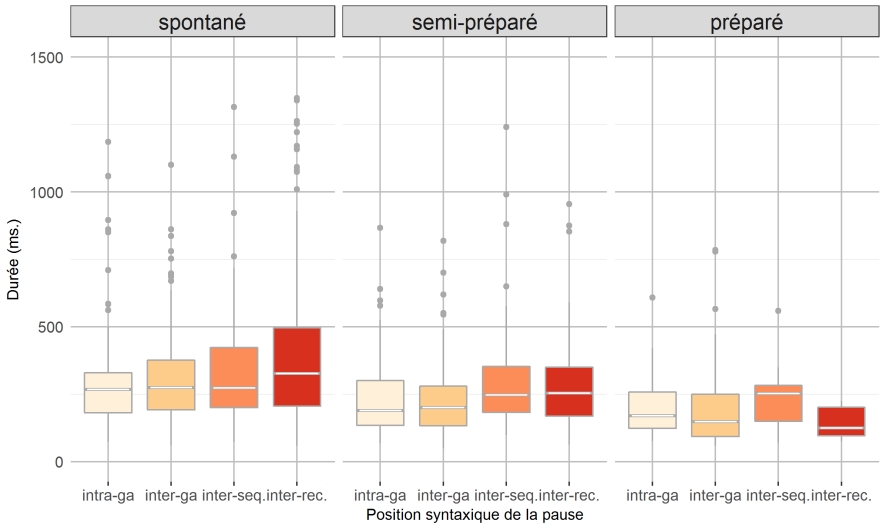


FIGURE 5.13 – Durées des pauses pleines en fonction de la position syntaxique et du degré de préparation.

Enfin, les diagrammes en boîte à moustaches (Figures 5.14 et 5.15) révèlent la variabilité des durées en fonction des chaînes de (dis)fluences contenant des pauses pleines et des allongements au regard de leur position au sein des unités syntaxiques[21].

Pour les pauses pleines, toutes positions syntaxiques confondues, on remarque que les chaînes (md + pause pleine), (dis + pause pleine) sont caractérisées par une plus gande variabilité que leurs homologues (pause pleine + md) et (pause pleine + dis). Au contraîre, les pauses en *interregnum* sont marquées par une faible variabilité, à l'exception des inter-rectionnelles.

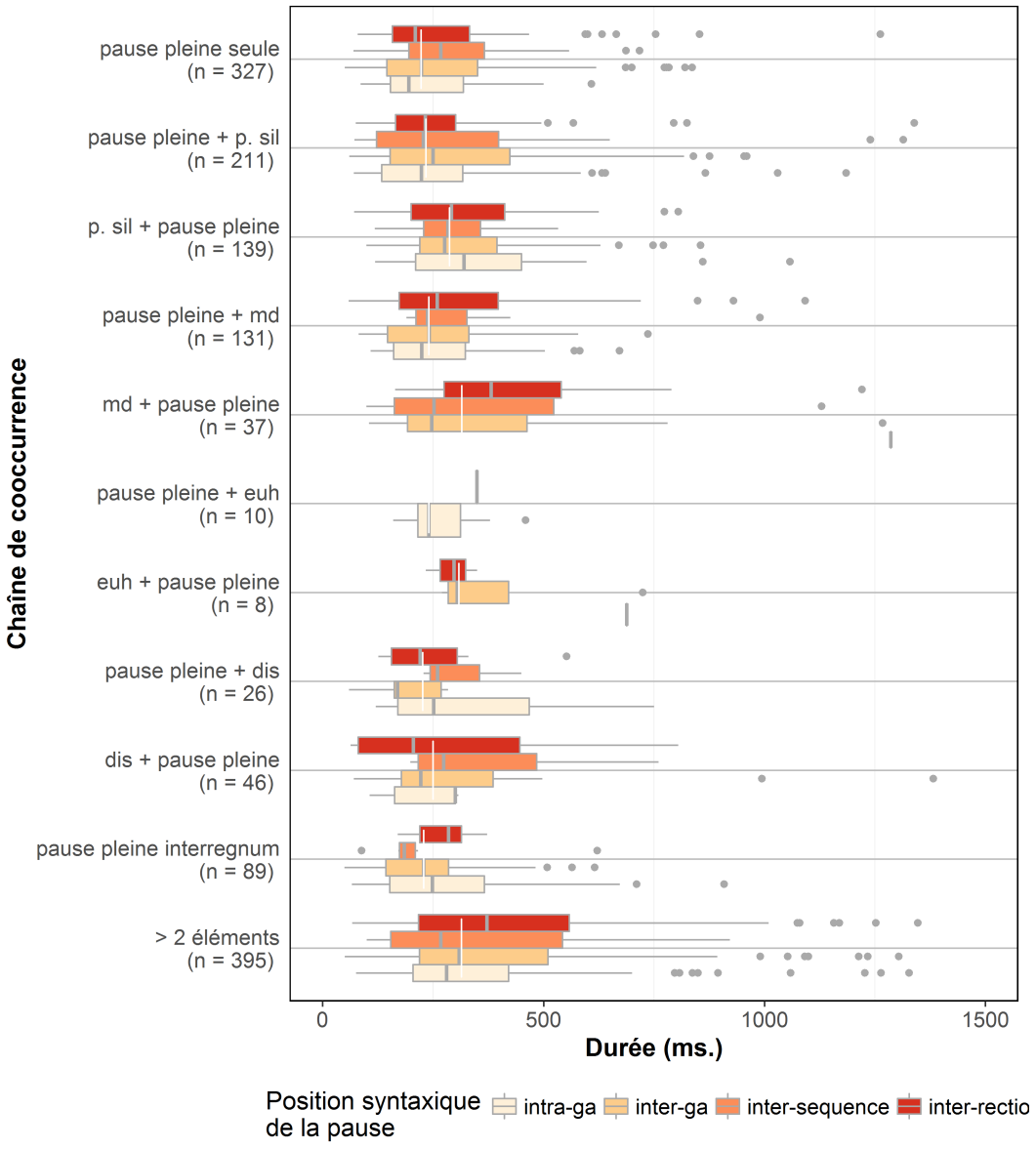


FIGURE 5.14 – Diagramme en boîte à moustaches de la distribution des durées des pauses pleines en fonction de leurs cooccurrences et de leur localisation syntaxique.

21. Le nombre d'occurrences de certaines configurations est inférieur à 20, notamment celui des chaînes de pauses pleines cooccurrentes ou de md cooccurrent à un allongement. Nous nous focalisons ci-après uniquement sur l'observation des chaînes les plus courantes.

Pour les allongements, leur durée en position intra-ga est particulièrement supérieure lorsqu'**en cooccurrence avec une autre (dis)fluence**. Les allongements intra-ga semblent plus longs que les allongements inter-rectionnels, bien que cette observation ne puisse être confirmée par des analyses statistiques plus poussées en raison de la faible fréquence de certaines modalités.

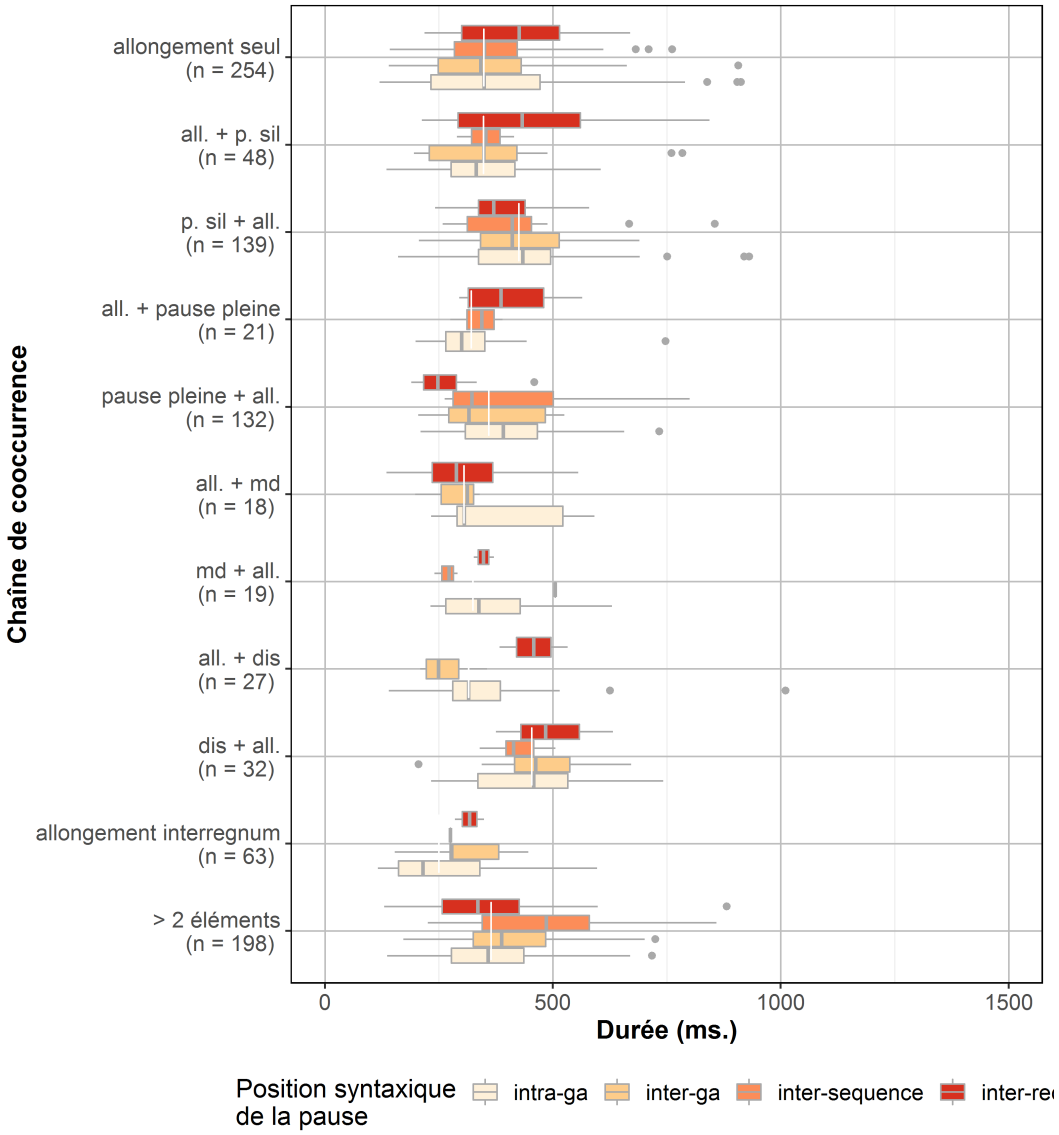


FIGURE 5.15 – Diagramme en boîte à moustaches de la distribution des durées des allongements en fonction de leurs cooccurrences et de leur localisation syntaxique.

Ces analyses de chaînes cooccurrentes sont pertinentes pour la compréhension des phénomènes au sein de l'interaction. Nous cherchons ici à savoir si l'actualisation de ces marques est associée à la **tenue /vs/ passation d'un tour de parole**[22]. Pour évaluer la position des pauses pleines et des allongements dans l'interaction, nous établissons un sous-corpus **excluant les monologues** des corpus LOCAS-F et C-HUMOUR[23].

Méthodologiquement, nous avons attribué à chaque marqueur une des **trois (macro)positions** : (1) début de tour, (2) interne au tour et (3) fin de tour de parole. Le « début » et la « fin » de tour étant inadaptés à la description des tours monosyllabiques ou monolexicaux, nous avons distingué ces catégories. Afin d'obtenir une **granularité plus fine** des analyses, nous avons ensuite identifié au sein de ces catégories de « débuts » et de « fins » de tour les marqueurs étant en **frontière de tour /vs/ proche d'une frontière de tour**. La proximité par rapport à une frontière de tour est définie en fonction de la durée relative du tour de parole[24] : une position initiale de début de tour englobe alors les 10 premiers pourcents de syllabes ou mots, tandis qu'une position finale en fin de tour 10 derniers pourcents de segments au sein d'un tour[25].

Le Tableau 5.2 révèle la distribution des pauses pleines par rapport aux tours de parole au sein des dialogues et des multilogues de LOCAS-F et de C-HUMOUR. On relève une différence significative de la position des pauses pleines au sein des tours de parole. Soulignons une grande proportion de pauses pleines intratours ($\chi^2 = 32.9, df = 6, p < 0.014$), avec (1) une actualisation relativement plus fréquente en début de tour qu'en fin de tour, et (2) des proportions relativement similaires en frontière de début de tour qu'en frontière de fin de tour. Ainsi, **la présence d'une pause pleine, isolée ou en chaîne ne semble *pas* à elle seule un indice suffisant de rupture du tour de parole**.

Position	**initiale**		**médiane**			**finale**	
	frontière	début	tour mono	tour court	tour long	fin	frontière
pas de pause pleine	1466	2197	534	3215	27749	3264	2064
pause pleine	44	62	20	40	645	63	62
% par tour	3.00	2.82	3.75	1.24	2.32	1.93	3.00
% par pause pleine	4.70	6.62	2.14	4.27	68.91	6.73	6.62

TABLEAU 5.2 – Position des pauses pleines par rapport aux tours de parole des dialogues et des multilogues.

22. L'annotation en tour de parole est définie au chapitre 3 section 3.1.
23. Une étude similaire pourrait être effectuée sur la position des (dis)fluences en fonction des unités inter-pausales incluant ainsi les monologues.
24. i.e. Position de la syllabe d'intérêt divisée par le nombre de syllabes au sein du tour.
25. Cette distinction semble moins pertinente pour les tours de moins de 10 syllabes ou de moins de 10 mots. Nous les avons ainsi distingués en catégorie « tour court ». Pour ces derniers, nous n'avons pas appliqué de distinction frontière /vs/ début.

La cooccurrence de marques au sein des chaînes du sous-corpus [26] révèle que : les pauses pleines seules ($n = 135$) sont relativement présentes en début et en fin de tour et plus rares au sein de tour de plus de 10 mots, tandis qu'elles sont, relativement aux autres chaînes, très présentes au sein de tours courts (4.2% des occurrences). Les chaînes > à 2 éléments ($n = 163$) ne provoquent pas plus de frontières de fin de tour que les autres chaînes. Par ailleurs, on trouve relativement moins de chaînes (pause silencieuse + pause pleine) ($n = 60$) et (pause pleine + pause silencieuse) ($n = 71$) en frontière de début ou de fin de tour de parole. Ces dernières semblent ainsi **davantage une marque de continuité**.

Les chaînes (md + pause pleine) ($n = 19$) et (pause pleine + md) ($n = 59$) sont relativement bien présentes en frontières de fin et de début de tour. Elles pourraient marquer des **lieux propices de transition**. La chaîne (dis + pause pleine) ($n = 29$) semble davantage un lieu propice de transition (présence importante en début et en fin de tour) que la chaîne (pause pleine + dis), relativement plus observée au sein de tour court (11.54%) quoique le nombre d'occurrences réduit de ce type de chaînes ne permette pas de généraliser ces observations ($n = 11$). Enfin, les pauses au sein d'une phase d'édition ($n = 42$) sont davantage présentes en frontière de début et en début de tour et plus marquées qu'en fin de tour, ce qui confirme l'hypothèse selon laquelle une (dis)fluence, même porteuse d'une phase d'édition, est synonyme de reprise ou de continuation.

Le Tableau 5.3 révèle la distribution des allongements par rapport aux tours de parole au sein des dialogues et des multilogues de LOCAS-F et de C-HUMOUR. Il montre que les allongements ne semblent pas associés plus particulièrement à la fin d'un tour de parole ou au début. Au contraire, pour les tours supérieurs à 10 syllabes, près de **79.45% des allongements se trouvent en position médiane de tour**.

Position	**initiale**		**médiane**			**finale**	
	frontière	début	tour mono	tour court	tour long	fin	frontière
syll. sans all.	1523	2843	520	2032	35029	4087	2245
syll. allongée	16	63	0	4	580	48	19
% par tour	1.05	2.22	0.00	0.20	1.66	1.17	0.85
% par all.	2.19	8.63	0.00	0.55	**79.45**	6.58	2.60

TABLEAU 5.3 – Position des allongements par rapport aux tours de parole des dialogues et des multilogues.

Les trois chaînes les plus observées **en fin et en frontière de tour**, sont les **allongements isolés** ($n = 29$), les chaînes **> à 2 éléments** ($n = 22$), les chaînes **(pauses pleine + allongement)** ($n = 18$). Cependant, les allongements de ces chaînes représentent respectivement 1.24%, 0.77% et 1.25% des marqueurs de catégories respectives. Ainsi, ces configurations ne sont pas particulièrement associées à la rupture. Ce qui ressort du fond sont les catégories (md + allongement) et (dis + allongement) qui ont

26. Les tableaux relatifs à ces données sont repris en Annexe E section 2.1.

tendance à s'actualiser en début et en fin de groupe. En effet, 23.05% des occurrences **(md + allongement) sont en position finale de tour de parole** ($n = 6$ pour un total de chaînes (md + allongement) de $n = 26$ dans ce sous-corpus.).

En conclusion, le croisement des occurrences de pauses pleines et d'allongements ne semble pas indiquer que ces marques soient à elles seules associées à une rupture du tour de parole. La cooccurrence de marque n'est pas plus porteuse de rupture interactionnelle que d'autres chaînes isolées. Les chaînes les plus présentes en frontière de fin de tour de parole sont celles associées à des marqueurs discursifs, mais celles-ci sont davantage des lieux de transition que des lieux de rupture, étant donné leur présence en début de tour. Dans le cadre de la modélisation de l'évaluation des pauses pleines et des allongements, nous mettons en relation ces résultats avec la durée même des cooccurrences, leur position syntaxique, et les contours aux frontières intonatives[27].

27. Voir section 3 de ce chapitre.

1.5 Fréquence et durée des pauses pleines et des allongements en fonction des contours aux frontières intonatives

Nous testons si la **durée d'une pause pleine ou d'un allongement est corrélée au contour intonatif de la syllabe la/le précédant**. Toute situation confondue, l'observation de la Figure 5.16 indique que la durée médiane des pauses pleines de LOCAS-F fluctue de 229 à 270 ms. en fonction du contour intonatif précédant la pause pleine [28].

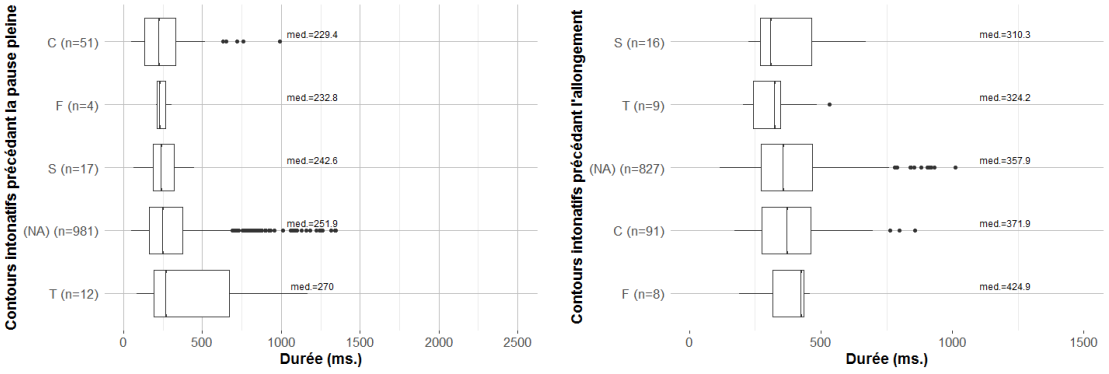


FIGURE 5.16 – Nombre d'occurrences et médianes de la durée des pauses pleines (*à gauche*) et des allongements (*à droite*) en fonction du contour intonatif la précédant.

Une analyse de la variance (ANOVA) sur ces résultats indique qu'on ne peut confirmer une variation significative entre les conditions pour les pauses pleines ($F(4, 1060) = 23.734, p < 0.34$). La durée des pauses pleines ne varie pas en fonction du contour la précédant. Lorsque l'annotation de la pause est précédée d'un contour, il est **majoritairement continuatif** ($n = 51$). Au contraire, on peut confirmer une variation significative entre les conditions pour les allongements, ($F(4, 1806) = 3.2, p < 0.02$). Ainsi, **la durée des allongements varie en fonction du contour le précédant**. Lorsque l'allongement est précédé d'un contour, celui-ci est majoritairement continuatif ($n = 251$). Cependant, ces résultats sont dus au fait que le protocole d'annotation des contours [29] empêche l'annotation des contours lors de la production d'une « hésitation ». En raison des choix du protocole d'annotation des contours, nous ne possédons pas les données nécessaires à la validation de nos hypothèses.

Nous enquêtons donc plutôt sur la forme intonative des *euh* en eux-mêmes plutôt qu'aux mouvements des contours à leurs frontières. Selon (DUEZ 1991:39), les pauses pleines se distinguent selon 9 patrons [30]. D'un point de vue théorique, chez MOREL & DANON-BOILEAU (1998), le *euh* et l'allongement ont des caractéristiques prosodiques communes et donc des similarités iconiques telles que une plage de F0 basse, associée à un **repli**

28. Dans cette section, les durées des pauses sont représentées dans les graphiques en millisecondes. Cependant, nos statistiques sont effectuées sur les données transformées en logarithme de base 10. Nous les reportons donc comme tel. Lorsque nécessaire, nous retransformons les résultats en chiffres naturels à des fins de lisibilité.

29. Voir Annexe C section 19.

30. i.e. montant, descendant, descendant+plat, descendant+montant, plat, plat+descendant, plat+montant, montant+descendant et montant+plat.

sur soi, dû au fait que l'énonciateur se concentre sur la **formulation de son propre discours**. Avant de comparer les marques entre elles, nous nous arrêtons sur leur forme prosodique interne en fonction des chaînes de cooccurrences dans lesquelles elles s'intègrent.

Dans nos corpus, sur 1419 pauses pleines, un tiers ont un contour plat ($n = 908$) (i.e. un mouvement intrasyllabique de moins de 2.5 demi-tons). Seulement 472 pauses pleines présentent un contour dynamique descendant (33.3%) et 39 un contour dynamique montant (2.7%). La majorité des pauses pleines affichent de ce fait des **contours stables** (64%). Nous ne confirmons donc pas les observations selon lesquelles les courbes intonatives internes des pauses pleines et des allongements tendraient majoritairement à reproduire une courbe descendante linéaire droite (Duez 2001b ; O'Shaughnessy 1992 ; Shriberg 1999).

Ainsi, si au niveau des cooccurrences des pauses pleines, la répartition des contours montants, plats ou descendants est significativement différente en fonction des chaînes de (dis)fluences ($\chi^2 = 128.72, df = 20, p < 0.001$), nous relevons au sein de chaque coocccurence les particularités suivantes :

- Les chaînes exhibant significativement le plus de contours montants sont celles (1) > à 2 éléments, (2) de cooccurrences de pauses pleines, et dans une moindre mesure (résidus entre [0.2-2.4[), les chaînes (dis + pause pleine), (4) les chaînes (pause silencieuse + pause pleine), et (5) les chaînes (pause pleine + pause silencieuse).
- La chaîne exhibant significativement le plus de contours plats est celle de (pause pleine + pause silencieuse) qui rejette l'actualisation de contours descendants.
- Au contraire, la chaîne (pause silencieuse + pause pleine) rejette l'actualisation d'une courbe plate, et favorise une courbe descendante (résidu > 4).

Cependant, ces variations de pente restent mineures et sont donc à relativiser. En accord, avec Guaïtella (1991), O'Shaughnessy (1992) et Duez (2001a), nous confirmons les observations selon lesquelles, lorsque descendante, la pente de F0 est marquée par moins de 20% de déclinaison entre le début et la fin de la pause. La Figure 5.18 présente la hauteur relative [31] de chaque pause pleine (début et fin de segment) en fonction de son contexte de cooccurrence [32]. Nous remarquons que la chute de plus de 2.5 tons reste marginale. Nous pouvons remettre en cause le fait que cette chute soit perceptible par l'oreille humaine.

Du côté des 857 allongements analysés, les contours montants (10.1%) sont plus couramment observés que pour les pauses pleines. Tandis que respectivement 54.4% et 33.5% de contours plats et descendant sont observés.

31. La formule pour convertir des Hz aux demi-tons (s.t., semi-ton) relatifs à 1 Hz est : s.t. = 12*log2(X), où X est la valeur en Hz. La formule pour convertir des semi-tons en Hz est : Hz = 2(X/12), où X est le s.t. relatif à 1 Hz. La conversion Hz semi-ton est un artefact mathématique permettant de rapprocher la représentation perceptuelle de l'intonation à celle de la musique. Nous renvoyons à Nolan (2003) et Patel, Iversen & Rosenberg (2006) pour une justification plus détaillée.

32. Nous avons exclu les cooccurrences de pauses pleines qui représentent moins de 20 occurrences.

Il existe par ailleurs une différence significative des types de mouvements intonatifs (montant, plat, descendant) en fonction des cooccurrences ($\chi^2 = 68.9, df = 20, p < 0.001$), bien que, comme en témoigne la Figure 5.17, ces variations restent mineures. Les tendances les plus notables sont observées pour les allongements d'hésitation précédés d'une pause, qui sont majoritairement marqués par un contour descendant (résidu > 4) et très rarement par un contour plat (résidu < -4.2). Enfin, la cooccurrence (pause pleine + allongement) est elle caractérisée par moins de contours descendants que toutes les autres cooccurrences (résidu < -4.2).

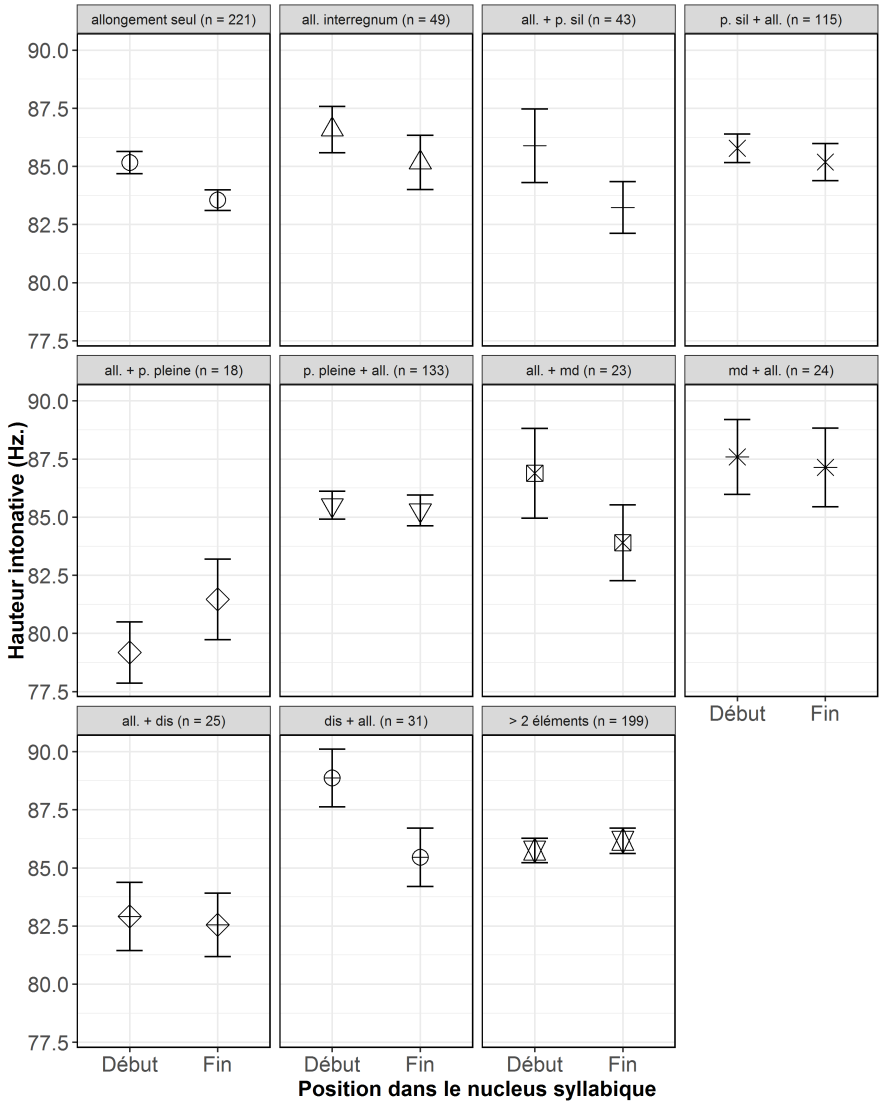


FIGURE 5.17 – Représentation du mouvement intonatif intrasyllabique des allongements en fonction des cooccurrences.

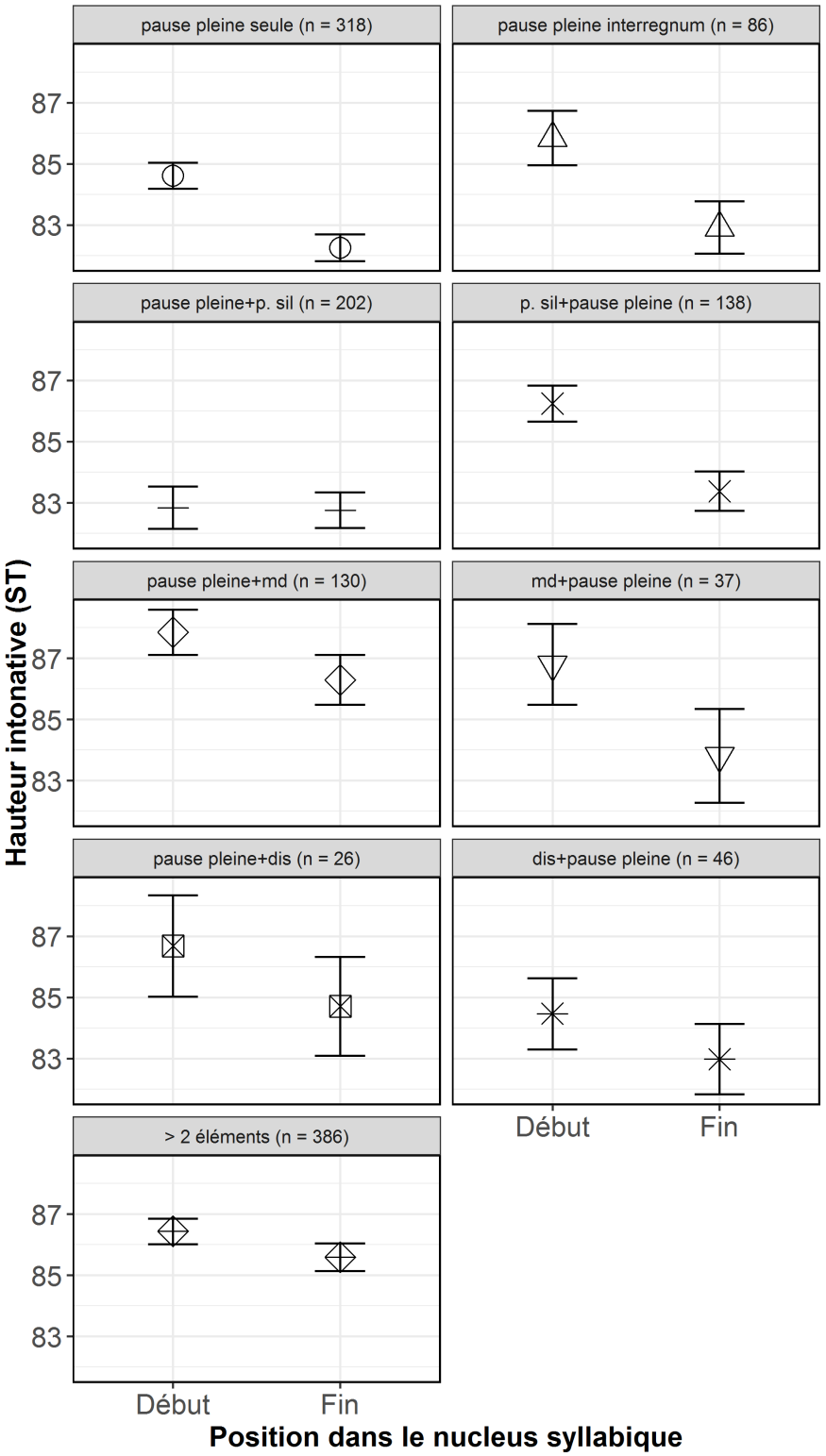


FIGURE 5.18 – Représentation du mouvement intonatif intrasyllabique des pauses pleines en fonction des cooccurrences.

2 Evaluation perceptive des pauses pleines et des allongements

La section antérieure a mis en évidence la **variation distributionnelle** des pauses pleines et des allongements au regard de **facteurs situationnels, prosodiques et syntaxiques**. En confrontant les résultats des données de corpus et les **différentes campagnes d'annotation** (i.e. *online, offline, manuelle, automatique*), cette section précise les facteurs favorisant la perception et/ou le jugement (dis)fluent des pauses pleines et des allongements comme « marqueurs d'hésitation ».

Pour faire écho aux analyses des données de corpus, du côté des auditeurs, les pauses pleines seraient jugées « fluentes » lorsqu'elles s'actualisent aux frontières de constituants prosodiques et/ou syntaxiques, et lorsque leur fréquence fondamentale est stationnaire. Moniz, Mata & Viana (2007) montrent que les auditeurs jugent *dis*fluentes des marques produites avec un **contour intonatif descendant** ou dans un **registre intonatif** bas, et surtout plus bas que le *reparandum* et le *reparans* des (dis)fluences cooccurrentes. En effet, seuls ces contours de **F0 stationnaires** et en frontières de constituant(s) prosodique(s) et/ou syntaxique(s) seraient jugés acceptables par les auditeurs. Le profil acoustique de la pause pleine serait alors **proche de celui des parenthétiques**. Nous évaluons l'influence des paramètres expérimentaux (annotation *online* /vs/ *offline*), acoustiques (allongement objectif de la syllabe, courbes intonatives), syntagmatiques (cooccurrence des (dis)fluences), morphosyntaxiques (types de lexème), syntaxiques et prosodiques (position du phénomène dans une unité) sur la détection et la perception des (dis)fluences syllabiques. La seconde section de ce chapitre teste l'impact de la méthodologie d'annotation des pauses pleines et des allongements sur leur annotation, notamment en fonction de leur durée, forme acoustique et des cooccurrences.

Cette recherche permet alors de spécifier dans **quels contextes** les pauses pleines et les allongements ont été **annotés de manière univoque**, et quels sont les corrélats acoustiques entraînant la perception d'une hésitation. Pour ce faire, deux expériences ont été réalisées[33]. La première compare les annotations de trois linguistes « en écoute segmentée » (*offline*) (section 2.1). La seconde compare les annotations d'une annotatrice, effectuées de manière continue en ligne (*online*) à deux autres méthodologies d'annotation : celle de nos corpus (manuelle et en écoute segmentée) et celle résultant d'une annotation automatique des allongements (section 2.2). Pour ces deux expériences, nous mettons en parallèle les résultats d'annotation avec le matériel du corpus. Nous évaluons donc ce qui soutient, de manière générale, l'annotation et la perception de la pause pleine ou de l'allongement, mais également son évaluation comme « marque d'hésitation ».

33. Les résultats de ces expériences sont le fruit de riches collaborations. Un remerciement particulier revient aux chercheurs Lucie Rousie-Vercruyssen, George Christodoulides et Emilie Duvivier. Les publications et les présentations dont elles ont fait l'objet sont pertinemment citées à chaque section.

2.1 Perception *offline* des pauses pleines et des allongements par des annotateurs experts

Cette étude est doublement motivée. Partiellement présentée au chapitre 3 section 2.1.2, elle offre un récapitulatif et une discussion méthodologique transparente autour de l'annotation des pauses pleines, allongements et autres phénomènes d'hésitation [34] dans le cadre d'évaluations interannotateurs. Cette étude de cas éclaire par là même l'influence des paramètres prosodiques sur la perception des pauses pleines et des allongements dits d'hésitation [35].

2.1.1 Objectif et hypothèses

Quelles variables acoustiques, prosodiques et syntaxiques prédisent le mieux l'annotation des pauses pleines et des allongements en français ? L'allongement est associé de manière inhérente à au moins deux phénomènes de parole en français. Les données de corpus montrent que **l'allongement est marqueur de frontière prosodique** ; par conséquent, il interagit aussi très souvent avec d'autres (dis)fluences.

On formule l'hypothèse selon laquelle la **présence d'une frontière et/ou d'une (dis)fluence influence la perception des allongements** ; et par là même, que la durée n'est pas le seul paramètre permettant de prédire l'actualisation d'un allongement en français (e.g. *le mouvement intonatif syllabique, la cooccurrence avec d'autres marques d'hésitation, la position par rapport aux constituants prosodiques et syntaxiques,* etc.). Nous formalisons les hypothèses qui seront vérifiées par la comparaison du degré d'accord interannotateurs issu de la perception et de l'évaluation d'une hésitation :

- **influence de la catégorie grammaticale** : augmentation de l'accord interannotateurs sur les items lexicaux ;
- **influence de la durée relative** : augmentation de l'accord interannotateurs sur les syllabes plus longues ;
- **influence du mouvement intonatif interne** : augmentation de l'accord interannotateurs sur les syllabes à courbe intonative stationnaire ;
- **influence de la fréquence fondamentale relative** : augmentation de l'accord interannotateurs sur les syllabes intonativement marquées (avec mouvement descendant, très hautes ou très basses) ;
- **influence des autres hésitations** : augmentation de l'accord interannotateurs lors de cooccurrence(s) de (dis)fluences ;
- **influence de la position de l'unité séparée par des pauses** : augmentation de l'accord interannotateurs en position initiale ou médiane (car l'allongement final de groupe est quant à lui attendu).

34. Les résultats de cette étude sont le fruit d'une collaboration. Ils ont fait l'objet d'une présentation : I. Grosman, L. Rousier-Vercruyssen & E. Duvivier. 2016. « Perception des hésitations : liens entre accord inter-annotateurs et contextes prosodiques. » *Journée annuelle du Cercle Belge de Linguistique.* Louvain-la-Neuve, Belgique.

35. Dans cette section, le mot allongement se réfère à « allongement d'hésitation » et non « allongement grammatical » (voir chapitre 2 section 2.2).

2.1.2 Données et méthode

Le corpus d'expérience regroupe **six enregistrements** extraits de **récits narratifs** issus de LOCAS-F (Martin, Degand & Simon 2014) et des **narrations d'images séquentielles** du corpus de thèse de L. Rousier-Vercruyssen (Fossard et al. 2018)[36]. Ces enregistrements ont été sélectionnés en raison de leur **relative similarité discursive** : dans les termes du paradigme de la communication référentielle, la narration d'images séquentielles et le récit conversationnel sont des **interactions verbales collaboratives** (Clark 1986). Il s'agit de monologues non préparés, non médiatisés et semi-interactifs – les locuteurs ayant des statuts asymétriques. Les fichiers `rcv_1`, `rcv_2` et `rcv_3` sont extraits de LOCAS-F tandis que les fichiers `barque_S37`, `chaussure_S09` et `chaussures_S49` font partie du corpus de thèse de Rousier-Vercruyssen (Tableau 5.4).

Fichier	Durée (sec.)	Prop (%)
`barque_S37`	55.9	7.6
`chaussure_S09`	27.4	3.7
`chaussures_S49`	30.6	4.2
`rcv_1`	206.5	28.1
`rcv_2`	217.6	29.6
`rcv_3`	197.9	26.9
Total	**735.9**	**100**

Tableau 5.4 – Durées et proportions des enregistrements du sous-corpus.

Le sous-corpus est constitué de 12 minutes de parole annotée (3266 syllabes). Les corpus sont transcrits, segmentés au niveau lexical, syllabique et phonétique, et alignés au son. Les enregistrements ont été annotés sous Praat (Boersma & Weenink 2015) par trois linguistes francophones. Les fichiers originaux fournis aux annotateurs contiennent les couches d'annotations `phones, syll, hesi` et `words` au format TextGrid (Figure 5.19). Les données TextGrid de chaque annotateur ont été fusionnées afin de créer un unique Textgrid rassemblant les annotations des différentes annotatrices LRV, ED & IG (Figure 5.20). Les Textgrids ont ensuite été analysés via les plugin Prosogram (Mertens 2004) & Prosoprom (Goldman, Avanzi, Auchlin & Simon 2012) afin d'obtenir une **stylisation de la fréquence fondamentale**[37] et une **annotation graduelle des proéminences**[38].

36. Ces résultats sont en partie présentés dans le travail doctoral de L. Rousier-Vercruyssen. 2017. « Prosodie, Syntaxe, et Référence : processus cognitifs et marqueurs linguistiques ». Thèse de doct. Paris, France : Université Paris 10, projet FNS n°140269.

37. i.e. Simplification mathématique permettant de discrétiser l'analyse du signal sonore en segments syllabiques montants /vs/ descendants.

38. Voir Annexe C section 4.

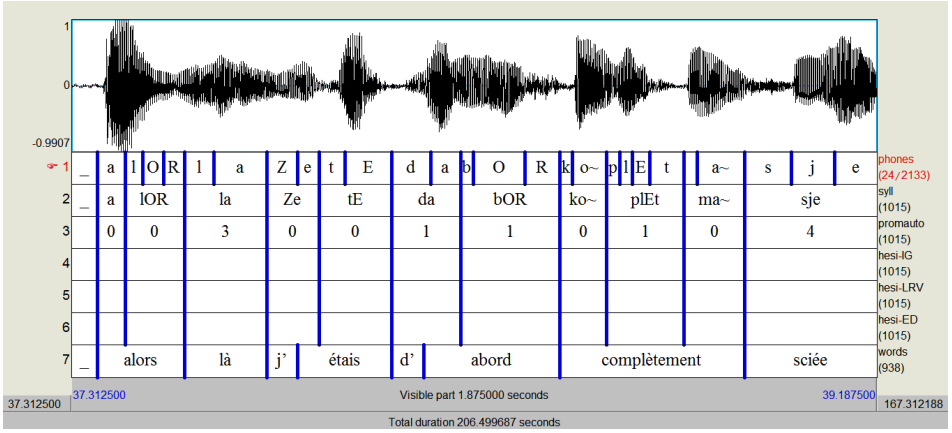


FIGURE 5.19 – TextGrids avant annotation.

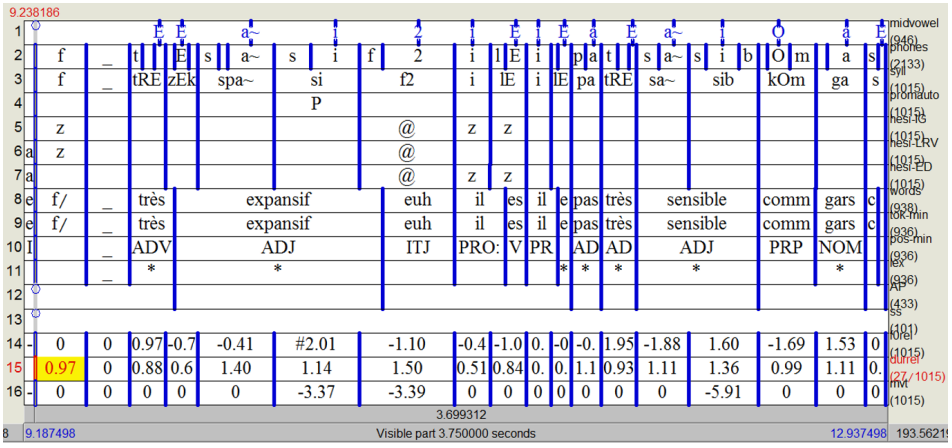


FIGURE 5.20 – TextGrids après analyse prosodique.

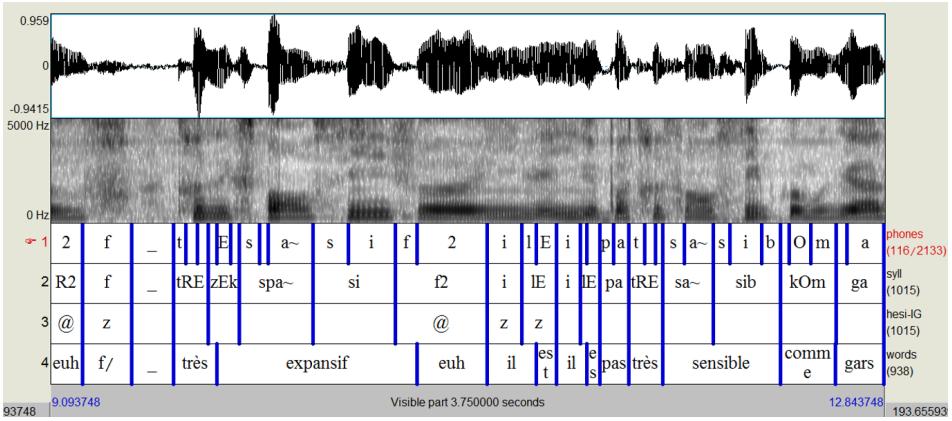


FIGURE 5.21 – TextGrids après annotation.

Quatre catégories de syllabes ont été annotées : les (1) syllabes non disfluentes, (2) les syllabes jugées porteuses d'un allongement dit d'hésitation, (3) les syllabes représentant une pause pleine [39], et (4) les syllabes du *reparandum* de quatre autres (dis)fluences (*répétitions, amorces, autocorrections, insertions*). Le protocole de définition des phénomènes demandait aux annotateurs d'annoter de la manière suivante :

Phénomène	**Etiquette**	**Définition**
Pauses pleines	@	Annotation des *euh* et de leurs variantes phonétiques (*euhm*, *hmm*, etc.), structurants comme d'hésitation (i.e. sans interprétation de leur fonction langagière), mais en excluant les « mm » motivés par un « *back-channel* ». Cette annotation inclut également celle des « *euh* » dits post-toniques (e.g. « *elle est sorti•e* »), qu'ils soient détachés intonativement ou non.
Allongements d'*hésitation*	a	Annotation de la perception subjective des voyelles et des consonnes tenues au sein ou en fin de mot, en excluant les allongements inhérents à la grammaire de la langue (allongement de fin de groupe) ou les allongements dus à une emphase (voir notamment Candea (2000a:262)).
Autres phénomènes d'hésitation	z	Tout autre phénomène dit « d'hésitation » (i.e. la première partie d'une répétition, ou d'une autocorrection).

Tableau 5.5 – Protocole d'annotation des pauses pleines et des allongements dits d'hésitation.

2.1.3 Instructions

Les consignes d'annotations étaient les suivantes :

- L'annotateur est requis d'**écouter 50 à 100 secondes** de l'enregistrement afin de se familiariser avec l'**idiostyle du locuteur** – surtout dans le cas d'un accent francophone qui lui est étranger [40]. Cette étape est nécessaire à la bonne distinction entre les allongements inhérents à l'idiostyle du locuteur et ceux d'hésitation. Elle permettra à l'annotateur de ne pas adopter un **point de vue normatif** décontextualisé de l'allongement mais au contraire de se rapprocher dans **la norme du parler du locuteur en question**.
- L'annotateur utilise une fenêtre d'écoute allant de **3.5 à 6 secondes** autour de la syllabe d'intérêt.
- L'annotateur se limite au maximum à **trois écoutes** par fenêtre d'annotation.

39. Schwa post-tonique, *euh* isolé, mais également *hum* et *mh*.
40. Les trois annotatrices ayant des passes linguistiques différents (belge, français et suisse)

– Dans le cas où un phénomène d'hésitation est allongé (« le:: le petit enfant ») l'annotation de l'allongement « a » prime sur celle de l'hésitation.
– Dans le cas d'une répétition, on annote le(s) segment(s) répétable(s) en utilisant l'étiquette « z ».
– Une couche d'annotation supplémentaire optionnelle est mise à disposition des annotateurs pour d'éventuels commentaires (e.g. *doutes sur l'annotation, écoutes supplémentaires, passages inaudibles,* etc.).

2.1.4 Analyses et résultats

Accords interannotateurs

Les résultats sur l'**accord global** indiquent un **fort taux d'accord interannotateurs** avec un score Kappa Fleiss de 0.715 ($p < 0.001$), ce qui représente une convergence substantielle sur la tâche exécutée de manière générale, suggérant une perception globale semblable des hésitations.

Les résultats du tableau croisé de l'accord interannotateurs permettent de différencier les syllabes pour lesquelles aucun des annotateurs n'était d'accord, deux sur trois l'étaient et finalement un accord parfait où tous les annotateurs étaient d'accord (Tableau 5.6). Cet accord parfait atteint un taux élevé (3026 cas soit 92,68%). Enfin 233 marques ont été perçues par deux annotateurs tandis qu'à peine 6 marques l'ont été par une seule annotatrice.

	Accord nul	Accord 2 annotateurs	Accord parfait
n – Freq	6	233	3026
% – Pourcent	0.18	7.14	92.68

TABLEAU 5.6 – Tableau croisé d'accord interannotateurs sur toutes les marques.

Avant de se concentrer sur l'influence des paramètres acoustiques sur l'évaluation d'un allongement dans le cadre d'une modélisation de l'accord interannotateurs, nous nous focalisons sur (A) l'accord par facteur et sur (B) les résultats détaillés par annotateurs et par fichiers.

A. Accord interannotateurs par facteur

Les détails de ces scores par (dis)fluence sont également concluants (Tableau 5.7 [41]). L'accord sur les pauses pleines est le plus élevé ($K = 0.86$), sans doute en raison du biais graphique lors de l'annotation. En regardant les données, les erreurs des pauses pleines se concentrent majoritairement sur l'évaluation de schwa post-toniques. Le taux d'accord est relativement bon pour les allongements perçus comme des hésitations ($K = 0.63$) ainsi que pour les *reparandum* des autres (dis)fluences ($K = 0.44$).

41. *** indique une significativité de $p < 0.001$.

Phénomène	Kappa	Z	Sig.
Tout segment	0.72	75.53	***
Segment non annoté	0.76	74.78	***
Pause pleine	0.86	85.55	***
Allongement d'hésitation	0.64	63.11	***
Reparandum d'autres (dis)fluences	0.44	43.19	***

TABLEAU 5.7 – Accords interannotateurs par type de (dis)fluences au niveau syllabique.

L'accord interannotateurs des allongements est lui marqué par un accord substantiel. De plus, parmi les syllabes annotées comme allongement **aucune n'a fait le cas d'un désaccord total** entre les annotateurs (Tableau 5.8). Nous y reviendrons dans notre modélisation.

Allongement	Accord nul	Accord 2 annotateurs	Accord parfait
n – Freq	0	61	64
% – Pourcent	0	49	51

TABLEAU 5.8 – Tableau croisé d'accords interannotateurs des allongements.

B. Résultats détaillés – Biais annotateurs et tendances par fichiers

Nous observons maintenant les résultats par annotateur et par fichier. Existe-t-il un biais par annotateur et/ou interannotateurs ? La tendance par annotateur sur les trois phénomènes d'intérêt révèle globalement que l'annotateur LRV a plus tendance à annoter les allongements que tous les autres annotateurs (Figure 5.22).

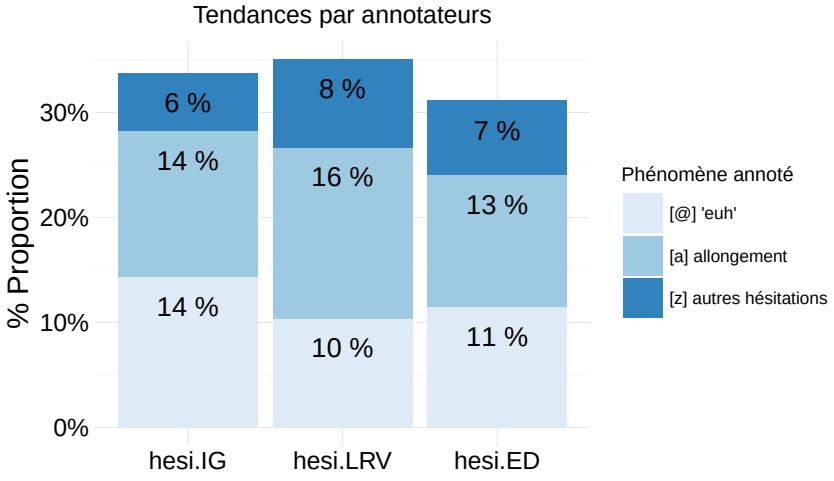


FIGURE 5.22 – Diagrammes en barres des tendances d'annotation par annotateur et par phénomène.

L'annotatrice IG tend par contre à annoter davantage de schwa post-toniques. Enfin, ED a tendance à moins annoter et ce de manière globale, à l'exception de la catégorie « autres hésitations ». On peut formuler l'hypothèse que l'incertitude d'une annotation l'a conduite soit à ne pas annoter soit à annoter dans la catégorie « autre ».

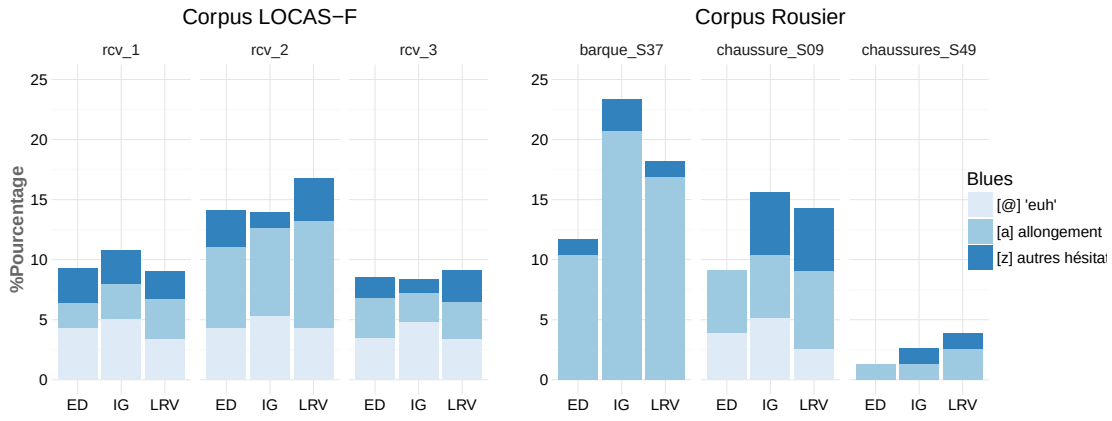


FIGURE 5.23 – Diagramme en barres des tendances d'annotation par annotateur, par phénomène et par fichier.

Certains enregistrements provoquent plus d'accord que d'autres, c'est le cas de : `barque_S37` qui contient le plus d'annotations et le plus de désaccords contre `rcv_3` qui contient globalement peu d'annotations et un taux d'accord concluant (Figure 5.23). De manière générale, les annotations produites par ED et IG ont l'air plus similaire que celles produites par LRV. On peut formuler l'hypothèse selon laquelle la perception et l'évaluation des évaluatrices belges seraient proches de celles de l'évaluatrice suisse (LRV).

In extenso, certains couples d'annotateurs obtiennent un score d'accord plus élevé (Figure 5.24). Les différences de résultats restent minimes. C'est le cas du couple d'annotateurs IG-ED ($K = 0.746$) contrairement au couple LRV-ED ($K = 0.698$), tandis que le couple IG-LRV est dans la moyenne ($K = 0.702$). Dans le panel supérieur gauche, on observe que 29 syllabes non annotées par LRV ont été marquées comme allongées par IG, tandis que 32 non annotées par IG ont été annotées comme allongement par LRV. Ce désaccord est encore plus marqué entre ED et LRV.

2.1.5 Modélisation des allongements en fonction des résultats d'expérience

La suite de cette étude examine si la forme acoustique et/ou l'éventuelle présence d'autres hésitations influence(nt) le jugement des annotateurs. Comment prédire l'allongement d'hésitation ? Quelles sont ses particularités prosodiques lorsque celui-ci est perçu ? La **présence d'une frontière intonative**, marquée habituellement en français par une augmentation de la durée syllabique, a-t-elle des effets sur l'annotation de l'allongement ? Le modèle s'appuie sur une partie des données précédemment analysées : les syllabes **annotées par au moins un annotateur** comme porteuses d'un allongement.

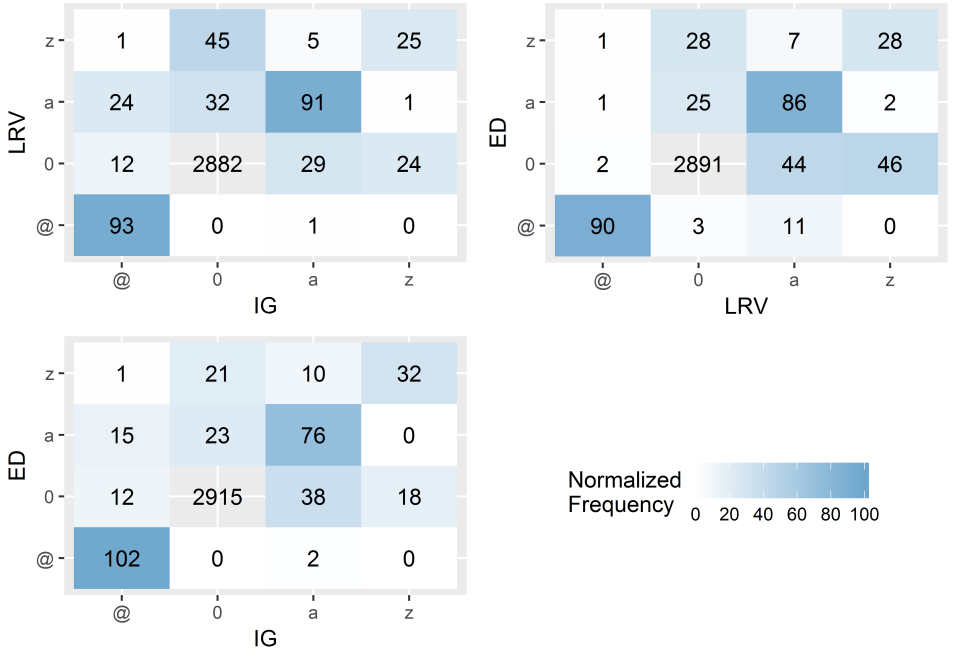


FIGURE 5.24 – Matrices de confusions interannotateurs.

Méthodologiquement, afin de déterminer la potentielle part explicative de ces paramètres sur le jugement des annotateurs, un **modèle mixte pour les variables ordinales** a été appliqué [42]. La particularité du design implique qu'on peut retrouver une même syllabe avec un même contexte prosodique dans un seul fichier. Reconnaître la variable fichier comme **effet aléatoire** permet d'évaluer la variance intra- et interfichier [43].

Nous sélectionnons ensuite les variables potentielles influençant l'accord interannotateurs. Nous questionnons non seulement les **caractéristiques acoustiques intrinsèques** de la syllabe mais aussi le lien entre l'annotation des allongements et la présence de **chaînes de cooccurrence**. Les variables testées sont présentes dans l'encadré ci-après [44].

Les résultats de la modélisation révèlent que les trois paramètres corrélés positivement avec la détection de l'allongement, et ayant une taille d'effet importante sont sans surprise (1) la **durée relative de la syllabe** par rapport à son environnement, (2) la **présence d'une autre hésitation aux alentours** et (3) la **présence d'une pause après la syllabe**. Le Tableau 5.9 se lit ainsi : lorsqu'on passe d'une syllabe sans allongement d'hésitation à une syllabe marquée comme allongée, il y a 93.01% de probabilité que celle-ci ait effectivement une durée plus longue.

42. Les fonctions appliquées sont extraites du module `lme4` (BATES, MÄCHLER, BOLKER & WALKER 2015) sous l'environnement de traitement statistique `R` (*R Core Team* 2016). Ce package permet d'effectuer des modèles linéaires logistiques à effets mixtes sur des données binaires (syllabe allongées /vs/ non allongées). Les deux approches sont testées ici et au chapitre 7.

43. L'apport d'information de la variable aléatoire `fichier` sur la modélisation du score d'accord interannotateurs permet de faire descendre le maximum de likelihood du modèle de manière significative. Les résultats de leur apport est disponible au Tableau E.12 en Annexe E section 2.2.

44. La présence /vs/ absence d'une pause silencieuse après la syllabe d'intérêt, est significativement moins porteuse d'information. Elle est donc retirée du modèle. Les détails de sélection du modèle et le code relatif au modèle mixte sélectionné sont disponibles en Annexe E section 2.2 Tableaux E.13 et E.14.

Par ailleurs, on remarque (étonnamment) que les syllabes dont la durée environnante est égale sont plus marquées par la perception et l'annotation des allongements que celles de durées différentes (2ème ligne du Tableau 5.9). Ce phénomène laisse à penser que certaines syllabes sont effectivement de **durées similaires mais sont annotées comme allongées en vertu d'autres critères prosodiques**. Dans cet ordre d'idées, on note que la **présence d'une autre marque** d'hésitation dans l'environnement syllabique favorise également l'annotation des allongements d'hésitation, avec une taille d'effet aussi très importante (+110.22). De la même façon, la présence de pauses silencieuses avant (et non après) a une taille d'effet tout autant non négligeable sur la tâche de détection.

Certains effets fixes sont **corrélés négativement** avec l'annotation de l'allongement. Dans les tailles d'effet les plus importantes, les syllabes allongées sont plus marquées par **des mouvements intonatifs descendants** (mais stables) (ligne 4). Les allongements sont aussi plus annotés sur les **syllabes aux frontières de groupes accentuels**, surtout **polysyllabiques** – et moins lorsqu'en position médiane et/ou au sein de groupes monosyllabiques (lignes 5 et 7). Ces données sont convergentes avec l'étude sur corpus [45]. La Figure 5.25 résume graphiquement la taille des effets fixes pour la prédiction de la détection des allongements dits d'hésitation.

Certains paramètres sont non significatifs (en rouge dans le Tableau 5.9). Par exemple, l'annotation de l'allongement n'est pas influencée par le caractère **stable /vs/ dynamique** du mouvement intrasyllabique. Le niveau de hauteur différente de la F0 par rapport aux syllabes environnantes n'est pas non plus retenu par le modèle ($p < 0.09$). Les syllabes marquées par l'allongement ont de manière marginale tendance à être caractérisées par un niveau de F0 inférieur aux syllabes environnantes (ligne 6) mais cette différence n'est elle non plus pas significative. Enfin, de manière encore plus marginale et non significative, les syllabes en position initiale de ga sont effectivement davantage détectées comme des allongements d'hésitation que les syllabes en **finale de ga** (5ème ligne), mais elles ne participent pas de manière significative à l'augmentation de l'annotation du phénomène.

45. Voir section 1.4 de ce chapitre.

	Odds Ratios	Estimations	ET	z	Sig	[2.5%	97.5]%
(Intercept)	-94.09	-2.83	0.26	-10.73	0.00	-3.58	-2.08
durrel:sup. /vs/ inf.	93.01	0.66	0.23	2.87	0.00	0.00	1.31
durrel:égal /vs/ diff.	120.21	0.79	0.09	8.67	0.00	0.53	1.05
mvt:montant /vs/ descendant	-59.86	-0.91	0.12	-7.44	0.00	-1.26	-0.56
mvt:stable /vs/ dynamique	5.95	0.06	0.09	0.63	0.53	-0.21	0.32
F0rel:sup. /vs/ inf.	-21.30	-0.24	0.14	-1.71	0.09	-0.64	0.16
F0rel:égal /vs/ diff.	-2.80	-0.03	0.08	-0.36	0.72	-0.25	0.20
ga pos:initial /vs/ final	14.44	0.13	0.17	0.79	0.43	-0.35	0.62
ga pos2:median /vs/ periph.	-24.89	-0.29	0.10	-3.00	0.00	-0.56	-0.01
ga pos:monosyll. /vs/ polysyll.	-20.58	-0.23	0.06	-3.91	0.00	-0.40	-0.06
cooc:cooc. /vs/ non cooc.	110.22	0.74	0.11	6.94	0.00	0.44	1.05
pause-avant:pause /vs/ no pause	59.55	0.47	0.17	2.83	0.00	0.00	0.94

TABLEAU 5.9 – Tableau des effets fixes.

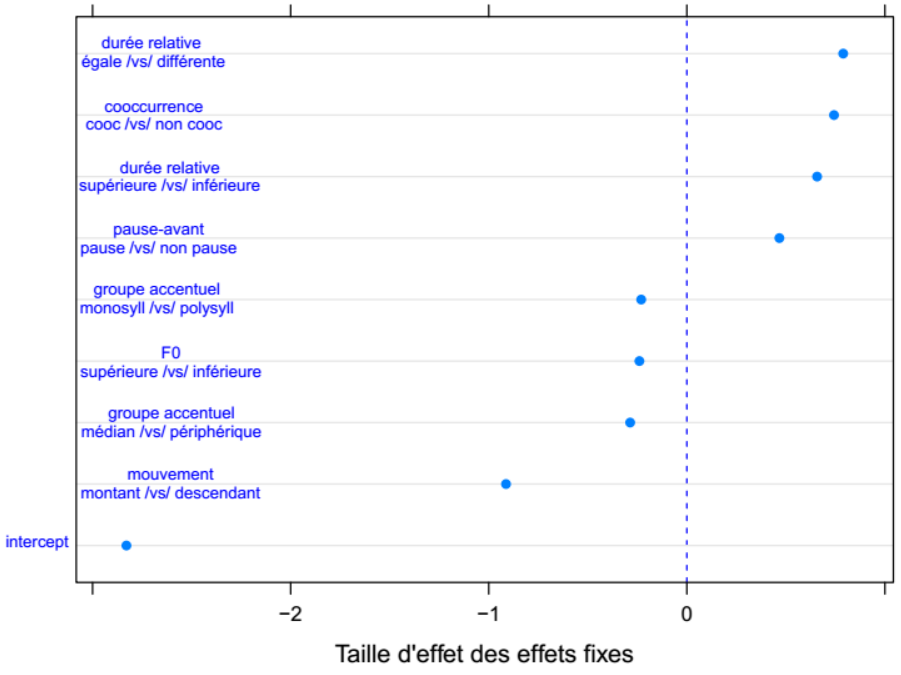


FIGURE 5.25 – Taille des effets fixes pour la prédiction de l'annotation des allongements dits d'hésitation.

Explication des paramètres du modèle :

- `durrel` : `sup.` /vs/ `inf.` – La syllabe d'intérêt est de durée supérieure ou inférieure par rapport aux syllabes environnantes.
- `durrel` : `égal` /vs/ `diff.` – La syllabe d'intérêt est de durée égale ou différente par rapport aux syllabes environnantes.
- `F0 rel` : `sup.` /vs/ `inf.` – La fréquence fondamentale de la syllabe d'intérêt est supérieure ou inférieure par rapport aux syllabes environnantes.
- `F0 rel` : `égal` /vs/ `diff` – La fréquence fondamentale de la syllabe d'intérêt est égale ou différente par rapport aux syllabes environnantes.
- `mvt` : `montant` /vs/ `descendant` – Le mouvement intrasyllabique de la syllabe d'intérêt est montant ou descendant.
- `mvt` : `stable` /vs/ `dynamique` – Le mouvement intrasyllabique de la syllabe d'intérêt est stable ou dynamique.
- `ga pos` : `initial` /vs/ `final` – La syllabe d'intérêt est en position initiale ou finale du groupe accentuel actualisé.
- `ga pos` : `median` /vs/ `periph` – La syllabe d'intérêt est en position médiane ou aux frontières du groupe accentuel actualisé.
- `ga pos` : `monosyll` /vs/ `polysyll` – La syllabe d'intérêt est incluse dans un groupe accentuel actualisé monosyllabique ou polysyllabique.
- `cooc` : `cooc` /vs/ `non cooc` – La syllabe d'intérêt est en cooccurrence ou non avec d'autres (dis)fluences environnantes (+2/-2 syllabes)
- `pause-avant` : `pause` /vs/ `pas de pause` – La syllabe d'intérêt est précédée d'une pause silencieuse.
- `pause-après` : `pause` /vs/ `pas de pause` – La syllabe d'intérêt est succédée par une pause silencieuse.

2.1.6 Conclusions et discussion

Cette étude a cherché à mettre en valeur le fait qu'au **niveau méthodologique** l'accord interannotateurs est plus qu'une simple corrélation et que le retour aux données est essentiel pour évaluer (1) le biais interannotateurs et (2) les paramètres acoustiques et distributionnels prédictifs des allongements d'hésitation. Enfin, dans le cadre d'une modélisation, elle a aussi tenté de se servir des résultats des expériences d'annotation pour une approche plus compréhensive du phénomène.

En résumé, cette étude a montré que l'accord interannotateurs dans une **tâche d'annotation *offline*** est substantiel pour tous les types d'hésitations, particulièrement pour les *euh* et les allongements dits d'hésitation. Nous avons par ailleurs **peu de biais intra-annotateurs et interannotateurs** dans l'expérience. Les **variables les plus informatives** à la modélisation de l'accord interannotateurs pour les allongements sont : la durée relative, la cooccurrence avec d'autres (dis)fluences, pauses antéposées, ainsi que la position dans le ga (et l'information lexicale /vs/ grammaticale), les mouvements intonatifs et la hauteur de F0.

L'augmentation du nombre de données permettrait sans doute de voir émerger des résultats plus significatifs et peut-être des tailles d'effet plus modérées. Cela autoriserait aussi à tester une **modélisation ordinale** (modèle clm) et à évaluer les paramètres prosodiques favorisant la distinction des **allongements dits d'hésitation /vs/ allongements grammaticaux**. Ou encore à différencier les syllabes qui, dans un même contexte, sont allongées mais parfois marquées d'hésitation ou non, et correspondent donc à d'autres allongements – dits grammaticaux. C'est ce que nous effectuons ci-après.

2.2 Comparaison des méthodologies *online* et *offline* de l'annotation des pauses pleines et des allongements d'hésitation

Nous comparons les résultats de trois méthodologies d'annotation des phénomènes liés à l'hésitation dans la parole. Comment les annotations en ligne en continu et hors ligne en différé peuvent-elles enrichir notre connaissance de la perception des pauses pleines et des allongements ? L'analyse comparative permet de révéler **quels sont les corrélats acoustiques d'un phénomène relié à l'hésitation pour chaque campagne d'annotation** [46].

2.2.1 Objectif et hypothèses

Nous présentons une comparaison des trois ensembles de données :

1. La première campagne d'annotation est une méthode **hors-ligne, en écoute segmentée**, du corpus par des annotateurs **ayant accès à des informations supplémentaires**, la représentation visuelle du signal, et la possibilité d'**écouter plusieurs fois des segments**. La plupart des annotations manuelles dans les corpus de parole sont produites en utilisant cette méthodologie.
2. La deuxième campagne s'appuie sur un logiciel permettant l'**annotation en ligne** d'un phénomène en temps réel, sans possibilité d'arrêter la lecture ou d'écouter de façon répétée des segments.
3. La troisième campagne est le résultat d'une **annotation automatique**, basée sur des modèles statistiques formés sur un corpus oral différent. Cette annotation a ensuite été vérifiée manuellement.

L'analyse comparative des annotations informe des corrélats syllabiques qui ont déclenché l'annotation et/ou la perception d'une pause pleine et d'un allongement [47]. Ces corrélats comprennent : la **durée de la syllabe**, le **contour intonatif**, la **présence** et la **durée d'une pause silencieuse avant ou après l'allongement**, le **débit de parole local**, la *position* de la syllabe dans un **item lexical** et au sein d'un **constituant syntaxique**.

Cette triangulation, c'est-à-dire « la contre-vérification des données provenant de sources multiples pour rechercher des régularités dans les données de recherche » (O'Donoghue & Punch 2003:78) [48] tente de répondre à ces questions :

1. Existe-t-il une différence significative de distribution des durées entre les pauses pleines détectées selon les types de campagne d'annotation ?

46. Une partie de ces résultats, respectant la même méthodologie, a été présentée dans le cadre de la collaboration E. Duvivier, G. Christodoulides & I. Grosman. 2017. « Comparing Online and Offline Methodologies for Spoken Corpus Annotation. The case of Filled Pauses and Hesitation-Related Lengthening in French ». EN. Heraklion, Crete, Greece.

47. Dans les trois cas, les annotateurs peuvent être considérés comme « experts » i.e. Détenteurs d'une formation en linguistique et d'expérience en phonétique, prosodie et annotation de corpus.

48. "[...] (a) method of cross-checking data from multiple sources to search for regularities in the research data." Notre traduction.

2. Est-ce que le mouvement intonatif intrasyllabique et son étendue [49] influencent la probabilité d'être détectés/annotés ?
3. Les pauses pleines et les allongements ont-ils une probabilité de détection différente en fonction de leurs cooccurrences (e.g. précédées ou suivies de pauses silencieuses) ? Si oui, la durée de cette pause influence-t-elle la détection ?

2.2.2 Données et méthode

Pour chaque campagne d'annotation, nous décrivons brièvement les protocoles d'annotation des pauses pleines [50] et des allongements [51]. Toutes les annotations ont été effectuées sur le corpus LOCAS-F (Martin, Degand & Simon 2014).

– **Campagnes d'annotation hors ligne en écoute segmentée – *offline*.**
Le corpus entier a été annoté en écoute segmentée selon **deux méthodologies indépendantes**. La première, applique manuellement le schéma d'annotation de Crible, Dumont, Grosman & Notarrigo (2016). Elle correspond aux données présentes dans cette thèse. La seconde, est automatique et vérifiée manuellement, elle utilise les modèles statistiques de l'annotateur en partie du discours et en (dis)fluences DisMo (Christodoulides, Avanzi & Goldman 2014).

Bien que les deux campagnes utilisent différents schémas d'annotation, toutes deux partageaient une **même définition de la pause pleine et de l'allongement** dit d'hésitation. Durant les deux campagnes d'annotation, aucune limite minimum ou maximum d'écoute n'a été fixée pour l'annotation des marqueurs.

Pour rappel, l'annotation *offline* est dépendante de (et influencée par) l'alignement et la transcription du corpus. Les conventions de transcription utilisées dans le corpus identifient facilement les pauses autonomes et les syllabes se terminant par un schwa post-tonique (éventuellement perçus comme des « e » épenthétiques).

– **Campagne d'annotation en ligne en continu – *online*.**
Les résultats de la campagne d'annotation *online* consistent en la comparaison de deux sessions d'annotation de LOCAS-F [52]. L'annotatrice avait pour instruction d'appuyer sur la barre d'espace du clavier dès qu'elle entendait une **pause pleine et/ou un phénomène d'allongement dit d'hésitation**.

49. i.e. La différence entre la fréquence fondamentale en début et en fin de syllabe.
50. i.e. Les *euh* et leurs variantes phonétiques (e.g. *euhm, hm*), en excluant les « *mm* » entraînés par des backchannels. L'annotation inclut également les schwa post-toniques.
51. Allongements vocaliques perçus comme des hésitations, à l'intérieur ou en fin de mots.
52. Plus précisément, l'annotatrice a réalisé 6 mini-sessions contiguës ; processus qu'elle a répété à deux reprises.

Méthodologiquement, en utilisant le temps de réaction moyen de l'annotateur comme ligne de base [53], chaque point de touche appuyé a été associé automatiquement à un instant t et à un item lexical en fonction du temps de réaction correspondant à la session [54].

L'annotatrice n'avait pas de **représentation écrite ou de visualisation** du signal sonore, afin d'approcher le plus possible de la perception naturelle du phénomène. La campagne d'annotation a employé le logiciel « Tapping Annotator » (Figure 5.26) [55]. L'interface permet à l'utilisateur d'écouter des échantillons audio de corpus et d'appuyer sur une touche dès qu'il perçoit le phénomène étudié. Toutes les frappes (pressions et relâchements) sont enregistrées pour l'analyse [56].

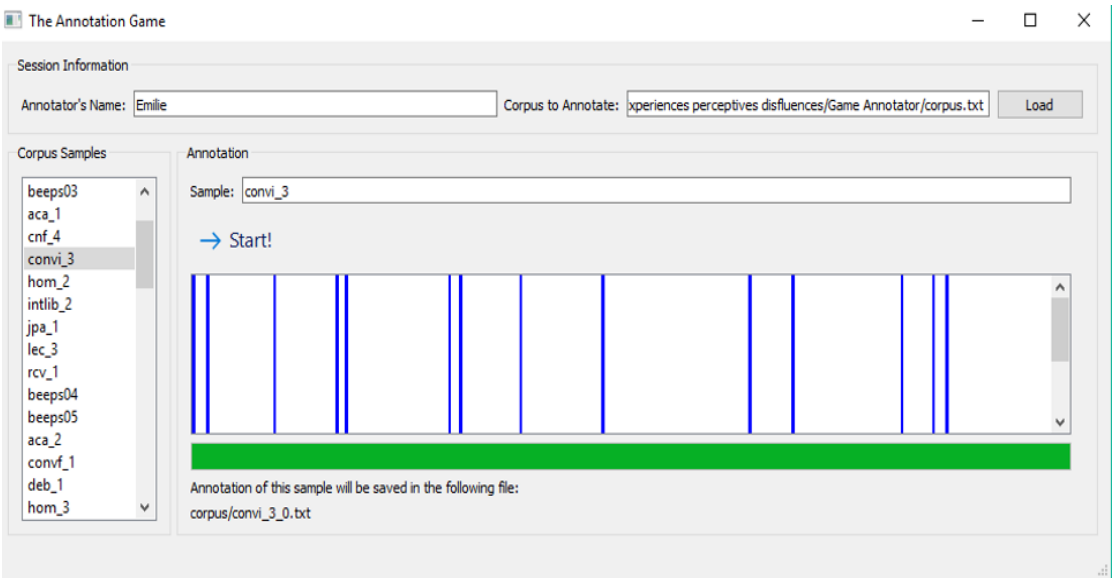


FIGURE 5.26 – Fenêtre de dialogue du programme « Tapping Annotator ». *L'utilisateur sélectionne un échantillon de corpus (à gauche) et lance la lecture. Une barre de progression (en bas, en vert) indique le temps de lecture restant. Chaque touche enregistrée est également affichée en tant que ligne bleue verticale (au milieu). Les résultats sont sauvergadés dans un fichier XML pour une analyse plus approfondie.*

En **croisant ces deux sessions d'annotations**, 1091 phénomènes d'hésitation ont été retenus. Chaque élément annoté a été relié à une des 4 étiquettes possibles :

– *pause pleine* : la touche est associée à la présence d'une pause pleine dans la transcription.

53. Le temps de réaction de l'annotatrice a été calculé premièrement de manière globale (temps de réaction moyen) et au début de chaque session. Il a été effectué via l'écoute d'une série de sons purs auxquels elle devait réagir le plus rapidement en appuyant sur la barre d'espace du clavier.

54. Le temps de réaction constitue donc ici la différence entre un phénomène perçu et la pression de la barre d'espace.

55. Précédemment utilisé pour étudier la perception des frontières prosodiques dans l'étude A. C. SIMON & G. CHRISTODOULIDES. 2016. « Frontières prosodiques perçues : corrélats acoustiques et indices syntaxiques ». *Langue française* (3), 83–106.

56. Le volume sonore au casque était constant, toute autre application de l'ordinateur était fermée afin d'éviter tout ralentissement du processeur.

- *allongement* : la touche est associée à la présence d'un allongement dans un environnement proche de la touche.
- *pause pleine + allongement* : le phénomène pourrait être associé à la présence d'une pause pleine et d'un allongement.
- *non expliqué* : la touche n'est pas clairement associée à une pause pleine ou un allongement (e.g. présence d'une pause subjective, erreur de frappe).

L'annotatrice a obtenu un score intra-annotateur fort de $K = 0.64$ entre ses deux campagnes d'annotation. Par facteur, l'accord le plus important a été observé dans l'ordre décroissant pour les pauses pleines ($K = 0.70, p < 0.001$), les phénomènes à étiquette double de pause pleine + allongement ($K = 0.67, p < 0.001$), les tokens non annotés ($K = 0.64, p < 0.001$), les allongements ($K = 0.33, p < 0.001$) et les phénomènes non expliqués ($K = 0.04, p < 0.001$). Ces résultats sont donc relativement faibles pour les allongements, nous pouvons émettre l'hypothèse d'une moindre sensibilité aux allongements d'hésitation en écoute continue.

Le Tableau 5.10 reprend le nombre de phénomènes annotés en fonction des campagnes d'annotation. Le corpus entier analysé regroupe ainsi les données des deux campagnes *offline* et de la campagne *online*.

Phénomène associé	**online**	**offline-Dismo**	**offline-ARC**
Pause pleine (FP)	823	1167	1299
Allongement (LG)	113	285	700

Tableau 5.10 – Fréquence des phénomènes d'hésitation dans LOCAS-F en fonction des campagnes d'annotation.

Parmi ces résultats, nous distinguons les pauses pleines et les allongements qui n'ont été détectés que dans le cadre de la campagne d'annotation *offline* seulement, de ceux détectés dans la campagne *offline* et *online* (Tableau 5.11).

Phénomène	*offline* seulement	*online* et *offline*	**Total**
Pause pleine	474	612	1086
Allongement	247	32	279

Tableau 5.11 – Résultat du croisement des campagnes d'annotation.

2.2.3 Analyses et résultats

Nous analysons la **probabilité de détection** d'une pause pleine en fonction de sa durée. La campagne d'annotation *offline* permet, de manière significative, de détecter **des pauses pleines plus courtes**, qui ne sont donc pas détectés en campagne d'annotation *online* ($\chi^2 = 157.1, df = 1, p < 0.001$) (Figure 5.27). Un test de Wilcoxon compare les deux échantillons indépendants : il révèle une différence des durées des syllabes détectés comme pauses pleines en fonction des campagnes d'annotation ($W = 80165.5$, p < 0.001, $95\%, C.I[-0.12, -0.09]$).

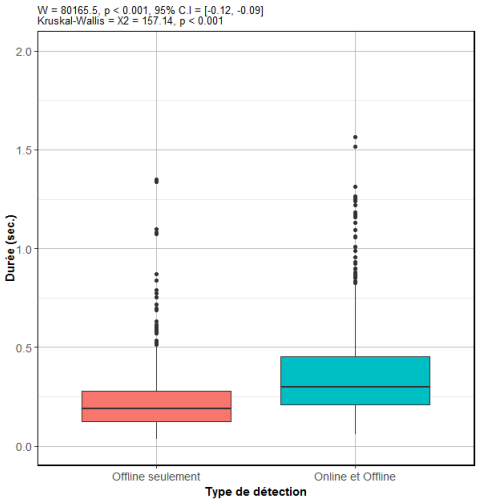


FIGURE 5.27 – Diagramme en boîte à moustaches des pauses pleines par type de détection.

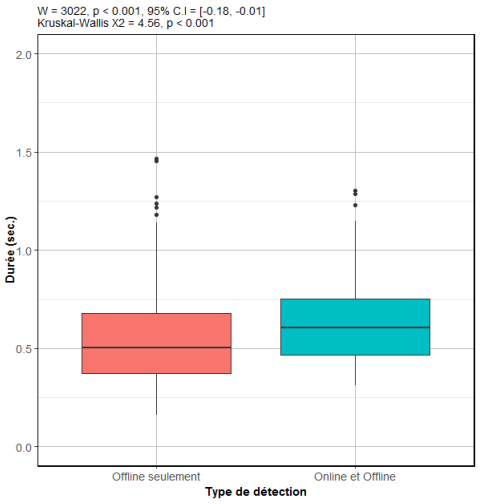


FIGURE 5.28 – Diagramme en boîte à moustaches des allongements par type de détection.

La campagne d'annotation *offline* permet, de manière significative, de **détecter des allongements plus courts**, qui ne sont donc pas détectées en campagne d'annotation *online* comme des allongements d'hésitation ($\chi^2 = 157.1, df = 1, p < 0.001$) (Figure 5.28). Un test de Wilcoxon compare les deux échantillons indépendants : il révèle une différence des durées des allongements dits d'hésitation détectés en fonction des campagnes d'annotation ($W = 30225$, p < 0.001, $95\%, C.I[-018, -0.01]$). Il semblerait ainsi que les campagnes d'annotation *online* ne détectent pas autant d'allongements d'hésitation et que ceux-ci ont tendance à être plus longs.

Nous voyons aussi que la durée des pauses pleines a une **plus grande variabilité de durées** en campagne *online* et *offline* que *offline* seulement (Figure 5.29). Cependant, nous ne disposons pas d'assez de données pour émettre les mêmes observations que pour les allongements (Figure 5.30).

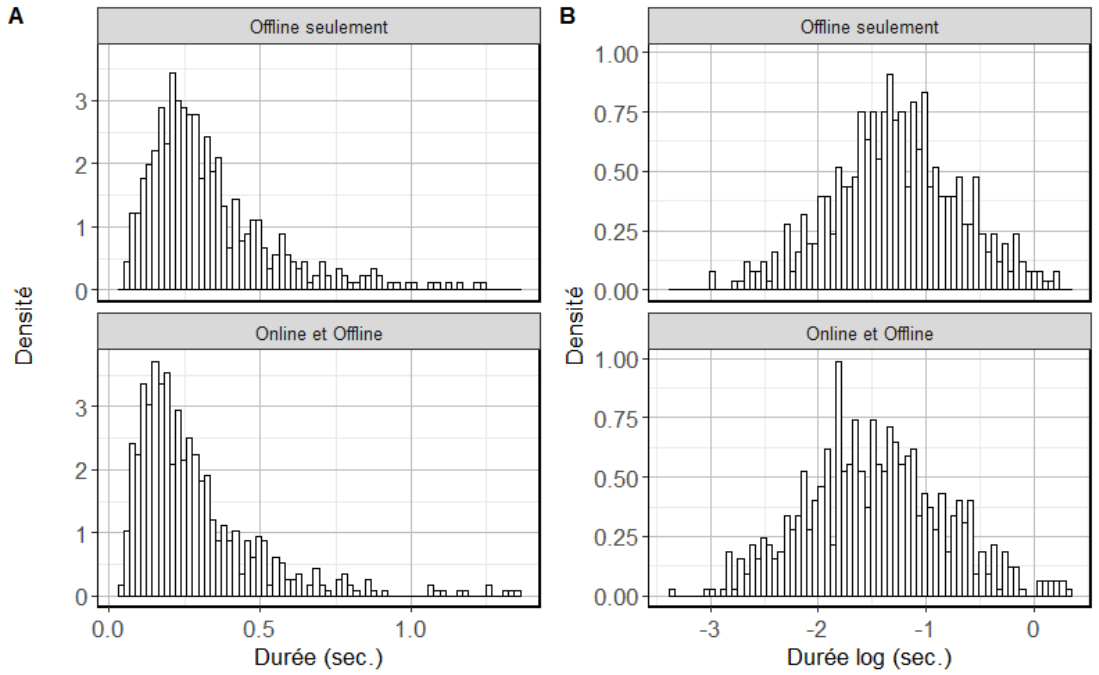


FIGURE 5.29 – Distribution des pauses pleines par type de détection en durée absolue (*panel A*) et log10 (ms.) (*panel B*).

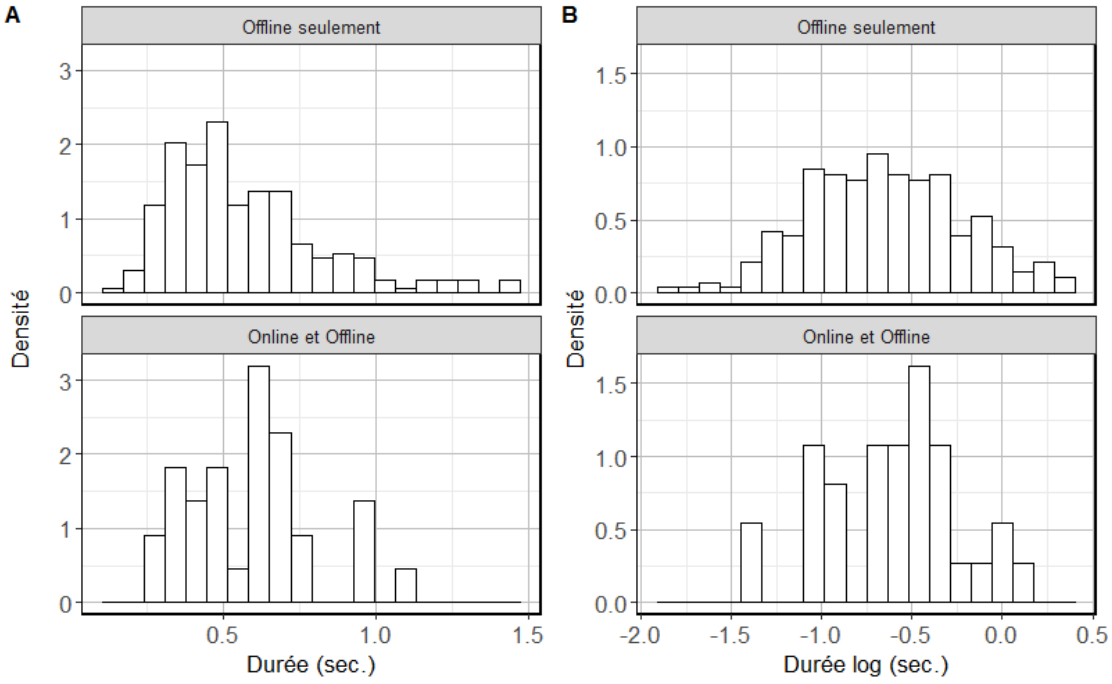


FIGURE 5.30 – Distribution des allongements par type de détection en durée absolue (*panel A*) et log10 (ms.) (*panel B*).

En analysant le **mouvement intonatif interne** des pauses pleines et des allongements, nous remarquons que les marques annotées dans la campagne *online* ont une plus grande probabilité de chute intonative que celles en campagne *offline* et *online*, que ce soit en durée absolue ou relative (respectivement Figures 5.31, 5.32 et Figures 5.33, 5.33). Pour les deux marques, le test de Wilcoxon prouve une différence significative à $p < 0,003$: **les mouvements intonatifs des allongements sont plus étendus que ceux des pauses pleines**.

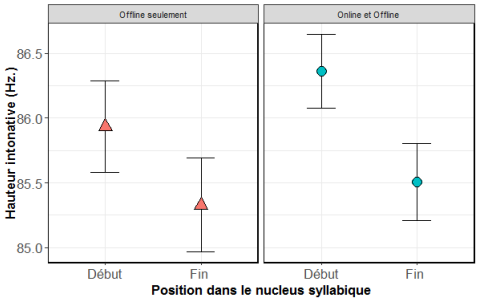


FIGURE 5.31 – Diagramme en boîte à moustaches du mouvement intrasyllabique des pauses pleines par type de détection.

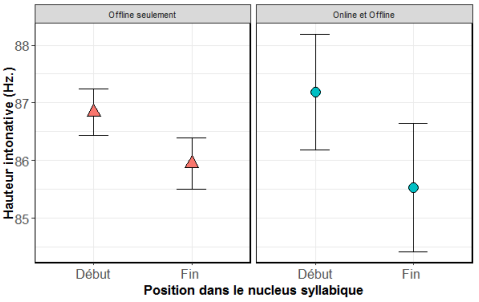


FIGURE 5.32 – Diagramme en boîte à moustaches du mouvement intrasyllabique des allongements par type de détection.

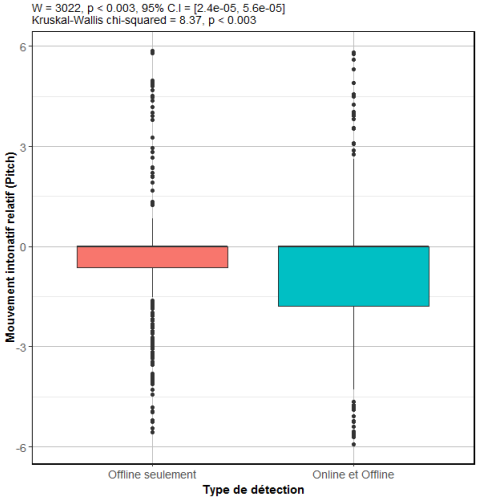


FIGURE 5.33 – Diagramme en boîte à moustaches du mouvement intrasyllabique des pauses pleines par type de détection.

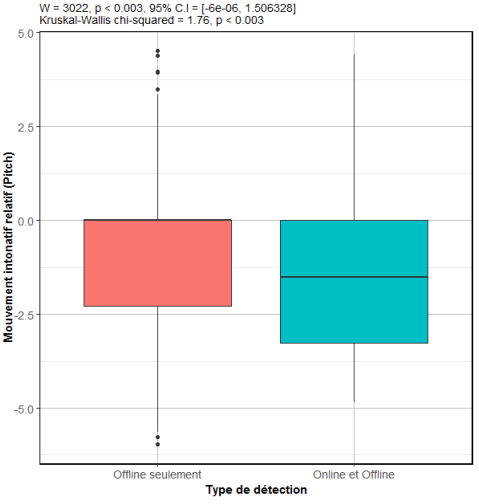


FIGURE 5.34 – Diagramme en boîte à moustaches du mouvement intrasyllabique des allongements par type de détection.

Ce type d'étude permet également d'analyser les différences de détection non seulement en fonction des durées mais **du contexte d'actualisation** de la pause pleine et de l'allongement. A la Figure 5.35, les allongements [all. (auto)] représentent la comparaison des allongements annotés automatiquement ; les allongements [all. (manu)] ressortent de la comparaison avec ceux annotés selon le protocole ARC.

Les panneaux supérieurs représentent la durée des pauses pleines et des allongements lorsqu'ils sont **précédés d'une pause silencieuse** et le panneau inférieur lorsqu'ils sont **suivis d'une pause silencieuse**. Nous avons répété cette méthode pour les pauses pleines et les allongements lors des deux campagnes d'annotation hors ligne. Par conséquent, nous pouvons voir en haut à gauche qu'**une pause pleine est plus susceptible d'être détectée quand précédée d'une pause silencieuse** (principe de proéminence). En bas à gauche, une pause pleine est davantage susceptible d'être détectée lorsqu'elle est suivie d'une pause silencieuse (mais moins).

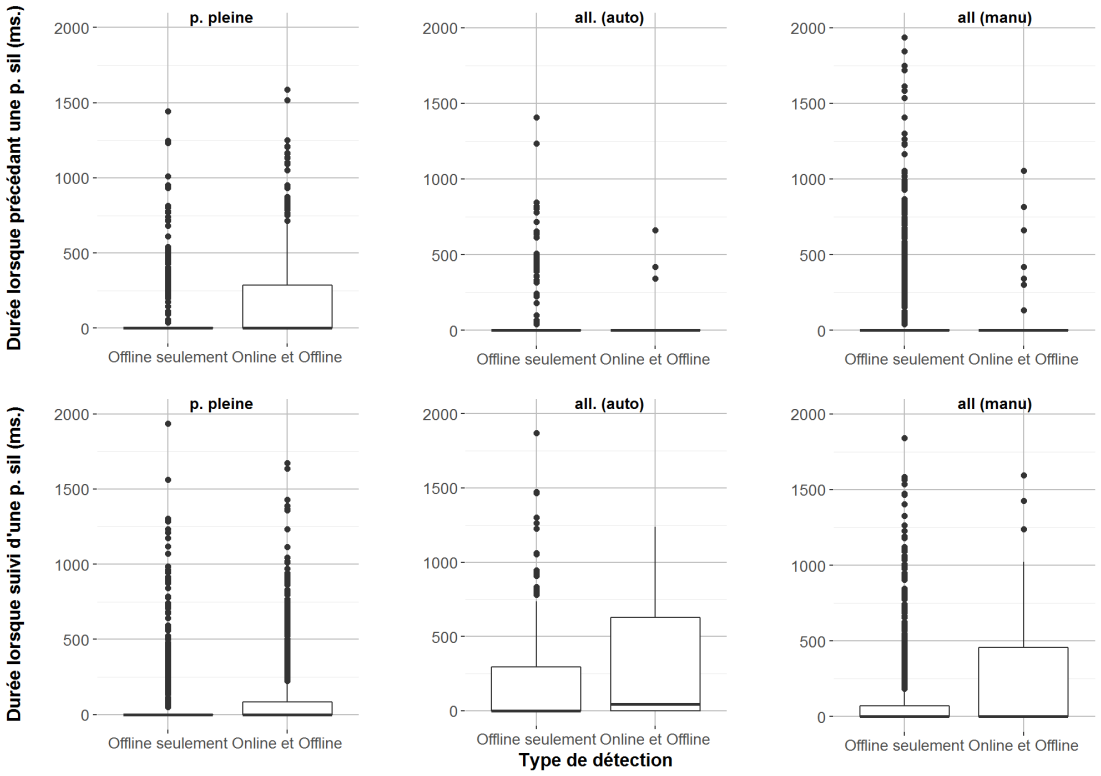


FIGURE 5.35 – Diagramme en boîte à moustaches de la durée des marqueurs détectés en fonction des pauses précédant/succédant et de la méthodologie de détection.

Les panneaux d'allongement en haut à droite nous permettent de voir que **la durée des pauses précédentes n'influence pas la détection d'un allongement**. Au contraire, les panneaux du bas montrent qu'un **allongement est plus susceptible d'être détecté lorsqu'il est suivi d'une pause**, que lorsqu'il est précédé d'une pause, surtout dans la campagne d'annotation en ligne. Nous discutons de ces résultats au regard de l'expérience précédente [57] en conclusion de ce chapitre.

57. Voir section 2.1 de ce chapitre.

2.2.4 Discussion sur la comparaison des méthodes

Cette section a testé l'impact de l'utilisation de différentes méthodologies d'annotation sur la détection et l'évaluation des pauses pleines et des allongements.

Une première observation est que les pauses pleines et les allongements ont été détectés **en dessous du seuil des 100 ms.** ; ce qui est inférieur aux seuils de détection de la littérature. Nous avons aussi montré que de manière générale les pauses pleines et les allongements détectés par l'annotation en ligne ont une étendue intrasyllabique supérieure, et que les pauses pleines et les allongements suivis d'une pause silencieuse sont plus remarqués en campagne d'annotation en ligne. Enfin, il n'y a pas de différence entre les campagnes *online* et *offline* lorsque l'allongement est précédé d'une pause silencieuse.

Soulignons que chaque méthode d'annotation a ses avantages et ses inconvénients. En linguistique de corpus, les transcriptions suivent des directives strictes (e.g. *euh* /vs/ schwa) qui pourraient biaiser les résultats. Dans l'annotation hors ligne typique, il est possible d'écouter le segment pour annoter plusieurs fois. Nous avons observé une tendance à **« surannoter » les phénomènes** : la présence d'une visualisation et l'accès à des informations supplémentaires semblerait induire un biais d'annotation. C'est notamment pourquoi l'annotation en temps réel est plus proche des **contraintes de perception de la parole**. Cependant, dans l'annotation en ligne proposée ici, l'**attention des annotateurs** est axée sur un phénomène en particulier, ce qui les y sensibilise considérablement. Lors du traitement de la parole en contexte, certaines de ces (dis)fluences pourraient échapper. Le croisement des corpus hors ligne et les annotations expérimentales en ligne peuvent aider à mieux en comprendre les raisons.

3 Conclusions et discussion : la rupture et la continuité des pauses pleines et des allongements

Dans ce chapitre, nous avons analysé les distributions et les formes acoustiques des pauses pleines et des allongements en fonction de leurs durées et de leurs cooccurrences. Sur cette base, nous avons comparé les similarités et les complémentarités distributionnelles (acoustiques, prosodiques, syntaxiques et interactionnelles) des deux marques, de manière isolée et au sein de cooccurrences. Enfin, nous avons évalué leurs caractéristiques dans des tâches de perception *online* et *offline* en les comparant aux résultats des corpus de production. Nous nous sommes notamment posé les questions suivantes :

- Ces marques et leurs cooccurrences relèveraient-elles davantage d'une mise en attente discursive pour les allongements et d'un retardement à venir pour les pauses pleines ?
- Ces marques et leurs cooccurrences sont-elles positionnées à des endroits de rupture ou de cohésion discursive ?
- Certaines actualisations constituent-elles un écart envers une ou des norme(s) et une ou des attente(s) phrastique(s) ou situationnelle(s) ?
- Quels intérêts spécifiques les différentes méthodologies d'annotation ont-elles au regard de l'évaluation de la fluence ?

Notre conclusion principale soutient que, bien que les distributions des durées de pauses pleines et d'allongements soient toutes les deux lognormales unimodales et asymétriques, l'observation des caractéristiques distributionnelles **supporte l'hypothèse de deux ensembles de données différents**. Nous résumons ci-après les observations supportant cette conclusion.

Analyse distributionnelle en clusters

- *Sur les corpus entiers.* Les durées moyennes et les médianes des allongements sont supérieures à celles des pauses pleines. Les allongements présentent une distribution de leur durée plus étroite que celle des pauses pleines (et des pauses silencieuses). Si la durée des allongements est soumise à moins de variation, leur profil intonatif – les mouvements intrasyllabiques – est relativement plus marqué.
- *Par situation et par locuteur.* Nos analyses révèlent des durées de pauses pleines et des allongements différentes en fonction des situations. Toutefois, cette différence de durée reste minime et présente de faibles variations intrasituationnelles.
- Ainsi, le caractère unimodal de la distribution de durée et la faible variation intrasituationnelle des pauses pleines et des allongements plaident en faveur d'une modélisation unique et uniforme de leur durée pour chacune des marques. Au contraire des pauses silencieuses, rien n'oblige de distinguer des pauses pleines « courtes » /vs/ « longues » ou des allongements « courts » /vs/ « longs », quels que soit la situation ou le locuteur. Enfin, rien ne permet de regrouper deux marques sous le terme de pause « sonore » étant donné la différence importante de durée les distinguant.

Cooccurrences avec d'autres marques de (dis)fluences. Un quart des pauses pleines et des allongements apparaît dans des contextes non cooccurrents à d'autres (dis)fluences. Un point commun émerge : en combinant les différents types de cooccurrences, les pauses pleines comme les allongements sont potentiellement moins structurants que les pauses silencieuses.

– La moitié des pauses pleines s'associe avec des pauses silencieuses (antéposées ou postposées). Par ailleurs, la pause pleine se marie plus avec les marqueurs discursifs que l'allongement. On relève ainsi davantage de chaînes (md + pause pleine) et (pause pleine + md) que de chaînes (md + allongement) ou (allongement + md). Dans le même ordre d'idée, les pauses pleines sont plus souvent suivies de (dis)fluences que ne le sont les allongements, eux majoritairement précédés de (dis)fluences. Ces observations rejoignent celles de GUAÏTELLA (1991b citée par CANDEA 2000:138). On retrouve une proportion importante dans nos corpus de (pause pleine + pause silencieuse) et (pause pleine + md) d'une part, et de (pause silencieuse + allongement) et (pause pleine + allongement) d'autre part.

Durées et cooccurrences. Nous avions formulé l'hypothèse selon laquelle les durées des pauses pleines et des allongements augmentent lorsqu'elles sont en cooccurrence avec d'autres (dis)fluences plutôt qu'isolées. Cette différence de durée est globalement confirmée :

– Dans les **configurations initiées par une pause pleine**, la **durée de la pause pleine est inférieure** aux configurations où celle-ci succède à une autre marque (e.g. pause silencieuse + pause pleine, md + pause pleine, dis + pause pleine). Ainsi, à l'extrême du continumm, on oppose les chaînes (md + pause pleine), (pause silencieuse + pause pleine) et (> 2 éléments), significativement plus longues que les chaînes (pause pleine + *interregnum*), (pause pleine + dis) et (pause pleine isolée).
– Au niveau des allongements, seules deux chaînes se distinguent. Les chaînes (allongement + md) et (dis + allongement) exhibent des durées plus importantes d'allongements que les 8 chaînes de cooccurrences additionnelles décrites dans ce chapitre.

Interactions entre les durées et les situations. Il existe une triple association entre situations de parole, durées des marques et chaînes de cooccurrences pour les deux marques (et davantage encore pour les pauses pleines).

– Alors que les pauses pleines sont relativement plus longues en discours spontané que préparé, les allongements sont eux plus longs en discours préparé que spontané.
– On remarque ainsi que les cooccurrences (pause silencieuse + allongement) et (pause silencieuse + pause pleine) sont typiques des discours spontanés, tandis que les cooccurrences (allongement + pause silencieuse) et (pause pleine + pause silencieuse) sont plus fréquentes en situations préparées.

Interaction de la fréquence et dela durée en fonction des frontières syntaxiques. Nous relevons une différence des distributions des durées des pauses pleines et des allongements ainsi que de leurs cooccurrences relativement au degré de frontières syntaxiques (i.e. intra-ga, inter-ga, inter-séquence, inter-rection) :

- Près d'un tiers des pauses pleines s'actualise en position inter-rectionnelle et en position inter-ga, suivi d'un quart des pauses pleines en intra-ga et une faible minorité en position inter-séquences. Nous confirmons les observations de CANDEA (2000a:262) selon lesquelles la moitié des *euh* sont observés « **en dehors d'un bloc syntaxique** » (entre deux pauses silencieuses, entre deux constituants juxtaposés ou après un connecteur). Le passage d'une séquence fonctionnelle à une autre n'est pas propice à l'actualisation des pauses pleines.
- Les pauses pleines, qui font le lien entre « ce qui a été dit » et « ce qui va être dit », s'actualisent en frontière d'unités syntaxiques qui appellent à être coordonnées et liées. Les *euh* sont alors des **ligateurs discursifs**, témoins du processus d'encodage. Ils indiquent l'inachèvement mais pas nécessairement une rupture, ou une *dis*fluence. Ils marquent une **intention de continuation** plus que de clôture.
- En contraste, les allongements sont davantage des actualisations microsyntaxiques, en grande majorité au sein de groupes syntaxiques minimaux en médiane d'unité plus qu'en frontière (plus de la moitié en intra-ga et près d'un quart en inter-ga).
- Par ailleurs, cette position de la pause pleine en inter-rectionnelle est davantage exploitée en spontané, tandis que les positions inter-ga ou inter-séquentielles sont observées en situations de parole semi-préparées et spontanées. Et cette position des allongements intra-ga est renforcée en discours préparé plutôt qu'en discours spontané.
- Les allongements opèrent une distinction énonciative avec les *euh*. La fonction modale et prospective des allongements s'incarne dans le fait qu'ils s'attachent à **la gestion immédiate** du discours (intra-ga, inter-ga), inhérente à un **travail de planification** et de conceptualisation (e.g. ré-évaluation du référent) ou de **maintien de la parole**[58].

Interactions de la durée et des cooccurrences au regard du degré des frontières syntaxiques.

- Pour rappel, CANDEA (2000a:230) n'a pas pu mettre en évidence de différence de durée moyenne entre les pauses pleines en position **initiale** /vs/ **médiane de constituant**. Cependant, à l'instar de GUAÏTELLA, l'auteur constate bien une différence significative entre la durée des pauses pleines en position médiane de constituant et la durée de celles produites entre deux pauses silencieuses (isolées des constituants syntaxiques), les secondes étant plus longues que les premières. Dans nos corpus, les annotations syntaxiques permettent de spécifier les éléments suivants :
  - **Isolées**, on oppose deux groupes significativement différents de durée des pauses pleines : (inter-ga et intra-ga) /vs/ (inter-séquence et inter-rection), ces dernières étant plus longues. La durée des allongements, elle, qui variait bien en fonction des cooccurrences, ne varie pas en fonction des frontières syntaxiques.
  - **En cooccurrence**, on observe que les pauses pleines insérées (> 2 éléments) sont effectivement plus longues, et que certaines chaînes ont des profils plus homogènes que d'autres (e.g. allongement + pause silencieuse, allongement seul, pause pleine + md, pause silencieuse + pause pleine). Pour les allongements, nous soulignons le retardement notable provoqué par les durées importantes des chaînes (allongement+dis)

58. Ces éléments ont été abordés plus en détails au chapitre 2 section 2.3.1.

et (dis + allongement), surtout en intra-ga. Cette observation confirme notre hypothèse que les allongements en cooccurrence (particulièrement avec des (dis)fluences composées) sont **plus longs au sein de groupes microsyntaxiques**.

Interaction de la durée et de la position au sein des tours de parole. Les durées supérieures des pauses pleines et des allongements ne sont pas motivées interactionnellement par la passation de tour :

- La pause pleine est majoritairement intra-tour et globalement plus en début de tour, initiatrice de tour de parole, qu'en fin de tour. Les chaînes > à 2 éléments ne marquent pas plus de changements de tour de parole que les autres chaînes. En outre, les (dis)fluences porteuses d'une phase d'édition *(interregnum* simple) ne sont pas synonymes de transition de tour. Au contraire, elles permettent la continuation du tour de parole.
- Au niveau des pauses pleines, les chaînes (md + pause pleine) et (pause pleine + md) sont, elles, relativement présentes en frontière de tour de parole, ce que nous associons donc à des lieux propices de transition, surtout au regard de leur positionnement syntaxique. Les chaînes (dis + pause pleine) sont par ailleurs davantage synonymes de transition que les chaînes (pause pleine + dis), intra-tours.
- Au niveau des allongements, la position médiane de tour est majoritairement observée. L'allongement seul, les chaînes (> à 2 éléments) et les chaînes (pause pleine + allongement) sont actualisés en fin de tour, mais de manière significativement marginale. A l'instar des pauses pleines, les chaînes (md + allongement) et (dis + allongement) sont davantage observées en frontière de tour[59].

Variation des mouvements mélodiques en fonction des durées.

- Concernant l'interaction des **formes de contours prosodiques**, nous ne postulions pas d'une distribution significativement différente des contours de pauses en fonction des groupes intonatifs au sein desquels elles s'insèrent (Duez 2001b:39). Dans nos données, la **durée** de la pause pleine et des allongements n'est en effet pas associée au contour de la syllabe les précédant. L'hypothèse de fonction continuative des deux marques est appuyée par le fait que les pauses pleines comme les allongements semblent précédés par des contours continuatifs.
- Les données permettent de distinguer les marques en fonction de leur dynamisme intrasyllabique : l'étendue du mouvement intonatif est de plus grande ampleur pour les allongements que pour les pauses pleines. Les allongements se distinguent car ils présentent des mouvements intrasyllabiques plus dynamiques que les pauses pleines. Ils accueillent aussi davantage de mouvements montants contre moins d'un tiers de descendants.
- De plus, nous confirmons une association du type de mouvement intrasyllabique (i.e F0 montante, plate, descendante) en fonction des cooccurrences pour les deux marques. La majorité des pauses pleines et des allongements ont un mouvement intrasyllabique plat et stable, à l'encontre des observations de Duez (2001a) et O'Shaughnessy (1992).

59. Ces éléments ont été abordés plus en détails au chapitre 2 section 2.3.1.

– Par ailleurs, certaines cooccurrences sont caractérisées par des schémas intonatifs montants (e.g. pause silencieuse + pause pleine) ou descendants (e.g. pause silencieuse + allongement) et rejettent les mouvements intonatifs stables. Au contraire la pause pleine précédant une pause silencieuse exhibe des contours majoritairement plats. Enfin, la chaîne (pause pleine + allongement) est particulièrement propice à la réception d'un mouvement intrasyllabique descendant [60].

Critères pertinents à la détection d'une hésitation dans une tâche d'annotation *offline*. Nous avons montré que :

– L'accord interannotateurs dans une **tâche d'annotation *offline*** est robuste et substantiel pour tous les types d'hésitation, particulièrement pour les *euh* et les allongements dits d'hésitation. On observe **peu de biais intra-annotateurs et interannotateurs** dans l'expérience.
– Nous avons modélisé les profils (acoustiques, prosodiques, syntaxiques) des allongements pour lesquels le taux d'accord interannotateurs s'avère élevé dans les corpus. Au niveau méthodologique, cette détection et cette modélisation des « meilleurs exemples d'allongements d'hésitation » permettent d'obtenir une approche plus compréhensive du phénomène d'allongement.
– Les **variables les plus informatives** relatives à la modélisation de l'accord interannotateurs pour les allongements sont : la durée relative (ce qui confirme les observations de Shriberg, Stolcke & Baron (2001:161)), la cooccurrence avec d'autres (dis)fluences, les pauses précédant ainsi que la position dans le ga (et l'information lexicale /vs/ grammaticale), le mouvement intonatif et la hauteur de F0. Cette dernière considération soutient également que malgré la faible amplitude du mouvement intonatif au sein des allongements d'hésitation, ces paramètres sont clés dans leur détection automatique et leur perception/annotation manuelle humaine.
– Une perspective de cette étude est d'effectuer sur un plus grand nombre de données une **modélisation ordinale** [61] prenant en compte la variation situationnelle, les paramètres prosodiques et surtout, nous l'avons vu lors de l'analyse de corpus, les unités de rection, afin de : (1) distinguer **les allongements dits d'hésitation /vs/ les allongements grammaticaux**, (2) distinguer les segments, les syllabes et les types de cooccurrences qui, dans un même contexte, sont de durées relatives allongées similaires, mais évalués comme des hésitations ou comme des allongements « grammaticaux », « structurants ».

Critères pertinents à la détection et les caractéristiques acoustiques d'une hésitation dans une tâche d'annotation *online* Nous avons développé notre perspective sur la variation des courbes intonatives des pauses pleines et des allongements

60. Bien que nous n'ayons pas croisé ces données ici, la modélisation finale prend en compte l'interaction potentielle entre les formes acoustiques et les localisations syntaxiques. Mais, pour rappel, en introduction, on ne postulait pas une différence des courbes intonatives des pauses pleines en fonction de leur localisation syntaxique (O'Shaughnessy 1992).

61. Une modélisation ordinale considèrerait un allongement comme non disfluent /vs/ peu disfluent /vs/ très disfluent et non comme une catégorisation binaire fluent /vs/ disfluent.

en analysant les formes détectées par une expérience confrontant trois méthodologies différentes d'annotation des pauses pleines et des allongements : en ligne, différée, automatique.

- Des pauses pleines et des allongements ont été détectés **en-dessous du seuil des 100 ms.** en annotation *online*, bien que la littérature applique souvent des seuils avoisinant les 200 ms.
- Les pauses pleines et les allongements relevés par l'annotation en ligne ont une **étendue du mouvement intonatif intrasyllabique plus importante** que ceux détectés et annotés manuellement en *offline* ou automatiquement. Ces résultats sont en accord avec les données de corpus de production.
- Cependant, ce mouvement reste de faible amplitude pour les deux marques (souvent inférieur à 2.5 ST). Nous confirmons donc les conclusions de Guaïtella (1991:114) : les allongements d'hésitation sont un phénomène où sur une syllabe « se manifeste un allongement de durée, sans qu'il y ait modification du contour de F0 ». En effet, si l'allongement s'accompagnait d'un **décrochage de la F0**, il serait assimilé à un accent. Seul le **contour plat** lui permet d'être associé à un allongement vocalique d'hésitation.
- Enfin, les pauses pleines précédées ou suivies d'une pause silencieuse sont plus remarquées en campagne d'annotation *online* qu'*offline* manuelle et automatique. **L'allongement suivi d'une pause silencieuse** est également plus annoté en campagne *online*, tandis qu'il n'y a pas de différence entre les campagnes *online* et *offline* lorsque l'allongement est précédé d'une pause silencieuse. Ainsi, la présence d'une pause silencieuse après un allongement favorise la **perception d'une hésitation et/ou d'une frontière**.

Méthodologiquement, dans l'annotation *offline* typique, il est possible d'écouter le segment pour annoter plusieurs fois. Il y a une **tendance à surannoter les phénomènes**. La présence d'une visualisation et l'accès à des informations supplémentaires sur le signal sonore présentent un risque de biais d'annotation supplémentaire. L'annotation en temps réel est plus proche des contraintes de perception de la parole. Cependant, dans l'annotation en ligne proposée ici, l'attention des annotateurs est axée sur un phénomène en particulier, ce qui les y sensibilise considérablement. Lors du traitement de la parole, certaines de ces (dis)fluences pourraient échapper. **Le croisement des corpus hors ligne et des annotations expérimentales en ligne** peut aider à mieux en comprendre les raisons.

Notre conclusion principale soutient que, bien que les distributions des durées de pauses pleines et d'allongements soient **toutes les deux lognormales unimodales et asymétriques**, l'observation des caractéristiques distributionnelles supporte l'hypothèse de **deux ensembles de données différents** en vertu de leurs paramètres de durées sur l'ensemble du corpus, de leur différences de positionnement syntaxique, du dynamisme distinct de leur courbe mélodique, et des différentes chaînes de cooccurrence qui leur sont associées. Là où **la pause pleine initie un nouveau référent ou un nouvau**

tour, **l'allongement prend soin du maintien et de la cohésion discursive**. Les deux marques assurent des lieux propices de transition et ont effectivement pour point commun de ne pas laisser entrevoir le signe d'une rupture.

Chapitre 6

Evaluation de la (dis)fluence des répétitions à l'identique

Ce chapitre 6 aborde une (dis)fluence structurée spécifique (composée d'un *reparandum*, d'un *interregnum* et d'un *reparans*) : les **répétitions à l'identique**. Les répétitions sont une des (dis)fluences les plus fréquentes après les pauses pleines et silencieuses : elles représentent jusqu'à 25% des marques de certains discours[1]. On s'interroge ici sur la **manière d'analyser et d'évaluer** les répétitions à l'identique en production et en perception. Plus précisément, nous prenons en compte les acquis des chapitres précédents et nous nous attardons sur la **production et la perception des pauses silencieuses, des pauses pleines et des allongements au sein des répétitions**.

En production, nous présentons les caractéristiques distributionnelles des répétitions à l'identique observées dans les corpus. Nous considérons pour chaque échantillon la proportion de segments répétés par rapport à celle de segments non répétés et de segments (dis)fluents et leurs variations intersituationnelles. Nous observons ensuite les distributions de ces répétitions en chaînes de **cooccurrences**, en fonction de leur **localisation syntaxique et interactionnelle**, et de leur **forme acoustique** (section 1). En perception, deux expériences ont été réalisées afin d'estimer l'effet de ces structures ainsi que l'impact de la prise en compte du contexte de production sur l'évaluation de la (dis)fluence (section 2).

1. Voir notamment chapitre 3 section 3.3.

1 Analyse des données de corpus

Cette section présente les types de structures de répétitions et leurs cooccurrences[2]. Après avoir exposé les objectifs et les hypothèses de l'analyse en production (1.1), nous étudions les variations situationnelles des répétitions à travers les situations du corpus (1.2). Emergent ainsi des typologies de répétitions en fonction de leur structure interne et de leurs cooccurrences sur base des régularités du corpus (1.3). Nous évaluons alors l'interaction de ces structures avec les unités syntaxiques et les tours de parole (1.4). Enfin, dans le cadre de deux études distinctes, nous observons les corrélats de durées et d'intonation et nous comparons nos résultats à la littérature antérieure (1.5). En conclusion, nous croisons ces différents résultats et questionnons l'évaluation des répétitions en fonction des paramètres offerts par les corpus de production (1.6).

Ponctuellement, nous mentionnons dans cette section le corpus « Metacorpus ». Cette base de données regroupe 5 corpus alignés phonétiquement : LOCAS-F (Martin 2014), C-Humour (Grosman 2016), le *Driving Simulator Cognitive Load Corpus* (Christodoulides 2016), C-Phonogenre (Pršir, Goldman & Auchlin 2013) et le corpus Rhapsodie (Lacheret et al. 2014). Dans le cadre de deux études (Christodoulides, Grosman, Degand & Simon 2017 ; Grosman, Christodoulides, Degand & Simon 2017), les corpus ont été compilés en un, en supprimant les éventuels doublons d'échantillons, en homogénéisant leurs métadonnées et la définition des traits situationnels, et en effectuant une annotation automatique des (dis)fluences avec le plugin DisMo (Christodoulides, Avanzi & Goldman 2014). L'utilisation de ce regroupement de corpus est justifié par une taille d'échantillon considérable nécéssaire à l'analyse acoustique des répétitions.

1.1 Objectifs et hypothèses

Les répétitions extraites des corpus LOCAS-F et C-HUMOUR sont des répétitions de parole, monologiques et lexicales[3] :

Par répétitions de parole, nous entendons :

Toutes les répétitions lexicales de forme(s) à classe morphosyntaxique identique[4]. Ces répétitions à l'identique ont été marquées sans interprétation à priori de leur fonction communicative (phrastique, situationnelle, interpersonnelle) fluente ou disfluente. Cette définition est en accord avec Dister (2007:130). Elles s'opposent aux répétitions « faits de langue », où la répétition est fait du bon déroulement syntagmatique de l'énoncé (e.g. « nous nous sommes amusés »).

2. Voir chapitre 2 section 2.2 pour un rappel sur la méthodologie de description des (chaînes de) répétitions et (dis)fluences.

3. Pour un rappel théorique et méthodologique sur les répétitions à l'identique voir respectivement le chapitre 2 section 2.4 et le chapitre 3 section 2.1.

4. C'est-à-dire, en excluant les cas d'homonymie.

Par répétitions monologiques, nous entendons :

Toutes les répétitions qui sont le fait d'un même locuteur et qui sont intratour de parole : la prise de parole par un autre locuteur bloque l'annotation d'une répétition à l'identique. On oppose les répétitions monologiques intratours aux répétitions transtours (reprise à l'identique après une alternance de tour) et/ou dialogique (reprise des mots de l'interlocuteur).

Par répétitions lexicales, nous entendons :

Les répétitions sont annotées premièrement sur base de leur forme lexicale. Elles peuvent cependant être accompagnées d'une répétition prosodique impliquant une reprise de patron acoustique (durée, contour) entre le répétable et le répété [5], ce critère ne prime cependant pas dans notre annotation.

Un objectif sous-jacent de cette section est l'analyse des phénomènes de cooccurrence et de leur distribution pour permettre la formulation d'hypothèses sur leur perception et sur leur évaluation. Nous tentons donc de répondre aux questions suivantes :

1. Au niveau de **leur structure interne**, quels types de répétitions existe-t-il dans les corpus ?
2. Au niveau de **leurs cooccurrences**, les types de répétitions sont-ils associés à des marques différentes ? Ces cooccurrences varient-elles en fonction des situations de parole ?
3. Au **niveau discursif**, les répétitions s'insèrent-elles à des positions syntaxiques, interactionnelles et prosodiques précises ?
4. Au **niveau acoustique**, les répétitions montrent-elles des régularités prosodiques internes en fonction de leur structure ?

Les sections suivantes testent ainsi des hypothèses de variations situationnelles portant sur le contexte de production, des hypothèses de distribution de la structure interne, d'association avec des cooccurrences, de distributions syntaxiques et interactionnelles, et enfin de formes acoustiques. Chacune des hypothèses sera développée dans des sections spécifiques (sections 1.2 à 1.5).

5. Nous testons en section 1.5 l'association entre les structures de répétitions et les patrons prosodiques observés.

1.2 Fréquences et proportions des répétitions en fonction des situations de parole et traits situationnels

Sur 61834 mots simples[6], 1167 répétitions ont été annotées. Proportionnellement, on décompte **18.9 répétitions tous les 1000 mots**. Bien entendu, une répétition peut survenir en même temps que d'autres (dis)fluences et des répétitions peuvent s'actualiser au sein d'autres répétitions. En outre, ces 1167 chaînes de répétitions mettent en jeu un total de 4225 mots (répétables, répétés, phases d'édition, (dis)fluences associées et marques périphériques). En prenant en compte ces éléments, ce sont au total **68.9 mots tous les 1000 mots** qui sont concernés par une répétition.

Ces observations confirment que la répétition est la troisième marque la plus fréquente en parole, après les pauses silencieuses et les pauses pleines. Ces résultats sont comparables à ceux de Candea (2000a) et de Dister (2007:131). Pour cette dernière, la fréquence de chaînes de répétition(s) va de 10 à 102 répétitions tous les 1000 mots[7]. Du reste nous émettons les hypothèses suivantes :

Hypothèses sur le contexte de production (i.e. situationnelles)

- La proportion de répétitions diffère en fonction du contexte de production.
- A proportion égale de (dis)fluences, les situations préparées sont plus affectées par la répétition que les situations spontanées.
- De plus, les **caractéristiques des structures de répétitions**[8] sont associées à des caractéristiques situationnelles.

En **temps d'articulation**[9], les répétitions représentent un total de **5.5%** du corpus occupé par un locuteur en train de se répéter. Cette proportion moyenne fluctue de 0.53% à 13.9% selon les situations de parole. La Figure 6.1 illustre cette répartition des répétitions en fonction de leur durée cumulée par situation de parole.

Sur l'ensemble du corpus la proportion de répétitions semble à priori significativement différente en fonction des situations de parole ($F(16, 86) = = \ 0.4, p \ < \ 0.0052$), les détails de l'ANOVA révèlent une proportion moindre de temps d'articulation des répétitions pour **4 situations en particulier** : le discours académique, le dialogue conversationnel informel, la conférence scientifique et le sketch humoristique. Cependant, un test de Tukey DSH révèle que cette différence (ou taille d'effet) de proportion de production des répétitions **varie peu selon les situations**. En d'autres termes, lorsqu'on compare

6. Ce décompte de LOCAS-F et C-HUMOUR combinés exclut les pauses pleines et troncations qui ne sont pas considérées comme des *répétables*, car exempts de contenu sémantique (voir chapitre 3, section 2.1). La justification d'exclusion réside également dans le fait que les pauses pleines et les troncations apparaissent souvent en périphérie et au sein de répétitions, et que par conséquent ces éléments pourraient même en grande partie être inclus au sein d'une répétition.

7. Par ailleurs, ces derniers résultats sont davantage comparables aux nôtres étant donné qu'ils prennent en compte les répétitions à l'identique polylexicales, et surtout, les répétitions de parole, sans jugement à priori de leur fonction fluente ou disfluente.

8. e.g. éléments périphériques, présence d'un terme d'édition, durée et forme intonative intrinsèque, positionnement prosodico-syntaxique et interactionnel.

9. Pour rappel, le temps d'articulation représente la somme du temps de parole d'un locuteur duquel ont été soustraites les pauses silencieuses.

de manière pairée chaque proportion de répétitions par situation l'une à l'autre, les différences ne sont pas suffisantes pour distinguer la proportion de parole répétée d'une situation par rapport à une autre.

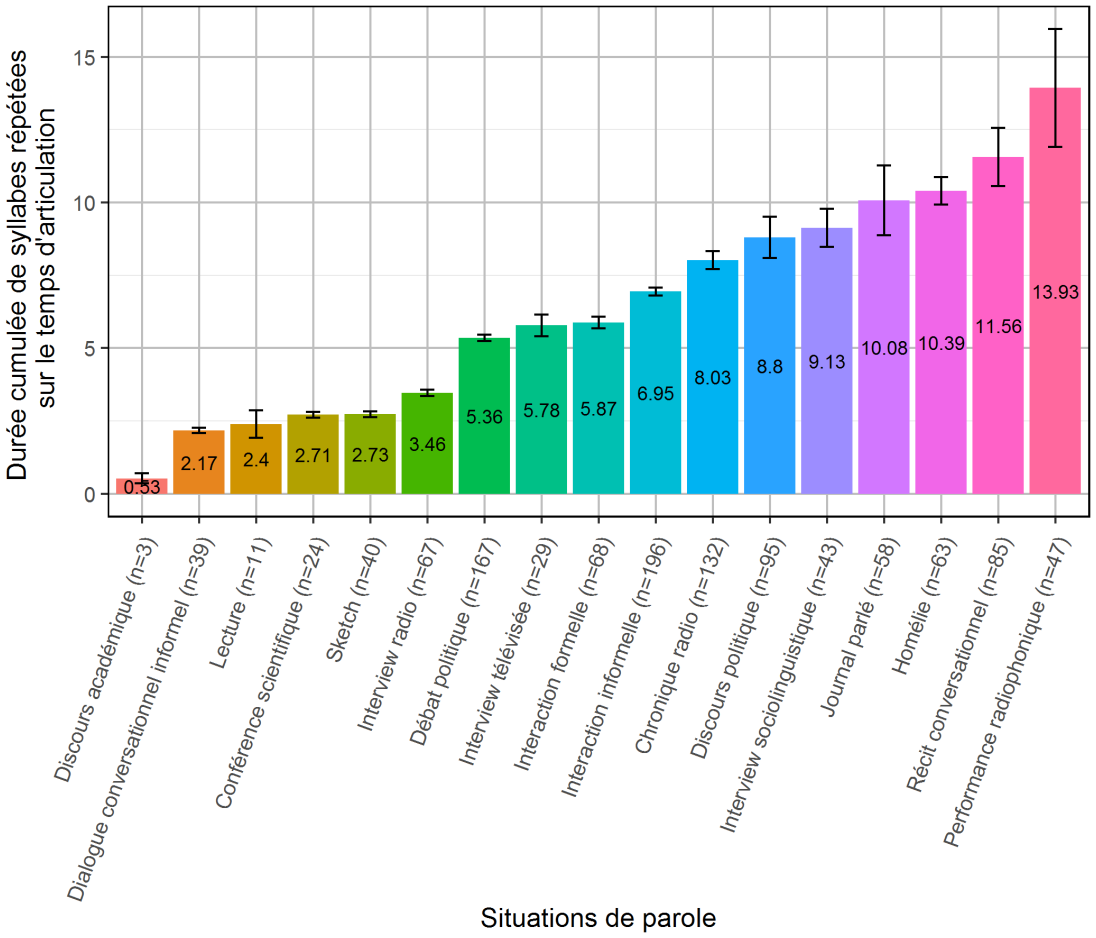


FIGURE 6.1 – Proportion de temps de répétitions sur le temps d'articulation (n = *nombre de répétitions*).

Bien que la proportion de parole répétée d'une situation à une autre ne semble pas différente, nous avons formulé la question suivante : **à proportion de (dis)fluences égale, certaines situations privilégient-elles les répétitions plutôt que d'autres marques** ? Pour tester cette distinction, nous différencions **trois types de segments** au sein des corpus : (1) les segments concernés par une répétition (`rep`), (2) les segments concernés par une (dis)fluence autre que la répétition (`dis-nrep`), et (3) les segments qui ne sont pas porteurs d'une (dis)fluence (`ndis`). Le ratio (`rep` / `dis-nrep` $\times$ 100) exprime la proportion relative de répétitions par rapport à la production de (dis)fluences.

La Figure 6.2 associe la proportion de chacun de ces types de segments aux situations de parole. Le graphique regroupe en ordonnée les situations de parole, tandis qu'on retrouve en abscissse les proportions de chaque type de segment. La taille des boîtes est proportionnelle au nombre d'occurrences et la **taille de l'effet** (corrélation positive ou négative) est indiquée par la **mesure de résidu** (à droite). De manière générale, les

situations les moins marquées par la **(dis)fluence (lexicale)** sont la chronique radiophonique (`chrorad`), le journal parlé (`jpa`), les discours académiques (`aca`), les lectures (`lec`) et les sketchs humoristiques (`sketch`) (Tableau 6.1).

Situation	ndis	nrep	rep
jpa	**84.80**	14.82	0.38
aca	**77.20**	22.15	0.65
lec	**77.34**	21.64	1.02
pol	74.24	24.22	1.54
hom	72.16	26.27	1.57
cnf	67.20	30.72	2.08
sketch	75.79	20.63	3.58
chrorad	**82.17**	14.24	3.59
single	**59.78**	34.44	5.78
rcv	63.13	30.55	6.32
deb	68.28	23.53	8.19
conv-i	63.21	28.14	8.65
intrad	63.37	27.08	9.55
conv-f	**58.48**	31.51	10.02
intfor	65.57	23.93	10.49
intlib	**58.74**	28.82	12.44

TABLEAU 6.1 – Proportion de répétitions par rapport aux autres (dis)fluences et segments non (dis)fluents par situations de parole.

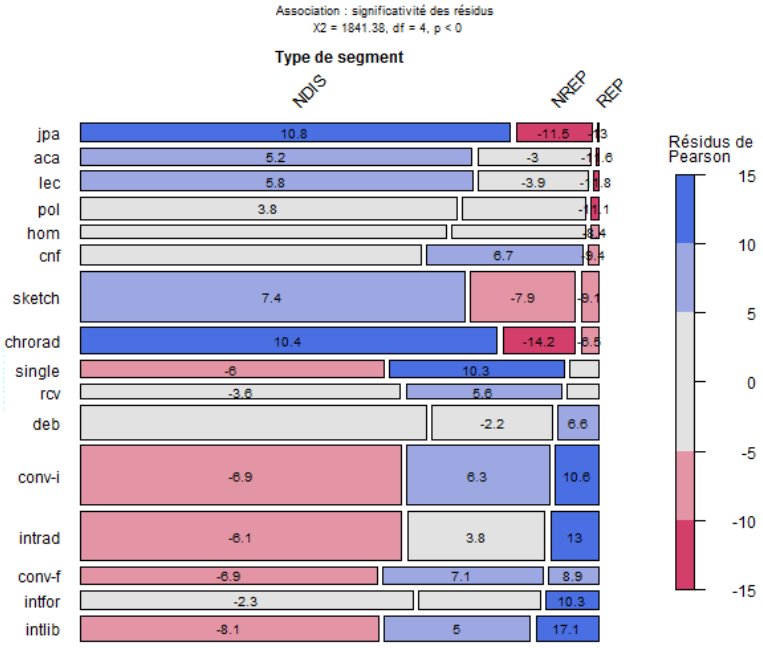


FIGURE 6.2 – Association de répétitions par rapport aux autres (dis)fluences et segments non (dis)fluents par situations de parole.

La **proportion absolue** de répétitions à l'identique est particulièrement importante pour les situations d'interviews libres, d'interviews formelles et de conversations informelles (trois dernières lignes du Tableau). En calculant la **fréquence relative** de segments répétés par rapport aux autres (dis)fluences, nous faisons face aux mêmes constats. Les données attestent d'une différence entre les situations ($F(16, 86) = 2.67$, $p < 0.0019$), avec significativement moins de répétitions pour la situation de journal parlé par rapport à tous les autres types de (dis)fluences (résidu $= -18.5$, $p < 0.03$), pour la lecture (résidu$= -17.17$, $p < 0.07$), le discours académique, le discours politique et l'homélie (avec des résidus et des seuils de significativité comparables à ceux de la lecture). Toutefois, un test de Tukey DSH révèle que **nous ne pouvons opposer un groupe de situations comme plus enclin à faire apparaître la répétition comme marque de (dis)fluence plus qu'un autre groupe de situations**.

Sur base des mêmes critères, l'observation des fréquences relatives et absolues des répétitions en fonction du **degré de préparation** des situations s'avère plus pertinente. Les proportions absolues sont effectivement différentes en fonction du degré de préparation ($\chi^2 = 1841$, $df = 4$, $p < 0.001$) (Tableau 6.2). De plus, la fréquence relative de segments répétés par rapport aux segments (dis)fluents est également significativement différente en fonction des modalités ($F(2, 100) = 8.83$, $p < 0.001$). En effet, un test post-hoc de Tukey DSH permet d'établir que les situation préparées ont, à proportion

de (dis)fluences égale, tendance à moins contenir de répétitions que les situations semi-préparées et spontanées (Figure 6.3). Au reste, aucune différence n'existe entre ces deux dernières modalités.

Degré de préparation	ndis	rep	rep
préparé	78.45	19.96	1.59
semi-préparé	68.25	25.89	5.85
spontané	62.79	27.80	9.40

TABLEAU 6.2 – Proportion de répétitions par rapport aux autres (dis)fluences et segments non (dis)fluents par degré de préparation.

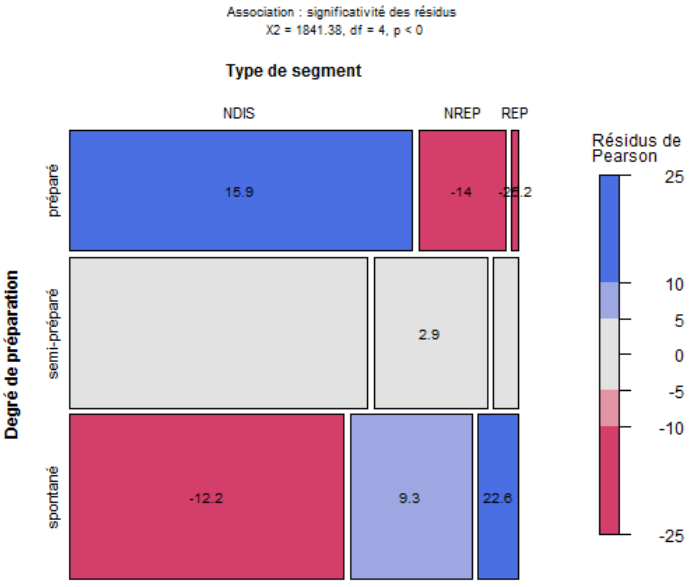


FIGURE 6.3 – Association de répétitions par rapport aux autres (dis)fluences et segments non (dis)fluents par degré de préparation.

Enfin, nous effectuons le même type de comparaison pour deux traits situationnels additionnels. Le **degré d'orientation médiatique** d'une situation ne s'avère pas pertinent à la prédiction de la proportion relative de (dis)fluences produites ($F(2, 100) = 0.21$, $p = 0.81$). Par contre, la différence en proportion relative de répétitions produites en fonction du **degré d'interactivité** présente une taille d'effet considérable ($F(2, 100) = 9.36$, $p < 0.001$). Un test post-hoc de Tukey DSH distingue 2 groupes : les situations semi-interactives et interactives s'opposent aux situations non interactives, avec une tendance pour les premières à contenir plus de répétitions que les situations non interactives de manière absolue et à taux de (dis)fluences égal.

En conclusion, les résultats révèlent que si la proportion de répétitions varie effectivement en fonction des situations, les différences sont **trop faibles pour permettre d'opposer un groupe de situations à un autre**. Le regroupement par traits situationnels est nécessaire. Il met en avant que le **dégré de préparation** d'une situation et son **degré d'interactivité** sont les traits les plus révélateurs des **proportions absolues et relatives de répétitions** produites dans une situation. Les répétitions sont, à proportion égale de (dis)fluences, davantage des (dis)fluences caractéristiques de situations de paroles spontanées et (semi-)interactives.

1.3 Fréquences et structures des répétitions en fonction des cooccurrences

Nous continuerons notre réflexion en évaluant si le type de situation influence aussi **la structure des répétitions observées**. Avant cela, revenons sur les caractéristiques définitoires des répétitions : (1) nombre de répétés, (2) catégorie du mot concerné par la répétition, (3) longueur du répétable, (4) présence et type de phase d'édition et (5) type de segment précédant et suivant la (séquence de) répétition(s). Nous formulons les hypothèses suivantes :

Hypothèses de structure interne

- Les répétitions sont majoritairement **monolexicales**, **monosyllabiques** et s'actualisent sur des **mots clitiques**.
- Les répétitions peuvent être **directes** comme **associées** (i.e. comprendre ou non une phase d'édition).
- Les répétitions associées comprennent, en phase d'édition, majoritairement des **pauses silencieuses** et/ou des **pauses pleines**.

Hypothèses de cooccurrences

- Les répétitions polylexicales amènent à davantage de rupture discursive. Elles sont plus suivies de pauses que les répétitions monolexicales.
- Les répétitions directes, sans phase d'édition, sont davantage précédées de cooccurrences que les autres types de répétitions. Au contraire, les **répétitions associées** avec une ou plusieurs marques en phase d'édition sont moins précédées ou suivies de pauses ou de (dis)fluences : elles sont **liées au contenu qui les précède et les suit**.

(a) Nombre de répétés : répétition simple, double et multiple

Dans les corpus à l'étude, la majorité des répétitions annotées sont des répétitions simples (de 76% à 88% dans nos corpus). Une part non négligeable des répétitions contient deux répétés (17% dans C-HUMOUR contre 9.5% dans LOCAS-F). Le nombre maximal de répétés est de 6 dans nos corpus, contre 5 chez Candea (2000a:28) et de 10 chez Dister (2007:131). Enfin, seule une minorité de répétitions sont supérieures à deux éléments (Tableau 6.3).

Corpus	**simple**		**double**		**multiple**	
	(n.)	(%)	(n.)	(%)	(n.)	(%)
LOCAS-F	607	87.7	66	9.5	19	2.8
C-HUMOUR	261	75.8	58	16.9	25	7.3
METACORPUS	2025	90.1	174	7.7	48	2.1

Tableau 6.3 – Proportion du nombre de répétés par corpus.

Ces résultats sont comparables à ceux de Dister (2014) et de Candea (2000a). Cependant, on retrouve en général dans les corpus LOCAS-F et C-HUMOUR **moins de répétitions simples** que dans Metacorpus qui résulte pour rappel d'une annotation automatique des répétitions dans 5 corpus de parole. Une hypothèse est que l'annotation manuelle humaine se focalise moins sur la répétition de mots grammaticaux et/ou de répétitions à structure très régulière.

Sur LOCAS-F et C-HUMOUR, contrairement aux résultats relatifs à la proportion de répétitions, aucune différence significative n'est observée entre le nombre de segments répétés et le degré de préparation ($\chi^2 = 6$, $df = 4$, $p = 0.2$) ou le degré d'interactivité ($\chi^2 = 9$, $df = 4$, $p = 0.05$) d'une situation. Par contre, il existe bien des associations avec le **type de situation** ($\chi^2 = 50$, $df = 30$, $p < 0.01$), le **nombre de locuteurs** ($\chi^2 = 20$, $df = 4$, $p < 0.001$) et le **degré d'orientation médiatique** ($\chi^2 = 18.2$, $df = 4$, $p < 0.001$). Toutes choses égales par ailleurs, les répétitions doubles s'actualisent davantage en **dialogue** qu'en monologue ou multilogue. Cet effet reste cependant marginal ($\chi^2 = 20$, $df = 4$, $p = 0.001$). De même l'actualisation de **répétitions simples** est relativement plus importante en situation non orientée médiatiquement qu'**en situation à orientation médiatique**. Nous reviendrons sur le **croisement des critères** pertinents orientant la structure d'une répétition en section 1.6.

(b) Nature grammaticale du mot concerné par la répétition : mots pleins /vs/ mots outils

Le Tableau 6.4 [10] montre que les répétitions simples sont majoritairement actualisées sur des mots outils (64.1% des répétitions simples). C'est aussi le cas pour les répétitions doubles et multiples (61% et 56% respectivement).

Nature	Corpus	**simple**		**double**		**multiple**		**Total**	
		(n.)	(%)	(n.)	(%)	(n.)	(%)	(n.)	(freq. rel)
MO	LOCAS-F	1327	46.6	163	32.9	67	30.0	1557	67.7
	C-HUMOUR	515	18.1	134	27.0	58	26.0	707	64.6
MP	LOCAS-F	621	21.8	105	21.2	39	17.5	765	23.1
	C-HUMOUR	386	13.5	94	19.0	59	26.5	539	37.8

Tableau 6.4 – Fréquences des unités lexicales faisant l'objet d'une répétition par nombre de répétés et fréquences relatives par 1000 mots sur l'ensemble du corpus.

Ces chiffres doivent être appréhendés au regard de la **proportion relative de cliticques** dans les corpus. Ainsi, la dernière colonne du Tableau ci-avant reprend le ratio (`clitique en répétition` / `clitique dans le corpus` × 1000). Elle exprime la proportion relative de mots clitiques dans la répétition par rapport à la production totale

10. Ces chiffres sont comparables avec ceux de Candea (2000a:316), car celle-ci opère comme nous une analyse en fréquence relative des unités faisant l'objet d'une répétition. Cependant, ces répétitions ne concernent que celles reconnues comme marque du travail de formulation.

de clitiques. Les données montrent ainsi également que pour 1000 mots prononcés, **64.6 et 67.7 seront concernés par une répétition de clitiques** contre 37.8 et 23.1 par une répétition de mots lexicaux. Ces résultats sont sans surprise, ils confirment les observations déjà faites par Grosjean & Deschamps (1975), Henry & Pallaud (2002), Candea (2000b:316) et Dister (2007:160). S'ils apportent peu d'informations en tant que tels, leur association à d'autres caractéristiques internes de la répétition permet de vérifier l'hypothèse selon laquelle **les répétitions de MO obéissent à des associations plus systématiques que les répétitions de MP, qui auraient un caractère plus accidentel**[11]. Nous revenons sur cette distinction dans notre modélisation finale (section 1.6)[12].

(c) Longueur du répétable

Nous distinguons trois catégories de répétables en fonction de leur longueur en mot graphique[13]. Dans nos corpus, les cas les plus fréquents sont ceux où le répétable est composé d'**un seul mot : 50% des répétitions** observées contre 80% chez Dister (2007:136) (Tableau 6.5). A titre de comparaison, Metacorpus révèle des proportions légèrement différentes, soit 62% de répétitions monolexicales, 24.4% de répétables composés de deux mots et 13.6% de plus de deux mots. Ces observations sont peu comparables à celles d'Henry & Pallaud (2002), qui révèlent 86% de répétitions d'un mot contre 17.7% de 2 mots. Du reste, il y a cependant proportionnellement moins de répétitions à mot unique dans nos corpus que dans l'étude d'Henry, Campione & Véronis (2004).

Corpus	**1 mot**		**2 mots**		**> 2 mots**		**Total**	
	n	%	n	%	n	%	n	%
LOCAS-F	388	56.2	133	19.2	171	24.2	344	100
C-HUMOUR	176	51.2	79	23.0	89	25.9	692	100
Total	564		212		260		1036	

Tableau 6.5 – Longueur du répétable en nombre de mots.

Dans nos corpus, il n'existe pas de relation entre la longueur du répétable en mot et le degré de préparation ($\chi^2 = 1$, $df = 4$, $p = 0.9$), mais bien entre la longueur du répétable en nombre de syllabes. Dans Grosman, Christodoulides, Degand & Simon (2017), nous avions observé, dans l'ensemble de données Metacorpus, **61.3% de répétitions monosyllabiques**, contre > 23% de bisyllabiques et 14% de plus de 2 syllabes. Les répétitions ont tendance à y être plus courtes en discours spontanés que

11. Voir section 1.5 de ce chapitre.

12. Pour ces mêmes raisons et par manque de place, nous ne nous attarderons pas sur les catégories grammaticales des mots répétés, nous envisagerons davantage les catégories syntaxiques de séquences fonctionnelles concernées par la répétition.

13. La longueur d'un répétable peut également se calculer en nombre de **syllabes articulées** ou en fonction de sa **durée** (absolue ou relative, i.e. durée du *reparans* / durée du *reparandum*). Voir section 2.4.3 de ce chapitre pour cette dernière approche.

préparés ou semi-préparés ($\chi^2 = 22.507$, $df = 3$, $p < 0.001$)[14]. Ainsi, dans les discours préparés, **les répétitions ne sont pas plus courtes en terme de nombre de mots répétés mais bien en nombre de syllabes**.

L'association par mot est également absente avec le degré d'orientation médiatique ($\chi^2 = 3$, $df = 4$, $p = 0.5$) ou le degré d'interactivité d'une situation ($\chi^2 = 5$, $df = 4$, $p = 0.3$). **Ainsi, si l'on se répète plus dans les situations non préparées et (semi-)interactives (voir section 1.2), on ne répète pas pour autant de segments plus courts ou plus longs**.

(d) Présence et type de phase d'édition

La phase d'édition peut comprendre des marques de différents types. Lorsque Candea (2000a:28-30) inclut dans ses analyses les séquences de répétitions porteuses de terme(s) d'édition, elle remarque que ces derniers peuvent être un son, une syllabe, un mot, une amorce de mot, de plusieurs syllabes ou de plusieurs mots, tant qu'ils ne portent « **aucune valeur sémantique** » (e.g. : « il est p/ p/ pas »). Au sein des répétitions des corpus C-HUMOUR et LOCAS-F, on décompte **78 types de termes d'édition possibles** dans la phase d'édition.

Cependant, cette variété ne doit pas dissimuler le fait que, selon nos observations, **près de la moitié** des répétitions sont produites **sans termes d'édition** (45.83%). De manière générale, nous observons relativement moins de répétitions avec une phase d'édition qu'Henry, Campione & Véronis (2004:2) qui notent que « sur 141 répétitions, 109 sont associées à (au moins) une pause, silencieuse ou remplie, soit 77,3% ». Du reste, la proportion de répétitions directes, sans terme d'édition, est comparable à celle observée dans Metacorpus (55.8%).

Parmi les répétitions incluant une phase d'édition, **trois marques** sont privilégiées : **la pause silencieuse (29.7%), le marqueur de discours (10.1%) et la troncation (9.6%)**. Une part marginale des répétitions inclue des marques composées de (dis)fluence (6.5%) ou une pause pleine (4.9%). Les autres combinaisons les plus fréquentes impliquent deux marqueurs, avec comme combinaison privilégiée (pause silencieuse + md) (4.5%), (troncation + pause silencieuse) (3.8%) et (pause pleine + pause silencieuse) (3.08%). Le reste des combinaisons représente de 3% des actualisations. Le type de phase d'édition est très fortement associé au degré de préparation ($\chi^2 = 88.24$, $df = 14$, $p < 0.001$).

Par ailleurs, Candea (2000a:28-30) observait déjà qu'**une pause silencieuse apparaît en moyenne une fois sur deux en phase d'édition** (en particulier pour les répétitions de MO) (Candea 2000a:344). Nos observations convergent : on décompte 306 répétitions émergeant avec une pause silencieuse sur les 552 répétitions contenant une phase d'édition, **soit 56%**. On relève également 28.5% des répétitions ($n = 157$) incluant un md en phase d'édition (contre 11.3% dans Metacorpus), et 15.9% incluant une pause pleine ($n = 88$) (contre 21% dans Metacorpus).

14. Cependant, cette statistique devrait être calculée par rapport à la fréquence de mots mono /vs/ polysyllabiques sur l'entièreté du corpus.

Chez Henry, Campione & Véronis (2004:2), le caractère simple ou complexe de la répétition n'a pas d'influence significative sur cette répartition. Nous évaluons les associations entre les types de répétitions (simples, doubles, multiples) et les phases d'édition à la section suivante.

(e) Type de segment précédant et suivant la (séquence de) répétition(s)

Nous avons extrait le type de segment suivant ou précédant une répétition. Dans les deux cas, le mot non annoté d'une autre (dis)fluence, **un item lexical, est le segment qui précède et suit majoritairement la répétition**. Ainsi, la répétition intervient comme un véritable phénomène de reprise après d'autres (dis)fluences et après laquelle le discours reprend son cours.

De manière générale, la répétition est moins souvent précédée (44.2%) que suivie (58.5%) d'un mot lexical. Une répétition est suivie donc majoritairement d'une reprise directe de l'énoncé. Cette observation est appuyée par le fait que la répétition est **rarement suivie d'un silence initiant un changement de tour de parole (*gap*) ou de la parole d'un autre locuteur** (1.11% et 3.76% des cas). En revanche, la répétition semble plus souvent initier la prise de parole et/ou intervient en chevauchement de prise de parole (8.13% des cas). Par ailleurs, si elle n'initie pas le tour, la répétition est précédée d'une marque initatrice de segmentation, en particulier les marqueurs de discours (14.71%) et la pause silencieuse (12.92%) (Tableau 6.6). Du reste, comme le fait remarquer Candea (2000a:322), la pause après la répétition est relativement rare quoique présente (7.61%), la structure ayant tendance à être davantage suivie de troncation(s) et de répétition(s) à l'identique (6.67% et 6.24% des occurrences).

L'hypothèse selon laquelle les répétitions directes (sans phase d'édition), sont plus souvent précédées de cooccurrences que les autres types de répétition est invalidée. Au contraire, **les répétitions associées avec une ou plusieurs marques en phase d'édition sont moins précédées ou suivies de pauses ou de (dis)fluences** : elles sont davantage **liées au contenu qui les précède et leur succède**. Cette hypothèse rappelle les distributions de Shriberg (1995) qui, sur un corpus de données expérimentales, distingue les répétitions prospectives des répétitions rétrospectives[15]. Les **répétitions rétrospectives**, associées à une phase d'édition et non suivies par une marque de (dis)fluence sont, chez Shriberg (1995), largement majoritaires.

Nous confirmons ces observations en français dans une étude portant sur les répétitions à l'identique annotées automatiquement (Grosman, Christodoulides, Degand & Simon 2017). Nous mettions effectivement en évidence que les répétitions sont majoritairement rétrospectives plutôt que prospectives. Mais sutout, la **proportion de répétitions sans *interregnum* et sans segment après le *reparans* domine largement les corpus** et ce, quel que soit la longueur du répétable (1 mot, 2 mot, > supérieur à 2 mots) ou le nombre de répétés (1 répété, 2 répétés, > à 2 répétés).

15. Voir chapitre 2 section 2.4.3.

Type de segment	**précédant**		**suivant**	
	n	%	n	%
gap	34	2.91	13	1.11
parole d'un autre locuteur	102	8.13	44	3.76
mot	517	44.23	684	58.51
pause silencieuse	151	12.92	89	7.61
pause pleine	43	3.68	23	1.97
marqueur discursif	172	14.71	53	4.53
troncation	30	2.57	78	6.67
répétition à l'identique	48	4.11	73	6.24
répétition avec modification	16	1.37	32	2.74
substitution propositionnelle	11	0.94	23	1.97
substitution morphologique	0	0	0	0
faux départ	10	0.86	3	0.26
insertion lexicale	1	0.09	22	1.88
insertion parenthétique	4	0.34	3	0.26
autre (chaîne de) (dis)fluence	30	2.54	29	2.74
Total	1169	100	1169	100

TABLEAU 6.6 – Type de segments qui précèdent et suivent les répétitions.

Nous soutenons ainsi que **réduire à ces deux paramètres** la répétition (présence /vs/ absence d'une pause après, et présence /vs/ absence d'une phase d'édition) ne permet de représenter qu'une faible partie des répétitions actualisées dans les corpus (moins de 10% des répétitions de Metacorpus). Pourquoi limiter l'analyse à la pause silencieuse subséquente lorsqu'on observe que la troncation et la répétition à l'identique sont postposées en proportion similiaire ? Pourquoi ne pas par exemple distinguer les répétitions en fonction des **segments précédant la pause** également ?

Nous proposons en Figure 6.4 une représentation des associations de répétitions en fonction des 4 paramètres analysés ci-avant [16] afin de faire émerger par les corpus des catégories de répétitions. La Figure 6.5 représente le même ensemble de données mais spécifie pour chaque modalité de variable son **apport informationnel** dans l'explication de la variance des structures.

16. Nous avons exclu de l'analyse la distinction clitique/non clitique à des fins de représentation.

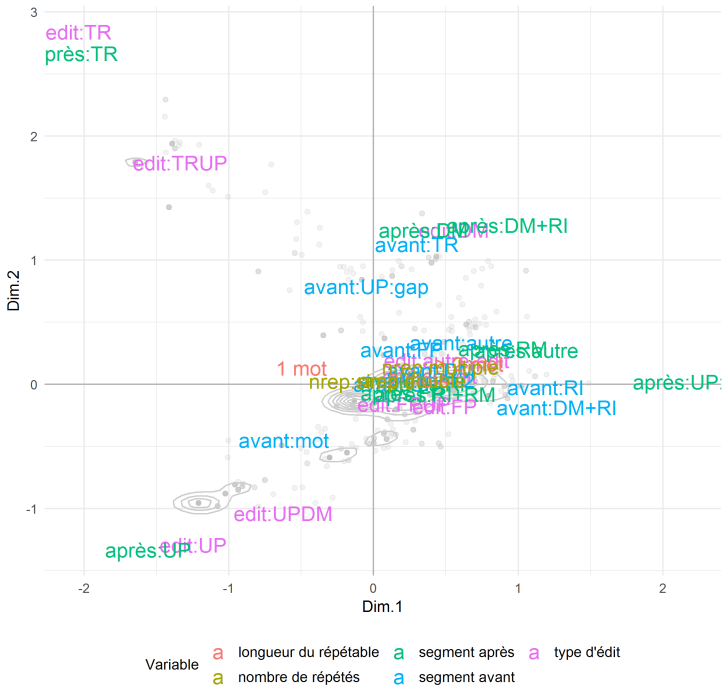


FIGURE 6.4 – Analyse en composantes multiples des structures de répétitions et des cooccurrences par variable.

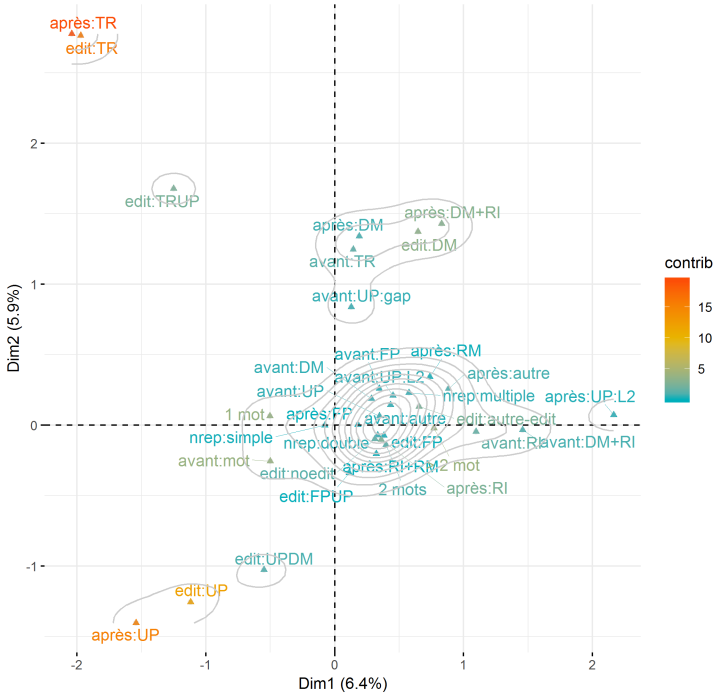


FIGURE 6.5 – Analyse en composantes multiples des structures de répétitions et des cooccurrences par rapport à leur pertinence informationnelle.

En d'autres termes, la Figure 6.5 rend explicite **les associations et les niveaux de variables les plus importants** lors de la caractérisation d'une répétition [17]. Une lecture en cluster révèle la présence d'au moins **4 *clusters* distincts** :

- **Le cluster « marqueur de discours et troncations »**. En haut à gauche, on remarque une tendance des répétitions avec phase d'édition comprenant une troncation et/ou un md à être suivies par un md et une troncation. Elles ont également tendance à initier un tour de parole (`avant:UP:gap`).
- **Le cluster « pause silencieuse et md »**. En bas à gauche, les répétitions porteuses d'une phase comprenant une pause silencieuse ou (une pause silencieuse + md) ont tendance à être suivies d'une pause silencieuse également.
- **Le cluster « un segment = un répété »**. A gauche, on retrouve les répétitions composées d'un seul répété d'une longueur de répétable équivalente à un mot. Ces répétitions ont tendance à être précédées d'un item lexical, d'une pause pleine ou d'une pause silencieuse.
- **Le cluster « plusieurs segments, plusieurs répétés »**. A droite, on retrouve les répétitions composées de deux segments ou plus et de plus d'un répété. Celles-ci peuvent être précédées d'une pause silencieuse également, mais sont généralement caractérisées par des phases d'édition plus complexes (e.g. (pause pleine + pause silencieuse), autres termes d'édition, etc.), et sont suivies par la parole de L2, et des répétitions.

On note que les **modalités essentielles dans l'établissement de catégories** de répétition en fonction de leur structure interne et leurs associations sont (1) la répétition de plus de deux répétés, (2) la présence d'une pause silencieuse postposée à la répétition. On note également une importance relative des modalités, (3) la répétition composée d'un seul répété, (4) la répétition dont le répété est monolexical, (5) la présence d'une pause silencieuse en phase d'édition et, (6) la présence d'une troncation après la répétition.

Nous avons séparé les différentes variables à l'étude dans la Figure 6.6 et nous avons tracé les **ellipses de confiance** pour leurs différentes modalités. Cette approche permet d'examiner si les modalités sont significativement différentes les unes des autres. L'analyse exploratoire révèle ainsi que 4 variables ont des modalités particulièrement différentes et quelles modalités pourraient être regroupées :

> **Type de phase d'édition**. On distingue au moins 3 ellipses différentes, l'une isolant les phases d'édition comprenant une **troncation** (en haut à gauche), l'autre les **phases d'édition avec pause silencieuse** et la troisième distinguant les cooccurrences de **(pause silencieuse + md)** au sein des phases d'édition. Ensuite, les structures porteuses d'un marqueur de discours en phase d'édition sont isolées des autres. Au centre, les quatres modalités restantes affichent peu de variations et peuvent donc ne pas être distinguées dans les analyses.

17. Brièvement, l'analyse en composantes multiples est une méthode multifactorielle permettant ici de décrire l'ensemble des répétitions en fonction des différentes variables qualitatives. Nous renvoyons le lecteur à Kassambara (2017) pour une revue extensive des méthodes et des applications.

Longueur du répétable. On peut distinguer clairement les répétitions constituées d'un unique mot des répétitions de 2 mots. Cependant la proximité des deux derniers clusters laisse à penser qu'il existe **peu de différences** entre les structures comportant **deux ou plus de deux répétables**.

Segment après. A l'instar des phases d'édition, il est possible de distinguer les structures en fonction du fait qu'elles soient **suivies d'une pause pleine, d'une troncation ou d'un marqueur discursif**. Au centre, les modalités ne sont pas significativement différentes. A l'extrême droite cependant, les répétitions suivies directement de la **parole d'un autre locuteur** se distinguent dans leur structure (elles sont par ailleurs associées à des répétitions avec des phases d'éditions plus complexes (`autre-edit`), à longueur de répétable multiple et précédées d'autres répétitions).

Nombre de répétés. Si l'on peut distinguer les répétitions simples des doubles et multiples, l'analyse ne permet pas de distinguer clairement deux modalités distinctes avec d'une part les répétitions au répété double et d'autre part les répétitions avec un répété multiple. Ce type d'observation incite au **regroupement en une catégorie binaire** pour les analyses de modélisation.

En conclusion, l'analyse en composantes multiples sur les seuls paramètres de structure permet, sur les deux première dimensions, d'expliquer 14% de la variance des répétitions en cooccurrence. Bien que cette analyse en composantes multiples mette en lumière des associations entre les types de structure et leur environnement, elle ne prend pas en compte l'association de ces chaînes avec les **frontières d'unités syntaxiques**, ni l'association de ces chaînes avec des **profils acoustiques distincts**. C'est ce que nous nous proposons de faire aux sections suivantes.

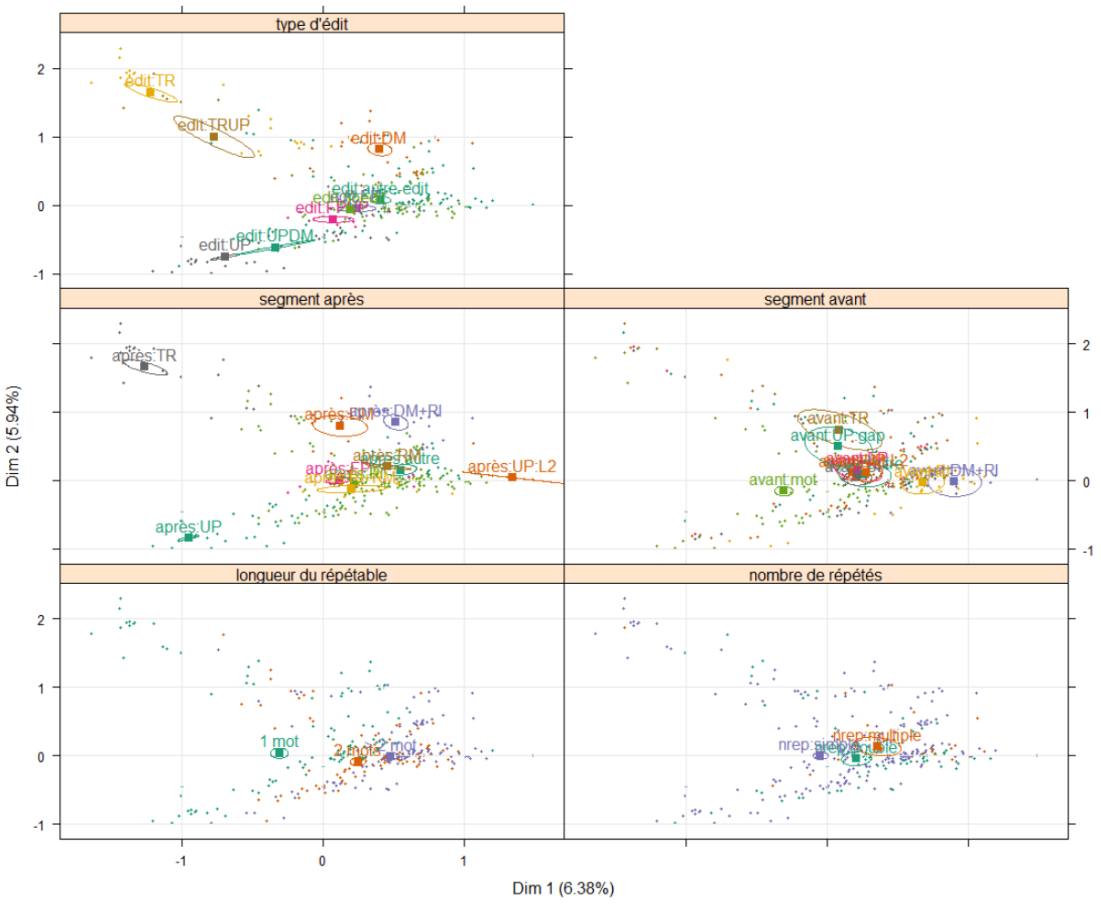


FIGURE 6.6 – Analyse en composantes multiples par variable et par modalité avec représentation des ellipses de confiance.

1.4 Fréquences et positions des répétitions en fonction des unités syntaxiques et tours de parole

Nous questionnons la tendance des répétitions à segmenter le flux discursif en fonction de leurs distributions syntaxiques et de leur position au sein des tours de parole :

1. Dans **quelles positions syntaxiques** se trouvent les mots concernés par la répétition (*reparandum*, phase d'édition et *reparans* compris) ? Que représentent ces fréquences relativement à l'entièreté du corpus ?
2. Quelles positions syntaxiques sont associées à l'**initiation et à la fin d'une répétition** ?
3. Le début et la fin d'une répétition sont-ils associés au **début et/ou à la fin d'un tour de parole** ?

Au contraire des chapitres précédents, il ne s'agit donc pas ici d'observer la durée d'une marque par rapport au degré de frontière syntaxique mais bien de savoir si les répétitions ont tendance à s'actualiser dans certaines positions au sein des unités syntaxiques et des tours de parole. Nous formulons les hypothèses suivantes :

Hypothèses syntaxiques et interactionnelles

- Les répétitions peuvent s'actualiser à toute frontière syntaxique. Elles sont particulièrement présentes **en début d'unité**.
- Les fins de répétitions ne sont pas associées à une rupture syntaxique ou à une alternance de tour de parole.
- Cependant, plus la structure d'une répétition est complexe plus celle-ci est **associée au marquage d'une rupture syntaxique**.

Le Tableau 6.7 distingue le nombre d'unités syntaxiques répétables [18] et le nombre d'unités syntaxiques concerneés par une répétition dans LOCAS-F. Ces données attestent du fait que les répétitions concernent 5.08% à 5.68% des unités syntaxiques du corpus. On ne relève aucune différence significative entre les types d'unités.

Unités	Répétables (n.)	Impliquant au moins un répétition (n.)	Proportion relative (%)
rection	6179	351	5.68
séquence	10303	541	5.25
ga accentué	9512	483	5.08

TABLEAU 6.7 – Proportion relative d'actualisation des répétitions au sein des unités syntaxiques de LOCAS-F.

18. i.e. Nous créons donc un sous-corpus, dont nous excluons les *gap* et les silences dus à la parole d'un second locuteur. Nous excluons donc les unités composées uniquement d'items non lexicaux de ce décompte (e.g. pauses silencieuses, pauses pleines, troncations et phénomènes paralinguistiques tels que le rire, la toux et les segments incompréhensibles).

Pour aller plus loin, le Tableau 6.8 représente les frontières syntaxiques associées à l'initiation et à la finalisation d'une répétition. Les fréquences relatives révèlent qu'on observe respectivement **44 et 31 initialisations de répétitions tous les 1000 mots occupant la position de frontière inter-rectionnelle et inter-séquentielle**. A l'extrême opposé, on ne trouve que **6 répétitions tous les 1000 mots en dehors de toute frontière, au sein d'un ga**. Cette dernière observation supporte le fait que lorsqu'une unité est initiée par une répétition, la répétition est bien marque d'**initiation d'une nouvelle unité syntaxique**, majoritairement inter-rectionnelle ou inter-séquentielle.

Position	**intra-ga** (20755)		**inter-ga** (10524)		**inter-seq** (3917)		**inter-rec.** (5786)	
	(n.)	(f.rel.)	(n.)	(f.rel)	(n.)	(f.rel)	(n.)	(f.rel.)
Début de répétition	127	6.12	178	16.9	124	31.66	256	44.24
Fin de répétition	479	22.02	153	11.90	16	3.28	43	3.64

TABLEAU 6.8 – Fréquences absolues et relatives des répétitions en fonction du degré de frontière syntaxique dans LOCAS-F.
Les fréquences relatives sont normalisées pour 1000 unités syntaxiques.

Les fins de répétitions permettent, elles, d'éclairer sur l'éventuelle fonction de rupture discursive de la répétition. On remarque, toujours au Tableau 6.8, que les fins de répétitions ne sont pas associées aux degrés de frontières maximales : moins de 6 frontières inter-rectionnelles et inter-séquentielles sur 1000 sont provoquées par la finalisation d'une répétition. Au contraire, on observe une tendance des répétitions à ne pas s'associer avec des frontières syntaxiques fortes : 22.02 et 11.9 fin de répétitions sur 1000 mots – ce qui représente les fréquences relatives les plus élevées. Les **positions de fin de répétitions** sont exemptes de frontières syntaxiques ou sont **positionnées entre des ga** : elles **maintiennent la cohésion séquentielle et donc fonctionnelle**. Ces observations sont bien entendu liées au fait que les répétables ont tendance à s'actualiser sur « les petits mots du discours » et sont donc **propices à la réinitialisation syntagmatique**. Comme le faisait remarquer DISTER (2007), ce n'est pas sur le lexème seul que se manifeste le travail de formulation mais sur toute l'unité dont il est la tête (DISTER 2007:163). Cette constatation favorise la reconnaissance de la répétition comme marque (ou moyen) de **continuité syntaxique** dans l'énoncé, puisque les constructions ouvertes par un mot outil sont au final achevées au niveau syntaxique [19].

Au niveau interactionnel, la continuité syntaxique pourrait être fragilisée par **l'alternance de tour après la répétition**. Cependant, les résultats des sections antérieures ont déjà révélé que moins de 5% des répétitions étaient suivies par une alternance de tour.

19. D'ailleurs, dans un sous-corpus de DISTER (2007), moins de 6% des répétitions (66 occurrences) occasionnent une réelle **rupture syntaxique** de l'énoncé. L'auteure remarque par là-même qu'en réalité les ruptures de construction ne sont précédées d'aucune (dis)fluence lexicalisée précise, nous y revenons en conclusion de ce chapitre.

Ces observations sont confirmées plus en détail dans le Tableau 6.9. La position interactionnelle des répétitions sur les dialogues et multilogues de LOCAS-F et C-HUMOUR, ne semble pas mettre à la mal la continuité discursive.

Position		**initiale**		**médiane**			**finale**	
		frontière	début	tour mono	tour court	tour long	fin	frontière
Début REP	(n.)	74	62	6	70	546	49	26
	(f.rel.)	53.8	28.8	13.7	22.6	20.0	15.3	13.2
Fin REP	(n.)	9	50	6	89	575	67	52
	(f.rel.)	6.55	23.19	13.67	28.67	21.04	20.86	26.55

TABLEAU 6.9 – Position des répétitions par rapport aux tours de parole des dialogues et des multilogues (effectifs et fréquences normalisées pour 1000).

Au contraire, la répétition est **initiatrice de tour de parole**. Les proportions relatives les plus élevées s'observent en effet en frontière initiale et en début de tour de parole (respectivement 54 et 29 tours sur 1000). De plus, leur actualisation en fin ou en frontière de tour est particulièrement marginale (15 et 13 tours sur 1000) : lorsqu'un locuteur initie une répétition, **il ne cède que très rarement la parole**[20]. En outre, les frontières de *reparans* (fin de répétition), bien que davantage sujettes à l'alternance de tours (21 et 27 cas sur 1000), restent marginales.

En conclusion, nous avons vu dans cette section que lorsqu'une unité est initiée par une répétition, la répétition est bien marque d'initiation d'une nouvelle unité syntaxique, majoritairement inter-rectionnelle ou inter-séquentielle. De plus, lorsqu'un locuteur initie une répétition, il n'est que rarement interrompu et ne cède pas la parole. Enfin, les fins de répétitions ne sont pas associées aux degrés de frontières maximales. Les positions de fin de répétitions sont exemptes de frontières syntaxiques ou sont positionnées entre des ga : elles maintiennent la cohésion séquentielle et donc fonctionnelle mais aussi interactionnelle.

20. Notons que ce résultat est à relativiser en vertu du processus d'annotation : les répétitions initiées reprises en frontières finale dee tour représentent ici toutes des cas de chevauchement de parole. En effet, le protocole d'annotation n'étudiant que la répétition monologale, l'annotation est bloquée lors de l'alternance de tours. Cette note est également valable à la présence marginale de fin de répétition en frontière initiale de tours de parole, qui sont toutes des occurrences apparaissant en parole chevauchée.

1.5 Forme acoustique des répétitions à l'identique

Nous étudions deux aspects des formes acoustiques des répétitions : (a) l'influence des marques cooccurrentes sur **la durée** des segments de la répétition et (b) l'influence de ces mêmes cooccurrences sur la forme **des courbes intonatives** du *reparandum*, de la phase d'édition et du *reparans*[21]. La catégorisation binaire opposant les répétitions prospectives aux répétitions rétrospectives ne couvrant pas l'entièreté de la variabilité des cooccurrences observables de nos corpus[22]. Nous analysons la durée et l'intonation des répétitions au-delà de la pause silencieuse (i.e. associations avec les pauses pleines, les marqueurs de discours et les troncations).

Les analyses des deux sous-sections suivantes concernent les répétitions à l'identique annotées automatiquement par le logiciel DisMo (Christodoulides, Avanzi & Goldman 2014) dans Metacorpus et vérifiées manuellement (Christodoulides, Grosman, Degand & Simon 2017 ; Grosman, Christodoulides, Degand & Simon 2017).

(a) Allongement dans le *reparandum*

Dans cette section, nous nous focalisons sur les répétitions à l'identique d'**un seul répété**[23]. Contrairement à la littérature antérieure, nous incluons les répétitions **de mots outils et de mots pleins** (MO et MP), ainsi que de **tout contenu syllabique** afin d'appréhender la variété des structures possibles[24]. Nos analysent portent sur la **durée relative du *reparandum* par rapport au *reparans***. Nous testons les hypothèses suivantes :

Hypothèses acoustiques de durées relatives

- Le répétable est en général plus allongé que le répété (Duez 1993 ; Shriberg 1999 ; Candea 2000a:325)[25].
- L'allongement du répétable est davantage marqué pour les répétitions associées à une pause pleine en phase d'édition (Henry, Campione & Véronis 2004).
- Les répétitions suivies d'une pause ou d'une (dis)fluence sont plus longues que celles qui ne le sont pas (Shriberg 1995:386).
- Les répétitions simples de MO sont caractérisées par un allongement relatif moindre de leur durée par rapport aux répétitions de MP, plus caractéristiques du travail d'élaboration syntaxique (inspiré de Morel & Danon-Boileau (1998:85-87) et de Candea (2000a:325-327).

21. Cette étude s'inspire principalement des méthodologies et des hypothèses présentes dans Candea (2000a), Hieke (1985), Plauche & Shriberg (1999) et Shriberg (1995).
22. Voir section 1.3 de ce chapitre.
23. (il * il) /vs/ (il * il il). Ce sous-ensemble de données se justifie par la comparabilité avec les études antérieures et la représentativité de ces structures à travers le corpus.
24. L'étude ne se limite donc pas à la répétition d'articles définis monosyllabiques.
25. Le calcul de la durée relative est présenté en section 1.3c de ce chapitre.

Premièrement, dans nos Metacorpus, on remarque que dans seuls 11% des cas, le *reparans* est plus allongé que les *reparandum*. (Figure 6.7). Dans 43% des cas il est de durée comparable et dans 46% des cas le *reparandum* est même de durées supérieure au *reparans*. C'est surtout lorsqu'il est suivi d'une pause que le *reparans* a tendance à être plus long que le *reparandum* ($\chi^2 = 128.12$, $df = 2$, $p < 0.001$) (en rouge sur la Figure). Plus en détail, nous remarquons que :

– Les individus bleus du panel supérieur décrivent les répétitions **avec une phase d'édition mais avec continuation** directe après le *reparans* (répétitions dites rétrospectives). Ces dernières ont effectivement tendance à être composées majoritairement d'**un *reparandum* plus long que le *reparans***. Le *reparans* court supporte l'idée d'une **fonction de continuation** de ces répétitions.

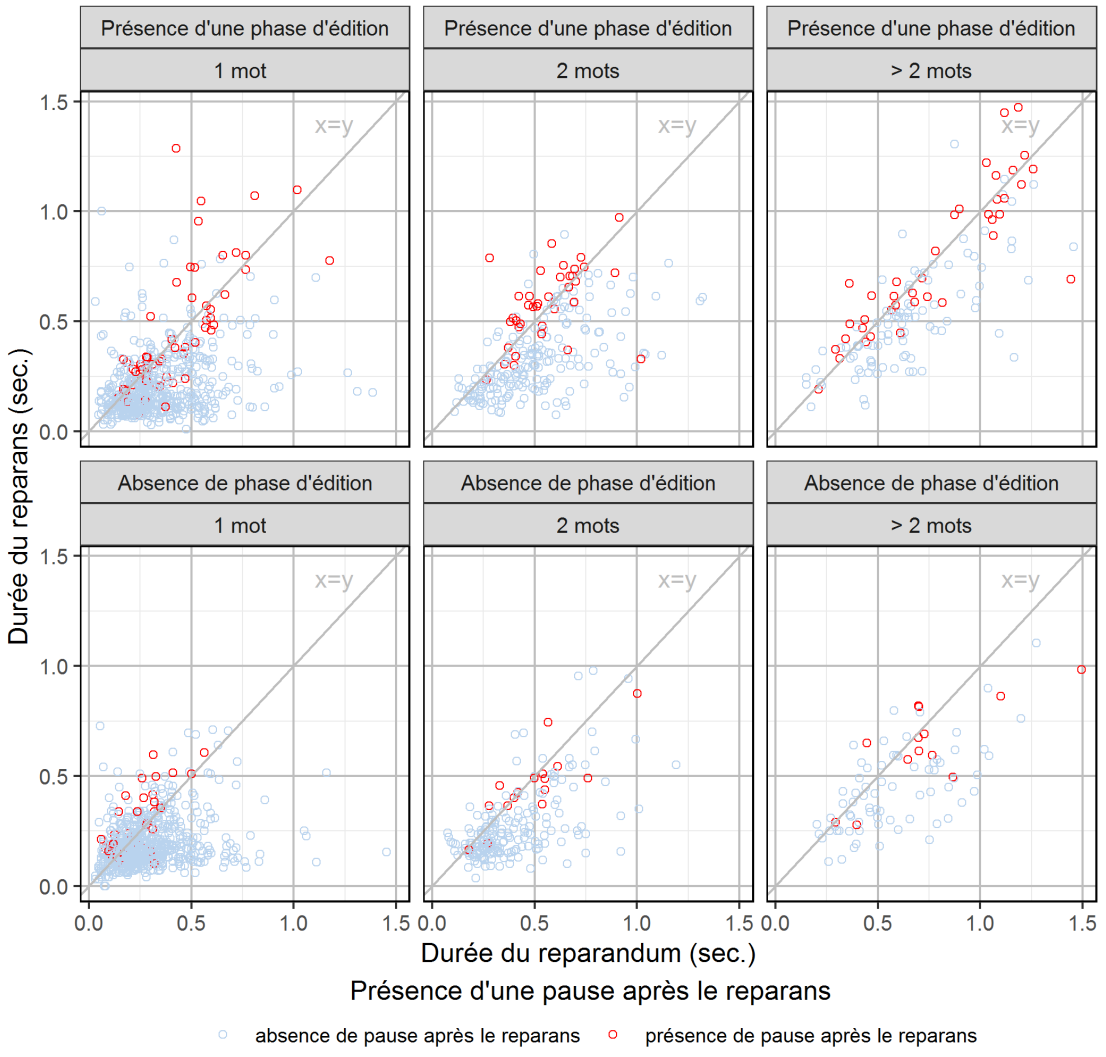


FIGURE 6.7 – Répartition des profils de répétitions en fonction du rapport (durée du *reparandum* / durée du *reparans*).
Sous-ensemble de données : répétable d'un mot (e.g. il (il) il)).

– Les individus rouges du panel inférieur décrivent les répétitions **sans terme d'édition et dont le *reparans* est suivi d'une pause** (répétitions dites prospectives). Malgré le faible nombre d'occurrences, nous observons une tendance de ces répétitions

à exhiber **un *reparans* de durée plus grande**, soutenant également la **fonction de maintien** de la parole et **éventuellement d'hésitation** supportant notamment les hypothèses de Shriberg (1999).

Néanmoins, on remarque que l'allongement du *reparans*, lorsqu'il est suivi d'une pause en particulier, s'associe à 2 autres phénomènes facteurs d'allongement en français :

1. « La **présence d'une frontière** pourrait coïncider avec l'actualisation de la pause ». La marque de frontière étant caractérisée acoustiquement par l'allongement de la dernière syllabe, cette dernière pourrait être la raison même de l'allongement du *reparans*. On pourrait penser que l'allongement du *reparans*, suivi d'une pause silencieuse, actualise une frontière prosodique en français. Or cette interprétation ne doit pas être retenue car la fin de répétition est rarement associée à une frontière syntaxique. Comme nous l'avons explicité en section antérieure, **une fin de répétition est très rarement associée au marquage de frontière syntaxique forte**, cet argument doit donc être pondéré.
2. « L'allongement relatif du *reparans* pourrait être dû à la tendance à prononcer le *reparans* avec un **nombre de syllabes supérieur** à celui du *reparandum* » [26]. Nous avons extrait toutes les répétitions identiques en nombre de mot, mais nous avons voulu garder celles qui, éventuellement, étaient marquées par une variation phonétique légère (nombre de syllabe, présence d'un schwa). Un chi-carré confrontant la durée relative des segments à la différence entre le nombre de syllabes articulées dans le *reparandum* et dans le *reparans* révèle une association entre les variables ($\chi^2 = 30, df = 8, p < 0.001$). Néanmoins, dans notre sous-ensemble de données, les répétitions montrant un *reparans* constitué d'un nombre de syllabes supérieur à celui du *reparandum* représentent une **minorité des données** ($3.5\%, n = 75$). Nous ne pouvons donc affirmer que la majorité des allongements du *reparans* par rapport au *reparandum* sont dus à ce phénomène articulatoire.

Nous vérifions l'hypothèse selon laquelle la répétition simple de MO **fonctionne de manière distincte** de la répétition en combinaison avec d'autres marques (e.g. allongement du répétable, pause silencieuse ou *euh*) (inspiré de Candea (2000a:325-327)). Plus précisément, la répétition simple de MO serait due à un **simple retard articulatoire** de la production vocale (par rapport à la programmation d'unité) et s'approcherait d'un **cas léger de bégaiement** ou de gestion (à très court terme) du travail d'élaboration syntaxique. La production de MP impliquerait **une recherche de formulation** plus conséquente pour le locuteur et une réinitialisation syntaxique (Morel & Danon-Boileau 1998:85-87).

26. e.g. petite dans le segment « la petite fille » serait prononcé en une syllabe dans le *reparandum* ([la ptit fij]) mais en deux syllabes dans le *reparans* ([la pe.tit fij]).

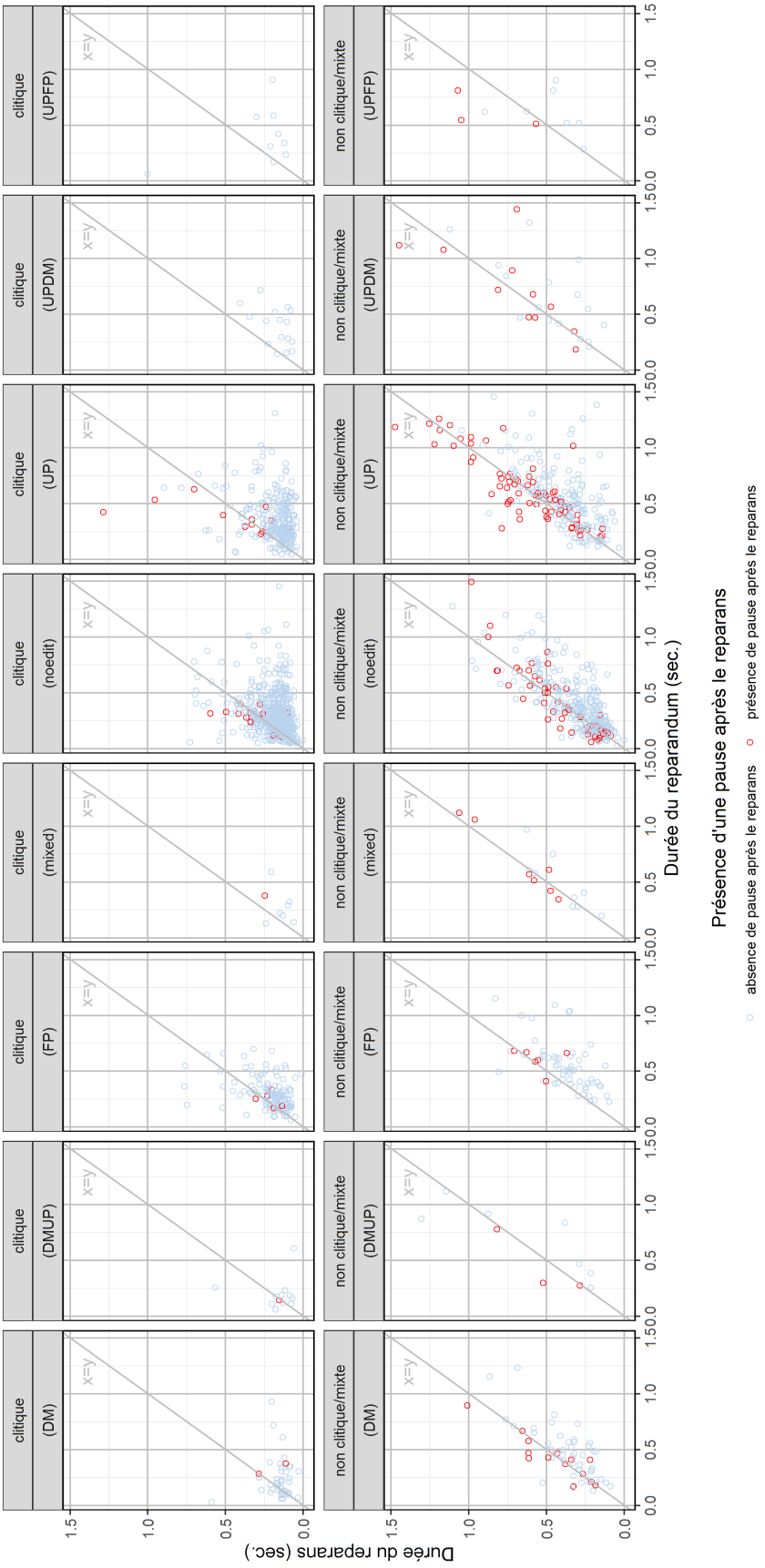


FIGURE 6.8 – Répartition des profils de répétitions en fonction du rapport (durée du *reparandum* / durée du *reparans*), en fonction du terme d'édition et de la nature du répétable.
Sous-ensemble de données : répétable d'un mot (e.g. il (il) il)).

Le croisement des données sur la durée relative avec la **nature grammaticale du répétable** permet d'étayer ces hypothèses. L'étiquette non clitique/mixte a été assignée aux répétitions impliquant un MP et éventuellement le(s) clitique(s) associé(s) au sein du groupe accentuel (e.g. « le patron * le patron »). La Figure 6.8 croise ainsi le ratio (`durée du` ***reparans***`/durée du` ***reparandum***) avec la **nature grammaticale du mot concerné par la répétition** et le **type de phase d'édition**. Les résultats sont évocateurs. La tendance des répétitions à être **suivies d'une pause après** le *reparans* (en rouge dans la Figure) ne s'observe pratiquement que pour les **répétitions de non-clitiques** (panel du bas). Ces répétitions sont majoritairement caractérisées pas un ***reparans*** plus long. Ces données étayent l'hypothèse selon laquelle les répétitions de MP nécessitent un temps de formulation plus important chez le locuteur, associé à un moment d'une pause.

En outre, on remarque une association entre ces deux variables (la nature grammaticale de la répétition, la présence d'une pause après le *reparans*) et le degré de préparation de la situation (test de Cochran-Mantel-Hazenszel, $M^2 = 100, df = 2, p < 0.001$). **Plus le discours est préparé, moins la tendance à allonger le *reparans* et à produire des répétitions de MP est attestée**. En terme d'attentes, on peut émettre l'hypothèse selon laquelle les répétitions de clitiques suivies d'une (dis)fluence sont davantage **porteuses de ruptures discursives**, car moins communes et/ou représentées à travers les corpus.

(b) Courbes intonatives des *reparandum* et *reparans*

A l'instar de Savova & Bachenko (2003) et de Shriberg (1995, 1999), nous nous focalisons sur la relation entre l'*onset* et l'*offset* du *reparandum* et du *reparans*[27]. Ces analyses se basent sur un sous-ensemble spécifique de Metacorpus : **1385 répétitions au *reparandum* monosyllabique**, soit directes (absence de terme d'édition), soit associées à un terme d'édition (une pause pleine ou une pause silencieuse) (i.e. <R0 R1>, <R0 UP R1>, <R0 FP R1>). Au reste, les MO comme les MP sont considérés, ainsi que tous les nombres de répétés. Sur base de la littérature, nous testons l'hypothèse suivante :

Hypothèse prosodique

- Comparées aux répétitions sans phase d'édition, les répétitions comportant une phase d'édition sont caractérisées (a) par une chute plus importante de la fréquence fondamentale, (b) par un parallélisme entre la courbe intonative du *reparandum* et du *reparans* (pente et amplitude similaires) et donc (c) par une équivalence de la hauteur des débuts de *reparandum* et de *reparans* (Shriberg 1999).

Par ailleurs, les observations de Shriberg (1999) sont corroborées par la **structure en parallélisme** observée par Cole et al. (2005)[28]. En assimilant les interprétations de la courbe de F0 continuellement descendante de ces répétitions (dites rétrospectives)

27. Ces mesures sont extraites automatiquement. Elles sont établies sur la F0 moyenne du noyau syllabique.

28. Voir chapitre 2 section 2.4.3.

aux interprétations faites des courbes descendantes des pauses pleines, les auteurs maintiennent que ces configurations ont une fonction de *filler*. Cependant, il n'existe pas de consensus dans la littérature. Plauche & Shriberg (1999) et Savova & Bachenko (2003) ont par exemple trouvé des répétitions qui fonctionnent comme parallélismes et d'autres comme contrastives. Ces hypothèses n'ont, à notre connaissance, jamais été testées sur des corpus francophones.

Nos données sont présentées séparément pour les femmes et les hommes vu la différence moyenne de hauteur de leur pitch. Elles permettent de formuler plusieurs observations (Figure 6.9) :

- En l'absence de terme d'édition, les répétitions exhibent **globalement des courbes plates** dans le *reparandum* et le *reparans*.
- En présence d'une pause silencieuse en phase d'édition, **le *reparans* est réalisé dans une plage plus basse lorsqu'il est suivi d'une pause silencieuse** et dans une plage de même hauteur ou plus haute lorsqu'il n'est pas suivi d'une pause.
- En présence d'une pause pleine en phase d'édition :
  - Pour les répétitions qui ne sont pas suivies par une pause, **le *reparans* est contrastif**. Il est réalisé dans **une plage intonative plus haute que celle du *reparandum***. La pause pleine est alors elle effectuée dans une plage plus basse.
  - Pour les répétitions suivies par une pause, les données sont plus difficilement interprétables en raison de l'absence de pitch. A fortiori, justement, **rien ne permet de soutenir des profils similaires avec les autres répétitions** ; la tendance irait davantage à l'absence de pitch, ou à la formulation d'un *reparans* pitch dans une plage de hauteur similaire ou basse.

Nos résultats confirment globalement ceux de la littérature anglophone : la présence de terme(s) d'édition a tendance à abaisser la plage intonative dans la phase d'édition ou sur le *reparans*. Cependant, les profils divergent selon la présence d'une pause pleine/-silencieuse en phase d'édition. La pause pleine en phase d'édition sans pause après le *reparans* favorise un contrast mélodique plus marqué : l'*interregnum* est en effet encore plus bas lorsque le *reparans* est suivi d'une pause (répétitions dites prospectives), ce qui est en accord avec les analyses de Shriberg (1995). Du reste, lors de l'absence de terme(s) d'édition, nos observations ne témoignent pas de variations significatives des courbes intonatives en se basant sur le pitch moyen. Ces conclusions vont dans le sens de Levelt (1983), selon qui un **parallélisme prosodique** entre les deux régions est relié au caractère « approprié » d'une séquence (et s'observe dans le cas de répétition et d'insertion) tandis qu'un **marquage contrastif est présent dans le cas de la correction d'erreur** (surtout pour les substitutions) ou dans notre cas, pour les répétitions associées et en cooccurrence avec d'autres éléments.

Toutefois, nous tenons à préciser que ce type d'analyse mérite qu'on accorde une attention plus particulière à la méthodologie de description de la hauteur de F0 moyenne au centre d'une syllabe, qui ne relève pas d'une même représentativité pour les syllabes au mouvement intonatif stable /vs/ les syllabes au mouvement intonatif dynamique. Par ailleurs, ces données, au contraire des analyses précédentes, ne présentent pas les possibles variations intersituationnelles ou interlocuteurs.

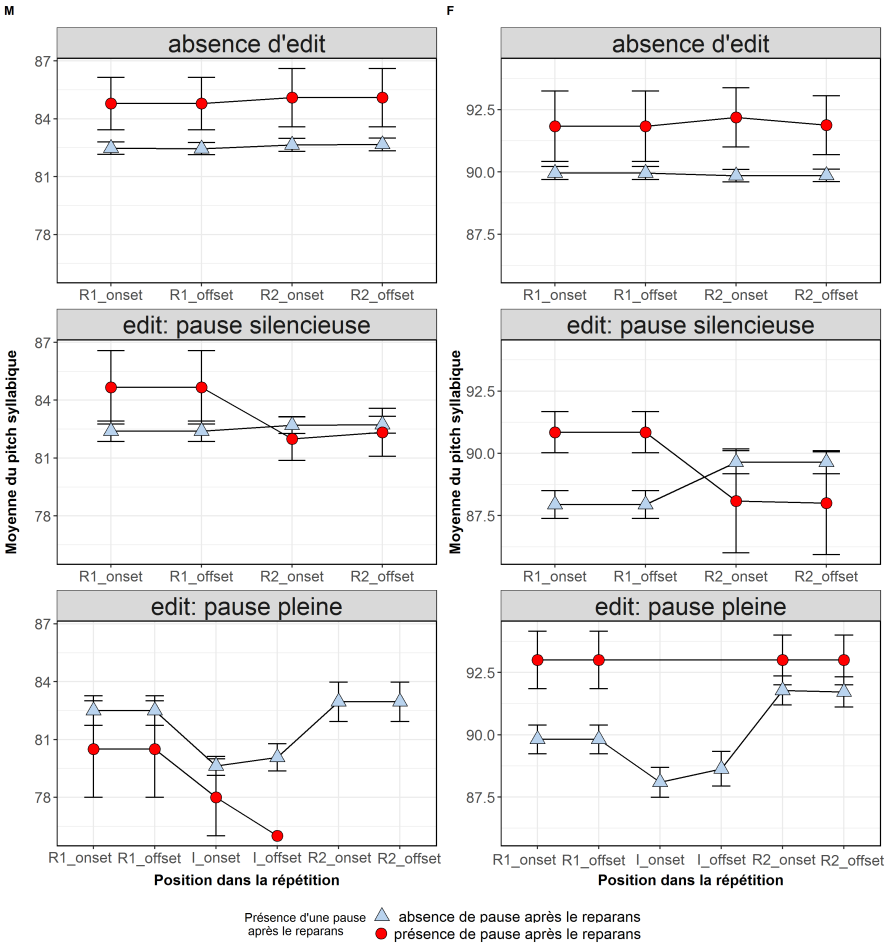


FIGURE 6.9 – Hauteur moyenne de l'*onset* et de l'*offset* du *reparandum* et *reparans* des répétitions pour les hommes (H *à gauche*) et femmes (F *à droite*).
Note : La hauteur moyenne de la F0 est mesurée sur la première et la dernière syllabe du *reparandum* (R1_onset et R1_offset) ainsi que sur la première et la dernière syllabe du *reparans* (R2_onset et R2_offset). De même, au début et en fin de phase d'édition (I_onset, I_offset).

1.6 Conclusions sur l'évaluation des répétitions produites

Dans cette section, nous avons proposé une analyse des structures de répétitions en fonction de leurs variations situationnelles, de leurs cooccurrences, de leurs localisations syntaxiques et interactionnelles et de leurs formes acoustiques. Nous concluons sur la production des répétitions en mettant en exergue les relations entre les principaux résultats.

A travers le corpus, les répétitions s'actualisent en moyenne **tous les 18.9 mots sur 1000 et au total 68.9 mots sur 1000 sont impliqués dans une structure de répétition, ce qui représente en moyenne 5.5% du temps d'articulation et 5% des unités syntaxiques**. Nous n'avons pas observé d'opposition possible entre les situations en fonction de la proportion de la durée cumulée du temps d'articulation de répétés. Par contre, à proportion de (dis)fluence égale, certaines situations privilégient l'actualisation de répétitions plutôt que d'autres (dis)fluences : c'est le cas des situations spontanées et (semi-)interactives.

Dans nos corpus, près de 60% des répétitions sont monosyllabiques et/ou s'actualisent sur les clitiques et/ou relèvent de répétitions simples. S'il n'existe pas d'influence du degré d'orientation médiatique sur la quantité de répétitions produites, **la structure interne des répétitions dans les situations orientées médiatiquement varie**. Ainsi, il y a plus de répétitions simples en situation à orientation médiatique. En outre, la structure des répétitions varie en fonction du nombre de locuteurs : la répétition double est plus fréquente en dialogue.

Par ailleurs, **notre recherche remet en cause la distinction de trois modalités de longueur du répétable et du nombre de répétés**. Nos résultats d'analyses en composantes multiples ont effectivement mis en évidence que peu de distinctions de structure interne existent entre les répétitions caractérisées par un nombre de répétés double /vs/ multiple, ou entre les structures caractérisées par une longueur de répétable de 2 mots /vs/ $>$ à 2 mots. Elle encourage les recherches suivantes à regrouper ces modalités.

De manière surprenante, nous n'avons **pas trouvé d'association entre la longueur du répétable en nombre de mots et le degré de préparation ou l'interactivité**. Néanmoins, on observe bien que le nombre de syllabes du répété, lui, est bien plus élevé en situations de parole préparées qu'en spontanées. En d'autres termes, si l'on effectue relativement plus de répétitions dans les situations préparées et (semi-)interactives, on ne répète pas pour autant des répétables de longueur de mots inférieure, mais les répétitions s'actualisent bien sur des mots constitués de moins de syllabes.

Près de la moitié des répétitions dans le corpus sont actualisées sans terme d'édition. Les corpus témoignent d'une grande variabilité des combinaisons de termes d'édition au sein de la phase d'édition (+ de 78 combinaisons), et celles-ci sont fortement associées au dégré de préparation. Les marques privilégiées des phases d'édition sont

majoritairement la pause silencieuse, suivie des marqueurs discursifs et des troncations. Moins de 10% des répétitions en phase d'édition sont ainsi concernées par une (dis)fluence ou une pause pleine en phase d'édition.

Plusieurs éléments convergent en faveur de **la reconnaissance de la répétition comme ligateur du flux discursif**. Au niveau syntagmatique, les mots lexicaux exempts de (dis)fluences sont la marque majoritaire précédant et suivant la répétition. En outre, la répétition est dans moins de 5% des cas suivie d'une pause marquant l'alternance de tour. La seule combinaison favorisant l'alternance de tour est un cluster spécifique où la répétition doit être elle-même précédée par une répétition, inclure une phase d'édition composée et être composé d'un répétable de longueur double ou multiple. Au reste, la répétition est initiatrice de tour même lorsqu'elle est précédée d'un marqueur discursif et/ou lorsqu'elle inclut un marqueur discursif ou une troncation en phase d'édition. Par ailleurs, les répétitions associées à des marques en phases d'édition sont également liées au contenu qui les précède et leur succède : malgré une phase d'édition complexe, elles sont majoritairement actualisées en continuation directe avec la suite de l'énoncé (i.e. répétitions rétrospectives).

Les répétitions ne brisent ni le déroulé syntaxique ni le déroulé interactionnel. Les débuts de répétitions sont en général inter-séquentiels et inter-rectionnels, seules 6 répétitions sur 1000 s'initialisent au sein même d'un groupe accentuel. Ainsi, la répétition initie les unités syntaxiques et les tours de parole. Les fins de répétitions ne sont par ailleurs pas associées à des frontières syntaxiques majeures : seules 6 répétitions sur 1000 s'actualisent dans une position inter-rectionnelle ou inter-séquentielle. Au contraire, les fins de répétitions s'actualisent au sein de groupes accentuels ; elles maintiennent la cohésion discursive et sont très rarement marques de fin de tour de parole.

Au niveau acoustique, **les répétitions sans terme d'édition et suivies par une pause sont plus longues que les autres**. On soutient qu'elles ont une fonction de maintien de la parole lors de l'hésitation. Elles sont d'ailleurs plus caractéristiques d'une situation de parole spontanée. A l'appui de ce point de vue, les répétitions de mots pleins (non clitiques) sont plus suivies de temps de pause, c'est-à-dire d'un temps de reformulation conséquent. En résumé, plus le degré de préparation est élevé, moins il y a de répétitions de MP et moins il y a d'allongement du *reparans* par rapport au *reparandum*.

Enfin, **l'analyse de hauteur mélodique** relative des segments de répétitions monosyllabiques (d'un ou de plusieurs répétés) **a révélé que les *reparans* et les *reparandum* ont globalement un contour mélodique plat**. Les profils de hauteur se distinguent en fonction du type de phase d'édition et de l'élément suivant la répétition. Le *reparans*, surtout lorsqu'il est suivi d'une pause silencieuse, est dans une plage plus haute que le *reparandum*. De plus, le *reparans* des répétitions impliquant une pause pleine en phase d'édition et d'une continuation directe a une hauteur mélodique contrastive bien plus

haute que le *reparandum*. Au contraire, pour les répétitions avec une pause pleine en phase d'édition qui sont suivies par une pause après le *reparans*, celui-ci a tendance à s'actualiser dans une plage de même hauteur ou plus basse que le *reparandum*.

Dans la suite de ce chapitre, nous évaluons les effets de ces différentes structures de répétitions (directes, associées à une pause pleine ou à une pause silencieuse) sur leur perception et leur évaluation par les auditeurs. Une première expérience vise à tester la capacité des répétitions à augmenter l'attention des locuteurs : à structure prosodique et syntaxique similaire, elle teste l'interaction entre la variation des structures des répétitions à l'identique et les temps de réaction des locuteurs (section 2.1). Une seconde expérience ancre les répétitions de manière situationnelle ; elle invite les participants à l'évaluation holistique et componentielle *offline* des répétitions à l'identique en contexte et hors contexte (section 2.2).

2 Perception des répétitions avec et sans marqueur d'édition

Les chapitres précédents et la littérature antérieure témoignent du potentiel structurant des pauses pleines et silencieuses. Elles ont en effet la capacité de déclencher chez l'auditeur des attentes quant au niveau de complexité d'un événement à venir et/ou à sa pertinence [29]. Cependant, rares sont les travaux questionnant l'évaluation de la fonction fluente ou disfluente de **marques structurées plus complexes**, telles que les répétitions [30]. Les résultats sont encore plus maigres si l'on s'attarde sur les recherches relatives à la parole non pathologique de locuteurs francophones natifs et/ou de résultats issus de données expérimentales.

Néanmoins, la linguistique reconnaît que « les (dis)fluences » sont potentiellement porteuses d'informations métalinguistiques pour l'auditeur (e.g. Brennan & William 1995). Toutefois, les recherches neurolinguistiques attestent que « **toutes les disfluences ne sont pas égales** » : les différentes (dis)fluences n'affecteraient pas les processus cognitifs sous-jacents à la perception de la parole de manière identique – en termes de prédiction, d'attention, de compréhension ou de mémorisation (MacGregor, Corley & Donaldson 2009).

Cette section présente les résultats de deux protocoles d'expériences portant sur la perception et l'évaluation de répétitions à l'identique par des auditeurs francophones natifs. Ces contributions s'attachent à une meilleure compréhension de la **fluence perceptuelle** par les auditeurs francophones, c'est-à-dire des « inférences que les auditeurs font à propos de la fluence cognitive en se basant sur leur perception de la fluence phrastique » (Segalowitz 2010:165) [31]. Les deux protocoles expérimentaux lient ainsi la ***fluence phrastique*** à la ***fluence perceptuelle*** en questionnant l'impact de la présence qu'une répétition, notamment sur l'**aisance de traitement** d'un énoncé par l'auditeur et l'évaluation contextuelle qu'il en fait.

La première expérience restreint cette évaluation au **niveau cognitif individuel** hors contexte situationnel (i.e. le **contexte interne**). Elle mesure le temps de latence dans le cadre d'une tâche de sélection d'image et évalue ainsi la **fluence cognitive linguistique**. Nous y émettons spécifiquement des hypothèses sur les **processus d'attention et de prédiction** (section 2.1). Nous effectuons une **évaluation décontextualisée des structures des répétitions sur l'attention**. La présente étude lie la fluence phrastique et la fluence perceptuelle par la mesure des temps de réaction, i.e. la fluence cognitive associée à une tâche de décision incluant des phrases porteuses de répétitions lexicales

29. Voir Arnold, Fagnano & Tanenhaus (2003) et Clark & Tree (2002) et les chapitres 3 et 4.

30. Citons cependant quelques recherches (non francophones) sur les effets des répétitions : Clark & Bernicot (2008), Clark & Wasow (1998), Fox Tree (1995), Moniz, Trancoso & Mata (2009) et Watanabe, Hirose, Den & Minematsu (2008).

31. Texte original : "the inferences hearers make about speakers' cognitive fluency based on their perceptions of the utterance fluency." (Segalowitz 2010:165). Notre traduction.

et de termes cooccurrents en phase d'édition. Cette étude reproduit un design expérimental réalisé par Watanabe, Hirose, Den & Minematsu (2008) évaluant l'impact d'une information métalinguistique que la répétition peut constituer pour les auditeurs et participer ainsi à la fluence perceptive.

Le second design questionne l'impact des **attentes des auditeurs** vis-à-vis d'une situation sur leur perception et l'évaluation d'un énoncé porteur d'une répétition. Ce design inclut le contexte situationnel (i.e. **contexte externe**) dans l'évaluation individuelle. Il amorce ainsi une réflexion sur l'évaluation de la **fluence linguistique situationnelle** (section 2.2). Nous testons des énoncés comprenant différents niveaux de *fluence phrastique* (énoncé sans répétitions, avec répétition contiguë, répétition avec une pause pleine en phase d'édition et répétition avec pause silencieuse en phase d'édition). La première originalité de cette expérience est de proposer l'écoute de stimuli naturels – existant dans les corpus oraux.

2.1 Expérience 1 : fluence cognitive des répétitions

Notre démarche s'attache à la **fonction positive** des (dis)fluences et s'accorde avec le fait qu'elle puisse refléter une gestion cognitive et/ou interactionnelle de la production langagière (Shriberg 2005:1783). Cette perspective suggère que, entre autres choses, les (dis)fluences peuvent servir d'indice et/ou de **signal d'un besoin de traitement cognitivement demandeur**. Cela suggère que la présence de (dis)fluences pourrait produire des « résultats supérieurs » (e.g. augmentation de l'attention, rapidité de traitement, meilleure compréhension, etc.) (Alter 2013:440 ; Arnold, Fagnano & Tanenhaus 2003:32 ; Clark & Tree 2002:107).

Cette étude reproduit un design expérimental réalisé par Watanabe, Hirose, Den & Minematsu (2008) évaluant l'impact d'une **information métalinguistique**, telle que la répétition, peut constituer pour les auditeurs, participant ainsi à la fluence perceptive. La fluence phrastique est **mesurée acoustiquement** à partir des stimuli audio. Le design expérimental la lie à la fluence cognitive (ou « l'efficacité de traitement de la fluence cognitive responsable pour la production des énoncés ») (Segalowitz 2010:165, notre traduction) [32] en mesurant les **temps de latence** relatifs à une **tâche de décision** à l'écoute d'énoncés incluant une (dis)fluence (i.e. la répétition lexicale et son éventuelle cooccurrence avec des termes d'édition). Dans le cadre de cette expérience, la mesure du temps de réaction (temps d'exhaustion de la capacité libre) (log(RT)) est utilisée comme une définition concrète de la « **(dis)fluence de récupération** », c'est à dire la rapidité avec laquelle on accède à un objet.

32. Texte original : "efficiency of operation of the underlying processes responsible for the production of utterances." (Segalowitz 2010:165).

Ce travail lie ainsi la **fluence phrastique** et **la fluence perceptuelle** par la mesure des temps de réaction [33]. Il évalue par cette mesure la **fluence cognitive** associée à une tâche de décision incluant des **énoncés porteurs de répétitions de mots grammaticaux** et de termes cooccurrents en phase d'édition (i.e. pauses vides et pleines).

2.1.1 Matériel et méthode

Design expérimental. Le design invite chaque auditeur à écouter **96 énoncés** décrivant des formes géométriques simples et composées. Chaque énoncé n'est écouté qu'une fois. Chaque énoncé est une requête de type : « Allez chercher un papier de couleur `(x)` et de forme `(y)` dans une pièce `(z)` », permettant à l'interlocuteur de pouvoir constituer une guirlande. Ainsi, chaque énoncé décrit une couleur (i.e. `rouge, bleu, orange`), une forme (i.e. `circulaire, triangulaire`) et une pièce dans laquelle se trouve cette forme (i.e. dans le bureau bureau, dans la chambre, dans le salon). Les variables de couleurs et de pièces agissent comme **distracteurs**, les paires de formes étant toujours de la même couleur. Les énoncés se distinguent donc en fonction du **degré de complexité de la forme décrite** et du **type de (dis)fluence prononcé**.

– **Facteur complexité.** Le degré de complexité de la forme oppose une condition complexe (`CC` – *avec les flèches*) à une condition simple (`CS` – ∅, *sans flèches*). A cette complexité visuelle de forme est ainsi associée une « **complexité linguistique** » de surface : on lie ici un nombre supérieur de mots à un degré de prototypicalité inférieure. La Figure 6.10 illustre les paires de stimuli visuels présentés aux participants.

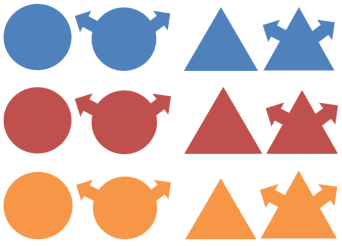


FIGURE 6.10 – Exemple de six paires de stimuli.

– **Facteur fluence.** Les énoncés se distinguent également en fonction du degré de (dis)-fluence phrastique. Le lexème de forme y est toujours précédé soit (1) d'une répétition contiguë (`REP-C`), soit (2) d'une répétition avec une pause pleine en phase d'édition (`REP-FP`), soit (3) d'une répétition avec une pause vide de même durée en phase d'édition (`REP-UP`), ou (4) d'aucune répétition (`No-REP`).

Instructions. A chaque écoute unique d'énoncé, le plus rapidement possible, le participant doit appuyer sur une touche de gauche ou de droite du clavier, correspondant à la forme demandée présentée à l'écran (voir exemple 22).

33. Les résultats de cette étude sont présentés dans GROSMAN (2015), ils font partie des actes de conférences de DiSS 2015 – *The 7th workshop on Disfluency in Spontaenous Speech* (8-9 Août 2015, Edinburgh, Scotland, UK).

Exemple (22)

`[Forme-simple + REP-UP]`

a. « *Apporte-moi le euh le petit papier carré rouge qui est dans la chambre* »

`[Forme-complexe + REP-C]`

b. « *Apporte-moi le le petit papier orange circulaire avec des flèches qui est dans le bureau* »

Enregistrement des stimuli. Les stimuli originaux enregistrés constituent 24 énoncés différents :

```
3 couleurs * 2 formes * 2 niveaux de complexité * 2 pièces = 24 stimuli
```

Les énoncés d'origine correspondent tous à la condition `REP-FP` : un énoncé contenant une répétition avec une pause pleine en phase d'édition. Ils ont été présentés à une locutrice francophone native qui a eu pour instruction de les prononcer, sans nécessairement lire la liste afin qu'ils paraissent provenir de parole relativement naturelle [34].

Sélection des stimuli. Chaque énoncé a été prononcé deux fois afin de **sélectionner ceux représentant le moins de variation prosodique entre eux**. Pour ce faire, tous les énoncés ont été transcrits phonétiquement sous Praat (Boersma & Weenink 2015) en utilisant EasyAlign (Goldman 2011). L'alignement et la phonétisation ont été vérifiés deux fois en respectant strictement les règles de segmentation syllabique de Bigi, Meunier, Bertrand & Nesterenko (2010). Les **24 stimuli présentant le moins de variation prosodique entre eux** ont été sélectionnés parmi les 48 énoncés produits par la locutrice [35].

Vérification des stimuli. Des *tests-t* d'indépendance des moyennes pour les deux niveaux de complexité (`simple` /vs/ `complexe`) ont été effectués séparément pour chaque paramètre prosodique. Aucun d'eux ne s'est avéré significativement différent entre les conditions, à l'exception de la durée absolue du temps d'enregistrement entre les deux conditions de complexité. En moyenne, la durée des stimuli complexes était de 600 ms supérieure à celle des stimuli non complexes. Les pauses pleines dans la condition simple n'étaient pas plus longues que celles dans la condition complexe ($\mu = 360$ms., $\sigma = 40$ms., $dur.rel = 1.40$). Enfin, les schwas avant le *reparans* ont été prononcés avec une durée homogène à travers les conditions de complexité ($\mu = 350$ms, $\sigma = 38$ ms), en accord avec la tendance observée sur la première partie des répétitions et des segments précédant une pause pleine [36].

34. L'enregistrement au format WAV a été effectué en utilisant un enregistreur `Roland R-09HR` (24 bit, 96kHz) et remasterisé au format MP3 (16 bit, 44kHz).

35. Il s'agit donc de mesurer la distance relative la plus petite entre chaque énoncé. Les mesures de comparaison sont : la moyenne et la médiane du débit de parole, de l'étendue tonale (en demi-ton), de la fréquence fondamentale minimum et maximum, leur moyenne et médiane de fréquence fondamentale, la proportion moyenne de glissando, de syllabes montantes et descendantes de chaque échantillon de parole, ainsi qu'en fonction de la durée du schwa du *reparandum* (le premier « le » dans « *Apporte-moi le euh le papier bleu triangulaire qui est dans le bureau* »), et la durée relative de la pause pleine. Les phrases avec les distances relatives les plus faibles ont été sélectionnées.

36. Voir notamment chapitre 2 section 3 et section 1 de ce chapitre.

Génération des niveaux de facteur fluence. Les 24 stimuli originaux sélectionnés ont été utilisés pour générer les **trois conditions de fluence supplémentaires** en supprimant le segment audio approprié sous Audacity et ProTools. Il s'agit en cascade de remplacer le *euh* par un silence (`REP-UP`), d'éliminer le silence (`REP-C`) et d'éliminer le *reparandum* (`No-REP`) (Figure 6.12). L'impression de naturel des stimuli a été vérifiée et validée par 4 locuteurs francophones natifs.

Participants. Tous les participants à l'expérience étaient des locuteurs francophones natifs volontaires et non rémunérés pour leur contribution. Ils ont été contactés en ligne (e.g. cercles académiques, mailing list de participants, médias sociaux, etc.). Avant l'expérience, tous les participants ont signé un formulaire de consentement. Après, il leur était demandé de faire mention de toute déficience auditive, entraînement musical, et si un dérangement éventuel avait eu lieu pendant le test. In fine, sur les 80 participants, 18 ont été retirés de l'analyse : 5 pour des défauts d'audition déclarés, 13 pour des perturbations.

Les **62 participants** sont âgés de 15 à 58 ans ($\mu = 28, \sigma = 11.04$). Parmi eux, 35 sont originaires de Belgique et 21 de France. Les autres participants venaient de Suisse, du Canada, du Maroc et d'Israël. Les participants étaient autorisés à utiliser des enceintes ou des écouteurs (30 et 32 participants respectivement). La majorité d'entre eux n'était pas musiciens (40) ou pratiquait la musique comme hobby (20), seuls 2 des participants étaient des musiciens professionnels.

Questionnaire pré-test. Avant l'expérience, les participants ont également complété un **test de réflexion cognitive** (*"Cognitive Reflection Test"*, désormais CRT) (Frederick 2005). Les résultats du CRT sont « une simple mesure d'un type d'abilité cognitive » fortement corrélés **aux capacités cognitives analytiques** (Frederick 2005:26, notre traduction)[37]. Toutes les questions demandent de la part des participants une **mobilisation d'un système de traitement profond** (« système 2 » dans la terminologie de Stanovich & West (2000)), tandis qu'une réponse intuitive et impulsive (ne mobilisant que le « système 1 ») résulterait en une réponse erronée. Les trois questions suivantes ont été posées aux participants. En vert, la réponse valide.

1. Une batte et une balle coûtent la somme de 1.10 euros. La batte coûte 1 euro de plus que la balle. Combien la balle coûte-t-elle ?

   Réponse : (a) **0.5 centimes** (b) 10 centimes (c) 1 euro

2. Si 5 machines prennent 5 minutes pour fabriquer 5 objets, combien cela prendrait-il de temps à 100 machines de fabriquer 100 objets ?

   Réponse : (a) 100 minutes (b) 500 minutes (c) **5 minutes**

3. Dans un lac, il y a un lit de nénuphars. Chaque jour, ce lit double de taille. S'il faut 48 jours au lit de nénuphars pour recouvrir entièrement le lac, combien de temps lui faut-il pour recouvrir la moitié du lac ?

   Réponse : (a) 24 jours (b) **47 jours** (c) 23 jours

37. Texte original : "*a simple measure of one type of cognitive ability*" (Frederick 2005).

1. Testez le volume de votre casque ou de vos enceintes

2. Lisez les consignes

Dans ce test, vous allez entendre un son par page. Il vous est demandé de choisir à chaque fois entre deux images. Ces images représentent une forme et une couleur.

Choisissez **le plus rapidement possible** celle demandée dans l'enregistrement sonore. Vous devez presser la touche **A** ou **Q** pour choisir l'image de gauche ou la touche **P** pour l'image de droite.

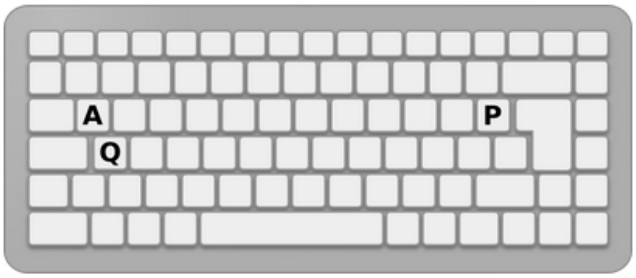


FIGURE 6.11 – Illustration accompagnant les instructions de l'expérience 1.

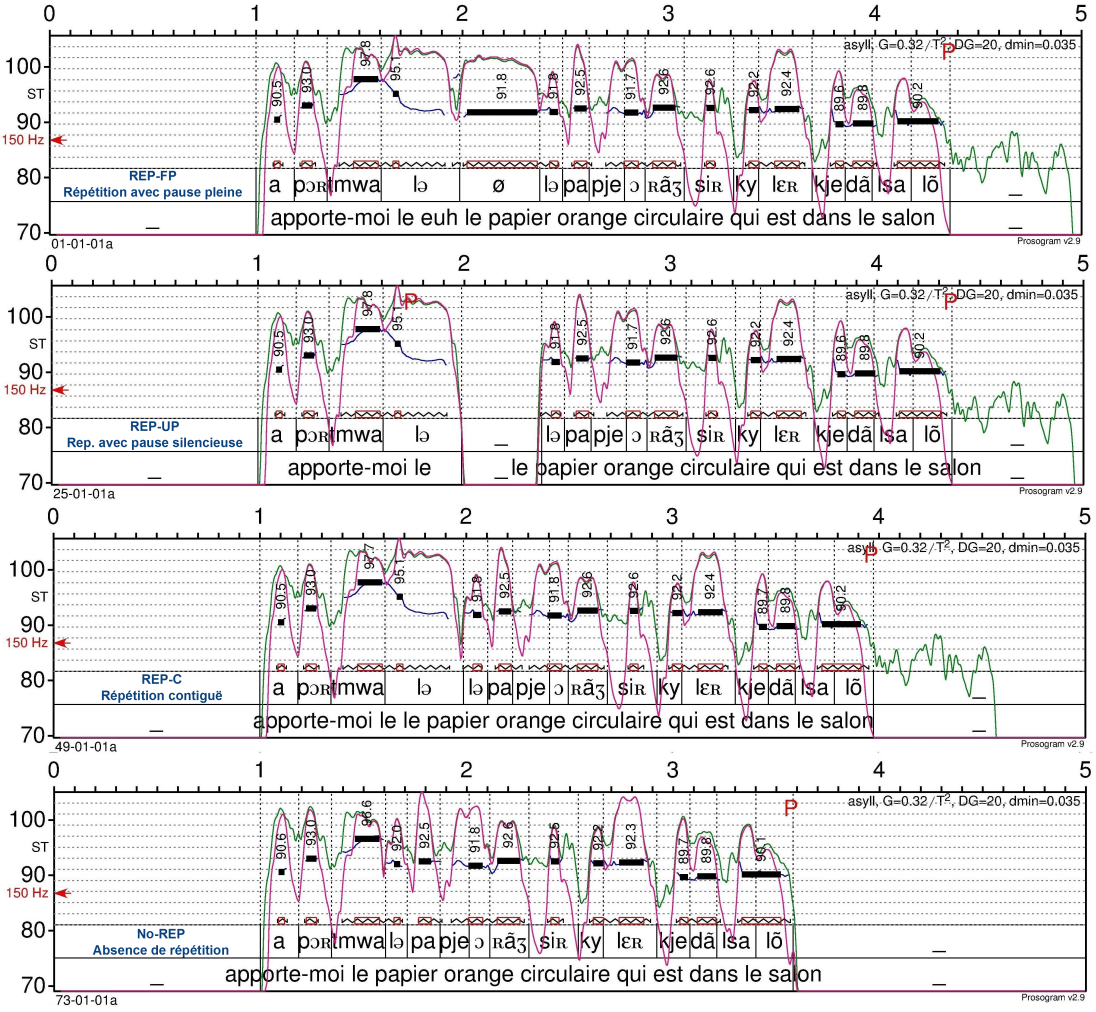


FIGURE 6.12 – Prosogramme de 4 stimuli correspondant au même énoncé.

Déroulement. Le design n'autorisait pas un énoncé d'une condition A à être suivi du même énoncé dans sa condition B. Hormis cette contrainte, l'**ordre d'apparition des stimuli est aléatoire** pour chaque participant. Les stimuli étaient joués jusqu'à la fin quel que soit le moment où le participant accomplissait la tâche. L'intervalle inter-stimuli était de 3 secondes. Si un participant a pressé une touche plus d'une fois, **seule la première pression est considérée**. L'expérience durait de 15 à 20 minutes incluant une courte session d'entraînement et 2 pauses de 30 secondes aux premier et second tiers.

Mesure & temps de réaction. Les temps de réaction (RT) ont été développés comme une mesure objective indirecte dans le cadre de la **théorie de charge cognitive** (*"Cognitive Load Theory"*) (Brunken, Plass & Leutner 2003:57). Ils peuvent être utilisés comme une définition concrète de la « (dis)fluence de récupération », la rapidité avec laquelle on accède à un objet. Des études antérieures ont montré le **lien direct entre la complexité d'une information et la charge cognitive** : cette dernière est le résultat intrinsèque de la charge intrinsèque et extrinsèque du matériel (Brunken, Plass & Leutner 2003:56). Les temps de réaction mesurent donc la **charge mentale**, c'est-à-dire « les aspects de la charge cognitive qui proviennent de l'interaction entre une tâche et les caractéristiques du sujet »(Paas, Tuovinen, Tabbers & Van Gerven 2003:64, notre traduction)[38].

Au niveau méthodologique, les temps de réaction ont été **automatiquement mesurés** depuis le début de chaque stimulus. Pour chacun, le point de mesure du début du **premier mot permettant la désambiguïsation entre les deux choix de formes** a été extrait et reporté sur des tableurs Praat. Dans les conditions `simple` et `complexe`, les points d'extraction étaient respectivement les phones [k] (de « **qui** est dans le bureau ») et [a] (dans « **avec** les flèches »). Les temps utilisés pour la présentation des résultats ont été extraits en soustrayant le point de mesure du mot de désambiguïsation au point de mesure de pression de la touche (réponse). La **médiane des temps de latence de chaque condition** a été extraite pour chaque participant et les mesures ont été comparées à travers les conditions.

38. Texte original : "the aspect of the cognitive load that originates from the interaction between task and subject characteristics." (Paas, Tuovinen, Tabbers & Van Gerven 2003:64).

2.1.2 Hypothèses

En accord avec les hypothèses d'Arnold, Fagnano & Tanenhaus (2003), le design teste le fait que les répétitions soient utilisées comme des outils pour **capturer l'attention** et/ou pour **faciliter l'amorçage conceptuel d'entités complexes**. Selon celles-ci l'amorçage conceptuel est facilité par le traitement perceptif. Par conséquent, la présence de certaines (dis)fluences rendrait des segments de discours **plus saillants perceptivement** et donc, dans un sens, plus aisé et rapide à traiter. Les hypothèses formulées posent ainsi que :

1. **Hypothèse de complexité** : Les RTs à des instructions requérant aux auditeurs de sélectionner des formes simples seront plus courts que ceux exigeant de choisir des formes composées (e.g. une sphère non agrémentée de flèche /vs/ une sphère avec des flèches).

   Cette hypothèse part de la prémisse que l'exemplaire le plus simple d'une catégorie visée, tel qu'un stimulus visuel non composé est de manière inhérente plus « fluente », dans le sens de **plus aisé au traitement et traité plus rapidement** (Alter & Oppenheimer 2009).

2. **Hypothèse de présence d'une marque**. Les RTs à des instructions requérant aux auditeurs de sélectionner des formes composées seront **plus courts** si les instructions de sélection sont **précédées d'une répétition** qu'en l'absence de répétition.

   Les répétitions serviraient par là même d'indices précieux : elles enclencheraient des hypothèses sur la **difficulté de planification** du contenu à venir mais offriraient aussi un **temps supplémentaire** pour l'auditeur lui permettant d'émettre des prédictions sur le contenu du discours en préparation (Watanabe, Hirose, Den & Minematsu 2008).

3. **Hypothèse de structure de la marque**. Les répétitions contenant un terme d'édition donneront lieu à des **réponses plus rapides** que les répétitions simples directes (i.e. contiguës), quelle que soit la condition de forme. Cette hypothèse d'un **résultat additif** des différentes structures de la répétition est formulée en vertu de la **saillance relative des pauses pleines** dans le flux de la parole.

4. **Hypothèse de réflexion cognitive**. Les participants s'appuyant davantage sur un **traitement intuitif**, c'est-à-dire ceux **répondant moins correctement** au CRT, auront une tendance à s'appuyer davantage sur la présence d'une disfluence que les participants ayant des résultats favorables au CRT.

2.1.3 Résultats

Taux et types d'erreurs. Globalement, le taux moyen de réponses correctes est de 98.147% ($\sigma = 1.73$). Ce résultat est comparable avec celui de Watanabe, Hirose, Den & Minematsu (2008). Les réponses précoces (i.e. réponses anticipées réalisées avant le son désambiguïsant) ont été écartées et traitées comme des erreurs. Les taux d'erreurs étaient significativement différents entre les conditions simples et composées, confortant l'idée que les **formes composées** étaient bien d'une **complexité supérieure** ($t(4, 92), p < 0.005$).

Temps de réaction globaux. Les participants ont pressé une touche entre 0.55 et 0.70 sec. **après l'onset de la syllabe désambiguïsante** (Figure 6.13). A l'instar de Watanabe, Hirose, Den & Minematsu (2008), une analyse de mesures répétées à deux niveaux (ANOVA) a révélé un **effet principal du facteur de complexité** sur les temps de réaction ($F(1, 61) = 26.76, p < 0.001$). Comme la première hypothèse le prédisait, les temps de latence sont en effet plus courts pour les items simples que pour les items composés. De plus, contrairement à l'étude japonaise homologue, nous avons trouvé un **effet principal du facteur de fluence** lors des comparaisons post-hoc des *tests-t* avec une correction de Bonferrroni[39] indiquant que **les scores moyens à l'intérieur des conditions de fluences sont significativement différents** ($F(3, 183) = 3.87, p < 0.02$).

De manière générale, peu de variance est attestée entre les RTs des trois types d'items comportant un marqueur (`REP-C`, `REP-UP`, `REP-FP`). Cependant, une analyse *post-hoc* montre que ces **différences de niveaux de fluence** restent significatives **en condition de forme simple** ($F(3, 58) = 2.79, p < 0.05$), mais non dans le cas d'une forme composée ($F(3, 58) = 1.9, p = 0.131$). Ces résultats sont en contraste avec ceux de Watanabe, Hirose, Den & Minematsu (2008) qui ont reporté des significativités inverses (i.e. validant l'hypothèse pour les formes composées mais non simples).

En conséquence, la seconde hypothèse selon laquelle les temps de réaction des formes composées seraient plus courts lorsque l'item inclut une répétition que s'il en est exempt, (et selon laquelle les RTs des formes simples pourraient même être plus longs lorsqu'il y a une répétition que quand il n'y en a pas) **ne peut pas être supportée dans le cadre de cette expérience**.

Les résultats obtenus témoignent également de l'**absence d'interaction** entre les facteurs de complexité et de fluence. Nous pouvons ainsi seulement conclure que, **peu importe la condition de complexité,** les réponses des **items portant des marqueurs de (dis)fluence témoignent de RTs plus courts que ceux sans marqueurs de (dis)fluence**.

Enfin, la troisième hypothèse est **partiellement confirmée**. Les items comportant une répétition associée à une pause pleine en phase d'édition présentent en effet des **temps de réaction plus courts** que leur version homologue avec pause silencieuse ($t(60) = 4.018$,

39. Méthode de correction statistique du seuil de significativité lors de comparaisons multiples.

$p < 0.001$). Ce résultat laisse à penser que la cooccurrence de marques à un **effet additif**. Néanmoins, la différence est minime entre les RTs de la condition de la répétition avec pause pleine vs. la répétition contiguë. Ainsi les temps de réaction plus courts des `REP-FP` ne pourraient être dûs qu'à la présence d'une répétition et non à la présence d'une pause pleine.

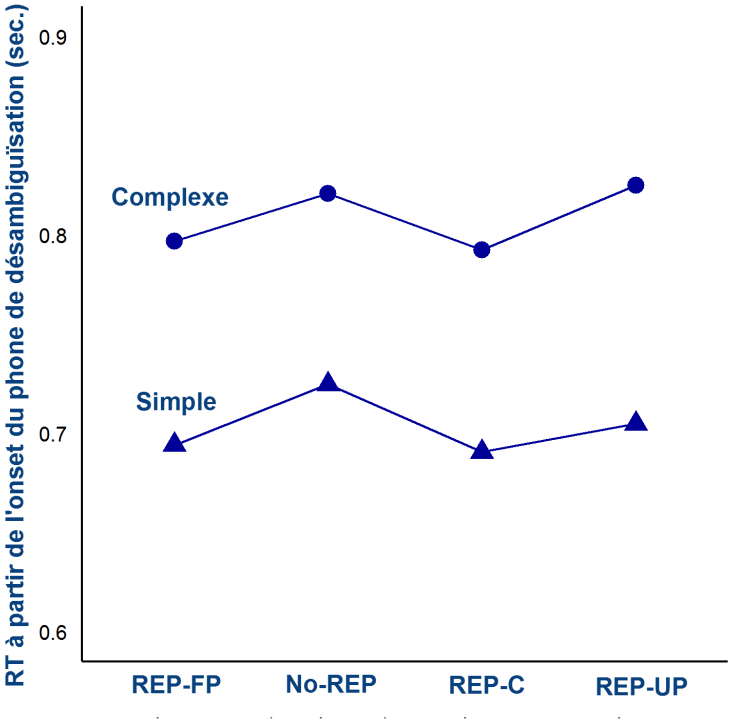


FIGURE 6.13 – Temps de réaction par condition de fluence.

Modélisation linéaire à effets mixtes. Le principe derrière une modélisation linéaire mixte est de tester la significativité statistique de contrastes expérimentaux entre les participants et les items, en utilisant des méthodes basées sur les maxima de log-likelihood (LogLik). Les durées sont modélisées en prenant en compte les éventuelles différences individuelles et en posant un *random intercept* singulier pour chaque participant et item, tandis que les conditions de complexité et de fluence sont considérées comme des facteurs à effets fixes (Baayen, Davidson & Bates 2008:264 ; Baayen & Milin 2010) [40].

Concrètement, en vertu du design de mesures répétées, tous les cas ne sont pas indépendants. A ce titre, 46,5% de la variance des RTs peut s'expliquer par les variations interparticipants, qui se différencient fortement en termes de durée des RTs (99% *Intraclass Correlation Coefficient*) : certains ont tendance à être lents et d'autres rapides. De plus, diverses stratégies non contrôlées peuvent être adoptées par les participants : certains attendent la fin du syntagme nominal, tandis que d'autres n'ont besoin que du premier phonème désambiguïsé pour réagir. Enfin, comme nous ne pouvons certifier avoir testé tous les éléments possibles, leur gamme de variabilité ainsi que leur position, nous sommes enclins à considérer l'objet, le participant et la position d'un item dans l'expérience (de 1 à 96) comme des effets aléatoires.

Un des objectifs majeurs de cette étude est d'évaluer si le degré de fluence et de complexité linguistique influence les temps de latence. Nous avons avant tout testé si la prise en compte d'effets aléatoires dûs à l'énoncé (`item`) est requise (i.e. à intégrer dans le modèle en vertu des calculs des ratios de test de likelihood).

A l'observation des effets aléatoires [41], on remarque qu'une part significative de la variance est motivée par la variation interparticipants. Le modèle permet d'expliquer plus de 57% de la variance des temps de réaction : 57% par les effets fixes et aléatoires ($pseudo-R^2 conditionnel = 0.57$) et 16% par les effets fixes seuls ($marginal Pseudo-R^2 = 0.16$) (Nagakawa 2013:136).

Groupe	Nom	Variance	Ecart-Type
`items`	(Intercept)	200.6	49.46
`subjects`	(Intercept)	18165.3	134.78
`nquest`	(Intercept)	2445.9	49.46
résidus		19169.7	138.45

Tableau 6.10 – Effets aléatoires de la modélisation des temps de réaction (`lmer.cog`).

40. La modélisation est faite grâce aux fonctions du module statistique `lme4` (Bates, Mächler, Bolker & Walker 2015) sous l'environnement de traitement statistique `R` (*R Core Team* 2016).

41. Le Tableau E.16, disponible en Annexe E section 3.1 reprend le détail des résultats du ratio du test de likelihood pour les modèles généraux et spécifiques.

Le modèle linéaire à effet mixte `lmer.cog`[42] a été sélectionné en se basant sur l'AIC et le log-likelihood (Nagakawa 2013:136). Bien que les différences de latence semblent modérées lors de la comparaison des variances (ci-avant), la méthode par effets mixtes confirme qu'ajouter le **facteur de fluence** (modèle `lm.min`) a un effet significatif sur la robustesse du modèle (Table E.16). De la même manière, les résultats sont considérablement améliorés lors de l'ajout d'informations telles que les items et leur ordre d'apparition dans l'expérience (modèle `lmer.nquest`). Au contraire, un modèle complet incluant toutes les données des participants (i.e. *âge, genre, aptitude musicale et bilinguisme*) (modèle `lmer.full`) n'améliore pas la part de variance expliquée dans les résultats. Comme présupposé, **les résultats au CRT fournissent bien des informations pertinentes**. Ce dernier paramètre renforce le modèle, étant hautement corrélé à l'intercept ($r = 0.821$). En conclusion, hormis les effets aléatoires des items et des sujets, le modèle inclut les facteurs de complexité et de fluence ainsi que la position de l'item au sein du design expérimental et le résultat des participants au test de réflexion cognitive.

Le Tableau 6.11 rassemble les résultats de la modélisation des effets fixes (traitements de contrastes). L'intercept représente la condition originelle : **répétition avec pause pleine dans une condition de complexité non composée** ($M = (e^{660.82/1000}) = 1.94$ sec.). Les informations supplémentaires sur les participants (i.e. *âge, sexe, connaissance musicale, bilinguisme linguistique*) n'ont pas été inclues dans le modèle comme elles n'augmentent pas la variance expliquée.

	Estimate (log)	Estimate (sec)	Df	Valeur t	Sig. (p)
(Intercept)	660.82	1.94	66.13	20.62	<0.001***
`REP-FP sans flèche`	-50.25	-1.65	90.52	-10.78	<0.001***
`No-REP`	+13.95	+1.014	90.20	2.12	0.03*
`REP-C`	+0.60	1.000	90.23	0.09	0.93
`REP-UP`	+10.79	+1.011	90.80	1.64	0.11
`CRT scores`	-1.79	-1.001	60.00	-4.01	<0.001***

Tableau 6.11 – Effets fixes de la modélisation des temps de réaction (`lmer.cog`).

In fine, les résultats du modèle confirment que les stimuli fluents sont en effet plus longs à traiter que ceux porteurs de disfluences. On observe des **RTs plus longs pour des énoncés fluents** ($+1,01$ *sec.*, $p = 0,03$) et il semble que globalement la diminution de RTs soit **plus due à la présence d'un marqueur qu'à la répétition**. Enfin, les estimations négatives sur `cogrisk` révèlent que **plus le score des participants au Test de Réflexion Cognitive est faible, plus les participants réagissent rapidement aux stimuli** (i.e. formulent des hypothèses et/ou anticipent des événements, c'est-à-dire font des suppositions ou anticipent le contenu à venir).

42. Le modèle `lmer.cog` reprend les paramètres suivants : `lmer(1000*logRT complex + cond + (1-subject)+ (1-item) + (1-nquest) + cogrisk`, où cond = facteur de fluence, complex = facteur de complexité, subject = participant, item = stimuli-énoncé, cogrisk = résultat du *Relative Risk Cognition Test*, nquest = ordre d'apparition du stumulus. Nous avons testé six modèles possibles résumés au Tableau E.15, disponible en Annexe E section 3.1.

Globalement, les RTs étaient en effet 1.65 sec. plus courts pour les formes simples (`sans flèche`) que pour les formes composées (`intercept`). Par ailleurs, les items exempts de (dis)fluences ont bien élicités des temps de réaction plus longs que ceux incluant des marqueurs de (dis)fluence, et ce quel que soit le type de forme présenté (simple vs. composé) (+1.01 *sec.*, $p < 0.05$). Ainsi au sein de la condition simple, remplacer une pause pleine par un silence n'accroît pas significativement les temps de latence. Il en va de même lorsque la pause est entièrement supprimée (`REP-C`) (+1.011 et +1.00 sec. respectivement).

La Figure 6.14 illustre les différences des moyennes des moindres carrés entre les conditions de fluence et leurs intervalles de confiance [43]. Les différences des moindres carrés sont plus faibles si elles incluent une répétition avec silence que pour un item exempt de toute marque de (dis)fluence, mais non de manière significative. Dans l'ordre décroissant, les items sans (dis)fluence, les répétitions avec pause silencieuse, les répétitions contiguës témoignent de temps de latence plus importants que celles avec pause pleine. De plus, les temps de latence des **répétitions contiguës sont significativement plus courts que ceux des items sans (dis)fluence** ($p < 0.001$ pour `No-REP > REP-UP`).

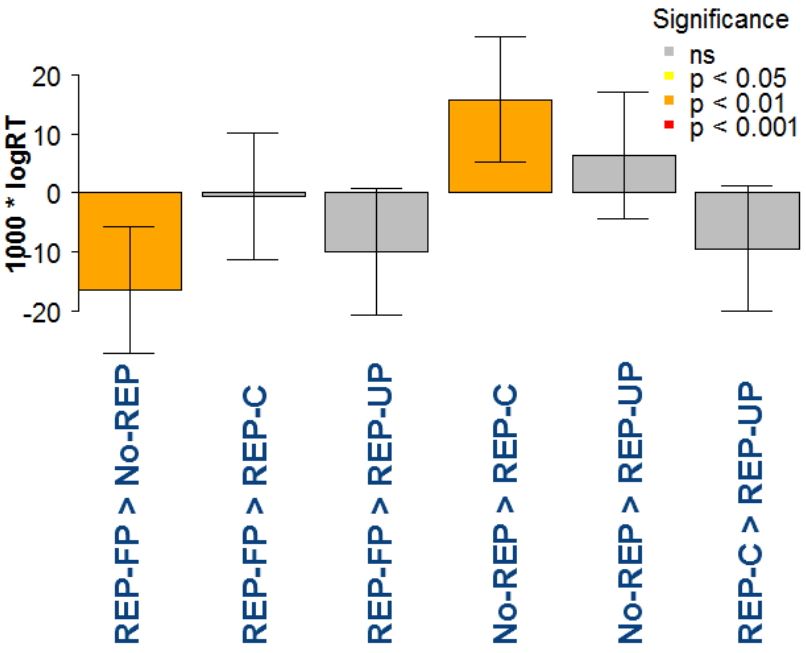


FIGURE 6.14 – Différences entre conditions par méthode des moindres carrés.

43. La méthode est basée sur une approximation de Kenward-Roger dans du package lmerTest (KUZNETSOVA, BROCKHOFF & CHRISTENSEN 2017) sous R (*R Core Team* 2016).

2.1.4 Conclusions et discussion

Cette étude a testé l'hypothèse selon laquelle les auditeurs s'**attendent à un contenu relativement plus long et/ou plus complexe si un segment est précédé par une marque de (dis)fluence**. Nos résultats ne convergent pas entièrement avec les études précédentes : notre expérience n'a fait apparaître **aucune interaction** entre les facteurs de fluence et de complexité. Les répétitions et les termes d'édition ne semblent ainsi pas faciliter le traitement ni biaiser les attentes des auditeurs lorsque la complexité augmente, ce qui reste en accord avec les résultats de MacGregor, Corley & Donaldson (2009).

Des différences significatives de temps de latence ont été observées pour les divers niveaux de fluence lorsque les formes étaient simples mais non pour les conditions en formes composées. Nos résultats montrent que les auditeurs francophones natifs **répondent de manière générale plus rapidement aux formes composées et simples lorsque les items contiennent une répétition contiguë ou une répétition avec pause pleine**. Dans l'ordre décroissant, les items sans (dis)fluence, les répétitions avec pause silencieuse, les répétitions contiguës affichent des temps de latence plus longs que les répétitions avec pause pleine. Ces résultats confirment l'hypothèse selon laquelle les répétitions avec pause pleine n'ont pas un effet additif ou cumulatif. En accord avec ceux-ci, la répétition avec pause pleine encouragerait les auditeurs à **réagir plus vite**, ce qui peut donc être interprétée comme **un outil pour capter leur attention plus que la marque d'un élément complexe à venir**.

Il est important de noter que dans la présente étude, nous avons seulement examiné l'effet des répétitions lexicales et leur terme d'édition au **début d'un constituant**. On ne sait pas quelle influence les répétitions dans les syntagmes nominaux pourraient avoir sur l'auditeur[44]. De plus, lors de la conception de l'expérience, nous étions conscients qu'il pouvait y avoir un **risque d'apprentissage spécifique** à l'expérience : les participants pouvaient réaliser que la valeur d'information des répétitions n'était plus un signal fiable d'un événement complexe à venir. En conséquence, ils le rejetteraient comme s'il provenait d'une **source non pertinente** (Alter & Oppenheimer 2009:230). De futures recherches pourraient faire l'hypothèse que les RTs dans les conditions au début de ce type d'expérience montreraient de plus grandes différences qu'en fin.

44. C'est ce que nous testons en section 2.2 de ce chapitre.

2.2 Expérience 2 : évaluation contextualisée des structures des répétitions

Cette expérience mesure l'impact du **contexte de production** d'une marque de (dis)-fluence sur l'évaluation de la fluence des énoncés par des locuteurs francophones natifs. L'unité de mesure est le **degré d'accord** des auditeurs par rapport à six affirmations portant sur chaque énoncé écouté. **Trois dimensions de la fluence** sont considérées par trois paires de questions : la fluence linguistique-phrastique (componentielle), la fluence linguistique situationnelle (holistique) et la fluence socio-interpersonnelle [45]. L'évaluation est donc avant tout **qualitative**.

En modifiant le contexte de production de l'énoncé, l'expérience vérifie aussi si l'évaluation d'un même énoncé dépend du **contexte de production** de la parole (Günther, Müller & Geyer 2017 ; Pradas Macías 2006 ; Schmid & Günther 2016). L'hypothèse principale est que l'évaluation de la fluence phrastique est moins soumise aux contraintes contextuelles que l'évaluation de la fluence situationnelle. Ensuite, en établissant un **profil de réactivité interpersonnelle** des participants, le design permet de tester l'importance des profils individuels sur l'évaluation de la (dis)fluence (Martin 2015). L'hypothèse principale étant qu'un participant aux capacités de relations interpersonnelles développées sera plus clément vu son aptitude à se projeter à la place du locuteur. Enfin, une des forces du design de cette étude est notamment de proposer l'écoute de **stimuli naturels** extraits de corpus oraux francophones alignés.

2.2.1 Matériel et méthode

Le matériel analysé est composé des résultats d'un questionnaire pré-test de 28 items portant sur la **réactivité interpersonnelle** du participant, et des **réponses d'évaluation à 40 énoncés** (40 stimuli x 6 affirmations = 240 réponses par participant). Les énoncés sont présentés soit dans leur contexte original de production soit dans leur contexte modifié (médiatique /vs/ non médiatique). Nous testons ainsi plus spécifiquement l'impact de l'**orientation médiatique** de la situation de parole sur la fluence.

Nous apprécions également l'influence du **type de (dis)fluence** présent dans l'énoncé sur l'évaluation de la fluence par l'auditeur. A ce titre, 20 énoncés contiennent à l'origine des **répétitions contiguës à l'identique**. Ils sont manipulés afin de créer 20 énoncés **ne contenant pas de répétition**. En parallèle, 20 énoncés contiennent à l'origine une **répétition avec une pause pleine** et sont manipulés afin de créer 20 nouveaux énoncés contenant une **répétition avec une pause silencieuse** de même longueur. Il s'agit d'un **design croisé**. Chaque auditeur est amené à évaluer 40 stimuli : 10 stimuli par condition de fluence dont la moitié sont présentés avec leur contexte de production original et l'autre moitié en contexte modifié.

45. Cette segmentation est une application directe du modèle théorique présenté au chapitre 2 section 1.4.3.

Nous présentons ci-après le (A) test de réactivité interpersonnelle, (B) la méthodologie d'évaluation tridimensionnelle de la fluence, (C) la sélection des stimuli et la génération des conditions et (D) le profil des participants.

A. Test de réactivité interpersonnelle

La première partie de l'expérience consiste en un questionnaire de réactivité interpersonnelle (« *Inter-personal Reactivity Index* » ; désormais F-IRI) (Gilet et al. 2013). Nous suivons le protocole appliqué par le projet CorpAge (Bolly & Boutet en prep.) : nous utilisons le F-IRI comme mesure des « **réactions d'un individu sur les observations d'expérience d'un autre** » (Davis 1980, 1983). L'index évalue **le degré d'empathie d'un individu**, défini comme « la capacité à se mettre à la place d'autrui, à imaginer ses pensées et ses sentiments ». Le F-IRI inclut 28 questions induisant une réponse située sur une échelle Likert de 5 points. Le test s'articule autour de **quatre dimensions** : (1) la prise de perspective, (2) l'imagination, (3) l'implication empathique et (4) la détresse personnelle [46] :

1. **Prise de perspective** *(PT – perspective taking)* : évalue la tendance des participants à adopter spontanément le point de vue psychologique des autres.
2. **Imagination** *(F – Fantasy)* : évalue la tendance des participants à se transposer de manière imaginaire dans les sentiments et actions de personnages fictifs (e.g. livres, films, jeux, etc).
3. **Implication empathique** *(EC - Empathic Concern)* : évalue l'orientation des sentiments des participants envers les autres et leur empathie envers les moins chanceux.
4. **Détresse personnelle** *(PD - Personal Distress)* : évalue « l'auto-orientation » des sentiments des participants dans des environnements interpersonnels tendus.

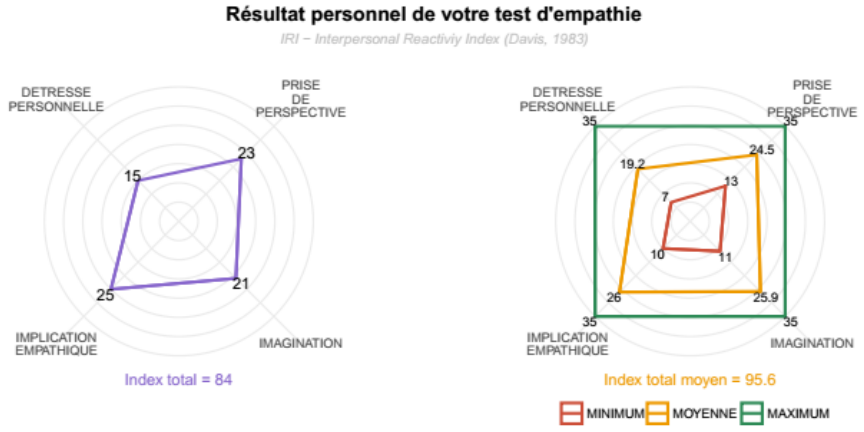


Figure 6.15 – Exemple de résultat (anonymisé) du test F-IRI envoyé à un participant.

Chaque participant intéressé reçoit ses résultats sous forme de graphe. Le panel A représente le résultat du participant sur chaque dimension et son index total (ici 84 sur un maximum de 140). Le panel B reprend les scores minimum, moyen et maximum des participants dans leur globalité.

46. Les 28 questions du test F-IRI sont reprises en Annexe E section 3.2.1.

B. Evaluation tridimensionnelle de la fluence

Chaque stimulus fait l'objet d'une **écoute unique** par l'auditeur. Après chaque écoute, l'auditeur exprime son **degré d'accord** avec **six affirmations** en utilisant une échelle de Likert allant de 1 à 5 [47]. Chaque paire d'affirmation porte sur une des **trois dimensions de la fluence** s'articulant autour de trois types de formulations des attentes de fluence(s) : phrastique, situationnelle et interpersonnelle.

- **Evaluation de la fluence componentielle linguistique-phrastique**. Il s'agit des attentes linguistiques décontextualisées. L'énoncé est-il bien constitué d'un tout cohérent et compréhensible, grammaticalement et syntaxiquement bien formé ?

  (A1) Les phrases sont bien formées.

  (A2) Les phrases contiennent des hésitations.

- **Evaluation de la fluence socio-interpersonnelle**. Elle dépend de l'ajustement avec les attentes sociales vis-à-vis du locuteur, reliée à la norme de fluence sociale ou interpersonnelle de l'auditeur. En d'autres termes, cette dimension questionne si la forme et le contenu du discours sont appropriés par rapport aux rôles sociaux imputables aux locuteurs et/ou à leurs relations interpersonnelles. Elle est intimement liée à l'évaluation de l'aisance relative du locuteur par l'auditeur.

  (A3) La manière de parler est fluide.

  (A4) La manière de parler est agréable à l'écoute.

- **Evaluation de la fluence linguistique holistique située (situationnelle)**. Il s'agit des attentes liées à l'utilisation d'un registre. Autrement dit, la forme et le contenu du discours sont-ils compatibles, en adéquation, avec les connaissances liées au genre et/ou avec les contraintes pragmatiques de la situation ?

  (A5) La manière de parler utilisée dans ce contexte me plaît.

  (A6) La manière de parler utilisée dans ce contexte est inappropriée.

Pour permettre à l'auditeur de se concentrer **davantage sur la forme** du message que sur son contenu, chaque stimulus audio est **accompagné d'un énoncé de contextualisation**. Tous les énoncés sont des groupes nominaux assez courts repris en exemple 23 [48].

Exemple (23) – Exemple d'énoncé de contextualisation.

a. Écriture d'un livre sur Jimmy Hendricks
b. Découverte de la tour de Pise
c. Vie de quartier à Paris
d. Choix d'un métier
e. Relation entre les arts et la connaissance
f. Réalisation d'un film
g. Reproduction des araignées

47. A ce titre, nous traitons les réponses des participants comme des données ordinales.
48. La liste complète des énoncés de contextualisation est donnée en Annexe E section 3.2.2.

C. Sélection des stimuli & génération des conditions

Les 40 stimuli d'origine proviennent des corpus oraux francophones présents dans Metacorpus (Christodoulides, Grosman, Degand & Simon 2017) [49]. Ils ont été sélectionnés manuellement en fonction de **critères stricts** : les énoncés sont issus de situations à orientation médiatique ou non médiatique mais non semi-médiatique, afin de toucher aux extrêmes du continuum lors de la manipulation de la condition de fluence situationnelle.

Les répétitions ne sont constituées que d'**un seul répété**. De plus, les segments répétables sont uniquement des **articles grammaticaux monosyllabiques**. Ce choix se justifie par leur distribution prépondérante dans les corpus (voir section 1.2). Afin de minimiser l'influence d'autres facteurs de fluence phrastique, la sélection des stimuli a exclu les énoncés incluant d'autres (dis)fluences que les répétitions (e.g. troncations, reformulations, pauses avant ou après la répétition, etc.). Enfin, nous avons veillé à ce que la répétition ne soit **ni en début ni en fin d'énoncé** pour éviter tout **effet de primauté ou de récence** lors de l'évaluation.

Avant l'expérience, un échantillon plus large de stimuli a été testé par des experts de la langue pour évaluer leur plausibilité à être énoncés dans des contextes médiatiques et non médiatiques. Cette étape pilote a permis de sélectionner les 40 stimuli finaux et d'effectuer les manipulation du facteur de fluence phrastique. Les stimuli ont une durée de **7 à 19 secondes** ($\mu = 11.34, \sigma = 3.1$). Le questionnaire post-test demandait aux participants d'évaluer le **côté naturel des stimuli**, 90% d'entre eux les ont considérés comme naturels.

Manipulation du facteur de fluence phrastique. Pour des raisons techniques, il existe une **distribution complémentaire** entre les 4 niveaux de fluence phrastique étudiés : répétition contiguë (`REP-C`), absence de répétition (`No-REP`) d'une part, répétition avec pause pleine (`REP-FP`) et répétition avec pause silencieuse (`REP-UP`) d'autre part. Les énoncés contenant une répétition contiguë ont servi à créer les énoncés exempts de répétition, par suppression du premier segment. Les énoncés à répétitions avec pause pleine ont eux servi à créer des énoncés avec des répétitions incluant une pause silencieuse par remplacement du *euh* par un bruit de fond (voir exemple 24). Un auditeur évaluant un énoncé dans sa condition `REP-C` ou `REP-FP` n'écoutera jamais son équivalent `No-REP` ou `REP-UP` respectivement, et vice-versa.

49. Pour rappel, Metacorpus comprend les corpus C-HUMOUR (Grosman 2016), LOCAS-F (Martin, Degand & Simon 2014), C-Phonogenre (Pršir, Goldman & Auchlin 2014), Rhapsodie (Gerdes, Kahane & Pietrandrea 2012) et DRIVE (Christodoulides 2016).

Exemple (24)

`REP-C` → `No-REP` **: rep. contiguë → absence de rep.**

... on s'est retrouvé sans eau sans gaz ni téléphone ni électricité et les pompiers, la ville est assez haute et **{la la}** Loire est en bas, les malheureux allaient puiser de l'eau dans la Loire et remontaient pour arroser (`REP-C`, stimuli 1521-2-2NM-Rhap-D0003)

... on s'est retrouvé sans eau sans gaz ni téléphone ni électricité et les pompiers, la ville est assez haute et **{la}** Loire est en bas, les malheureux allaient puiser de l'eau dans la Loire et remontaient pour arroser (`No-REP`, stimuli 1521-2-2NM-Rhap-D0003-manip)

`REP-FP` → `REP-UP` **: rep. avec pause pleine → rep. avec pause silencieuse**

... ces nobles vont en général chanter entre guillemet **{leur euh leur}** nostalgie leur mélancolie vis-à-vis de leurs privilèges perdus (`REP-FP`, stimuli 386-1-2NM-conv-f-2)

... ces nobles vont en général chanter entre guillemet **{leur (0.900 sec.) leur}** nostalgie leur mélancolie vis-à-vis de leurs privilèges perdus (`REP-UP`, stimuli 386-1-2NM-conv-f-2-manip)

Manipulation du facteur de contexte de production. Chaque stimulus écouté par l'auditeur est contextualisé par le biais de l'**affichage d'une image** lui indiquant le type de situation de parole dans lequel l'énoncé a été produit. Ainsi pour chaque stimulus de l'exemple précédent, **deux contextes sont possibles** : médiatique ou non médiatique (Figures 6.16 et 6.17). Le contexte est soit celui d'origine soit le contexte modifié. A l'instar des conditions de fluence, un auditeur écoutant un stimulus dans une condition A ne rencontrera pas ce même stimulus dans sa condition B.

FIGURE 6.16 – Image de contextualisation d'une situation de parole non médiatique.

FIGURE 6.17 – Image de contextualisation d'une situation de parole médiatique.

D. Profil des participants

Tous les participants à l'expérience étaient des locuteurs **francophones natifs** volontaires et **dédommagés pour leur contribution** à hauteur de 10 euros. Ils ont été **contactés en ligne** (e.g. cercles académiques, listes de diffusion de participants, réseaux sociaux, etc.). Avant l'expérience, tous les participants ont signé un formulaire de consentement et ont pris connaissance des conditions de participation : être exempt de toute déficience auditive, posséder un système d'écoute (casque ou enceintes) et se trouver dans un environnement calme. Sur les 204 participants, 2 ont été retirés des analyses en raison d'un **temps de lecture des instructions** inférieur à 5 secondes.

Les 202 participants sont âgés de **19 à 72 ans** ($\mu = 30.62, \sigma = 11.62$), 69% étaient de sexe féminin. Parmi eux, **23% sont originaires de Belgique et 76% de France**. Les participants étaient autorisés à utiliser un système d'enceintes (système son externe) ou des écouteurs/un casque audio (43% et 57% des participants respectivement). Près de la moitié des participants était bilingue (47%) contre un quart d'entre eux se déclarant monolingue ou trilingue (26% et 23% respectivement). Moins de 5% parlait couramment plus de trois langues.

Le niveau d'études des participants était réparti ainsi : 22.8% issus d'un parcours professionnel ou secondaire, 29.2% d'un diplôme de cycle court, 27.7% en possession d'un diplôme de cycle long et 20.3% détenteurs d'un diplôme de troisième cycle. Enfin, 68.8% des participants ont évalué que leur **métier était en relation avec la langue**, et 67.7% qu'ils étaient amenés à y **juger de la parole**.

Les scores moyen total du test F-IRI des participants est de 95.63 ($\sigma = 12.28$, $max = 123$, $min = 60$). L'évaluation de la dimension « implication empathique » a obtenu le score moyen le plus élevé ($\mu = 26.02$, $\sigma = 4.72$), suivi des dimensions « imagination » ($\mu = 25.88, \sigma = 5.15$), «prise de perspectives » ($\mu = 24.5, \sigma = 4.58$) et « détresse personnelle » ($\mu = 19.24, \sigma = 5.43$). En nous basant sur les écarts interquartiles, nous avons catégorisé les scores de chaque participant en 5 niveaux. Les scores catégorisés (1) très bas, (2) bas, (3) milieu, (4) haut et (5) très haut, correspondent respectivement aux scores des écarts inter-quartiles de la population de participants compris entre (1) [0%-20%[, (2) [20-40%[, (3) [40%-60%[, (4) [60%-80%[et (5) [80%-100%] [50].

50. Les intervalles ouverts n'incluent pas le seuil.

2.2.2 Hypothèses

Dans le cadre de notre étude, la ***dis***fluence correspondrait à une **sur-saillance par rencontre ou déception marquée des attentes de l'auditeur vis-à-vis du contexte**, externe et interne (i.e. *contenu linguistique, du registre et/ou de locuteur dans une situation*) [51]. Ainsi, différents types de (dis)fluence émergent en fonction du **type d'attente**, mais également en fonction du **type de contexte** pris en compte.

(A) **Hypothèse de structure de la marque :** la structure interne d'une répétition influence son évaluation (hors contexte situationnel). Plus précisément :

Les répétitions non contiguës avec un terme d'édition marqué sont jugées plus sévèrement que celles contiguës et sans terme d'édition marqué. Cette hypothèse soutient le continuum suivant, allant du plus fluent au plus disfluent : (fluent) → `No-REP` → `REP-FP` → `REP-UP` → `REP-C` → (disfluent).

L'orientation de cette hypothèse est motivée (1) par les résultats en production, notamment, la proportion relative importante de répétitions impliquant une pause vide, même en discours préparé, et (2) par les résultats d'observation en perception, notamment la facilité de traitement et la saillance de la structure observée dans l'expérience précédente [52].

(B) **Hypothèse du contexte de production de parole médiatique :**

Les stimuli en contextualisation formelle médiatique sont jugés plus sévèrement que ceux en contextualisation informelle non médiatique.

Une même marque de (dis)fluence dans deux contextes de production différents n'induit pas nécessairement la même évaluation chez les auditeurs.

Par ailleurs, nous n'émettons pas l'hypothèse d'une interaction entre le type de marque et le contexte de production médiatique ou non médiatique (i.e. une marque n'est pas plus acceptée qu'une autre en fonction du contexte de production). Ce dernier paramètre n'a effectivement pas révélé d'association en fonction du type de marque d'édition au sein de la structure de répétition [53].

(C) **Hypothèse de réactivité interpersonnelle :** le profil de la personne influence son évaluation des stimuli. Le score du test F-IRI est associé à l'évaluation des (dis)fluences.

Globalement, les auditeurs marquant un score F-IRI plus élevé sont plus sensibles à la présence de (dis)fluence(s).

En détail, nous testons l'association de l'index avec l'évaluation de trois dimensions de la fluence par les participants : la fluence componentielle linguistique-phrastique, la fluence linguistique holistique située et la fluence sociale interpersonnelle [54].

51. Cette segmentation est une application directe du modèle théorique présenté au chapitre 2 section 1.4.3
52. Voir les sections 1.2 et 2.1 de ce chapitre.
53. Voir section 1.3 de ce chapitre.
54. Retour à la proposition de modèle de la fluence-disfluence Chapitre 2 section 1.4.3.

Nous posons l'hypothèse selon laquelle le score F-IRI est important dans l'explication de la variance de la fluence perçue lorsqu'on questionne les participants sur les dimensions situationnelle (A3-A4) et interpersonnelle (A5-A6). **L'impact des profils personnels serait par contre indépendant pour une évaluation de questions portant sur la fluence phrastique** (A1-A2).

Parmi les dimensions du F-IRI évaluées, nous émettons l'hypothèse que la perception et l'évaluation des (dis)fluences[55] dépendent de la capacité de l'interlocuteur à anticiper du contenu. Plus les participants ont de bonnes capacités de projection plus ils seraient à même de traiter les (dis)fluences de manière efficace . Concrètement, nous faisons l'hypothèse formelle que les participants ayant des résultats hauts-très hauts dans les dimensions de l'implication empathique, de la prise de perspective et de l'imagination jugent en effet la fluence plus favorablement. De manière synthétique, obtenir des scores F-IRIs hauts augmenterait la probabilité d'évaluation d'un stimulus comme fluent sur le continuum de la fluence-disfluence.

Variable	Type de var.	Nb. de niveaux	Niveaux de var.
Degré d'accord	ordinale	5	pas du tout d'accord, 2, 3, 4, tout à fait d'accord
Degré de fluence perçue	ordinale	5	1, 2, 3, 4, 5
Dimension de la fluence	catégorielle	6 | 3	A1-A2 phrastique, A3-A4 situationnelle, A5-A6 sociale
Orientation médiatique	catégorielle	2	`Média, No-Média`
Fluence phrastique	catégorielle	4	`No-Rep, REP-C, REP-FP, REP-UP`
Index F-IRI Total	ordinale	5	très bas, bas, milieu, haut, très haut
F-IRI détresse personnelle	ordinale	5	très bas, bas, milieu, haut, très haut
F-IRI imagination	ordinale	5	très bas, bas, milieu, haut, très haut
F-IRI implication empathique	ordinale	5	très bas, bas, milieu, haut, très haut
F-IRI prise de perspective	ordinale	5	très bas, bas, milieu, haut, très haut

TABLEAU 6.12 – Résumé des variables et niveaux de variables.

55. ...et donc du continuum fluence-disfluence.

2.2.3 Importance du stimulus, du participant et des seuils équidistants

Dans un premier temps, nous construisons un **modèle mixte reconnaissant dans la structure de nos données deux effets aléatoires** : (1) le même stimulus audio a été évalué selon six différentes mesures de fluence perceptive, (2) le même participant a évalué plusieurs stimuli. Ce modèle permet ainsi de reconnaître l'éventuelle part explicative (i.e. l'influence) du stimulus audio lui-même et du profil du participant sur l'évaluation de la (dis)fluence.

Premièrement, nous testons cette hypothèse sous-jacente de l'influence des effets aléatoires sur l'évaluation de la fluence : *quelle part de la variance est expliquée par ces facteurs ?* Une ANOVA compare (a) un modèle ordinal vide avec pour seul facteur la variable indépendante le « degré d'accord » à (b) un modèle ordinal mixte avec pour variable indépendante le « degré d'accord » et pour effets aléatoires le stimulus audio et le participant.

Nos résultats montrent que **l'ajout d'effets aléatoires augmente la puissance du modèle** ($Pseudo - R^2 = 0.22438, p < 0.001$), soit 22.44% de variance expliquée par les effets aléatoires[56]. La Figure 6.18 reprend la variance des évaluations de chaque participant et de chaque stimulus par rapport à l'effet sur la puissance explicative du modèle. Les participants utilisent donc bien l'échelle de réponse différemment et les stimuli ont une influence sur l'annotation de la fluence.

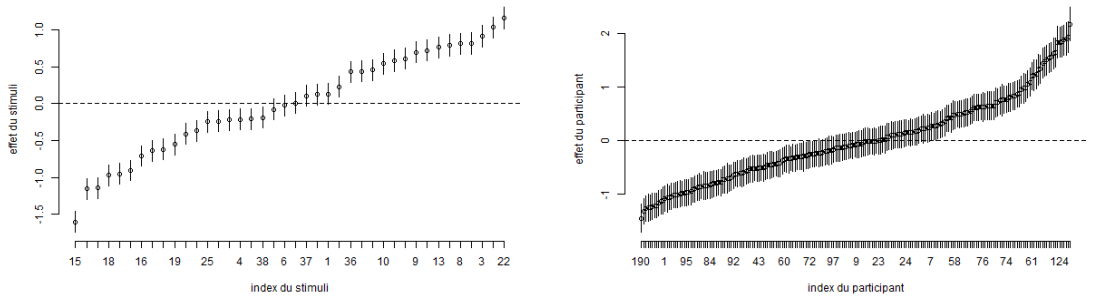


FIGURE 6.18 – Effets de la variance des stimuli (*panel gauche*) et desparticipants sur la taille des effets aléatoires du modèle (*panel droit*).

Lors de l'évaluation de nos trois hypothèses, nous présentons les résultats graphiques sur leur échelle d'origine, le `degré d'accord`. Pour la modélisation, nous avons transformé la variable `degré-d'accord` en `degré-de-fluence` en inversant les échelles des questions A2 et A6 qui étaient formulées de manière négative. Enfin, les espaces entre les catégories ordonnées de 1 à 5 sont partout considérés comme « équidistants ». C'est à dire qu'on reconnaît une distance égale entre le degré 3 (le neutre) et le degré 2 par rapport à la

56. $Pseudo - R^2$ basé sur la méthode de Cox and Snell. Le $Pseudo - R^2$ de McFadden est lui de 0.08557. Cet indice est parfois recommandé mais il est basé sur une méthode de calcul plus proche de celle effectuée pour des modèles linéaires.

distance entre les degrés 3 et 4. Un test de log-likelihood entre un **modèle à seuils équidistants** /vs/ symétriques a en effet prouvé l'apport significatif d'information de ce premier ($LRtest = 535.15, p < 0.001$).

Cependant, le degré de fluence perçue peut être pris comme une variable numérique et être associé à une médiane par affirmation (Figure 6.19). Quelques observations d'ordre général :

- Premièrement, l'affirmation A6[57] a suscité **les scores les plus élevés de fluence**. Cette affirmation est aussi évaluée avec le moins de variance ($Me = 5, \sigma = 1.031$)[58].

- Ensuite, les affirmations A1, A2, A3 ont toutes un score médian de 4. Parmi elles, on observe une variance plus importante des résultats pour l'affirmation A2 ($\sigma = 1.247$), surtout dans deux cas : (1) présence d'une répétition avec pause pleine ($\sigma = 1.268$) et (2) répétition avec pause silencieuse ($\sigma = 1.266$).
- Enfin, les affirmations A4 et A5 ont un score médian de 3. Dans le cas de l'affirmation A5, la variance dans le score de fluence perçue est moindre lors (1) de la présence d'une répétition avec pause silencieuse ($\sigma = 1.164$), et (2) d'une répétition contiguë ($\sigma = 1.168$), qui semble légèrement améliorer le score de fluence. Pour confirmer, cette hypothèse, nous devons avant tout tester l'impact de la structure interne d'une répétition sur son évaluation.

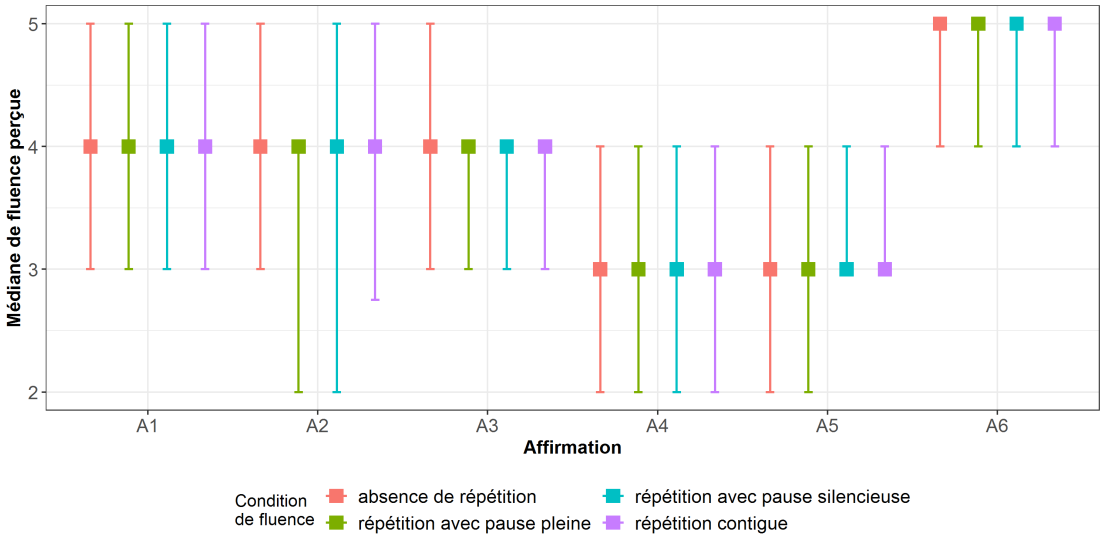


FIGURE 6.19 – Médianes de fluence perçue par affirmation et condition de fluence.

57. L'affirmation était « La manière de parler utilisée dans ce contexte est inappropriée » ; nous avons inversé ce score pour le présenter sur une échelle de degré de fluence comparable aux autres affirmations. Cependant, *post-hoc* les résultats soulèvent un questionnement sur la pertinence de l'interrogation. Le terme approprié n'apportant que peu d'informations.

58. Les scores ordinaux ne suivant pas une distribution normale, nous présentons les mesures centrales des données utilisant la médiane (Me) et non la moyenne.

2.2.4 Résultats

(A) Hypothèse de structure interne de la marque

Nous établissons un **modèle cumulatif mixte** ne considérant que la structure interne de la (dis)fluence, avec pour effets aléatoires le stimulus et le participant. Le Tableau 6.13 reprend les résultats de la modélisation sur l'ensemble du set de données, toutes affirmations confondues. L'intercept représente ici le niveau de facteur `No-REP`, l'absence de répétition.

	Estimate	Odds Ratio	Erreur Standard	Valeur z	Sig.
(intercept)					
`REP-C`	-0.126	0.883	0.024	-5.29	< 0.001
`REP-FP`	-0.166	0.848	0.218	-0.76	0.446
`REP-UP`	-0.138	0.872	0.218	-0.63	0.526

TABLEAU 6.13 – Effets fixes de la modélisation du degré de fluence en fonction du type de (dis)fluence.

On remarque que la seule différence significative de degré de fluence perçue est attestée lors de la comparaison entre les évaluations des modalités `No-Rep` vs. `REP-C`. **Le degré de fluence perçue tend à diminuer à l'apparition d'une répétition** ($estimate = -0.126, oddsratio = 0.88$), quelle que soit sa structure interne. En d'autres termes, un stimulus porteur de répétition a 88% plus de probabilité d'être évalué de manière plus disfluente qu'un stimulus exempt de répétition.

Nous avons ensuite testé l'influence des différentes structures internes sur l'**évaluation de chaque dimension de la fluence**[59]. Les résultats montrent aussi qu'en réalité, **quel que soit le type de fluence testé la présence d'une répétition semble suffire à la dévaluation de la fluence**. Aucune différence significative n'existe entre les conditions de fluence du stimulus. Néanmoins, **la force de cette dévaluation est plus importante pour la fluence socio-interpersonnelle** ($OR = 0.970, p < 0.001$), que pour la fluence situationnelle ($OR = 0.8585, p < 0.001$) ou phrastique ($OR = 0.8834, p < 0.001$).

Ainsi, notre première hypothèse est partiellement validée. La présence d'une répétition influence l'évaluation de la fluence et ce quelle que soit la structure de la répétition. Séparément, toutes les conditions de fluences sont impactées. En détail, la fluence socio-interpersonnelle est la plus touchée, c'est à dire l'impression de fluidité et d'aisance dans le discours. En d'autres termes, même si une phrase n'est pas perçue comme contenant des hésitations ou comme inappropriée, **l'énoncé est moins agréable à l'écoute et paraît moins fluide**.

59. Pour rappel, fluence phrastique (A1, A2), fluence sociale interpersonnelle (A3, A4) et fluence situationnelle (A5, A6).

Par souci de transparence des données, la Figure 6.20 reprend les donnés brutes des échelles de Likert (pour l'ensemble des échantillons audio).

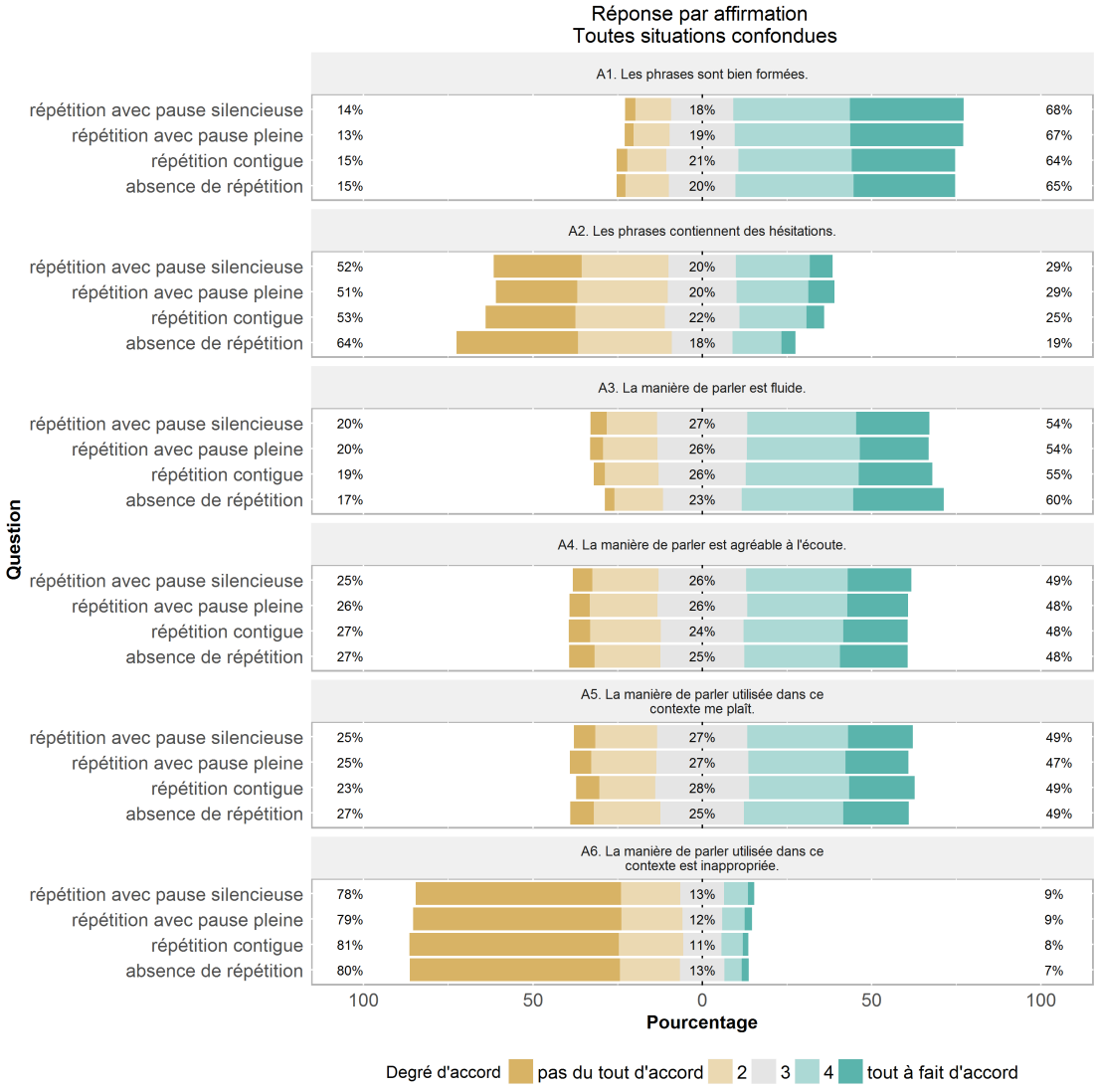


FIGURE 6.20 – Evaluation des participants (degré d'accord aux affirmations, échelle de Likert de 1 à 5) par degré de fluence phrastique (condition de fluence des stimuli) et par dimension de la fluence (affirmations A1-A6).

(B) Hypothèse du contexte de production de parole médiatique :

Nous testons ensuite si **deux contextes de production induisent une différence d'évaluation chez un même participant**. Notre hypothèse est que les stimuli en contextualisation formelle médiatique sont jugés plus sévèrement que ceux en contextualisation informelle non médiatique. Nous n'émettions par contre pas l'hypothèse d'une interaction entre le type de marque et le contexte de production (i.e. une marque n'est pas plus acceptée qu'une autre en fonction du contexte de production).

Un modèle cumulatif ordinal avec pour variable dépendante le degré de fluence perçue et pour variable indépendante l'orientation médiatique testée est établi selon la même logique que le modèle précédent. Les résultats sont présentés au Tableau 6.14. L'intercept représente le degré d'orientation médiatique.

	Estimate	Odds Ratio	Erreur Standard	Valeur z	Sig.
(intercept)					
`Non-média`	0.033	1.033	0.017	1.97	< 0.05

TABLEAU 6.14 – Effets fixes de la modélisation du degré de fluence en fonction du type de média testé.

Nous notons une différence significative de l'appréciation de la fluence perçue en fonction de l'orientation médiatique de la situation dans laquelle l'énoncé a été émis (*estimate* $= 0.33$, *oddsratio* $= 1.033$, $p < 0.05$). **Le degré de fluence perçue tend à augmenter dans une situation non médiatique**. Ainsi, pour une augmentation d'une unité dans le score de fluence perçue, la probabilité de se trouver dans une situation non médiatique augmente de 103%. En d'autres termes, les participants jugent de manière moins sévère les situations non médiatiques que médiatiques.

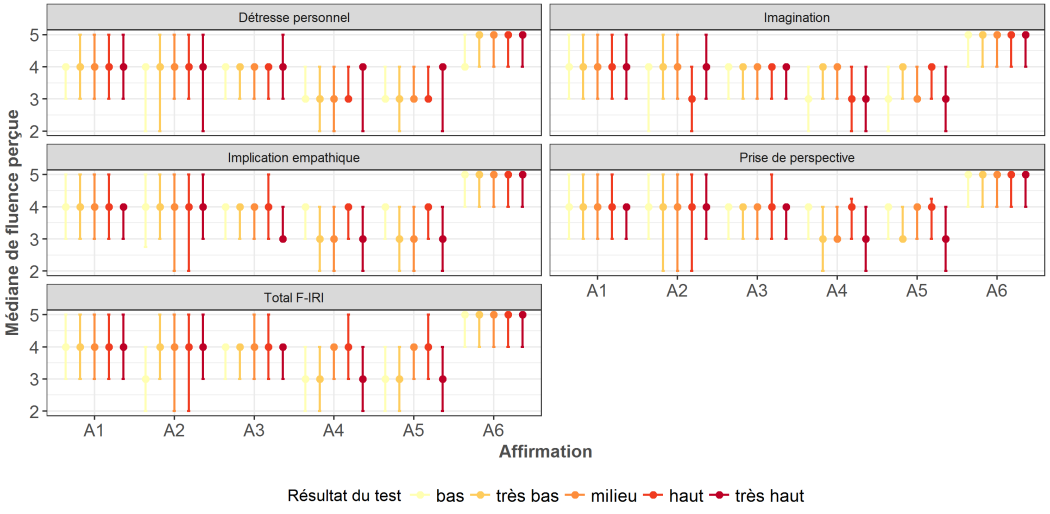


FIGURE 6.21 – Médianes de la fluence perçue par affirmation, en fonction des index F-IRI des participants.

Cependant, l'hypothèse n'est que partiellement confirmée. Si une différence existe dans son ensemble entre le degré de fluence perçue et l'orientation médiatique, **aucune interaction n'est observée entre la dimension de fluence questionnée** et le degré d'orientation médiatique. Autrement dit, la taille d'effet étant minime, on ne peut affirmer que la contextualisation a plus d'influence sur la dimension situationnelle. Elle impacte les trois dimensions de fluence de manière égale dans leur ensemble.

(C) Hypothèse de réactivité interpersonnelle

Nous testons enfin si le profil du participant établi à partir des index F-IRI influence son évaluation du degré de fluence des stimuli, et ce sur les différentes dimensions du continuum fluence-disfluence. En détail, nous testons l'association de l'index avec l'évaluation de trois dimensions de la fluence par les participants : la fluence linguistique-phrastique, la fluence linguistique situationnelle et la fluence sociale interpersonnelle [60]. Nous avions émis l'hypothèse que les participants ayant obtenu un score F-IRI plus élevé étaient plus sensibles à la présence de (dis)fluence(s).

Nous avons d'abord testé l'influence du score total F-IRI sur le degré de fluence perçue afin de voir si le profil du participant infléchit son évaluation des stimuli. Si sur l'ensemble des affirmations, **les résultats montrent que le score total F-IRI n'a pas d'impact sur l'évaluation de la fluence dans son ensemble** ($estimate = 0.006, p = 0.144$) ; **il influence toutefois chacune des dimensions de la fluence individuellement**.

En détail, nous testons donc l'hypothèse selon laquelle le score F-IRI est essentiel dans l'évaluation de la **fluence situationnelle** (A3-A4) et de la fluence **interpersonnelle** (A5-A6). Nous observons bien un lien de causalité sur ces deux dimensions de la fluence (situationnelle $estimate = 0.383$, $p < 0.03$; interpersonnelle $estimate = 0.266$, $p < 0.05$). Autrement dit, envisager les scores test F-IRI dans leur globalité est déjà un facteur explicatif de l'évaluation de la fluence situationnelle et interpersonnelle. **Plus le participant a un score F-IRI élevé, plus il jugera favorablement les fluences situationnelles et interpersonnelles mais non phrastiques**. Il n'existe en effet pas de relation entre les affirmations portant sur la fluence phrastique et les profils F-IRI des participants ($estimate = -0.038$, $p < 0.7$). La fluence phrastique se réfère en effet a une dimension requérant moins d'implication personnelle.

Parmi les dimensions du F-IRI évaluées, nous émettions l'hypothèse que l'évaluation des (dis)fluences dépend surtout de la capacité de l'interlocuteur à anticiper le contenu et à se projeter à la place d'autrui. En d'autres termes, **plus les participants ont de bonnes capacités de projection plus ils sont à même de traiter efficacement les (dis)-fluences**. Nous avons ainsi effectué **4 modélisations** basées sur la formule suivante où le degré de fluence varie et questionne (1) l'ensemble des dimensions (2) la dimension phrastique, (3) la dimension situationnelle et (4) la dimension interpersonnelle.

60. Retour à la proposition de modèle de la fluence-disfluence 2 section 1.4.3.

```
fluence = imagination + détresse personnelle + implication empathique +
      prise de perspective + (1|index audio) + (1|participant)
```
[61].

Nos observations livrent les résultats suivants :

- **Modèle 1. L'ensemble des dimensions de fluence.** On observe un effet quadratique de **la détresse personnelle** ($estimate = 0.29, p < 0.025$) et un effet linéaire de **l'implication empathique** ($estimate = 0.327, p < 0.03$). En d'autres termes, plus un individu fait preuve d'implication empathique, plus il jugera comme fluents les items. Plus un individu témoigne d'une détresse personnelle élevée, c'est-à-dire d'une capacité d'auto-orientation dans des environnements interpersonnels tendus, plus il jugera les items de manière fluente, mais jusqu'à un certain seuil (l'aspect quadratique de l'effet indique qu'au-delà d'un certain seuil la détresse personnelle n'a plus d'effet ou dévalue les scores de fluence).
- **Modèle 2. La dimension fluence phrastique**. On note un effet linéaire de l'**imagination** ($estimate = -0.34, p < 0.028$) et un effet quadratique de la **détresse personnelle** ($estimate = 0.33, p < 0.027$). Autrement dit, moins les participants ont tendance à se transposer dans les sentiments des autres plus ils jugent favorablement la fluence phrastique des énoncés. Vu d'une autre perspective, **plus les participants ont une imagination débordante moins ils évaluent défavorablement la construction grammaticale d'un énoncé**. Ensuite, à l'instar de l'observation sur l'ensemble des dimensions, plus un individu témoigne d'une détresse personnelle élevée, plus il jugera les items de manière fluente, mais jusqu'à un certain seuil. De même, au-delà d'un certain seuil la détresse personnelle n'a plus d'effet ou dévalue les scores de fluence.
- **Modèle 3. La dimension situationnelle**. On identifie un effet linéaire de taille plus importante de l'**implication empathique** ($estimate = 0.41, p < 0.01$). Conformément à notre hypothèse, plus un individu est empathique plus il évalue de manière favorable l'énoncé de parole dans son contexte.
- **Modèle 4. La dimension sociale interpersonnelle**. On observe à nouveau un effet quadratique de la **détresse personnelle** ($estimate = 0.377, p < 0.023$) et un effet linéaire de l'implication empathique ($estimate = 0.383, p < 0.05$). En résumé, plus l'individu fait preuve d'une détresse personnelle élevée (jusqu'à un certain seuil) plus il évaluera positivement les énoncés sur leur dimension interpersonnelle (i.e. perception fluide agréable de la parole).

En résumé, les résultats de l'expérience révèlent que ce sont **les participants ayant des résultats hauts et très hauts dans les dimensions de l'implication empathique et de la détresse personnelle qui jugent les items de manière plus favorable**. Ainsi, obtenir des scores F-IRI élevés dans ces dimensions augmente la probabilité d'évaluation d'un stimulus comme fluent sur le continuum de la fluence-disfluence.

61. Où « 1| » représente pour chaque facteur les effets aléatoires du modèle prédisant différentes ordonnées à l'origine selon les paramètres du modèle. Chaque stimulus audio a ses particularités et sa variance, et chaque participant a sa manière d'évaluer la fluence et la disfluence.

Par ailleurs, cette implication empathique est particulièrement présente dans le cas de l'évaluation de la fluence socio-interpersonnelle. Au reste, différents types d'évaluation de la fluence sont influencés différemment par les dimensions du F-IRI. L'effet le plus **cohérent à travers les dimensions est l'association positive avec la détresse personnelle. On remarquera que, contrairement à notre hypothèse de départ, le test F-IRI n'est pas indépendant de l'évaluation de la fluence phrastique**, l'influence de l'imagination du participant dans l'évaluation de la fluence phrastique est un résultat interprétable mais néanmoins inattendu.

2.2.5 Conclusions et discussion

Cette éude a testé différentes **définitions des types de fluence** et a évalué à quel point les **types d'évaluation** que ces fluences entraînent sont plus ou moins influencés par (1) **la définition du contexte** et (2) **les caractéristiques du participant**. Une question chapeaute cet exercice : « Par quelle(s) dimension(s) du continuum fluence-disfluence le degré de fluence perçue par les participants est-il le plus déterminé ? ». Pour y répondre, nous avons plus précisément questionné (A) « Comment le degré de fluence perçue est-il influencé par la répétition à l'identique ? », (B) « Comment le degré de fluence est-il influencé par la présence d'un contexte ? » et (C) « Dans quelle mesure le profil d'inter-réactivité (F-IRI) des participants augmente-t-il notre capacité à prédire le degré de fluence perçue ? ».

Nous posions l'hypothèse selon laquelle **la structure interne d'une répétition influence son évaluation**. Plus précisément que les répétitions non contiguës avec un terme d'édition marqué sont jugées plus sévèrement que celles contiguës et sans terme d'édition marqué. Nos résultats montrent en réalité que **les structures avec répétitions sont jugées plus sévèrement que les structures sans répétition**, et qu'il n'existe **pas de différence d'évaluation en fonction de la structure de cette répétition**. Nous avons montré que la force de cette dévaluation est pertinente en fonction de **toutes les définitions de la fluence envisagées** (phrastique, situationnelle et interpersonnelle).

Nous avons ensuite **posé l'hypothèse de contexte de production selon laquelle la parole médiatique est jugée plus sévèrement** que la parole non médiatique. Cette hypothèse est confirmée sur les dimensions de fluence dans leur ensemble. **Aucune interaction** n'est cependant observée **avec les différentes dimensions de la fluence**. En d'autres termes, l'effet minime ne permet pas d'affirmer que l'influence du contexte est plus importante dans le cadre de l'évaluation de la fluence situationnelle.

Enfin, nous posions l'hypothèse de réactivité interpersonnelle selon laquelle **le profil du participant influence son évaluation des stimuli**. Cette hypothèse est confirmée. Globalement, **les participants marquant un score F-IRI plus élevé sont plus sensibles à la présence de (dis)fluence(s)**. En détail, nous avons vu que différents types de fluences sont associées de manière plus ou moins forte avec différentes dimensions

de ce score F-IRI. Nous retenons notamment l'**importance du niveau d'implication empathique** (association linéaire) et de la **détresse personnelle** (association quadratique). Par ailleurs, le test F-IRI n'est pas indépendant de l'évaluation de la fluence phrastique, nous notions l'**influence de la capacité d'imagination dans l'évaluation de la fluence phrastique**.

Bien qu'un plus grand nombre de participants permettrait de mettre en relation toutes ces modalités et leurs interactions[62], ce type de modélisation nuirait sans doute à la lisibilité et à l'interprétation des résultats. En s'attachant à la résolution de ces questions séparément, les résultats sont tout à fait évocateurs. Cette expérience a permis de mettre en valeur une grande **variabilité des résultats en fonction du type de fluence considéré** et de souligner donc l'**importance du type de question énoncée** lors des études expérimentales portant sur la fluence, selon que l'objet d'étude est détaché ou non du contexte et de la personne. Enfin, notre design expérimental s'est confronté à un challenge notable. Il incarne une **double tentative de conciliation**. Il tente premièrement de concilier **la naturalité et la variabilité des stimuli** avec le contrôle même de cette variation. Il a également essayé de concilier la prise en compte de marques phrastiques spécifiques de la (dis)fluence avec des dimensions plurielles et plus abstraites de la (dis)fluence. En ce, il cherche à lier la **recherche appliquée** sur corpus, la **recherche expérimentale** et la **recherche théorique** conciliant production, perception et interprétation que nous développions en début de ce travail de thèse.

62. Voici, pour évaluer, le nombre de modalités et de paramètres mis en jeu dans notre design : (`5 degrés de fluence perçue`) + (`3 dimensions de la fluence`) + (`4 conditions de structure interne du stimulus`) + (`5 niveaux d'index F-IRI (* 4 dimensions)`) (1|`index stimuli audio`) + (1|`index participant`). Où « 1| » représente pour chaque facteur les effets aléatoires du modèle en prédisant différentes ordonnées à l'origine selon les paramètres du modèle. Chaque stimulus audio a ses particularités et sa variance, et chaque participant a sa manière d'évaluer la fluence et la disfluence.

3 Conclusions et discussions de l'évaluation des répétitions à l'identique sur les continuums fluence-disfluence

Dans ce chapitre, nous avons analysé les structures de répétitions en fonction de leurs variations situationnelles, de leurs cooccurrences, de leurs localisations syntaxiques et interactionnelles et de leurs formes acoustiques. Nous concluons sur la production des répétitions en mettant en exergue les relations entre les principaux résultats. Nous avons évalué les effets de ces différentes structures de répétitions (directes, associées à une pause pleine ou à une pause silencieuse) sur leur perception et leur évaluation par les auditeurs. Une première expérience a testé la capacité des répétitions à augmenter l'attention des locuteurs : à structure prosodique et syntaxique similaire, elle évalue l'interaction entre la variation des structures des répétitions à l'identique et les temps de réaction des locuteurs. Une seconde expérience ancre les répétitions de manière situationnelle ; elle invite les participants à l'évaluation holistique et componentielle des répétitions à l'identique en contexte et hors contexte. Nous résumons et discutons les principaux résultats de ces deux sections ci-après.

A travers le corpus, les répétitions s'actualisent au total **tous les 18.9 mots sur 1000, ce qui représente en moyenne 5.5% du temps d'articulation et 5% des unités syntaxiques**. Et nous avons révélé qu'à proportion de (dis)fluence égale, certaines situations privilégient l'occurrence de répétitions, c'est le cas des situations spontanées et (semi-)interactives. Nos recherches en production ont aussi montré qu'on effectue relativement plus de répétitions et celles-ci s'actualisent sur des mots plus courts en nombre de syllabes dans les situations préparées et (semi-)interactives. Pour autant, on ne répète pas des répétables plus courts en nombre de mots.

Au niveau acoustique, **les répétitions sans terme d'édition et suivies par une pause sont plus longues que les autres**. On soutient qu'elles ont une fonction de maintien de la parole lors de l'hésitation. Elles sont d'ailleurs plus caractéristiques d'une situation de parole spontanée. A l'appui de ce point de vue, les répétitions de MP (non clitiques) sont plus souvent suivies de temps de pause, c'est-à-dire d'un temps de reformulation conséquent. En résumé, plus le degré de préparation est élevé, moins il y a de répétitions de MP et moins il y a d'allongements du *reparans* par rapport au *reparandum*.

Au niveau fonctionnel, de nombreux éléments convergent en faveur de la **reconnaisance de la répétition comme un ligateur du flux discursif**. Ces éléments sont notamment d'ordre acoustique (comme mentionné ci-avant) mais aussi syntagmatiques (précédés et suivis majoritairement de mots lexicaux), syntaxiques (faible association avec des frontières d'unités syntaxiques fortes) et interactionnels (peu de passation de tour en fin de répétition). La répétition a tendance à initier unités syntaxiques les tours de parole. **Les fins de répétitions ne sont par ailleurs pas associées à des frontières syntaxiques majeures**.

Au niveau de la méthodologie de description des répétitions, **notre recherche a tenu à remettre en cause la distinction de trois modalités de longueur du répétable et du nombre de répétés**. Nos résultats d'analyses en composantes multiples ont effectivement mis en évidence que peu de distinctions de structure interne existent entre les répétitions caractérisées par un nombre de répétés double /vs/ multiple, ou entre les structures caractérisées par une longueur de répétable de 2 mots /vs/ > à 2 mots. Elle encourage les recherches suivantes à regrouper ces modalités.

De plus, bien que **près de la moitié des répétitions dans le corpus soient actualisées sans terme d'édition**, les corpus témoignent d'une grande variabilité des combinaisons de termes d'édition au sein de la phase d'édition (+ de 78 combinaisons). Les marques privilégiées des phases d'édition sont majoritairement la pause silencieuse, suivie des marqueurs discursifs et des troncations. Moins de 10% des répétitions avec phase d'édition sont ainsi concernées par une (dis)fluence ou par une pause pleine en phase d'édition. Nos expériences de perception ont permis de mettre en valeur l'influence (marginale) de ces termes d'édition sur la perception et l'évaluation des (dis)fluences.

En relation avec les données de production, la première expérience a testé l'hypothèse selon laquelle les auditeurs s'**attendent à un contenu relativement plus long et/ou plus complexe si un segment est précédé par une marque de (dis)fluence**. Nos résultats ne convergent pas entièrement avec les études précédentes : notre expérience n'a fait apparaître **aucune interaction** entre les facteurs de fluence et de complexité. Les répétitions et les termes d'édition ne semblent ainsi pas faciliter le traitement ni biaiser les attentes des auditeurs lorsque la complexité augmente, ce qui reste en accord avec les résultats de MacGregor, Corley & Donaldson (2009). Nos résultats montrent que les locuteurs francophones natifs **répondent de manière générale plus rapidement lorsque les items contiennent une répétition contiguë ou une répétition avec pause pleine**. Dans l'ordre décroissant, les items sans (dis)fluence, les répétitions avec pause silencieuse, les répétitions contiguës affichent des temps de latence plus longs que les répétitions avec pause pleine. Ces résultats confirment l'hypothèse selon laquelle les répétitions avec pause pleine n'ont pas un effet additif ou cumulatif. En accord avec ceux-ci, la répétition avec pause pleine encouragerait les auditeurs à **réagir plus vite**, ce qui peut donc être interprété comme **un outil pour capter leur attention plus que la marque d'un élément complexe à venir**.

Nous avions mis en avant les aspects bénéfiques des répétitions comme ligateurs discursifs et comme outils de facilitation du traitement des (dis)fluences. Nous avons par la suite tenté de comprendre l'influence du contexte et du participant sur le traitement de ces indices. Pour ce faire, nous avons également envisagé de réconcilier nos recherches fondamentales appliquées sur corpus avec notre cadre théorique considérant trois types de fluence (phrastique, situationnelle et sociale). Cette dernière expérience a révélé que c'était davantage la présence d'une répétition qui influençait son évaluation perceptive que sa structure interne. Une répétition avec une phase d'édition n'est donc pas jugée de manière plus sévère qu'une répétition sans phase d'édition et ce quel que soit son contexte de production. Néanmoins, les répétitions sont, de manière générale, évaluées

plus sévèrement dans des contextes à orientation médiatique que non médiatique. L'influence de ce contexte reste toutefois minime et ne permet pas de mettre en valeur quelles dimensions de la fluence (phrastique, situationnelle, sociale) sont le plus impactées par la présence d'un contexte particulier.

Enfin, nous avons tenté de poser dans nos analyses la variation intergenres et intralocuteur comme facteur essentiel de l'évaluation de la fluence-disfluence. Cette dernière expérience a permis de mettre en valeur que **le profil du participant influence son évaluation des stimuli**. Nous avons vu que différents types de fluences sont associés de manière plus ou moins forte avec le **degré du niveau d'implication empathique**, avec le **degré de détresse personnelle** ainsi qu'avec la mesure du degré d'imagination par le test de réactivité interpersonnelle F-IRI. Selon nous, le challenge principal de ce type d'approche est de concilier des hypothèses portant concrètement sur des marques phrastiques spécifiques (les structures de répétitions) avec une définition plus abstraite et pluridimensionnelle du continuum fluence-disfluence. Nous espérons que cette tentative nourisse la réflexion méthodologique sur le sujet.

Chapitre 7

Conclusion et discussion

L'action de recherche concertée dans laquelle s'inscrit cette thèse reconnaît que la fluence et la disfluence peuvent être **deux faces du même phénomène**, lequel s'incarne dans les marqueurs de (dis)fluence. Les formes et les fonctions de ces marqueurs varient, plus ou moins systématiquement, en fonction de la langue, des locuteurs et des types de discours concernés. Leur interprétation se module par rapport à ces mêmes critères. L'analyse de l'évaluation de ces (dis)fluences en fonction de ces indices linguistiques et contextuels a ainsi pour objectif de désambiguïser leur rôle dans un dispositif communicatif fluent ou disfluent. La mesure du degré de (dis)fluence dépend du contenu des énoncés (fluence phrastique) mais aussi des attentes des locuteurs par rapport à la situation (fluence situationnelle ou socio-interpersonnelle). Nous étudions ainsi ces marques linguistiques sur des données de production et de perception et cherchons à rendre compte de leurs variations **diaphasiques** (situations de communication), voire **diastratiques et idiostylistiques** (inhérentes au profil du locuteur).

Dans cette conclusions, nous résumons **les objectifs et les hypothèses** de ce travail de recherche (section 1). Nous abordons les défis d'une approche multidisciplinaire de la fluence et les difficultés de croiser des **méthodes plurielles** pour appréhender **les différentes facettes du continuum** fluence-disfluence. Nous soulignons ensuite les principaux **résultats méthodologiques, distributionnels, et fonctionnels** obtenus dans cette thèse (section 2). De notre discussion débouche une série de propositions pour les recherches portant sur une **approche contextualisée de l'évaluation de la fluence**, visant à concilier la **dimension théorique**, avec une **approche systématique des données**, nécessaire à toute analyse du discours, de l'interaction, et des corpus de parole (section 3).

1 Résumé des objectifs et des hypothèses

Ce travail de recherche doctorale se focalise sur **une définition de la (dis)fluence produite, perçue et interprétée** de discours **non pathologiques** produits par des **locuteurs francophones natifs** [1]. Lorsque la (dis)fluence n'est ni définie par un trouble pathologique, ni par un déficit langagier d'apprenant – comment peut-on la modéliser (détecter, définir, prédire et interpréter) au-delà d'une « expérience subjective d'aisance » avec laquelle les interlocuteurs traitent l'information ? Peut-on unifier les définitions et des types de fluence-disfluence évoqués (sociales, linguistiques, cognitives – produites et perçues – de manière située) observés dans la littérature d'un point de vue cognitif, psycholinguistique, discursif et computationnel, et si oui, comment ? Cette thèse pose un cadre théorique et méthodologique cherchant à répondre aux questions suivantes :

> Q1. Comment **définir** la fluence-disfluence pour des locuteurs natifs, au regard des différentes pratiques communicatives ?
>
> Q2. Quels sont les **paramètres essentiels à la catégorisation** des (dis)fluences sur le continuum fluence-disfluence ?
>
> Q3. Comment nos (dis)fluences s'agencent-elles **dans les discours** ?
>
> Q4. Quelles sont les structures qui entraînent une réelle **rupture** dans le discours et une **perception de *dis*fluence** en situation ?
>
> Q5. Quand est-ce qu'une **marque d'hésitation** est détectée, évaluée et catégorisée comme telle par les interlocuteurs ?
>
> Q6. Comment la **fréquence d'usage** et le **contexte d'occurrence** des (dis)-fluences (cooccurrences, paramètres prosodico-syntaxiques et situationnels) interagissent-ils en production et en perception ?

Les marques du travail d'élaboration, telles que les pauses silencieuses ou pleines, les répétitions, les amorces de mots et les marqueurs de discours, sont inhérentes à toute production de la parole. Ce travail réaffirme que la production et la perception d'« **aisance dans la parole** » ne se réduisent pas à la **linéarité ininterrompue du flux discursif** (i.e. à l'absence de (dis)fluence(s) ou de ruptures). Certaines « hésitations » semblent même acceptables, structurantes, voire **bénéfiques au traitement des discours dans des contextes donnés**. Nous cherchons à établir les « règles » ou **tendances qui régissent le fonctionnement de la (dis)fluence produite et perçue en contexte**.

Plus spécifiquement, le premier objectif de cette recherche est d'étudier l'interaction des **formes acoustiques** des (cooccurrences de) (dis)fluences avec les contextes prosodiques et syntaxiques. Nous avons cherché à éviter l'écueil d'une analyse soit trop holistique soit trop componentielle, qui omettent toutes deux la nature pragmatique et sociolinguistique de la communication.

1. Au contraire d'une majorité des recherches sur les disfluences qui se focalisent sur la parole pathologique ou non native, avec faible variation situationnelle.

Nous avons posé comme prémisse que le contexte de communication affecte la manière dont le discours prend forme et que, à nombre de marque(s) égal de (dis)fluence, certains discours paraissent plus fluides que d'autres. Notre recherche a donc intégré les **variations situationnelles et stylistiques** de l'évaluation des (dis)fluences, **par situation et parfois par locuteur**. Les analyses interprétatives ne tranchent ainsi pas dichotomiquement sur l'aspect fluent ou disfluent d'une marque. Au contraire, nous apportons de nombreux exemples de la **diversité d'interprétation** des (dis)fluences. Cette pluralité naît du fait que l'évaluation de la fluence est foncièrement dépendante des principes de ***self- and other- monitoring***. Or, les processus les gouvernant varient eux-mêmes en fonction (1) du **niveau d'attention** porté au discours et aux interlocuteurs, (2) de la conscience/**connaissance de la situation** et (3) des **niveaux d'empathie** de chacun. Synthétiquement, la production et la perception d'un discours comme *dis*fluent sont caractéristiques d'**une parole et d'une écoute (auto)-conscientes, surveillées et normatives**.

Notre second objectif théorique est de tester une nouvelle définition de la (dis)fluence.- Selon celle-ci : **la production et l'interprétation de (dis)fluences résultent de l'ajustement nécessaire entre les attentes liées à la situation et le discours effectivement produit, perçu et interprété**. La force de cet ajustement découlerait du degré de convergence/divergence avec nos attentes linguistiques, situationnelles ou interpersonnelles, et de l'adaptabilité des relations interlocuteurs-situations. Dans cette perspective, lorsque les attentes ne sont pas rencontrées, l'évaluation de la ***dis***fluence est fortement dépendante du niveau d'empathie et de flexibilité des discours et de leurs paramètres situationnels.

Notre troisième objectif résulte de cette réflexion théorique et vise à **formaliser cette nouvelle définition** de la disfluence : proposer des méthodes d'analyse prenant en compte sa production, sa perception et son interprétation par des locuteurs francophones natifs. Nous avons à ce titre enrichi cette réflexion théorique d'une proposition de **modèle de continuum** d'évaluation de la fluence à la (dis)fluence. Celui-ci envisage l'interprétation des marques de (dis)fluence selon (1) les contextes et (2) les attentes prises en compte par l'analyste. En d'autres termes, dans cette proposition, l'évaluation *dis*fluente consiste en une sur-/sous confirmation des attentes discursives des interlocuteurs, nécessitant un ajustement plus conséquent. Ce travail a donc proposé d'envisager le statut d'un phénomène sur le continuum fluence-disfluence en fonction de deux axes : (1) son **adéquation aux « normes discursives »** ou tendances générales et (2) son influence sur la **rupture/continuité discursive** (prosodique, syntaxique).

A fortiori, ce travail appelle un questionnement méthodologique plus large sur l'application de **méthodes mixtes en sciences humaines**. Notre recherche propose une **solution théorique** et pratique en croisant plusieurs méthodes de recherche. Nous avons mis « l'acceptabilité communicative de l'acte de parole produit en fonction **des attentes** des (inter)locuteurs-auditeurs » au centre de la définition l'évaluation de la fluence-disfluence.

Pour pouvoir définir ce que ces attentes peuvent être, nous avons chaque fois évalué l'homogénéité des pratiques communicatives et les variants et invariants de la marque de (dis)fluence en contexte (cooccurrences, distributions prosodique et syntaxique, variations individuelles et situationnelles).

Dans notre approche, cette contextualisation passe par la prise en compte d'abord de la relation entre les fréquences d'occurrence en production et la perception d'une continuité de l'axe syntagmatique et ensuite par la relation d'ajustement entre les attentes liées au genre et la perception-évaluation d'une production discursive. Pour autant, notre interprétation de la marque ne se limite pas à l'observation des corpus dans une approche fréquentielle. Nous avons effectivement émis l'**hypothèse** que l'évaluation d'une marque et/ou d'une production discursive par les auditeurs est associée aux fréquences objectives observées dans les corpus. Nous avons estimé le sens de cette association **à partir de tâches d'annotation et de tests perceptifs**, dont les résultats convergent ou non avec les **fréquences observées dans les corpus de production**. Un des enjeux de cette thèse a donc été de vérifier si l'homogénéité de production traduit une perception et une évaluation différentes des (dis)fluences. Ce travail précise au contraire que le raccourci « ce qui est le plus courant est le plus fluent » peut s'avérer dangereux. Il insiste sur le fait que les corpus sont une source d'évaluation des variations intra- et intersituationnelles.

2 Résumé des principaux résultats

Nous avons testé au sein de chaque chapitre d'analyse des hypothèses précises sur les formes des (cooccurrences de) (dis)fluences, ainsi que sur leurs interactions avec leurs distributions situationnelles, prosodiques et syntaxiques [2]. Nous présentons ici les principaux résultats des hypothèses testées pour chaque marque de (dis)fluences. Nous soulignons des résultats méthodologiques, distributionnels, et fonctionnels. Nous relevons notamment **où les marqueurs sont observés dans les corpus et où ils sont perçus en expérience**. Nos résultats ont attrait parfois à la **fluence phrastique produite**, et parfois aux **attentes des locuteurs**, à une fluence situationnelle et socio-interpersonnelle. Cette alternance illustre les deux défis et questionnements principaux de ce travail : l'interprétation issue de la **convergence de résultats** issus de données différentes, et le lien entre les **niveaux micro- et macro- analytiques** sont discutés à la section suivante.

Des méthodes mixtes répondant à des données multiples. Dans notre démarche, l'évaluation de la fluence perçue permet de réévaluer de manière motivée et de pondérer l'**importance de la fluence phrastique** effectivement produite. Nous avons donc

2. Méthodologiquement, nous avons tenu à annoter toutes les marques en surface, c'est-à-dire sans présumer à priori de leur(s) propriété(s) fonctionnelle(s). Pour rappel, les paramètres de description et de contextualisation pris en compte ont principalement été : la forme de la marque (durée, hauteur de F0, mouvement intonatif), le type de marque (simple ou composée), son contexte lexical (marque la précédant ou lui succédant), son environnement syntaxique (séquence fonctionnelle, unité de rection ou groupe accentuel), son environnement prosodique (frontières intonatives majeures, courbes mélodiques et contours intonatifs), la pratique communicative (la situation et les traits situationnels qui lui sont inhérents).

des objets d'études mixtes qui appellent des méthodes mixtes. Les données traitées reflètent cette démarche. Les corpus de production et de perception analysés dans cette thèse rassemblent **18 situations de parole** et plusieurs **tâches d'annotation et de perception** des hésitations par des **locuteurs experts et naïfs**. Notre étude a mis en évidence un certain nombre de régularités distributionnelles des marques d'hésitation, de leur perception et de leur interprétation.

Un double objectif de modélisation de la durée des pauses. L'évaluation de la production des pauses silencieuses en fonction de leur durée et de leur contexte d'actualisation suit un double objectif. Le premier est un objectif méthodologique, il est de proposer une modélisation de la durée des pauses de manière fidèle et sensible aux données. Il soutient le second objectif analytique qui est d'établir les variants et invariants de la durée des pauses en situation. Nous revenons sur ces points.

Le rejet des seuils et transformations uniques. Au risque de supprimer plus de 10% des pauses observables, nos résultats s'opposent, pour plusieurs raisons, à l'exclusion de pauses en fonction de seuils minimaux et maximaux uniques non motivés, c'est-à-dire « pris pour acquis ». Dans le même ordre d'idée, nous réfutons l'existence de la nécessité d'une catégorisation à priori des pauses en « pauses courtes » et « pauses longues ». Premièrement, le rejet arbitraire met en péril la comparabilité voire la validité des résultats des recherches. Deuxièmement, l'importante variabilité des débits entre les situations et locuteurs ne permet pas de prétendre qu'une pause de 100-200 ms, dans une situation pour un locuteur, a une valeur similaire partout ailleurs. Nous formulons aussi une mise en garde contre les artefacts mathématiques et méthodologiques gommant la variabilité interindividuelle. Nous valorisons les méthodes favorisant l'analyse et la catégorisation des pauses en fonction de plusieurs mesures : le logarithme doit s'associer à une mesure de débit de parole local, normalisé en fonction du contexte, et donc plus proche de la perception humaine.

La reconnaissance des variabilités intrasituationnelles et intersituationnelles. Notre objectif analytique de description des variants et invariants de la durée pausale fait émerger des différences situationnelles fortes. En fonction des locuteurs et des situations les distributions pausales sont tantôt unimodales, tantôt bimodales et, dans des cas plus rares, trimodales. Nous proposons de différencier les méthodologies d'analyse des pauses en fonction de cette variabilité intrasituationnelle. A l'observation des corpus, nos résultats révèlent que dans le cas de variabilité élevée dans un même genre de discours, il est plus intéressant de considérer des analyses distributionnelles par modèles mixtes[3]. Par ailleurs, les résultats de la modélisation multifactorielle confirment les recherches antérieures : la position des pauses par rapport aux unités syntaxiques et leurs cooccurrences au sein de chaînes de (dis)fluences sont des paramètres explicatifs robustes de prédiction de la durée des pauses.

3. Elles prennent en compte les traits du locuteur dans la paramétrisation des seuils de pauses minimaux et maximaux, des segmentations en unités séparées par des pauses et de la catégorisation éventuelle en pauses « courtes » /vs/ « longues ».

La durée de la pause augmente toujours en fonction de la frontière syntaxique quelle que soit la situation de parole. Les analyses mono-factorielles livrent tour à tour des éléments de la variation des fréquences et des durées des pauses silencieuses (1) par situation de parole et par locuteur, (2) selon leur localisation au sein d'une chaîne de (dis)fluences cooccurrentes à une pause, (3) selon la localisation syntaxique de la pause et (4) selon les contours intonatifs la précédant. Nos observations confirment la littérature antérieure en montrant que la durée moyenne d'une pause augmente en fonction du degré de frontière syntaxique [4].

Le profil des pauses courtes non structurantes. La durée de la pause augmente en fonction d'autres facteurs également. En ligne avec les résultats syntaxiques, la durée d'une pause augmente après un contour terminal /vs/ un contour continuatif. Par ailleurs, la durée des pauses croît particulièrement lorsque la pause est seule ou en cooccurrence avec des marqueurs de discours. L'augmentation de la durée en fonction de ces facteurs est caractéristique des pauses structurantes, en frontière d'unités. Nous confirmons par ailleurs que les pauses les plus courtes sont celles qui seraient qualifiées traditionnellement de « disfluentes ». Leur profil est d'être (1) plus brèves que les pauses structurantes, (2) internes aux groupes accentuels actualisés et (3) en cooccurrence avec d'autres marques de (dis)fluences.

Variation de la durée des pauses en fonction des cooccurrences. Nous avons pu distinguer un premier groupe caractérisé par des pauses relativement plus longues (*euh* + pause, pause + *euh*, pause en *interregnum*, et combinaison supérieure à 2 marques) d'un second regroupant des pauses relativement plus courtes (pause + dis, dis + pause, pause + md et md + pause, pause seule). Nos conclusions montrent que la chaîne (pause + *euh*) s'apparente davantage à un environnement disfluent que (*euh* + pause). Les combinaisons (*euh* + pause) et (md + pause) sont davantage continuatives et s'observent majoritairement en position inter-ga ou inter-rectionnelle. Les autres chaînes (dis + pause) ne sont pas porteuses de rupture syntaxique et encore moins marquées intonativement, elles entraînent davantage de continuité dans le flux discursif. Enfin, les chaînes de plus de 2 éléments en cooccurrence avec la pause sont souvent situées à la frontière entre deux unités syntaxiques (de rection) et associées à une rupture syntaxique, mais sans frontière intonative marquée.

Plus de distinctions que de similitudes pour les profils des pauses pleines et des allongements. Les allongements sont d'une durée supérieure aux pauses pleines et présentent moins de variation de durée. Selon la situation de parole, on observe que les pauses pleines sont plus longues en discours spontané que préparé, tandis que les allongements sont plus longs en discours préparé que spontané. Ces différences restent cependant minimes. Le peu de variation intra et intersituationnelle de la durée des pauses pleines et des allongements, ainsi que le caractère unimodal de leur distribution, incitent à ne pas distinguer des groupes de marques « courtes » /vs/ « longues ». Cependant, les résultats de nos analyses penchent en faveur de la reconnaissance de deux marques

4. e.g. il n'y a pas soudain une tendance à observer des pauses très courtes aux frontières d'unités de rections des discours spontanés.

distinctes entre la pause pleine et l'allongement, étant donné que ces deux marques n'ont ni les mêmes valeurs de durée, ni les mêmes cooccurrences. Par ailleurs, elles n'occupent pas les mêmes positions syntaxiques et n'exhibent pas les mêmes patrons d'amplitude mélodique.

Différences de durée en fonction des chaînes de pauses pleines et d'allongements. Un quart des pauses pleines et des allongements apparaissent en cooccurrence avec d'autres (dis)fluences. En cela, ces marques ont un potentiel moins structurant que les pauses silencieuses. La pause pleine s'associe particulièrement avec les pauses silencieuses et les marqueurs discursifs. La position privilégiée par les deux marques au sein des cooccurrences est différente : la pause pleine a tendance à être suivie de (dis)fluences, tandis que l'allongement a tendance à être précédé de (dis)fluences. Par ailleurs, certaines chaînes de cooccurrences ont effectivement une incidence sur la durée des pauses pleines et des allongements. On notera d'une part que les pauses pleines sont en général plus longues lorsqu'elles succèdent à une marque que lorsqu'elles la précèdent, et que d'autre part, les allongements sont particulièrement plus longs lorsqu'ils sont suivis d'un marqueur discursif ou précédés d'une (dis)fluence.

La pause pleine et l'allongement comme marque de cohésion. Une distribution complémentaire des marques s'observe en fonction de leur position syntaxique et au sein des unités de parole. La pause pleine est majoritairement positionnée entre des unités rectionnelles ou des groupes accentuels ; elle est « en dehors des unités syntaxiques (maximales) » et ne se trouve que rarement entre des constituants syntaxiques (annotés comme séquences fonctionnelles dans nos corpus). Au niveau des tours de parole, la pause pleine est majoritairement initiatrice de tours de parole ou en frontière de tour de parole ; même en cooccurrence avec d'autres éléments (dis)fluents, elle ne signale presque jamais une transition de tour. Les allongements sont s'actualisent au niveau microsyntaxique. Ils s'attachent à la gestion immédiate du discours et sont particulièrement plus longs lorsqu'ils sont en cooccurrence avec des (dis)fluences et placés entre les groupes accentuels, ce qui laisse à penser que leur actualisation est le signe d'un discours en élaboration. Au niveau interactionnel, on retrouve d'ailleurs davantage d'allongements en position médiane de tours de parole, à l'exception des cas où il sont en cooccurrence avec des marqueurs discursifs et des (dis)fluences.

Quatre critères pertinents à la détection de la pause pleine et de l'allongement d'hésitation. Deux tâches de perception *online* et *offline* des pauses pleines et des allongements ainsi que de leurs cooccurrences nous permettent de décrire les paramètres acoustiques et linguistiques pertinents à leur saillance perceptive. Les résultats permettent de mieux appréhender leur détection, leur distribution et leur classification. Les deux expériences de perception révèlent des accords intra et interannotateurs robustes lors de l'annotation des pauses pleines et des allongements. De manière générale, la présence d'une pause silencieuse aux alentours d'une marque la rend plus saillante et augmente sa probabilité de détection par l'annotateur humain dans les deux tâches d'annotation proposées. Hormis la durée relative de ces marques, les critères les plus

pertinents à leur catégorisation comme hésitation sont (1) la cooccurrence avec d'autres marques, (2) la présence d'une pause antéposée, (3) la position intraconstituant (de niveau groupe accentuel) et (4) la présence d'un mouvement intonatif plat ou descendant.

Le maintien de la parole par la pause pleine et l'allongement. En observant les corpus et en comparant des tâches d'annotation *online* et *offline*, nous avons mis en lumière des similarités contextuelles mais aussi des profils intonatifs différents pour les deux marques. Premièrement, les pauses pleines comme les allongements sont majoritairement précédés d'un contour continuatif, ce qui appuie leur fonction de maintien de la parole. Ensuite, l'étendue du mouvement syllabique des allongements est plus grand que celle des pauses pleines. Les allongements sont aussi en général mélodiquement plus dynamiques, particulièrement lorsqu'ils sont perçus comme des hésitations par le biais de tâches expérimentales. Ils accueillent plus de mouvements intonatifs montants que les pauses pleines. Cependant, dans nos observations, la majorité des deux marques exhibent peu de mouvement. Les contours étudiés sont majoritairement plats et stables, surtout pour les pauses pleines précédées par une pause silencieuse. On distinguera deux cooccurrences particulièrement associées à un contour descendant : la pause pleine lorsqu'elle est suivie d'un allongement, et l'allongement lorsqu'il est précédé d'une pause silencieuse. Le seul mouvement montant observé s'actualise sur la pause pleine lorsqu'elle est précédée d'une pause silencieuse et donc initiatrice d'une unité.

La répétition à l'identique produite différemment en fonction des situations. Nos recherches ont montré qu'à proportion de (dis)fluence égale, certaines situations privilégient davantage l'occurrence de répétitions. C'est le cas des situations spontanées et (semi-)interactives. De plus, on n'effectue pas nécessairement les mêmes répétitions en discours spontané que préparé. Plus le degré de préparation est élevé, moins il y a de répétitions de mots pleins et moins il y a d'allongements du *reparans* par rapport au *reparandum*. Par ailleurs, on soutient que les répétitions ont une fonction de ligateur lors de l'hésitation.

La répétition comme ligateur. Près de la moitié des répétitions dans le corpus sont actualisées sans terme d'édition. Les corpus témoignent d'une grande variabilité des combinaisons de termes d'édition au sein de la phase d'édition. Moins de 10% des répétitions en phase d'édition contiennent une (dis)fluence ou une pause pleine. Au niveau acoustique, les répétitions sans terme d'édition et suivies par une pause sont plus longues que les autres. Au niveau fonctionnel, de nombreux éléments convergent en faveur de la reconnaissance de la répétition comme un ligateur du flux discursif. Ces éléments sont notamment d'ordre acoustique mais aussi syntagmatique (précédés et suivis majoritairement de mots lexicaux), syntaxique (faible association avec des frontières d'unités syntaxiques fortes) et interactionnel (peu de passation de tour en fin de répétition). La répétition a davantage tendance à initier les unités syntaxiques et les tours de parole ; les fins de répétitions ne sont par ailleurs pas associées à des frontières syntaxiques majeures.

La répétition et la fluence cognitive. Une première expérience de perception révèle que les répétitions et termes d'édition semblent faciliter le traitement de la parole par les auditeurs mais n'influencent pas leur attentes sur la complexité du contenu à venir. Les résultats montrent que les auditeurs francophones natifs répondent de manière générale plus rapidement lorsque les items contiennent une répétition contiguë ou une répétition avec pause pleine. La répétition avec pause pleine encouragerait les auditeurs à réagir plus vite, ce qui peut donc être interprété comme un outil pour capter leur attention plus que comme la marque d'un élément complexe à venir.

La répétition est évaluée plus favorablement dans certaines situations. Une dernière expérience tente de comprendre l'influence du contexte et du profil de l'auditeur sur l'évaluation des (dis)fluences. Cette expérience, au design pourtant fort différent, a confirmé les résultats de la première : c'est davantage la présence d'une répétition qui influence l'évaluation d'un énoncé plutôt que la structure interne de cette répétition. Par ailleurs, en fonction du contexte, ces répétitions sont, de manière générale, évaluées plus sévèrement dans des contextes à orientation médiatique que non médiatique. Enfin, nous avons montré que différents types de fluence sont associés de manière plus ou moins forte au degré d'implication empathique, au degré de détresse personnelle ainsi qu'au degré d'imagination (mesuré par le test de réactivité interpersonnelle F-IRI). Dans la ligne de notre étude sur la pause silencieuse en début de ce travail de thèse, ces résultats marquent l'importance de l'interprétation des données en fonction des productions, mais aussi des perceptions individuelles.

Les résultats de cette thèse nous apprennent autant sur la rupture que sur la continuité dans le discours. Au niveau situationnel nous avons vu que le degré de préparation, l'orientation médiatique et le degré d'interactivité sesont souvent avéré être les indices les plus pertinents. Nous avons certainement évalué l'impact des marqueurs sur la production d'une **fluence phrastique linguistique** et conclu que peu de marques étaient fortement et uniquement associées à une rupture discursive. Nous avons confirmé que toutes les marques n'étaient pas interprétées comme disfluentes par les auditeurs. Nous avons insisté sur la nécessité de différencier dans les définitions de la fluence : l'évaluation de la fluence phrastique /vs/ l'évaluation de la fluence situationnelle ou interpersonnelle, cette dernière appelle à un contexte interne plus conséquent. Nos résultats montrent de manière intéressante que l'empathie et la compréhension d'une situation réduisent la sévérité de nos jugements. Nous pensons réellement que la recherche doit poursuivre l'étude des relations existantes entre ces traits de personnalité individuels et la perception des discours.

Nos données, et les résultats auxquels elles ont conduit, nous ont forcée à relever deux défis principaux. Le premier que nous évoquerons est le croisement d'observations portant concrètement sur des marques phrastiques spécifiques (de (dis)fluences) avec leur interprétation (plus conceptuelle) sur le continuum d'évaluation pluridimensionnel de la fluence-disfluence. Cette mise en relation dépend des contextes situationnels et des paramètres linguistiques considérés. En d'autres termes, comment lier les observations microlinguistiques et les considérations plus générales posées par la notion de « fluence »

dans les sciences humaines et sociales. Le second défi, en ligne avec le premier, est d'extraire des invariants du discours tout en ne gommant pas ce qui crée les particularités individuelles nécessaires à l'expression de la parole. En d'autres termes, cela pose la question de comment extraire des régularités tout en reconnaissant le besoin de variation. Nous espérons que nos analyses puissent nourrir la réflexion méthodologique à ce sujet. Notre discussion et nos perspectives vont dans ce sens.

3 Discussion et perspectives

Une discussion a constamment animé cette thèse. Elle confrontait **les manières d'écouter, d'annoter, d'analyser et de faire communiquer les données**. Une partie conséquente de notre travail a cherché à développer les techniques d'homogénéisation d'annotations et de **traitements intercorpus**[5]. D'autres sections ont insisté sur la **transparence des données**, en clarifiant notamment nos positions sur le traitement et la transformation des données « aberrantes » et leurs conséquences sur la reproductibilité et l'interprétation des résultats. La discussion a été jusqu'à tester les limites-mêmes de certains schémas d'annotation utilisés[6]. La contribution concrète de cette thèse s'illustre par ce travail de constitution et d'annotation de corpus ainsi que par les méthodes reproductibles de traitement des résultats[7]. Nous espérons que les débats qui ont émergé enrichissent la recherche sur la parole orale francophone en général et sur les relations entre prosodie et évaluation de la parole en situation en particulier.

Notre apport au sein du projet fluence-disfluence est un **accroissement des ressources** pour mieux comprendre comment **la prosodie et ses variations phonostylistiques interagissent avec les phénomènes d'hésitation**. Notre recherche a tenté de surmonter certaines limites généralement observées lors de l'étude de la (dis)fluence. Nous pensons notamment avoir (1) évité une **vision réductrice** d'un unique idéal de fluence parfait, mais aussi (2) saisi l'impact tout relatif d'un phénomène local sur la **perception globale** d'un discours entier. Pour orienter nos interprétations, nous avons (3) inclu les **paramètres situationnels** dans l'analyse des marqueurs de désambiguïsation des interprétations, (4) étendu notre réflexion au-delà d'un débat d'**association monofactorielle** unique de type forme-fonction-contexte. Au contraire, nous reconnaissons la **diversité des corpus**, **des locuteurs-auditeurs** et **des formes** étudiées. Pour relever ces défis, nous avons discuté de la possibilité de mettre en commun des littératures et des approches parfois très variées mêlant psycholinguistique expérimentale, linguistique computationnelle et sciences humaines et sociales.

Notre cadre théorique pluridisciplinaire et la définition multidimensionnelle de la fluence appelaient à l'emploi de méthodes mixtes. **Croiser les résultats** de recherches appliquées sur corpus, d'expériences d'annotation, de perception et d'évaluation est une **méthode d'opérationnalisation** pour répondre à cette demande. Si le croisement offre l'avantage d'une perspective multiple sur les données, il comporte **deux dangers**. Premièrement, l'on pourrait être séduit par une **sur-comparaison des résultats issus de différentes méthodes** : on ne peut comparer des données mesurant des choses différentes. Les résultats obtenus au moyen d'une méthode ne confirment pas ceux obtenus par

5. A côté de C-HUMOUR et LOCAS-F, nos analyses ont parfois impliqué plus de 5 corpus de parole ainsi que 6 sets de données issus d'expériences d'annotation et/ou de perception.

6. Ce fut notamment le cas dans Crible, Dumont, Grosman & Notarrigo (2018), où le protocole de (dis)fluences a été appliqué aux langues parlées et signées (anglais, français, langue des signes belge francophone) produites à la fois par des locuteurs autochtones et non natifs. Ce travail commun est lié au développement de 4 thèses du projet « ARC-Fluency – Evaluation croisée multilingue et multimodale de la (dis)fluence » (Crible 2017 ; Dumont 2018 ; Notarrigo 2017:, et le présent travail).

7. En témoignent notamment les Annexes A à E.

une autre, ils peuvent tout au plus les faire converger en les croisant. Deuxièmement, les approches et les objets mobilisés sont vastes. Ils croisent des niveaux d'analyse totalement différents. Du **conceptuel** au **microacoustique** en passant par le **macrosituationnel**, il est parfois périlleux pour le chercheur de poser une **conclusion proportionnée** et inébranlable.

Nous avons tenté d'atténuer ces risques par une analyse contextuelle **dont la granularité varie selon le questionnement**. Celle-ci mobilise une ou des dimension(s) spécifique(s) du continuum fluence-disfluence (phrastique, situationnel, socio-interpersonnel). En fonction, nos analyses étaient parfois situées à un niveau macrosituationnel et parfois ancrées dans un niveau syntagmatique local ou acoustique. Nous avons justement observé qu'au niveau intersituationnel, il est intéressant pour l'évaluation de la disfluence de considérer des **paramètres métalinguistiques** (descriptions d'attentes linguistiques, phonostylistiques), macrosyntaxiques ou interactionnels (syntaxe et tour de parole) et des **unités microlinguistiques** (les groupes accentuels, les relations acoustiques des segments de répétitions). Tous ces niveaux semblent pertinents à l'analyse de l'évaluation de la (dis)fluence. Nous avons délibérément choisi d'étudier de petites marques ayant trait à la **linguistique locale** (les marques de (dis)fluences) pour évaluer leur **impact à des niveaux plus larges de l'interaction**. Cet objet d'étude est intéressant à un niveau sémiotique et disciplinaire : les **petits indices métacognitifs** sont des précieux objets complexes de réflexion pour les **recherches métalinguistiques** et métacritiques.

Du local au global, du générique au particulier, ce cheminement est aussi un essai visant à concilier l'interprétation des **adaptations phonostylistiques quotidiennes** des interlocuteurs d'une part et, d'autre part, la prise en compte des profils de chaque locuteur-auditeur en puissance [8]. Nous encourageons les recherches ultérieures à étudier de manière individuelle, en considérant l'**influence des paramètres situationnels** sur le discours de chaque locuteur. Cette approche minimise le nombre de variables indépendantes étant donné que **les caractéristiques sont inhérentes aux locuteurs** [9].

Bien entendu, cet objectif analytique et méthodologique amène à discuter de la **taille de corpus nécessaire** à ce type d'analyse [10]. Les corpus exploités dans ce travail ont parfois encore été lacunaires ; notamment lorsque les cooccurrences de (dis)fluences précises apparaissaient **en faible nombre dans des contextes forts éloignés**. Pour pallier le manque actuel, nous encourageons la constitution de corpus reprenant de nombreuses fois le même locuteur dans une situation différente, comme le corpus phonogenre, le corpus PFC ou le corpus C-Humour. Selon nous, il ne faut pas neutraliser mais au contraire **valoriser la variable locuteur** lors de la constitution de corpus. Ce type de corpus a l'avantage de diminuer le nombre de variables indépendantes moins descriptibles. Enfin, quels que soient les particularismes individuels de production, soulignons que les

8. Une étude de cas sur C-Humour embrasse cette approche dans Grosman (2016).

9. Bien que chaque discours aura toujours sa variabilité en fonction par exemple de l'état de fatigue ou de l'humeur du locuteur ainsi que de la de la charge cognitive de la tâche effectuée.

10. Nous ne pouvons que spéculer sur la taille nécessaire des échantillons. En basant nos estimations sur C-Humour, qui reprend 4 humoristes dans 3 situations en 2.5 heures de paroles, l'analyse idiostylistique de la disfluence comme nous l'envisageons requérerait de tripler la taille du corpus actuel.

expériences de perception et d'annotation de ce travail révèlent des **taux d'accords** interjuges/interparticipants très honorables. Les auditeurs peuvent s'entendre sur ce qu'est la disfluence, ils témoignent de comportements réguliers. Nous avons également montré que de simples indices sur leur **profil d'empathie** influencent leur évaluation. Nous pensons que nos résultats, leur transparence et leur reproductibilité seront d'intérêt pour d'autres domaines de la recherche.

Seule la prise en compte de l'**interaction dans les discours** peut mener à une analyse intégrée de la fluence (monologique et dialogique). Nous discutons ici de l'intérêt de concevoir en parallèle **trois segmentations du discours** pouvant parachever les débats amorcés et enrichir les perspectives de recherche. Une approche affinée de ce croisement des annotations permettrait de contextualiser encore davantage les frontières d'unités analysées dans le présent travail. Nous proposons de croiser les analyses reprises dans ces corpus avec une **segmentation reconnaissant l'interaction des tours de parole, des unités interpausales (IPU) et des unités de base du discours**.

Dans cette thèse, nous avons commencé par analyser la durée des pauses silencieuses pour établir les seuils de segmentation pour les unités séparées par des pauses et les tours de parole. Ces seuils sont inhérents au profil idiostylistique ou aux caractéristiques phonostylistiques de la situation [11]. Ils permettent l'analyse interactionnelle des (dis)fluences aux niveaux intra et intersituationnels. Enfin, les unités de base du discours sont des informations précieuses sur le processus d'encodage et de décodage de la parole [12]. Maximiser la cohérence d'annotation entre ces trois types d'annotation/niveaux d'analyse permettrait une plus fine granularité des profils interactionnels des (dis)fluences en contexte : savoir quelles marques favorisent la prise de parole, évaluer les interruptions selon leur longueur, caractériser les temps de transition, etc.

Au-delà des débats sur les segmentations ou sur les interruptions, **un début alternatif** de la recherche serait de se focaliser sur la fluidité, un archétype de l'« extra-fluence » symbolisé par un **discours isochrone en flux continu**, exempt de marques. Il requérerait un traitement couteux de l'information : absence du regroupement d'unités, monotonie, **aucune forme ne ressortirait du fond**. Ce type de discours serait sans doute peu intelligible et/ou peu agréable à l'écoute. Il serait **dérangeant et disfluent**. Et on peut alors s'interroger : quelles sont les variations minimales des paramètres prosodiques et des (dis)fluences nécessaires à une **évaluation positive** du discours par les (inter)locuteurs ? Nous pourrions ainsi regarder le continuum de la fluence du côté « fluence » du spectre, et tester (notamment) des scores de fluence pour les unités du discours exemptes de marques mais qui ne répondent pas aux attentes [13].

11. Voir Annexe C section 22.

12. Nos essais de contextualisation des marques en fonction des unités de base du discours au sein desquelles elles s'actualisent sont ébauchés dans Grosman (2014).

13. Voir Loy, Rohde & Corley (2017) pour un exemple de design d'évaluation de la déception créée par un discours fluent.

Ces dernières conclusions sur **la segmentation et la fluidité** peuvent être utilisées en psycholinguistique. Ces données de durée et de cooccurrences peuvent être mises en œuvre pour approfondir l'impact des (dis)fluences sur les processus de décodage, d'attention, de mémorisation et d'évaluation de la parole. Nos recherches donnent aussi une **représentation des seuils et « des normes » dans la parole de francophones natifs**. Cette représentation de la variété acquise des usages, ces résultats peuvent asseoir également la **nécessité d'un enseignement reconnaissant les normes objectives** de (dis)fluences dans la pratique de la parole, intégrant l'usage des pauses et des marqueurs (Gilquin et al. 2008 ; Stenström 2011, 2014). De plus, décrire ces disfluences en parole non pathologique enrichit les données de comparaison pour la **parole pathologique** (**vieillissement langagier**, détection des cas précoces de démence) (Dubois-dindien 2017 ; Pinget, Rutger Bosker, Quené & de Jong 2014). Pour tous ces cas, la description de la disfluence comme « norme » de fluence à travers sa variation doit faire partie intégrante du système de description de la langue.

Comme amorcé dans notre cadre théorique et dans notre dernière expérience, ces recherches peuvent nourrir les réflexions sur la parole ou le **diagnostic d'empathie**, ainsi que sur le perfectionnisme. Les disfluences affectant l'interprétation pragmatique des messages, l'impact de la production des disfluences sur la **perception de l'honnêteté**, du mensonge, de la ruse ou de la nuance sont des domaines intéressés par les recherches expérimentales en la matière (Loy, Rohde & Corley 2017). Les résultats s'appliqueraient aux méthodes d'évaluation pour la synthèse de la parole. En témoignent les résultats convergents de Dall, Yamagishi & King (2014) relatifs à l'effet des attentes sur l'**évaluation de la naturalité de la parole de synthèse en situation**. L'évaluation des régularités de distributions (de durées et de cooccurrences) de (dis)fluences peut alors trouver des applications lors de recherches en traitement automatique de la parole. Nos conclusions permettent en effet de faire ressortir la probabilité qu'une marque soit jugée fluente ou disfluente, soit allongée ou non, en fonction des contextes prosodiques et syntaxiques. Les résultats sont intéressants pour les chercheurs en **linguistique fondamentale** et offrent des propositions pour des **approches pluriméthodologiques** des **sciences humaines** de manière plus générale.

Nous avons donc observé un phénomène, disséqué et croisé la littérature, conçu de nouvelles associations de méthodes, testé ces méthodes dans nos analyses, interprété des résultats, que nous avons essayés de transmettre à la communauté scientifique. Nous sommes convaincue : croiser des données d'origines différentes offre le luxe de répondre à des questionnements conceptuels comme concrets et matériels plus complexes. Nous comptons sur le lecteur pour approfondir les perspectives de recherche sur l'évaluation de la fluence et de la disfluence à laquelle nous avons contribué ici. Malgré certains passages disfluents, nous espérons que cette recherche a clairement mis en évidence que les marques du travail de formulation sont des indices précieux de la parole à plus d'un niveau ; et qu'il en faut beaucoup pour qu'un locuteur-auditeur, ou une auteure, interrompe l'élaboration de son discours.

Bibliographie

Adda, G. et al. 2007. « Speech overlap and interplay with disfluencies in political interviews ». *International Workshop on Paralinguistic Speech-between models and data, ParaLing 2007*, 41–46.

Adda-Decker, M. et al. 2008. « Annotation and analysis of overlapping speech in political interviews ». *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'11)*. European Language Resources Association (ELRA).

Adell, J., A. Bonafonte & D. Escudero-Mancebo. 2008. « On the generation of synthetic disfluent speech: Local prosodic modifications caused by the insertion of editing terms ». *Proceedings of Interspeech 2008, 9th Annual Conference of the International Speech Communication Association*. Brisbane, Australia : ISCA.

Akehurst, L., G. Kohnken, A. Vrij & R. Bull. 1996. « Lay persons' and police officers' beliefs regarding deceptive behaviour ». *Applied Cognitive Psychology* 10 (6), 461–471.

Allen, J. B. 1977. « Short-term spectral analysis, and modification by discrete Fourier Transform ». *IEEE Transactions on Acoustics Speech and Signal Processing*, 235–238.

Allopenna, P. D., J. S. Magnuson & M. K. Tanenhaus. 1998. « Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models ». *Journal of Memory and Language* 38 (4), 419–439.

Alter, A. L. 2013. « The Benefits of Cognitive Disfluency ». *Current Directions in Psychological Science* 22 (6), 437–442.

Alter, A. L. & D. M. Oppenheimer. 2007. « Overcoming intuition: Metacognitive difficulty activates analytic reasoning ». *Journal of Experimental Psychology : General* 136 (4), 569–576.

Alter, A. L. & D. M. Oppenheimer. 2009. « Uniting the Tribes of Fluency to Form a Metacognitive Nation ». *Personality and Social Psychology Review* 13 (3), 219–235.

Arciuli, J., D. Mallard & G. Villar. 2010. « "Um, I can tell you're lying": Linguistic markers of deception versus truth-telling in speech ». *Applied Psycholinguistics* 31 (3), 397–411.

Arnold, J. E., M. Fagnano & M. K. Tanenhaus. 2003. « Disfluencies signal theee, um, new information ». *Journal of psycholinguistic research* 32 (1), 25–36.

Arnold, J. E., A. Losongco, T. Wasow & R. Ginstrom. 2000. « Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering ». *Language* 76, 28–55.

Arnold, J. E. & M. K. Tanenhaus. 2011. « Disfluency isn't just um and uh: the role of prosody in the comprehension of disfluency ». *The processing and acquisition of reference*. Sous la dir. d'E. Gibson & N Permultter. Boston, MA : MIT Press, 197–218.

Arppe, A. et al. 2010. « Cognitive corpus linguistics: five points of debate on current theory and methodology ». *Corpora* 5 (1), 1–27.

Attardo, S., éd. 2014. *Encyclopedia of Humor Studies*. Thousand Oaks, CA : Sage Publications.

Auchlin, A., J. Goldman, A. C. Simon & M. Avanzi. 2018. « Pauses avec et sans prise de souffle: typologie acoustique et fonctionnelle ». *Typologie acoustique et fonctionnelle*. Sous la dir. d'E. Richard. Peter Lang, 57–72.

Auer, P. 2009. « On-Line Syntax: Thoughts on the Temporality of Spoken Language ». *Language Sciences* 31 (1), 1–13.

Auer, P. & S. Pfänder. 2007. « Multiple retractions in spoken French and spoken German. A contrastive study in oral performance styles ». *Cahiers de praxématique* (48), 57–84.

Avanzi, M. 2010. « Rattachement et fragmentation de la syntaxe par la prosodie ». *Travaux de linguistique* 60 (1), 145–166.

Avanzi, M., G. Bordal & N. Obin. 2011. « Typological Variations in the Realization of the French Accentual Phrase ». *Proceedings of ICPHS 17*. Honk Kong, Japan, 1–4.

Avanzi, M., A. Lacheret-Dujour, N. Obin & B. Victorri. 2011. « Vers une modélisation continue de la structure prosodique: le cas des proéminences syllabiques ». *Journal of French Language Studies* 21 (1), 53–71.

Avanzi, M., N. Obin, A. Bardiaux & G. Bordal. 2012. « Speech Prosody of French Regional Varieties ». *Proceedings of the 6th International Conference on Speech Prosody 2012*. Shanghai, China, May 22-25 2015 : Tongji University Press.

Avanzi, M., A. C. Simon, J. Goldman & A. Auchlin. 2010. « C-PROM : An Annotated Corpus for French Prominence Study ». *Proceedings of Prosodic Prominence, Speech Prosody 2010 Workshop*. Chicago, Illinois, USA.

Baayen, R. H., D. Davidson & D. M. Bates. 2008. « Mixed-effects modeling with crossed random effects for subjects and items ». *Journal of Memory and Language* 59 (4), 390–412.

Baayen, R. H. & P. Milin. 2010. « Analyzing reaction times ». *International Journal of Psychological Research* 3 (2), 12–28.

Bachy, S. et al. 2007. *Conventions de transcription régissant les corpus de la banque de données VALIBEL*. Rapp. tech. Université catholique de Louvain.

Barik, H. C. 1968. « On Defining Juncture Pauses: A Note On Boomer's "Hesitation and Grammatical Encoding" ». *Language and Speech* 11 (3), 156–159.

Barlow, M. & S. Kemmer, éds. 2000. *Usage Based Models of Language*. University of Chicago Press.

Baron, D., E. E. Shriberg & A. Stolcke. 2002. « Automatic punctuation and disfluency detection in multi-party meetings using prosodic and lexical cues ». T. 20. 61. Denver, Colorado, 949–952.

Barto, A., M. Mirolli & G. Baldassarre. 2013. « Novelty or Surprise? » *Frontiers in Psychology* 4 (907), [en ligne].

Bates, D. M., M. Mächler, B. Bolker & S. Walker. 2015. « Fitting Linear Mixed-Effects Models Using lme4 ». *Journal of Statistical Software* 67 (1), 1–48.

Bazillon, T. et al. 2008. « La parole spontanée: transcription et traitement ». *Revue TAL. Volume* 49 (3), 47–76.

Bazzanella, C. 2011. « Redundancy, repetition, and intensity in discourse ». *Language Sciences* 33 (2), 243–254.

Bear, J., J. Dowding & E. E. Shriberg. 1992. « Integrating Multiple Knowledge Sources for Detection and Correction of Repairs in Human-computer Dialog ». *Proceedings of the 30th Annual Meeting on Association for Computational Linguistics*. ACL '92. Stroudsburg, PA, USA : Association for Computational Linguistics, 56–63.

Beliao, J., S. Kahane & A. Lacheret. 2013. « Modéliser l'interface intonosyntaxique: ratio et synchronisation entre périodes intonatives et unités illocutoires ». *Actes d'IDP 2013*. Louvain, Belgique, 21–26.

Beliao, J. & A. Lacheret. 2013. « Disfluency and discursive markers: When prosody and syntax plan discourse ». *Proceedings of DiSS'06, Disfluency in Spontaneous Speech Workshop*. T. 54. Stockholm, Sweden : ISCA, 5–8.

Bell, A. 1984. « Language style as audience design. » *Language in Society* 13 (2), 145–204.

Bell, A. 2001. « Back in style: Reworking audience design ». *Style and sociolinguistic variation*. Sous la dir. de P. Eckert & J. R. Rickford. Cambridge : Cambridge University Press, 139–169.

Bell, L., R. Eklund & J. Gustafson. 2000. « A comparison of disfluency distribution in a unimodal and a multimodal speech interface. » *Proceedings of the Sixth International Conference on Spoken Language (ICSLP 2000)*. Beijing, China : ISCA, 626–629.

Belz, M. & U. Reichel. 2015. « Pitch characteristics of filled pauses ».

Berthold, A. & A. Jameson. 1999. « Interpreting Symptoms of Cognitive Load in Speech In-put. » *Proceedings of the Seventh International Conference on User Modeling, UM99 User Modeling, 20–24 June 1999*. Sous la dir. de J. Kay. Banff, Canada. Vienna. New York : Springer, 235–244.

Bertrand, R., G. Ferré & M. Guardiola. 2013. « French Face-to-Face Interaction: Repetition as a Multimodal Resource ». *Coverbal synchrony in Human-Machine Interaction*. Sous la dir. de M. R. N. Campbell. Science Publishers/CRC Press, 141–172.

Bertrand, R. et al. 2008. « Le CID-Corpus of Interactional Data-Annotation et exploitation multimodale de parole conversationnelle ». *Traitement Automatique des Langues* 49 (3), 1–30.

Bertrand, R. 1996. *The Collected Papers of Bertrand Russell - Vol. 10 A Fresh Look at Empiricism: 1927-42*. London et New York : Routledge.

Besser, J. & J. Alexandersson. 2007. « A Comprehensive Disfluency Model for Multi-Party Interaction ». *Proceedings of SIGDial 2007*. Antwerp, Belgium, 182–189.

Bigi, B., C. Meunier, R. Bertrand & I. Nesterenko. 2010. « Annotation automatique en syllabes d'un dialogue oral spontané ». *XXVIIIèmes Journées d'étude sur la Parole*. Mons, Belgique, 181–184.

Bilger, M. & E. Campione. 2002. « Proposition pour un étiquetage en séquence fonctionnelles ». *Recherches sur le français parlé* 17, 117–136.

Blackmer, E. R. & J. L. Mitton. 1991. « Theories of monitoring and the timing of repairs in spontaneous speech ». *Cognition* 39 (3), 173–194.

Blakemore, S., J. Winston & U. Frith. 2004. « Social cognitive neuroscience: where are we heading? » *Trends in Cognitive Sciences* 8 (5), 216–222.

Blanche-Benveniste, C. 1984. « La dénomination dans le français parlé: une interprétation pour les répétitions et les hésitations ». *Recherches sur le Français Parlé Aix-en-Provence* (6), 109–130.

Blanche-Benveniste, C. 1987. « Syntaxe, choix de lexique, et lieux de bafouillage ». *Documentation et Recherche en linguistique Allemande Contemporaine* 37-38, 123–157.

Blanche-Benveniste, C. 1990a. « Présentation des travaux du G.A.R.S. » *Travaux de linguistique*. La linguistique de l'oral 21, 42–45.

Blanche-Benveniste, C. 1990b. « Un modèle d'analyse syntaxique 'en grilles' pour les productions orales ». *Anuario de Psicologia Liliane Tolchinsky* 47, 11–28.

Blanche-Benveniste, C. 1997. *Approches de la langue parlée en français*. Paris : Ophrys.

Blanche-Benveniste, C. 2003. « Le recouvrement de la syntaxe et de la macro-syntaxe ». *Macro-syntaxe et pragmatique. L'anlayse linguistique de l'oral*. Sous la dir. d'A. Scarano. Università deglo studi di Firezi : Bulzoni, 53–75.

Blanche-Benveniste, C. 2010. *Approches de la langue parlée en français - Nouvelle édition*. Gap u.a. : Ophrys.

Blanche-Benveniste, C. & C. Jeanjean. 1986. *Le français parlé - Transcription et édition*. Paris : Didier Erudition.

Blanche-Benveniste, C. et al. 1979. « Des grilles pour le français parlé ». *Recherche sur le français parlé* 2, 123–206.

Blumer, H. 1986. *Symbolic interactionism: perspective and method*. Berkeley : University of California Press.

Boersma, P. & D. Weenink. 2015. *Praat: doing phonetics by computer [Computer Program]*. url : http://www.praat.org/.

Bolly, C. 2010. « Flou phraséologique, quasi-grammaticalisation et pseudo marqueurs de discours : un no man's land entre syntaxe et discours ? » *Linx. Revue des linguistes de l'université Paris X Nanterre* (62-63), 11–38.

Bolly, C. & D. Boutet. en prep. « The multimodal CorpAGEst corpus: Keeping an eye on pragmatic competence in later life. » *Corpora* 13 (2), [Draft paper].

Bolly, C., L. Crible, L. Degand & D. Uygur-Distexhe. 2015. « MDMA. Un modèle pour l'identification et l'annotation des marqueurs discursifs « potentiels » en contexte ». *Discours. Revue de linguistique, psycholinguistique et informatique.A journal of linguistics, psycholinguistics and computational linguistics* (16).

Boomer, D. S. 1965. « Hesitation and grammatical encoding ». *Language and speech* 8 (3), 148–158.

Bortfeld, H. et al. 2001. « Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender ». *Language and Speech* 44 (2), 123–147.

Bosker, H. R., H. Quené, T. Sanders & N. H. de Jong. 2014. « The Perception of Fluency in Native and Nonnative Speech: Native and Nonnative Fluency Perception ». *Language Learning* 64 (3), 579–614.

Bosshardt, H. 2002. « Effects of concurrent cognitive processing on the fluency of word repetition: comparison between persons who do and do not stutter ». *Journal of Fluency Disorders* 27 (2), 93–114.

Bothuyne, M. 2013. « Profil de fluence d'élèves du secondaire : de l'analyse linguistique à l'évaluation par l'enseignant ». Mémoire de Master. Louvain-la-Neuve, Belgique : Université catholique de Louvain.

Boula de Mareüil, P. 2014. « Qu'est-ce qu'un (phono) style? » *Cahiers de linguistique française* (31), 9–19.

Boula de Mareüil, P. et al. 2005. « A quantitative study of disfluencies in French broadcast interviews ». *Proceedings of DiSS'2005, Disfluency in Spontaneous Speech Workshop*. Aix-en-Provence, France : ISCA, 27–32.

Boula de Mareüil, P. et al. 2013. « Une étude quantitative des marqueurs discursifs, disfluences et chevauchements de parole dans des interviews politiques ». *TIPA. Travaux interdisciplinaires sur la parole et le langage* (29).

Bouraoui, J. M. 2008. « Analyse, modélisation et détection automatique des disfluences dans le dialogue oral spontané contraint: le cas du contrôle aérien ». Thèse de Doctorat. Université de Toulouse, Université Toulouse III-Paul Sabatier.

Bouraoui, J. M. & N. Vigouroux. 2006. « Etude de dysfluences dans un corpus linguistiquement contraint ». *Journée d'Etude sur la Parole (JEP2006)*. Dinard, France, 429–432.

Branca-Rosoff, S. 1999. « Types, modes et genres : entre langue et discours ». *Langage et société* 87 (1), 5–24.

Brennan, S. E. & M. William. 1995. « The feeling of another's knowing. Prosody and filled pauses as cues to listeners ». *Journal of Memory and Language* 34, 383–398.

Broen, P. A. & G. M. Siegel. 1972. « Variations in normal speech disfluencies ». *Language and Speech* 15 (3), 219–231.

Brunken, R., J. L. Plass & D. Leutner. 2003. « Direct Measurement of Cognitive Load in Multimedia Learning ». *Educational Psychologist* 38 (1), 53–61.

Bybee, J. L. 2006. « From Usage to Grammar: The Mind's Response to Repetition ». *Language* 82 (4), 711–733.

Byrd, C. T., M. Vallely, J. D. Anderson & H. Sussman. 2012. « Nonword repetition and phoneme elision in adults who do and do not stutter ». *Journal of Fluency Disorders* 37 (3), 188–201.

Campione, E. & J. Véronis. 2002. « A Large-Scale Multilingual Study of Silent Pause Duration ». *Proceedings of the Speech Prosody 2002 Conference*. Sous la dir. de B. Bel & I. Marlien. Aix-en-Provence, France : ISCA, 199–202.

Campione, E. & J. Véronis. 2004. « Pauses et hésitations en français spontané ». *Actes des XXVèmes Journées d'Etude sur la Parole (JEP'2004)*. Fes, Maroc : ISCA-AFCP, 1–4.

Campione, E. & J. Véronis. 2005. « Pauses and hesitations in French spontaneous speech ». *Proceedings of DISS 05, Disfluency in spontaneous speech workshop*. Aix-en-Provence, France : ISCA, 43–46.

Candea, M. 2000a. « Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané. Etude sur un corpus de récits en classe de français. » Thèse de doctorat. Université de la Sorbonne nouvelle - Paris III.

Candea, M. 2000b. « Les euh et les allongements dits d'hésitation deux phénomènes soumis à certaines contraintes en français oral non lu ». *Actes des XXIIIèmes Journées d'Etude sur la Parole (JEP'2000)*. Aussois, France, 73–76.

Candea, M. 2000c. « Répétition, altération, reformulation dans les textes et discours: Colloque international, [Besançon], 22-24 juin 1998 ». *Annales littéraires de l'Université de Franche-Comté*. T. Répétition, altération, reformulation. Série Linguistique et Sémiotique. Presses Univ. Franche-Comté, 119–130.

Candea, M., I. Vasilescu & M. Adda-Decker. 2005. « Inter-and intra-language acoustic analysis of autonomous fillers ». *Proceedings of DISS 05, Disfluency in Spontaneous Speech Workshop*. Aix-en-Provence, France : ISCA, 47–52.

Carton, F. 1999. « L'épithèse vocalique en français contemporain : étude phonétique ». *Faits de langues* 7 (13), 35–45.

Chafe, W. 1980. *The pear stories : cognitive, cultural, and linguistic aspects of narrative production*. Advances in discourse processes 3. Norwood : Ablex.

Chafe, W. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago et London : University of Chicago Press.

Chambers, J. 1997. « What do we mean by fluency? » *System* 25 (4), 535–544.

Cholin, J. & W. J. M. Levelt. 2009. « Effects of syllable preparation and syllable frequency in speech production: Further evidence for syllabic units at a post-lexical level ». *Language and Cognitive Processes* 24 (5), 662–684.

Christodoulides, G. 2014. « Praaline: Integrating Tools for Speech Corpus Research. » *Proceedings of the 9th International Conference on Language Resources and Evaluation - LREC'14*. Reykjavik, Iceland : European Language Resources Association (ELRA), [en ligne].

Christodoulides, G. 2016. « Effect of Cognitive Load on Speech Production and Perception ». Thèse de doct. Louvain-la-Neuve, Belgique : Université catholique de Louvain.

Christodoulides, G., M. Avanzi & J. Goldman. 2014. « DisMo: A Morphosyntactic, Disfluency and Multi-Word Unit Annotator. An Evaluation on a Corpus of French Spontaneous and Read Speech ». *Proceedings of the 9th International Conference on Language Resources and Evaluation - LREC'14*. Reykjavik, Iceland : European Language Resources Association (ELRA), [en ligne].

Christodoulides, G., I. Grosman, L. Degand & A. C. Simon. 2017. « Methodologies for modelling silent pause length. Insights on individual and situational variation from a large-scale corpus study ». Louvain-la-Neuve, Belgique.

Clark, A. 2013. « Whatever next? Predictive brains, situated agents, and the future of cognitive science ». *The Behavioral and Brain Sciences* 36 (3), 181–204.

Clark, E. V. & J. Bernicot. 2008. « Repetition as ratification: How parents and children place information in common ground ». *Journal of Child Language* 35 (2), 349–371.

Clark, H. H. 1994. « Managing problems in speaking ». *Speech Communication* 15 (3-4), 243–250.

Clark, H. H. 1996. *Using Language*. Standford Univeristy, California : Cambridge University Press.

Clark, H. H. 2002. « Speaking in time ». *Speech Communication* 36, 5–13.

Clark, H. H. & E. V. Clark. 1977. *Psychology and Language: An Introduction to Psycholinguistics*. Harcourt, Brace, Jovanovich. New York.

Clark, H. H. & J. E. F. Tree. 2002. « Using uh and um in spontaneous speaking ». *Cognition* 84 (1), 73–111.

Clark, H. H. & T. Wasow. 1998. « Repeating Words in Spontaneous Speech ». *Cognitive Psychology* 37 (3), 201–242.

Claverie, B. 2010. « Pluri-, inter-, transdisciplinarité : ou le réel décomposé en réseaux de savoir ». *Projectics / Proyéctica / Projectique* 4 (1), 5.

Clerc-Renaud, J., I. Vasilescu, M. Candea & M. Adda-Decker. 2004. « Étude acoustique et perceptive des hésitations autonomes multilingues ». *Actes des XXVèmes Journées d'Etude sur la Parole (JEP'2004)*. Fes, Maroc.

Cole, J. S. et al. 2005. « Prosodic parallelism as a cue to repetition and Hesitation disfluency ». *Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop*. Aix-en-Provence, France : ISCA, 53–58.

Corley, M. 2017. « What do listeners think of disfluencies? » Louvain-la-Neuve, Belgium.

Corminboeuf, G. & C. Benzitoun. 2014. « Approches non gradualistes des liaisons de prédications : l'exemple des modèles macro-syntaxiques ». *Langue française* (182), 43–58.

Cowan, J. M. & B. Bloch. 1948. « An Experimental Study of Pause in English Grammar ». *American Speech* 23 (2), 89.

Crible, L. 2014. « Crible_2014_Disfluency in low-interactivity situations My contribution.pdf ».

Crible, L. 2017. « Discourse markers and (dis)fluency across registers: A contrastive usage-based study in English and French ». Thèse de doct. Louvain-la-Neuve, Belgique : Université catholique de Louvain.

Crible, L., L. Degand & G. Gilquin. 2017. « The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis)fluency ». *Languages in Contrast* 17 (1), 69–95.

Crible, L., A. Dumont, I. Grosman & I. Notarrigo. 2014. *Set d'étiquettes des données et métadonnées*. Rapport technique. Technical report, Université catholique de Louvain & Université de Namur.

Crible, L., A. Dumont, I. Grosman & I. Notarrigo. 2016. *Annotation manual of fluency and disfluency markers in multilingual, multimodal, native and learner corpora. Version 2.0.* Technical Report. Université catholique de Louvain et Université de Namur.

Crible, L., A. Dumont, I. Grosman & I. Notarrigo. 2018. « (Dis)fluency across spoken and signed languages : Application of an interoperable annotation scheme ». *Corporal and Language in Use – Proceedings 4*.

Crystal, D. 1988. *The Cambridge encyclopedia of language*. Repr. Cambridge : Cambridge university press.

Cucchiarini, C., H. Strik & L. Boves. 2000. « Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology ». *The Journal of the Acoustical Society of America* 107 (2), 989–999.

Cutler, A., D. Dahan & W. Van Donselaar. 1997. « Prosody in the comprehension of spoken language: A literature review ». *Language and Speech* 40 (2), 141–201.

Dall, R., J. Yamagishi & S. King. 2014. « Rating Naturalness in Speech Synthesis: The Effect of Style and Expectation ». *Proceedings Speech Prosody, Dublin, Ireland*. Dublin, Ireland.

Davis, M. H. 1980. « A multidimensional approach to individual differences in empathy ». *JSAS Catalog of Selected Documents in Psychology* 10, 85–99.

Davis, M. H. 1983. « Measuring individual differences in empathy: Evidence for a multidimensional approach ». *Journal of personality and social psychology* 44 (1), 113–126.

De Jong, N. H. & H. R. Bosker. 2013. « Choosing a threshold for silent pauses to measure second language fluency ». *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech (DiSS)*. Sous la dir. de R. Eklund. ISCA, 17–20.

De Cock, S. et al. 2004. « Preferred sequences of words in NS and NNS speech ». *Belgian Journal of English Language and Literatures (BELL)* 2, 225–46.

de Medeiros, H. 2014. « Automatic Detection of Disfluencies in a Corpus of University Lectures ». Thèse de doct. University Institute of Lisbon. School of Technology et Architecture.

Debaisieux, J. & J. Deulofeu. 2001. « Grammatically unacceptable utterances are communicatively accepted by native speakers, why are they? » *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech, DiSS'01*. Edinburgh, Scotland, UK, 69–72.

Degand, L. 2016. « Spoken discourse segmentation and the paradox of discourse markers. » *L&C Talk*. MPI. Nijmegen.

Degand, L., J. Eychenne, A. C. Simon & T. Van Damme. 2011. « L'interprétation des hésitations dans la délimitation des unités de base du discours ». *Conférence internationale : (Dés)organisation de l'oral*. Rennes, France.

Degand, L. & G. Gilquin. 2013. « The clustering of 'fluencemes' in French and English ». fr.

Degand, L., L. J. Martin & A. C. Simon. 2011. « Unités discursives de base et leur périphérie gauche dans LOCAS-F, un corpus oral multigenres annoté ». *SHS Web of Conferences*. T. 8. EDP Sciences, 2613–2625.

Degand, L. & A. C. Simon. 2005. « Minimal Discourse Units: Can we define them, and why should we ». *Proceedings of SEM-05. Connectors, discourse framing and discourse structure : from corpus-based and experimental analyses to discourse theories*, 65–74.

Degand, L. & A. C. Simon. 2009. « On identifying basic discourse units in speech: theoretical and empirical issues ». *Discours* (4), [en ligne].

Degand, L. & G. van Bergen. 2016. « Discourse Markers as Turn-Transition Devices: Evidence From Speech and Instant Messaging ». *Discourse Processes : a multidisciplinary journal* 00 (0). Routledge (Philadelphia), 1–25.

Delais-Roussarie, E. 1999. « Accentuation et réalisation des clitiques en français ». *Cahiers de Grammaire* 24, 17–37.

Delais-Roussarie, E. 2000. « Vers une nouvelle approche de la structure prosodique ». *Langue française* 126 (1), 92–112.

Delattre, P. 1966. « Les Dix Intonations de base du français ». *The French Review* 40 (1), 1–14.

Demol, M., W. Verhelst & P. Verhoeve. 2007. « The duration of speech pauses in a multilingual environment. » *Proceedings of Interspeech 2007, 8th Annual Conference of the International Speech Communication Association*. Antwerp, Belgium : ISCA, 990–993.

Denke, A. 2005. *Nativelike Performance: A Corpus Study of Pragmatic Markers, Repair and Repetition in Native and Non-native English Speech*. Stockholm : Department of English, Stockholm University.

DePaulo, B. M. et al. 2003. « Cues to deception. » *Psychological Bulletin* 129 (1), 74–118.

Di Cristo, A. 1985. *De la microprosodie à l'intonosyntaxe*. Broché, Université de Provence.

Di Cristo, A. 2011. « Une approche intégrative des relations de l'accentuation au phrasé prosodique du français ». *Journal of French Language Studies* 21 (1), 73–95.

Di Cristo, A. 2013. *La prosodie de la parole*. Voix, parole, langage. De Boeck Supérieur.

Dister, A. 2007. « De la transcription à l'étiquetage morphosyntaxique. Le cas de la banque de données textuelles orales VALIBEL ». Thèse de doctorat. Université catholique de Louvain.

Dister, A. 2014. « « parler sans accent pour moi c'est sans sans sans bafouiller » Quelles répétitions de formes en français parlé ? » *SHS Web of Conferences* 8, 2017–2031.

Dister, A. & A. Lacheret-Dujour. 2011. « Annotation intonosyntaxique de corpus oraux ». *Ecole thématique annotation de données langagières*. Biarritz, France.

Divjak, D. 2016. « The Role of Lexical Frequency in the Acceptability of Syntactic Variants: Evidence From that-Clauses in Polish ». *Cognitive Science* 41 (2), 354–382.

Doutrich, D. 2000. « Cultural Fluency, Marginality and the Sense of Self ». *Perspectives on fluency*. Sous la dir. de H. Riggenbach. Ann Arbor, MI : University of Michigan Press, 141–159.

Dowding, J. et al. 1994. « GEMINI: A Natural Language System for Spoken-Language Understanding ». *CoRR* abs/cmp-lg/9407007. arXiv : `cmp-lg/9407007`.

Duboisdindien, G. 2017. « Analyse multimodale des marqueurs pragmatiques à fonction intersubjective au sein du vieillissement langagier, en tant qu'indices précoces de démence ». Thèse de doctorat. Modyco, Université Paris Ouest Nanterre, Paris (France).

Duez, D. 1982. « Silent and Non-Silent Pauses in Three Speech Styles ». *Language and Speech* 25 (1), 11–28.

Duez, D. 1985. « Perception of silent pauses in continuous speech ». *Language and Speech* 28 (4), 377–389.

Duez, D. 1991. *La Pause dans la parole de l'homme politique*. Paris : Editions du CNRS.

Duez, D. 1993. « Acoustic correlates of subjective pauses ». *Journal of Psycholinguistic Research* 22 (1), 21–39.

Duez, D. 1995. « Perception of hesitations in spontaneous french speech ». *Proceedings of ICPhS 95*. Sous la dir. de K. Elenius & P. Auer. T. 2. Stockholm, Sweden, 498–501.

Duez, D. 1997. « Effet de l'interaction de la vitesse d'élocution et de la localisation dans le syntagme et l'énoncé sur la durée vocalique en français lu ». *Proceedings of Linguistic Studies Workshop Vowels of All Shapes, Kinds and Sizes*. Nantes, France, 124–130.

Duez, D. 1998. « How articulation rate and position in utterance and phrase affect segmental duration : Within- and cross-subject variability ». *Proceedings of Speaker Recognition and its Commercial and Forensic Application*. Avignon, France, 16–20.

Duez, D. 1999. « Effects of articulation rate on duration in read French speech ». *Proceedings of Eurospeech '99*. Budapest, Hungary : ISCA, 715–718.

Duez, D. 2001a. « Acoustic-phonetic characteristics of filled pauses in spontaneous French speech: preliminary results ». *Proceedings of DiSS'01, Disfluency in Spontaneous Speech Workshop*. Edinburg, UK : ISCA, 41–44.

Duez, D. 2001b. « Caractéristiques acoustiques et phonétiques des pauses remplies dans la conversation en français ». *Travaux interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence (TIPA)* 20, 31–48.

Duez, D. 2001c. « Signification des hésitations dans la production et la perception de la parole spontané ». *Revue Parole* 17-19, 113–138.

Dumont, A. 2018. « Fluency and disfluency: a corpus study of non-native and native speaker (dis)fluency profiles. » Thèse de doct. UCL - Faculté de philosophie, arts et lettres.

Dutoit, T. 2013. *An Introduction to Text-to-Speech Synthesis*. Springer Science & Business Media.

Dutrey, C. et al. 2014. « Disfluences dans la parole spontanée conversationnelle: détection automatique utilisant des indices lexicaux et acoustiques ». *Actes des XXXe Journées d'Étude sur la Parole (JEP'14)*. Le Mans, France : 366-373.

Duvivier, E., G. Christodoulides & I. Grosman. 2017. « Comparing Online and Offline Methodologies for Spoken Corpus Annotation. The case of Filled Pauses and Hesitation-Related Lengthening in French ». EN. Heraklion, Crete, Greece.

Edlund, J., J. B. Hirschberg & M. Heldner. 2009. « Pause and gap length in face-to-face interaction ». *Proceedings of Interspeech 2009, 10th Annual Conference of the International Speech Communication Association*. Brighton, UK : ISCA, 2779–2782.

Eklund, R. 2004. « Disfluency in Swedish human-human and human-machine travel booking dialogues ». PhD Thesis. Linköping : Linköping studies in science et technology.

Eklund, R. & E. E. Shriberg. 1998. « Crosslinguistic disfluency modelling: a comparative analysis of Swedish and american English human-human and human-machine dialogues. » *Proceedings of the Fifth International Conference on Spoken Language (ICSLP 98)*. T. 98. Citeseer, 2631–2634.

Eskenazi, M. 1993. « Trends in speaking styles research ». *Proceedings of Eurospeech '93*. T. 1. Berlin, Germany : ISCA, 501–509.

Estes, W. K. 1994. *Classification and Cognition*. New York/Oxford : Oxford University Press.

Faygal, Z. 1999. « Combien de clichés mélodiques ? Révision de l'inventaire des contours intonatifs stylisés en français ». *Faits de langues* 13, 17–25.

Ferreira, F. 1993. « Creation of prosody during sentence production. » *Psychological review* 100 (2), 233.

Fillmore, C. J. 2000. « On Fluency ». *Perspectives on Fluency*. Sous la dir. de H. Riggenbach. Ann Arbor, MI : University of Michigan Press, 43–61.

Fischer, K. 2000. *From Cognitive Semantics to Lexical Pragmatics: The Functional Polysemy of Discourse Particles*. Berlin, Boston : De Gruyter.

Fleiss, J. L. 1971. « Measuring nominal scale agreement among many raters. » *Psychological Bulletin* 76 (5), 378–382.

Fleiss, J. L., J. Cohen & B. S. Everitt. 1969. « Large sample standard errors of kappa and weighted kappa ». *Psychological Bulletin* 72, 323–327.

Fónagy, I. 1989. « Le français change de visage? » *Revue Romane* 24 (2), 225–254.

Fors, K. L. 2011. « An investigation of intra-turn pauses in spontaneous speech ». *Fonetik TMH - QPSR* 51, 35–68.

Fossard, M. et al. 2018. « Referential Choices in a Collaborative Storytelling Task: Discourse Stages and Referential Complexity Matter ». *Frontiers in Psychology* 9 (176).

Fox Tree, J. E. 1995. « The effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech ». *Journal of Memory and Language* 34 (6), 709–738.

Fraisse, P. 1974. *Psychologie du rythme*. Paris, France : Presses Universitaires de France (PUF).

Fraley, C. & A. E. Raftery. 2007. « Model-based methods of classification: using the mclust software in chemometrics ». *Journal of Statistical Software* 18 (6), 1–13.

Frederick, S. 2005. « Cognitive reflection and decision making ». *Journal of Economic perspectives*, 25–42.

Freed, B. F. 2000. « Is Fluency, Like Beaty, in the Eyes (and Ears) of the Beholder ». *Perspectives on Fluency*. Sous la dir. de H. Riggenbach. Ann Arbor, MI : University of Michigan Press, 243–265.

Friginal, E. & J. A. Hardy. 2014. *Corpus-based Sociolinguistics: A Guide for Students*. Routledge.

Geeraerts, D. 1989. « Introduction: Prospects and problems of prototype theory ». *Linguistics* 27 (4), 587–612.

Geeraerts, D. 2017. « Entrenchment as onomasiological salience ». *Entrenchment and the psychology of language learning : How we reorganize and adapt linguistic knowledge*. Language and the human lifespan series. Boston, MA, US : De Gruyter Mouton, 153–174.

Georgila, K. 2009. « Using integer linear programming for detecting speech disfluencies ». *Proceedings of Human Language Technologies : The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume : Short Papers*. Association for Computational Linguistics, 109–112.

Georgila, K., N. Wang & J. Gratch. 2010. « Cross-domain speech disfluency detection ». *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, 237–240.

Gerdes, K., S. Kahane & P. Pietrandrea. 2012. « Intonosyntactic data structures: the Rhapsodie treebank of spoken French ». *Proceedings of the Sixth Linguistic Annotation Workshop*. Association for Computational Linguistics, 85–94.

Gilet, A. et al. 2013. « Assessing dispositional empathy in adults: A French validation of the Interpersonal Reactivity Index (IRI). » *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 45 (1), 42–48.

Gilquin, G. 2006. « The place of prototypicality in corpus linguistics: Causation in the hot seat ». *Corpora and Cognitive Linguistics : Corpus-based Approchaes to Syntax and Lexis*. T. 172. Trends in Linguistics. Berlin : Mouton de Gruyter, 159–191.

Gilquin, G. 2013. « Making sense of collostructional analysis: On the interplay between verb senses and constructions ». *Constructions and Frames* 5 (2), 119–142.

Gilquin, G., S. De Cock & S. Granger. 2010. *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM.*

Gilquin, G. & S. T. Gries. 2009. « Corpora and experimental methods: A state-of-the-art review ». *Corpus Linguistics and Linguistic Theory* 5 (1), 1–26.

Gilquin, G. et al. 2008. « Hesitation markers among EFL learners: Pragmatic deficiency or difference? » *Pragmatics and corpus linguistics : A mutualistic entente*, 119–149.

Ginzburg, J., R. Fernández & D. Schlangen. 2007. « Unifying self-and other-repair ». *Proceeding of DECALOG, the 11th International Workshop on the Semantics and Pragmatics of Dialogue (SemDial07)*, [en ligne].

Ginzburg, J., R. Fernández & D. Schlangen. 2014. « Disfluencies as intra-utterance dialogue moves ». *Semantics and Pragmatics* 7, 1–64.

Giora, R. 2003. *On Our Mind: Salience, Context, and Figurative Language*. Oxford University Press.

Glynn, D. & K. Fischer, éds. 2010. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. Berlin, New York : De Gruyter Mouton.

Goldman, J. 2011. « EasyAlign: an automatic phonetic alignment tool under Praat ». *Proceedings of Interspeech 2011, 12th Annual Conference of the International Speech Communication Association*. Firenze, Italy : ISCA.

Goldman, J., A. Auchlin & A. C. Simon. 2011. « Discrimination de styles de parole par analyse semi-automatique ». *Actes d'IDP 2009*. Sous la dir. de H. Yoo & E. Delais-Roussarie. Paris, France, 287–301.

Goldman, J., A. Auchlin & A. C. Simon. 2013. « Les variables temporelles dans le dialogue ». *Actes d'IDP 2013*. Louvain, Belgique, 39–44.

Goldman, J., M. Avanzi & A. Auchlin. 2010. « Hesitations in read vs. spontaneous French in a multi-genre corpus ». *Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop*. Tokyo, Japan : ISCA, 101–104.

Goldman, J., M. Avanzi, A. Auchlin & A. C. Simon. 2012. « A continuous prominence score based on acoustic features ». *Proceedings of the 13th Annual Conference of the International Speech Communication Association*. Portland, Oregon : ISCA.

Goldman, J., T. François, S. Roekhaut & A. C. Simon. 2010. « Étude statistique de la durée pausale dans différents styles de parole ». *Actes des XXVIIIeme Journée d'Etudes sur la Parole (JEP'2010)*. Mons, Belgique : ISCA-AFCP, 161–164.

Goldman, J. et al. 2010. « Prominence perception and accent detection in French. A corpus-based account ». *Proceedings Speech Prosody 2010*. Chicago, USA.

Goldman, J. et al. 2014. « Phonogenre Identification: A Perceptual Experiment with 8 Delexicalised Speaking Styles. » *Cahiers de linguistique française* 31, 51–62.

Goldman-Eisler, F. 1958. « Speech production and the predictability of words in context ». *Quarterly Journal of Experimental Psychology* 10, 96–106.

Goldman-Eisler, F. 1961. « A Comparative Study of Two Hesitation Phenomena ». *Language and Speech* 4 (1), 18–26.

Goldman-Eisler, F. 1968. *Psycholinguistics: Experiments in Spontaneous Speech*. Academic Press of London.

Goldman-Eisler, F. 1972. « Pauses, clauses, sentences ». *Language and Speech* 15, 103–113.

Götz, S. 2011. « Fluency in Native and Nonnative English Speech: Theory description, implications ». Ph.D. Thesis. Gießen, Germany : Justus-Liebig-Universität Gießen.

Götz, S. 2013a. « Fluency in Native and Nonnative English Speech: An integrated approach ». Louvain-la-Neuve, Belgium.

Götz, S. 2013b. « How fluent are advanced German learners of English (perceived to be)? Corpus findings versus native-speaker perception ». *Corpus Linguistics and Variation in English : Focus on non-native Englishes*. Sous la dir. de M. Huber & J. Mukherjee. Studies in Variation, Contacts and Change in English 13. University of Helsinki : Varieng Electronic Series, 1–13.

Gries, S. T. 2011. « Methodological and interdisciplinary stance in Corpus Linguistics ». *Perspectives on Corpus Linguistics*. Sous la dir. de V. Viana, S. Zyngier & G. Barbrook. Studies in Corpus Linguistics 48. Amsterdam & Philadelphia : John Benjamins Publishing Company, 81–98.

Grondelaers, S., M. Brysbaert, D. Speelman & D. Geeraerts. 2002. « Er als accessibility marker: on-en offline evidentie voor een procedurele interpretatie van presentatieve zinnen ». *Gramma/TTT* 9 (1), 1–22.

Grosjean, F. & A. Deschamps. 1972. « Analyse des variables temporelles du français parlé ». *Phonetica* 26, 129–156.

Grosjean, F. & A. Deschamps. 1973. « Analyse des variables temporelles du français spontané ». *Phonetica* 28 (3-4), 191–226.

Grosjean, F. & A. Deschamps. 1975. « Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation ». *Phonetica* 31 (3-4), 144–184.

Grosman, I. 2013a. « Synthèse de travail - Marqueurs prosodiques de (disfluence) en Français ». Louvain, Belgique.

Grosman, I. 2013b. « Variations phonostylistiques et patrons de bipolarisations rythmiques. Etude sur corpus : la parole des humoristes français ». Mémoire de Master. Université catholique de Louvain.

Grosman, I. 2014. « Dynamics of disfluencies within Basic Discourse Units. » *Cahiers de linguistique française* (31), 45–50.

Grosman, I. 2015. « Complexity Cues or Attention Triggers? Repetitions and Editing Terms for Native Speakers of French. » *Proceedings of the The 7th Workshop on Disfluency in Spontaneous Speech (DiSS 2015) - ICPHS Satellite Conference*. Edinburgh, Scotland, UK.

Grosman, I. 2016. « How do French humorists manage their persona across situations? A corpus study on their prosodic variation ». *Metapragmatics of humor : current research trends*. Sous la dir. de L. Ruiz-Gurillo. JB IVITRA. John Benjamins Publishing Company, 147–177.

Grosman, I., G. Christodoulides, L. Degand & A. C. Simon. 2017. « Prosodic variation of identical repetitions as a function of their properties and editing terms : A large-scale corpus study on French speech ». International Conference on Fluency & Disfluency Across Language and Language Varieties. Louvain-la-Neuve, Belgium.

Grosman, I., L. Rousier-Vercruyssen & E. Duvivier. 2016. « Perception des hésitations : liens entre accord inter-annotateurs et contextes prosodiques. » *Journée annuelle du Cercle Belge de Linguistique*. Louvain-la-Neuve, Belgique.

Guaïtella, I. 1991. « Rythme et parole : comparaison critique du rythme de la lecture oralisée et de la parole spontanée ». Thèse de Doctorat. Aix-en-Provence : Université de Provence.

Guénot, M. 2005. « Parsing de l'oral: traiter les disfluences ». *Actes des TALN2005 - Traitement Automatique des Langues Naturelles*. Dourdan, France, 323–332.

Günther, F., H. J. Müller & T. Geyer. 2017. « Salience, attention, and perception ». *Entrenchment and the psychology of language learning : How we reorganize and adapt linguistic knowledge*. Language and the human lifespan series. Boston, MA, US : De Gruyter Mouton, 289–312.

Halliday, M. 1987. « Spoken and written modes of meaning ». *Comprehending Oral and Written Language*. Sous la dir. de R. Horowitz & S. Samuels. Academic Press. New York, 55–82.

Hambye, P. & A. C. Simon. 2004. « The Production of Social Meaning Via the Association of Variety and Style: A Case Study of French Vowel Lengthening in Belgian ». *Canadian Journal of Linguistics/Revue canadienne de linguistique* 49 (3-4), 397–421.

Hartsuiker, R. J. 2014. « Monitoring and control of the production system ». *The Oxford handbook of language production*. Oxford University Press, 417–436.

Hartsuiker, R. J. & H. H. Kolk. 2001. « Error Monitoring in Speech Production: A Computational Test of the Perceptual Loop Theory ». *Cognitive Psychology* 42 (2), 113–157.

Heeman, P. A. & J. F. Allen. 1999. « Speech repairs, intonational phrases, and discourse markers: modeling speakers' utterances in spoken dialogue ». *Computational Linguistics* 25 (4), 527–571.

Heldner, M. & J. Edlund. 2010. « Pauses, gaps and overlaps in conversations ». *Journal of Phonetics* 38 (4), 555–568.

Henderson, A., F. Goldman-Eisler & A. Skarbek. 1966. « Sequential Temporal Paterns in Sponteaneous Speech ». *Language and Speech* 9 (4), 207–216.

Henry, S., E. Campione & J. Véronis. 2004. « Répétitions et pauses (silencieuses et remplies) en français spontané ». *Actes des XXVèmes Journées d'Etude sur la Parole (JEP'2004)*. Fes, Maroc : ISCA-AFCP, 1–4.

Henry, S. & B. Pallaud. 2002. « Étude des répétitions en français parlé spontané pour les technologies de la parole ». *Actes des 4e Rencontres des étudiants Chercheurs en Informatique pour le Traitement Automatique des Langue*. Nancy, France : Association pour le Traitement Automatique des Langues, 467–476.

Henry, S. & B. Pallaud. 2003. « Word fragments and repeats in spontaneous spoken French ». *Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop*. Sous la dir. de R. Eklund. Goldberg University, Sweden : Gothenburg Papers in Theoretical Linguistics, ISCA, 77–80.

Henry, S. & B. Pallaud. 2004. « Amorces de mots et répétitions dans les énoncés oraux ». *Recherches sur le français parlé* 18, 201–229.

Hieke, A. E. 1981. « A Content-Processing View of Hesitation Phenomena ». *Language & Speech* 24 (2), 147–160.

Hieke, A. E. 1985. « A Componential Approach to Oral Fluency Evaluation ». *The Modern Language Journal* 69 (2), 135–142.

Hieke, A. E., S. Kowal & D. C. O'Connell. 1983. « The trouble with "articulatory" pauses ». *Language and Speech* 26 (3), 203–214.

Hindle, D. 1983. « Deterministic Parsing of Syntactic Non-fluencies ». *Proceedings of the 21st Annual Meeting on Association for Computational Linguistics*. ACL '83. Stroudsburg, PA, USA : Association for Computational Linguistics, 123–128.

Hough, J., L. de Ruiter, S. Betz & D. Schlangen. 2015. « Disfluency and laughter annotation in a light-weight dialogue mark-up protocol ». *The 7th Workshop on Disfluency in Spontaneous Speech (Diss 2015)*. T. 1. Edinburg, Scotland, 8.

Jaffe, J. & S. Feldstein. 1970. *Rhythms of dialogue*. Personality and psychopathology. New York : Academic Press Inc.

Janda, L. A. 2010. « Cognitive linguistics in the year 2010 ». *International Journal of Cognitive Linguistics* 1 (1), 1–30.

Johnson, M. & E. Charniak. 2004. « A TAG-based Noisy Channel Model of Speech Repairs ». *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*. ACL '04. Stroudsburg, PA, USA : Association for Computational Linguistics.

Johnson, W. 1959. *The Onset of Stuttering: Research Findings and Implications*. Minneapolis, MN : University of Minnesota Press.

Jun, S. & C. Fougeron. 2002. « Realizations of accentual phrase in French intonation ». *Probus* 14 (1), 147–172.

Kapatsinski, V. 2010. « Frequency of Use Leads to Automaticity of Production: Evidence from Repair in Conversation ». *Language and Speech* 53 (1), 71–105.

Karniol, R. 1995. « Stuttering, language, and cognition a review and a model of stuttering as suprasegmental sentence ». *Psychological Bulletin* 117 (1), 104–124.

Kasl, S. V. & G. F. Mahl. 1965. « Relationship of disturbances and hesitations in spontaneous speech to anxiety ». *Journal of Personality and Social Psychology* 1 (5), 425–433.

Kassambara, A. 2017. *Practical Guide To Principal Component Methods in R*. English. STDA.

Kihlstrom, J. F. & P. E. Nathan. 2013. « The Person–Situation Interaction ». *The Oxford Handbook of Social Cognition*. Oxford : Oxford University Press, 786–806.

King, J. P., J. E. Loy & M. Corley. 2017. « Contextual effects on online pragmatic inferences of deception. » *Discourse Processes* 55, 123–135.

Kirsner, K., J. Dunn & K. Hird. 2003. « Fluency: Time for a paradigm shift ». *Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop*. Goldberg University, Sweden : Gothenburg Papers in Theoretical Linguistics, ISCA, 13–16.

Kirsner, K., J. Dunn & K. Hird. 2005. « Language Production: A complex dynamic system with a chronometric footprint », [en ligne].

Koch, P. & W. Osterreicher. 2001. « Gesprochene Spracche und geschriebene Sprache / Langage parlé et langage écrit ». *Lexikon der Romanistischen Linguistik*. Sous la dir. de G. Holtus, M. Metzeltin & C. Schmitt. T. 1. Tübinge : Niemeyer, 584–627.

Koponen, M. & H. Riggenbach. 2000. « Overview: Varying Perspectives on Fluency ». *Perspectives on fluency*. Sous la dir. de H. Riggenbach. Ann Arbor, MI. Michigan : University of Michigan Press, 5–24.

Kousidis, S., T. Pfeiffer & D. Schlangen. 2013. « MINT. tools : tools and adaptors supporting acquisition, annotation and analysis of multimodal corpora. » *Proceedings of Interspeech 2013, 14th Annual Conference of the International Speech Communication Association*. Lyon, France : ISCA, 2649–2653.

Kuznetsova, A., P. B. Brockhoff & R. H. B. Christensen. 2017. « lmerTest Package: Tests in Linear Mixed Effects Models ». *Journal of Statistical Software* 82 (13), 1–26.

Labroo, A. A., R. Dhar & N. Schwartz. 2007. « Of Frog Wines and Frowning Watches: Semantic Priming, Perceptual Fluency, and Brand Evaluation ». *Journal of Consumer Research* 34.

Lacheret, A. & B. Victorri. 2002. « La période intonative comme unité d'analyse pour l'étude du français parlé: modélisation prosodique et enjeux linguistiques ». *Verbum*. T. 1, 55–72.

Lacheret, A. et al. 2014. « Rhapsodie: a prosodic-syntactic treebank for spoken french ». *Proceedings of the 9th International Conference on Language Resources and Evaluation - LREC'14*. Reykjavik, Iceland : European Language Resources Association (ELRA), [en ligne].

Lacheret-Dujour, A. & F. Beaugendre. 1999. *La Prosodie du français*. Paris : Editions du CNRS.

Lai, C., K. Gorman, J. Yuan & M. Liberman. 2007. « Perception of disfluency: language differences and listener bias ». *Proceedings of Interspeech 2007, 8th Annual Conference of the International Speech Communication Association*. Antwerp, Belgium : ISCA, 2345–2348.

Langacker, R. W. 2000. « A dynamic usage-based model ». *Usage based models of language*. Sous la dir. de M. Barlow & S. Kemmer. Stanford : CSLI, 1–63.

Lefeuvre, F. 1999. *La phrase adverbiale en français*. Paris : L'Harmattan.

Lennon, P. 1990. « Investigating Fluency in EFL: A Quantitative Approach* ». *Language Learning* 40 (3), 387–417.

Lennon, P. 2000. « The lexical element in spoken second language fluency ». *Perspectives on fluency*. Sous la dir. de H. Riggenbach. Ann Arbor, MI : University of Michigan Press, 25–42.

Léon, P. R. 1969. « Principes et méthodes en phonostylistique ». *Langue française* 3 (1), 73–84.

Léon, P. R. 1993. *Précis de phonostylistique: Parole et expressivité*. Paris : Nathan.

Levelt, W. J. M. 1983. « Monitoring and self-repair in speech ». *Cognition* 14 (1), 41–104.

Levelt, W. J. M. 1989. *Speaking: From Intention to Articulation*. MIT Press.

Levelt, W. J. M. 1995. « The ability to speak: from intentions to spoken words ». *European Review* 3 (1), 13–23.

Levelt, W. J. M. & A. Cutler. 1983. « Prosodic Marking in Speech Repair ». *Journal of Semantics* 2 (2), 205–218.

Levinson, S. C. 2012. « The Original Sin of Cognitive Science ». *Topics in Cognitive Science* 4 (3), 396–403.

Levow, G. 2004. « Prosodic cues to discourse segment boundaries in human-computer dialogue ». *Proceedings of The 5th SIGdial Workshop on Discourse and Dialogue*. Boston, MA : ACL - Association for Computational Linguistics.

Lickley, R. J. 1994. « Detecting disfluency in spontaneous speech. » Ph.D. Thesis. University of Edinburg.

Lickley, R. J. & E. G. Bard. 1998. « When Can Listeners Detect Disfluency in Spontaneous Speech? » *Language and Speech* 41 (2), 203–226.

Little, D. R. et al. 2013. « Fluency Profiling System: An automated system for analyzing the temporal properties of speech ». *Behavior Research Methods* 45 (1), 191–202.

Liu, Y. et al. 2006. « Enriching speech recognition with automatic detection of sentence boundaries and disfluencies ». *IEEE Transactions on Audio, Speech, and Language Processing* 14 (5), 1526–1540.

Llisteri, J. 1992. « Speaking styles in speech research ». *Proceedings ELSNET/ESCA/-SALT Workshop on Integrating Speech and Natural Language*. Dublin, Ireland.

Loty, L. 2005. « Pour l'indisciplinarité ». *The Interdisciplinarity Century ; Tensions and convergences in 18th century Art, History and Literature* (4). edited by Julia Douthwaite and Mary Vidal, Oxford, Studies on Voltaire and the Eighteenth Century Voltaire Foundation, 2005, 245–259.

Loy, J. E., H. Rohde & M. Corley. 2016. « Effects of Disfluency in Online Interpretation of Deception ». *Cognitive Science* 41 (6), [en ligne].

Loy, J. E., H. Rohde & M. Corley. 2017. « Effects of Disfluency in Online Interpretation of Deception ». en. *Cognitive Science* 41 (S6), 1434–1456.

Luzzati, D. 1985. « Analyse périodique du discours ». *Langue française* 65 (1), 62–73.

MacGregor, L. J., M. Corley & D. I. Donaldson. 2009. « Not all disfluencies are are equal: The effects of disfluent repetitions on language comprehension ». *Brain and Language* 111 (1), 36–45.

Maclay, H. & C. E. Osgood. 1959. « Hesitation Phenomena in Spontaneous English Speech ». *Word* 15 (1), 19–44.

Malinin, L. H. 2016. « Creative Practices Embodied, Embedded, and Enacted in Architectural Settings: Toward an Ecological Model of Creativity ». *Frontiers in Psychology* 6, [en ligne].

Martin, L. J., L. Degand & A. C. Simon. 2014. « Forme et fonction de la périphérie gauche dans un corpus oral multigenres annoté ». *Corpus*. Eléments initiaux dans la phrase : approches inter-genres et inter-langues 13, 243–265.

Martin, P. 2014. « Analyse automatique de la structure prosodique d'énoncés de styles variés ». *Journal of Speech Sciences* 4 (1), 5–12.

Martin, R. 2015. « Analyse multimodale de la fluence discursive en production dans le cadre du vieillissement normal ». Mémoire de master en Linguistique générale et appliquée. Paris, France : Université Paris-Ouest Nanterre La Défense.

Maschler, Y. & D. Schiffrin. 2015. « Discourse Markers Language, Meaning, and Context ». *The Handbook of Discourse Analysis*. Sous la dir. de D. Tannen, H. E. Hamilton & D. Schiffrin. DOI : 10.1002/9781118584194.ch9. John Wiley & Sons, Inc., 189–221.

Mazur-Palandre, A. 2009. « Le flux de l'information, aspects syntaxiques et discursifs. Une étude fonctionnaliste et développementale ». Thèse de Doctorat. Lyon, France : Université Lumière Lyon 2.

McCroskey, J. C. & R. S. Mehrley. 1969. « The effects of disorganization and nonfluency on attitude change and source credibility ». *Speech Monographs* 36 (1), 13–21.

Mertens, P. 1993. « Accentuation, intonation et morphosyntaxe ». *Travaux de linguistique* 26, 21–69.

Mertens, P. 2004. « The prosogram: Semi-automatic transcription of prosody based on a tonal perception model ». *Proceedings of Speech Prosody 2004*. Nara, Japan, 1–4.

Mertens, P. 2008. « Syntaxe, prosodie et structure informationnelle : une approche prédictive pour l'analyse de l'intonation dans le discours ». *Travaux de linguistique* 56 (1), 97–124.

Mertens, P. 2011a. « A system for automatic tonal annotation of speech corpora ». *Actes d'IDP 2011*. Salford, Manchester, UK.

Mertens, P. 2011b. « Prosodie, syntaxe, discours: autour d'une approche prédictive ». *Actes d'IDP2009*. Sous la dir. de H. Yoo & E. Delais-Roussarie. Paris, 19–32.

Mertens, P. 2014. « Polytonia: a system for the automatic transcription of tonal aspects in speech corpora ». *Journal of Speech Sciences* 4 (2), 17–57.

Mertens, P. & A. C. Simon. 2013. « Towards automatic detection of prosodic boundaries in spoken French ». *Actes d'IDP 2013*. Louvain, Belgique, 81–87.

Mertens, P., P. Swiggers & K. Van den Eynde. 1998. « Structuration segmentale et suprasegmentale en syntaxe: vers un modèle intégrationniste de l'écrit et de l'oral ».

Meurant, L. 2015. *Corpus LSFB. A digital open access corpus of movies and annotations in French Belgian Sign Language (LSFB)*. Rapp. tech. Laboratoire de Langue des signes de Belgique francophone (LSFB-Lab). FRS-F.N.R.S et Université de Namur.

Mildner, V. 2008. *The cognitive neuroscience of human communication*. New York : Lawrence Erlbaum Associates.

Moniz, H. 2013. « Processing Disfluencies in European Portuguese. » Ph.D. Thesis. Univesidaded de Lisboa.

Moniz, H., F. Batista, I. Trancoso & A. I. Mata. 2012. « Prosodic contex-based analysis of disfluencies. » *Proceedings of Interspeech 2012, 13th Annual Conference of the International Speech Communication Association*. Portland, Or, USA : ISCA.

Moniz, H., A. I. Mata, I. Trancoso & C. Viana. 2008. « How can you use disfluencies and still sound as a good speaker? » *Proceedings of Interspeech 2008, 9th Annual Conference of the International Speech Communication Association*. Brisbane, Australia : ISCA, 1687.

Moniz, H., A. I. Mata & C. Viana. 2007. « On filled-pauses and prolongations in european portuguese. » *Proceedings of Interspeech 2007, 8th Annual Conference of the International Speech Communication Association*. Antwerp, Belgium : ISCA, 2645–2648.

Moniz, H., I. Trancoso & A. I. Mata. 2009. « Classification of disfluent phenomena as fluent communicative devices in specific prosodic contexts ». *Proceedings of Interspeech 2009, 10th Annual Conference of the International Speech Communication Association*. Brighton, UK : ISCA, 1719–1722.

Morel, M. & L. Danon-Boileau. 1998. *Grammaire de l'intonation l'exemple du français*. Bibliothèque de Faits et Langues : Linguistique. Paris, France : Ophrys.

Mosegaard-Hansen, M. 1997. « 'Alors' et 'donc' in spoken French: a reanalysis ». *Journal of Pragmatics* 28, 153–187.

Nakatani, C. & J. B. Hirschberg. 1993. « A speech-first model for repair detection and correction ». *Proceedings of the 31st annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 46–53.

Nakatani, C. & J. B. Hirschberg. 1994. « A Corpus-based study of repairs cues in spontaneous Speech ». *Journal of the Acoustical Society of America* 95 (3), 1603–1616.

Nolan, F. 2003. « Intonational equivalence : an experimental evaluation of pitch scales ». *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*. T. 39.

Norwine, A. C. & O. J. Murphy. 1938. « Characteristic Time Intervals in Telephonic Conversation ». *Bell System Technical Journal* 17 (2), 281–291.

Notarrigo, I. 2017. « Les marqueurs de (dis)fluence en Langue des signes de Belgique francophone ». Thèse de doct. UCL - Faculté de philosophie, arts et lettres.

Notarrigo, I., L. Meurant, M. Van Herreweghe & M. Vermeerbergen. 2016. « Repetition of Signs in French Belgian Sign Language (LSFB) and in Flemish Sign Language (VGT) ». *Typology and Annotation protocol, poster presented at the TISLR12*, 4–7.

Obin, N., X. Rodet & A. Lacheret. 2008. « Un modele de durée des syllabes fondé sur les propriétés syllabiques intrinseques et les variations locales de débit ». *Actes des 27ème Journées d'étude sur la Parole JEP-TALN 08*.

O'Connell, D. C. & S. Kowal. 2004. « The history of research on the filled pause as evidence of The Written Language Bias in Linguistics (Linell, 1982) ». *Journal of Psycholinguistic Research* 33 (6), 459–474.

O'Connell, D. C. & S. Kowal. 2008. *Communicating with One Another*. New York, NY : Springer New York.

O'Donoghue, T. & K. Punch, éds. 2003. *Qualitative Educational Research in Action: Doing and Reflecting*. English. London : Routledge.

Oehmen, R., K. Kirsner & N. Fay. 2010. « Reliability of the manual segmentation of pauses in natural speech ». *Advances in Natural Language Processing* 6253, 263–268.

Olive, J. P., A. Greenwood & J. Coleman. 1993. *Acoustics of American English Speech: A Dynamic Approach*. Springer Science & Business Media.

Osborne, J. 2010. « Fluency, complexity and informativeness in native and non-native speech ». *International Journal of Corpus Linguistics* 16 (2), 276–298.

O'Shaughnessy, D. D. 1992. « Analysis of false starts in spontaneous speech ». *Proceedings of the Second International Conference on Spoken Language (ICSLP 92)*. Banff, Alberta, Canada : [Washington, D.C.] Distributed by ERIC Clearinghouse, 931–934.

Oviatt, S. 1995. « Predicting spoken disfluencies during human-computer interaction ». *Computer Speech and Language* 9, 19–35.

Paas, F., J. E. Tuovinen, H. Tabbers & P. W. M. Van Gerven. 2003. « Cognitive Load Measurement as a Means to Advance Cognitive Load Theory ». *Educational Psychologist* 38 (1), 63–71.

Pallaud, B., S. Rauzy & P. Blache. 2013. *Identification et annotation des auto-interruption et des disfluences dans le corpus du CID*. Rapport Technique. Laboratoire Parole et Langage.

Pasdeloup, V. 1992. « Durée intersyllabique dans le groupe accentuel en Français ». *Actes des XIXèmes Journées d'Etudes sur la Parole*. Bruxelles, Belgique, 531–536.

Pasdeloup, V. 2005. « Figures et fond dans la scène prosodique: leur résistance face aux variations du débit de parole ». *Actes d'IDP 2005*.

Patel, A. D., J. R. Iversen & J. C. Rosenberg. 2006. « Comparing the rhythm and melody of speech and music: the case of British English and French ». eng. *The Journal of the Acoustical Society of America* 119 (5 Pt 1), 3034–3047.

Perfetti, C. A. & M. Bertuccelli Papi. 1985. « Towards a cognitive typology of pause phenomena ». *Communication & Cognition* 18 (4), 339–351.

Peshkov, K., L. Prévot & R. Bertrand. 2013. « Evaluation of automatic prosodic segmentations ». *Actes d'IDP 2013*. Louvain, Belgique, 95.

Petrova, P. K. & R. B. Cialdini. 2005. « Fluency of Consumption Imagery and the Backfire Effects of Imagery Appeals ». *Journal of Consumer Research* 32 (3), 442–452.

Pierrehumbert, J. B. 1980. « The phonology and phonetics of English intonation ». Thèse de doct. Indiana University Linguistics Club.

Pike, K. L. 1945. *The Intonation of American English*. University of Michigan Press.

Pinget, A.-F, H. Rutger Bosker, H. Quené & N. de Jong. 2014. « Native Speakers' Perceptions of Fluency and Accent in L2 Speech ». *Language Testing* 31, 349–365.

Plauche, M. C. & E. E. Shriberg. 1999. « Data-driven subclassification of disfluent repetitions based on prosodic features ». *Proceedings of the International Congress of Phonetic Sciences*. T. 2. San Francisco, CA, 1513–1516.

Postma, A. 2000. « Detection of errors during speech production: A review of speech monitoring models ». *Cognition* 77 (2), 97–132.

Postma, A. & H. H. Kolk. 1993. « The covert repair hypothesis: prearticulatory repair processes in normal and stuttered disfluencies ». *Journal of Speech and Hearing Research* 36 (3), 472–487.

Pradas Macías, M. 2006. « Probing quality criteria in simultaneous interpreting: The role of silent pauses in fluency ». *Interpreting* 8 (1), 25–43.

Pršir, T., J. Goldman & A. Auchlin. 2013. « Variation prosodique situationnelle: étude sur corpus de huit phonogenres en français ». *Actes d'IDP 2013*. Louvain, Belgique, 107–111.

Pršir, T., J. Goldman & A. Auchlin. 2014. « Prosodic features of situational variation across nine speaking styles in French ». *Journal of Speech Sciences* 4 (1), 41–60.

R Core Team. 2016. Vienna, Austria.

Radden, G. 1992. « The cognitive approach to natural language ». *Thirty Years of Linguistic Evolution*. Sous la dir. de M. Pütz. Amsterdam : John Benjamins Publishing Company, 513.

Rappaport, J. 2004. *365 Yoga*. Penguin.

Reber, R. & N. Schwartz. 1999. « Effects of Perceptual Fluency on Judgments of Truth ». *Consciousness and Cognition* 8 (3), 338–342.

Reber, R., N. Schwartz & P. Winkielman. 2004. « Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? » *Personality and Social Psychology Review* 8 (4), 364–382.

Reich, S. S. 1980. « Significance of pauses for speech perception ». *Journal of Psycholinguistic Research* 9 (4), 379–389.

Rendle-Short, J. 2004. « Showing structure: Using um in the academic seminar ». *Pragmatics* 14 (4), 479–498.

Richard, E. 2004. « La répétition : syntaxe et interprétation ». *L'Information Grammaticale* 100 (1), 53–54.

Richard, E. 2015. « A propos de répétition : entre continuité et rupture ». *Semen. Revue de sémio-linguistique des textes et discours* 38.

Riggenbach, H. 1991. « Toward an Understanding of Fluency: A Microanalysis of Nonnative Speaker Conversations ». *Discourse Processes* 14, 423–441.

Rosch, E. 1975. « Cognitive Representations of Semantic Categories ». *Journal of Experimental Psychology : General* 104 (3), 192–233.

Rosen, K. M. 2005. « Analysis of speech segment duration with the lognormal distribution: A basis for unification and comparison ». *Journal of Phonetics* 33 (4), 411–426.

Rossi, M. 1999. *L'intonation, le système du français : description et modélisation*. Paris, France : Oprys.

Rousier-Vercruyssen, L. 2017. « Prosodie, Syntaxe, et Référence : processus cognitifs et marqueurs linguistiques ». Thèse de doct. Paris, France : Université Paris 10.

Sacks, H., E. A. Schegloff & G. Jefferson. 1974. « A simplest systematics for the organization of turn-taking for conversation ». *Language* 50 (4), 696–735.

Sasisekaran, J. 2013. « Nonword repetition and nonword reading abilities in adults who do and do not stutter ». *Journal of Fluency Disorders* 38 (3), 275–289.

Saussure, F. d. 1989. *Cours De Linguistique Generale: Edition Critique*. Otto Harrassowitz Verlag.

Savova, G. & J. Bachenko. 2003. « Prosodic features of four types of disfluencies ». *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech DiSS'03*. Göteborg, Sweden.

Schegloff, E. A. 1998. « Reflections on studying prosody in talk-in-interaction. » *Language and Speech* 41 (3-4), 235–63.

Schegloff, E. A., G. Jefferson & H. Sacks. 1977. « The Preference for Self-Correction in the Organization of Repair in Conversation ». *Language* 53 (2), 361–382.

Schiffrin, D. 1987. *Discourse markers*. Cambridge : Cambridge University Press.

Schmid, H. 2010. « Does frequency in text really instantiate entrenchment in the cognitive system? » *Quantitative methods in cognitive semantics : Corpus-driven approaches* 46, 101.

Schmid, H. 2012. *Cognitive Pragmatics*. Berlin : Walter de Gruyter.

Schmid, H. & F. Günther. 2016. « Toward a Unified Socio-Cognitive Framework for Salience in Language ». *Frontiers in Psychology* 7 (1110).

Schmid, H. 1994. « Probabilistic Part-of-Speech Tagging Using Decision Trees ». *Proceedings of International Conference on New Methods in Language Processing*. Manchester, UK, 44–49.

Schmidt, R. 1992. « Psychological Mechanisms Underlying Second Language Fluency ». *Studies in Second Language Acquisition* 14 (4), 357–385.

Schönefeld, D., éd. 2011. *Converging Evidence: Methodological and theoretical issues for linguistic research*. T. 33. Human Cognitive Processing. Amsterdam : John Benjamins Publishing Company.

Schourup, L. 1999. « Discourse markers ». *Lingua* 107, 227–265.

Segalowitz, N. 2000. « Automaticity and attentional skill in fluent performance ». *Perspectives on Fluency*. Sous la dir. de H. Riggenbach. Ann Arbor, MI : University of Michigan Press, 200–219.

Segalowitz, N. 2010. *Cognitive Bases of Second Language Fluency*. New York : Routledge.

Segalowitz, N. 2016. « Second language fluency and its underlying cognitive and social determinants ». *International Review of Applied Linguistics in Language Teaching* 54 (2), 79–95.

Selting, M. 2000. « The construction of units in conversational talk ». *Language in Society* 29 (4), 477–517.

Sereno, K. K. & G. J. Hawkins. 1967. « The effects of variations in speakers' nonfluency upon audience ratings of attitude toward the speech topic and speakers' credibility ». *Speech Monographs* 34 (1), 58–64.

Shriberg, E. E. 1994. « Preliminaries to a Theory of Speech Disfluencies ». Thèse de doct. Berkeley, USA. : University of California.

Shriberg, E. E. 1995. « Acoustic properties of disfluent repetitions ». *Proceedings of the International Congress of Phonetic Sciences*. T. 4, 384–387.

Shriberg, E. E. 1999. « Phonetic consequences of speech disfluency ». *Proceedings of the 14th ICPhS conference*. San Francisco : DTIC Document, 619–622.

Shriberg, E. E. 2001. « To 'errrr' is human: ecology and acoustics of speech disfluencies ». *Journal of the International Phonetic Association* 31 (1), 153–169.

Shriberg, E. E. 2005. « Spontaneous speech: how people really talk and why engineers should care. » *Proceedings of Interspeech 2005, 9th European Conference on Speech Communication and Technology*. Lisbon, Portugal : ISCA, 1781–1784.

Shriberg, E. E., R. A. Bates & A. Stolcke. 1997. « A prosody only decision-tree model for disfluency detection. » *Proceedings of the 5th European Conference on Speech Communication and Technology*. Rhodes, Greece, 2383–2386.

Shriberg, E. E. & R. J. Lickley. 1993. « Intonation of clause-internal filled pauses ». *Phonetica* 50 (3), 172–179.

Shriberg, E. E., A. Stolcke & D. Baron. 2001. « Can Prosody Aid the Automatic Processing of Multi-Party Meetings? Evidence from Predicting Punctuation, Dis uencies, and Overlapping Speech ». *ISCA Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding*.

Simon, A. C. 2004. *La structuration prosodique du discours en français: une approche multidimensionnelle et expérientielle*. Sciences pour la communication. Bern : Peter Lang.

Simon, A. C. 2010. « Variation intono-syntaxique selon les discours : question de genre ou question de style ? » [présentation orale] Journées Rhapsodie. Toulouse.

Simon, A. C. 2012. *La variation prosodique régionale en français*. Champs linguistiques. De Boeck Supérieur.

Simon, A. C., A. Auchlin, M. Avanzi & J. Goldman. 2010. « Les phonostyles: une description prosodique des styles de parole en français. » *Les voix des Français. En parlant, en écrivant*. Sous la dir. de M. Abecassi & G. Ledegen. T. 2. Bern : Peter Lang, 71–88.

Simon, A. C., A. Auchlin, J. Goldman & G. Christodoulides. 2013. « Tendances prosodiques de la parole radiophonique ». *Cahiers de praxématique* 61, [en ligne].

Simon, A. C., M. Avanzi & J. Goldman. 2008. « La détection des proéminences syllabiques. Un aller-retour entre l'annotation manuelle et le traitement automatique ». *Actes du Congrès Mondial de Linguistique Française 2008*. EDP Sciences.

Simon, A. C. & G. Christodoulides. 2016. « Frontières prosodiques perçues : corrélats acoustiques et indices syntaxiques ». *Langue française* (3), 83–106.

Simon, A. C. & L. Degand. 2011. « L'analyse en unités discursives de base : pourquoi et comment ? » *Langue française* 170 (2), 45.

Simon, A. C., M. Francard & P. Hambye. 2014. « The VALIBEL Speech Database. » *The Oxford Handbook of Corpus Phonology*. Sous la dir. de J. Durand, U. Gut & G. Kristoffersen. Oxford : Oxford University Press, 552–531.

Simon, A. C. & P. Mertens. 2013. « Interface Discours - Posodie (IDP-2013) ». *Actes d'IDP 2013*. Sous la dir. d'A. C. Simon & P. Mertens. Louvain, Belgique.

Stanovich, K. E. & R. F. West. 2000. « Advancing the rationality debate ». *Behavioral and brain sciences* 23 (5), 701–717.

Stefanowitsch, A. & S. T. Gries. 2003. « Collostructions: Investigating the interaction of words and constructions ». *International Journal of Corpus Linguistics* 8 (2), 209–243.

Stenström, A. 2011. « Pauses and hesitations ». *Pragmatics of Society*. Berlin, Boston : De Gruyter Mouton, 537–568.

Stenström, A. 2014. « Avoid silence! Keep talking!: Pragmatic markers as phatic devices in teenage conversation ». *Functions of Language* 21 (1), 30–49.

Stolcke, A. et al. 1998. « Automatic detection of sentence boundaries and disfluencies based on recognized words. » *Proceedings of the Fifth International Conference on Spoken Language (ICSLP 98)*.

Strassel, S. 2004. *Simple Metadata Annotation Specification Versions 2.6, 5.0, 6.2*. Rapp. tech. Linguistic Data Consortium.

Swerts, M. 1998. « Filled pauses as markers of discours structure ». *Journal of Pragmatics* 30, 485–496.

Tanguy, N. 2012. « De la prosodie à la syntaxe Vers une description sémantico-syntaxique des périodes intonatives binaires ». *Représentations du sens linguistique* 5, 1–20.

Tanguy, N., T. Van Damme, D. Liesbeth & A. C. Simon. 2012. « Projet FRFC «Périphérie gauche des unités de discours» Protocole de codage syntaxique ».

Tashakkori, A. & C. Teddlie. 2003. *Handbook of Mixed Methods in Social & Behavioral Research*. Sous la dir. de T. Oaks. Sage.

Tesnière, L. 1988. *Eléments de syntaxe structurale*. 2e édition. Paris : Klincksieck.

Tomasello, M. et al. 2005. « Understanding and sharing intentions: the origins of cultural cognition ». *The Behavioral and Brain Sciences* 28 (5), 675–691.

Vaissière, J. 1991. « Rhythm, accentuation and final lenghtening in French ». *Music, Language, Speech and Brain*. Sous la dir. de J. Sundberg, L. Nord & R. Carlson. London : Macmillan Press, 108–120.

Vasilescu, I., M. Candea & M. Adda-Decker. 2004. « Hésitations autonomes dans 8 langues: une étude acoustique et perceptive ». *Workshop Modélisations pour l'IDentification des Langues (MIDL04)*. Paris, France.

Vasilescu, I., M. Candea, M. Adda-Decker et al. 2005. « Perceptual salience of language-specific acoustic differences in autonomous fillers across eight languages ». *Proceedings of Interspeech 2005, 9th European Conference on Speech Communication and Technology*. Lisbon, Portugal : ISCA.

Voss, B. 1979. « Hesitation phenomena as sources of perceptual errors for non-native speakers ». *Language and Speech* 22 (2), 129–144.

Vrij, A. & F. W. Winkel. 1991. « Cultural patterns in Dutch and Surinam nonverbal behavior: An analysis of simulated police/citizen encounters ». *Journal of Nonverbal Behavior* 15 (3), 169–184.

Watanabe, M., K. Hirose, Y. Den & N. Minematsu. 2008. « Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners ». *Speech Communication* 50 (2), 81–94.

Wheeldon, L. & A. Lahiri. 1997. « Prosodic Units in Speech Production ». *Journal of Memory and Language* 37 (3), 356–381.

Wingate, M. E. 1984. « Fluency, disfluency, dysfluency and stuttering ». *Journal of Fluency Disorders* 9 (2), 163–168.

Winkielman, P. & J. T. Cacioppo. 2001. « Mind at Ease Puts a Smile on the Face Psychophysiological Evidence That Processing facilitation Elicits Positive Affect ». *Journal of Personality and Social Psychology* 81 (9), 989–1000.

Wlodarczak, M. & P. Wagner. 2013. « Effects of talk-spurt silence boundary thresholds on distribution of gaps and overlaps ». *Proceedings of Interspeech 2013, 14th Annual Conference of the International Speech Communication Association*. Lyon, France : ISCA, 1434–1437.

Zarcone, A., M. van Schijndel, J. Vogels & V. Demberg. 2016. « Salience and Attention in Surprisal-Based Accounts of Language Processing ». *Frontiers in Psychology* 7, 844.

Zechner, K. & I. I. Bejar. 2006. « Towards Automatic Scoring of Non-native Spontaneous Speech ». *Proceedings of the Human Language Technology : North American Chapter of the Association for Computational Linguistic HLT/NAACL*. HLT-NAACL '06. Stroudsburg, PA, USA : Association for Computational Linguistics, 216–223.

Zellner, B. 1994. « Pauses and the temporal structure of speech ». *Fundamentals of speech synthesis and speech recognition*. Sous la dir. d'E. Keller. Chichester : John Wiley, 41–62.

Zufferey, S. & L. Degand. 2013. « Explicit and implicit discourse relations across languages ». *13th International Pragmatics Conference*. New Delhi, India.

Index des auteurs

Adda, Gilles, 16
Adda-Decker, Martine, 45, 53
Adell, Jordi, 12
Akehurst, Lucy, 16
Alexandersson, Jan, 12, 122
Allen, James F., 52, 54, 55
Allen, Jont B., 34
Allopenna, Paul D., 17
Alter, Adam L., 2, 3, 6, 37, 67, 280, 287, 293
Anderson, Julie D., 7
Arciuli, Joanne, 16
Arnold, Jennifer E., 15, 16, 133, 279, 280, 287
Arppe, Antti, 62, 135, 136
Auchlin, Antoine, 18, 34, 35, 37, 38, 72, 80, 112, 139, 147, 221, 250, 297
Auer, Peter, 15, 16
Avanzi, Mathieu, 18, 34, 35, 37, 52, 72, 80, 112, 118, 221, 233, 250, 269

Baayen, R. Harald, 290
Bachenko, Joan, 49–51, 273, 274
Bachy, Sylviane, 80
Baldassarre, Gianluca, 64
Bard, Ellen Gurman, 14
Bardiaux, Alice, 35
Barik, Henri C., 74
Baron, Don, 12, 13, 44–47, 52, 79, 87, 245
Barto, Andrew, 64
Bates, Douglas M., 227, 290
Bates, Rebecca A., 12–14, 52
Batista, Fernando, 86
Bazillon, Thierry, 80
Bazzanella, Carla, 88
Bear, John, 13
Beaugendre, Frédéric, 32, 34, 78, 80, 82, 100, 126, 127, 189
Bejar, Isaac I., 11
Beliao, Julie, 22, 56
Bell, Allan, 37, 38
Bell, Linda, 36, 75
Belz, Malte, 49
Benzitoun, Christophe, 58
Bernicot, Josie, 279
Berthold, André, 37
Bertrand, Roxane, 80, 90, 141, 282
Bertrand, Russell, 118
Bertuccelli Papi, Marcella, 71
Besser, Jana, 12, 122
Betz, Simon, 122
Bigi, Brigitte, 282
Bilger, Mireille, 22, 110
Blache, Philippe, 118
Blackmer, Elizabeth R., 29
Blakemore, Sarah-Jayne, 63
Blanche-Benveniste, Claire, 20–22, 58, 71, 80, 88, 91, 92, 97, 116
Bloch, Bernard, 72, 75
Blumer, Herbert, 63
Boersma, Paul, 221, 282
Bolker, Ben, 227, 290
Bolly, Catherine, 22, 57, 295
Bonafonte, Antonio, 12
Boomer, Donald S., 37
Bordal, Guri, 35

Bortfeld, Heather, 133
Bosker, Hans Rutger, 8, 141, 145, 146
Bosshardt, Hans-Georg, 7
Bothuyne, Marie, 48
Boula de Mareüil, Philippe, 16, 35, 38, 45, 199, 206
Bouraoui, Jean-Léon Mehdi, 10, 11, 15, 45
Boutet, Dominique, 295
Boves, Lou, 11
Branca-Rosoff, Sonia, 111
Brennan, Susan E., 30, 80, 82, 279
Brockhoff, Per B., 292
Broen, Patricia A., 45
Brunken, Roland, 286
Brysbaert, Marc, 136
Bull, Ray, 16
Bybee, Joan L., 61
Byrd, Courtney T., 7

Cacioppo, John T., 6
Campione, Estelle, 22, 57, 71, 79, 85, 89, 92, 110, 141–144, 150, 258–260, 269
Candea, Maria, 1, 12, 22, 45–49, 52–54, 58, 59, 61, 69–86, 88–96, 98–104, 118, 144, 150, 156–158, 162, 164, 167, 189, 223, 243, 252, 256–260, 269, 271
Carton, Fernand, 46, 83, 84
Chafe, Wallace, 18, 30, 31, 63
Chambers, John, 70
Charniak, Eugene, 12
Cholin, Joana, 118
Christensen, Rune H. B., 292
Christodoulides, George, 31, 37, 52, 55–57, 80, 139, 147, 150, 154, 186, 233, 250, 258, 260, 269, 297
Cialdini, Robert B., 3
Clark, Andy, 64
Clark, Eve V., 83, 279
Clark, Herbert H., 15, 16, 31, 71, 73, 83, 96, 103, 188, 279, 280
Claverie, Bernard, xx
Clerc-Renaud, Jeanne, 53
Cohen, Jacob, 114
Cole, Jennifer S., 49, 94, 96, 273
Coleman, John, 50
Corley, Martin, 6, 15, 17, 279, 293, 312, 327, 328
Corminboeuf, Gilles, 58
Cowan, J. Milton, 72, 75
Crible, Ludivine, 11, 55, 57, 114, 128, 200, 233, 325
Crystal, David, 19
Cucchiarini, Catia, 11
Cutler, Anne, 34, 44, 50, 82, 96, 122

Dahan, Delphine, 34, 82
Dall, Rasmus, 328
Danon-Boileau, Laurent, 30, 48, 54, 57, 58, 74, 75, 79, 81, 83, 85, 91–93, 95, 189, 215, 269, 271
Davidson, Doug J., 290
Davis, Mark H., 295
De Jong, Nivja, 328
De Jong, Nivja H., 8, 141, 145, 146
de Medeiros, Henrique R. B, 45
De Ruiter, Laura, 122
Debaisieux, Jeanne-Marie, 20
Degand, Liesbeth, 11, 57, 61, 109, 110, 114, 147, 150, 154, 166, 199, 200, 221, 233, 250, 258, 260, 269, 297
Delais-Roussarie, Elisabeth, 34
Delattre, Pierre, 33, 35
Demberg, Vera, 31
Demol, Mike, 142, 148, 153
Den, Yasuharu, 16, 100, 279, 280, 287, 288
Denke, Annika, 8
DePaulo, Bella M., 16
Deschamps, Alain, 47, 57, 69, 71, 76, 80, 84, 86, 92–94, 100, 103, 144, 150, 156, 162, 198, 258
Deulofeu, José, 20
De Cock, Sylvie, 114

Dhar, Ravi, 3
Dister, Anne, 22, 57, 76, 79–86, 88, 89, 91–94, 97, 189, 206, 250, 252, 256–258, 267
Divjak, Dagmar, 64
Di Cristo, Albert, i, ii, 32–36, 52, 77, 126
Donaldson, David I., 279, 293, 312
Doutrich, Dawn, 26
Dowding, John, 12, 13
Duboisdindien, Guillaume, 328
Duez, Danielle, 9, 12, 18, 46–48, 52, 53, 59, 69–72, 74, 75, 77, 79–82, 91, 97, 101, 135, 139, 144, 145, 150, 189, 190, 215, 216, 244, 269
Dumont, Amandine, 114, 128, 233, 325
Dunn, John, 142
Dutoit, Thierry, 34
Dutrey, Camille, 52
Duvivier, Emilie, 123

Edlund, Jens, 72, 131, 145
Eklund, Robert, 7, 13, 36, 52, 71, 73, 75, 118, 187, 188
Escudero-Mancebo, David, 12
Eskenazi, Maxime, 112
Estes, William Kaye, 61
Everitt, Brian S., 114
Eychenne, Julien, 110

Fónagy, Ivan, 80
Fagnano, Maria, 279, 280, 287
Fay, Nicolas, 141
Faygal, Zsuzsanna, 99
Feldstein, Stanley, 127
Fernández, Raquel, 19, 26, 28, 90, 119
Ferré, Gaëlle, 90
Ferreira, Fernanda, 30
Fillmore, Charles J., 19, 20
Fischer, Kerstin, 11
Fleiss, Joseph L., 114
Fors, Kristina Lundholm, 72, 131
Fossard, Marion, 123, 221
Fox Tree, Jean E., 1, 16, 70, 91, 187, 279
Fraisse, Paul, 77
François, Thomas, 142, 143, 146, 148, 150, 153, 162
Frederick, Shane, 283
Freed, Barbara F., 18
Friginal, Eric, 131
Frith, Uta, 63

Götz, Sandra, i, 8–11
Günther, Franziska, i, 4, 5, 23, 31, 64, 65, 294
Geeraerts, Dirk, 62, 64, 136
Georgila, Kallirroi, 44, 52
Gerdes, Kim, 297
Geyer, Thomas, 294
Gilet, Anne-Laure, 295
Gilquin, Gaëtanelle, 62, 67, 114, 133, 135, 136, 199, 200, 328
Ginstrom, Ryan, 15
Ginzburg, Jonathan, 19, 26, 28, 90, 119
Giora, Rachel, 64
Goldman, Jean-Philippe, 18, 34, 35, 37, 38, 52, 72, 80, 110–112, 139, 142, 143, 146–148, 150, 153, 162, 221, 233, 250, 269, 282, 297
Goldman-Eisler, Frida, 18, 23, 30, 71, 72, 74
Gorman, Kyle, 16
Granger, Sylviane, 114
Gratch, Jonathan, 44
Greenwood, Alice, 50
Gries, Stefan Th., 63, 133, 135, 136
Grondelaers, Stefan, 136
Grosjean, François, 47, 57, 69, 71, 76, 80, 84, 86, 92–94, 100, 103, 144, 150, 156, 162, 198, 258
Grosman, Iulia, 60, 109, 111, 114, 123, 128, 147, 150, 154, 233, 250, 258, 260, 269, 281, 297, 325–327
Guénot, Marie-Laure, 16, 22
Guaïtella, Isabelle, 43, 49, 71, 79, 83, 86, 190, 216, 246
Guardiola, Mathilde, 90
Gustafson, Joakim, 36, 75

Halliday, M.A.K, 30, 31
Hambye, Philippe, 63
Hardy, Jack A., 131
Hartsuiker, Robert J., 26, 58
Hawkins, Gary J., 16
Heeman, Peter A., 52, 54, 55
Heldner, Mattias, 72, 131, 145
Henderson, Alan, 18
Henry, Sandrine, 47, 71, 89, 92, 93, 258–260, 269
Hieke, Adolf E., i, 20, 26, 70, 94, 95, 144, 269
Hindle, Donald, 13, 44, 51, 55, 122
Hird, Kathryn, 142
Hirose, Keikichi, 16, 100, 279, 280, 287, 288
Hirschberg, Julia Bell, i, 12–14, 45, 72, 73, 131
Hough, Julian, 122

Iversen, John R., 216

Jaffe, Joseph, 127
Jameson, Anthony, 37
Janda, Laura A., 62, 135
Jeanjean, Colette, 22, 71
Jefferson, Gail, 18, 26, 72, 74, 131
Johnson, Mark, 12
Johnson, Wendell, 7

Kahane, Sylvain, 56, 297
Kapatsinski, Vsevolod, 30
Karniol, Rachel, 7
Kasl, Stanislav V., 36
Kassambara, Alboukadel, 263
Kihlstrom, John F., 63
King, J. P., 15
King, Simon, 328
Kirsner, Kim, 141, 142
Koch, Peter, 112
Kohnken, G., 16
Kolk, Herman H.J., 26, 58
Koponen, Matti, 19
Kousidis, Spyros, 122
Kowal, Sabine, 18, 70, 72, 144
Kuznetsova, Alexandra, 292

Léon, Pierre Roger, 15, 18, 32, 33, 35–37, 70, 77, 80, 190
Labroo, A. Aparna, 3
Lacheret, Anne, 22, 56, 143, 144, 147, 148, 250
Lacheret-Dujour, Anne, 32, 34, 78, 80, 82, 100, 126, 127, 189
Lahiri, Aditi, 30
Lai, Catherine, 16
Langacker, Ronald W., 62, 64
Lefeuvre, Florence, 110, 127
Lennon, Paul, 18–20
Leutner, Detlev, 286
Levelt, Willem J. M., i, 7, 12, 13, 23, 24, 26, 29, 30, 35, 44, 47, 50, 55, 58, 72, 96, 118, 119, 122, 274
Levinson, Stephen C., 63
Levow, Gina-Anne, 50
Liberman, Mark, 16
Lickley, Robin J., 14, 47, 49, 50, 73, 80, 190
Little, Daniel R., 142, 146, 153
Liu, Yang, 14, 16, 52
Llisteri, Joachim, 112
Losongco, Anthony, 15
Loty, Laurent, xvii, 135
Loy, Jia E., 6, 15, 17, 327, 328
Luzzati, Daniel, 85

Müller, Hermann J., 294
MacGregor, Lucy J., 279, 293, 312
Maclay, Howard, 18, 30, 71, 86
Magnuson, James S., 17
Mahl, George F., 36
Malinin, Laura H., i, 3
Mallard, David, 16
Martin, Laurence J., 109, 110, 114, 147, 166, 221, 233, 297
Martin, Philippe, 250
Martin, Raphaëlle, 294
Maschler, Yael, 11

Mata, Ana Isabel, 14, 16, 49, 52, 73, 86, 187, 188, 190, 198, 199, 219, 279
Mazur-Palandre, Audrey, i, 23–25
McCroskey, James C., 16
Mehrley, R. Samuel, 16
Mertens, Piet, 34, 60, 69, 110, 126, 127, 221
Meunier, Christine, 282
Meurant, Laurence, 91, 114
Mildner, Vesna, 63
Milin, Petar, 290
Minematsu, Nobuaki, 16, 100, 279, 280, 287, 288
Mirolli, Marco, 64
Mitton, Janet L., 29
Moniz, Helena, 14, 16, 49, 52, 73, 86, 118, 187, 188, 190, 198, 199, 219, 279
Morel, Mary-Annick, 30, 48, 54, 57, 58, 74, 75, 79, 81, 83, 85, 91–93, 95, 189, 215, 269, 271
Mosegaard-Hansen, Maj-Brit, 80, 83
Mächler, Martin, 227, 290

Nakatani, Christine, i, 12–14, 45, 73
Nathan, P. E., 63
Nesterenko, Irina, 282
Nolan, Francis, 216
Notarrigo, Ingrid, 91, 114, 128, 233, 325

O'Connell, Daniel C., 18, 70, 72, 144
O'Shaughnessy, Douglas D., 49, 189, 190, 216, 244, 245
Obin, Nicolas, 35, 143, 148
Oehmen, Raoul, 141
Olive, Joseph P., 50
Oppenheimer, Daniel M., 2, 3, 6, 67, 287, 293
Osborne, John, 8
Osgood, Charles E., 18, 30, 71, 86
Osterreicher, Wulf, 112
Oviatt, Sharon, 58

Paas, Fred, 286
Pallaud, Berthille, 47, 89, 93, 118, 258
Pasdeloup, Valérie, 53
Patel, Aniruddh D., 216
Perfetti, Charles A., 71
Peshkov, Klim, 141
Petrova, Petia K., 3
Pfänder, Stefan, 16
Pfeiffer, Thies, 122
Pierrehumbert, Janet Breckenridge, 51
Pietrandrea, Paola, 297
Pike, Kenneth Lee, 51
Pinget, A.-F, 328
Plass, Jan L., 286
Plauche, Madelaine C., 94, 269, 274
Postma, Albert, i, 26, 27
Prévot, Laurent, 141
Pršir, Tea, 147, 250, 297
Pradas Macías, Macarena, 294

Quené, Hugo, 8
Quené, Hugo, 328

Radden, Günter, 62
Rappaport, Julie, 139
Rauzy, Stéphane, 118
Reber, Rolf, 6, 26
Reich, Shuli S., 74
Reichel, Uwe, 49
Rendle-Short, Johanna, 16
Richard, Élisabeth, 88, 90, 97
Riggenbach, Heidi, 19, 70, 75
Rodet, Xavier, 143, 148
Roekhaut, Sophie, 142, 143, 146, 148, 150, 153, 162
Rohde, Hannah, 6, 17, 327, 328
Rosch, Eleanor, 61
Rosen, Kristin M., 142, 148
Rosenberg, Jason C., 216
Rossi, Mario, 47, 77
Rousier-Vercruyssen, Lucie, 123
Rutger Bosker, Hans, 328

Sacks, Harvey, 18, 26, 72, 74, 131
Sanders, Ted, 8
Sasisekaran, Jayanthi, 7

Saussure, Ferdinand de, 21
Savova, Guergana, 49–51, 273, 274
Schegloff, Emanuel A., 18, 26, 72, 74, 131
Schiffrin, Deborah, 10, 11
Schlangen, David, 19, 26, 28, 90, 119, 122
Schmid, Hans-Jörg, i, 4, 5, 23, 31, 62, 64, 65, 69, 294
Schmidt, Richard, 19
Schourup, Lawrence, 10
Schwartz, Norbert, 3, 6, 26
Segalowitz, Norman, i, 8, 9, 18, 37, 279, 280
Sereno, Kenneth K., 16
Shriberg, Elizabeth E., i, 11–14, 44–52, 58, 59, 73, 79, 80, 86, 87, 89, 91, 93–96, 98, 100, 101, 118, 119, 122, 189, 190, 216, 245, 260, 269, 271, 273, 274, 280
Siegel, Gerald M., 45
Simon, Anne Catherine, 18, 32, 34, 35, 37, 38, 55–57, 60, 61, 63, 80, 109, 110, 112, 114, 139, 142, 143, 146–148, 150, 153, 154, 162, 166, 186, 221, 233, 250, 258, 260, 269, 297
Skarbek, Andrew, 18
Speelman, Dirk, 136
Stanovich, Keith E., 283
Stefanowitsch, Anatol, 133
Stenström, Anna-Brita, 328
Stolcke, Andreas, 12–14, 44–47, 52, 79, 87, 245
Strassel, Stepanie, 12, 118
Strik, Helmer, 11
Sussman, Harvey, 7
Swerts, M., 58, 72, 73, 80, 83
Swiggers, Pierre, 69

Tabbers, Huib, 286
Tanenhaus, Michael K., 16, 17, 133, 279, 280, 287
Tanguy, Noalig, 127
Tashakkori, Abbas, 69
Teddlie, Charles, 69
Tesnière, Lucien, 57
Tomasello, Michael, 20
Trancoso, Isabel, 14, 16, 49, 52, 86, 187, 190, 198, 199, 279
Tree, Jean E. Fox, 15, 16, 31, 73, 188, 279, 280
Tuovinen, Juhani E., 286

Uygur-Distexhe, Deniz, 57

Véronis, Jean, 57, 71, 79, 85, 89, 92, 141–144, 150, 258–260, 269
Vaissière, Jacqueline, 76, 78
Vallely, Megann, 7
Van Bergen, Geertje, 11
Van Schijndel, Marten, 31
Van Damme, Thomas, 110
Van den Eynde, Karel, 69
Van Donselaar, Wilma, 34, 82
Van Gerven, Pascal W. M., 286
Van Herreweghe, Mieke, 91
Vasilescu, Ioana, 45, 53
Verhelst, Werner, 142, 148, 153
Verhoeve, Piet, 142, 148, 153
Vermeerbergen, Myriam, 91
Viana, Céu, 73, 86, 188, 198, 199, 219
Victorri, Bernard, 144
Vigouroux, Nadine, 45
Villar, Gina, 16
Vogels, Jorrig, 31
Voss, Bernd, 97
Vrij, Aldert, 16

Wagner, Petra, 127
Walker, Steve, 227, 290
Wang, Ning, 44
Wasow, Thomas, 15, 73, 96, 279
Watanabe, Michiko, 16, 100, 279, 280, 287, 288
Weenink, David, 221, 282
West, Richard F., 283
Wheeldon, Linda, 30
William, Maurice, 30, 80, 82, 279

Wingate, Marcel E., 7
Winkel, Frans Willem, 16
Winkielman, Piotr, 6, 26
Winston, Joel, 63
Wlodarczak, Marcin, 127

Yamagishi, Junichi, 328
Yuan, Jiahong, 16

Zarcone, Alessandra, 31
Zechner, Klaus, 11
Zellner, Brigitte, 71, 73, 98, 143, 145
Zufferey, Sandrine, 11

Grosman, Iulia • 2018
Evaluation contextuelle de la (dis)fluence en production et perception. Pratiques communicatives et formes prosodico-syntaxiques en français •
Thèse de doctorat • Université catholique de Louvain.

Université catholique de Louvain

Valibel - Centre de recherche Discours et Variation
ILC - Institut Langage et Communication

Thèse de doctorat

Evaluation contextuelle de la (dis)fluence en production et perception

Pratiques communicatives et formes prosodico-syntaxiques en français

Annexes

Auteure :
Iulia Grosman

Promotrices :
Prof. Anne Catherine Simon
Prof. Liesbeth Degand

Thèse soutenue en vue de l'obtention du grade de
Docteur en Langues et Lettres

Membres du jury :
Prof. Lambert Isebaert (Université catholique de Louvain, président)
Prof. Anne Catherine Simon (Université catholique de Louvain, promotrice)
Prof. Liesbeth Degand (Université catholique de Louvain, promotrice)
Prof. Maria Candea (Université Paris Sorbonne Nouvelle - Paris 3)
Prof. Philippe Boula de Mareüil (CNRS, LIMSI)
Prof. Laurence Meurant (Université de Namur)

Septembre 2018

Annexe A

Description de C-Humour : La parole des humoristes français

1 Constitution et structure du corpus

Le corpus francophone de paroles d'humoristes (C-Humour ; Grosman 2016) comprend 26 échantillons de parole de 8 humoristes français [1], récoltés entre 2009 et 2011. Sa durée totale est de 1h56 minutes. Il compte 6 monologues (les chroniques radiophoniques) et 20 dialogues. Parmi les huit humoristes, quatre prononcent des dialogues dans le cadre de la chronique radiophonique contre 4 en interviews. [2]. Ces derniers incluent 4 interviews radiophoniques, 8 dialogues conversationnels informels et 8 sketchs sur scène. La Figure A.1 reprend la structure du corpus.

L'échantillonnage de ce corpus permet de comparer le profil linguistique d'un même locuteur au travers de trois situations. La structure équilibrée du corpus encourage une étude contrastive selon plusieurs axes : le sexe de l'humoriste (4H/4F), son degré d'expérience (+10 années / -10 années sur scène) et la situation de parole. Il permet d'étudier la variation phonostylistique de la parole de chaque humoriste au sein des différentes situations de parole (Grosman 2013b). Le corpus a été aligné automatiquement au moyen d'EasyAlign (Goldman 2011) et comprend 19 couches d'annotations. Cette annexe décrit leur contenu. La description détaillée de chaque tire est reprise en Annexe C.

1. Il serait ardu de garder leur anonymat. Les échantillons reprennent la voix d'Anne Roumanoff, de Bérangère Krief, d'Emma Gattuso, de Florence Foresti, d'Arnaud Tsamère, de Kev Adams, d'Olivier De Benoist et de Stéphane Guillon.

2. Deux échantillons ont été ajoutés pour deux des humoristes en raison de la brièveté de leurs chroniques.

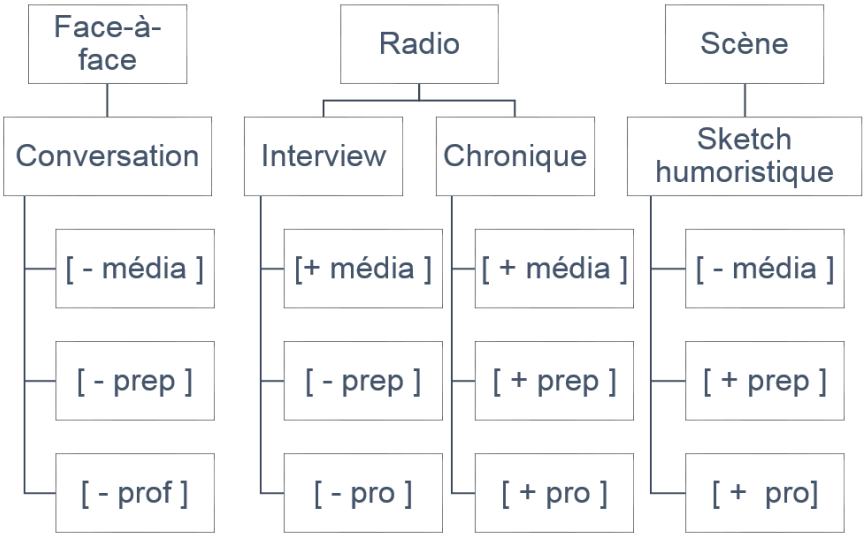


FIGURE A.1 – Structure du corpus C-Humour.

Le corpus C-HUMOUR est structuré en deux dossiers :

1. Le dossier [CORPUS] regroupe les enregistrements (fichiers sons) et la stylisation du pitch de chaque fichier. Il contient une série d'annotations : les TextGrids et tables associés, le profil prosodique de chaque enregistrement et sa transcription orthographique.
2. Le dossier [METADATA] comprend ce manuel ainsi que les fichiers de métadonnées consignant les caractéristiques de chaque situation d'enregistrement et des locuteurs. Un fichier synthétise ces informations.

2 Droits et origines du corpus

Les enregistrements du corpus C-HUMOUR proviennent de diverses sources référencées au Tableau A.1.

Nr	Fichier	Date d'enregistrement	Titre	Diffusion
1	chfAR1	01.09.11	Anne Roumanoff, en vrai : une interview en toute intimité	Philppe Vaillant
2	chfAR2a	10.09.12	Les inédits d'Anne Roumanoff : La hausse des impôts c'est maintenant !	Europe 1
3	chfAR2b	19.09.12	Les inédits d'Anne Roumanoff : De l'huile sur le feu	Europe 1

Nr	Fichier	Date d'enregistrement	Titre	Diffusion
4	chfAR3	01.09.09	Radio Bistro : Anne Roumanoff allume Sarkozy and Friends	France 2 Radio Bistro
5	chfBK1	02.06.10	Bérengère Krief : "Je pleurais quand j'étais seule sur scène"	Mademoizelle
6	chfBK2	31.01.12	Le point virgule fait l'Olympia : Bérangère Krief parle de ses potes gay	Rire et Chanson
7	chfBK3	01.01.10	Ma mère mon chat et Doctor House	Théatre le bout
8	chfEG1	06.02.12	Emma Gattuso dans les cahiers du Rire	Les cahiers du rires
9	chfEG2	04.09.12	Ma journée à ESTROSI-LAND	France Bleu Azur
10	chfEG3	01.06.12	Gattuso m'a tuer	Théâtre des minimes
11	chfFF1	02.06.11	Interview Facebook itw Live	TF1 Vision (Social V0D)
12	chfFF2	01.12.10	Le prince charmant c'est de la connerie	Europe 1
13	chfFF3	26.06.10	Mother Fucker	(DVD)
14	chmAT1	01.06.12	Les chéris d'Anne au Murano : spéciale Arnaud Tsamère	PVO Your Humour
15	chmAT2	01.05.11	Michel Drucker : les talents de demain	Europe 1
16	chmAT3	01.09.11	On ne demande qu'à en rire : Le cours d'économie	France 2
17	chmKA1	01.07.11	Kev-Adams - L'interview rafraîchissante par Première.fr	Première.fr
18	chmKA2	27.06.10	On va s'gêner (Part 2)	Europe 1
19	chmKA3	01.06.10	The Young Man Show	(DVD)
20	chmOB1	12.10.11	Femmes d'Aujourd'hui : Interview d'Olivier de Benoist	Femme d'aujourd'hui
21	chmOB2a	21.06.11	La Bretagne ça vous gagne	Europe 1
22	chmOB2b	19.09.11	La soupe à la grimace	Europe 1
23	chmOB3	11.03.11	Olivier de Benoist : seul face aux femmes !	Festival du Rire TV
24	chmSG1	05.11.11	Interview de Stéphane Guillon du 05 11 11	Duval Sarah

Nr	Fichier	Date d'enregistrement	Titre	Diffusion
25	chmSG2	22.03.10	L'humeur de Stephane Guillon	France inter
26	chmSG3	01.06.06	En avant la musique	(DVD)

TABLEAU A.1 – Droits des enregistrements de C-Humour.

3 Fichiers sons

C-Humour compte 26 enregistrements sonores au format `wav`[3]. Pour chaque enregistrement, seule la parole du locuteur principal (humoriste) a été transcrite et alignée.

4 Métadonnées

Les métadonnées de C-Humour sont présentées dans quatre fichiers. Le principal regroupe toutes les informations par locuteur dans chaque situation d'enregistrement (`04_CHUM_Meta_all.xlsx`). Il présente l'avantage de reprendre les informations des trois autres systèmes d'agrégation :

1. `01_CHUM_Meta_Genre.xlsx`, couple les caractéristiques situationnelles aux 12 situations de communication ;
2. `02_CHUM_Meta_Situation.xlsx`, spécifie celles relatives à la situation d'enregistrement ;
3. `03_CHUM_Meta_Speaker.xlsx`, décrit les caractéristiques de chaque locuteur présent dans chaque situation d'enregistrement.

4.1 Les locuteurs

Le Tableau A.2 reprend un échantillons des données relatives aux locuteurs.

Code	Prénom - Nom	Age	Expérience	Situation familliale	Domicile actuel
chfAR	Anne Roumanoff	47	25	Mariée, deux filles	Paris
chfBK	Bérangère Krief	29	10	Célibataire, sans enfant	Paris
chfEG	Emma Gattuso	30	4	Célibataire, sans enfant	Paris, Cannes
chfFF	Florence Foresti	39	15	Divorcée, un enfant	Paris

3. La majorité des enregistrements du corpus (70%) a une fréquence d'échantillonnage à 44100 Hz. La fréquence d'échantillonnage minimum est de 16000 Hz (20% des enregistrements). Enfin, 10% des enregistrements ont une fréquence d'échantillonnage de 22050 Hz.

chmAT	Arnaud Tsamère	37	8	Célibataire, sans enfant	Paris
chmKA	Kev Adams	21	4	Célibataire, sans enfant	Paris
chmOB	Olivier De Benoist	38	10	Marié, trois enfants	Reims
chmSG	Stephane Guillon	49	23	Marié, sans enfant	Paris

TABLEAU A.2 – Echantillon des métadonnées relatives aux locuteurs.

4.2 Les traits situationnels et situations

Les traits situationnnels et situations de C-Humour sont décrits au Chapitre 3 section 1.3.

5 Les tires d'annotations de C-Humour

L'Annexe C reprend l'ensemble des annotations du corpus C-Humour exploitées dans le cadre de cette étude [4] (manuelle, automatique ou semi-automatique), les protocoles d'annotations utilisés, les prérequis techniques à leur réalisation et assurant leur reproductibilité. Le Tableau A.3 liste les tires présentes dans le corpus C-Humour, leur définition sommaire et le contexte de leur développement.

n.	nom	définition
1	**phones**	Segmentation et alignement phonétique (caractères phonétiques SAMPA) (EasyAlign)
2	**syll**	Segmentation syllabique (caractères phonétiques SAMPA) (EasyAlign)
3	**promauto**	Annotation automatique des degrés de proéminence des syllabes (ProsoProm)
4	**if**	Annotation syllabique de la position de la syllabe au sein du mot (lex) (plugin ProsoBox)
5	**delivery**	Annotation manuelle des proéminences et des phénomènes d'hésitation qui interfèrent avec les proéminences (selon le protocole C-PROM)
6	**tok-min**	Segmentation lexicale automatique à partir de la tire de transcription (DisMo)
7	**words**	Segmentation lexicale automatique à partir de la tire de transcription (EasyAlign)
8	**pos-dismo**	Annotation lexicale automatique des catégories grammaticales (DisMo)

4. D'autres annotations ont été produites dans le cadre de GROSMAN (2013a).

n.	nom	définition
9	**pos-treetag**	Annotation lexicale automatique des catégories grammaticales (TreeTagger)
10	**lex**	Annotation lexicale automatique en items lexicaux vs. grammaticaux à partir de l'annotation grammaticale (plugin ProsoBox)
11	**ga-accentuable**	Segmentation automatique en groupes accentuables (*chunks*) à partir de la tire « lex » (plugin ProsoBox)
12	**ga-accentué**	Segmentation automatique en groupes accentués (proéminence finale) à partir de la tire « lex » et « promauto » (script Praat)
13	**disfluency**	Annotation manuelle des (dis)fluences au niveau lexical (selon le protocole de Crible *et al.* (2016))
14	**diacritics**	Annotation manuelle contextualisée des (dis)fluences au niveau lexical ((selon le protocole de Crible *et al.* (2016))
15	**disfluency-seq**	Identification automatique des séquences/cooccurrences de (dis)fluences à partir de la tire « disfluency » (script Praat)
16	**speaker**	Segmentation et annotation manuelles en tours de parole et identification des chevauchements
17	**ipu-file**	Segmentation automatique des tours de parole du locuteur principal en unités inter-pausales en fonction du seuil de pauses individuel moyen du locuteur
18	**ipu-genre**	Segmentation automatique des tours de parole du locuteur principal en unités inter-pausales en fonction du seuil de pauses individuel moyen par situation de parole
19	**comment**	Annotation manuelle libre d'éléments lexicaux et supra-lexicaux sujets à discussion

Tableau A.3 – Résumé des tires d'annotations présentes dans C-Humour.

6 Diffusion et utilisation du corpus

Le corpus est disponible et diffusé sous forme de table `csv` ou TextGrid. Pour accéder aux annotations originales au format TextGrid, aux sons et vidéos, veuillez prendre contact directement avec la responsable du corpus : Iulia Grosman et/ou avec le centre Valibel (UCL).

Pour citer ce corpus :

– Grosman I. (2013) *Variations phonostylistiques et patrons de bipolarisations rythmiques. Etude sur corpus : la parole des humoristes français*. Mémoire de fin d'études de Master en Linguistique – Finalité approfondie. Université Catholique Louvain.

– Grosman I. (2016) *How do French Humorists manage their persona across situations ? A corpus study on their prosodic variation*. In *Metapragmatics of humor : Current Research Trends*. John Benjamins. pp. 147-177.

Annexe B

Description de LOCAS-F : *Louvain Corpus of Annotated Speech - French*

1 Enregistrements et annotations

Le corpus LOCAS-F (*LOuvain Corpus of Annotated Speech – French*) (Martin, Degand & Simon 2014) est un corpus oral multigenre en français. Il a été développé au cours de deux projets :

- FRFC – Périphérie gauche des unités de discours (analyse syntaxique, prosodique et fonctionnelle et étude de l'impact sur la macro-organisation du discours) (2010-2014 – convention FRFC 2.4524.11 (17468261))
- ARC « Fluency –Disfluency » : Fluence et disfluence. Étude comparative des marques qui facilitent l'interprétation (2012-2017 – Action de recherche concertée, convention n ° 12/17-044).

Ce développement a été permis grâce à l'implication des chercheurs suivants : Liesbeth Degand, Anne Catherine Simon, Thomas Van Damme, Julien Eychenne, Noalig Tanguy, et Laurence Martin. Par la suite, il a été exploité dans diverses thèses et a été annoté notamment par Georges Christodoulides, Ludivine Crible, Emilie Duviver, Lucie Rousier-Vercruyssen et Iulia Grosman.

2 Structure du corpus

Le corpus LOCAS-F est structuré en deux dossiers :

1. Le dossier [CORPUS] regroupe les enregistrements (fichiers sons) et la stylisation du pitch de chaque fichier . Il contient une série d'annotations : les TextGrids et tables associés, le profil prosodique de chaque enregistrement et sa transcription orthographique.
2. Le dossier [METADATA] comprend ce manuel ainsi que les fichiers de métadonnées consignant les caractéristiques de chaque situation d'enregistrement et des locuteurs. Un fichier synthétise ces informations.

3 Droits et origines du corpus

Les enregistrements du corpus LOCAS-F proviennent de plusieurs projets :

Date	**Projet**	**Titre**	**Bailleurs de fonds & n° de convention**	**Chercheurs**
2007-2008	MDU	Procédure de segmentation du discours oral en unités minimales (MDU) sur la base de critères syntaxiques et prosodiques	FNRS-FRFC n°2.4535.02	L. Degand A.C Simon A. Dister
2007-2010	C-PROM	Un corpus prosodique annoté de référence pour l'étude du français	-	M. Avanzi A.C Simon J-P. Goldman A. Auchlin
2010-2014	Périphérie Gauche	Périphérie gauche des unités de discours (analyse syntaxique, prosodique et fonctionnelle et étude de l'impact sur la macro-organisation du discours)	FRFC 2.4524.11 (17468261)	AC. Simon L. Degand T. Vandamme N. Tanguy L. Martin

Date	Projet	Titre	Bailleurs de fonds & n° de convention	Chercheurs
2012-2017	Fluency-Disfluency	Fluence et disfluence. Étude comparative des marques qui facilitent l'interprétation.	ARC FWB n° 12/17-044	L. Degand AC. Simon I. Grosman C. Christodoulides S. Roekhaut L. Crible

Tableau B.1 – Liste des projets contributeurs de LOCAS-F.

Ainsi, si certains enregistrements sont originaux, d'autres proviennent de la banque de données Valibel (Simon, Francard & Hambye 2014) et de C-PROM (Avanzi, Simon, Goldman & Auchlin 2010).

4 Fichiers sons

LOCAS-F compte 48 enregistrements sonores au format `wav`[1]. Ils se répartissent en 17 monologues, 27 dialogues et 5 multilogues. La durée totale du corpus s'élève à 3.59 heures (215.76 minutes).

Pour chaque enregistrement, on a séparé les annotations correspondant à la parole d'un locuteur dans des fichiers différents. Par conséquent, on retrouve autant de paires fichier sons-textgrids que de locuteurs dans une situation de parole (dans le cas d'un dialogue : `wav_louteur1-fichier_locuteur1` et `wav_louteur2-fichier_locuteur2`). Au total, on compte 17 locuteurs en situation de monologue, 41 locuteurs en situation dialogale et 19 locuteurs en situation multilogale. Le corpus LOCAS-F regroupe ainsi 63 locuteurs différents.

5 Metadonnées

Les métadonnées de LOCAS-F sont présentées dans quatre fichiers. Le principal regroupe toutes les informations par locuteur dans chaque situation d'enregistrement (`04_LOCASF_Meta_all.xlsx`). Il présente l'avantage de reprendre les informations des trois autres systèmes d'agrégation :

1. La majorité des enregistrements du corpus (58%) a une fréquence d'échantillonnage à 44100 Hz. La fréquence d'échantillonnage minimum est de 16000 Hz (23% des enregistrements). Enfin, 19% des enregistrements ont une fréquence d'échantillonnage de 22050 Hz.

1. `01_LOCASF_Meta_Genre.xlsx`, couple les caractéristiques situationnelles aux 12 situations de communication ;
2. `02_LOCASF_Meta_Situation.xlsx`, spécifie celles relatives à la situation d'enregistrement ;
3. `03_LOCASF_Meta_Speaker.xlsx`, décrit les caractéristiques de chaque locuteur présent dans chaque situation d'enregistrement.

5.1 Les locuteurs

En accord avec les conventions VALIBEL (BACHY et al. 2007), chaque locuteur de LOCAS-F est identifié par un code unique composé de deux initiales et un chiffre. Les métadonnées supplémentaires incluent leur nom lorsqu'il s'agit de personnalités publiques[2], leur sexe (`F/H`), leur nationalité (`be-fr-ch`), leur langue maternelle, leur relation aux métiers de la communication (`com`/`non-com`), leur date de naissance, leur âge et appartenance à une tranche d'âge (au moment de l'enregistrement), leur lieu de résidence (pays, et ville lorsque possible), leur profession et les études poursuivies.

Les locuteurs de LOCAS-F sont tous francophones natifs. Ils sont d'origines géographiques différentes. Au sein des 77 fichiers, on retrouve dans 37 situations des locuteurs belges, contre 35 situations à locuteurs français et 5 à locuteurs suisses. La répartition du nombre total de locuteurs est relativement équilibrée au niveau du sexe : les femmes sont présentes dans 43 % des échantillons contre 57 % pour les hommes. La proportion des catégories d'âges par échantillon est représentée au Tableau B.2.

Catégorie d'âge	**0-25**	**26-44**	**46-65**	**66+**	**n.a**	**Total %**
Nombre de fichiers	21	17	29	1	32	100

TABLEAU B.2 – Proportion d'échantillons classés par catégories d'âge.

5.2 Les traits situationnels

LOCAS-F envisage l'étude de la variation stylistique et des registres au travers des situations de parole. Si celles-ci se distinguent en fonction des spécificités d'une *tâche communicative* (e.g. discours politique, homélie, interview radiophonique, etc.), c'est la description componentielle en traits situationnels qui a guidé la constitution du corpus.

Ce processus de description des paramètres qui influencent le type de langage produit est hérité de travaux antérieurs (CRIBLE, DUMONT, GROSMAN & NOTARRIGO 2014 ; GOLDMAN, FRANÇOIS, ROEKHAUT & SIMON 2010 ; KOCH & OSTERREICHER 2001 ;

2. L'anonymisation des transcriptions et des `wav` n'a pas encore été effectuée à ce stade du développement du corpus.

Llisteri 1992 ; Simon, Auchlin, Avanzi & Goldman 2010). Le Tableau B.3 reprend la définition de chaque trait situationnel utilisé pour la catégorisation de situations de parole.

Traits	**Définitions**
D. ELICIT	**Degré d'élicitation** La présence et le poids du protocole expérimental agissant comme une contrainte sur l'interaction
dirigé	Production artificielle due à un protocole expérimental rigide dans le cadre d'une recherche scientifique exerçant un contrôle fort sur la production du locuteur et sur des variables linguistiques.
semi-dirigé	Production naturelle supervisée par un protocole expérimental souple dans le cadre d'une recherche scientifique guidant le choix du topique mais ne contraignant pas le locuteur à choisir une formulation préétablie. Le locuteur est libre du choix de ses mots.
naturel	Production authentique du locuteur dénuée de tout protocole expérimental, non suscitée par une action de recherche scientifique.
Nombre de locuteurs	**Catégorisation du nombre de locuteurs** Le nombre de locuteurs prenant une part active dans l'interaction
monologue	Une seule personne est le locuteur principal désigné.
dialogue	Deux personnes sont les locuteurs principaux désignés ;
multilogue	Plus de deux locuteurs principaux prennent part à l'interaction.
D. PREP	**Degré de préparation** Le degré de la préparation orale et/ou écrite du discours du locuteur principal
préparé	Le locuteur a entièrement préparé avant l'interaction le contenu et la forme de son discours. Le discours a été rédigé et est formulé avec ou sans support textuel.
semi-préparé	Le locuteur a préparé la direction globale de son discours, mais le discours n'a pas été rédigé dans sa totalité. L'interaction peut inclure un support textuel partiel.
spontané	Le locuteur improvise son discours au fur-et-à-mesure de l'interaction et n'en a préparé aucune partie.
D. INTERACT	**Degré d'interactivité** La capacité d'adaptation des interlocuteurs au regard des attentes sur le statut des locuteurs dans l'interaction
interactif	Relation symétrique entre les locuteurs. La situation en cours permet aux interlocuteurs de prendre la parole.

Traits	Définitions
semi-interactif	Relation asymétrique où un locuteur occupe plus la parole qu'un autre. La situation de parole n'exclut pas des interventions ponctuelles des locuteurs secondaires.
non interactif	Relation asymétrique où un des locuteurs prend et garde de manière quasi-continue la parole, sans possibilité d'intervention des interlocuteurs.
D.MEDIA	**Degré de médiatisation** L'importance de la diffusion sur les grands médias comme objectif principal de l'interaction
médiatisé	Le but principal de l'interaction est d'être diffusé par un média
semi-médiatisé	L'interaction aurait lieu même si elle n'était pas diffusée par un média
non médiatisé	L'interaction n'est diffusée par aucun média
C.PRO	**Catégorie de professionnalisation** La situation est ou n'est pas suscitée par la profession des interlocuteurs
pro	La situation est suscitée par la profession d'au moins un des interlocuteurs.
privé	La situation n'est pas suscitée par la profession des interlocuteurs.

Tableau B.3 – Définitions des traits situationnels utilisés – (Cribble 2014).

5.3 Les situations de parole

Quatorze types de situations de parole sont représentés : des discours académiques, des conférences scientifiques, des dialogues conversationnels formels (professionnels), des dialogues conversationnels informels, des débats politiques, des homélies, des interviews sociolinguistiques, des interviews télévisées, des interviews radiophoniques, des journaux parlés télévisés, des lectures, des discours politiques, des récits conversationnels et des performances radiophoniques. Le Tableau B.4 résume les particularités de chacun de ces types de situation ainsi que les traits situationnels associés.

Situation de parole	Traits situationnels et caractérisation en faisceaux
Discours académiques	Monologue - non interactif - préparé - semi-médiatique - professionnel - naturel Discours d'ouverture de la rentrée académique des représentants UCL (recteur, président de l'Assemblée Générale des étudiants de Louvain, président du corps scientifique) (belge)
Conférences scientifiques	Monologue - non interactif - semi-préparé - non médiatique - professionnel - naturel

Situation de parole	Traits situationnels et caractérisation en faisceaux
	Conférences scientifiques par des linguistes enseignants-chercheurs (belges, français et suisse)
Interactions formelles *(asymétriques)*	Monologue/dialogue – interactif/semi-interactif - semi-préparé/non préparé - non médiatique - professionnel - naturel
	Dialogues en contexte professionnel : entretien entre un étudiant et son professeur, entretien d'orientation professionnel, cours magistral en secondaire (belge)
Interactions informelles *(symétriques)*	Dialogue/multilogue - interactif - non préparé - non médiatique – privée & professionnel - naturel
	Discussions informelles entre des collègues ou amis (discussion de sortie, d'une réunion de KOT, d'un voyage scolaire) (belge)
Débats politiques	Multilogue - interactif - semi-préparé - médiatique - professionnel – naturel
	Débats télévisés (France 2, France Inter et I-Télé) de personnages politiques et d'un romancier (français)
Homélies	Monologue - non interactif - préparé - semi-médiatique - professionnel – naturel
	Homélies de 2008 prononcées lors d'une cérémonie religieuse (belge, français, suisse)
Interviews sociolinguistiques	Dialogue - semi-interactif - non préparé - non médiatique - professionnel - semi-élicité
	Entretiens sociolinguistiques étudiants menés auprès d'enseignants à propos de la perception et l'enseignement de la norme langagière du français (belge)
Interviews télévisées	Dialogue - semi-interactif - non préparé - médiatique - professionnel – naturel
	Discussions d'un animateur avec trois pour une émission de la télévision belge francophone informelle (RTBF - Hep Taxi) (belge, français)
Interviews radiophoniques	Dialogue - semi-interactif - non préparé - médiatique - professionnel – naturel
	Interviews radiophoniques d'écrivains à propos de leur dernier livre paru (belge, français)
Journaux parlés	Dialogue - non interactif - préparé - médiatique - professionnel – naturel
	Journaux parlés radiophoniques de chaînes belges, françaises et suisses
Lectures (radiophoniques)	Dialogue - non interactif - préparé - médiatique - professionnel – naturel

Situation de parole	**Traits situationnels et caractérisation en faisceaux**
	Lectures radiophoniques d'un extrait d'œuvre par leur auteur (français)
Discours politiques	Dialogue - non interactif - préparé - semi-médiatique/médiatique - professionnel – naturel
	Discours télévisé officiel de 3 chefs d'état ou de gouvernement (belge, français et suisse)
Récits conversationnels	Monologue - semi-interactif - non préparé - non médiatique - privée - semi-élicité
	Narrations spontanées issues de conversations (belge, français, suisse)
Performances radiophoniques	Monologue - non interactif - semi-préparé - médiatique - professionnel – naturel
	Emission « à micro-ouvert » sur Arte Radio de trois animateurs radio invités à parler librement à l'antenne pendant une durée de +/- 3 minutes

TABLEAU B.4 – Résumé des types de situations et traits situationnels associés.

6 Tires d'annotation de LOCAS-F

L'Annexe C reprend l'ensemble des annotations du corpus LOCAS-F (manuelle, automatique ou semi-automatique), les protocoles d'annotations utilisés, les prérequis techniques à leur réalisation et assurant leur reproductibilité. Le Tableau B.5 liste les tires présentes dans le corpus LOCAS-F, leur définition sommaire et le contexte de leur développement.

n.	nom	définition	projet
1	**phones**	Segmentation et alignement phonétique (caractères phonétiques SAMPA) (EasyAlign)	C-PROM ARC Fluency
2	**syll**	Segmentation syllabique (caractères phonétiques SAMPA) (EasyAlign)	C-PROM ARC Fluency
3	**promauto**	Annotation automatique des degrés de proéminence des syllabes (ProsoProm)	ARC Fluency
4	**if**	Annotation syllabique de la position de la syllabe au sein du mot (lex) (plugin ProsoBox)	ARC Fluency
5	**delivery**	Annotation manuelle des proéminences et des phénomènes d'hésitation qui interfèrent avec les proéminences (selon le protocole C-PROM)	ARC Fluency
6	**tok-min**	Segmentation lexicale automatique à partir de la tire de transcription (DisMo)	ARC Fluency
7	**words**	Segmentation lexicale automatique à partir de la tire de transcription (EasyAlign)	C-PROM FRFC MDU ARC Fluency
8	**pos-dismo**	Annotation lexicale automatique des catégories grammaticales (DisMo)	ARC Fluency
9	**pos-treetag**	Annotation lexicale automatique des catégories grammaticales (TreeTagger)	ARC Fluency
10	**lex**	Annotation lexicale automatique en items lexicaux vs. grammaticaux à partir de l'annotation grammaticale (plugin ProsoBox)	ARC Fluency
11	**ga-accentuable**	Segmentation automatique en groupes accentuables (*chunks*) à partir de la tire « lex » (plugin ProsoBox)	ARC Fluency
12	**ga-accentué**	Segmentation automatique en groupes accentués (proéminence finale) à partir de la tire « lex » et « promauto » (script Praat)	ARC Fluency
13	**disfluency**	Annotation manuelle des (dis)fluences au niveau lexical (selon le protocole de Crible, Dumont, Grosman & Notarrigo (2016))	ARC Fluency

n.	nom	définition	projet
14	**diacritics**	Annotation manuelle contextualisée des (dis)fluences au niveau lexical (selon le protocole de Crible, Dumont, Grosman & Notarrigo (2016))	ARC Fluency
15	**disfluency-seq**	Identification automatique des séquences/cooccurrences de (dis)fluences à partir de la tire « disfluency » (script Praat)	ARC-Fluency
16	**rection**	Segmentation et annotation manuelles en unités syntaxiques de rection (dépendance) (selon le protocole de Tanguy et al. (2012))	FRFC MDU
17	**sequences**	Segmentation et annotation manuelles en séquences syntaxiques fonctionnelles (selon le protocole de Tanguy *et al.* (2012))	FRFC MDU
18	**boundary**	Marquage (*point-tier*) manuel des frontières perçues (selon le protocole décrit dans Mertens & Simon (2013))	FRFC MDU
19	**contours**	Annotation manuelle des contours intonatifs perçus aux frontières prosodiques (selon le protocole décrit dans Mertens & Simon (2013))	FRFC MDU
20	**bdu-auto**	Segmentation et annotation automatiques en unités de base du discours à partir des unités de rection et des contours intonatifs (scripts Perl) selon les règles décrites dans Degand & Simon (2009))	FRFC MDU
21	**speaker**	Segmentation et annotation manuelles en tours de parole et identification des chevauchements	ARC Fluency
22	**ipu-file**	Segmentation automatique des tours de parole du locuteur principal en unités interpausales en fonction du seuil de pauses individuel moyen du locuteur	ARC Fluency
23	**ipu-genre**	Segmentation automatique des tours de parole du locuteur principal en unités interpausales en fonction du seuil de pauses individuel moyen par situation de parole	ARC Fluency
24	**comment**	Annotation manuelle libre d'éléments lexicaux et supra-lexicaux sujets à discussion	ARC Fluency

Tableau B.5 – Résumé des tires d'annotations présentes dans LOCAS-F.

Annexe C

Description et référence des annotations

Cette Annexe reprend la description détaillée de chaque couche d'annotation (*tire*). Par souci de clarté, la présentation respecte l'ordre d'apparition de la tire dans un fichier d'annotation [1] et non son ordre de production.

Chaque section décrit une couche d'annotation et ses dépendances à d'autres annotations (Tableau C.1). Pour chacune, on résume les étiquettes utilisées et les références bibliographiques mobilisées. Enfin, lorsque nécessaire, on informe du fonctionnement et de la localisation des scripts associés à la production de chaque couche d'annotation.

n.	nom	définition
1	**phones**	Segmentation et alignement phonétique (caractères phonétiques SAMPA) (EasyAlign)
2	**syll**	Segmentation syllabique (caractères phonétiques SAMPA) (EasyAlign)
3	**promauto**	Annotation automatique des degrés de proéminence des syllabes (ProsoProm)
4	**if**	Annotation syllabique de la position de la syllabe au sein du mot (lex) (plugin ProsoBox)
5	**delivery**	Annotation manuelle des proéminences et des phénomènes d'hésitation qui interfèrent avec les proéminences (selon le protocole C-PROM)
6	**tok-min**	Segmentation lexicale automatique à partir de la tire de transcription (DisMo)
7	**words**	Segmentation lexicale automatique à partir de la tire de transcription (EasyAlign)
8	**pos-dismo**	Annotation lexicale automatique des catégories grammaticales (DisMo)

1. Au format TextGrid (Boersma & Weenink 2015).

n.	nom	définition
9	**pos-treetag**	Annotation lexicale automatique des catégories grammaticales (TreeTagger)
10	**lex**	Annotation lexicale automatique en items lexicaux vs. grammaticaux à partir de l'annotation grammaticale (plugin ProsoBox)
11	**ga-accentuable**	Segmentation automatique en groupes accentuables (*chunks*) à partir de la tire « lex » (plugin ProsoBox)
12	**ga-accentué**	Segmentation automatique en groupes accentués (proéminence finale) à partir de la tire « lex » et « promauto » (script Praat)
13	**disfluency**	Annotation manuelle des (dis)fluences au niveau lexical (selon le protocole de Crible, Dumont, Grosman & Notarrigo (2016))
14	**diacritics**	Annotation manuelle contextualisée des (dis)fluences au niveau lexical (selon le protocole de Crible, Dumont, Grosman & Notarrigo (2016))
15	**disfluencyseq**	Identification automatique des séquences/cooccurrences de (dis)fluences à partir de la tire « disfluency » (script Praat)
16	**rection**	Segmentation et annotation manuelles en unités syntaxique de rection (dépendance) (selon le protocole de Tanguy (2012))
17	**sequences**	Segmentation et annotation manuelles en séquences syntaxiques fonctionnelles (selon le protocole de Tanguy (2012))
18	**boundary**	Marquage (*point-tier*) manuel des frontières perçues (selon le protocole décrit dans Mertens & Simon (2013))
19	**contours**	Annotation manuelle des contours intonatifs perçus aux frontières prosodiques (selon le protocole décrit dans Mertens & Simon (2013))
20	**bdu-auto**	Segmentation et annotation automatiques en unités de base du discours à partir des unités de rection et des contours intonatifs (scripts Perl) selon les règles décrites dans Degand & Simon (2009)
21	**speaker**	Segmentation et annotation manuelles en tours de parole et identification des chevauchements
22	**ipu-file**	Segmentation automatique des tours de parole du locuteur principal en unité interpausales en fonction du seuil de pauses individuel moyen du locuteur
23	**ipu-genre**	Segmentation automatique des tours de parole du locuteur principal en unité interpausales en fonction du seuil de pauses individuel moyen par situation de parole
24	**comment**	Annotation manuelle libre d'éléments lexicaux et supra-lexicaux sujets à discussion

Tableau C.1 – Sommaire du détail des tires d'annotation.

1 Tire 1 : phones

Nom de tire : phones

Tire de dépendance en amont du traitement :
transcription (absente des TextGrids)

Tire de dépendance en aval du traitement :
syll

Méthodologie d'annotation :
automatique (EasyAlign) – vérification manuelle

Référence bibliographique :
J. Goldman. 2011. « EasyAlign: an automatic phonetic alignment tool under Praat ». *Proceedings of Interspeech 2011, 12th Annual Conference of the International Speech Communication Association.* Firenze, Italy : ISCA

La segmentation phonétique est obtenue automatiquement à partir d'une transcription orthographique manuelle [2] via EasyAlign (Goldman 2011), plugin développé sous Praat (Boersma & Weenink 2015). Le plugin phonétise la transcription orthographique (SAMPA – *Speech Assessment Methods Phonetic Alphabet*) [3], alphabet phonétique informatiquement lisible en caractère ASCII. L'alphabet SAMPA phonétique inclut les caractères suivant :

- **Les plosives « p, b, t, d, k, g » :** pont, [po~] ; bon, [bo~] ; temps, [ta~] ; dans, [da~] ; quand, [ka~] ; gant, [ga~] ;
- **Les fricatives « f, v, s, z, S, Z »** :femme, [fam] ; vent, [va~] ; sans, [sa~] ; zone, [zon] ; champ [Sa~] ; gens, [Za~] :
- **La glide fricative « j »** : ion, [jo~] ;
- **Les nasales « m, n, J, N »** : mon, [mo~] ; nom, [no~] ; oignon [oJo~] ; camping [ka~piN] ;
- **Les liquides « l, R » et les voyelles glides « w, H, j »** : long, [lo~] ; rond, [Ro~] ; coin [kwe~] ; juin, [ZHe~] ; pierre, [pjER]
- **Les 12 voyelles « i, e, E, a, A, O, o, u, y, 2, 9, @ » et 4 voyelles nasales « e~, a~, o~, 9~ »** : si, [si] ; ses, [se] ; seize, [sEz] ; patte, [pat] ; pâte, [pAt] ; comme, [kOm] ; gros, [gro] ; doux, [du] ; du, [dy] ; deux, [d2] ; neuf, [n9] ; justement, [Zyst@ma~] ; vin, [ve~] ; vent, [va~] ; bon, [bo~] ; brun, [br9~]

Les annotateurs ont corrigé l'annotation phonétique pour tenir compte des phénomènes suivants : coarticulations, élisions, dévoisements, diphtongaisons, liaisons, etc.

2. Cette annotation orthographique n'est pas reprise dans les fichiers TextGrid. Les transcriptions sont disponibles au format « .doc » dans le dossier [CORPUS].
3. Contenu complet accessible à l'adresse suivante : *http://www.phon.ucl.ac.uk/home/sampa/french.htm*

2 Tire 2 : syll

Nom : syll

Tires de dépendances en amont du traitement :

phones - words

Tire de dépendance en aval du traitement :

delivery - promauto - if

Méthodologie d'annotation :

automatique (EasyAlign) - vérification manuelle

Références bibliographiques :

J. Goldman. 2011. « EasyAlign: an automatic phonetic alignment tool under Praat ». *Proceedings of Interspeech 2011, 12th Annual Conference of the International Speech Communication Association.* Firenze, Italy : ISCA B. Bigi, C. Meunier, R. Bertrand & I. Nesterenko. 2010. « Annotation automatique en syllabes d'un dialogue oral spontané ». *XXVIIIèmes Journées d'étude sur la Parole.* Mons, Belgique, 181–184

La tire syllabique est créée concurremment à la tire « phones » via EasyAlign (Goldman 2011). Chaque voyelle donne lieu à une syllabe. Les segments silencieux sont marqués par un tiret bas « _ ». La syllabation suit un système de règles énoncé par Bigi, Meunier, Bertrand & Nesterenko (2010) (Figure C.1) :

(1) Il n'y a qu'une voyelle par syllabe ;

(2) Tendance universelle à privilégier les syllabes ouvertes ;

(3) En conséquence, la consonne est assignée à la seconde syllabe ;

(4) Dans un groupe de consonnes intervocaliques, le maximum de consonnes doit être attribué à l'*onset* de la seconde syllabe plutôt qu'à la *coda* de la première.

Tab. 1: Règles générales

	Séquence	Règle	Exemples
1	VV	V.V	poëte : po.ɛt
			il y a un : i.ɑ.œ̃
			en haut : ɑ̃.o
2	VXV	V.XV	limité : li.mi.te
			et donc on : e.dɔ̃.kɔ̃
3	VXXV	VX.XV	jardin : ʒɑʁ.dɛ̃
			comme ça : kom.sɑ
			parce qu'il : pɑs.ki
4	VXXXV	VX.XXV	avec moi : ɑ.vek.mwa
			cheval noir : sə.val.nwɑʁ
5	VXXXXV	VX.XXXV	il se présentait : il.spʁe.zɑ̃.te
6	VXXXXXV	VXX.XXXV	alors je crois : ɑ.lorʒ.krwɑ

Tab. 2: Règles exceptions

	Séquence	Règle	Exemples
1	VXGV	V.XGV	baignoire : be.nwɑʁ
			spéciaux : spe.sjo
			tu vois : tu.vwɑ
2	VFLV	V.FLV	découvre : de.ku.vʁə,
3	VOLV	V.OLV	il trouve : i.tʁuv
			mais de la : me.dlɑ
4	VFLGV	V.FLGV	effroyable : e.fʁwɑ.jɑbl
5	VOLGV	V.OLGV	incroyable : ɛ̃.kʁwɑ.jɑbl
6	VOLOV	VOL.OV	connaître tu : ko.nɛtʁ.ty
			capable parce : ka.pabl.pɑs

Figure C.1 – Résumé schématique des règles de syllabation – extrait de (Bigi, Meunier, Bertrand & Nesterenko 2010). *V = voyelle, X = consonne, G = glide, F = fricative, L = liquide.*

3 Tire 3 : delivery

Nom de tire : delivery

Tire de dépendance en amont du traitement :

syll - disfluency

Tire de dépendance en aval du traitement :

promauto

Méthodologie d'annotation :

manuelle

Référence bibliographique :

- M. Avanzi, A. C. Simon, J. Goldman & A. Auchlin. 2010. « C-PROM : An Annotated Corpus for French Prominence Study ». *Proceedings of Prosodic Prominence, Speech Prosody 2010 Workshop*. Chicago, Illinois, USA
- M. Avanzi, A. Lacheret-Dujour, N. Obin & B. Victorri. Jan. 2011. « Vers une modélisation continue de la structure prosodique: le cas des proéminences syllabiques ». *Journal of French Language Studies* 21 (1), 53–71

Cette tire d'exclusion décrit les phénomènes para et extralinguistiques influençant l'identification des proéminences syllabiques. Dans le corpus C-PROM annoté manuellement pour les syllabes proéminentes, cette tire contenait, outre l'annotation des syllabes proéminentes elles-mêmes, l'identification des syllabes dont les caractéristiques pouvaient interférer avec la perception des proéminences (e.g. un marqueur de syllabe allongée à cause d'une hésitation est-elle proéminente ou non ?). Cette méthodologie est décrite dans Avanzi, Lacheret-Dujour, Obin & Victorri (2011), empruntant elle-même à celle de Avanzi, Simon, Goldman & Auchlin (2010).

Le protocole d'annotation d'Avanzi, Simon, Goldman & Auchlin (2010) est résumé au Tableau C.5.

Label	Description
a	Allongement connecté à une hésitation
_	Pause silencieuse
z	Phénomène d'hésitation sans actualisation d'allongement d'hésitation
!	Morphème ou constituant interrompu (répétitions, amorces de mots, etc.)
2	Syllabes post-toniques non accentuées et *euh*
$	Appendice prosodique
%	"junk" (bruits, rires, toux, etc.), chevauchement
#	Paraverbal (he, mh, hein), rire, ricanement, toux sans chevauchement
+	Chevauchement ou second locuteur

Tableau C.5 – Référence des annotations de la tire `delivery`.

4 Tire 4 : promauto

Nom de tire : promauto

Tire de dépendance en amont du traitement :
syll - delivery

Tire de dépendance en aval du traitement :
ga-actualisé

Méthodologie d'annotation :
automatique (prosograd)

Référence bibliographique :

- A. C. Simon, M. Avanzi & J. Goldman. 2008. « La détection des proéminences syllabiques. Un aller-retour entre l'annotation manuelle et le traitement automatique ». *Actes du Congrès Mondial de Linguistique Française 2008.* EDP Sciences
- J. Goldman, M. Avanzi, A. Auchlin & A. C. Simon. 2012. « A continuous prominence score based on acoustic features ». *Proceedings of the 13th Annual Conference of the International Speech Communication Association.* Portland, Oregon : ISCA

Reproduisant la méthodologie proposée par Simon, Avanzi & Goldman (2008) et Goldman, Avanzi, Auchlin & Simon (2012), la tire `promauto` caractérise le degré de proéminence de chaque syllabe en lui assignant un score de 0 (syllabe non proéminente) à 4 (proéminence forte).

L'assignation d'un score à une syllabe est basée sur la durée de la syllabe (en millisecondes) et le maximum de f0 (en semi-tons). Ces mesures sont ensuite relativisées en vertu du contexte syllabique immédiat (i.e. relativement aux deux syllabes précédentes et à la syllabe suivante). La détection de la proéminence se base ainsi sur :

- Les durées relatives des syllabes (en pourcentage),
- Les mesures de différences relatives de f0 (en semi-tons),
- La montée mélodique (ou mouvement mélodique intrasyllabique) dans le noyau syllabique stylisé lors d'un mouvement montant – par conséquent celle-ci est considerée nulle lors d'un mouvement descendant ou plat,
- La présence d'une pause subséquente. En raison de son rôle de marqueur de fin de groupe en francais (Lacheret-Dujour & Beaugendre 1999), la présence d'une pause de plus de 250 ms. est ignorée comme syllabe adjacente dans le calcul des paramètres relatifs.

Sur base de ces mesures, chaque paramètre est considéré comme un critère de décision autonome de la proéminence s'il dépasse un certain seuil choisi (e.g. dépassement du maximum relatif de la f0 supérieur au seuil determine). Les conditions de seuils sont suffisantes mais non exclusives – le logicien reconnaît la non-orthogonalité (i.e. l'interdépendance) des paramètres.

Par ailleurs, l'annotation `promauto` dépend des phénomènes annotés dans `delivery` et des quatre paramètres de seuils selectionnés. Afin d'éviter une colinéarité des données avec les informations de la tire `IPU.file` et sans justificatif perceptuel avéré, les paramètres pour la détection de la proéminence sont fidèles au script d'origine : le seuil de détection de la pause est fixé à 250 ms. de manière invariante. Cependant, on aura modifié en conséquence la paramétrisation des signes annotés dans la tire `delivery` (Figure C.2).

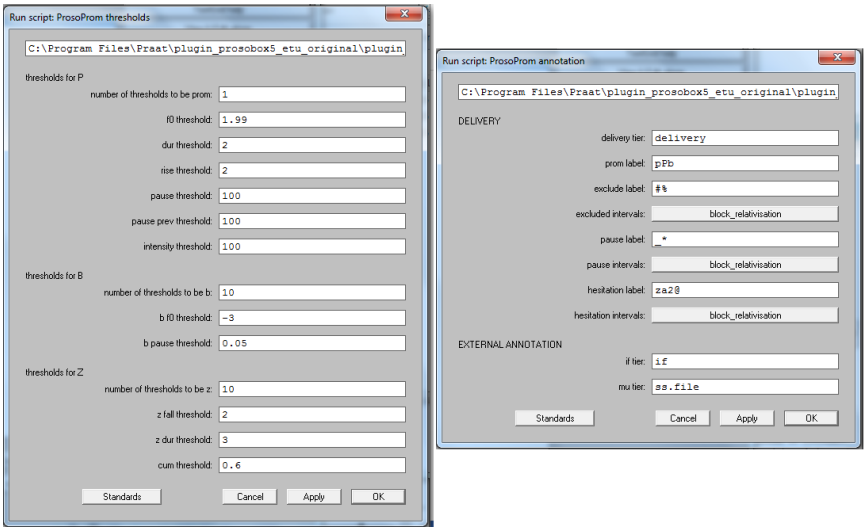


FIGURE C.2 – Paramètres et seuils d'annotation de Prosoprom.

5 Tire 5 : if

Nom de tire : if

Tire de dépendance en amont du traitement :

syll - words - lex

Tire de dépendance en aval du traitement :

ga-accentuable

Méthodologie d'annotation :

automatique

Référence bibliographique :

J. Goldman et al. 2010. « Prominence perception and accent detection in French. A corpus-based account ». *Proceedings Speech Prosody 2010*. Chicago, USA.

La tire syllabique `if` est annotée automatiquement via l'utilisation du plugin ProsoBox 5. Celle-ci assigne à chaque syllabe un code en fonction de sa position dans le mot et dans le `ga` (groupe accentuel)– elle-même dépendante de la tire `lex`. Lorsque mise en relation avec la tire `proéminence`, cette tire permet donc de caractériser le type d'accent (initial, final ou autre) dont la syllabe proéminente est porteuse.

Les phénomène annotés sont repris au Tableau C.8.

Code	Explication
.	Syllabe de mot monosyllabique ou syllabe non initiale ou finale
1	Première syllabe du groupe accentuel théorique
i	Syllabe initiale de mot bi- ou polysyllabique
p	Syllabe penultième de mot plein polysyllabique
f	Syllabe finale de mot (mono, bi et polysyllabique)

Tableau C.8 – Etiquettes d'annotation de la tire syllabique `if`.

6 Tire 6 : tok-min

Nom de tire : tok-min

Tire de dépendance en amont du traitement :
transcription (absente des TextGrids)

Tire de dépendance en aval du traitement :
POS-dismo

Méthodologie d'annotation :
automatique (DisMo)

Référence bibliographique :
G. Christodoulides, M. Avanzi & J. Goldman. 2014. « DisMo: A Morphosyntactic, Disfluency and Multi-Word Unit Annotator. An Evaluation on a Corpus of French Spontaneous and Read Speech ». *Proceedings of the 9th International Conference on Language Resources and Evaluation - LREC'14*. Reykjavik, Iceland : European Language Resources Association (ELRA), [en ligne]

La segmentation lexicale en *token* minimaux `tok-min` s'effectue automatiquement à partir du logiciel libre DisMo (v.0.4b3) (Christodoulides, Avanzi & Goldman 2014)[4]. La segmentation DisMo est celle ayant servi de base aux autres segmentations alignées au mot (e.g. annotation en parties du discours, annotation des disfluences, etc.).

Le logiciel génère une segmentation en unités lexicales minimales légèrement différente de celle dans la tire `words` produite par EasyAlign. On observe le regroupement des unités *parce* et *que* (en un intervalle, « *parce que* »), *aujourd'* et *hui* (*aujourd'hui*) ; d'autres unités polylexicales, regroupées dans la tire `words`, sont segmentées en deux items sur `tok-min` (e.g. *c'est* et *que*, *d'* et *antisemitisme*). La segmentation DisMo est parfois plus représentative de la cohésion lexicale.

Voici la paramétrisation utilisée lors du traitement des corpus dans DisMo :

(1) « Ajout des tires d'annotation DisMo au corpus sélectionné »

(2) « Créer un fichier séparé par des tabulation pour le corpus entier »

(3) « Ecraser les fichiers pré-existants sans demander »

(4) Aucun préfixe n'est ajouté aux tires générées par le programme

(5) Tires d'annotation selectionnées : (a) « Token Minimal » (d) « POS Minimal »

Aucune modification ou correction de la segmentation DisMo n'a été effectuée par la suite.

4. Logiciel libre diffusé à l'adresse *http://sourceforge.net/projects/dismo* et inclus dans le dossier `PROCESSING`.

7 Tire 7 : words

Nom de tire : words

Tire de dépendance en amont du traitement :
transcription (absente des TextGrids)

Tire de dépendance en aval du traitement :
POS-treetag – disfluency– diacritics-e-tc.

Méthodologie d'annotation :
automatique (EasyAlign) – vérification manuelle

Référence bibliographique :
J. Goldman. 2011. « EasyAlign: an automatic phonetic alignment tool under Praat ». *Proceedings of Interspeech 2011, 12th Annual Conference of the International Speech Communication Association.* Firenze, Italy : ISCA

Le plugin EasyAlign (Goldman 2011) crée une tire `words` au moment ou il génère l'alignement phonétique (`phones`) et syllabique (`syll`). Les unités graphiques (suite de caractères suivie d'un séparateur : espace, trait d'union, apostrophe) sont identifiées dans la transcription orthographique et dupliquées dans la tire `words`, en alignement avec les phonèmes correspondants.

8 Tire 8 : POS-dismo

Nom de tire : POS-dismo

Tire de dépendance en amont du traitement :
tok-min

Tire de dépendance en aval du traitement :
if – ga-théorique – ga-actualisé

Méthodologie d'annotation :
automatique (DisMo) – vérification manuelle

Référence bibliographique :
G. Christodoulides, M. Avanzi & J. Goldman. 2014. « DisMo: A Morphosyntactic, Disfluency and Multi-Word Unit Annotator. An Evaluation on a Corpus of French Spontaneous and Read Speech ». *Proceedings of the 9th International Conference on Language Resources and Evaluation - LREC'14*. Reykjavik, Iceland : European Language Resources Association (ELRA), [en ligne]

La couche d'annotation semi-automatique en parties du discours (POS – *Part Of Speech*) avec les étiquettes DisMo est alignée à la segmentation `tok-min`. Les étiquettes de DisMo sont présentées au Tableau C.13.

Cette annotation a été vérifiée manuellement à l'aide de filtres de tableur (Libre Office, Excel, etc.)[5]. Les corrections les plus fréquentes concernent les étiquettes portant sur :

- les pronoms relatifs *vs.* conjonctions de subordination (e.g. « que ») ;
- les déterminants définis *vs.* pronoms personnels d'objet direct (e.g. « le », « nous », « vous », etc.)[6].

5. La transformation des fichiers TextGrids en fichiers `csv` (*coma separated values*) lisibles par un logiciel-tableur et leur reconversion subséquente en TextGrids a été effectuée via l'utilisation de script PERL ecrits par Sophie Roekhaut et présents dans le dossier [PROCESSING].

6. Cette phase de correction ayant été effectuée après la création des tires `delivery`, `disfluency`, et `speaker`, elle a également permis la correction d'erreurs éventuelles d'annotations présentes dans ces tires (majuscules aux noms propres, oublis, fautes de frappes, fautes orthographiques, etc.).

9 Tire 9 : POS-treetag

Nom de tire : POS-treetag

Tire de dépendance en amont du traitement :
POS-dismo

Tire de dépendance en aval du traitement :
aucune

Méthodologie d'annotation :
automatique (Treetagger) – vérification manuelle

Référence bibliographique :
H. Schmid. 1994. « Probabilistic Part-of-Speech Tagging Using Decision Trees ». *Proceedings of International Conference on New Methods in Language Processing*. Manchester, UK, 44–49

La couche d'annotation `POS-Treetag` est une conversion des étiquettes DisMo en étiquettes TreeTagger (Schmid 1994). Leur nombre réduit est un avantage. Il autorise le regroupement de catégories détaillées du set d'étiquettes DisMo menant à posteriori à la simplification des analyses linguistiques impliquant les POS. Par ailleurs, leur utilisation répandue et leur étiquetage multilingue facilitent la comparaison des résultats entre différents corpus sans se restreindre à la sphère francophone.

La conversion des étiquettes DisMo vers TreeTagger a été effectuée après correction des POS de la tire `pos-dismo`, semi-automatiquement, par filtrage lors de la conversion des TextGrids en tableau. Le Tableau C.13 reprend les conversions d'étiquettes par catégorie grammaticale.

Phénomène	**DisMo**	**TreeTagger**
Silences	_	_
Eléments non grammaticaux (amorce, bruit, etc.)	0	0
Adjectif	ADJ	ADJ
Adjectif numéral	ADJ:num	
Adverbe	ADV	ADV
Adverbe de comparaison	ADV:comp	
Adverbe de gradation	ADV:deg	
Adverbe interrogatif	ADV:int	
Adverbe de négation	ADV:neg	
Conjonction de coordination	CON:coo	KON
Conjonction de subordination	CON:sub	
Déterminant défini	DET:def	DET:ART
Déterminant démonstratif	DET:dem	
Déterminant indéfini	DET:ind	
Déterminant interrogatif	DET:int	
Déterminant partitif	DET:par	

Phénomène	**DisMo**	**TreeTagger**
Déterminant possessif	DET:pos	DET:POS
Mot étranger	FRG	FRG
Interjection	ITJ	INT
Acronyme et/ou abréviation	NOM:acr	ABR
Nom commun	NOM:com	NOM
Nom propre	NOM:pro	NAM
Nombre cardinal déterminant	NUM:crd:det	NUM
Nombre cardinal nominal	NUM:crd:nom	
Nombre ordinal adjectival	NUM:ord:adj	
Elément paralinguistique (toux, rires, etc.)	PARA	PARA
Préfixe	PFX	PFX
Pronom	PRO	PRO
Pronom démonstratif	PRO:dem	PRO:DEM
Pronom indéfini	PRO:ind	PRO
Pronom interrogatif	PRO:int	PRO:INT
Pronom personnel objet direct	PRO:per:objd	PRO:PER
Pronom personnel objet indirect	PRO:per:obji	
Pronom personnel sujet	PRO:per:sjt	
Pronom personnel tonique	PRO:per:ton	
Pronom réflexif	PRO:ref	
Pronom relatif	PRO:rel	PRO:REL
Préposition	PRP	PRP
Combinaison d'une préposition et d'un déterminant	PRP:det	PRP:det
Verbe conditionnel simple	VER:cond	VER:cond
Verbe conditionnel auxiliaire	VER:cond:aux	
Verbe conditionnel imparfait	VER:cond:impf	
Verbe futur simple	VER:fut	VER:futu
Verbe futur auxiliaire	VER:fut:aux	
Verbe impératif	VER:impe	VER:impe
Verbe imparfait	VER:impf	VER:impf
Verbe imparfait auxiliaire	VER:impf:aux	
Verbe infinitif	VER:inf	VER:inf
Verbe infinitif auxiliaire	VER:inf:aux	
Verbe participe présent	VER:ppas	VER:ppre
Verbe participe présent auxiliaire	VER:ppas:aux	
Verbe participe passé	VER:ppre	
Verbe participe passé auxiliaire	VER:ppre:aux	
Verbe présent	VER:pres	VER:pres
Verbe présent auxiliaire	VER:pres:aux	
Structure présentative (« c'est », « il y a »)	VER:pres:entatif	
Verbe passé simple	VER:simp	VER:simp

Phénomène	**DisMo**	**TreeTagger**
Verbe passé simple auxiliaire	VER:simp:aux	
Verbe subjonctif imparfait	VER:subi	VER:subi
Verbe subjonctif présent	VER:subp	VER:subp
Verbe subjonctif présent auxiliaire	VER:subp:aux	

TABLEAU C.13 – Correspondance des étiquettes Dismo en étiquettes TreeTagger.

10 Tire 10 : lex

Nom de tire : lex

Tire de dépendance en amont du traitement :
POS-dismo

Tire de dépendance en aval du traitement :
ga-accentuable – ga-théorique

Méthodologie d'annotation :
automatique (ProsoBox)

Référence bibliographique :
- E. Delais-Roussarie. 2000. « Vers une nouvelle approche de la structure prosodique ». *Langue française* 126 (1), 92–112
- P. Mertens. 2011a. « A system for automatic tonal annotation of speech corpora ». *Actes d'IDP 2011*. Salford, Manchester, UK

La couche d'annotation automatique `lex` informe du caractère grammatical ou lexical d'un token. Elle indique si le mot est clitique (unité lexicale à fonctionnement accentuable) ou non clitique (unité grammaticale à fonctionnement non accentuable). Les *mots clitiques*, sont souvent monosyllabiques et *atones* (*non accentuables*). On oppose la catégorie des clitiques aux *non clitiques*, mots pleins reconnus comme des unités accentuables qui marquent la fin d'un groupe accentuel ou intonatif lorsque la proéminence s'actualise sur leur syllabe périphérique droite (Lacheret-Dujour & Beaugendre 1999:45-46 ; Delais-Roussarie 1999:21, 2000:104-106).

Cette catégorisation « clitique /non clitique » dépend de l'annotation des `pos-dismo` alignée niveau lexical de la tire. Suivant les directives de Mertens (2011a:21), la proposition de catégorisation adoptée est résumée au Tableau C.15.

Phénomène	Catégorie grammaticale associée
Clitiques « »	0, ADV:comp, ADV:deg, CON:coo, CON:sub, DET:def, DET:dem, DET:ind, DET:int, DET:par, DET:pos, NUM:crd:det, NUM:crd:nom, NUM:ord:adj, PARA, PFX, PRO:per:obj, PRO:per:objd, PRO:per:obji, PRO:per:sjt, PRO:ref, PRO:rel, PRP, PRP:det, VER:cond:aux, VER:fut:aux, VER:impf:aux, VER:inf:aux, VER:ppas:aux, VER:ppre:aux, VER:pres:aux, VER:simp:aux VER:subp:aux
Non Clitiques « * »	ADJ, ADJ:num, ADV, ADV:int, ADV:neg, FRG, ITJ, NOM:acr, NOM:com, NOM:pro, PRO:dem, PRO:ind, PRO:int, PRO:per:ton, VER:cond, VER:cond:impf, VER:fut, VER:impe, VER:impf, VER:inf, VER:pers, VER:ppas, VER:ppre, VER:pres, VER:pres:entatif, VER:simp, VER:subi, VER:subp

Tableau C.15 – Distinction clitiques /vs/ non clitiques en fonction de la catégorie grammaticale.

11 Tire 11 : ga-accentuable

Nom de tire : ga-accentuable

Tire de dépendance en amont du traitement :
lex – IPU.file

Tire de dépendance en aval du traitement :
ga-accentué

Méthodologie d'annotation :
automatique (ProsoBox)

Référence bibliographique :
P. Mertens. 2011b. « Prosodie, syntaxe, discours: autour d'une approche prédictive ». *Actes d'IDP2009*. Sous la dir. de H. Yoo & E. Delais-Roussarie. Paris, 19–32

La chaîne verbale se découpe en groupes accentuels (ga – *Groupe accentuel*) (Avanzi, Bordal & Obin 2011 ; Jun & Fougeron 2002). Les ga se définissent comme des unités constituées d'un ou de plusieurs mots lexicaux et leur(s) clitique(s) associé(s). La nature clitique d'un élément dépend à la fois de sa position dans la construction syntaxique et de sa catégorie grammaticale. Sur la dernière syllabe lexicale d'un groupe, une proéminence peut se manifester en un accent final de groupe ou accent primaire (Lacheret-Dujour & Beaugendre 1999:45).

De manière schématique, la tire `ga-accentuable` segmente le flux de parole en unités incluant un mot plein et ses éventuels clitiques dépendants au sein d'une Unité Separée par des Pauses (tire `IPU.file`). Ces segments symbolisent des groupes rythmiques potentiellement porteurs d'un accent final.

12 Tire 12 : ga-accentué

Nom de tire : ga-accentué

Tire de dépendance en amont du traitement :
ga-accentuable – promauto – if

Tire de dépendance en aval du traitement :
aucune

Méthodologie d'annotation :
automatique (script Praat)

Référence bibliographique :
P. Mertens. 2011b. « Prosodie, syntaxe, discours: autour d'une approche prédictive ». *Actes d'IDP2009*. Sous la dir. de H. Yoo & E. Delais-Roussarie. Paris, 19–32

L'accent final démarcatif de fin de groupe frappe sur la dernière syllabe pleine d'un élément lexical (Mertens 2011b:21). Il est constitutif des groupes accentuels actualisés. A la différence des groupes accentuables, potentiellement accentués, les groupes accentuels actualisés sont obligatoirement marqués par cet accent final démarcatif. En ce, il s'apparente aux frontières intono-syntaxiques de Rossi (1999:207).

La segmentation en groupe accentuel actualisé dépend des deux annotations précédentes en ce qu'elles sont définies par la combinaison d'un mot plein porteur d'un accent final et de ses clitiques associés. En francais, l'accent final démarcatif de fin de groupe, constitutif des groupes accentuels, et par extension des groupes intonatifs, frappe sur la dernière syllabe pleine d'un élément lexical précédé éventuellement de syllabes atones (clitiques) (Mertens 2011b:21).

L'accent final démarcatif est ici défini par la congruence de deux annotations distinctes :

- **la tire `promauto`** : la présence d'une annotation de degré égale ou supérieure à 2 – le degré 1 représentant un niveau trop faible de précision de l'algorithme.
- **la tire `if`** : l'annotation « f » symbolisant la présence d'une syllabe finale de groupe accentuable.

Un exemple d'annotation est donné au Tableau C.18 : trois groupes accentuels accentuables fusionnent en un seul groupe accentuel accentué.

Accentuable : (Je reçois) (un coup) (de téléphone) (donc déjà) (rien que ça c'était) (un peu) (une surprise)
Accentué : (Je reçois un coup de téléphone) (donc déjà rien que ça c'était un peu une surprise)

Tableau C.18 – Comparaison des groupes accentuels accentuables et accentués entre parenthèses. (extrait de `rcv_1`).

13 Tire 13 : disfluency

Nom de tire : disfluency

Tire de dépendance en amont du traitement :
words

Tire de dépendance en aval du traitement :
disfluency-seq – diacritics

Méthodologie d'annotation :
semi-automatique (pause vide et pleine) et manuelle (autres phénomènes)

Référence bibliographique :
L. Crible, A. Dumont, I. Grosman & I. Notarrigo. 2016. *Annotation manual of fluency and disfluency markers in multilingual, multimodal, native and learner corpora. Version 2.0.* Technical Report. Université catholique de Louvain et Université de Namur.

La couche d'annotation manuelle `disfluency` est une couche lexicale alignée sur la segmentation lexicale de DisMo (`tok-min`). Cette couche d'annotation annote de manière componentielle les phénomènes de (dis)fluences. Cette annotation reflète une méthodologie basée purement sur la forme de surface, au niveau du mot. Elle invite ainsi l'analyste à envisager la fluence et/ou la disfluence « comme une combinaison de traits qui, pris isolément, peuvent être impliqués tantôt dans la fluence tantôt dans la disfluence d'un discours selon leur fréquence, leur fonction, leur position et leur combinaison » (Crible, Dumont, Grosman & Notarrigo 2016:5).

L'objectif de ce protocole d'annotation est de pouvoir prendre en charge les marqueurs de (dis)fluence en français, en anglais (L1 et L2) et en langue des signes de Belgique francophone (LSFB). Il est le fruit d'un travail collaboratif des recherches doctorales menées au sein de l'ARC-Fluency. Le protocole distingue les « fluencèmes simples » constitués d'une seule partie (marqueurs de discours, différents types de pauses, faux départ) et les « fluencèmes composés » qui s'actualisent en au moins deux parties (répétitions et substitutions). La troncation est intermédiaire car elle peut être abandonnée (simple) ou complétée (composée)[7].

Un systême de crochets délimite les frontières de chaque fluencème simple ou composé. La position du crochet est significative : « < » indiquant le début d'un phénomène tandis que « > » en indique la fin. La numérotation de certains éléments ne concerne que les différentes parties d'un fluencème composé, ainsi que les troncations répétées. Enfin, en présence de fluencèmes enchâssés, les fluencèmes simples sont placés en position initiale de telle sorte que les tags UP, FP, DM, ET et TR précèdent toujours les tags de répétition et de substitution. Le Tableau C.20 résume la liste des fluencèmes annotés dans les corpus. L'Annexe D reprend la dernière version du protocole de manière complète et détaillée.

7. Le terme « séquence » se réfère à tout segment de la parole contenant au moins un fluencème, simple ou composé, isolé, juxtaposé ou enchâssé.

Etiquettes	**Fluencèmes**
UP	Pause silencieuse
FP	Pause pleine
DM	Marqueur de discours
ET	Terme d'édition
FS	Faux départ
TR	Troncation
RI	Répétition à l'identique
RM	Répétition avec modification
SP	Substitution propositionnelle
SM	Substitution morphologique
IL	Insertion lexicale
IP	Insertion parenthétique

TABLEAU C.20 – Liste des fluencèmes annotés.

14 Tire 14 : diacritics

Nom de tire : diacritics

Tire de dépendance en amont du traitement :
disfluency

Tire de dépendance en aval du traitement :
aucune

Méthodologie d'annotation :
manuelle

Référence bibliographique :
L. Crible, A. Dumont, I. Grosman & I. Notarrigo. 2016. *Annotation manual of fluency and disfluency markers in multilingual, multimodal, native and learner corpora. Version 2.0.* Technical Report. Université catholique de Louvain et Université de Namur.

En complément à la tire `disfluency`, la tire `diacritique` décrit le fonctionnement des (dis)fluences en contexte (i.e. les modifications qu'elles peuvent apporter aux fluencèmes simples et composés). Cette annotation vient ainsi ajouter une information sur un autre élément de la chaîne linguistique. Tous ces « diacritiques » sont annotés sur une seule tire.

Allongement (LG). Cette catégorie couvre tout maintien d'un phonème en début, fin ou intérieur de mot, relativement à sa valeur attendue. Cette annotation est perceptive et, par conséquent, subjective. L'étiquette s'applique au mot affecté par l'allongement sans égard à la position de l'allongement dans le mot [8].

Enchâssement de fluencèmes (WI). Cette catégorie couvre tout fluencème simple encadré par un autre fluencème composé. Par exemple, une pause au milieu d'une répétition (entre le répétable et le répété ou entre deux répétés d'un même répétable) porte le tag <WI> (pour *within*) sur la tire réservée aux diacritiques. L'information ainsi conservée permet d'extraire automatiquement les deux types d'environnements (isolé et enchâssé) pour une éventuelle analyse contextuelle des marqueurs concernés. Les insertions lexicales et parenthétique ne recoivent jamais le trait <WI> étant donné qu'il est définitoire de cette catégorie.

Changement d'ordre (OR). Cette catégorie couvre tout élément répété à l'identique tant au niveau formel que sémantique, mais dans un nouvel ordre syntagmatique. Puisque les éléments répétés sont identiques, hormis le changement d'ordre, ces cas font toujours partie du premier type de répétition à l'identique (RI). Le tag s'applique à tous les éléments affectés par le changement d'ordre, et uniquement à ceux-ci, et ce dans les deux parties du fluencème (RI0 et RIn). Les crochets s'ouvrent sur le premier des termes dans le répétable et se ferment sur le dernier des termes ayant subi un changement d'ordre dans le répété.

8. Cette annotation de l'allongement au niveau du mot est complémentaire à l'annotation des allongements au niveau syllabique (notés « a » sur la tire `delivery`).

15 Tire 15 : disfluency-seq

Nom de tire : disfluencySeq

Tire de dépendance en amont du traitement :
disfluency – IPU.file

Tire de dépendance en aval du traitement :
aucune

Méthodologie d'annotation :
automatique (script Praat)

Référence bibliographique :
L. Crible, A. Dumont, I. Grosman & I. Notarrigo. 2016. *Annotation manual of fluency and disfluency markers in multilingual, multimodal, native and learner corpora. Version 2.0.* Technical Report. Université catholique de Louvain et Université de Namur.

La « chaîne de (dis)fluence(s) » se réfère à tout segment de la parole contenant au moins une (dis)fluence, simple ou composée, isolée, juxtaposée ou enchâssée (Crible, Dumont, Grosman & Notarrigo 2016:7).

La tire `disfluency-seq` comprend la segmentation de ces chaînes de (dis)fluences au sein de chaque tour de parole du locuteur (`IPU.file`). En conséquence, l'annotation exclut la concaténation de pauses interlocuteurs. Chaque (dis)fluence cooccurrente est séparée par le symbole « + ».

16 Tire 16 : rection

Nom de tire : rection

Tire de dépendance en amont du traitement :
words

Tire de dépendance en aval du traitement :
sequences – bdu-auto

Méthodologie d'annotation :
manuelle

Référence bibliographique :
N. Tanguy, T. Van Damme, D. Liesbeth & A. C. Simon. 2012. « Projet FRFC «Périphérie gauche des unités de discours» Protocole de codage syntaxique »

La tire `rection` segmente le texte en unités de rection (ur). Toute unité rectionnelle est organisée autour d'un noyau, c'est-à-dire un élément recteur. Il s'agit généralement d'un verbe tensé mais le noyau de l'unité peut également être un constituant averbal : nom, pronom, adverbe ou groupe prépositionnel prédicatif (Lefeuvre 1999). On distingue trois types d'unités de rection : (i) les unités de rection verbales (urv), (ii) les unités de rection averbales (ura) et (iii) les unités de rection elliptiques (ure). Ces trois types peuvent être précisés des mentions « inachevée » (ur$x$ -I) et « plus » (ur$x$ +), ce que résume le tableau ci-dessous.

Unité de rection	**Code**	**Unité inachevée**	**Unité « plus »**
Unité de rection verbale	urv	urv-I	urv+
Unité de rection averbale	ura	ura-I	ura+
Unité de rection elliptique	ure	ure-I	ure+

Tableau C.24 – Synthèse du codage des trois types d'unité de rection.

On annote également dans la tire `rection` les marqueurs de discours (md), les adjoints (a), certains silences (sil) et certains *euh* (euh), quand ceux-ci se trouvent en dehors de l'unité de rection. Tous les éléments codés dans cette tire sont notés en minuscule : urv, ura, ure, md, ur$x$ -I, sil, a, euh, etc. L'élément « -I » y apparaît néanmoins en majuscule.

Code	**Explications et remarques**
md	marqueur de discours
a (ag ou ad)	l'adjoint est intégré ou périphérique à l'UR à laquelle il se rattache
euh	pause pleine *euh*
sil	silence > 250 ms. si le silence est inférieur à 250 ms., il doit être rattaché à ce qui suit
#	non codable

Tableau C.25 – Synthèse du codage des éléments non rectionnels.

17 Tire 17 : séquences

Nom de tire : séquences

Tire de dépendance en amont du traitement :
rection

Tire de dépendance en aval du traitement :
bdu-auto

Méthodologie d'annotation :
manuelle

Référence bibliographique :
N. Tanguy, T. Van Damme, D. Liesbeth & A. C. Simon. 2012. « Projet FRFC «Périphérie gauche des unités de discours» Protocole de codage syntaxique »

La tire `sequence` segmente les unités rectionnelles en séquences de type fonctionnel ou catégoriel, selon le type d'unité rectionnelle (verbale, averbale ou elliptique). Ce découpage s'inspire de Bilger & Campione (2002) : (i) la séquence sujet (SS), (ii) la séquence verbale (SV), (iii) la séquence objet (SO) et (iv) la séquence régie (SR). Une séquence peut être interrompue par une autre séquence (insert, SR, associé, etc.) avant de se poursuivre. La suite de la séquence interrompue est codée par « -S » : SO-S, SV-S, SR-S, etc.

Code	Séquence fonctionnelle	Exemple
SS	Sujet apparaissant sous une forme lexicale.	(1) [(monsieur Barnier)$_{SS}$ (nous annonce)$_{SV}$ (une commission d'enquête carrément sur le prix de la salade)$_{SO}$]^{urv}
SO	Objet réalisé par une forme lexicale pleine. Inclus les compléments directs, les compléments indirects, dont les locatifs, les compléments d'agent, les attributs du sujet nominaux déterminés (SN définis et indéfinis) et les séquences de constructions impersonnelles.	(2)[(mes élèves)$_{SS}$ (disent)$_{SV}$ (« sauf »)$_{SO}$]^{urv} (3)[(si nous avons la responsabilité du pays)$_{SRg}$ (nous donnerons)$_{SV}$ (des papiers)$_{SO}$ (à tous ceux qui n'en ont pas)$_{SO}$]^{urv} (4)[(on est)$_{SV}$ (dans un salon de coiffure)$_{SO}$]^{urv}

Code	Séquence fonctionnelle	Exemple
SV	Séquence verbale se composant du verbe et de ses différents actants clitiques et/ou pronominalisés. Elle peut également comporter les marques verbales de négation et de restriction, les modaux, les infinitifs sélectionnés par le verbe, certaines formes liées au verbe, figées ou non, les attributs adjectivaux et les attributs nominaux non déterminés.	(5) <mais>$_{md}$ [(je le crois)$_{SV}$]^{urv} (6) [(l'un)$_{SS}$ (n'empêche pas)$_{SV}$ (l'autre)$_{SO}$]^{urv} (7) <donc>$_{md}$ [(le gouvernement)$_{SS}$ (peut agir)$_{SV}$]^{urv} (8) [(on peut ajouter)$_{SV}$ (une détection automatique des syllabes proéminentes)$_{SO}$]^{urv}
SR	Séquence régie relevant de la rection verbale mais n'étant pas un élément appelé syntaxiquement par le verbe (appartenant à la valence verbale). Tout comme les séquences associées, selon leur place (par rapport à l'unité rectionnelle ou le verbe), les séquences régies sont précisées dans la tire `sequence` par des mentions « g » et « d », en minuscules. De la même manière que la distinction élément associé /vs/ élément régi est parfois délicate, la séparation entre séquence régie et séquence objet à valeur locative n'est pas toujours évidente.	(9) [(dans le voyage au bout de la nuit)$_{SRg}$ (il commence)$_{SV}$ (une ligne)$_{SO}$ (en disant)$_{SRd}$]^{urv} (10) <finalement>$_{ag}$ [(quand j'ai eu enfin déchiffré tout le truc)$_{SRg}$ (je lui ai dit)$_{SO}$]^{urv} (11) <mais>$_{md}$ [<est-ce que>$_{mi}$ (vous lisez)$_{SV}$ (aussi mal que vous écrivez)$_{SRd}$]^{urv} <Monsieur Dantzig>$_{ad}$

TABLEAU C.27 – Résumé des types de séquences fonctionnelles.

Les unités rectionnelles averbales sont précisées en termes de catégories et non de fonctions : SN, SPron, SAdj, etc. Les unités rectionnelles elliptiques, selon les cas, sont précisées en séquences fonctionnelles ou catégorielles. Enfin, nous doublons dans cette tire les marqueurs de discours (md), les adjoints (a), certains silences (sil), certains *euh* et les éléments paraverbaux (#).

Par ailleurs, les séquences régies (SR) et les adjoints (a), selon leur place par rapport à l'unité de rection ou le verbe auxquels ils se rapportent, sont précisés par les mentions « g » et « d » en minuscules (gauche et droit respectivement). Tous les autres éléments codés dans la tire `sequences` y sont notés en majuscules (sauf l'insertion, les marqueurs de discours et les adjoints) : SS, SV, SO, SRg, SRd, ag, ad, insert, etc.

Les séquences de type catégoriel précisent les unités rectionnelles averbales et certaines unités rectionnelles elliptiques si celles-ci ne peuvent être « fonctionnalisées ».

Code	**Séquence catégorielle**	**Exemple**
SN	nominale	[(plus six euros)$_{\text{SN}}$ (sur les retraites)$_{\text{SRd}}$]$^{\text{ura}}$
SAdj	adjectivale	[(pas évident)$_{\text{SAdj}}$ (dans notre societe une société ou on se demande mais Dieu est-ce qu'Il existe)$_{\text{SRd}}$]$^{\text{ura}}$
SPron	pronominale	<en tout cas> $_{\text{md}}$ [(moi)$_{\text{SPron}}$ (parce que j'avais choisi de vivre dans une famille d'accueil)$_{\text{SRd}}$]$^{\text{ure}}$ [(comment)$_{\text{SAdv}}$ (ca)$_{\text{SPron}}$]$^{\text{ura}}$ [(ça n'a)$_{\text{SV}}$ (aucune importance)$_{\text{SO}}$]$^{\text{urv}}$
SAdv	adverbiale	[(oui)$_{\text{SAdv}}$]$^{\text{ura}}$ [(allez-y)$_{\text{SV}}$]$^{\text{urv}}$ [(non {non, o})$_{\text{SAdv}}$]$^{\text{ura}}$ [(allez-y)$_{\text{SV}}$]$^{\text{urv}}$
Sprep	prépositionnelle	[(pour devenir quelqu'un)$_{\text{Sprep}}$]$^{\text{ure}}$ [(pour avoir un papier)$_{\text{Sprep}}$]$^{\text{ure}}$
SInf	infinitive	[(à suivre)$_{\text{SInf}}$]ure [(absolument)$_{\text{SAdv}}$]$^{\text{ura}}$
Ssub	subordonnée	[(bien sur)$_{\text{SAdj}}$ (que nous ne serions pas socialistes si nous n'avions pas de l'optimisme de l'espérance)$_{\text{Ssub}}$]$^{\text{ura}}$
SInt	interjective	[(il klaxonne)$_{\text{SV}}$]$^{\text{urv}}$ [(tut)$_{\text{SInt}}$]$^{\text{ura}}$
SPart	participale	[(photo de Chavez)$_{\text{SN}}$]$^{\text{ura}}$ [(trapu)$_{\text{SAdj}}$]$^{\text{ura}}$ [(les épaules ramassées)$_{\text{SPart}}$]$^{\text{ura}}$

Tableau C.28 – Résumé des séquences de type catégoriel.

18 Tire 18 : boundary

Nom de tire : boundary

Tire de dépendance en amont du traitement :
syll

Tire de dépendance en aval du traitement :
contours – bdu-auto

Méthodologie d'annotation :
validation perceptive d'une pré-annotation automatique

Référence bibliographique :
P. Mertens & A. C. Simon. 2013. « Towards automatic detection of prosodic boundaries in spoken French ». *Actes d'IDP 2013.* Louvain, Belgique, 81–87

Dans cette tire on annote les frontières prosodiques perçues, avec deux niveaux de frontière : frontière majeure (///) ou intermédiaire (//). Une frontière est toujours alignée sur une fin de syllabe et sur une fin de mot.

La tire a été annotée sur une base perceptive uniquement en suivant le protocole suivant :

- Ecouter des fragments d'environ 5 à 8 secondes ;
- Ne pas écouter un extrait plus de 3 fois (sous peine de sur-coder) ;
- Commencer par noter les frontières majeures « /// », puis les intérmédiaires « // » ;
- Puis noter les autres phénomènes (*hesi, foc, ai*).

Par la suite, Mertens & Simon (2013) ont tenté d'entraîner un modèle d'annotation automatique de ces frontières à partir de corrélats acoustiques. La seconde partie de l'annotation a consisté à visualiser la pré-annotation automatique et à la valider perceptivement.

Le Tableau suivant donne quelques indications des valeurs prosodiques associées aux frontières perçues par les experts :

Code	**Type de phéno-mène**	**Actualisation prosodique – lexicale**
///	Frontière majeure	changement de locuteur pause silencieuse 200 ms. non précedée d'une marque d'hésitation, ou durée relative, $>$ 3 ou montée mélodique $>$ 10 ST
//	Frontière intermé-diaire	durée relative $>$ 2, ou f0 rise (intrasyll) $>$ 4, ou f0 moyenne $>$ 5 ST,
hesi	Hésitation	*euh* ou allongement vocalique d'hésitation (parfois proche d'une frontière avec contour de suspens)
ai	Accent initial	accent initial (syllabe non finale) non allongé et haut
foc	Accent de focalisation	accent de focalisation sans effet de frontière, sur un mot
#	Backchannel / paraverbal	signal de l'interlocuteur au locuteur en cours, pour lui manifester qu'il peut continuer à parler

TABLEAU C.30 – Résumé de l'annotation manuelle/auditive des frontières.

19 Tire 19 : contours

Nom de tire : contours

Tire de dépendance en amont du traitement :
syll – boundary

Tire de dépendance en aval du traitement :
bdu-auto

Méthodologie d'annotation :
automatique – validée manuellement

Référence bibliographique :
P. Mertens & A. C. Simon. 2013. « Towards automatic detection of prosodic boundaries in spoken French ». *Actes d'IDP 2013.* Louvain, Belgique, 81–87

À chaque frontière prosodique perçue, on a associé un contour intonatif perçu parmi les 4 suivants : Continuatif, Terminal, Suspens ou Focalisation (suivi ou non d'un Appendice, suite de syllabes désaccentuées). Les annotateurs avaient reçu la consigne de noter le mouvement mélodique perçu en faisant au maximum abstraction de la syntaxe (par ex. une forme syntaxique interrogative peut être réalisée avec un contour descendant, qui est alors annoté comme Terminal).

Code	Type de contours	Actualisation prosodique
C	Continuatif	Montée intrasyllabique ou un intervalle positif par rapport à la syllabe qui précède. Peut correspondre à un contour montant interrogatif qui signale une fin de phrase interrogative
S	Suspens	Contour plat, parfois allongé ; pas d'écart mélodique par rapport à la syllabe qui précède
T	Terminal	Contour descendant
F	Focalisation	Mouvement descendant partant d'une cible haute, descente rapide et raide
App	Appendice	Suite de syllabes atones et plates, débit rapide qui suit un T ou un F

Tableau C.33 – Codage simplifié des contours intonatifs.

20 Tire 20 : bdu-auto

Nom : bdu-auto

Tires de dépendance en amont du traitement :

bdu-auto (rection – boundary – contours)

Tire de dépendance en aval du traitement :

aucune

Méthodologie d'annotation :

automatique – validee manuellement

Reference bibliographique :

L. Degand & A. C. Simon. 2009. « On identifying basic discourse units in speech: theoretical and empirical issues ». *Discours* (4), [en ligne]

La mise en correspondance de ces deux segmentations est réalisée comme suit : chaque fois qu'une frontière d'unité de rection coïncide avec une frontière prosodique majeure, une BDU est identifiée. Cela donne lieu à cinq types de BDU distinctes, représentés en Figure C.3 :

- les BDU-c – congruentes : une unité de rection correspond à une unité prosodique majeure
- les BDU-r – régulatives : elles sont constituées d'un élément syntaxiquement non régi (connecteur, adverbe de phrase ou marqueur discursif) et isolé par une frontière prosodique
- les BDU-i – groupées par l'intonation : une unité prosodique regroupe plusieurs unités de rection
- les BDU-s groupées par la syntaxe : une unité de rection est segmentée en plusieurs unités prosodiques majeures
- les BDU-x mixtes : BDU regroupant plusieurs unités syntaxiques et plusieurs unités prosodiques majeures (en attente de complétion Selting (2000)).

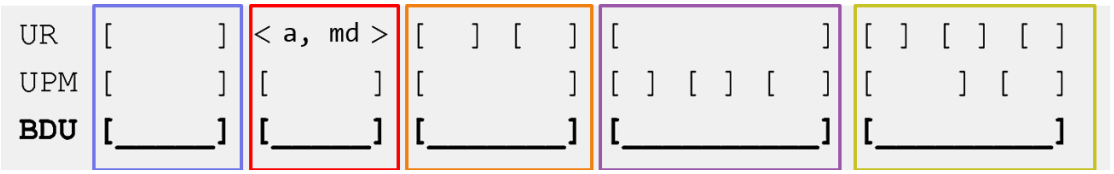


Figure C.3 – Segmentation en unités de bases du discours (BDU) issue de la segmentation en unités de rection (UR) et unités prosodiques majeures (UPM).

21 Tire 21 : speaker

Nom de tire : speaker

Tire de dépendance en amont du traitement :
transcription (absente des TextGrids)

Tire de dépendance en aval du traitement :
IPU.file – IPU.genre

Méthodologie d'annotation :
manuelle

Référence bibliographique :
J. Goldman, A. Auchlin & A. C. Simon. 2013. « Les variables temporelles dans le dialogue ». *Actes d'IDP 2013*. Louvain, Belgique, 39–44

L'objectif de cette annotation est double : (1) l'indication des chevauchements de parole est nécessaire à l'exclusion de ces segments du discours de l'analyse prosodique, (2) pouvoir distinguer l'appartenance des pauses par locuteur et exclure celles qui ne sont que le fruit d'un réel silence (*gap*) et arrêt de parole pour les deux locuteurs (pause intralocuteurs). Le calcul des durées moyennes des pauses intralocuteurs permises par la tire `speaker` sert directement à la définition des Unités Séparées par des Pauses (IPU – *Inter Pausal Units*). L'annotation de la tire `speaker` est effectuée manuellement. Elle s'inspire des travaux Edlund, Hirschberg & Heldner (2009) et de Goldman, Auchlin & Simon (2013).

En raison de la structure des échantillons dans LOCAS-F (un fichier par locuteur et non un fichier par enregistrement) ainsi que des prérequis d'annotation exigés pour l'analyse prosodique automatique des enregistrements via Prosogram, la méthodologie d'annotation a subi des adaptations mineures. Dans le cas d'un dialogue, la tire `speaker` sera identique dans les deux fichiers de locuteurs et les tires d'origine « LOC1 », « LOC2 » et « *gap* » sont fusionnées pour former la tire `speaker`.

Dans le cas de chevauchement, la tire `speaker` reprend le code des locuteurs concernés afin de ne pas perdre l'information des locuteurs concernés par le chevauchement (surtout dans le cas de multilogues, présents dans LOCAS-F). L'annotation permet ainsi d'envisager qui a « forcé » la prise du tour de parole *vs.* qui a été interrompu mais aussi qui a pu garder ce tour. Le fonctionnement de l'annotation des chevauchements suit ces deux règles : (1) le locuteur qui interrompt est toujours le premier annoté dans les segments de chevauchement L1+L2/L2+L1 – l'interrompu est le second.) ; dans de rares cas où les deux locuteurs entament leur prise de parole exactement ensemble, l'annotation de ces deux-ci est séparée par une barre oblique « / » (L1/L2) spécifiant ainsi que l'ordre de chevauchement n'est pas pris en compte.

Loc 1	pv1					pv1		pv1
Loc 2			pv2		pv2	pv2		
occupation		gap				overlap	gap	
speaker	L1	gap	L2	L2	L2	L2+L1	gap	L1

Tableau C.36 – Principe d'annotation de la tire `speaker` – adaptation de Goldman, Auchlin & Simon (2013).

22 Tire 23 & 24 : IPU-file & IPU-genre

Nom de tire : IPU-file, IPU-genre

Tire de dépendance en amont du traitement :
delivery – speaker

Tire de dépendance en aval du traitement :
disfluency-seq

Méthodologie d'annotation :
automatique

Référence bibliographique :
J. Goldman, T. François, S. Roekhaut & A. C. Simon. 2010. « Étude statistique de la durée pausale dans différents styles de parole ». *Actes des XXVIIIeme Journée d'Etudes sur la Parole (JEP'2010)*. Mons, Belgique : ISCA-AFCP, 161–164

La tire `IPU-file` segmente automatiquement le flux discursif d'un locuteur (marqués « L1 » dans la tire `speaker`) en Unités Séparées par des Pauses. L'IPU peut ainsi se définir comme une vocalisation unilatérale bornée par des pauses silencieuses au sein d'une bribe de parole émise par un locuteur (*talk-spurt*) (Jaffe & Feldstein 1970 ; Wlodarczak & Wagner 2013:1434). Elle est précédée et suivie par la parole d'un autre locuteur et perçue par celui-ci comme une bribe de parole (Norwine & Murphy 1938:282).

La définition de l'unité séparée par des pauses dépend de la détection et la catégorisation même de ce qui est considéré comme une pause. En vertu de Campione & Véronis (2002), le choix a été fait de ne pas fixer de seuil inférieur ou supérieur à la détection des pauses. En considérant l'importance de la variation intergenres et interlocuteurs dans la distribution pausale (Goldman, Avanzi & Auchlin 2010), le choix a été fait de ne pas établir un seuil de pause absolu lors de la segmentation en IPU.

La méthodologie en clustering (Fraley & Raftery 2007) proposée par Campione & Véronis (2002) et Goldman, Avanzi & Auchlin (2010) a amené à définir le nombre de type de pauses (courtes, moyennes, longues) et leurs seuils minimal et maximal, pour chaque locuteur (IPU-file) et chaque situation de parole.

En fonction des résultats de l'analyse en clustering deux méthodes de définition de seuils de pauses utiles aux IPU ont été distingués :

- Dans le cas d'une distribution pausale (unimodale) lognormale, il s'agit de la durée moyenne de la pause augmentée d'un écart type.
- Dans le cas d'une distribution bi ou trimodale, il s'agit du seuil maximum des pauses courtes.

Ces seuils adaptatifs (en fonction du locuteur dans sa situation de parole ou du type de situation de parole) permettent ainsi de segmenter le flux discursif en unités de tailles moyennes en écartant les pauses brèves (en général inférieures à 200 ms.). Cette procédure est motivée linguistiquement par le fait que les pauses brèves ne porteraient jamais de fonction démarcative (Barik 1968:156 ; Campione & Véronis 2004:2).

23 Tire 25 : comment

Nom de tire : comment

Tire de dépendance en amont du traitement :
toutes

Tire de dépendance en aval du traitement :
aucune

Méthodologie d'annotation :
(manuelle)

Référence bibliographique :
(aucune)

La tire `comment` regroupe des informations relatives au traitement des autres tires des corpus. On y retrouve en particulier :

- au niveau lexical, des hésitations de transcription notées entre crochets extraites de la tire `words` et préservées en vue de potentiels travaux sur la transcription de données ;
- au niveau supra-lexical des hésitations ou commentaires sur l'annotation des disfluences.

Annexe D

Protocole ARC d'annotation des (dis)fluences (EN)

Annotation manual of fluency and disfluency markers in multilingual, multimodal, native and learner corpora

Version 2.0

Crible L., Dumont A., Grosman I., Notarrigo I.

9th Feb. 2016

To cite this paper: Crible L., Dumont A., Grosman I., Notarrigo I. 2016. Annotation manual of fluency and disfluency markers in multilingual, multimodal, native and learner corpora. Version 2.0. *Technical report*. Université catholique de Louvain and Université de Namur.

This research is the outcome of the collaboration between the
Université catholique de Louvain and the Université de Namur

Fluency and disfluency markers: A multimodal contrastive perspective

"Action de Recherche Concertée" funded by the Fédération Wallonie-Bruxelles
(Grant number 12/17-044)

Contents

1 Introduction 4
1.1 Research context 4
1.2 Purpose of the manual 4
2 Overview of the protocol : design and applications 5
2.1 Definitions and terminology 5
2.2 Technical aspects of the annotation system 6
3 Categories of fluencemes covered by the annotation protocol 9
3.1 Simple fluencemes 9
3.1.1 Pauses (UP, FP, S1, S2, S3) 9
3.1.2 Palm-up in SL (PU) 11
3.1.3 Discourse marker (DM) 12
3.1.4 Explicit editing term (ET) 14
3.1.5 False start (FS) 15
3.1.6 Truncation (TR) 16
3.2 Compound fluencemes 17
3.2.1 Repetitions (RI, RM, RE, RG) 18
3.2.2 Substitutions (SM, SP) 22
3.3 Insertions 23
3.3.1 Lexical insertion (IL) 23
3.3.2 Parenthetical insertion (IP) 25
3.3.3 Deletion (DE) 26
3.4 Diacritic signs 26
3.4.1 Misarticulation (AR) 26
3.4.2 Lenghtening (LG) 26
3.4.3 Embedding of simple fluenceme (WI) 27
3.4.4 Re-ordering (OR) 28
3.4.5 Syntactic completion - functional and total (CF et CT) . 28
4 Queries 29
5 Conclusion 29

1 Introduction

1.1 Research context

Research on fluency and disfluency phenomena is flourishing, multi-disciplinary and adopts different approaches depending on the scientific goals. This situation has led to a panel of annotation protocols available in the literature which are nevertheless rarely comparable or generalisable to different data types. Since Shriberg (1994) and her seminal typology of markers, several authors have tried to adapt her coding scheme and terminology, with a view to allowing automatic annotation or to modifying her categories. Major inspiring studies in our research are: Shriberg (1994), Meteer (1995), Candéa (2000), Eklund (2004), Dister (2007) and Götz (2013). Other ongoing projects following similar approaches include Christodoulides, Avanzi, and Goldman (2014), Moniz et al. (2014) and Pallaud, Rauzy, and Blache (2013). We will refer to more specific studies focusing on particular phenomena in the following sections when relevant.

The interest for componential approaches to fluency as proposed by Götz (2013) tends to spread. However, the specific elements participating in this vision of fluency such as repetition, substitution, pause etc., as well as annotation formats often differ on the scope of covered phenomena. Concretely, differences between typologies pertain to the number and types of phenomena, language- and modality-related specificities, technical aspects of the annotation and quantitative treatment (labelling, data extraction), etc. Generally, most protocols show a number of pitfalls, either on practical aspects such as the replicability of the annotation system and its efficiency in quantitative analyses, or on theoretical considerations such as the validity of the categories and discriminating criteria and the overall cognitive-pragmatic relevance of the whole model.

Against this backdrop, the present protocol aims at addressing a number of these issues, by taking into account the contributions of previous research but also a wealth of theoretical frameworks and data types involved in our concerted research action. The originality of the present approach is to try to overcome the specificities of a particular framework by offering an exhaustive and flexible model, which can potentially adapt to many research questions.

The following sections will first present our theoretical and methodological framework, before turning to operational definitions of all phenomena covered by the annotation.

1.2 Purpose of the manual

The purpose of this document is to provide the annotators with a coding scheme for (dis)fluency markers in line with the componential approach to (dis)fluency as defined by Götz (2013) and flexible enough to account for such markers in French, English (L1 and L2) and French Belgian Sign Language (LSFB). This protocol has been developed in

the framework of a collaboration between four PhD students working with different data types in terms of language (French, English, LSFB), modality (speech vs. sign language), speakers (native speakers vs. learners), and, more generally, in their analytical approach (semi-automatic vs. manual, semasiological vs. onomasiological). This plurality of approaches not only vouches for the flexibility of the protocol, but also provides the foundation for a number of theoretical and practical decisions at the core of this work. By adopting this protocol and by using the same labels for the same category across different types of corpora, the different teams will ensure the comparability and inter-operability of their research. We would like to argue that this linguistic, modal and theoretical adaptability strengthens the robustness and relevance of the annotated categories and of the protocol as a whole.

In order to ensure replicability, (i) each phenomenon covered by the coding scheme is systematically defined, (ii) discriminating and/or application criteria are explained, (iii) labels and other technical aspects are considered, (iv) and illustrations with authentic corpus examples are provided.

The definitions and criteria presented here emerged from the comparison of existing models and their application to authentic corpus data. This corpus-based method vouches for the operationality of the protocol by ensuring the applicability of the criteria and minimizing inter-annotator disagreement. It remains that some categories may be more prone to individual interpretation than others, given the inevitable – but limited – semantic-pragmatic considerations involved in any approach to (dis)fluency. Nevertheless, the majority of the phenomena and criteria selected here are related to formal features or surface cues that considerably constrain the analysis, as will be shown for the different categories in section 3.

2 Overview of the protocol : design and applications

2.1 Definitions and terminology

(Dis)fluency can be defined in a holistic perspective as the smooth, fast and effortless use of language (Crystal 1988). In this view, the fluent or disfluent interpretation results from the hearer's perception and the overall impression of the interaction. The componential approach, on the other hand, views fluency and disfluency as a combination of features that, taken in isolation, can either be involved in fluent or disfluent speech depending on their frequency, function, position and combination patterns. In other words, these markers can be considered either as felicitous and deliberate cues to support production and comprehension, or as signals of online planning and production difficulties. These difficulties stem from lexical access trouble ("tip-of-the-tongue" phenomenon), semantic and/or syntactic complexity, cognitive or emotional load, stalling for planning time, or more directly the perception of a "mistake" by the speaker leading to its reformulation. In contrast, positive effects of (dis)fluency markers include anchoring an utterance in

discourse and the interaction, informative salience, or stylistic effect. In general, they can be understood as "relevant" in terms of an effort-effect ratio (Sperber and Wilson 1995).

Furthermore, our definition of (dis)fluency is fundamentally situational, i.e. fluency is only evaluated with respect to expectations in particular interaction settings. Thus, the recurrence of the same phenomenon (e.g. frequent silent pauses) can either create a fluent effect in a formal situation such as a political speech, or contribute to an impression of disfluency in a more informal setting. This ambivalence led us to use the term "(dis)fluency" in this manual and other related publications, in order to avoid any interpretative bias.

Apart from this general view of fluency and disfluency, other key notions referring to the different structures observed need to be briefly defined as well, namely fluenceme (simple and compound), phrase and sequence.

Following Götz (2013), we use the term "fluenceme" to refer to any (dis)fluency marker, with no a priori evaluation of its fluent or disfluent character. In the coding scheme, two groups of fluencemes are distinguished according to their internal structure: "simple" fluencemes consist of one part only (discourse markers, different types of pauses, false starts, editing terms); "compound" fluencemes show a two-part structure (repetitions and substitutions). Truncation is either simple if it is left incomplete, or compound when it is completed.

The term "phrase" is used to refer to lexicalized units consisting of several graphic words or several signs. It can for instance be used for some discourse markers (e.g. "tu vois", "in other words"). The term "sequence" refers to any segment of speech containing at least one fluenceme, simple or compound, either isolated, juxtaposed or embedded.

2.2 Technical aspects of the annotation system

Corpus annotation is mainly manual, sometimes combined with automatic annotation, and is not restrictive in technical terms so that different data configurations can be accounted for. Transcriptions need to be segmented at word or sign level and to be aligned with the sound (or video), in a format that is compatible with a multi-layered annotation software (EXMARaLDA (Schmidt and Wörner 2012), ELAN (Sloetjes and Wittenburg 2008), PRAAT (Boersma and Weenink 2014), etc.).

This protocol organizes the annotation in two layers (or "tiers"). Simple and compound fluencemes are both annotated on the first tier, and their annotation can be completed when necessary by "diacritics"[1] on the second tier. When a word belongs

[1]By "diacritics" we mean elements that are hierarchically dependent on the main annotation tier. In other words, they form a secondary annotation level giving additional information to previously annotated fluencemes (cf. section 3.4)

to several structures, it will receive a double tag, still on the first tier. We argue that the gain in information and efficiency that results from a unique annotation tier is a substantial asset for contextual analysis of fluencemes. However, thanks to the labelling system, it still remains possible to query each individual fluenceme or to extract them on separate tiers if necessary.

Each word or sign within a fluenceme is annotated. All elements included in a fluenceme are tagged with a label containing brackets, two-letter initials and sometimes numbers (cf. section 2.2). For instance, the phrase "you see" receives a tag at each graphic unit (i.e. one for "you" and one for "see"), even though it constitutes only one discourse marker (frequency counts are not altered thanks to the bracketing system). However, within a sequence of fluencemes, only the elements defined in the present protocol are annotated, i.e. fluencemes and related phenomena defined in the protocol such as certain parenthetical asides or inserted words; the annotation thus excludes any other element located in a sequence which is not explicitly defined in this protocol.

The annotation is mainly linear, but back-and-forth movements to the left and right contexts are not excluded if they are required for the interpretation. Moreover, when a sequence includes several compound fluencemes, the annotation focuses on a global interpretation of the phenomena: although every fluenceme is individually accounted for, some compound structures can be identified as supporting the whole sequence if they provide a coherent and representative interpretation of the segment. Concretely, in the following example, the higher-order fluenceme is the partial repetition of "it's a long process" repeated twice with internal modulations:

Backbone : Bb_en009 [2]

it's	a	long	process	(345)
<RM0	**RM0**	**RM0**	**RM0**<SP0	<UP>
				<WI>

it	's	a	long	haul
RM1	**RM1**	**RM1**	**RM1**<SP0	SP1>

and	it	's	a	very	expensive	process
<DM>	**RM2**	**RM2**	**RM2**	<IL>	SP1>	**RM2>**
<WI>						

[2] Examples are authentic and their reference corresponds to the name of the text in its corpus. Corpora used for the examples in the present document are: the French component of LINDSEI (Gilquin, De Cock, and Granger 2010), BACKBONE (Kurt 2012), LSFB (Meurant 2015), C-Humour (Grosman in press).

A bracketing system is used to delineate the boundaries of a (simple or compound) fluenceme. In other words, brackets are applied to all types of fluencemes, even those containing only one element (e.g. a pause or a false start). The position of the brackets is meaningful: open brackets "<" indicate the beginning of the phenomenon marked to their right (e.g. <RM0), whereas closing brackets " >" show the end of the phenomenon marked to their left (e.g. RM2>).

Labels	**Fluencemes**
	Fluenceme tier
UP	Unfilled pause
FP	Filled pause
S1/S2/S3	Manual stop
PU	Palm Up (LSFB)
DM	Discourse marker
ET	Explicit editing term
FS	False start
TR	Truncation
RI	Identical repetition
RM	Modified repetition
RE	Framing repetition
RG	Grammatical repetition (LSFB)
SP	Propositional substitution
SM	Morphosyntactic substitution
IL	Lexical insertion
IP	Parenthetical insertion
DE	Deletion
	Diacritics tier
AR	Misarticulation
LG	Lengthening
WI	Within
OR	Re-ordering
CF	Functional completion
CT	Total completion

Table 1 : Overview of labels in the protocol

A numbering system is used for compound fluencemes, namely repetitions, substitutions and repeated truncations. In case of embedded fluencemes, simple fluencemes are always tagged before others: the labels UP, FP, DM, ET and TR always precede those of repetition and substitution. Diacritics (OR, AR, LG, CT, CF, WI) are likewise also annotated on a separate tier. Examples in the remainder of this document will thus contain two tiers in addition to the transcription tier.

The annotation procedure presented here is neither constraining nor restrictive: not all phenomena need to be annotated if they do not contribute to a particular research question, and additional labels for specific modalities can be appended (e.g. pauses, cf. 3.1.1 or repetitions in LSFB, cf. 3.2.1).

3 Categories of fluencemes covered by the annotation protocol

This section defines and illustrates every annotation label. It summarizes similarities and differences with the phenomena as defined in the literature and specifies where the present approach stands. When no reference is mentioned, the phenomenon has, to the authors' knowledge, not been documented before.

3.1 Simple fluencemes

As a reminder, simple fluencemes are markers consisting of only one element. These phenomena can occur in isolation (e.g. a pause used alone), juxtaposed to other markers (e.g. a pause directly followed by a discourse marker) or embedded in compound fluencemes (e.g. a pause within a repetition).

3.1.1 Pauses (UP, FP, S1, S2, S3)

Despite the multimodal perspective of this protocol, pauses in spoken languages and sign languages (henceforth SL) are distinguished in order to avoid any premature mapping between different concepts. More specifically, although hand stops do occur in LSFB between signs by relaxing or fixing hands, the non-manual channel (i.e. eyes, gaze, eyebrows, mouth and head movements) keep on communicating information. These hand stops are therefore filled by the non-manual channel (Notarrigo 2016; Meurant and Notarrigo 2014). Moreover, contrarily to spoken languages where the sound vs. silence distinction is clear, in SL, hands do not "disappear" and the articulators are still visible: they have a specific location and shape in space. It is therefore premature, even perhaps not relevant, to refer to "empty" or "filled" pauses in SL. The analysis will in time shed more light unto the behavior of pauses in LSFB, thus opening the perspective of multimodal comparisons.

Unfilled pause (UP) This category covers any interruption of the sound signal above a certain threshold fixed by the annotation conventions already existing in the corpora[3] or fixed by taking into account the speaking rate in a given text[4]. No distinction is made *a priori* between different types of pauses based on their duration (Little et al. 2013).

[3]ICE and VALIBEL have different annotation conventions regarding pauses (variable thresholds from 100 to 250 ms).

[4]Either using an optional function of the annotation software DisMo (Christodoulides, Avanzi, and Goldman 2014) or with a system similar to Little et al. (2013).

LINDSEI : FR001-F

an	institution	(330)	really	er	necessary	is
		<**UP**>		<FP>		

Backbone : Bb_fr013

mon	ami	Georget	(315)[5]	monsieur	Georget	Daumas
<SP0	SP0	<RM0	<**UP**>	SP1>	RM1>	
			<**WI**>			

The protocol does not distinguish, in the first tier, pauses (or any other simple fluenceme) located within a compound fluenceme from pauses that are not embedded. However, it is possible to specify this information in the diacritics tier, in order to identify these cases afterwards. In those cases, the "WI" (for "within") tag is used (cf. section 3.4.3).

Filled pause (FP) This category covers active and conventional vocalizations traditionally seen as supporting or maintaining on-going speech (Duez 2001; Strassel 2003; Vasilescu, Candea, and Adda-decker 2004).

These vocalizations maintain the audio signal in a neutral position (e.g. *ah, eh, er, uh, um, euh, mh*). In case of hesitation between a filled pause and a lengthening, the latter is used for any post-vocal schwa that is not preceded by a silent pause.

The filled pause category can be transcribed in several forms (*erm*, *er*, *eh*, etc.). However, vocalizations such as *mh*, *mm*, *mhm* are only annotated when they constitute hesitation markers, not backchannelling devices.

Backbone :Bb_fr003

football	en	France	en	fait	euh	s'appuie
			<DM	DM>	<**FP**>	

LINDSEI : FR005-F

er	it	's	eh	it	's	our	our	project
<FP>	<RI0	RI0	<**FP**>	RI1	RI1>	<RI0	RI1>	
			<WI>					

[5]In the examples, duration (in milliseconds) of pauses is annotated in the transcription tier. Examples from different corpora will thus share the same format.

Hand stop in SL (S1/S2/S3) The three categories used to describe pauses in LSFB cover: hand stop during a sign (S1), hand stop between signs (S2), hesitation or lexical search moves (S3). The first group includes hands fixation in the starting position (<S1:ST>) or ending position (<S1:EN>) of a sign. The second group covers the moments when the signer is not signing (S2). Hands configuration is not significant in those cases: crossed (<S2:CR>), alongside the body (<S2:BO>) or relaxed in the neutral space in front of the signer (<S2:NE>). The last group (<S3>) includes the conventional sign corresponding to French *euh* as well as more idiosyncratic configurations. Because of overlap with other labels, S1 are annotated on a separate tier distinct from that of other fluencemes and other types of hand stops (see Notarrigo (2016) for more details).

LSFB[6] : CLSFBE CAM1002
« yes it's better to change for instance (320) *there are too many spelling signs for example the sign for USB it's better to use the sign for key »*

AVENIR	MEILLEUR	CHANGER	PAR-EXEMPLE	(320)	TROP
			<RM0	**<S2:CR>**	<SP0
				<WI>	

EPELLATION	PAR-EXEMPLE	USB	MIEUX	CLEF	CLEF
<S1:EN>		**<S1:EN>**			**<S1:EN>**
SP0	RM1>	SP1	SP1	SP1><RI0	RI1>

LSFB : CLSFBE
« yes it happened in Paris Paris uh three four years ago in Paris»

PARIS	PARIS	EUH	TROIS	QUATRE	ANS	PASSE	PARIS
	<S1 :EN>						
<RI0	RI1><RM0	**<S3 :EU>**	<SP0	SP1>			RM1>

3.1.2 Palm-up in SL (PU)

This category covers the following hand configuration: one or two hands are stretched, the five fingers separated from each other, making a rotating movement from the wrist to get the palms up. Palm-ups occur in the neutral space in front of the signer.

Palm-ups can have several functions (Loon 2012): they can indicate the beginning or end of a turn, show the signer's attitude towards his/her own discourse, invite the hearer

[6]In all LSFB examples, the sentence above the table is a translation in English without punctuation. The top line is a gloss of the video source, i.e. each sign appearing in the discourse in LSFB is assigned a label corresponding to a lemma.

to react, establish a relation between two units of discourse, fill a stop while showing the will to keep the turn. Palm-ups are thus quite similar to discourse markers (see Bolly, Gabarro-Lopez, and Meurant (2015) and Gabarro-Lopez (2015) for more detail).

LSFB : CLSFBE JMS22076
« that's it I sign now yes uh in my opinion well what is the goal of the deafs world day »

/	PT :PRO1	SIGNER	/	OUI	(220)	EUH	PT :PRO1
<**PU**>			<**PU**>		<S2 :CR>	<S3 :EU>	

/	JMS	VRAI	SOURD	MONDE	JOUR	POUR	QUOI
<**PU**>							

LSFB : CLSFBE JMS22076
« deaf people oh they are not really involved enough in the world of deafness »

SOURD	/	PAS	ASSEZ	DANS	PLUS	MONDE	SOURD
	<**PU**>						

3.1.3 Discourse marker (DM)

This category covers grammatically heterogeneous and multifunctional elements that are syntactically optional and oriented towards interpretation processes. They function at the metadiscursive level as cues to situate their host-unit in a dialogically co-built representation of speech (Brinton 1996; Hansen 2006; Crible in press). They can either signal a discourse relation between the host-unit and its context, make the structural organizations of segments explicit, express the speaker's metacomments on their speech, or contribute to interpersonal collaboration.

Discourse markers thus cover very diverse elements such as coordinating and subordinating conjunctions ("donc", "parce que"), adverbs ("enfin"), particles ("ben"), etc. A limited number of interjections can also be considered as discourse markers when they fulfill a text-structuring function (e.g. "oh" to introduce reported speech, "eh ben" as turn-opener). Backchannelling[7] devices are excluded from this category given their non-lexical form ("mh mh") and specific function.

[7]We define backchannel on an interactional criterion: they are the signal of understanding, vocalized by the interlocutor, i.e. not the current main speaker.

To keep a word-level annotation, each element in a discourse marker (for instance "je veux dire") receives the two-letter label. This does not mean that each separate element is granted the status of discourse markers on its own. Thanks to the bracketing system (< on the first and > on the last word of a phrase), it is possible to concatenate the labels assigned to each word in order to get a unique unit in post-treatment: the phrase "je veux dire" will be annotated <DM DM DM>.

The difference between discourse markers and explicit editing terms can sometimes be subtle. Therefore, the following criteria borrowed from Crible (2014) will be used to distinguish them and exclude some pragmatic expressions from the DM category:

- explicit reference to lexical access trouble (thus excluding expressions such as "comment", "comment dire", "comment dirais-je", "si je peux dire")
- low degree of grammaticalization (free juxtapositions and semantic transparency)
- presence of propositional content (thus excluding "if I may say so"' for instance).

Borderline cases are, for instance, "si vous voulez", "si tu veux", "if you will", "je dirais", "on va dire", "I don't know, "I suppose", which present a high degree of fixation but remain explicit references to the act of speaking or thinking. In some contexts, they can still be considered as discourse markers. We acknowledge that the identification of discourse markers is never free from language-specific considerations and remains partly subjective when dealing with items at the border of the category.

LINDSEI : FR014-F

I	don	't	know	but	in	fact	I	went	to	er	the
				<DM>	**<DM**	**DM>**				<FP>	

Backbone : Bb_en019

which	is	/	sort	of	/	I	suppose	would	be
	<SM0	<UP>	**<DM**	**DM>**	<UP>	**<DM**	**DM>**	SM1	SM1>
		<WI>	**<WI**	**WI>**	<WI>	**<WI**	**WI>**		

Backbone : Bb_fr021

loc/	euh	enfin	locales	euh	enfin	du	coin	enfin
<TR	<FP>	**<DM>**	TR><SP0	<FP>	**<DM>**	SP1	SP1>	**<DM>**
	<WI>	**<WI>**		<WI>	**<WI>**			

Backbone : Bb_fr013

mon	ami	Georget	(315)	monsieur	Georget	Daumas	plutôt	quoi
<SP0	SP0	<RM0	<UP>	SP1>	RM1>		<ET>	**<DM>**
			<WI>					

LSFB : S041 CLSFBI1905
« first ASL it's signs with iconicity it's visual »

PT :UN	FS :ASL	LS	ICONICITE	VISUELLE	LS	PT :UN
<DM><RE0		<RE0			RE1>	**<DM>**RE1>

LSFB : S041 CLSFBI1905
« if you didn't understand I can sign»

SI	PT:PRO2	COMPRENDRE-NEG	PT:PRO1	LS
<DM>				

3.1.4 Explicit editing term (ET)

This category covers any element that explicitly signals the perception of a disfluency by the speaker (Strassel 2003 : 8). It excludes discourse markers and filled pauses, which have their own label. A sequence of fluencemes can include more than one explicit editing term. The latter are only annotated when they are located in the immediate left or right context of another fluenceme of which they form the explicit acknowledgment or when they are framed by other fluencemes. Explicit editing terms can be verbal or non-verbal in SL.

Backbone : Bb_fr004

beaucoup	de	euh	comment	(242)	d'	exploitants
	<SM0	<FP>	**<ET>**	<UP>	SM1>	
		<WI>	**<WI>**	<WI>		

Backbone : Bb_en0014

over	a	year	(186)	what	is	it ?	(311)	fifteen	months
<SP0	SP0	SP0	<UP>	**<ET**	**ET**	**ET>**	<UP>	SP1	SP1>
			<WI>	**<WI**	**WI**	**WI>**	<WI>		

LINDSEI : FR015-F

there	is	(380)	less	no	it	's	er	less	noisier
<SP0	<RM0	<UP>	RM0	**<ET>**	SP1>	RM1	<FP>	RM1>	
		<WI>		**<WI>**			<WI>		

*LSFB :*CLSFBE JMS20060
« there are conferences that inform us sorry I'm confused uh »

CONFERENCE	INFORMATION	/	PARDON	CONFUSION	EUH
		<PU>	**<ET>**	**<ET>**	<S3>

3.1.5 False start (FS)

This category covers cases of auto-interruption leaving a discourse segment syntactically and/or semantically incomplete and abandoned (Pallaud, Rauzy, and Blache 2013). No element of the false start can be taken up in the next segment at lemma level (Biber et al. 2000), in which case it will be categorized as morphosyntactic (SM) and/or propositional substitution (SP) (see section 3.2.2). Only the last word in the incomplete sequence is tagged.

Backbone : Bb_fr008

nous	sommes	devenus	donc	vraiment	mais	la	Provence
			<DM>	**<FS>**	<DM>		

Backbone : Bb_en014

for	women	possibly	to	have	you	know	(231)	they	're	getting
				<FS>	<DM	DM>	<UP>			

LSFB : CLSFBE-JMS20064
« they learn ne/ it's as if they discover »

OUI	APPRENDRE	OUI	NOUVEAU	COMME	DECOUVRE
			<TR><FS>		

3.1.6 Truncation (TR)

This category covers any word fragment, either completed (immediately or with delay) or abandoned. This phenomenon reveals formal incompletion of a morpheme or word that can be left incomplete or be taken up and modified (Pallaud and Henry 2004). Similarly, in LSFB, a truncation is the onset of an interrupted sign. The sign must be recognizable by the hands configuration and the initial localization (albeit incomplete). The truncated word/sign can be directly completed with the repetition of the fragment, or completed after the insertion of a few elements, or be abandoned (Henry and Pallaud 2003 : 78). In this last case only, truncations are simple fluencemes; otherwise in all the other cases, they are compound.

The fragment is annotated along with its completed form, when applicable. An abandoned fragment takes opening and closing brackets. Without counter-evidence, we assume that the next word is the complete form of the fragment as soon as the first phoneme is in common between the fragment and its restart. When a truncation is combined with a false start, the former fluenceme is annotated before the latter, as in the LSFB example in section 3.1.5 (CLSFBE-JMS20064).

LSFB : S070 CLSFBI3406
« sometimes a person is mad so they sign fast »

DEPEND	LS	DEPEND	PERSONNE	NERVEUX	LS
<TR	**<TR**	**TR>**			**TR>**

LSFB : CLSFBE-JMS22-068
« to meet members of the ASBL and get lessons, explanations[8]

ASBL	RENCONTRER	ENSEIGNER	ENSEIGNER	EXPLIQUER
		<TR	**TR>**	
		<AR>		

In accordance with our definition of identical repetition, according to which a repetition only applies to a lexical segment (see section 3.2.1), a truncation cannot be affected by a repetition. Successive fragments will be annotated by a numbering system, as in the following example:

[8]The signer begins the standard sign "ENSEIGNER". The sign starts from the signer's body to the interlocutor. Then, the signer stops and changes the standard sign "ENSEIGNER" (in the sense of "I teach") by the derived version of the verb ("I am taught"). The signer's hands change their direction and the sign is completed from the interlocutor towards the signer.

(Built example)

la	m/	m/	maman
	<TR0	**TR1**	**TR>**

When a fragment is completed, it can be immediately or after a lexical insertion or another fluenceme.

Backbone : Bb_fr018

ils	ét/	euh	ils	étaient	tout	gênés
<RM0	**<TR**	<FP>	RM1>	**TR>**		
		<WI>				

The following two cases are examples where the audio recording is necessary in order to determine the status of the truncated word, i.e. whether it is a misarticulation or a propositional substitution.

Backbone : Bb_fr003

intégration	de	la	poli/	de	la	population	d'origine
	<RI0	RI0	**<TR**	RI1	RI1>	**TR>**	
			<AR>				

Backbone : Bb_en009

and	po/	after	that	it'	s	partnership
<DM>	**<TR**	<IL	IL	IL	IL>	**TR>**
	<AR>					

3.2 Compound fluencemes

As mentioned above (section 2.1), the structure of a compound fluenceme consists in at least two parts. It is not excluded that these phenomena can apply to elements previously annotated as simple fluencemes, namely discourse markers that can be repeated or substituted.

For the annotation of compound fluencemes, a numbering system is used: identical numbers correspond to words from the same part of the fluenceme while increasing numbers represent the number of times the segment has been repeated and/or substituted, as will become clear with the examples of the following sections.

3.2.1 Repetitions (RI, RM, RE, RG)

Bearing in mind that this annotation system makes no a priori judgment between potentially fluent or disfluent markers, all types of repetitions are annotated, including "rhetorical" repetitions and repetitions that are caused by external interventions in the interaction, provided they occur in the same speech turn. By extension, this restriction also applies to reported speech, thus excluding from the annotation cases where the repetition spreads over both direct and reported speech.

The protocol accounts for cases where the repeated elements appear in a different order than their first occurrence: in the diacritics tier, the label <OR> can be added to mark a re-ordering. Moreover, other labels on the diacritics tier, namely <CF> for functional completion and <CT> for total completion, signal the degree of syntactic completion of the repeated or substituted segment (see sections 3.2.1 and 3.2.2).

Identical repetition (RI) This category covers cases where one or several (quasi) contiguous words are repeated in their exact same form and without any semantic addition (Candéa 2000). "Quasi-contiguity" refers to the possibility to insert an element with low or no propositional content between the two parts of the repetition, namely an UP, FP, DM, ET or an IP (since the content of a parenthesis does not affect the content of the repeated elements). As mentioned above, repetitions only apply to complete propositional elements, thus excluding truncations and filled pauses (which are not lexical).

Backbone : Bb_fr004

suite	à	euh	suite	à	quelques	euh	comment	dire
<RI0	**RI0**	<FP>	**RI1**	**RI1>**		<FP>	<ET	ET>
		<WI>						

LINDSEI : FR002-F

they	er	they	go	(630)	eh	they	go	to	bed
<RI0	<FP>	**RI1><RI0**	**RI0**	<UP>	<FP>	**RI1**	**RI1>**		
	<WI>			<WI>	<WI>				

LSFB : S050 CLSFB2406
« or or there a lot of times, too often, when I meet oralists [9] *»*

OU	OU	/	BEAUCOUP	/	MOMENT
<S1 :EN>					<S1 :EN>
<DM><RI0	**<DM>RI1>**	<S2 :NE>	<RM0	<S2 :NE>	

[9] Deafs communicating in a spoken language

/	BEAUCOUP	/	ORAL	RENCONTRER	RENCONTRER
			<S1 :ST>		
<S2 :NE>	RM1>	<S2 :NE>		<RG0	RG1>

Modified repetition (RM) This category covers cases where one or several (quasi) contiguous words are repeated either partially or with a change in content. This phenomenon is based on a less restrictive definition than that of identical repetitions since it also includes the possibility of having syntactico-semantic modification. This modification can be a lexical insertion, a deletion or a substitution. In other words, a RM differs from a RI by the fact that its content is modified by one or several propositional elements. This category, however, cannot be applied to discourse markers and explicit editing terms.

During the analysis of sequences, formal redundancy primes over semantic interpretation. In other words, if the semantic change (caused by an insertion for instance) is "strong" as in "I ate some chocolate, some chocolate cake" (change of referent), the sequence is annotated as a repetition (and not as a propositional substitution) since formal parameters are prioritized over semantic considerations, in an attempt to enhance reliability and objectivity.

Clitics are also excluded from this category in order not to overload the annotation: a partial repetition where only a clitic (e.g. a subject pronoun) is repeated and the other words (e.g. the verb) are substituted will not be annotated, as in "**they** come **they** go".

Backbone : Bb_en021

asian	speakers	well	no	asian	people	living	in	the	UK
<RM0	<SP0	<DM>	<ET>	**RM1>**	SP1>				

Backbone : Bb_en009

a	lot	of	time	a	lot	of	money
<RM0	**RM0**	**RM0**	<SP0	**RM1**	**RM1**	**RM1>**	SP1>
		<CF	CF>				

Backbone : Bb_fr008

c'	était	défendu	à	l'	époque	c'	était	défendu	de	parler
<RM0	**RM0**	**RM0**	<IL	IL	IL>	**RM1**	**RM1**	**RM1>**		
		<CF>								

LSFB : S041 CLSFBI1905

« here in Europe more in the region of Belgium and France we are against dactylology we don't like dactylology »

EUROPE	CA-VEUT-DIRE	NS :Belgique	PLUS.P	NS :Belgique	NS :FRANCE
					<S1 :ST>
		<RM0	<IL>	**RM1>**	

ENDROIT	(320)	DACTYLOLOGIE	CONTRE	DACTYLOLOGIE-neg
			<S1:EN>	
	<S2:NE>	**<RM0**	<IL>	**RM1>**

LSFB : S007 CLSFBI306

« and moreover if there is a little mistale they blame me for it and moreover the mistake is related to a communication problem »

PLUS	APPARAITRE	PETITE	ERREUR	ACCUSER	ERREUR	ACCUSER
<DM>			**<RM0**[10]	**RM0**	**RM1**	**RM1<RG0**

ACCUSER	PLUS	ERREUR	COMM_PAS	COMM_PAS	ERREUR	COMM_PAS
<S1:EN>						
RM1>RG1>	<DM>	**<RM0**	**RM0**<RG0	RG1>	**RM1**	**RM1>**

Framing repetition in SL (RE) This category covers repetitions referred to in the SL literature as "doubling" (Kimmelman 2013) or "reduplication" (Vermeerbergen and Vriendt 1994). According to these authors, this type of repetition belongs to syntactic structures available in SL (for instance, there could be two subject slots for one verb (SVS)) or has discourse functions (for example in the topic-comment structure, in contrast or parenthetical constructions). They appear in a X Y X form, following the model by Kimmelman (2013). This type of repetition can apply to any class of signs (nouns, verbs, question words, adjectives, pronouns, adverbs etc.) and take scope over a sign within a unit or over the unit itself if it frames another unit. There are thus two parts of a repetition framing a propositional element. The repeated components act as symmetrical "braces" around a central element made up of one sign to one or several clauses

[10]In this example, the presence of a modified repetition is motivated by the presence of a grammatical repetition in one part of the repetition only. This absence of the grammatical repetition in the other part prevents a categorization as identical repetition since grammatical repetitions involve a nuance of number (plural) or aspect.

so that the repeated component closes the chunk the first one had begun (see Notarrigo (2016) and Notarrigo, Meurant, and Van Herreweghe (2016) for more details).

For practical reasons, we do not annotate insertions (either lexical or parenthetical) between the two parts of a framing repetition since it can cover very large segments, in which case the whole text would receive a label <IL>, rendering the annotation counter-productive.

LSFB : S007 CLSFBI306
« we used to pretend to fight for play »

RIRE	JOUER	BOXE	RIRE
<RE0			**RE1>**

Grammatical repetition in SL (RG) This category covers so-called "morphological" repetitions (Filpczak and Mostowski 2013) or "morphosyntactic reduplication" (Wilbur 2005; Pfau and Steinbach 2006). Those have grammatical functions such as marking a change in number or gender, marking the scope of the verb or marking aspect. For example, to express the fact that there are several houses, the signer can repeat n times in a row the sign for "HOUSE" either at the same location or placing them in the space in front of him/her. Another case is that of an action that is performed repeatedly with a sense of boredom. For instance, the signer can produce n times in a row the sign for "WORK" to express the idea that they work a lot.

Some degree of variation between the two parts of the repetition is accepted: "repeated signs are not necessarily identical; one occurrence can present some modulation of the citation form or be accompanied by non-manual features that are absent in the other occurrence" (Vermeerbergen and Vriendt 1994). Kimmelman (2013) also observes phonological, phonetic and grammatical variations between two parts of a repetition. For example, the second occurrence can be articulated in a more relaxed way and with a shorter and weaker move than the first one. The author further notes a difference of placement of the sign in the second occurrence in order to add some information of aspect, modality or other (see Notarrigo (2016) and Notarrigo, Meurant, and Van Herreweghe (2016) for more details).

LSFB : S041 CLSFBI1905

« but the structure is the same as in French "De, la, à" they are taken out the structure of signs follows that of French »

MAIS	(830)	STRUCTURE	AUSSI	FRANÇAIS	FS :ADELA	RIEN	ENLEVER
				<S1 :EN>			
<DM>	<SE :NE>	<RE0	RE0	RE0			**<RG0**

ENLEVER	ENLEVER	MOT	STRUCTURE	AUSSI	LS	FRANCAIS
						<S1 :EN>
RG1	**RG2>**		RE1	RE1	<IL>	RE1>

3.2.2 Substitutions (SM, SP)

Morphosyntactic substitution (SM) This category covers any morphosyntactic modification. Any modification of a full lemma (thus excluding truncations) is annotated. In other words, any addition, change or elision of morpheme from the same lemma is annotated. Morphosyntactic substitutions often involve partial repetitions.

Backbone : Bb_fr004

qui	présentent	les	meilleurs	euh	la	meilleure	volaille
		<SM0	**SM0**	<FP>	**SM1**	**SM1>**	
				<WI>			

Backbone : Bb_en022

you	can	make	have	anything	made	in	Birmingham
		<SM0	**SM1**	<IL>	**SM1**[11]**>**		

Propositional substitution (SP) This category covers any take-up of a discourse segment by another which includes a semantic nuance or modification. Substitution can include cases of modified repetitions (Duez 2001). In other words, this category concerns any element being replaced, including in an interrupted sequence. Appositions and reference chains (e.g. "this man he went there") are however excluded from propositional substitutions in this protocol.

As for the different categories of repetition, different types of substitution are distinguished in this protocol, namely substitutions taking scope over complete segments

[11]The two forms "make" and "have made" can be considered as a change of structure rather than change of lemma (SP). In case of hesitation, a bias towards SM is preferred.

(see diacritics tier) and substitutions taking scope over incomplete segments. However, this incompletion is not annotated: only completion gets a label. This makes it possible to distinguish a sentence such as "he goes uh we go to the market" (substitution of an incomplete structure by a complete one) from "he goes to the market uh we go to the market" (substitution of complete structures). Note that cases of abandon with no formal or semantic take-up are analyzed as false starts (not as substitutions).

To determine the end of a substitution (right boundary), we use the grammatical criterion of word-to-word equivalence as scope of the annotation. If such an equivalence cannot be found in context, the interpretation of the end of the substitution is left to the annotator's judgment on the basis of semantic and/or syntactic criteria, with a minimal bias (anything forming a complete, coherent content, a functional sequence). This choice aims at standardizing the annotation and prevents over-annotations of large segments.

Backbone : Bb_fr008

douze	boulevard	Jean	Jaurès	/	non	douze	boulevard
<RM0	RM0	**<SP0**	**SP0**	<UP>	<ET>	RM1	RM1>
				<WI>	<WI>		

du	vieux	pont
SP1	**SP1**	**SP1>**

Backbone : Bb_en021

asian	speakers	well	no	asian	people	living	in	the	UK
<RM0	**<SP0**	<DM>	<ET>	RM1>	**SP1>**				
		<WI>	<WI>						

LSFB : S007 CLSFBI306
« in order for me to work well I need to be controlled »

BIEN	TRAVAILLER	BIEN	ENCADRER	ENCADRER
<RM0	**<SP0**	RM1>	**SP1><RG0**	RG1>

3.3 Insertions

3.3.1 Lexical insertion (IL)

This category covers the cases when a lexical element is inserted within a compound fluenceme, i.e. a partial repetition or a truncation. This phenomenon is not considered as a fluenceme *per se*, but its annotation is useful for the analysis in context of repetitions and truncations. Lexical insertions modify the co-text by adding some propositional

information. In an attempt not to overload the annotation, all elements in an insertion are framed by one opening and one closing bracket, although this grouping might not be meaningful in a phraseological perspective.

Lexical insertions are only annotated in the second part of a repetition, between the two parts of a repetition, or between a word fragment and its completion. If a word appears in the first part of a repetition and is no longer present in the second part, it is rather annotated as a deletion (Shriberg 1994) (see section 3.3.3). Any element appearing after the closing bracket of a compound fluenceme, thus outside the sequence, is not annotated.

Lexical insertions semantically modify a pre-existing content, and are thus different from simple fluencemes embedded in a sequence of compound fluencemes. Insertions do not get a "WI" diacritic label (see section 3.4.3).

Backbone : Bb_en009

I	deal	with	disputes,	so	civil	disputes
			<RM0	<DM>	**<IL>**	RM1>
				<WI>		

Backbone : Bb_fr008

c'	était	défendu	à	l'	époque	c'	était	défendu	de	parler
<RM0	RM0	RM0	**<IL**	**IL**	**IL>**	RM1	RM1	RM1>		
		<CF>								

LSFB : S007 CLSFBI306

« you see as if for instance as if we would write in French »

VOIR	TITRE	EXEMPLE	TITRE	ECRIRE	FRANÇAIS
					<S1 :EN>
	<RM0	**<IL>**	RM1>		

LSFB : S007 CLSFBI306

« it's normal even the oldest think hard to search I think about the answers and then I prepare before I type them »

NORMAL	PLUS	AGE	CONCENTRER	CHERCHER	PT :PRO1	CHERCHER
			<S1 :EN>			
				<RM0		<TR

PENSER	CHERCHER	REPONDRE	AUSSI	PREPARER	CLAVIER
<IL>	TR> RM1>		<DM>		

3.3.2 Parenthetical insertion (IP)

This category covers any propositional segment functioning as a "parenthetical aside" (Shriberg (1994 : 61) and Strassel (2003 : 42)). Parentheses are not considered as fluencemes *per se* and they are only annotated when they are placed within a sequence (i.e. in the immediate right or left context of a fluenceme, or embedded within a compound fluenceme). The content of the segment, although contextually related to the adjacent utterance, only indirectly affects this utterance by providing additional background information. It remains independent and is not syntactically integrated. As for lexical insertions, the brackets open and close on the first and last word of the parenthesis.

Backbone : Bb_en009

normally	would	take	you	(257)	before	you	're	a	fully
<RI0	RI0	RI0	RI0	<UP>	**<IP**	**IP**	**IP**	**IP**	**IP**
<OR	OR			<WI>					

qualified	solicitor	(222)	would	normally	take	you	a	minimum	of
IP	**IP>**	<UP>	RI1	RI1	RI1	RI1>			
		<WI>	OR	OR>					

Backbone : Bb_fr019

nous	avons	je	l'	ai	pas	dit	dans	la	présentation
		<IP	**IP**	**IP**	**IP**	**IP**	**IP**	**IP**	**IP**

au	début	mais	euh	un	euh	un	sous-sol
IP	**IP>**	<DM>IP>	<FP>	<RI0	<FP>	RI1>	

LSFB : S040 CLSFBI1905

« before I me my true way of expressing it's LSFB I think it's good and enjoyable why because I used to talk before »

AVANT	PT:PRO1	PT:PRO1	VRAI	EXPRESSION	LSFB	BIEN
<RI0	RI0	**<IP**	**IP**	**IP**	**IP**	**IP**
<OR	OR					

AGREABLE	POURQUOI	PARCE QUE	PT:PRO1	AVANT	PARLER
IP	**<DM> IP**	**<DM>IP>**	RI1	RI1>	
				OR	OR>

3.3.3 Deletion (DE)

This category covers the cases where a lexical element present in the first part of a modified repetition is absent in the second part. Deletions are conceptually the opposite of insertions, and as such do not constitute fluencemes either. Brackets apply to the first and last word of the deleted segment.

C- Humour : chfAR3

et	Rocard	qui	a	été	nommé	par	Sarkozy	ambassadeur
<DM>	<RM0	**<DE**	**DE**	**DE**	**DE**	**DE**	**DE**	**DE>**

au	pôle	nord	Rocard	au	pôle	nord
RM0	RM0	RM0	RM1	RM1	RM1	RM1>

3.4 Diacritic signs

A number of local phenomena are not treated as proper fluencemes but add some information to another element in the co-text. All diacritics are annotated in the second tier.

3.4.1 Misarticulation (AR)

This category covers any element identified as different from a "standard" or "correct" pronunciation and identified as such by the speaker him/herself. This misarticulation must be explicitly noticed by the speaker through a fluenceme (ET, DM, TR, etc.) otherwise it is not annotated (this is to avoid any reference to a linguistic "norm").

Backbone : Bb_en009

to	do	resiv/	residential	conveyancing
		<TR0	TR1>	
		<AR>		

3.4.2 Lenghtening (LG)

This category covers any lengthening of a phoneme at the beginning, middle or end of a word, relatively to its expected duration. This annotation is perceptive and therefore arguably subjective. The label is applied to the word affected by the lengthening regardless of the position of the lengthening within the word.

LINDSEI : FR002-F

in	relation	(310)	to	(480)	er	(720)	a	disease
		<UP>		<UP>	<FP>	<UP>		
			<LG>					

3.4.3 Embedding of simple fluenceme (WI)

A simple fluenceme framed by a compound fluenceme – for example, a pause in the middle of a repetition (between the two parts of the repetition) – gets a <WI> label (for *within*) on the diacritics tier. This information can be useful to compare the contexts of annotated fluencemes. Deletions, lexical and parenthetical insertions, however, never get the <WI> label given their "intrinsic" embeddedness.

Backbone : Bb_fr004

beaucoup	de	euh	comment	(231)	d'	exploitants
	<SM0	<FP>	<ET>	<UP>	SM1>	
		<WI>	**<WI>**	**<WI>**		

Backbone : Bb_en019

wapping	which	is	/	sort	of	/	I	suppose
		<SM0	<UP>	<DM	DM>	<UP>	<DM	DM>
			<WI>	**<WI**	**WI>**	**<WI>**	**<WI**	**WI>**

would	be	classified	as	the	East
SM1	SM1>				

LSFB : S007 CLSFBI306

« I was sad and frustrated I felt frustrated. But it's as if I had felt that with a lot of delay [with respect to the other members of his family] »

FRUSTRER	TRISTE	FRUSTRER	/	SENTIR	FRUSTRER
<RE0		RE1><RM0	<SE :NE>	<IL>	RM1>
			<WI>		

/	MAIS	PT : PRO1	TITRE	RETARD	RETARD
		<S1 :EN>			
<PU>	<DM>			<RG0	RG1>

3.4.4 Re-ordering (OR)

This category covers any element repeated identically – formally and semantically – but in a different syntagmatic order. Since the repeated elements are identical (except for the change of order), these cases will always be tagged "RI" in the first tier. The diacritic OR applies to all elements affected by the change of order in both parts of the fluenceme (RI0 and RIn). Brackets open on the first word (first part of the repetition) and close on the last (last part of the repetition).

Backbone : Bb_en009

normally	would	take	you	before	you're	a	fully	qualified	solicitor
<RI0	RI0	RI0	RI0	<IP	IP	IP	IP	IP	IP>
<OR	**OR**								

would	normally	take	you	a	minimum	of
RI1	RI1	RI1	RI1>			
OR	**OR>**					

LSFB : S007 CLSFBI306
« when my uncle died my mother my brother and I were sitting side by side »

PT :PRO1	MAMAN	FRERE	A-COTE	MAMAN	A-COTE	PT :PRO1	PT :PRO3	MORT
					<S1:EN>			<SI:ST> <S1:EN>
<RI0	RI0		RI0	RI1	RI1	RI1>		
<OR	**OR**		**OR**	**OR**	**OR**	**OR>**		

3.4.5 Syntactic completion - functional and total (CF et CT)

This category covers two levels of completion: (i) a minimal level corresponding to a full functional sequence, for instance "a lot of money a lot of money" (as opposed to "a lot of a lot of money"), and (ii) a total completion corresponding to a whole utterance, a dependency structure with a predicate and its governed elements, typically a clause. These two labels are applied to repeated or substituted elements and are only given to the last word of the first part of the compound fluenceme. Incompletion (a case where the repetition or substitution takes scope over incomplete segments) is not annotated in order not to overload the annotation.

Backbone : Bb_fr021

alors	un	canton	euh	un	canton	c'est
	<RI0	RI0	<FP>	RI1	RI1>	
		<CF>	<WI>			

Backbone : Bb_en021

and	that	's	I	use	that	I	use	that
<DM>		<FS>	<RI0	RI0	RI0	RI1	RI1	RI1>
					<CT>			

4 Queries

The annotation protocol presented here is designed with the perspective of efficient extraction and query by the researcher, following the principle that when one type of fluenceme is queried, all relevant items are extracted except for the internal elements of a compound or phrasal structure (e.g. "parce que", "in other words" will be counted as one DM and not respectively 2 and 3). Similarly, a segment repeated three times counts as one occurrence of a repetition. Since all labels contain two letters, querying by label should never retrieve unwanted occurrences.

Combining labels, brackets and numbers makes it possible to automatically extract groups and subgroups of phenomena without complex queries. Concretely, to extract all fluencemes of a specific type, the simple formula [<TAG] will suffice. For instance, querying **[<RM]** will return all occurrences of modified repetitions and identify the first element of this type of fluenceme.

Other queries can also be made on both tiers or on double labels in the same tier to specify particular contexts. For instance, all silent pauses in a "within" position, i.e. embedded in a compound fluenceme, can be extracted by filtering the <WI> tags applying to <UP>. Similarly, to extract all repeated discourse markers, the following formula can be used: **[<DM<R]**. The use of automatic scripts can help with more complex queries that are beyond the scope of this protocol.

5 Conclusion

This protocol aims at offering a methodological framework for the annotation of more than twenty (dis)fluency phenomena. It was tested on aligned multimodal and spoken corpora in English, French and LSFB.

Its exhaustivity pairs up with its flexibility: it is likely to adapt to many research questions while providing operational definitions of the phenomena under scrutiny. By avoiding any preconceived judgment on the fluency or disfluency of some markers, it substantially broadens the coverage of annotated phenomena compared to previous annotation systems, without additional difficulty. We hope to have stressed the usefulness of using formally replicable criteria for the analysis that do not partake in particular theories of syntax, prosody or semantics, thus limiting the impact of subjective semantic-pragmatic considerations in our approach to (dis)fluency. This protocol is still being tested on corpus, and inter-rater agreement analysis is in progress.

Bibliography

Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. 2000. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.

Boersma, P. and D. Weenink. 2014. *Praat: doing phonetics by computer [Computer program]. Version 5.3.80*.

Bolly, C., S. Gabarro-Lopez, and L. Meurant. 2015. "Pragmatic gestures at the gesture-sign interface. Nonmanuals and palm-up gestures among older Belgian French speakers and French Belgian Sign Language signers". In: *Workshop "Nonmanuals at the Gesture Sign Interface (NaGSI), Göttingen, Germany*.

Brinton, L. J. 1996. *Pragmatics markers in English. Grammaticalization and discourse functions*. New York: Mouton de Gruyter.

Candéa, M. 2000. "Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané. Etude sur un corpus de récits en classe de français." PhD thesis. Université Paris III Sorbonne Nouvelle.

Christodoulides, G., M. Avanzi, and J.-P. Goldman. 2014. "DisMo : A Morphosyntactic , Disfluency and Multi-Word Unit Annotator An Evaluation on a Corpus of French Spontaneous and Read Speech Presentation of DisMo". In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC) 2014, Reykjavik, Iceland, 26-31 May 2014*, 3902–3907.

Crible, L. in press. "Towards an operational category of discourse markers: A definition and its model". In: *Discourse markers, pragmatic markers and modal particles: New perspectives*. Ed. by C. Fedriani and A. Sanso. Amsterdam : John Benjamins.

Crible, L. 2014. *Identifying and describing discourse markers in spoken corpora. Annotation protocol v.8*. Tech. rep. Université catholique de Louvain.

Crystal, D. 1988. "Another look at, well, you know..." In: *English Today* 4.1, 47–49.

Dister, A. 2007. "De la transcription à l'étiquetage morphosyntaxique. Le cas de la banque de données textuelles orales VALIBEL". PhD thesis. Université catholique de Louvain.

Duez, D. 2001. "Signification des hésitations dans la parole spontanée". In: *Revue Parole* 17-18-19.33, 113–138.

Eklund, R. 2004. "Disfluency in Swedish human-human and human-machine travel booking dialogues". PhD thesis. Linkoping Studies in Science and Technology.

Filpczak, J. and P. Mostowski. 2013. "Repetition in Polish Sign Language (PHM). Discourse grammar information structure". In: *Theoretical Issues in Sign Language Research Conference (TISLR 11)*. London, United Kingdom.

Gabarro-Lopez, S. 2015. “Les marqueurs du discours en langue des signes de Belgique francophone (LSFB) et en langue des signes catalane (LSC): les "balise-liste" et les "palm-ups"”. In: *4th International Symposium "Discourse markers in Romance languages: A contrastive approach", Heidelberg, Germany.*

Gilquin, G., S. De Cock, and S. Granger. 2010. *LINDSEI Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM.* Louvain-la-Neuve, Belgique: Presses Universitaires de Louvain.

Götz, S. 2013. *Fluency in native and nonnative English speech.* Amsterdam : John Benjamins.

Grosman, I. in press. “How do French humorists mangage their persona across situations? A corpus study on their prosodic variation”. In: *Metapragmatics of humor: Current research trends.* Ed. by L. Ruiz-Gurillo. Amsterdam : John Benjamins.

Hansen, M.-B. M. 2006. “A dynamic polysemy approach to the lexical semantics of discourse markers (with an exemplary analysis of French toujours)”. In: *Approaches to discourse particles.* Ed. by K. Fischer. Amsterdam: Elsevier, 21–41.

Henry, S. and B. Pallaud. 2003. “Word fragments and repeats in spontaneous spoken French”. In: *Gothenburg Papers in Theoretical Linguistics 90.* Ed. by R. Eklund, 77–80.

Kimmelman, V. 2013. “Doubling in RSL and NGT: a Pragmatic Account”. In: *Interdisciplinary Studies on Information Structure* 17, 99–118.

Kurt, K. 2012. “Pedagogic corpora for content and language integrated learning. Insight from the BACKBONE Project”. In: *The Eurocall Review* 20.2, 3–22.

Little, D. R., R. Oehmen, J. Dunn, K. Hird, and K. Kirsner. Mar. 2013. “Fluency Profiling System : An automated system for analyzing the temporal properties of speech”. In: *Behavioral Research Methods* 45.1, 191–202.

Loon, E. van. 2012. “What’s in the palm of your hands? Discourse functions of Palm-up in Sign Language of the Netherlands”. MA thesis. University of Amsterdam.

Meteer, M. 1995. “Dysfluency annotation stylebook for the Switchboard corpus. Linguistic Data Consortium.”

Meurant, L. 2015. *Corpus LSFB. A digital open access corpus of movies and annotation in French Belgian Sign Language (LSFB).* Laboratoire de Langue des signes de Belgique francophone (LSFB-Lab), FRS-F.N.R.S et Université de Namur, www.corpus-lsfb.be. URL: `www.corpus-lsfb.be`.

Meurant, L. and I. Notarrigo. 2014. “Nonmanuals and markers of (dis)fluency in French Belgian Sign Language (LSFB)”. In: *Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel*, 135–142.

Moniz, H., F. Batista, A. I. Mata, and I. Trancoso. 2014. “Speaking style effects in the production of disfluencies”. In: *Speech Communication* 65, 20–35.

Notarrigo, I. 2016. “Les marqueurs de (dis)fluence en Langue des Signes de Belgique Francophone (LSFB)”. PhD thesis. University of Namur, Belgium.

Notarrigo, I., L. Meurant, and M. Van Herreweghe M.and Vermeerbergen. 2016. “Repetition of signs in French Belgian Sign Language (LSFB) and Flemish Sign Language (VGT): Typology and Annotation protocol”. In: *Poster présenté à TISLR 12*

- 12th International Conference on Theoretical Issues in Sign Language Research, Melbourne, Australie.

Pallaud, B. and S. Henry. 2004. "Troncations de mots, reprises et interruption syntaxique en français parlé spontané". In: *Le poid des mots*. Ed. by G. Prunelle, C. Fairon, and A. Dister. Louvain-la-Neuve 10-12 mars 2004 / March 10-12, 2004 Gérald: Presses Universitaire de Louvain, 707–715.

Pallaud, B., S. Rauzy, and P. Blache. 2013. "Auto-interruption et disfluences en français parlé dans quatre corpus du CID". In: *TIPA. Travaux interdisciplinaires sur la parole et le langage* 29.

Pfau, R. and M. Steinbach. 2006. "Pluralization in sign and in speech: A cross-modal typological study". In: *Linguistic Typology* 10.2, 135–182.

Schmidt, T. and K. Wörner. 2012. "EXMARaLDA". In: *Handbook on Corpus Phonology*. Ed. by J. Durand, G. Ulrike, and G. Kristoffersen. Oxford: Oxford University Press, 402–419.

Shriberg, E. E. 1994. "Preliminaries to a Theory of Speech Disfluencies." PhD thesis. University of California at Berkeley.

Sloetjes, H. and P. Wittenburg. 2008. "Annotation by category – ELAN and ISO DCR". In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. URL: `http://tla.mpi.nl/tools/tla-tools/elan/`.

Sperber, D. and D. Wilson. 1995. *Relevance. Communication and Cognition [2nd Edition]*. Oxford: Blackwell.

Strassel, S. 2003. *Simple Metadata Annotation Specification*, 1–43.

Vasilescu, I., M. Candea, and M. Adda-decker. 2004. "Hésitations autonomes dans 8 langues : une étude acoustique et perceptive". In: *Colloque MIDL 2004, Modélisations pour l'identification des langues et des variétés dialectales, 29-30 novembre*, 29–30.

Vermeerbergen, M. and S. de Vriendt. 1994. "The repetition of signs in Flemish Sign Language". In: *Perspectives on sign language usage. Proceedings of The Fifth International Symposium on Sign Language Research*. Ed. by I. Ahlgren, B. Bergman, and M. Brennan. England: University of Durham, 201–214.

Wilbur, R. B. 2005. "A Reanalysis of Reduplication in American Sign Language". In: *Studies on Reduplication*. Ed. by B. Hurch. Empirical Approaches to Language Typology 28. Berlin: Mouton de Gruyter, 595–623.

Annexe E

Données de corpus & matériel d'expérience

1 Annexes du chapitre 4

1.1 Tableaux croisés des pauses silencieuses

Cooc.	Loc. syntaxq.	Contour précédant la pause							Propriété de la pause					
		F	T	S	C	(NA)	Total	Prop	Durée moyenne		Ecart type		Médiane	Kurtosis
		(n.)						(%)	log(ms)	(ms)	log(ms)	(ms)	(ms)	
pause interregnum	intra-ga	0	1	35	17	62	115	30.67	2.22	210.01	0.31	144.46	180.9	-0.61
	inter-ga	4	7	45	19	105	180	48.00	2.5	440.75	0.38	379.27	363.02	-0.37
	inter-seq.	1	0	4	4	4	13	3.47	2.67	503.11	0.18	229.83	421.8	-1.04
	inter-rec.	3	5	8	7	37	60	16.00	2.73	637.98	0.23	477.53	526.2	0.76
	Total (n.)	8	13	92	47	208	368	100						
	Prop(%)	2.17	3.53	25.00	12.77	56.52	100							

TABLEAU E.1 – Propriétés et ditrisbutions des pauses silencieuses interregnum en fonction du contour précédant la pause et de la localisation syntaxique.

Cooc.	Loc. syntaxq.	Contour précédant la pause							Propriété de la pause					
		F	T	S	C	(NA)	Total	Prop	Durée moyenne		Ecart type		Médiane	Kurtosis
		(n.)						(%)	log(ms)	(ms)	log(ms)	(ms)	(ms)	
> 2 éléments	intra-ga	0	1	9	7	27	44	11.73	2.07	152.3	0.32	117.25	116.5	-0.12
	inter-ga	3	8	25	11	72	119	31.73	2.48	421.87	0.39	309.5	338.9	-0.36
	inter-seq.	0	0	2	2	12	16	4.27	2.73	572.78	0.16	224.76	543.85	0.44
	inter-rec.	6	27	24	44	95	196	52.27	2.73	654.35	0.27	462.07	531.8	0.5
	Total (n.)	8	13	92	47	208	368	100						
	Prop(%)	2.17	3.53	25.00	12.77	56.52	100							

TABLEAU E.2 – Propriétés et distributions des pauses en chaîne > à 2 éléments en fonction du contour précédant la pause et de la localisation syntaxique.

Cooc.	Loc. syntaxq.	Contour précédant la pause							Propriété de la pause					
		F	T	S	C	(NA)	Total	Prop	Durée moyenne		Ecart type		Médiane	Kurtosis
		(n.)						(%)	log(ms)	(ms)	log(ms)	(ms)	(ms)	
dis + pause	intra-ga	0	0	1	1	5	7	6.73	2.1	138.45	0.22	62.86	121.7	-0.91
	inter-ga	2	1	3	12	33	51	49.04	2.42	358.93	0.36	290.59	282.45	0.09
	inter-seq.	1	0	1	4	13	19	18.27	2.66	484.27	0.15	155.3	450.2	-0.59
	inter-rec.	1	4	1	1	20	27	25.96	2.77	750.48	0.28	686.74	539	0.5
	Total (n.)	4	5	6	18	71	104	100						
	Prop(%)	3.85	4.81	5.77	17.31	68.27	100							

Tableau E.3 – Propriétés et distributions des chaînes (dis + pause) en fonction du contour précédant la pause et de la localisation syntaxique.

Cooc.	Loc. syntaxq.	Contour précédant la pause							Propriété de la pause					
		F	T	S	C	(NA)	Total	Prop	Durée moyenne		Ecart type		Médiane	Kurtosis
		(n.)						(%)	log(ms)	(ms)	log(ms)	(ms)	(ms)	
pause + dis	intra-ga	0	1	7	0	7	15	13.04	2.14	149.51	0.18	61.84	138	-1.6
	inter-ga	1	2	9	6	37	55	47.83	2.45	375.86	0.33	377.62	333.3	0.03
	inter-seq.	1	1	1	3	12	18	15.65	2.6	413.84	0.13	129.47	397.8	-0.81
	inter-rec.	2	8	1	4	12	27	23.48	2.71	568.04	0.21	283.87	438.6	-1.24
	Total (n.)	4	12	18	13	68	115	100						
	Prop(%)	3.48	10.43	15.65	11.30	59.13	100							

Tableau E.4 – Propriétés et distributions des chaînes (pause + dis) en fonction du contour précédant la pause et de la localisation syntaxique.

Cooc.	Loc. syntaxq.	Contour précédant la pause							Propriété de la pause					
		F	T	S	C	(NA)	Total	Prop	Durée moyenne		Ecart type		Médiane	Kurtosis
		(n.)						(%)	log(ms)	(ms)	log(ms)	(ms)	(ms)	
pause + euh	intra-ga	1	0	4	1	54	60	41.10	2.19	186.77	0.27	119.43	145.3	-0.89
	inter-ga	0	0	2	1	30	33	22.60	2.73	575.55	0.17	206.06	581.4	0.17
	inter-seq.	0	0	0	2	15	17	11.64	2.67	535.57	0.23	327.29	361.3	-1.1
	inter-rec.	0	0	5	1	30	36	24.66	2.72	584.08	0.2	274.18	537.25	-1.11
	Total (n.)	1	0	11	5	129	146	100						
	Prop(%)	0.68	0.00	7.53	3.42	88.36	100							

TABLEAU E.5 – Propriétés et distributions des chaînes (pause + euh) en fonction du contour précédant la pause et de la localisation syntaxique.

Cooc.	Loc. syntaxq.	Contour précédant la pause							Propriété de la pause					
		F	T	S	C	(NA)	Total	Prop	Durée moyenne		Ecart type		Médiane	Kurtosis
		(n.)						(%)	log(ms)	(ms)	log(ms)	(ms)	(ms)	
euh + pause	intra-ga	0	0	4	3	4	11	5.21	2.2	207.47	0.37	131.24	221.1	-1.17
	inter-ga	5	3	10	12	71	101	47.87	2.51	382.44	0.27	214.4	356.2	0.05
	inter-seq.	1	1	0	3	23	28	13.27	2.64	458.17	0.12	151.71	432.9	0.77
	inter-rec.	1	6	1	11	52	71	33.65	2.74	611.77	0.2	323.5	477.9	-0.7
	Total (n.)	7	10	15	29	150	211	100						
	Prop(%)	3.32	4.74	7.11	13.74	71.09	100							

TABLEAU E.6 – Propriétés et distributions des chaînes (euh + pause) en fonction du contour précédant la pause et de la localisation syntaxique.

Cooc.	Loc. syntaxq.	Contour précédant la pause							Propriété de la pause					
		F	T	S	C	(NA)	Total	Prop	Durée moyenne		Ecart type		Médiane	Kurtosis
		(n.)						(%)	log(ms)	(ms)	log(ms)	(ms)	(ms)	
pause + md	intra-ga	0	1	10	22	8	41	16.47	2.11	148.05	0.22	91.94	126.6	-0.06
	inter-ga	1	5	5	14	49	74	29.72	2.35	296.27	0.35	241.88	243.37	0.67
	inter-seq.	1	1	0	5	16	23	9.24	2.64	480.83	0.18	248.58	372.8	0.32
	inter-rec.	2	22	25	27	35	111	44.58	2.7	601.54	0.23	507.76	450.7	1.79
	Total (n.)	4	29	40	68	108	249	100						
	Prop(%)	1.61	11.65	16.06	27.31	43.37	100							

Tableau E.7 – Propriétés et distributions des chaînes (pause + md) en fonction du contour précédant la pause et de la localisation syntaxique.

Cooc.	Loc. syntaxq.	Contour précédant la pause							Propriété de la pause					
		F	T	S	C	(NA)	Total	Prop	Durée moyenne		Ecart type		Médiane	Kurtosis
		(n.)						(%)	log(ms)	(ms)	log(ms)	(ms)	(ms)	
md + pause	intra-ga	1	3	0	6	8	18	2.63	2.1	145.47	0.25	80.28	136.8	-0.74
	inter-ga	5	15	6	33	155	214	31.29	2.39	335.68	0.36	283.61	245	-0.65
	inter-seq.	3	1	1	2	22	29	4.24	2.64	461.55	0.16	183.3	414.2	-0.26
	inter-rec.	15	50	5	54	299	423	61.84	2.76	673.09	0.25	419.44	549.1	0.93
	Total (n.)	4	69	12	95	484	684	100						
	Prop(%)	0.58	10.09	1.75	13.89	70.76	100							

Tableau E.8 – Propriétés et distributions des chaînes (md + pause) en fonction du contour précédant la pause et de la localisation syntaxique.

		Contour précédant la pause							Propriété de la pause					
		F	T	S	C	(NA)	Total	Prop	Durée moyenne		Ecart type		Médiane	Kurtosis
Cooc.	**Loc. syntaxq.**	(n.)						(%)	log(ms)	(ms)	log(ms)	(ms)	(ms)	
pause seule	intra-ga	5	11	45	24	58	143	5.51	2.17	182.52	0.29	118.72	143	-0.76
	inter-ga	8	43	86	205	939	1311	50.56	2.42	346.83	0.34	271.87	291	-0.46
	inter-seq.	6	9	21	51	268	355	13.69	2.69	561.72	0.21	357.13	443.8	0.38
	inter-rec.	9	120	12	69	554	784	30.24	2.81	796.55	0.28	549.45	632.9	0.87
	Total (n.)	8	183	164	349	1819	2593	100						
	Prop(%)	0.31	7.06	6.32	13.46	70.15	100							

TABLEAU E.9 – Propriétés et distributions des chaînes de pauses seules (isolées) en fonction du contour précédant la pause et de la localisation syntaxique.

2 Annexes du chapitre 5

2.1 Tableaux croisés des cooccurrences d'allongements et de pauses pleines

Position	initiale				médiane						finale			
	frontière		début		tour mono		tour court		tour long		fin		frontière	
	n.	%	n.	%	n.	%	n.	%	n.	%	n.	%	n.	%
p. pleine seule	15	4.57	24	7.32	4	1.22	14	4.27	225	68.60	31	9.45	15	4.57
p. pleine+p. sil	0	0.00	12	5.63	0	0.00	3	1.41	180	84.51	15	7.04	3	1.41
p. sil+p. pleine	4	2.88	9	6.47	0	0.00	1	0.72	112	80.58	13	9.35	0	0.00
p. pleine+euh	0	0.00	2	20.00	0	0.00	0	0.00	8	80.00	0	0.00	0	0.00
euh+p. pleine	1	12.50	0	0.00	0	0.00	0	0.00	3	37.50	0	0.00	4	50.00
p. pleine+md	5	3.82	12	9.16	5	3.82	6	4.58	80	61.07	7	5.34	16	12.21
md+p. pleine	5	13.51	3	8.11	3	8.11	3	8.11	16	43.24	4	10.81	3	8.11
p. pleine+dis	1	3.85	1	3.85	0	0.00	3	11.54	19	73.08	2	7.69	0	0.00
dis+p. pleine	4	8.51	2	4.26	2	4.26	1	2.13	30	63.83	4	8.51	4	8.51
p. pleine interregnum	3	3.37	8	8.99	2	2.25	1	1.12	67	75.28	5	5.62	3	3.37
$>$ 2 éléments	11	2.76	33	8.29	4	1.01	8	2.01	293	73.62	34	8.54	15	3.77

TABLEAU E.10 – Chaînes de cooccurrences des pauses pleines et leur proportion par rapport à leur position au sein du tour de parole pour les situations de dialogues et multilogues.

Position	**initiale**				**médiane**						**finale**			
	frontière		début		tour mono		tour court		tour long		fin		frontière	
	n.	%	n.	%	n.	%	n.	%	n.	%	n.	%	n.	%
all. seul	1	0.31	34	10.56	0	0	1	0.31	257	79.81	25	7.76	4	1.24
all.+UP	0	0	3	5.26	0	0	0	0	51	89.47	3	5.26	0	0
UP+all.	1	0.65	18	11.61	0	0	0	0	125	80.65	11	7.10	0	0
all.+FP	0	0	2	9.52	0	0	0	0	17	80.95	2	9.52	0	0
FP+all.	6	3.75	11	6.88	0	0	0	0	125	78.12	16	10.0	2	1.25
all.+DM	1	3.12	0	0	0	0	1	3.12	28	87.50	1	3.12	1	3.12
DM+all.	2	7.69	2	7.69	0	0	1	3.85	15	57.69	1	3.85	5	19.23
all.+dis	1	3.23	3	9.23	0	0	1	3.23	25	80.65	1	3.23	0	0
dis+all.	2	4.55	4	9.09	0	0	0	0	35	79.55	0	0	3	6.82
all.interregnum	1	0.93	13	12.04	0	0	0	0	89	82.41	3	2.78	2	1.85
$>$ 2 éléments	2	0.77	25	9.65	0	0	0	0	210	81.08	20	7.72	2	0.77

Tableau E.11 – Chaînes de cooccurrences des allongements et leur proportion par rapport à leur position au sein du tour de parole pour les situations de dialogues et multilogues.

2.2 Modélisation des pauses pleines allongements en fonction de la perception *offline* par des annotateurs experts

LogLik modèle vide	LogLik fichier en aléatoire
-520.89	-512.8

TABLEAU E.12 – Influence de l'effet aléatoire `fichier` sur la modélisation de l'annotation des allongements d'hésitation.

Le Tableau E.13 regroupe deux des modèles testés pour la variable dépendante « syllabe allongée /vs/ non allongée » : le modèle `minimal` s'oppose au `complet`. Ce dernier regroupe toutes les variables indépendantes mentionnées ci-avant [1].

Type de L	formule
`minimal m0`	`glmer(binary all ~ 1 + (1| file), data = dataCBL, family= binomial)`
`complet m1`	`glmer(binary all ~ durrel + mvt + f0rel + AP pos|+ cooc + 'pause-avant + pauseapres + (1| file), data=dataCBL, family= binomial(link=logit), contrasts=list(APpos=contr.helmert(4), f0rel=contr.helmert(3), durrel=contr.helmert(3), mvt=contr.helmert(3), cooc=contr.helmert(2), pause-avant=contr.helmert(2), pauseapres=contr.helmert(2))`

TABLEAU E.13 – Exemple de modélisation de la durée des allongements.

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(> *Chisq*)
effets aléatoires seuls	2	1029.61	1041.65	-512.81	1025.61	NA	NA	NA
m2=m1-pauseapres	13	738.15	816.35	-356.07	712.15	313.46	11	0.00
m1=full model	14	740.07	824.29	-356.04	712.07	0.08	1	0.78

TABLEAU E.14 – Comparaison de modèles d'effets mixtes.

1. Les coefficients pour les régresseurs Helmert comparent chaque niveau des variables indépendantes avec la moyenne des niveaux précédents. Dans ce type de modèle, l'intercept représente la moyenne de toutes les moyennes.

2.3 Protocole d'expérience par Duvivier (2017)

Nous reprenons du mémoire Duvivier (2017) l'explication des phénomènes et le protocole utilisé.

1. **Phénomène Déclencheur** : L'étiquette utilisée permet de décrire le phénomène qui a supposément entraîné la perception (et donc la pression de la barre d'espace).
- Euh : Pause pleine
- All : Allongement vocalique
- Inexpliqué : Faux positif
- Doute : Phénomène ambigu entre une pause pleine et un allongement du e (le locuteur ajoute un e épenthétique ou prononce le e final avec insistance)
- NA : Données non valides (exclues des analyses)

2. **Code position** : Le code indique si l'hésitation a été perçue lors de la première, de la seconde ou des deux sessions d'annotation.
- 11 : Une hésitation a été perçue durant la première session et une hésitation a été perçue durant la seconde session
- 10 : Une hésitation a été perçue à la première session mais aucune à la seconde
- 01 : Rien n'a été perçu à la première annotation mais une hésitation a été perçue durant la seconde
- NA : Données non valides (exclues des analyses)

3. **Rapidité** : Lorsque l'hésitation a été perçue à chacune des deux sessions, cette colonne permet de savoir durant quelle session d'annotation l'annotatrice été plus rapide pour effectuer la marque.
- 2 puis 1 : plus rapide pour presser la barre d'espace durant la seconde annotation
- 1 puis 2 : plus rapide pour presser la barre d'espace durant la première annotation
- NA : Le phénomène n'a reçu qu'une seule marque ou n'est pas valide

4. **Retard** : La pression a été associée à un mot qui suit le phénomène d'hésitation ayant généré la perception (et donc la pression de la barre espace). Dans ces cas, la pression a donc été faite en retard par rapport à la production du phénomène d'hésitation
- R1 : Le phénomène ayant entraîné la marque a été produit un token avant la marque (e.g. il y avait une euh maison *press* jaune).
- R2 : Le phénomène ayant entraîné la marque a été produit deux tokens avant la marque (e.g. il y avait une euh maison jaune *press*).
- R3/4/5 : Le phénomène ayant entraîné la marque a été produit 3/4/5 tokens avant la marque.
- Non : La pression a été associée au bon phénomène d'hésitation l'ayant entraînée.
- Sil avant : L'étiquette utilisée permet de savoir si le phénomène observé est, ou non, précédé d'un silence.

3 Annexes du chapitre 6

3.1 Expérience 1 : fluence cognitive des répétitions

Name	Formula
`lmer.comp`	`lmer(1000*logRT complex + (1-subject))`
`lmer.min`	`lmer(1000*logRT complex + cond + (1-subject))`
`lmer.specific`	`lmer.min + (1-item)`
`lmer.nquest`	`lmer.min + (1-item) + (1-nquest)`
`lmer.cog`	`lmer.min + (1-item) + (1-nquest) + cogrisk`
`lmer.full`	`lmer.min + (1-item) + (1-nquest) + cogrisk+age+gender+music+bil`

TABLEAU E.15 – Incrémentation progressive de la complexité du modèle de RT.

Notes : cond = facteur de fluence, complex = facteur de complexité, subject = participant, item = stimuli-énoncé, cogrisk = résultat du *Relative Risk Cognition Test*, nquest = ordre d'apparition du stimuli, bil = aptitude bilingue.

Model	Df	AIC	logLik	deviance	Chi	Df	Pr(>Chisq)
`lmer.comp`	4	75179	-37586	75171			
`lmer.min`	7	75173	-37579	75159	12.67	3	0.01 **
`lmer.specific`	8	75170	-37577	75154	4.63	1	0.03 *
`lmer.nquest`	9	74694	-37338	74676	477.95	1	<0.0001 ***
`lmer.cog`	10	74681	-37331	74661	14.75	1	< 0.0001 ***
`lmer.full`	16	74682	-37325	74650	11.65	6	0.07

TABLEAU E.16 – Résultat des tests de Likelihood ratio test pour les modèles généraux et spécifiques.

3.2 Expérience 2 : évaluation contextualisée des structures des répétitions

3.2.1 Questionnaire F-IRI

Instructions Merci d'avoir accepté de participer, nous allons maintenant vous demander de répondre à quelques questions nous permettant de situer votre profil de participant. Les énoncés suivants apportent des informations sur vos pensées et sentiments dans diverses situations de la vie quotidienne. Indiquez à quel point chaque phrase vous décrit en choisissant le chiffre approprié sur l'échelle indiquée : 1, 2, 3, 4 ou 5. Lisez attentivement chaque énoncé avant de répondre. Une fois que vous aurez décidé votre réponse, notez le chiffre dans la boîte placée devant cet énoncé. Il n'y a pas de bonne ou mauvaise réponse, essayez de répondre aussi sincèrement et précisément que vous le pouvez.

Questions

1. J'imagine et songe assez régulièrement aux choses qui pourraient m'arriver.
2. J'éprouve souvent de la sollicitude et me sens préoccupé(e) pour les personnes ayant moins de chance que moi.
3. Je trouve parfois difficile de voir les choses du point de vue de quelqu'un d'autre.
4. Il m'arrive de ne pas être désolé(e) pour les gens qui ont des problèmes.
5. Je me laisse vraiment prendre par les sentiments ressentis par les personnages d'un roman.
6. Dans les situations d'urgence, je me sens inquièt(e) et mal à l'aise.
7. Je suis habituellement objectif(ve) quand je regarde un film ou une pièce de théâtre, et je ne me laisse pas complètement prendre par l'histoire.
8. Lors d'un désaccord, j'essaie d'écouter le point de vue de chacun avant de prendre une décision.
9. Quand je vois qu'on profite d'une personne, j'ai envie de la protéger.
10. Je me sens parfois désemparé(e) quand je me trouve au beau milieu d'une situation fortement émotionnelle.
11. J'essaie parfois de mieux comprendre mes amis en imaginant comment ils voient les choses de leur propre point de vue.
12. Être complètement plongé(e) dans un bon livre ou film m'arrive assez rarement.
13. Quand je vois quelqu'un qui est blessé, j'ai tendance à rester calme.
14. Le malheur des autres ne me perturbe habituellement pas beaucoup.
15. Quand je suis sûr(e) d'avoir raison à propos de quelque chose, je ne perds pas mon temps à écouter les arguments des autres.
16. Après avoir regardé une pièce de théâtre ou un film, il m'arrive de me sentir comme si j'étais un des personnages.
17. Me retrouver dans une situation de tension émotionnelle me fait peur.

18. Quand je vois quelqu'un être traité de façon injuste, il m'arrive de ne pas ressentir beaucoup de pitié pour lui.
19. D'habitude je suis plutôt efficace pour gérer les situations d'urgence.
20. Je suis souvent touché(e), affecté(e) par les événements dont je suis témoin.
21. Je crois qu'il y a toujours deux facettes à chaque question ou problème et j'essaie de les prendre en compte toutes les deux.
22. Je me décrirais comme une personne au cœur tendre, plutôt compatissante.
23. Quand je regarde un bon film, je peux très facilement me mettre à la place d'un des personnages principaux.
24. J'ai tendance à perdre mes moyens dans des situations d'urgence.
25. Quand je suis très contrarié(e) par quelqu'un, j'essaie en général de me mettre à sa place pendant un instant.
26. Quand je lis une histoire ou un roman intéressant, j'imagine ce que je ressentirais si les événements de l'histoire m'arrivaient.
27. Quand je vois que quelqu'un a fortement besoin d'aide en situation d'urgence, je perds mes moyens.
28. Avant de critiquer quelqu'un, j'essaie d'imaginer ce que je ressentirais si j'étais à sa place.

3.2.2 Enoncé de contextualisation des stimuli audio

Stimuli Audio	Enoncé de contextualisation
104_1_1M_chfFF1	Projets de scène
1089_1_1M_intrad_3	Écriture d'un livre sur Jimmy Hendricks
1121_1_1M_intrad_4	Écriture d'un roman
1149_1_2NM_NARP01H	Importance des gestes dans l'enseignement
1151_1_2NM_NARP01H	Importance des gestes dans l'enseignement
1204_1_2NM_NARP03H	Découverte de la tour de Pise
1230_2_2NM_NARP04H	Graphisme des affiches dans un festival
1262_2_2NM_NARP05F	Gazette de quartier à Besançon
127_1_1M_chfFF2	Ecriture d'un scénario de cinéma
1326_1_2NM_NARP07F	Démission quand on travaille dans le marketing
1346_1_2NM_NARP08H	Voyage à Gap
1521_2_2NM_RhapD0003	Extinction des incendies
1556_2_2NM_RhapD0004	Vie de quartier à Paris
1605_1_2NM_RhapD0006	Choix d'un métier
183_2_1M_chmAT2	Spectacle de théâtre
1994_1_2NM_RhapM0019	Film récemment vu
2000_1_2NM_RhapM0021	Film récemment vu
2011_2_2NM_RhapM0023	Film récemment vu
2044_1_1M_RhapM2002	Relation entre les arts et la connaissance
2054_1_1M_RhapM2002	Production et réception d'une œuvre d'art
2136_2_1M_single_1	A propos de ceux qui écoutent la radio
2151_2_1M_single_2	Gestion du temps à la radio
2177_2_1M_single_3	Écriture d'un roman
220_1_1M_chmKA2	Réalisation d'un film
386_1_2NM_convf_2	Déclin de la noblesse française
566_1_1M_DIDI22Hrein2	Ressenti d'un médecin en consultation
581_2_1M_DIDP03FHgrenouille	Reproduction des grenouilles
606_2_1M_DIDP05FHterre	Cheval de trait et fonctionnement de la roue
615_2_1M_DIDP06FFcastes	Organisation des castes indiennes
622_2_1M_DIDP07FFaraignee	Reproduction des araignées
626_1_1M_DIDP08FFpharaon	Multiplication des dieux égyptiens
632_1_1M_DIDP08FFpharaon	Multiplication des dieux égyptiens
643_2_1M_DIDP09FFvenin	Distinction des types de serpents
663_2_2NM_G06A_RADIO4_01_SUMMARY	Service minimum en cas de grève
678_1_2NM_G07A_RADIO2_01_SUMMARY	Achat de vêtements
702_2_2NM_G09A_RADIO2_01_SUMMARY	Grèves des transports
736_2_2NM_G12B_RADIO2_01_SUMMARY	La "semaine du bonjour"
742_2_2NM_G12B_RADIO4_01_SUMMARY	Magouilles financières et déficits
838_2_2NM_intfor_1	Variations de la norme
959_2_1M_intlib_2	Écriture d'un livre

Tableau E.17 – Liste des énoncés de contextualisation de stimuli audio.

Annexe F

Structure des données de la thèse et accès aux données

Cette dernière annexe reprend la structure de données et scripts associés à ce travail doctoral. Vous pouvez contacter le centre de recherche VALIBEL.

STRUCTURE DU REPERTOIRE

```
|   +---00_Avant_propos
|   +---00_Remerciements
|   +---02_1_Chapitre_1_INTRO
|   +---02_2_Chapitre_2_DEF
|   +---02_3_Chapitre_3_CORP
|   |   +---2014-02_Crible_al_AIA_Kappa_situational_features
|   |   +---2016-09_Corpus_Summaries
|   |   \---2017-03_Crible_Grosman_AIA_DIS
|   +---02_4_Chapitre_4_SIL
|   |   +---2017_LANGAGE_pauses-silencieuses
|   |   +---UP_contexte
|   |   \---UP_duration
|   |       +---2016-05_Pauses_treshold_cluster_BOTH
|   |       +---2016-09_Pauses_treshold_cluster_LOCAS
|   |       +---2016-09_Pause_Cluster_treshold
|   |       +---archive_R_scripts
|   |       +---Disfluency_2017
|   +---02_5_Chapitre_5_FIL_ALL
|   |   +---1_FP_ALL_GROSMAN_DISTRIB_DUREE_CONTOURS_PRODUCTION
|   |   +---2_FP_ALL_GROSMAN_VERCRUYSSEN_DUVIVIER
|   |   |   +---Analyse
|   |   |   +---Corpus
|   |   |   \---Protocole
|   |   +---3_FP_ONLINE_DUVIVIER_CHRISTO_GROSMAN
|   |   |   +---01_GAMEANNOTATOR_ANALYSIS
|   +---02_6_Chapitre_6_REP
|   |   +---2014-06_Grosman_EXPE_Repetitions-complexity
|   |   +---2016-09_Repetition_sequence_ARC_Disfluency2017
|   |   +---2017-02_METACORPUS_Grosman_al_REP_proso_analysis
|   |   +---2017-03_Grosman_al_EXPE_rep_sit
|   |   +---2017_LOCAS_CHUM_REP_Distribution
|   +---02_7_Chapitre_7_OVERALL
|   |   +---2014-11-Interactivefluency
|   |   +---2016-06_Visualize_disfluencies_LOCAS
|   |   +---2017-02_ARC_worshop
|   |   +---2017_MCA__BDU_TREE
|   |   +---Crible_al_2017_Compare_Disfluency2017_presentation
|   |   +---Grosman_2014_DisBduLocas_NouvCahierLingFranc
|   +---03_Appendix
|   |   |   00_Directory-Dendogram.R
|   |   |   03_A_AppendixCHUMOUR.Rnw
|   |   |   03_B_AppendixLOCASF.Rnw
|   |   |   03_C_AppendixTIER.Rnw
|   |   |   03_D_AppendixARC.Rnw
|   |   |   03_E_AppendixExpe.Rnw
|   |   |   Crible-et-al_2016_ARC_protocole.pdf
|   |   +---2013-2017_other_mini-studies
```

CORPORA_PROCESSING

```
|       +---00_PrepareData_CHUMOUR
|       +---00_PrepareData_LOCAS
|       +---2013-2017_MAKE_TRANSCRIPTS_ALL_KINDS
|       +---2013-2017_READ_WAV
|       +---2014-03_MAKE_LEX_PERL
|       +---2014-06_EXTRA_DIS_BDU_PERL
|       +---2016-05_MAKE_DIS_SEQ_PRAAT
|       +---2016-09_FIX_SPEAKER-DELIVERY-
|       +---2016-09_MAKE-ACCENTUAL-GROUP
|       +---2016-09_MAKE_IPU_TIER_PRAAT
|       +---other_script_program
```

DATA

```
|   +---CHUMOUR_grosman_thesis
|   |   +---Corpus
|   |   +---Metadata
|   |   \---Transcript
|   +---LOCAS_2016
|   |   +---Corpus
|   |   +---Metadata
|   +---LOCAS_grosman_thesis
|   |   +---table_syll
|   |   \---table_word
|   +---MERGED_Praaline
|   |   +---metadata_new
|   |   \---metadata_original
|   \---METACORPUS
|       +---CHUMOUR_2016
|       +---cprom-txt
|       +---DRIVING_2016
|       +---LOCAS_2016
|       +---MERGED
|       +---Metadata
|       +---PHONOGENRE_2016
|       \---RHAPSODIE_2016
```

ADMIN

```
|   +---Academic_reports
|   |   +---2012-2017_Appendices_protocoles_corpus
|   |   +---2012-2017_Conferences_reports
|   |   |   +---2013-09_IDP2013_pdn
|   |   |   +---2014-00_Other_conferences
|   |   |   +---2014-02_FrantInalco
|   |   |   +---2014-05_Journee_des_doctorants
|   |   |   +---2014-05_PLINE_linguists_day
|   |   |   +---2014-11-ICLC13-England-expeprototype
|   |   |   +---2014_09_SWIP_report
|   |   |   +---2015-10_Journee-doctorants
|   |   |   \---2015-12_CLAREWorkshop
|   |   +---2012-2017_Formations_reports
|   |   |   +---2013-09-JMP_formation_SMCS
|   |   |   +---2013-10_ExMaRaLDA
|   |   |   +---2014-11_Tronto-Autumn-School
|   |   |   +---2015-02_AcademicWriting
|   |   |   +---2015-06_Bootcamp
|   |   |   +---2015-11_interrater-seminar
|   |   |   +---2015-11_RedactionProjet-LucGauthier
|   |   |   \---2015-12_Stat Multivariées avec R Commander 2014
|   |   +---2012-2017_Official_reports
|   |   |   +---2012-2017_ARC-annual-reports
|   |   |   +---2013-11_Synthese-rapport-these
|   |   |   \---2015-06_Epreuve-de-confirmation
|   |   +---Grosman_PhD_followup
|   |   \---Readings_2015
|   +---contracts
|   +---Grosman_CV
|   \---publications
|       +---Christo_al_2017_Pauses-Distribution_Disfluency2017_presentation_DIAL-ok
|       +---Crible_al_2015_ARC-One-protocol-rule_3rdndARCworkshop_presentation_DIAL-ok
|       +---Crible_al_2015_Protocole-Annot-EN_DIS_report_DIAL-noneed
|       +---Crible_al_2015_Protocole-Annot-FR_DIS_report_DIAL-ok
|       +---Crible_al_2017_AcrossLggARC_Disfluency2017_presentation-DIAL-ok
|       +---Crible_al_2017_ProcDISFV_article
|       +---Duvivier_al_2017-Euh-all_Exling2017_abstract-submitted
|       +---Grosman_2013a_Master-thesis-text_report-DIAL-ok
|       +---Grosman_2013b_Dis-thesis_intro_Dubrovnik_Fall_school_poser-DIAL-ok
|       +---Grosman_2014_Dis-thesis_intro_IlcPosterDay_poster_DIAL-ok
|       +---Grosman_2014_DisBduContours_2ndArcWorkshop_presentation_DIAL-ok
|       +---Grosman_2014_DisBduLocas_NouvCahierLingFranc_article-DIAL-ok
|       +---Grosman_2014_DisBduLocas_SWIP03_Geneve_presentation_DIAL-ok
|       +---Grosman_2015_Chumour-DisVar_IPRA2015_presentation_DIAL-ok
|       +---Grosman_2015_ CausalityConvergence_MPI2015_poster_DIAL-ok
|       +---Grosman_2015_ConvergeMixMethods_3rdARCworkshop_presentation_DIAL-ok
|       +---Grosman_2015_Repetition-as-complexity-cue_ICPHS_DISS2015_article_DIAL-ok
|       +---Grosman_2015_Repetition-as-complexity-cue_ICPHS_DISS2015_poster_DIAL-ok
|       +---Grosman_2016_DisVar-Chumour_IPRA2014_JohnBenjamin_chapter_DIAL-ok
|       +---Grosman_al_2016_Percp-annot-hesi_CBL2016_presentation
|       +---Grosman_al_2016_Percp-annot-hesi_ILCARCseminar_presentation
|       +---Grosman_al_2017_Repetitions_prosody_Disfluency2017_presentation_DIAL-ok
|       \---Simon_al_2017_Pauses_article-abstract-submitted
```

Grosman, Iulia • 2018
Evaluation contextuelle de la (dis)fluence en production et perception. Pratiques communicatives et formes prosodico-syntaxiques en français •
Thèse de doctorat • Université catholique de Louvain.