

ben perso j' ai déjà es~ j' ai déjà essayé deux trois fois

Les amorces de mots en français parlé : de la transcription à l'analyse

VERSION NON DEFINITIVE

Résumé

Dans cet article, nous analysons 300 amorces issues du corpus *Les Vocaux*. Dans un premier temps, notre réflexion porte sur les choix de transcription et les effets qu'ils peuvent avoir sur le phénomène que nous étudions. Nous analysons ensuite nos données, en les répartissant en trois grandes catégories selon que l'élément amorcé est complété, modifié ou abandonné au profit d'une autre construction. Nous nous intéressons à la longueur du fragment ainsi qu'à sa classe grammaticale. Nous verrons que les résultats que nous obtenons ne sont pas différents de ceux issus de l'étude d'autres types de données orales non planifiées.

Abstract

In this article, we analyse 300 sentence fragments from the *Les Vocaux* corpus. We begin by examining transcription choices and the effects they may have on the phenomenon under study. We then analyse our data, dividing it into three broad categories depending on whether the initiated element is completed, modified or abandoned in favour of another construction. We examine the length of the fragment as well as its grammatical category. We shall see that the results we obtain are no different from those derived from the study of other types of unplanned spoken data.

1. Introduction

C'est une banalité aujourd'hui de dire que l'étude de la langue parlée ne peut se faire que via le passage par l'écrit, et le paradoxe a été souligné depuis longtemps par de nombreux chercheurs (Blanche-Benveniste et Jeanjean 1987, Gadet 2003, Raingeard et Lorscheider 1977). Comme le précisent Blanche-Benveniste et Sabio (2023 : 24) :

On ne peut pas étudier l'oral par l'oral, en se fiant à la mémoire qu'on en garde. On ne peut pas, sans le secours de la représentation visuelle, parcourir l'oral en tous sens et en comparer les morceaux.

On sait aussi que les choix de transcription engagent souvent les analyses qui ensuite seront faites de certains phénomènes, et que cette étape est donc cruciale dans la constitution des données, qui sont construites dès cette première étape (Mondada 2000). C'est notamment le cas des amorces de mots que nous allons étudier.

Nous définissons les amorces de mots comme des auto-interruptions par le locuteur d'un lexème en cours d'énonciation. L'exemple suivant est un cas typique d'amorce. Le lexème interrompu – noté dans *Les Vocaux* par un tilde collé directement à la droite de celui-ci – est complété ensuite dans l'énoncé, où il est repris sous sa forme pleine :

- (1) enfin pour moi sans **sans re~ sans respect** il y a pas de sport en fait
(61_74, F, 23, B)

Le transcripteur¹ a opté pour la graphie *re~* pour transcrire la suite sonore /ɛ/. Les graphies *rè~*, *rê~*, *rai~* ou *rei~* auraient aussi pu être possibles car conformes à la correspondance phonie/graphie du français (comme dans *règne*, *rêne*, *raison*, *reine*). Si le transcripteur choisit *re~*, c'est parce qu'il l'analyse comme l'anticipation de *respect*.

Dans cet article, nous nous pencherons d'abord sur la question de la transcription des amorces, après avoir replacé celles-ci dans le contexte général des phénomènes disfluents. Nous analyserons enfin un sous-corpus constitué de 300 amorces issu du corpus *Les Vocaux*.

2. Les disfluences, mode de production de l'oral

Les études sur la langue parlée ont permis de mettre en évidence une série de phénomènes propres à l'oral qu'on regroupe traditionnellement sous l'appellation générique de *disfluences*. On fait ainsi référence à un certain nombre de marques du discours en cours d'élaboration, d'« achoppements » dans la linéarité de l'énoncé. Ces phénomènes sont inhérents à la production de la langue parlée non planifiée et manifestent des « ratés dans la planification » de l'énoncé (Cappeau 1998 : 307).

Maclay et Osgood (1959), dans une étude pionnière sur la question, utilisent le terme *hesitation phenomena* (et non *disturbances* ou *disfluencies*). Ils ont établi une typologie des phénomènes dits d'*hésitation* qui a fréquemment été reprise dans les études postérieures. Selon eux, l' hésitation se manifeste dans la parole des locuteurs par les pauses remplies (qui correspondent en français au *euh*), les syllabes allongées, les faux départs (repris ou non), les répétitions (non sémantiques) et les pauses silencieuses non syntaxiques.

¹ Nous utilisons *transcripteur* au masculin singulier générique. Il va de soi qu'il peut s'agir d'une transcriptrice, mais aussi de plusieurs transcripteurs qui se sont succédé. Les transcriptions sont en général le fait de plusieurs étapes, et seule la version finale est disponible, qui ne tient plus compte différents choix qui ont pu avoir lieu.

Ainsi, l'énoncé suivant comprend, dans la terminologie de Maclay et Osgood, 4 faux départs (ç- / qui a / il a / de qu-), 2 pauses remplies (euh), 2 répétitions (il a il a / des des).

- (2) **ç- ça** dépend quand même enfin **m-** j'ai mon petit frère quand même **qui a il a il a** cinq ans maintenant et **euh** lui c'est vrai que il a récupéré mes jouets **de qu-** enfin des voitures **des des** petits personnages (Matthieu, CFPB-1000-1)

Initialement, ce sont des psycholinguistes (Maclay et Osgood 1959, Levelt 1989), des psychocliniciens ou encore des psychanalystes (on sait l'intérêt de Freud pour le lapsus, où l'attention est portée davantage sur le lexique) qui ont étudié ces différents phénomènes. C'est en effet l'aspect « raté » qui a d'abord intéressé les chercheurs, ainsi que les mécanismes d'encodage et de décodage des énoncés (cf. Duez 2001 pour un historique de ces notions). Aujourd'hui encore, le colloque DISS (*Disfluency in spontaneous speech*), qui se déroule tous les deux ans depuis 1999, regroupe des spécialistes de ces différents domaines de recherche, ainsi que des linguistes et des spécialistes du traitement automatique des langues, notamment à cause des défis que posent les éléments disfluents aux technologies de la parole.

L'intérêt plus tardif des linguistes pour ces phénomènes s'explique sans doute par le fait que les premières recherches sur l'oral se sont concentrées sur les aspects phonétiques ou phonologiques, à travers de la parole de laboratoire, parole planifiée et donc moins soumise aux disfluences.

La littérature sur le sujet manque de terme neutre pour parler de ces phénomènes courants. Si nous utilisons le terme général de *disfluente*², le préférant à celui d'*hésitation*, c'est que les phénomènes qu'il recouvre sont à ce point fréquents que l'on ne peut penser que les locuteurs sont constamment en train d'hésiter. *Disfluente* s'oppose à *fluence*, dans le sens où la linéarité de l'énoncé est à un moment suspendue.

En effet, au lieu de la disfluence, le déroulement linéaire de l'énoncé est brisé et l'on assiste à un piétinement sur l'axe syntagmatique (Blanche-Benveniste *et al.* 1990). Dans sa thèse consacrée au sujet, Shriberg (1994 : 7-9), à la suite notamment de Levelt (1983), a modélisé la séquence disfluente en la décomposant en quatre éléments distincts, qui correspondent à trois régions :

² La graphie *disfluente* a été préférée par les chercheurs à celle de *dysfluente*, réservée au domaine des pathologies du langage (Pallaud 2004).

- **reparandum** : le *reparandum* (RM) est la partie produite par le locuteur qui ne sera pas conservée et sera remplacée ultérieurement au profit du *repair* ;
- **interrupting point** : le *point d'interruption* (IP) est le moment de l'énoncé qui coïncide avec la fin du *reparandum*. Ce point d'interruption est vide ;
- **interregnum**³ : l'*interregnum* (IM) est la région qui commence à la fin du *reparandum* et s'achève au début du *repair*. L'*interregnum* peut ou non contenir un terme d'édition (*editing term*), c'est-à-dire une pause silencieuse, une pause remplie, etc., ou plusieurs autres tentatives de formulation inachevées également ;
- **repair** : le *repair* (RR) indique le retour à la « fluence », par la réparation, la correction du *reparandum*.

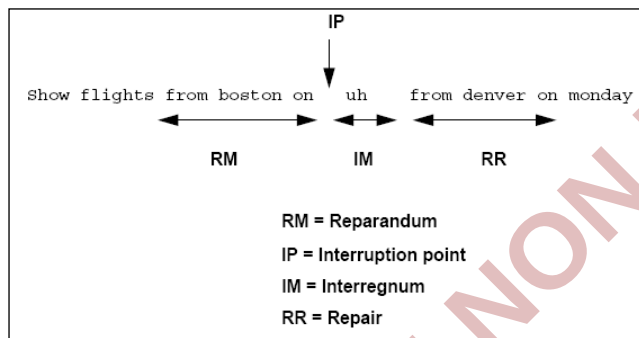


Figure 1 : Modélisation de la séquence disfluente (Shriberg 1994 : 8)

L'amorce est l'élément sur lequel nous centrons notre attention dans cet article ; elle appartient au *reparandum*. Signalons néanmoins que l'empan de celui-ci est souvent plus grand que le seul mot amorcé, et que ce n'est qu'en analysant le *repair* que l'on peut déterminer les éléments compris dans le *reparandum*.

Ainsi, les deux cas suivants :

- (3) du coup c' était le weekend et le weekend **hon~ honnêtement** ... j'
avais pas envie de faire ça (07_10, F, 19, F)
- (4) ben perso **j' ai déjà es~ j' ai déjà essayé** deux trois fois (07_09, F,
19, F)

³ L'*interregnum* correspond à l'*editing phase* de Levelt.

Dans l'énoncé (3), le *reparandum* est composé uniquement de l'amorce, et le *repair* est la forme entière de l'amorce. Par contre, dans (4), la locutrice, dans sa reprise de l'amorce qu'elle complète (cf. 6.1), revient au début du groupe syntaxique : c'est donc la séquence depuis le pronom sujet *j'* qui constitue le *repair*. Dans ces énoncés, *l'interregnum* est vide, mais le locuteur peut se livrer à des tentatives de reformulations, ou marquer sa recherche lexicale par certains éléments :

- (5) ouais // un petit problème de fisc pour le moment j'attends (rires)
 attention / non c'est / c'est pas vraiment ça c'est / c'est notre
 comptable / qui qui / qui a fait une euh / **une encéphalite à her/ à**
herpès une encéphalite herpétique (...) (Valibel, blaNB11)

Pour marquer visuellement cette stagnation en un point du déroulement de l'énoncé, la représentation en grilles développée par l'équipe du GARS (Blanche-Benveniste et al. 1979, Bilger 1982), le Groupe aixois de recherche en syntaxe, est particulièrement éclairante :

- (6) et c'était pas pour regarder le film
 c'était pour embra-
 embrasser les filles (Grand-Jojo, CFPP-1000-1)

Les grilles marquent visuellement le déroulement syntagmatique, représenté sur l'axe horizontal, et le déroulement paradigmatique, noté sur un axe vertical. La représentation en grilles note de manière superposée des segments qui occupent une même place syntaxique. Il est à noter que cette représentation n'est pas adoptée uniquement pour les séquences disfluences. Ainsi, par exemple, une suite d'appositions ou une énumération seront notées également à la verticale, pour visualiser l'effet de liste (Pietrandrea et Kahane 2013), qu'il s'agisse de disfluence ou non⁴.

3. L'amorce

L'amorce est le phénomène langagier qui consiste en l'interruption d'un lexème en cours d'énonciation. Elle participe d'un phénomène d'anticipation, marqué par des allées et venues sur l'axe syntagmatique :

⁴ Le GARS utilise la configuration en grilles notamment pour pallier les effets que produit sur l'interprétation des productions orales la linéarité de l'écrit traditionnel.

Ces anticipations montrent que le locuteur tient en réserve, lorsqu'il parle, les syntagmes qu'il vient de dire aussi bien que ceux qu'il projette de dire. (Blanche-Benveniste et al. 1990 : 23)

L'énoncé suivant, qui contient deux amorces, illustre bien ces allées et venues sur l'axe syntagmatique, où le locuteur anticipe une première fois l'adjectif *propre* et ensuite le nom *caractère*, avant de reprendre ces deux termes dans un syntagme complet :

- (7) oui pour moi tous tous les personnages qui viennent à l'encontre de la personnalité euh **pro**/ enfin des / du caractère de son **car**/ de son **propre caractère** c'est toujours beaucoup plus difficile à interpréter parce qu'on doit on doit / on doit sortir beaucoup plus de soi je veux dire (...) (Valibel, ilrLP1r)

D'emblée, on distinguera ce phénomène de l'apocope – qu'elle concerne des formes relativement standard comme *prof* ou *ciné*, ou des formes plus rares et plus individuelles comme *mayo(nnaise)* ou encore *pyj(ama)* –, où la troncation du mot est volontaire (cf. 4.3).

Jeanjean (1984) et Martinie (1999) désignent sous le terme d'*amorce* non seulement les amorces de lexèmes mais également les amorces de syntagmes, qui correspondent aux faux départs de Maclay et Osgood (1959). Ainsi, pour ces auteurs, un groupe syntaxique incomplet est qualifié d'amorce. Dans l'exemple (8), deux constructions sont laissées inachevées⁵ :

- (8) mais ma sœur pf elle a seize ans et euh elle est très petite dans sa tête tu vois donc elle sort euh vite fait elle va en ville mais euh pf avec ses amis mais le jour enfin elle sort jamais le soir ma sœur donc ça va **elle est pas ils sont pas euh** (Sonia, CFPB-1030-2)

Dans la littérature sur le sujet en français, on trouve également d'autres termes pour désigner l'amorce, telle que nous l'entendons ici, ou des phénomènes proches. Pour Cappeau (1998), les amorces de syntagmes sont désignées sous le terme de *bribes*, et relèvent plutôt de la répétition de déterminants. Bres et Gardes-Madray (1991 : 96 et sv.) parlent quant à eux de *tâtonnements syntagmatiques*. Martinie (1999) utilise les termes d'*énoncés réparés* et n'isole pas les amorces d'autres disfluences dans son analyse. Dans les études phonétiques pionnières de Grosjean et Deschamps (1972, 1973 et 1975), les amorces sont incluses dans les répétitions de syllabes. La terminologie n'est donc pas unifiée, ce qui rend la comparaison des résultats de différentes recherches souvent périlleuse. Il en est de même dans la terminologie anglaise, où les termes

⁵ Notons que l'inachèvement ne fait pas nécessairement partie de la définition de l'amorce chez ces auteurs.

self-repair et *self-correction* désignent tantôt l'amorce de mots, tantôt un processus plus large dans lequel l'amorce de mots n'est qu'une des réalisations possibles parmi d'autres.

4. Les choix de transcription

4.1. Transcrire, c'est interpréter

La transcription est loin d'être une activité neutre et transparente que pourrait faire un quelconque copiste (Blanche-Benveniste et Jeanjean 1987 ; Bilger 2000 : 77). Au contraire, le travail de transcription est sélectif et interprétatif :

Far from being an objective and exhaustive mirroring of events of an interaction, a transcript is fundamentally selective and interpretative.
(Edwards 1995 : 19)

Depuis l'article fondateur d'Ochs (1979), on s'accorde pour dire qu'elle engage, ou révèle, une théorie. La transcription, ainsi que l'établissement par le chercheur de conventions qui la sous-tendent, font l'objet d'une véritable réflexion dans laquelle apparaissent des enjeux théoriques⁶ :

(...) transcription is a selective process reflecting theoretical goals and definitions. (Ochs 1979 : 44)

Dans le cadre de transcriptions orthographiques faites pour des analyses linguistiques, cette correspondance entre les choix de transcription d'une part et les présupposés théoriques et les objectifs de la recherche d'autre part posent parfois un problème de circularité mis en évidence par Mondada (2000 : 133) :

De nombreuses interprétations de phénomènes se trouvent en effet incorporées dans la transcription a priori, alors qu'elles sont le but déclaré de l'analyse censée s'exercer a posteriori sur cette transcription.

Il faut donc être attentif au fait que la mise à plat opérée par la transcription tend à donner à celle-ci une apparente objectivité sur laquelle se base ensuite l'analyse.

Dans son activité cognitive de segmentation et de catégorisation, le transcripateur cherche à faire sens :

⁶ Outre à l'article de Ochs (1979), on se reportera, entre autres, à Blanche-Benveniste et Jeanjean (1987), Edwards (1993 et 1995), Gumperz et Berenz (1993), Mondada (2000).

Phonéticiens et psychologues s'accordent à dire que la perception est un processus actif qui s'appuie sur la compréhension des énoncés (Blanche-Benveniste et Jeanjean 1987 : 103).

L'homme a un rôle actif dans la perception des énoncés oraux. Il cherche à construire, voire à reconstruire du sens. Le transcripateur, dans ce travail de recherche de sens, part en général du principe que le locuteur tient des propos sensés, qu'il est cohérent, qu'il cherche lui-même à faire sens, etc. Ce crédit accordé au locuteur⁷ permet bien souvent de choisir entre des interprétations concurrentes, et ainsi, de ne pas multiplier les transcriptions en parallèle, appelées aussi *multitranscriptions*, utilisées dans certains corpus (cf. DELIC 2004, Raingeard et Lorscheider 1977).

Ainsi, devant une suite sonore homophone, c'est après une analyse, qui explore notamment le contexte et garantit une cohérence textuelle, que le transcripateur sélectionnera une solution parmi les multiples possibilités qui théoriquement s'offraient à lui. Les amorces ne font pas exception. Le transcripateur fait des hypothèses sur le fragment pour que sa transcription orthographique soit cohérente avec le mot entier qui n'est qu'amorcé. Cette transcription n'est donc pas que phonétique, elle reconstruit la forme pleine et, ce faisant, est déjà une analyse.

L'interprétation de l'amorce, et donc du choix de sa graphie, dépendent de plusieurs facteurs : le cotexte, les connaissances du transcripateur – notamment de la fréquence du lexique et des champs lexicaux (Roubaud 2004) – la variété de langue du locuteur, le fait que l'amorce soit reprise sous sa forme pleine dans un contexte proche, mais aussi la longueur du fragment.

Dans les énoncés suivants, le cotexte guide le transcripateur :

(9) je sais plus qu' est -ce qu' il disait le mec au sujet de cette société du divertissement **homo mediaticus** peut-être je je sais plu~ aussi ouais ça c~ ça doit être ça **ho~ homo** mediaticus ça me dit quelque chose enfin bref (437_27, H, 26, F)

(10) elle elle expliquait aussi que que enfin elle avait toujours des très beaux des très beaux cheveux trop bien entretenus elle expliquait justement que ça prenait des heures et des heures et que enfin voilà c'était **ho~ affreux** (433_55, H, 24, F)

⁷ Pallaud (2002 : 80-81) pointe le fait que les hypothèses de cohérence et de congruence ont pour effet de sous-estimer les phénomènes de lapsus.

Dans (10), *ho~* est l'anticipation de *homo*, repris directement sous sa forme pleine mais qui vient également d'être énoncé. La transcription semble évidente et la graphie du fragment ne laisse guère de doute. Dans (11), le transcripateur fait l'hypothèse qu'il s'agit de l'amorce de *horrible*, corrigé en *affreux*. C'est le sens global de l'énoncé et la connaissance du champ lexical qui autorisent cette supposition, et transcrire *ho~* semble plus cohérent, du point de vue orthographique, que le seul *o~*. Par contre, le choix posé pour l'énoncé (11) nous paraît beaucoup moins évident.

- (11) mon père écoute mon père est **ho~** ... a ... subi une première opération
(01_02, F, 41, F)

Le *repair* est *a subi une première opération*. Le champ lexical est celui de l'hôpital et la place syntaxique amorcée peut être interprétée comme celle d'un adjectif ou d'un participe passé. Choisir de transcrire *ho~* oriente l'interprétation du lecteur vers l'ébauche de *hospitalisé*. Mais il est tout aussi possible qu'il s'agisse de l'amorce de *opéré*, le lexème *opération* apparaissant dans le contexte droit du *repair*. Dans ce cas, comment trancher ? Dans de tels cas, la perception du transcripateur peut ne pas être la même que celle du locuteur ni celle du lecteur qu'est le chercheur. Ainsi, dans de nombreux exemples, « si [l']interprétation est hautement probable, elle ne peut être posée avec certitude. » (Pallaud 2003 : 69)⁸

4.2. La longueur du fragment

Plus le mot amorcé est long, plus son interprétation et donc le choix de la graphie est aisé. Ainsi, un fragment de 4 syllabes se laisse aisément interpréter, même quand celui-ci n'est pas complété dans la suite :

- (12) imaginons que je rate et je vais me dire putain je suis trop
conne j'aurais dû prendre des heures d'auto-école ... mais en même
temps si je réussis sans heures d'auto-école je vais me dire ah ben
c' est cool j' ai **économi~** enfin après c' est pas pour l' argent que ça
coute euh surtout que je vais pas prendre beaucoup d' heures je vais
prendre que quatre heures ou je sais pas combien mais pas beaucoup
quoi (105_24, F, 22, B)

⁸ Pour les amorces, le guide de transcription de *Les Vaux* indique : « On transcrit le phonème entendu suivi d'un tilde (~) : j'av~ ; l'amorce choisie est un mot existant dans la langue, correspondant a priori à l'intention ; en cas de doute, on a opté dans tous les cas pour un mot existant (ces t~/ c'êt~ (= c'était) vs sêt~) »
Voir le guide complet : https://hdl.handle.net/11403/lesvoux/v0.0.2/corpus_voux_v0.0.2_presentation.pdf

Mais les amorces aussi longues sont exceptionnelles. Notre corpus de 300 occurrences d'amorces n'en comporte qu'une seule. Les amorces de plus d'une syllabe sont en fait très minoritaires, comme le montre le tableau suivant :

Longueur	n=300
1 consonne	171 (57 %)
1 voyelle	20 (6,6 %)
1 syllabe	90 (30 %)
2 syllabes	16 (5,3 %)
3 syllabes	2 (0,6 %)
4 syllabes	1 (0,3 %)

Tableau 1 : Longueur du fragment amorcé (n=300)

On le voit, plus de la moitié des fragments (57 %) sont constitués d'une seule consonne. Quand on additionne les amorces d'une seule consonne et d'une seule voyelle à celles d'une seule syllabe, on constate que ce sont 93,7 % des amorces qui sont monosyllabiques. On verra que ce sont ces fragments très brefs qui sont le plus difficiles à interpréter. Au niveau de la transcription, ces amorces permettent de relever certaines incohérences dans les choix posés par les transpositeurs, comme on le voit dans ces deux exemples, où le *ça* amorcé aurait dû avoir la même graphie :

(13) et puis bon de base elle a une personnalité très intense mais euh un peu un peu violente dans le type un peu un peu bon *ça s~ s~ ça* vient du fait qu' il il y a il y a plein de trucs entre entre sa petite sœur qui a toute l' attention et sa grande sœur qui avait toutes les louanges (437_27, H, 26, F)

(14) je sais plus qu' est-ce qu' il disait le mec au sujet de cette société du divertissement homo mediaticus peut-être je je sais plu~ aussi ouais *ça c~ ça* doit être ça ho~ homo mediaticus ça me dit quelque chose enfin bref (<437_27>, H, 26, F)

L'établissement de conventions de transcription n'est pas toujours simple, et ce type de cas, qui peut sembler anecdotique, est difficile à anticiper sans vue globale des phénomènes. Les incohérences n'apparaissent souvent que lors de relevés plus systématiques, comme c'est le cas dans cette étude.

4.3.L'amorce vs l'apocope

L'apocope est un phénomène de troncation comparable à celui de l'amorce, si ce n'est qu'il ne participe pas d'une séquence disfluente. En effet, l'apocope ne relève pas d'un mécanisme d'auto-interruption : le lexème apocopé est planifié comme tel par le locuteur. Certaines apocopes sont d'ailleurs largement répandues, au point de concurrencer la forme pleine, et parfois même de s'y substituer dans l'usage : *kilo* pour *kilogramme* ; *ciné* ou *cinéma* pour *cinématographe* (Cerquiglini 2019 ; Dister 1995). C'est le cas dans les énoncés suivants, qui ne sont d'ailleurs pas marqués comme des amorces dans le corpus :

- (15) je lui ai fait une dizaine de petits exercices hum euh rapides
à faire pour voir s' il avait bien intégré les formules de **trigo**
(106_43, F, 37, F)
- (16) OK ça va j' ai rien dit #noise ben si tu veux faire un **tuto** vidéo
je veux bien (196_11, F, 24, I)
- (17) tu vas voir je t' ai envoyé un mail ... et du coup il y a mon
CV et ma lettre de **motiv** mais bon ça faut juste ... jeter un coup d'
œil mais a priori euh il y a pas de faute ... (02_08, F, 20, F)

Dans l'exemple suivant, le transcripteur alterne les analyses, alors qu'il semble clairement qu'il s'agit d'une apocope et non d'une amorce de mot :

- (18) c'est-à-dire un paquet euh ... ben il était pas très grand donc
à mon avis il y a pas beaucoup de drogue mais genre ... c'était un
paquet carré ... et quand même pas trop petit ... entouré de gros
scotch euh brun ... et il y avait trois flics et un des gros chiens euh
qui sentaient #noise comme ça mais #noise je sais pas comment ils
sentaient mais ils sentaient et après ils reniflaient y reniflaient et
après on a croisé **une combi** de police **un combi**~ je sais pas qui ar~
une combi #noise comment on dit (<105_21>, F, 22, B)

Pour la locutrice, la recherche lexicale porte sur le genre du mot (*comment on dit*), mais la forme *combi* est bien celle qu'elle a sélectionnée. S'il y a disfluence ici et auto-interruption, avec allée et venue sur l'axe syntagmatique et empilement paradigmatique, ce n'est pas l'amorce de mot qui est concernée mais la sélection du genre grammatical du nom, et donc un ajustement morphosyntaxique, comme on en trouve dans un type particulier de répétition déterminant (Dister 2008).

4.4. Les amorces de liaison

Si les transcriptions orthographiques n'ont pas comme objectif de noter des phénomènes phonétiques, on constate que la liaison n'y échappe pourtant pas, comme l'illustrent les exemples 19 et 20 :

(19) il y avait des moments où je passais pile au moment du lever de soleil et du coup il y avait de la brume sur la plaine et **les z~ les Alpes** au loin et le soleil se levait au-dessus de l'Allemagne (437_16, H, 26, F)

(20) du coup de ce que j' ai compris ... elle hum ... elle va ... elle fait ... un **un n~ un entretien** genre en vrai avec quelques personnes <61_18, F, 23, B>

On peut poser l'hypothèse sans craindre de se tromper que *Alpes* et *entretien* sont anticipés ; néanmoins ni la lettre *z* ni la lettre *n* ne sont les ébauches du mot orthographique qui sera complété. Or comme les locuteurs n'ébauchent ni le /a/ de *Alpes* ni le /ɑ̃/ de *entretien*, le transcripateur en est réduit à noter la liaison. Dans une transcription qui ne marque pas la liaison, ni aucun élément phonétique de manière générale, les notations proposées ci-dessus invitent le lecteur à interpréter la séquence comme une amorce corrigée (cf. 6.2), ce qui n'est pas le cas. Néanmoins, si le transcripateur décide de classer cette liaison comme un phénomène purement phonétique et de ne pas le transcrire, cette marque de l'élaboration de l'énonciation sera perdue.

L'exemple (21) peut sembler étonnant. Ici, le transcripateur choisit l'amorce *on~*. En réalité, ce n'est pas le pronom qui est amorcé, mais la construction syntaxique entière, qui est reprise ensuite avec *nous on est chauds*. Dans la cohérence des choix faits, il n'aurait donc pas fallu marquer l'ébauche de construction, même si l'écoute du passage suggère une auto-interruption de la locutrice marquée sur le pronom.

(21)... euh d'ailleurs euh **on~** si vous êtes chauds euh nous **on est** chauds euh enfin on est dispo en fait hein euh vendredi j' avais proposé à #anon et tout ça donc si vous êtes tentés euh on se fait un petit truc tranquille (...) (<106_02>, F, 37, F)

5. Le corpus

Le corpus *Les Vocaux* comporte 965 occurrences du signe ~ indicateur d'une amorce de mot. Nous avons travaillé sur la version V 002 qui comporte quelques erreurs, inhérentes à tout projet d'envergure transcrivant de l'oral (p. ex. *grand~ chose*, où le tilde est mis à la place d'un trait d'union).

Nous avons constitué un sous-corpus de 300 amorces sélectionnées aléatoirement. En effet, les études antérieures sur l'amorce (Pallaud 2002, Dister 2008) ont montré qu'on observe une forte variation inter-locuteurs. Pallaud, dont les résultats mettent en évidence le caractère individuel du phénomène, avance l'hypothèse que l'amorce est une « composante stable du style du locuteur » (Pallaud 2002 : 83). Elle observait un rapport de 1 à 10 selon le locuteur. Dans un corpus de 400.000 mots que nous avons analysé, composé majoritairement d'entretiens semi-dirigés, la moyenne, tout locuteur confondu, était d'une amorce tous les 123 mots. La répartition de la fréquence par locuteur⁹ allait de 20,6 amorces tous les 1000 mots (ce qui équivaut à une amorce tous les 49 mots) à aucune amorce chez l'une des intervieweuses qui comptabilisait pourtant 867 mots. Ainsi, nous avons un rapport de 1 à 20. A la différence des données de Pallaud, certains intervieweurs étaient présents dans plusieurs entretiens, et nous avons pu également observer une variation intra-locuteurs non négligeable (Dister 2007 : 192-193).

Quoi qu'il en soit, étant donné la composition du corpus *Les Vocaux* – des locuteurs ont envoyé plusieurs vocaux, d'autres un seul ; les vocaux vont de quelques secondes à plusieurs minutes (Glikman et al., dans ce volume) –, et afin de ne pas surreprésenter la parole de certains locuteurs, nous avons sélectionné aléatoirement les 300 amorces analysées. Nous avons néanmoins veillé à ce que chaque locuteur dont le discours en contient soit représenté par au moins une occurrence.

6. Typologie des amorces

Nous reprenons la typologie de Pallaud (2002) qui ventile les amorces de mots en trois catégories distinctes, selon les effets produits au lieu syntaxique qu'occupe l'amorce : l'amorce complétée, l'amorce corrigée et l'amorce laissée inachevée.

6.1. Les amorces complétées

Dans l'amorce complétée, le mot amorcé est repris en entier après l'interruption et ce sur une même place syntaxique. Dans cette catégorie, nous pouvons distinguer deux cas différents :

Le mot amorcé est complété avec reprise de la partie déjà énoncée

(22) il est parti ce matin très tôt et il revient pas avant **vendr~ vendredi**
matin (285_19, F, 29, F)

⁹ Nous avons retiré les données non significatives de ces résultats (lorsque le locuteur comptabilise peu de mots).

Le lexème complet suit immédiatement le fragment. En formulant des hypothèses quant à la cohérence et la congruence des propos du locuteur, nous considérons qu'un *repair* qui reprend à l'initiale la même forme que l'amorce est bien la forme complète de celle-ci. C'est d'ailleurs en se basant sur une telle hypothèse que le transcripteur a posé des choix de transcription pour certains morphèmes, principalement quand ils ne sont que des phonèmes (cf. 4.1).

Le mot amorcé est complété avec reprise de la partie amorcée mais aussi d'élément(s) antéposé(s) à l'amorce

(23) mais euh ... mais **on on on discu~ on discutera** de tout ça et surtout de ... de ... de ce qui se passe dans ta tête (433_11, H, 24, F)

Ici, la reprise concerne non seulement le verbe amorcé mais également le pronom sujet qui le précède. C'est tout le groupe syntaxique pronom sujet + verbe qui est repris.

Lorsque le mot amorcé est un substantif, le déterminant qui le précède est fréquemment repris dans le *repair* :

(24) mais c' est **c' est un ma~ un matériel** euh impressionnant (437_02, H, 26, F)

Lorsque le syntagme auquel appartient le lexème amorcé commence par une préposition, celle-ci est souvent reprise également :

(25) j' ai tellement l' habitude que personne n' utilise de point **à la fin de ses me~ à la fin de ses messages** (437_02, H, 26, F)

Un 3^e cas est possible, celui d'une auto-interruption suspensive. L'énoncé est alors momentanément interrompu, puis il se poursuit comme si cette interruption n'avait pas eu lieu (Pallaud et Bertrand 2020). Le mot amorcé est complété immédiatement, sans reprise de la partie ébauchée. Voici une transcription issue de la banque de données Valibel, où la barre oblique symbolise l'amorce¹⁰ :

(26) **iljDV1** je blague / enfin // je / à Gand parce que parce qu/ j'ai j'ai env/ j'ai envie de dire Gand parce que les les Flamands qui ont appris le français l'ont l'ont appris euh dans des milieux sociaux très très privilégiés probablement dans des milieux dans lesquels on on **a/-ccordait** énormément d'importance au au au au français (silence) (...)
[Valibel, iljDV1r]

¹⁰ La barre simple (/) indique une pause brève, la double barre (//) une pause plus longue.

Sans autre élément (un *euh* par exemple), l'*interregnum* est vide et seul l'indice phonétique d'une pause silencieuse indique la suspension de l'énonciation. Dans des transcriptions orthographiques de l'oral, où aucun recours à des appareils de mesure ne se fait, l'interprétation est laissée à l'appréciation du transcripteur. S'il n'y voit qu'un simple phénomène phonétique, et que la transcription n'est pas censée fournir d'indication relevant de la phonétique, cette suspension à l'intérieur d'un mot ne sera pas notée. Le corpus *Les Vocaux* ne fournit pas d'exemple de ce type d'amorce de mots. Ces interruptions, sorte de « haltes », ne sont pas notées non plus dans les transcriptions du CFPP2000 (Branca et al. 2011) ou du CFPB (Dister et Labeau 2017). Dans d'autres données qui les prennent en considération, celles-ci sont ultra minoritaires : 14 cas sur 3612 amorces étudiées dans Dister (2006) et 3 sur 1000 occurrences chez Pallaud (2002).

Notons par ailleurs que ces auto-interruptions suspendent momentanément le déroulement linéaire de l'énoncé sans que la reprise de la fluence ne présente de correction. Il n'y a donc à proprement ni *reparandum* ni *repair* dans les séquences de ce type où la fluence est momentanément suspendue.

6.2. Les amorces corrigées

Dans le cas des amorces corrigées, l'énonciation de la forme pleine du fragment se fait également sur la même place syntaxique, mais à la différence de l'amorce complétée, le lexème utilisé dans la correction diffère de celui qui a été amorcé par le locuteur. Comme pour la catégorie précédente, la correction peut se faire soit immédiatement (27), soit avec reprise d'éléments présents dans le contexte gauche de l'amorce (28), comme l'illustrent respectivement les deux exemples suivants :

(27) j'arrive pas à faire des longs vocaux il faut que **je réfl~ respire** entre les deux hum ... (329_07, F, 34, F)

(28) puis **elle a répon~ elle a dit** faut faire une danse de la pluie (365_26, A, 34, B)

Ce type d'amorces relève d'une erreur de sélection lexicale où le locuteur se livre à un réajustement dans le choix du lexique. Selon Pallaud, ce phénomène est à interpréter comme un lapsus interrompu et corrigé (2002 : 87). Dans (29), l'auto-variation est le fait d'une forme de lapsus (« la langue fourche »), plus que le témoin d'une recherche lexicale :

(29)... au pire après euh ... **on alu~ on annulera** ton trajet euh TER et ... (442_05, F, 23, F)

6.3. Les amorces inachevées

Ce qui distingue les amorces inachevées des deux précédentes classes, c'est que la correction de l'amorce se fait sur une autre place syntaxique. En fait, il serait plus correct de dire qu'il n'y a pas véritablement de correction : le déroulement syntagmatique de l'énoncé est interrompu et la place amorcée est laissée vide. La linéarité de l'énoncé se poursuit sur une autre place syntaxique :

- (30) ouais puis maman et papa ils regardent ce soir ... pour réserver avec l' avion et puis euh ... et puis l' hôtel enfin c' est le principal après les activités sur place ... **il y a pas besoin de m~ ça demande pas de réservation** (131_11, F, 28, S)

On le voit, c'est en fait toute la structure que la locutrice va modifier, reprenant la linéarité de l'énoncé avec une autre construction syntaxique. La construction qui s'achève avec l'amorce n'est pas reprise, et la locutrice privilégie une autre formulation. On a ici une rupture dans de la construction verbale amorcée après le point d'interruption.

6.4. Les séquences d'amorces

Notre corpus contient des lieux de disfluences dans lesquels cooccurrent plusieurs amorces, comme l'illustrent les exemples d'amorces suivants :

- (31) euh on boit un coup on fête ton anniv euh tout ça euh **on s~ b~** on se boit un truc on mange un petit truc tous ensemble euh et puis euh et puis voilà quoi (106_02, F, 37, F)

- (32) mais je je n' y arrivais pas ... euh ... je ... je sais pas j~ je d~ j~ j~ j' étais dans l' incapacité euh de de le faire (271_03, F, 22, B)

Ces cas mériteraient d'être relevés exhaustivement sur l'ensemble de *Les Vocaux* afin d'être analysés plus spécifiquement.

6.5. Cas inclassables

Si certaines amorces se laissent aisément classer et ne font aucun doute ni sur le lexème plein, ni sur la classe grammaticale ou la position syntaxique du fragment, d'autres par contre nécessitent une analyse plus fine. Ainsi, contrairement à ce qu'avance Jeanjean (1984) qui n'a

travaillé que sur des cas relativement simples¹¹, tous les fragments ne se laissent pas facilement appréhender par l'analyste.

Dans notre corpus, 11 amorces n'ont pu être classées. Elles sont composées d'une seule consonne, et il nous a semblé qu'il aurait fallu forcer l'interprétation pour restituer le lexème amorcé.

(33) j'ai pas vu le documentaire mais s~ je sais qu'il faut euh leur courir après
et tout mais ... moi je t'avoue perso euh j'avais la flemme (364_03, F,
32, B)

Dans cet exemple, on pourrait faire l'hypothèse qu'il s'agit de l'amorce de *sais*, mais nous n'avons guère d'indice qui plaiderait plus pour cette solution que pour une autre. En effet, bien que dans le contexte droit du fragment, cette anticipation du verbe *sais* sans la présence du pronom sujet ne correspond pas au format des amorces et à la planification des énoncés. En effet, dans les autres cas d'amorces du verbe présents dans le corpus, le pronom sujet n'est pas omis.

Dans le cas suivant, on ne voit pas quelle hypothèse sérieuse on pourrait formuler :

(34) et quoi te dire bé en gros j' ai passé euh pff trois heures chez lui euh
ben enfin euh t~ en fait je suis arrivée à onze heures (285_02, F, 29, F)

Nous avons donc préféré ne pas prendre en compte ces fragments dans la typologie¹².

6.6. Répartition des amorces en corpus

Les 300 amorces que nous avons analysées se ventilent de la manière suivante :

Types d'amorces	n = 300
complétées	47,6 % (143)
corrigées	19 % (57)

¹¹ L'auteure ajoute cependant : « La référenciation lexicale est donc seulement une question de probabilité, liée à la cohérence sémantique du texte. » (Jeanjean 1984 : 172)

¹² Dans Dister (2003), 25 % des 348 amorces étudiées n'ont pu être classées dans aucune des catégories, l'identification non seulement du lexème mais également de sa place syntaxique étant totalement impossible.

inachevées	18,6 % (56)
inclassables	3,6 % (11)
séquences	11 % (33)

Tableau 2 : Typologie des amorces

Si l'on ne tient pas compte des séquences d'amorces et des amorces inclassables, la répartition est la suivante :

Types d'amorces	N = 256
complétées	55,86 % (143)
corrigées	22,26 % (57)
inachevées	21,88 % (56)

Tableau 3 : répartition des 3 types d'amorces

Les amorces complétées sont les plus fréquentes, suivies des amorces corrigées à parts égales avec les amorces inachevées. Ainsi, c'est sur une même place syntaxique que le locuteur, majoritairement, poursuit son énoncé, que ce soit en complétant le lexème anticipé ou en le modifiant, puisque cette configuration se produit dans 78,13% des cas. Cela montre chez les locuteurs une bonne gestion de ces anticipations.

Chez Pallaud (2002), les amorces les plus nombreuses sont également les amorces complétées avec près de deux tiers des effectifs (59,9 %). Viennent ensuite les amorces inachevées (28,8 %) et enfin les amorces corrigées (11,3 %). Dans une autre étude sur 150 amorces (Pallaud 2004), la répartition est également en faveur des amorces complétées (59,6 %), avec une ventilation plus égale dans les deux autres catégories (18,5 % pour les amorces modifiées et 21,9 % pour les amorces achevées). Des résultats différents sont obtenus dans une étude précédente (Pallaud 1999) sur un corpus de 120 amorces ; l'auteure constatait que les amorces corrigées étaient les plus nombreuses avec 52,5 % des occurrences. Selon Pallaud, ces différences de résultats tendent à montrer une grande variation inter-locuteur dans ce phénomène langagier (cf. 5).

Si l'on se penche maintenant sur le format du *repair*, les résultats sont les suivants :

		Complétées (n=143)	Modifiées (n=57)	Total (n=200)
--	--	-----------------------	---------------------	---------------

1	immédiatement	37,06 % (53)	29,82 % (17)	35 % (70)
1bis	Immédiatement - avec <i>euh</i> ou ponctuant dans l'interregnum	2,80 % (4)	/	2 % (4)
2	reprise à l'initiale du groupe	27,27 % (39)	40,35 % (23)	31% (62)
2bis	reprise à l'initiale du groupe - avec <i>euh</i> ou ponctuant dans l'interregnum	2,10 % (3)	5,26 % (3)	3 % (6)
3	Ajout à gauche de l'amorce	20,98 % (30)	17,54 % (10)	20 % (40)
4	Reprise en modifiant des éléments antérieurs	6,99 % (10)	3,51 % (2)	6 % (12)
8	Séquence disfluente	1,40 % (2)	3,51 % (2)	2 % (4)
9	Suppression d'un élément	0,70 % (1)	/	0,5 % (1)
10	Ajout direct avec insertion	0,70 % (1)	/	0,5 % (1)

Tableau 4 : repair des amorces complétées et modifiées

La configuration la plus fréquente est celle où l'amorce est complétée immédiatement, sans reprise d'élément(s) antérieur(s) au fragment, avec un *interregnum* vide. Plus d'une séquence sur trois présente ce format (35 %), qui est aussi celui qui est le plus fréquent pour les amorces complétées. Les amorces modifiées, par contre, sont plus nombreuses avec reprise à l'initiale du groupe syntaxique (40,35% des occurrences dans cette catégorie). La classe 3 est illustrée par les séquences où le *reparandum* contient des éléments ajoutés soit à gauche de l'amorce (35) et (36), soit à gauche du début du *reparandum* (37) :

(35)... euh **la temp~ la fameuse tempête** de grêle ... (358_01, F, 31, F)

(36) d' ailleurs enfin je suis en train de me dire que peut-être je vais mettre un truc **sur T~ sur ma bio Tinder** à ce propos (433_102, H, 24, F)

(37) je suis en train de réfléchir à ce que je priorise pour euh remplir le camion ... et en fait mon père euh il m~ il m' a dit que il pourrait euh que en novembre euh il comptait euh ... louer une camionnette et me ramener euh des trucs et passer euh chez ma grand-mère récupérer les trucs les autres trucs qu' elle nous file ... et euh et il m' a dit aussi de faire attention euh enfin ... que voilà genre si jamais euh finalement euh ... le canapé on se dit putain c' est chaud ça va pas rentrer et tout euh ... comme on sait pas finalement euh ... comment ça

va se passer demain pour tes parents ce qu' ils vont pouvoir mettre et tout ... que il peut aussi nous le ramener donc **est -ce que si j~** parce que là il y a le frigo en plus qu' il va falloir mettre dans la ... camionnette quand même ... **est -ce que si euh si j' emmène** pas le canapé ça pose un problème ou pas parce que ... euh moi j' ai peur que tout tienne pas ... dans ma voiture ... (442_10, F, 23, F)

Les retouches ici ne concernent pas le choix du lexique anticipé, puisque l'amorce est complétée. L'auto-interruption est liée au fait que le locuteur souhaite mieux préciser son propos, en ajoutant des éléments d'information. Il garde en mémoire ce qu'il vient de dire, et le *reparandum* reprend ces éléments en y ajoutant d'autres.

Dans l'exemple (37), la locutrice s'interrompt pour ajouter une longue proposition subordonnée. Durant cet ajout, elle gardant en mémoire la structure qu'elle avait amorcée, structure qu'elle reprend à l'identique pour la compléter. On le voit, ces amorces participent pleinement de la planification des énoncés en cours d'élaboration.

7. La classe grammaticale de l'amorce

Outre les cas inclassables et les séquences d'amorces (44 occurrences), il n'a pas été possible de donner la classe grammaticale de 31 fragments sans forcer l'analyse. Le décompte se fait donc sur 225 amorces.

Classe grammaticale	N=225
nom	19,11 % (43)
verbe conjugué	15,56 % (35)
pronom sujet	18,89 % (29)
verbe à infinitif	10,22 % (23)
déterminant	8 % (18)
adverbe	6,22 % (14)
participe passé	6,22 % (14)
pronom clitique	5,78 % (13)
préposition	4 % (9)
adjectif	3,56 % (8)
adv composé	3,11 % (7)
nom propre	1,33 % (3)

conjonction de coordination	1,33 % (3)
conjonction de subordination	0,89 % (2)
ponctuant	0,89 % (2)
pronom relatif	0,89 % (2)

Tableau 5 : Classe grammaticale des amorces

Les lexèmes les plus fréquemment amorcés appartiennent à la classe du verbe avec 72 occurrences (en incluant la forme du participe), soit 32 % des amorces étudiées. Viennent ensuite les noms avec 47 occurrences (20,89 %), suivis des pronoms avec 44 fragments (19,55 %). Les mots lexicaux sont les plus nombreux, avec 147 occurrences¹³ (65,33 %), contre 78 occurrences de mots grammaticaux (34,66 %). Dans Dister (2003), nous obtenions 70,6% de mots lexicaux. Pallaud (2002), sur les 925 amorces analysées, a un taux encore plus élevé de mots grammaticaux, avec de 82% des occurrences. Si le pourcentage est plus élevé, il n'en reste pas moins que ce sont une minorité de mots grammaticaux qui sont amorcés. Pallaud émet cette hypothèse :

Que cela soit dû à un lien entre la taille du mot et le phénomène d'amorce reste à vérifier. Cela suggère en effet que les mots plurisyllabiques puissent être plus fréquemment l'objet d'achoppements que les mots monosyllabiques. (Pallaud 2002 : 93)

On assiste donc pour les amorces à l'inverse de ce que l'on constate pour un autre phénomène disfluent qui participe aussi à la structuration de l'énoncé : les répétitions de mots. Dans ces séquences, ce sont justement les mots grammaticaux, principalement monosyllabiques, qui font majoritairement l'objet de répétitions (Dister 2014, Pallaud et Henry 2004a et 2004b).

Conclusion

Si elle peut sembler facile à appréhender (« un locuteur commence un mot puis s'arrête en cours d'énonciation »), l'amorce ne se laisse pas si facilement identifier dans les faits. On en veut pour preuve les difficultés rencontrées lors de la transcription. De plus, indépendamment de son identification, l'amorce pose aussi des problèmes quant au choix de transcription

¹³ Nous avons comptabilisé les adverbes (21 occurrences) dans les mots lexicaux pour permettre les comparaisons avec les études antérieures.

orthographique, qui devraient idéalement être pris en compte en amont, lors de l'établissement des conventions de transcription.

Pour particuliers que soient les messages vocaux – des données médiées, différées, écologiques –, l'analyse des amorces montre des tendances tout à fait comparables à celles que l'on rencontre dans des corpus où les interactions sont plus formelles, comme des entretiens semi-dirigés : ce sont les amorces dont la forme complétée ou corrigée se poursuit sur une même place syntaxique qui sont majoritaires. Les locuteurs, dans ses allées et venues sur l'axe syntagmatique, anticipent de la même manière dans leur discours en cours d'élaboration, gardant en mémoire les séquences amorcées.

BIBLIOGRAPHIE

BILGER M., 1982, « Contribution à l'analyse en grilles », *Recherches sur le français parlé* 4, Université de Provence, p. 195-215.

BILGER M., 2000, « Petite typologie des conventions de transcription de l'oral. Quelques aspects pratiques et théoriques », M. Bilger (Éd.), *Linguistique sur corpus. Études et réflexions*, *Cahiers de l'Université de Perpignan* 31, p. 77-92.

BLANCHE-BENVENISTE Cl. BOREL B., deulofeu J., durand J., giacomini A., loufrani Cl., MEZIANE B. et PAZERY N., 1979, « Des grilles pour le français parlé », in *Recherches Sur le Français Parlé* 2, p. 163-209.

BLANCHE-BENVENISTE Cl., BILGER M., ROUGET Chr., VAN DEN EYNDE K., 1990, *Le Français parlé. Études grammaticales*, Paris, CNRS Éditions.

BLANCHE-BENVENISTE Cl., 1987, « Syntaxe, choix de lexique et lieux de bafouillage », *DRLAV* 36-37, pp. 123-157.

BLANCHE-BENVENISTE Cl., 2000, « Transcription de l'oral et morphologie », M. Gille et R. Kiesler (Éds.), *Romania Una et diversa, Philologische Studien für Theodor Berchem*, Tübingen, Gunter Narr, p. 61-74.

BLANCHE-BENVENISTE Cl., 2002, « Réflexions sur les transcriptions de corpus de français parlé », *Revue Parole* 22-23-24, p.91-117.

- BLANCHE-BENVENISTE Cl., 2003, « La naissance des syntagmes dans les hésitations et répétitions du parler », J.-L. Araoui (Éd), *Le sens et la mesure. Hommages à Benoît de Cornulier*, Paris, Honoré Champion, p. 40-55.
- BLANCHE-BENVENISTE Cl. et JEANJEAN C., 1987, *Le Français parlé. Transcription et édition*, Paris, Didier Érudition.
- BLANCHE-BENVENISTE Cl. et sabio Fr., 2023, *Approches de la langue parlée en français*, Paris, Ophrys (3^e édition).
- BRANCA-ROSOFF S., FLEURY S., LEFEUVRE Fl. et PIRES M. (2011), *Constitution et exploitation d'un corpus de français parlé parisien*, *Corpus* 10, 81-98.
- BRES J. et GARDES-MADRAY Fr., 1991, « Ratages et temps de l'à-dire », H. Parret (Éd.), *Le Sens et ses hétérogénéités*, Paris, Éditions du CNRS.
- CAPPEAU P., 1998, « Quelques mots sur quelques bribes liés au genre », M. Bilger, K. van den Eynde et Fr. Gadet (Éds.), *Analyse linguistique et approches de l'oral. Recueil d'études offert en hommage à Claire Blanche-Benveniste*, *Orbis Supplementa* 10, Leuven, Peeters, pp. 301-311.
- CERQUIGLINI B., 2019, *Parlez-vous tronqué ? Portrait du français d'aujourd'hui*, Paris, Larousse.
- DELIC, 2004, « Présentation du Corpus de référence du français parlé », *Recherches sur le français parlé* 18, pp. 11-42.
- DISTER A., 1995, *L'Apocope, un phénomène vivant du français contemporain*, Département d'Études romanes, Université de Liège, Mémoire inédit de licence.
- DISTER A., 2007, *De la transcription à l'étiquetage morphosyntaxique. Le cas de la banque de données textuelles orales VALIBEL*, UCLouvain, Thèse non publiée.
- Dister A., 2008, « L'autocorrection immédiate en français parlé : le cas des déterminants », *Actes des 9^{es} Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)*.
- DISTER A., 2014, « parler sans accent pour moi c'est sans sans sans bafouiller. Quelles répétitions de formes en français parlé ? », *Congrès mondial de linguistique française*, Berlin, 19-23 juillet 2014.

https://www.shs-conferences.org/articles/shsconf/pdf/2014/05/shsconf_cmlf14_01386.pdf

DISTER A. et LABEAU E., 2017, « Le Corpus de français parlé à Bruxelles : origine, hypothèses et développements », *Cahiers de l'AFLS* Vol. 21, no.1, p. 1-22

DUEZ Danielle (1991). *La Pause dans la parole de l'homme politique*, Paris, Éditions du CNRS.

EDWARDS Jane A., 1995, « Principles and alternative systems in the transcription, coding and mark-up of spoken discourse », G. Leech, G. Myers, J. Thomas (Éds.), *Spoken English on Computer. Transcription, Mark-up and Application*, New York, Longman, p. 19-34.

GADET Fr., 2003, *La Variation sociale en français*, Paris, Ophrys.

GROSJEAN Fr. et DESCHAMPS A., 1972, « Analyse des variables temporelles du français spontané », *Phonetica* 26, p. 129-156.

GROSJEAN Fr. et DESCHAMPS A., 1973, « Analyse des variables temporelles du français spontané. II. Comparaison du français oral dans la description avec l'anglais (description) et avec le français (interview radiophonique) », *Phonetica* 28, p. 191-226.

GROSJEAN Fr. et DESCHAMPS A., 1975, « Analyse contrastive des variables temporelles de l'anglais et du français : vitesse de parole et variables composantes, phénomènes d'hésitation », *Phonetica* 31, p. 144-184.

GUMPERZ J. J. et BERENZ N., 1993, « Transcribing Conversational Exchanges », J.A. Edwards and M.D. Lampert (Éds.), *Talking Data. Transcription and Coding in Discourse Research*, Hillsdale, Lawrence Erlbaum Associates, p. 91-120.

JEANJEAN C., 1984, « Les ratés c'est ça fabuleux. Étude syntaxique des amorces et des répétitions », *LINX* 10, pp. 171-177.

GLIKMAN J., MAZZIOTTA N., BENZITOUN Chr. et FAUTH C., 2026, « Le corpus Les Vaux : de la constitution à l'exploitation », *Travaux de linguistique*.

LEVELT W. J.M., 1989, *Speaking : from intention to articulation*. Cambridge, MIT Press.

LICKLEY R., 2001, « Dialogue Moves and Disfluency Rates », *Proceedings of DISS'01* (August 29-31, Edinburgh), pp. 93-96.

MACLAY H. et OSGOOD Ch. E., 1959, « Hesitation Phenomena in Spontaneous English Speech », *Word* 15:1, p. 19-44.

- MONDADA L., 2000, « Les effets théoriques des pratiques de transcription », *LINX* 42, revue de l'Université de Paris X-Nanterre, p. 131-150.
- MARTINIE Br., 2001, « Remarques sur la syntaxe des énoncés réparés en français parlé », *Recherches sur le français parlé* 16, Université de Provence, p. 189-206.
- OCHS E., 1979, « Transcription as theory », E. Ochs et B. B. Schieffelin (Éds.), *Developmental pragmatics*, New York-San Francisco-London, Academic Press, p. 43-72.
- PALLAUD B., 2002, « Les amorces de mots comme faits autonymiques en langage oral », *Recherches sur le français parlé* 17, Université de Provence, p. 79-101.
- PALLAUD B., 2003, « Splendeurs et misères de la transcription », *Cliniques méditerranéennes* 68, p. 59-73.
- PALLAUD B. et HENRY S., 2004a, « Amorces de mots et répétitions dans les énoncés oraux », *Recherches sur le français parlé* 18, pp. 201-229.
- PALLAUD B. et HENRY S., 2004b, « Amorces de mots et répétitions : des hésitations plus que des erreurs en français parlé », G. Purnelle, C. Fairon et A. Dister (Éds.), *Le Poids des mots. Actes des 7^{es} Journées internationales d'Analyse statistique des Données textuelles*, Louvain-la-Neuve, Presses universitaires de Louvain, p. 849-858.
- PALLAUD B. et XUEREB R., 2008, « Les troncations de mots chez un locuteur bègue », *Travaux Interdisciplinaires de Parole et Langage* 26, p. 93-113.
- PALLAUD B. et BERTRAND R., 2020, « Auto-interruptions et disfluences à l'oral. C'était euh tu vois complètement loufoque comme si ouais euh comme situation », in Fabrice Hirsh, Ivana Dirdiskova, Christelle Dodane (Dirs), *Manuel de pausologie. Recueil de recherches sur la parole et le discours*, Paris, L'Harmattan, p. 21-47
- PIETRANDREA P. et KAHANE S., 2012, « La typologie des entassements en français », Actes du 3^e congrès mondial de linguistique française, Lyon. <https://shs.hal.science/halshs-01086467/file/piles-typologie-cmlf2012.pdf>
- RAINGEARD M. et LORSCHIEDER U., 1977, « Édition d'un corpus de français parlé », *Recherches sur le français parlé* 1, Université de Provence, p. 14-29.
- ROUBEAU M.-N., 2004, « Du bon usage des amorces dans la transcription des corpus », *Recherches sur le français parlé* 18, Université de Provence, p. 163-184.

SHRIBERG E., 1994, *Preliminaries to a Theory of Speech Disfluencies*, Université de Berkeley,
Thèse non publiée.

VALIBEL, <https://www.uclouvain.be/fr/instituts-recherche/ilc/valibel/valibel-acces/access>

VERSION NON DÉFINITIVE