

Les unités de discours : pourquoi et comment?

Anne Catherine Simon (anne-catherine.simon@uclouvain.be)

basé sur un travail conjoint avec **Liesbeth Degand**

28-29/05/2015 – Jétou 2015

« Le(s) Discours en Sciences du Langage:
unités et niveaux d'analyse »

Université Jean Jaurès, Toulouse, France



Contexte de la recherche

- le discours comme production langagière multimodale
→ l'interface prosodie-syntaxe et son rôle dans la production discursive
- la superposition d'unités entretenant des relations d'interdépendance forte
→ la segmentation du discours en Unités Discursives de Base (BDUs)
- les fonctions (planification, segmentation, organisation...)
→ éléments en périphérie gauche et localisation / distribution des marqueurs de discours



Contexte institutionnel de la recherche

Collaboration avec Liesbeth Degand, projets financés par le Fonds national de la recherche scientifique (FNRS)

- Projet FRFC 2007-2008: Établissement d'une procédure de segmentation du discours oral en unités minimales (MDU) sur la base de critères syntaxiques et prosodiques (A. Dister)
- Projet FRFC 2011-2014: Périphérie gauche des unités de discours: analyse syntaxique, prosodique et fonctionnelle et étude de l'impact sur la macro-organisation du discours (J. Eychenne, Th. Van Damme, N. Tanguy, L. Martin)



Plan de l'exposé

1. Le modèle des *Basic Discourse Units*
2. Le corpus LOCAS-F (*Louvain Corpus of Annotated Speech – French*)
 - Un corpus multigenres
 - Annotations syntaxiques et prosodiques
 - Aperçu plus détaillé des unités prosodiques
3. Évidences empiriques
 - Distribution des BDU selon les genres de discours
 - Types et fonctions des éléments en périphérie gauche
 - Fonctionnement des marqueurs de discours (DM)
4. Conclusions



Le modèle des *Basic Discourse Units*

- modèle développé pour traiter les productions orales; applicable à l'écrit
- assumer le paradoxe du discours
 - objet structuré, organisé (cohérence, relations de discours; ex. Grosz, Sidner, 1986; Mann, Thompson, 1988; Polanyi, 1988; Roulet et al., 2001)
 - événement temporel (linéarisation; caractère transitoire; achèvement interactionnel; ex. Couper-Kulhen 2001, Auer 2009)

“discourse planning involves at least two types: conceptual planning and strategic planning” (Hannay, Kroon 2005: 103)

Le modèle des *Basic Discourse Units*

- produire/interpréter un discours implique de manipuler
 - unités conceptuelles (idées, événements, etc. > *situation/event model*, Van Dijk 1997)
 - unités stratégiques (étapes de progression > *intonation units*, Chafe 1994; *communicative event*, Van Dijk 1997)
- BDU en **production**: paramètres situationnels (conceptionnels) favorisant des stratégies discursives
 - “different types of BDUs [exist] on the basis of their prosodic and syntactic shaping. We expect that these different types fulfill different functional roles in the construction of the ongoing discourse, especially in terms of information management and activity type” (Degand & Simon 2009)
- BDU en **réception**: indices de surface syntaxiques et prosodiques pour segmenter le flux discursif en unités, et pour les articuler entre elles
 - the segments that speakers use to build a representation (interpretation) of the discourse, i.e. a kind of “minimal discourse interpretation segment” (Degand & Simon 2009)

Principes méthodologiques

- les BDU sont identifiées en combinant indices syntaxiques et prosodiques, **sans donner la priorité à un niveau ou à un autre**
- approches défendant le **primat de la prosodie**
 - e.g. **Intonation Unit** (Chafe 1994) = "a unit of mental and linguistic processing" [...] "that seems to be of exactly the right size to be processed in its entirety with the help of echoic memory" (Chafe 1994:55)
 - "It is in terms of [**Tone Units**] – rather than any specific grammatical unit – that speakers organize and present information in discourse, and it is through TUs that listeners perceive and understand this information" (Altenberg 1987: 46)

Principes méthodologiques

- les BDU sont identifiées en combinant indices syntaxiques et prosodiques, **sans donner la priorité à un niveau ou à un autre**
- approches défendant le **primat de la syntaxe**
 - “In actual speech, **the processing of syntactic structures in it-self is largely independent of prosodic information.** [...] In our view, it appears to be a truism that **prosody**, unlike syntax, **is not inherently essential to the conveyance of information** in human communication.” (Monschau, Kreyer and Mukherjee 2003: 584-585)
 - “the basic elementary discourse unit (EDU) **is a clause**”(Carlson & Marcu 2001)

L'épreuve des faits...



Unités prosodiques

elle est à la masse elle m'énervé///
franchement je vois ma grand-mère
elle ///
elle a mal partout ///
mai:s elle a sa t/ enfin tu vois
elle est encore juste quoi ///

- 4 unités prosodiques
- 2 marques de disfluence

Unités syntaxiques

[elle est à la masse]
[elle m'énervé]
<franchement>
[je vois ma grand-mère]
[elle + elle a mal partout]
<mais>
[elle a sa t/]
<enfin>
<tu vois>
[elle est encore juste]
<quoi>

- 6 unités de rection complètes,
dont 1 inachevée
- éléments non régis (associés, MD...)



L'épreuve des faits...



Unités prosodiques

nous en arrivons ainsi à l'espace
national et régional///
et nous devons nous demander comment
celui-ci va s'inscrire///
dans le projet européen lui-même///
pour l'UCL///
cet espace c'est d'abord celui de
ses relations avec les autres
universités de la Communauté
française///
et avec les hautes écoles///

- 6 unités prosodiques

Unités syntaxiques

[nous en arrivons ainsi à l'espace
national et régional]
<et>
[nous devons nous demander comment
celui-ci va s'inscrire dans le
projet européen lui-même]
[pour l'UCL cet espace est d'abord
celui de ses relations avec les
autres universités de la
Communauté française et avec les
hautes écoles]

- 3 unités de rection complètes
- 1 élément non régi (connecteur)

La coïncidence intono-syntaxique n'est pas la norme

- 40% des Intonation Units [IU] ne correspondent pas à 1 clause (Chafe 1994)
- 97% des IU sont des Grammatical Units [GU] (de différents niveaux!) et 25% des GU ne correspondent pas à 1 IU (Croft 1995)
- 50% des transitions entre tours de parole ne correspondent pas à une complétude syntaxique, intonative et pragmatique (Ford & Thompson 1996)
- 15% des Syntagmes Nominaux disloqués n'ont pas de frontière prosodique obligatoire (Avanzi 2011)



Plan de l'exposé

1. Le modèle des *Basic Discourse Units*
2. Le corpus LOCAS-F (*Louvain Corpus of Annotated Speech – French*)
 - Un corpus multigenres
 - Annotations syntaxiques et prosodiques
 - Aperçu plus détaillé des unités prosodiques
3. Évidences empiriques
 - Distribution des BDU selon les genres de discours
 - Types et fonctions des éléments en périphérie gauche
 - Fonctionnement des marqueurs de discours (DM)
4. Conclusions



LOuvain Corpus of Annotated Speech - French

- 14 genres (ou activités communicatives), décrits selon les paramètres (cf. Koch & Oesterreicher 2001)
 - degré de préparation
 - degré d'interaction
 - caractère médiatique
 - etc.
- nombre d'échantillons 48
- durée totale 3h38
- nombre total de tokens 41 322

Un corpus multi-genres

Genre	Interaction	Préparation	Caractère médiatique	Tokens
Discours académique	non-interactif	préparé	semi-media	2331
Conférence scientifique	non-interactif	préparé	non-media	2850
Conversation formelle (professionnelle)	interactif	non-préparé	non-media	2360
Conversation informelle	interactif	non-préparé	non-media	2648
Débat politique	semi-interactif	semi-préparé	media	5110
Homélie	non-interactif	préparé	semi-media	1673
Interview sociolinguistique	semi-interactif	semi-préparé	non-media	2817
Interview télévisée	interactif	non-préparé	semi-media	3861
Interview radio	semi-interactif	semi-préparé	media	4142
Journal parlé	non-interactif	préparé	media	2817
Lecture	non-interactif	préparé	non-media	2890
Discours politique	non-interactif	préparé	media	3144
Séquence narrative (en conversation)	interactif	non-préparé	non-media	2233
Performance radiophonique	non-interactif	semi-préparé	semi-media	2446

Un corpus multi-genres

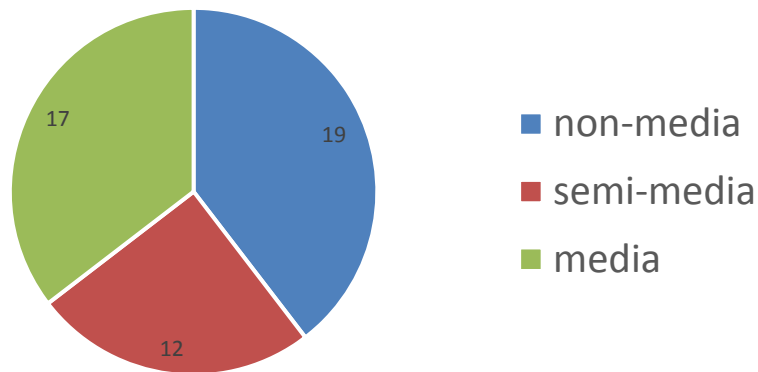
Genre	Interaction	Préparation	Caractère médiatique	Tokens
Discours académique	non-interactif	préparé	semi-media	2331
Conférence scientifique	non-interactif	préparé	non-media	2850
Conversation formelle (professionnelle)	interactif	non-préparé	non-media	2360
Conversation informelle	interactif	non-préparé	non-media	2648
Débat politique	semi-interactif	semi-préparé	media	5110
Homélie	non-interactif	préparé	semi-media	1673
Interview sociolinguistique	semi-interactif	semi-préparé	non-media	2817
Interview télévisée	interactif	non-préparé	semi-media	3861
Interview radio	semi-interactif	semi-préparé	media	4142
Journal parlé	non-interactif	préparé	media	2817
Lecture	non-interactif	préparé	non-media	2890
Discours politique	non-interactif	préparé	media	3144
Séquence narrative (en conversation)	interactif	non-préparé	non-media	2233
Performance radiophonique	non-interactif	semi-préparé	semi-media	2446

Un corpus multi-genres

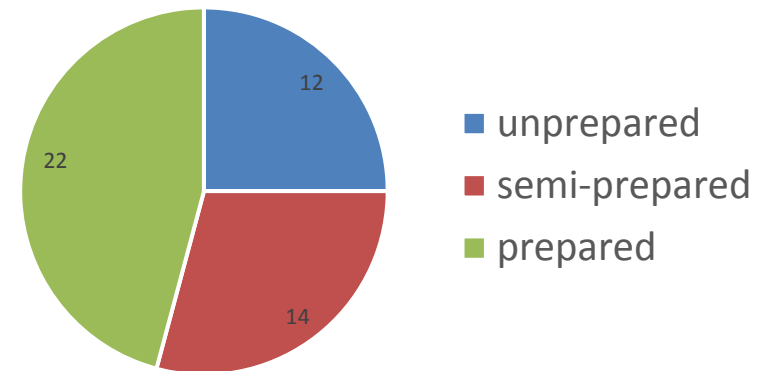
Genre	Interaction	Préparation	Caractère médiatique	Tokens
Discours académique	non-interactif	préparé	semi-media	2331
Conférence scientifique	non-interactif	préparé	non-media	2850
Conversation formelle (professionnelle)	interactif	non-préparé	non-media	2360
Conversation informelle	interactif	non-préparé	non-media	2648
Débat politique	semi-interactif	semi-préparé	media	5110
Homélie	non-interactif	préparé	semi-media	1673
Interview sociolinguistique	semi-interactif	semi-préparé	non-media	2817
Interview télévisée	interactif	non-préparé	semi-media	3861
Interview radio	semi-interactif	semi-préparé	media	4142
Journal parlé	non-interactif	préparé	media	2817
Lecture	non-interactif	préparé	non-media	2890
Discours politique	non-interactif	préparé	media	3144
Séquence narrative (en conversation)	interactif	non-préparé	non-media	2233
Performance radiophonique	non-interactif	semi-préparé	semi-media	2446

Répartition des 48 échantillons selon des paramètres situationnels

Caractère médiatique



Degré de préparation



- corpus pas complètement équilibré
 - préparé > non préparé
 - monologique > dialogique
 - certaines situations difficiles à documenter (ex. non interactif, non préparé)

Annotation syntaxique: unités de rection (UR)

- références
 - Blanche-Benveniste et al. (1990), *Le français parlé. Études grammaticales*
 - Bilger & Campione (2002), « Propositions pour un étiquetage en 'séquences fonctionnelles' »
 - guide de codage: Tanguy, Van Damme, Degand & Simon (2012)
- une UR est organisée autour d'un noyau (en général un verbe tensé) et se compose de l'élément recteur et de ses dépendants
- des éléments non régis peuvent entrer linéairement dans la composition d'une UR (inserts, marqueurs de discours...)

urv	unité de rection verbale (complète)
ure	unité de rection elliptique
ura	unité de rection averbale (nominale, adverbiale, etc.)

Annotation syntaxique: séquences fonctionnelles

- séquences maximales « qui représentent les constituants fonctionnels rencontrés dans les textes sans entrer dans le détail de leur composition » (Bilger & Campione 2002: 118)
- toute séquence peut faire l'objet
 - d'expansions: liste, coordination, reformulations, etc.
 - de mise en dispositif (extraction, binarisation, pseudoclivée...)
- toute unité (rection, séquence) peut être inachevée

SS	séquence sujet
SV	séquence verbe
SO	séquence objet
SR	séquence rection
SA	adjoint
insert	unité de rection qui interrompt une unité en cours

Exemple d'annotation syntaxique

- <alors>_{md} <bon>_{md} <pour ce qui est de notre euh représentation de l'intonation>_{ag} <en fait>_{md} [(on reprend)_{SV} (le terme de profil mélodique)_{SO}]_{urv} <c'est-à-dire qu'>_{md} [(on considère)_{SV} (que tout énoncé a un profil mélodique qui lui est associé)_{SO}]_{urv} [(c'est)_{SV} (un terme qu'avait utilisé par Dell quatre-vingt-quatre)_{SO}]_{urv} (cnf-1)
- [(visibilité)_{SN}]_{ura} <première caractéristique>_{ag} [(c'est)_{SV} (la ///S visibilité de la langue)_{SO}]_{urv} ///T (cnf-4)
- <alors>_{md} <pour finir>_{ag} [(on se dit)_{SV}]_{urv} <ben>_{md} [(on va aller voir)_{SV} (un autre film)_{SO}]_{urv} [(on n'allait pas)_{SV}]_{urv-l} <enfin>_{md} [(on va aller voir)_{SV} (un autre film)_{SO}]_{urv} (conv-inf-1)

Annotation prosodique

- peu de consensus autour d'un modèle des unités / frontières prosodiques (Portes & Bertrand 2011, Astésano, Bertrand, Espesser & Nguyen 2012, Mertens & Simon 2013)

"A problem for the categorical view of boundary strength, however, is that often when boundaries of different strength can be discerned, **the differences are quantitative rather than qualitative**. Experiments based on pitch accent scaling and on the various durational cues to prosodic boundary strength reviewed above suggest that **many correlates of boundary strength show gradient and cumulative effects**." (Wagner & Watson 2010: 911, je souligne)

Annotation prosodique

- annotation perceptive par deux annotateurs experts; réglage des désaccords (discussion) en vue d'arriver à une annotation de référence
- pour certaines parties du corpus
 - calcul du degré d'accord entre annotateurs et tentative de développer un modèle d'annotation automatique (Mertens & Simon 2013)
 - aide par une visualisation du signal (durée des pauses, mouvements mélodiques)

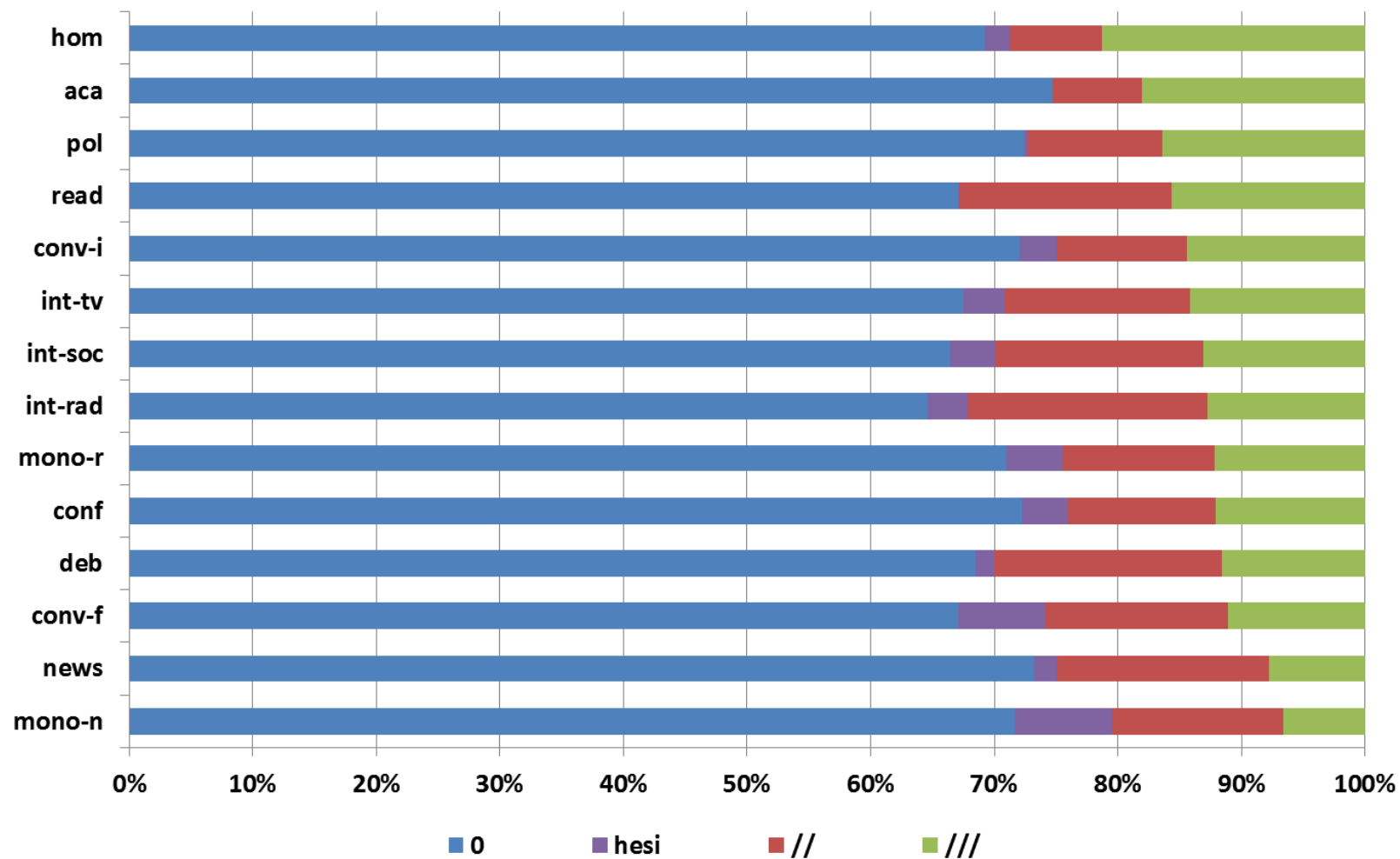
///	frontière prosodique majeure
//	frontière prosodique intermédiaire
hesi	hésitation (<i>euh</i> ou allongement vocalique)
interrupt	unité interrompue (par une prise de tour de l'interlocuteur)
#	paraverbal (rire, toux, etc.)
C, T, F, S	contour perçu (continuatif, terminal, focalisation, suspens)



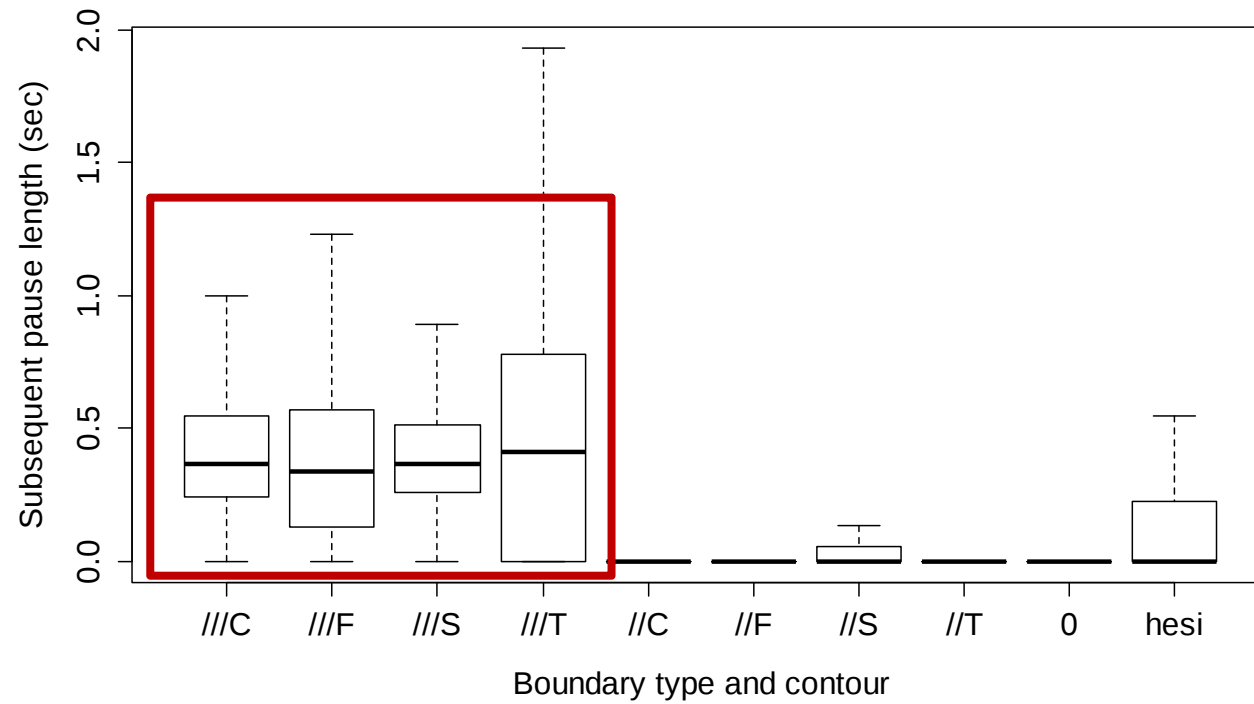
Un aperçu plus détaillé des corrélats acoustiques des FP (Christodoulides & Simon 2015, ICPHS)

- les frontières prosodiques se distribuent-elles différemment selon les genres?
- quels sont les prédicteurs des frontières prosodiques perçues (FP)?
 - corrélats acoustiques (durée relative de la syllabe, durée de la pause, mouvement mélodique)
 - POS et annotation syntaxique (séquences)

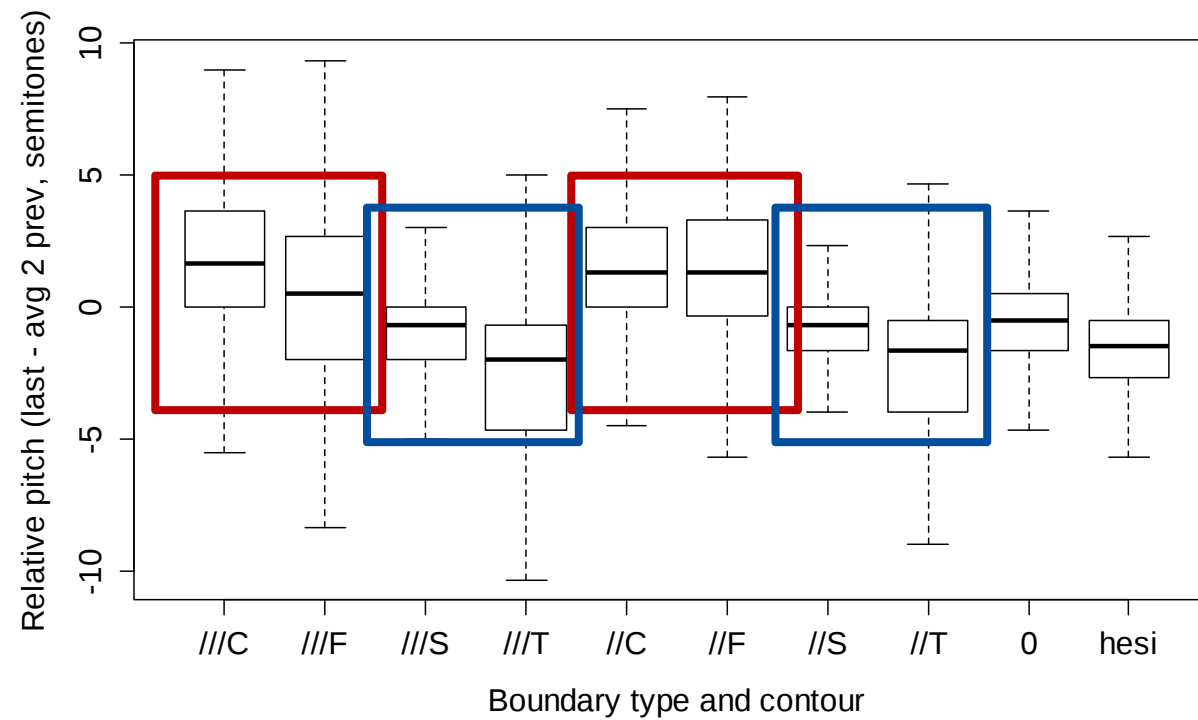
Types de frontières selon les genres



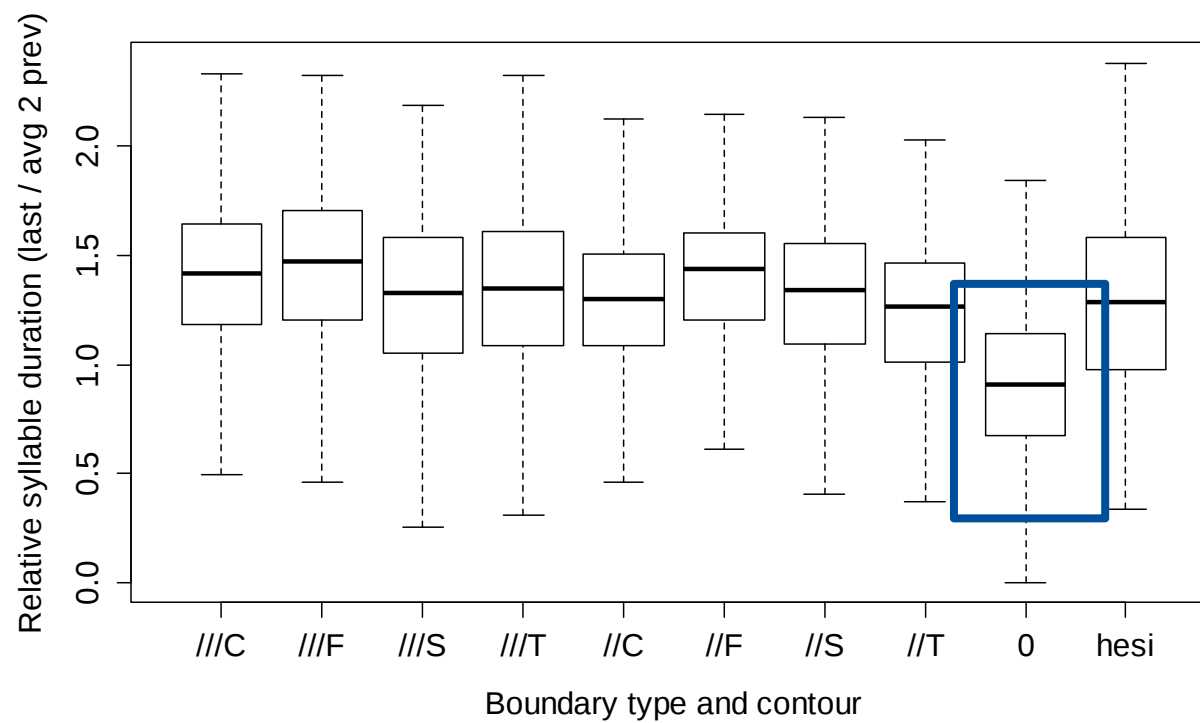
Durée de la pause subséquente (sec)



Hauteur relative (- 2 , + 2)



Durée relative (- 2 , + 2)





Mapping des deux annotations

- annotations réalisées sous Praat
- 2 tires d'annotation syntaxique + 2 tires d'annotation prosodique
- 1 script génère automatiquement une tire BDU
- ces analyses sont disponibles sous différents formats
 - TextGrid multi-tires
 - version textuelle annotée
 - fichier tableur (XLS)

La méthode pour l'identification des BDU

- la mise en relation de deux niveaux d'annotation réalisés de manière indépendante génère un niveau d'organisation spécifique (discursif)
 - syntaxe: unités de rection maximale
 - prosodie: frontières prosodiques majeures > unités prosodiques
- une BDU émerge chaque fois qu'une fin d'unité syntaxique coïncide avec une frontière prosodique majeure
 - "The talk unit is the maximal unit defined by syntax plus intonation" (Halford 1996: 33)
- aussi longtemps que les deux niveaux ne coïncident pas, le locuteur projette une suite, et l'auditeur attend une forme de complétion (Selting 2000)
 - excepté les unités inachevées / interrompues (Van Damme, Eychenne, Degand, Simon, à paraître)

La méthode pour l'identification des BDU

UR	[]	[] []	[]	<a, md>	[] [] []
UP	[]	[]	[] [] []	[]	[] []
BDU	[bdu-c]	[bdu-i]	[bdu-s]	[bdu-a]	[bdu-x]

(Degand & Simon 2009a)

- **BDU-c congruente**: une unité de rection correspond à une unité prosodique majeure
- **BDU-s groupée par la syntaxe**: plusieurs unités prosodiques majeures sont groupées dans une unité de rection
- **BDU-i groupée par l'intonation**: plusieurs unités de rection sont regroupées dans une unité prosodique majeure
- **BDU-a adjoint**: un élément non régi est autonomisé dans une unité prosodique majeure
- **BDU-x mixtes**



Exemples des types de BDU

BDU-c congruente

[(dans la majorité politique libanaise) (beaucoup de voix) (s'étaient élevées) (contre sa venue)]^{urv} ///C

BDU-s groupée par la syntaxe

[(pourquoi) (ce rôle majeur) (n'est-il pas dévolu) ///S (à la religion)]^{urv+} ///C

BDU-i groupée par l'intonation

<euh> [(on est devenu) (bien potes)]^{urv} [(tout le monde) (se connaissait)]^{urv} ///C

BDU-a régulative

<mais> <bon> ///T

BDU-x mixte

<donc>md euh [(je veux dire)SV]^{urv} euh [(l'employeur)SS (a)SV (le plus la possibilité de
///S de choisir vraiment euh la personne qu'il lui faut)SO]^{urv} <et>md ///S



Exemple

BDU #4 : **bdu-4-c_C**

Type d'EPG : **PGS-R**

NBUR : 1 NBSEQ : 2 NBUI : 1 NBtokens : 16 xmin : 26.079547537178513 xmax : 29.072538806818336

<en tout cas>_{ag} [(moi)_{SPron} //C (parce que j'avais choisi de vivre dans une famille d'accueil)_{SRd}]^{ure} ///C

BDU #5 : **bdu-5-c_C**

Type d'EPG : **PGS-R**

NBUR : 1 NBSEQ : 2 NBUI : 1 NBtokens : 5 xmin : 29.597453582079524 xmax : 30.749533289524585

<et>_{md} [(Pierre)_{SS} (à l'école)_{SO}]^{ure} ///C

BDU #6 : **bdu-6-i_C**

Type d'EPG : **PGC-R**

NBUR : 6 NBSEQ : 12 NBUI : 1 NBtokens : 57 xmin : 31.635038066540222 xmax : 42.95841240900619

<et>_{md} <donc>_{md} [(ils ont)_{SV} (d'ab/ d'abord)_{SRd} (déposé)_{SV-S} (Pierre)_{SO}]^{urv} //C <et>_{md} <puis>_{md} [(ils m'ont déposée)_{SV}]^{urv} <moi>_{ad} euh [(c'était)_{SV} (un taxi //C comme ça)_{SO}]^{urv} euh en fait>_{md} <le mec qui qui est venu nous chercher à l'aéroport>_{ag} //C [(il nous a juste mis)_{SV} (dans le taxi puis>_{md} <lui>_{ag} [(il est parti euh (je sais pas où)_{SO}]^{urv} //C euh [(en ville)_{SO}]^{ure} //C <ou quoi>_{md} ///C

BDU #7 : **bdu-7-i-int_S**

Type d'EPG : **PGC-R**

NBUR : 5 NBSEQ : 9 NBUI : 1 NBtokens : 27 xmin : 43.533778841014545 xmax : 49.9610316878724

<et>_{md} <puis>_{md} [(on était tous seuls)_{SV} (dans ce taxi)_{SO}]^{urv} //C SV]^{urv} //C [(je voyais)_{SV} (rien)_{SO}]^{urv} //C [(c'était)_{SV} (un une arrivée //C assez bizarre)_{SO} //C (je dois dire)_{insert}]^{urv+} //T [(parce que)_{ad-I}]^{ad-I} ///S



Plan de l'exposé

1. Le modèle des *Basic Discourse Units*
2. Le corpus LOCAS-F (*Louvain Corpus of Annotated Speech – French*)
 - Un corpus multigenres
 - Annotations syntaxiques et prosodiques
 - Aperçu plus détaillé des unités prosodiques
3. **Évidences empiriques**
 - **Distribution des BDU selon les genres de discours**
 - **Types et fonctions des éléments en périphérie gauche**
 - **Fonctionnement des marqueurs de discours (DM)**
4. Conclusions



3.1. Distribution des types de BDU selon les genres de discours

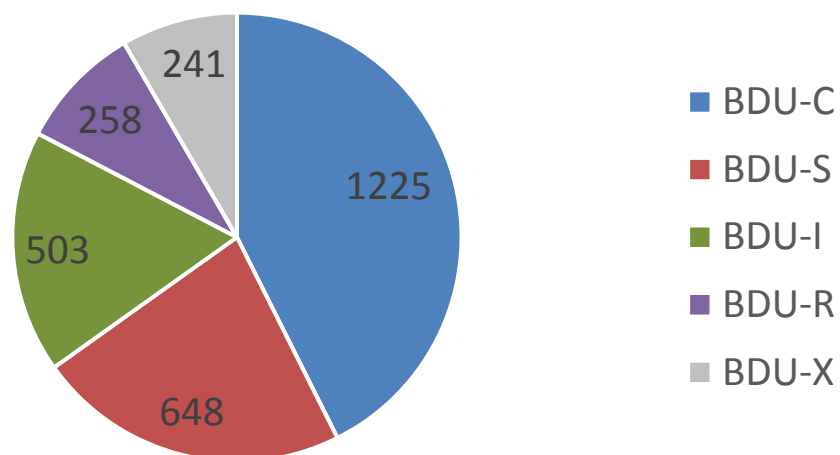
Hypothèses

- les BDU de type *congruent* (syntaxe = prosodie) ne constituent pas la « norme »
- en fonction des paramètres de la situation de production, les locuteurs utilisent des stratégies différentes pour construire leurs unités de discours
- par conséquent, les types de BDU se distribuent de manière significativement différente selon les genres / paramètres situationnels

Distribution des BDU dans l'ensemble du corpus LOCAS-F

Types de BDU	Nombre	Pourcentage
BDU-C	1225	43%
BDU-S	648	23%
BDU-I	503	17%
BDU-R	258	9%
BDU-X	241	8%
Total général	2875	100%

Nombre de TYPE BDU

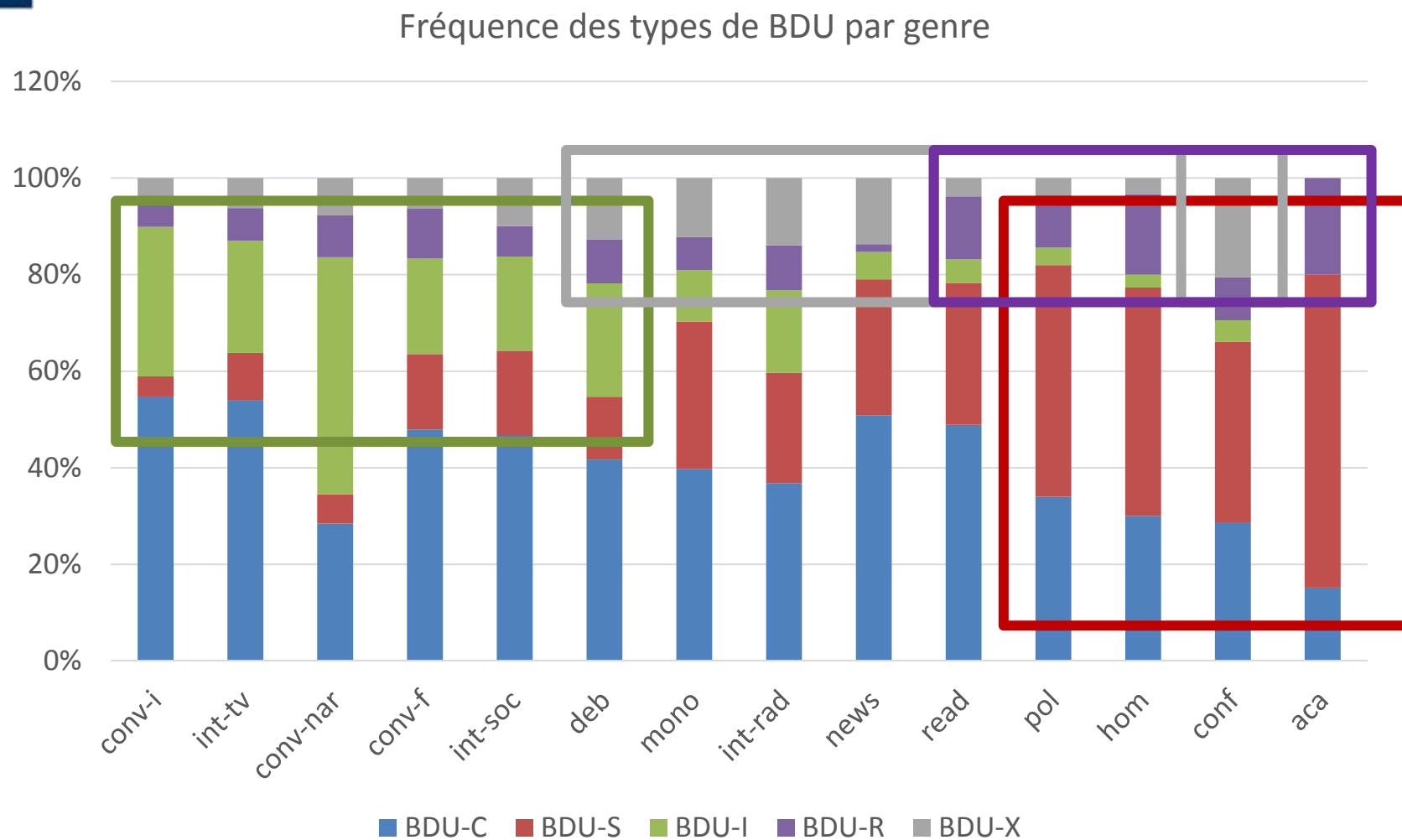


Distribution des types de BDU selon les genres de discours

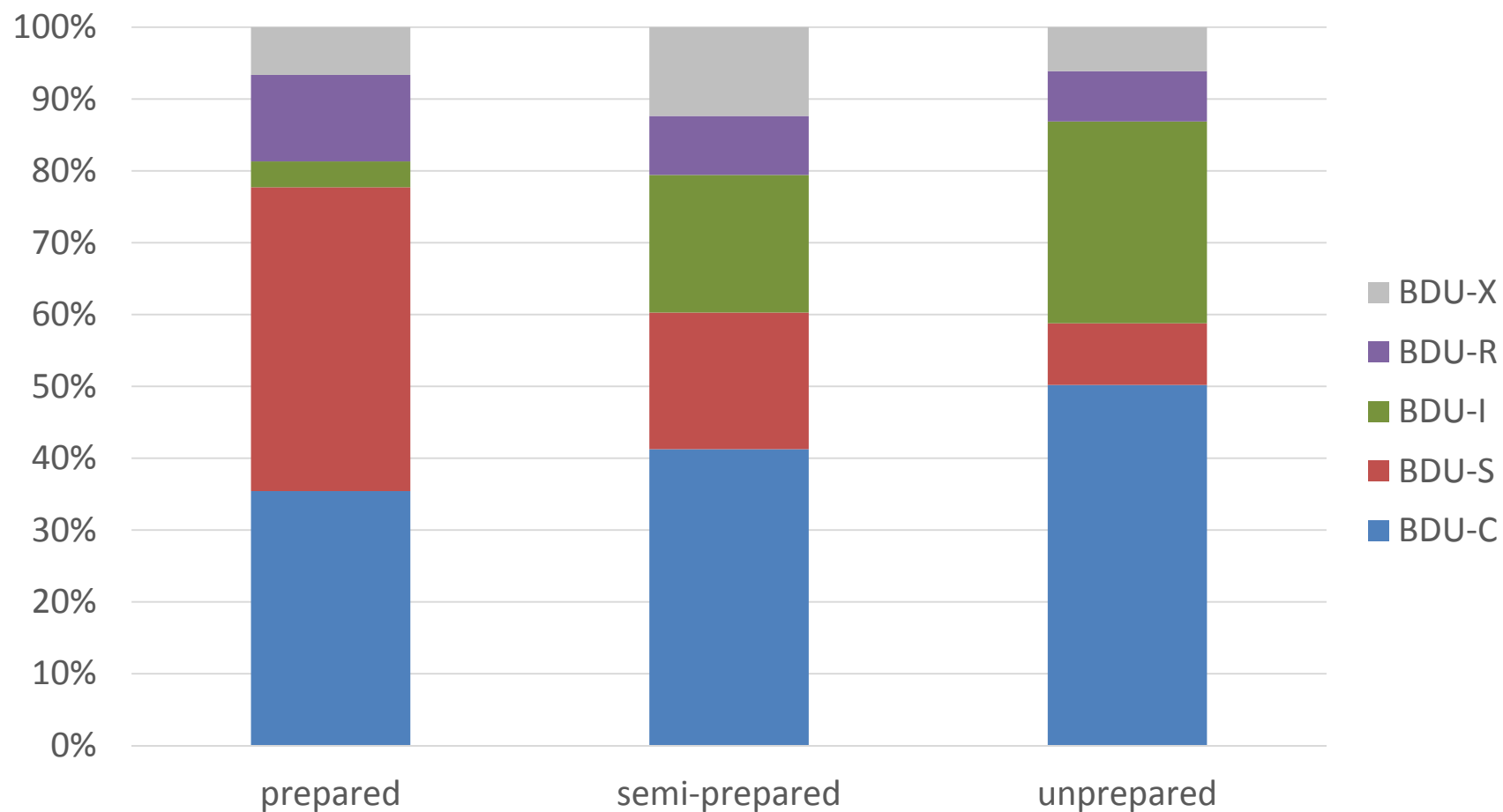
	Types de BDU (nombre)					
Genres	BDU-C	BDU-I	BDU-R	BDU-S	BDU-X	Total général
int-tv	200	86	25	36	23	370
conv-i	184	104	16	14	18	336
deb	151	85	33	47	46	362
int-soc	103	43	14	39	22	221
int-rad	95	44	24	59	36	258
conv-f	92	38	20	30	12	192
read	90	9	24	54	7	184
pol	66	7	21	93	7	194
news	63	7	2	35	17	124
mono	52	14	9	40	16	131
hom	45	4	25	71	5	150
conv-nar	33	57	10	7	9	116
conf	32	5	10	42	23	112
aca	19	0	25	81	0	125
Total général	1225	503	258	648	241	2875

X-squared = 747.0443, df = 52, p-value < 0.0001

Distribution des BDUs selon les genres (pourcentages)



Distribution des types de BDU en fonction du degré de préparation



X-squared = 478.5868, df = 8, p-value < 0,001



Distribution des types de BDU selon les genres de discours

Hypothèses

- ✓ les BDU de type congruent ne constituent pas la « norme »

en fonction des paramètres de la situation de production, les locuteurs utilisent des stratégies différentes pour présenter l'information

- ✓ par conséquent, les types de BDU se distribuent de manière significativement différente selon les genres / paramètres situationnels

Commentaires

- comparaison des statistiques (chi-carré) de distribution / genre
 - BDU
 - unités syntaxiques
 - unités prosodiques, par genre,
- la segmentation en BDU est la plus spécifique (Cramer's V approx. 1,5 plus haut)

3.2. Types et fonctions des éléments en périphérie gauche (EPG)

- intérêt pour la « périphérie gauche » des BDU (Degand, Martin & Simon 2014, Martin, Degand & Simon 2014)
 - les BDU remplissent une fonction cognitive dans la planification et l'interprétation du discours (Degand et Simon 2009a:56)
- la PG constituerait une **position stratégique** pour le développement du discours
 - lieu où le message prend forme, point d'ancrage (Virtanen 2004:80-81)
 - emplacement où un changement de tour peut se produire, négociation de la structure de la conversation
 - le contenu doit être connecté de manière cohérente au contenu élaboré précédemment (Beeching & Detges, 2014; Degand, 2014)

Définition large de la périphérie gauche

- Caractéristiques formelles
 - PG définie respectivement aux BDU
 - PG définie en termes de détachement syntaxique et / ou prosodique
 - PG plus restrictive que la position initiale

Élément intégré prosodiquement	Élément régi syntaxiquement	
+	+	EI
–	–	EPG
+	–	EPG-s
–	+	EPG-p

Exemples

- EI [**(on ne fait)**_{SV} (qu'accroître le fossé /// qui sépare /// le commun /// populaire de ceux qui parlent en son nom)_{SO}]^{urv} [(ils se disent)_{SV}]^{urv} <mais>_{md} <enfin>_{SV} [(ils nous prennent)_{SV} (pour)_{SO-I}]^{urv-I} <euh> [(je vais pas vous dire)_{SV} (le mot)_{SO}]^{urv} /// (INT)
- EPG-s [**(surtout)**_{SAdv} >^a [(ce que je voudrais vous dire /// c'est)_{SV} (qu'il y a eu une maturation politique incroyable du peuple français /// à travers ces événements)_{SO}]^{urv+} /// (INT)
- EPG-p [**(l'évolution)** /// entre prix et salaires)_{SS} /// (est couplée)_{SV} (en Belgique)_{SRd}]^{urv} /// (DIS)
- EPG <**bon**>_{md} /// <donc>_{md} <euh> [(non)_{SAdv}]^{ura} [(il s'agit pas)_{SV} (de ça)_{SO}]^{urv} /// (INT)



Hypothèse - EPG

Hypothèses relatives aux fonctions des EPG

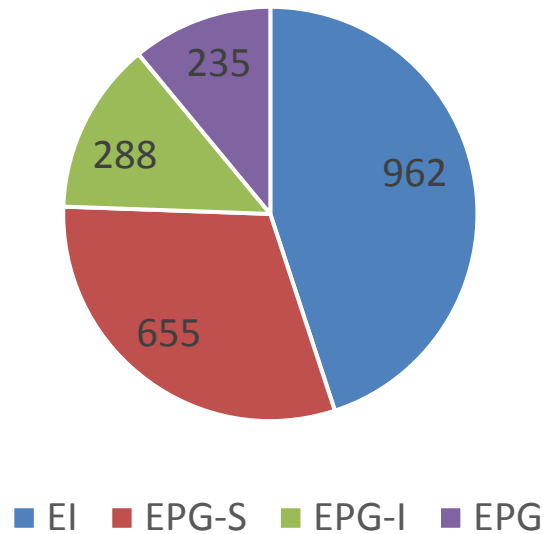
- les EPG résultent de stratégies discursives, et leurs **formes / fonctions** varient en fonction de la situation communicative

Fonctions

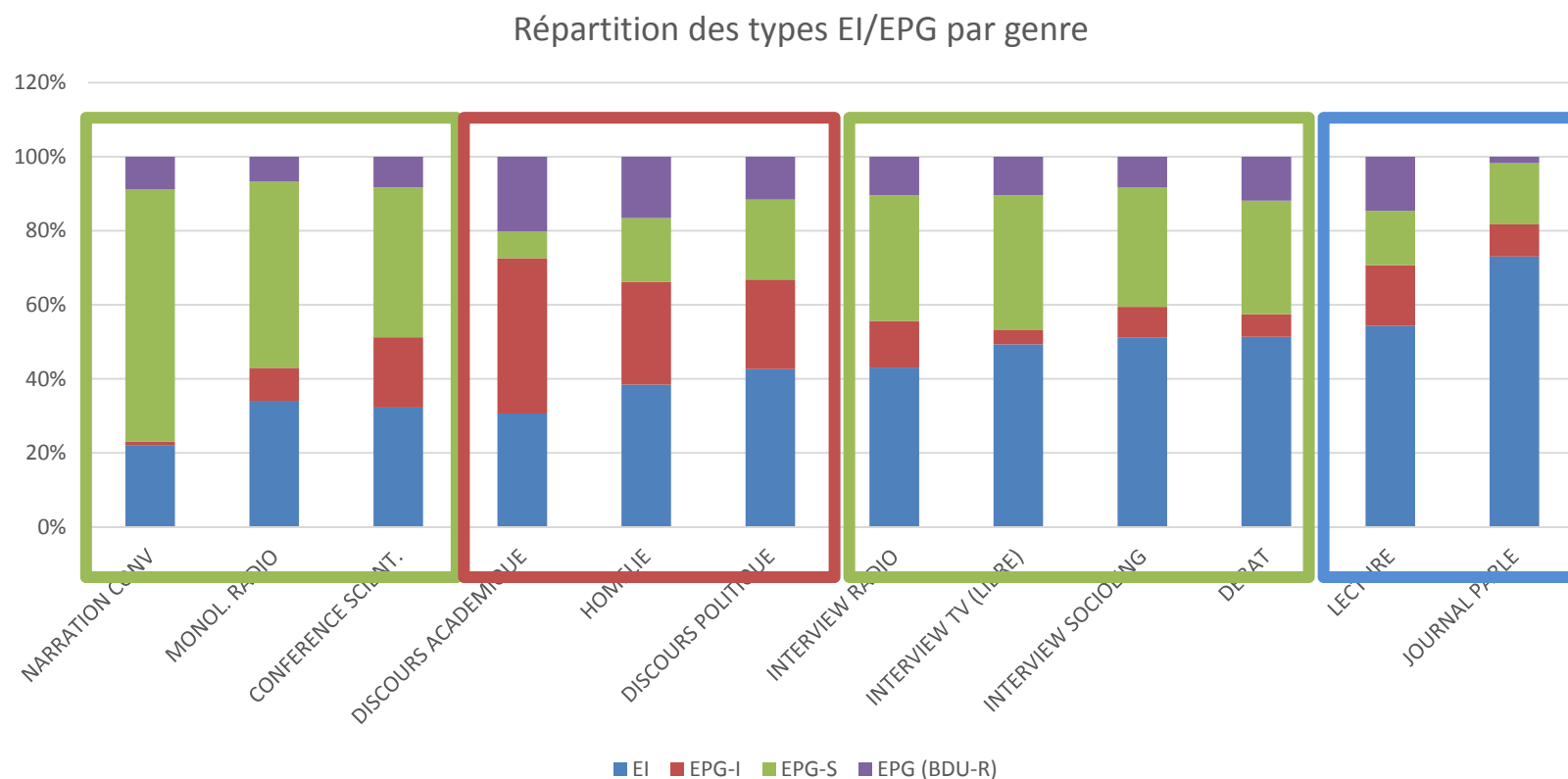
- annotation en double aveugle des fonctions discursives des EPG
- catégories fonctionnelles inspirées de Chafe (1994:64):
 - textuelle (relation de cohérence; gestion topicale; métadiscours)
 - interactive (tour de parole; interpellation)
 - subjective (point de vue; deixis)

Distribution des **formes** (EI vs EPG) dans le corpus

Types de périphérie gauche dans LOCAS



Distribution des **formes** (EI vs EPG) selon les situations



Annotation des **fonctions** (double codage: Kappa 0.83)

- **Textuelle**

- **Logique:** <mais>_{md} [(on ne voyait rien)_{sv}]^{urv} ///
- **Gestion topicale :** <ce génie national>_{ag} [(on le trouve)_{sv} (principalement)_{SRd} (dans les traditions ///_{so} qui doivent capter et formuler ce génie national)_{so}]^{urv} ///
- **Métadiscours :** [(ce sera le premier point de de mon exposé qui en port/ comportera deux)_{insert} ///_{so} (je voudrais essayer de réfléchir)_{sv} (à aux raisons pour lesquelles ///_{so} le facteur linguistique semble jouer ///_{so} une espèce de de de rôle primordial dans la constitution des identités)_{so}]^{urv+} ///

- **Interactive**

- **Interpellation :** <alors>_{md} <vous>_{ag} [(vous êtes plus doux vous êtes plus doux que moi)_{sv}]^{urv} ///
- **Turn-taking :** <en fait>_{md} [(j'étais)_{sv} (un enfant ///_{so} très très malheureux ///_{so} euh très renfermé et tout ça)_{so}]^{urv} ///

- **Subjective**

- **Point de vue:** <honnêtement>_{ag} ///_{so} [(je crois)_{sv} (que c'est pour faire bon garçon respectable)_{so}]^{urv} ///
- **Deixis:** <moi>_{ag} [(j'aime bien aussi)_{sv} ///_{so} (lui donner l'impression qu'il est libre)_{so}]^{urv} ///

Distributions des **fonctions** dans LOCAS-F

Fonction	Categorie	Fréquence
Textuel		625 (79.2 %)
	Cohérence	402
	Gestion topicale	197
	Métadiscours	26
Interactif		88 (11.2 %)
	Turn-taking	53
	Interpellation	33
	Backchannel	2
Subjectif		76 (9.6 %)
	Point of view	53
	Deixis	23

Distribution selon le degré d'interactivité

Fonction	Catégorie	Interactivité 0	Interactivité 1	Interactivité 2
Textuel	Cohérence	304	89	104
	Topic	150	30	35
	Métadiscours	17	8	15
Interactif	Turn-taking	0	13	49
	Interpellation	19	5	15
	Backchannel	0	0	4
Subjectif	Point de vue	22	12	30
	Deixis	10	6	8

Corrélations forme – fonction

	EPG	EPG-S	EPG-I
Textuel	127	332	166
Interactif	41	43	4
Subjectif	14	38	24

$X^2 = 41.1$; $df = 4$; $p < .0001$ (C's V: 0.16) 

$X^2 = 252$; $df = 12$; $p < .0001$ (C's V: 0.40) 

	EPG	EPG-S	EPG-I
Cohérence	80	273	49
Topic	42	40	115
Métadiscours	5	19	2
Turn-taking	16	37	0
Interpellation	23	6	4
Backchannel	2	0	0
Point de vue	14	27	12
Deixis	0	11	12



Conclusion sur la périphérie gauche

- la définition de la périphérie gauche en termes syntaxiques et prosodiques constitue une extension valide de la notion de périphérie gauche
- les EPG servent premièrement les fonctions textuelles
 - EPG syntaxiques: spécialisation comme marqueurs de la cohérence discursive
 - EPG intonatifs: spécialisation pour le marquage des topiques, particulièrement dans le discours non interactif



3.3. Fonctionnement des marqueurs de discours

Données

- 1357 MD annotés dans le corpus LOCAS-F (incluant des cooccurrences de MD)

Questions de recherche (en cours)

- les MD sont-ils majoritairement en position initiale?
- y a-t-il un fonctionnement spécifique des MD en fonction des critères situationnels
 - degré d'interactivité
 - degré de préparation
 - caractère médiatique

Forte présomption théorique pour la **position initiale** des MD

It has to be commonly used in initial position of an utterance (Schiffrin 1987)

DMs “prototypically introduce the discourse segments they mark” (Hansen 1997)

“most items considered DMs are at least possible in initial position, and many occur there predominantly” (Schourup 1999)

Almost all DMs occur in initial position ... , fewer occur in medial position and still fewer in final position (Fraser 1999: 938)

Forte présomption théorique pour la **position initiale** des MD

It has to be commonly used in initial position of an utterance (Schiffrin 1987)

DMs “prototypically introduce the discourse segments they mark” (Hansen 1997)

“most items considered DMs are at least possible in initial position, and many occur there predominantly” (Schourup 1999)

Almost all DMs occur in initial position ... , fewer occur in medial position and **still fewer in final position** (Fraser 1999: 938)

La position des MD en fonction des unités

Position	Initiale	Médiane	Finale	Isolée
BDU	660	459	104	138
unité de rection	1025	103	83	138
unité prosodique	538	617	199	*



Conclusions: unités de discours et corpus

small is beautiful

... when annotated

- un corpus multi-annoté permet de poser différemment des questions « classiques »
- un corpus multi-genre permet de déceler des fonctionnements spécifiques à certains paramètres situationnels
- possibilité (presque infinie) d'enrichir les annotations (! circularité)



Quelques publications en lien avec la présentation

Accessibles depuis: <http://www.uclouvain.be/anne-catherine.simon>

- Martin, L.; Degand, L.; Simon, A.C. 2014. Forme et fonction de la périphérie gauche dans un corpus oral multigenres annoté. *Corpus*, 13, 243-265
- Degand, L.; Martin, L.; Simon, A.C. 2014. LOCAS-F : un corpus oral multigenres annoté. *CMLF 2014 - 4 ème Congrès Mondial de Linguistique Française*, EDP Sciences: Berlin, 2613-2626.
- Tanguy, N.; Van Damme, Th.; Degand, L.; Simon, A.C. 2012. *Projet FRFC " Périphérie gauche des unités de discours " - Protocole de codage syntaxique*, 17 p.
- Degand, L.; Simon, A.C. 2009. *On identifying basic discourse units in speech: theoretical and empirical issues*. *Discours : revue de linguistique, psycholinguistique et informatique*, Vol. 4, [en ligne]
- Degand, L.; Simon, A.C. 2009. Mapping prosody and syntax as discourse strategies: How Basic Discourse Units vary across genres. In: A. Wichmann, D. Barth-Weingarten, N. Dehé (eds), *Where prosody meets pragmatics: research at the interface*, Emerald: Bingley, 79-105.



Merci de votre attention!

UCL

**Université
catholique
de Louvain**
