



HAL
open science

Amorces de mots et répétitions : des hésitations plus que des erreurs en français parlé

Berthille Pallaud, Sandrine Henry

► **To cite this version:**

Berthille Pallaud, Sandrine Henry. Amorces de mots et répétitions : des hésitations plus que des erreurs en français parlé. 7es Journées Internationales d'Analyse Statistique des Données Textuelles, Dec 2004, Louvain-la-Neuve, Belgique. pp.848-858. <hal-00285535>

HAL Id: hal-00285535

<https://hal.science/hal-00285535v1>

Submitted on 5 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Amorces de mots et répétitions : des hésitations plus que des erreurs en français parlé

Berthille Pallaud¹, Sandrine Henry²

¹ UMR 6057 Université de Provence, Aix-en-Provence, 13621 cedex 1 France

² DELIC, Université de Provence, Aix-en-Provence, 13621 cedex 1 France

Abstract

Word fragments are, like repeats, very frequent events in spoken language. Speakers and listeners alike are generally unaware of these phenomena as if they were not part of the communication process. Previous studies (Pallaud, 2002, Henry, 2002, Henry et Pallaud, 2003) have shown that word fragments and repeats encompass different categories of phenomena, and this enables us to define them as an heterogeneous group ruled by different types of constraints and mechanisms. These two types of disfluencies may be considered less as repairs than hesitation phenomena. Fragments of word differ from repetitions: they occur later in the syntagm and in the clause.

Résumé

Les amorces ou fragments de mots tout comme les répétitions sont des événements langagiers d'une grande fréquence à l'oral. Ces deux phénomènes sont toutefois le plus souvent ignorés du locuteur qui parle et de son récepteur comme si ces phénomènes étaient non-communicatifs. Des travaux antérieurs (Pallaud, 2002, Henry, 2002, Henry et Pallaud, 2003) ont montré que les amorces de mots comme les répétitions regroupent des phénomènes de catégories différentes ce qui permet de les caractériser comme un ensemble hétérogène obéissant à plusieurs types de contraintes ou de mécanismes.

L'analyse de ces deux sortes d'achoppement montre qu'ils sont moins le moment d'une réparation d'erreurs que celui d'une hésitation dans l'élaboration de l'énoncé. La comparaison des amorces de mots et des répétitions fait apparaître une nette différence : les répétitions se produisent plus tôt dans le syntagme que ne le font les amorces de mots.

Mots-clés : amorces de mots, répétitions, disfluences, français oral et spontané

Key-words : word fragments, repeats, disfluencies, oral and spontaneous French

1. Introduction

Notre étude s'intéresse à ces phénomènes caractéristiques de l'oral et souvent qualifiés de « scories » ou même « ratés » : les amorces de mots et les répétitions. Dans notre étude, le terme d'amorce (ce que Dister nomme les *mots tronqués* et Schegloff *et al.*, 1977 des *auto-corrections*) est conforme à l'emploi qui en est fait dans les conventions de transcription du *Groupe Aixois de Recherche en Syntaxe* (GARS) : il est réservé aux ratés dans la production de morphèmes. Les fragments de syntagme sont, quant à eux, plus souvent identifiés sous le terme de *bribes* (Cappeau, 1998). Il reste que ces trébuchements ne sont pas toujours distingués. Ils sont parfois analysés comme une seule catégorie de fragments (Jeanjean, 1984 et Martinie, 1999 pour ce qui est des études sur le français parlé ; Eklund, 2003, en ce qui concerne d'autres langues). Les conventions de transcription du GARS permettent un repérage automatique de ce type d'achoppement :

FEMMES 51,15 pour moi c'est pas *malho-* c'est pas *malhonnête* c'était euh

Dans la très grande majorité des cas, ces phénomènes étant identifiés comme des "réductions" de mots qui ne sont qu'amorcés, leur transcription est orthographique et atteste ainsi de cette interprétation.

Même si les études de ces phénomènes particuliers à l'oral ne les catégorisent pas toujours de façon semblable, elles constatent toutes que ces achoppements ne sont pas rares et ne se produisent pas de façon aléatoire (Duez, 2001b, Pallaud, 2002a). En moyenne (sur un corpus de 46 549 mots¹), on peut dire qu'on trouve environ 4 amorces tous les 1000 mots soit pour un débit moyen de 200 mots par minute une amorce toutes les 75 secondes. Avec un corpus de un million de mots (*conf. infra*), cette fréquence est encore plus grande puisqu'on retrouve une amorce toutes les 50 sec (Henry et Pallaud, 2003d). Ces phénomènes viennent loin derrière les pauses courtes silencieuses ou remplies qui sont six fois plus nombreuses (Pallaud, 2002a).

Une analyse comparative de ces douze corpus (un locuteur par corpus) a été faite sur *le nombre moyen de mots entre deux amorces* afin d'évaluer le nombre d'événements (ici les amorces) tous les 1000 mots. Elle montre qu'il existe une grande hétérogénéité entre les locuteurs puisque les valeurs peuvent varier d'un facteur 10. Cette variabilité observée dans la production des amorces suggère que ce paramètre reflète fortement une composante individuelle. Le phénomène d'amorce semble être une marque d'énonciation qui rend compte de caractéristiques subjectives : « Les hésitations peuvent fonctionner comme des marqueurs de personnalité » (Duez, 2001 p.129).

Notre étude sur les amorces répétées (Henry et Pallaud, 2003c) a confirmé, quant à elle, que la distinction entre les amorces et les répétitions de mots s'imposait. Elle a suggéré, de plus, que les répétitions de mots se produisent plus tôt que les interruptions de mots dans le cours du syntagme. Reste à savoir si cette préséance se retrouve également au niveau de la proposition (avant plutôt qu'après le verbe, par exemple). Nous avons constaté, par ailleurs que le phénomène d'amorce a été souvent considéré comme une auto-correction, un phénomène de réparation ou d'hésitation. Notre analyse de ces achoppements se propose de préciser en quoi ils sont une manifestation d'hésitation plutôt qu'une erreur d'énonciation autrement dit une pause dans l'énoncé plutôt qu'un trébuchement auquel le locuteur remédiera.

2. Corpus et méthodes

A partir des années 1970, sous l'impulsion et la direction de Jean Stéfanini puis de Claire Blanche-Benveniste, de nombreuses études sur le français parlé ont été entreprises dans le cadre du *Groupe Aixois de Recherche en Syntaxe*. Elles ont nécessité l'établissement de corpus oraux dont l'objectif était de recueillir des données représentatives d'un français parlé qu'il serait possible de qualifier d'un « usage général et courant ». Un premier ensemble a été constitué connu sous le nom de *Corpaix*

Depuis les années 1990, toutes les transcriptions ont été saisies sur ordinateur et mises à un format standard. Sont distinguées la partie "ancienne" du corpus (non-informatisée) et la partie "moderne" qui comprend systématiquement trois supports : enregistrement sonore (sur minidisk, puis transfert sur ordinateur), transcription sur papier et saisie informatisée du texte transcrit. La partie "moderne" comporte, au début de l'année 2000, environ 1 million de mots. Dans cet ensemble, ont été retenus des locuteurs ayant fait leur scolarité primaire en France. Le GARS ne privilégie pas l'étude des productions dites "spontanées" (conversations ou entretiens libres), un grand nombre de situations "contraintes" ont été retenues (demandes d'explications ou de renseignements, exposés, descriptions et récits etc.), avec des lieux très

¹ Cet ensemble est composé de 12 corpus recueillis par Pallaud dans le cadre du *Corpus de Référence de Français Parlé* (Pallaud, 2001). Par ailleurs, l'ensemble de la base de données sur les amorces est tirée du recueil *Corpaix* (une série de corpus d'entretiens référés par un nom) ; cet ensemble a été interrogé à l'aide d'un logiciel concordancier (*Contextes*) élaboré par Véronis (<http://up.univ-mrs.fr/~veronis/>).

variés: domicile d'une personne, lieu de travail, magasin, bureau, organisme public ou radio et télévision. Aucun enregistrement n'a été fait en laboratoire.

L'ensemble du corpus, enregistrements et transcriptions, a été fait, pour la plupart, grâce au travail d'étudiants. Deux exigences, souvent peu compatibles, ont prévalu: celle de la transcription, qui doit être fidèle (noter tout ce qui est dit, y compris les répétitions, les bribes, etc.) et la nécessité de rester lisible, au moins pour pouvoir être lue facilement par une communauté de chercheurs et pour ne pas rester confinée au groupe qui l'a réalisée. Comme le GARS étudie essentiellement des phénomènes de syntaxe qui se saisissent sur de longs passages, où la lisibilité est primordiale, le choix est celui d'une transcription "pauvre", sans surcharge pour noter les mimiques, les rires, les toux, les phénomènes d'interaction gestuelle, ni même les intonations.

Les transcriptions sont faites en orthographe standard, sans aucun "trucage". Le modèle orthographique est celui des dictionnaires, y compris pour les interjections (*euh*, *ben*, etc.). Des conventions particulières interviennent pour les éléments qui ne peuvent pas recevoir une orthographe française normée: mots étrangers, néologismes, etc. Les transcriptions ne comportent aucune ponctuation. En 1995, les nouvelles technologies, employées au sein de l'ESA-CNRS 6060, ont rendu possible le regroupement des transcriptions (le corpus *Corpaix*) et le développement d'un logiciel d'exploitation (Xcor un concordancier mis au point par JP Adam à partir de 1997).

Le « Corpus de référence de français parlé » répond à une requête de la Délégation à La Langue Française, Ministère de la Culture, qui l'a totalement financé. La réalisation de ce projet avait été confiée en 1998 à l'URA CNRS 6060, à partir de 2000, le projet a été pris en charge par la Jeune équipe DELIC (*DEscription Linguistique Informatisée sur Corpus*), dirigée par Jean Véronis. L'objectif de ce corpus était de mettre à la disposition de la communauté de linguistes des témoignages de la langue française parlée aujourd'hui, dans les principales villes de l'hexagone. Des enregistrements (132) de locuteurs choisis selon des critères précis (âge, niveau d'éducation, etc.), dans une quarantaine de villes de France, représentant des types de parole variés (parole publique, professionnelle ou privée) sont transcrits selon les normes du GARS. La totalité² est estimée à environ 50 heures de parole, soit environ 400 000 mots. La transcription et la segmentation ont été réalisées majoritairement à l'aide du logiciel *Transcriber 1.4.2*.³, programme développé par Claude Barras spécialement pour la transcription et l'alignement de corpus oraux.

Une fois l'intégralité du corpus obtenue, la suppression des balises XML et la concaténation de l'intégralité du corpus ont été effectuées grâce à des scripts en Perl afin de pouvoir effectuer des requêtes avec des concordanciers classiques. Le corpus a donné lieu au développement de plusieurs outils, notamment d'un concordancier (*Contextes*, développé par Jean Véronis) qui permet d'effectuer des requêtes sous forme de chaînes de caractères et d'obtenir la concordance sous forme textuelle.

La présente étude prend appui sur des données extraites de *Corpaix* auquel s'adjoignent des transcriptions inédites pour former un corpus de travail de 1 000 382 mots (283 situations, 794 locuteurs). Afin d'assurer une fiabilité dans l'extraction des répétitions, celles-ci sont

² Pour le français parlé, il offre une base de comparaison avec les corpus de français parlé hors hexagone, le corpus Valibel en Belgique (dix fois plus important) et le corpus d'Ottawa-Hull au Canada; il représente l'amorce de la réalisation d'un corpus comparable à celui du British National Corpus, pour l'anglais parlé.

³ Logiciel téléchargeable gratuitement sur le site: <http://www.etca.fr/CTA/gip/Projets/Transcriber/>

détectées à l'aide d'un programme informatique : écriture d'un script en langage *Perl*, puis application sous *Linux*. L'extraction des amorces s'est faite à l'aide du concordancier *Xcor* sur les mille premières amorces. L'ensemble des résultats sont ouverts dans un tableur *Excel* pour en faciliter le traitement.

2. Analyse des fonctions de l'amorce et de la répétition de mots

Pour tester notre hypothèse selon laquelle les phénomènes d'amorces ou de répétitions de mots sont moins le moment d'une erreur que la marque d'une élaboration de l'énoncé, nous avons retenu plusieurs paramètres : le poids respectif des trois types d'amorces, les effets de la répétition et de l'insertion d'éléments sur les amorces.

2.1 Les types d'amorces

Les amorces ne sont pas toutes semblables. Nous avons identifié (par annotation manuelle) trois sortes d'amorces dont la distinction repose sur l'effet de l'interruption du mot au lieu syntaxique qu'il occupe. Y a-t-il réduction définitive sur cette place syntaxique, auquel cas l'amorce n'est pas complétée ni modifiée mais simplement laissée inachevée, et ce qui suit l'amorce occupe une autre place syntaxique? Si ce qui suit l'amorce ne change pas de place syntaxique, deux cas sont possibles. Y a-t-il piétinement sur la place syntaxique et une simple reprise puis poursuite du morphème inachevé ou plutôt une modification lexicale? La réponse à ces questions permet de distinguer trois catégories d'amorces :

- amorce laissée inachevée

FEMMES 10,9 je pense que par rapport à M. C. on a souvent *tendan-* à à analyser un peu les choses euh à dire

- amorce complétée (dont celles qui s'accompagnent de modifications antérieures)

METALL 21,16 donc le cours *cons-* *consiste* en gros euh à à un exposé théorique
BUSABCD 20,3 elles suivent des cours d'aide-soignante *en d-* euh *au bout de dix* mois elles sortent diplômées

- amorce modifiée

ARBORIGN 13,4 pour les bras j'ai déjà *dét-* *retravaillé* les coudes et et allongé les bras

Les amorces complétées dans un corpus de plus d'un million de mots sont largement majoritaires (60%). Viennent ensuite celles qui sont inachevées (22%) et modifiées (18%). Ce fait nous incite à conclure que, dans les trois quart des cas (82%), l'amorce de mot est un achoppement marquant l'hésitation plus que l'erreur de langage. L'amorce complétée ou inachevée prélude le plus souvent à une poursuite de l'énoncé sans qu'une modification intervienne. L'étude de Cappeau (1998) sur les phénomènes de bribes à l'oral (c'est-à-dire des amorces répétées de syntagmes et non pas de mots) en relation avec le genre a montré que les bribes qui s'accompagnent d'un changement de genre sont 5 à 8 fois moins nombreuses que les précédentes. Son résultat confirme donc les nôtres : les bribes complétées sont beaucoup plus fréquentes que les bribes modifiées.

On constate par ailleurs (Figure1) que cette répartition est très différente selon que ces achoppements sont recueillis *à la volée* ou recherchés dans des corpus enregistrés (ici *Corpaix* ou le corpus *Bertuzzi97*⁴) :

Amorces complétées : deux exemples de ce type seulement (soit 5% des amorces) se trouvent dans les amorces relevées *à la volée*, alors qu'ils abondent dans les corpus enregistrés

⁴ Le corpus *Bertuzzi97* fait partie de l'ensemble *Corpaix* ; il comprend 25 460 mots et 100 amorces de mots.

(*Corpaix* et *Bertuzzi97*) où ce type d'amorce représente respectivement 60% et 30% des amorces émises.

Amorces laissées inachevées : dans le corpus *à la volée*, on ne relève aucune amorce de ce type, alors que dans les corpus enregistrés, ces amorces inachevées constituent près de 20%.

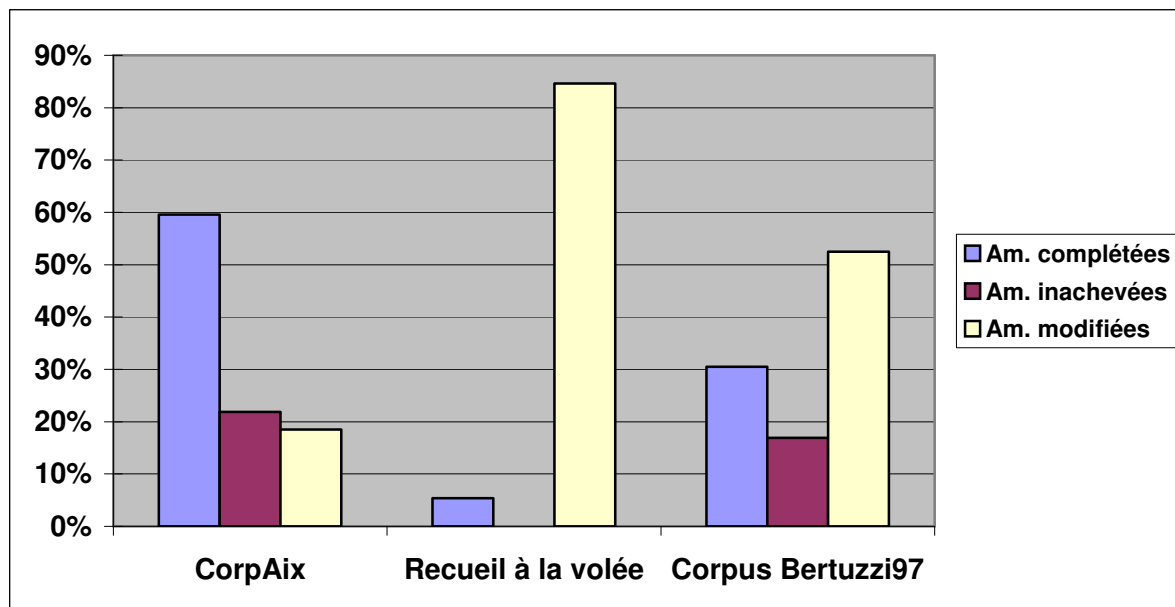


Figure 1. Répartition des trois types d'amorces en fonction du type de recueil

Amorces modifiées : ces phénomènes de modification sur l'élément touché par l'amorce représentent la quasi-totalité des amorces relevées *à la volée* alors que la proportion est de 18% seulement dans le grand corpus *Corpaix* et 52,5% dans le corpus *Bertuzzi97*.

Il n'y a donc pas que la fréquence des amorces qui peut varier d'un locuteur à l'autre. La répartition des trois types d'amorces peut également différer de façon non négligeable. La comparaison des résultats obtenus dans un gros corpus (*Corpaix*) avec les amorces relevées pour un locuteur (le corpus *Bertuzzi97*) confirme qu'il y a probablement une grande variabilité d'un corpus à l'autre c'est-à-dire d'un locuteur à l'autre.

Ces résultats montrent que, dans un recueil *à la volée*, non seulement les amorces de mots sont sous-représentées dans les « erreurs » de langage mais seules les amorces suivies d'une modification sont véritablement détectées. Comme, en fait, les amorces modifiées produites par les locuteurs ne sont pas les plus nombreuses (18,5% dans notre grand corpus), on peut penser que les amorces complétées et les amorces inachevées ne sont pas perçues ou ne sont pas considérées comme des « erreurs » de langage. Sur le versant de la perception, ces trébuchements ne seraient pas disfluents, du moins à ce stade de la discussion.

2.2 Les effets en contexte de l'achoppement

D'autres résultats viennent soutenir cette hypothèse. Il s'agit des effets en contexte de l'achoppement que constitue le phénomène d'amorce. Trois sortes de phénomènes seront pris en compte : les modifications antérieures à l'amorce de mot, les insertions d'éléments à l'occasion de phénomènes d'amorces et le cas des amorces répétées.

2.2.1. Modifications antérieures à l'amorce de mot.

Parmi les cas d'amorces complétées, toutes ne s'accompagnent pas d'une simple reprise. Nous avons en effet trouvé un certain nombre de modifications qui portent sur des éléments

antérieurs dans l'énoncé. Le locuteur arrête le cours de son énoncé au milieu d'un mot, reprend une partie de son énoncé et en corrige un élément :

HISTOIRE 3,14 comment on reconnaît un Belge **dans** un marchand de *chau- chez* un marchand de *chaussures*

L'amorce est bien complétée mais une modification intervient sur un élément qui a déjà été prononcé. L'amorce n'est donc pas le lieu de la réparation même si elle « annonce » dans ces cas une modification. Ces achoppements ne sont pas des amorces modifiées puisqu'en fait elles seront complétées après qu'une partie de l'énoncé antérieur ait été remplacé. Parfois, la modification antérieure consiste à ôter un élément présent dans l'énoncé précédent. Cette catégorie est rare ; 9 cas seulement sur les 1 000 amorces étudiées :

PUGET-VI 28,9 mais souvent c'est que + tous les **petits en-** + tous les *enfants* c'est le désir qu'ils ont de jouer

Ces cas de modification antérieure sont réservés aux amorces complétées et ne sont pas fréquents ; ils sont au nombre de 44 dans le corpus *Corpaix* et représentent 4,6% du total des amorces (44/948) et 7,7% (44/565) des amorces complétées. Rien ne s'opposerait à ce que ce type de phénomène soit observé également dans le cas des amorces modifiées mais, en fait, nous n'en avons observé aucun. Par ailleurs, ces résultats incitent à étudier jusqu'où le locuteur peut remonter dans son énoncé et produire une réparation. Il ne s'agit pas de déterminer un empan de la réparation antérieure mais de voir si cette modification "rétroactive" dépasse les limites du syntagme engagé. On s'aperçoit que sur les 44 amorces de cette catégorie, seules quatre modifications concernent un syntagme antérieur différent de celui où se trouve l'amorce "annonciatrice". Encore constate-t-on qu'éléments amorcés et éléments modifiés sont tous (sauf un) dans la rectification du même verbe :

BEAUMETT 2,15 quand les **gens** vont à *l'arm-* les **hommes** vont à *l'armée*

2.2.2. Les insertions d'éléments à l'occasion de phénomènes d'amorce

Dans d'autres cas encore, l'amorce n'est pas le lieu-cible de la réparation. Elle est suivie d'une reprise d'une partie de l'énoncé dans lequel se trouvent insérés un ou plusieurs éléments :

BENOT 3,13 tu vois + après nous *on li- on les livre* comme ça

Cette sous-catégorie n'est pas négligeable puisqu'elle accompagne un tiers des amorces (29,5%), toutes catégories confondues. De plus, ces insertions ne se produisent pas avec la même fréquence selon les types d'amorces : 71,2% sur les amorces complétées (210/295), le cinquième sur les inachevées (19% ; 56/295) et 9,8% sur les modifiées (29/295). La comparaison des pourcentages d'amorces complétées, inachevées et modifiées dans le grand corpus *Corpaix* avec la répartition de ces insertions selon ces types d'achoppements montre une différence significative ($\text{Khi}^2 = 13,58$; d.d.l. = 2 ; $p < .01$). Les insertions sont plus fréquentes lors d'amorces complétées et moins fréquentes lors d'amorces modifiées. Ces ajouts à l'occasion d'interruptions de mots sont particulièrement la marque d'une activité autonymique puisque les amorces concernées ne sont pas le lieu d'une modification mais annoncent une ré-élaboration sur ce qui a été énoncé antérieurement.

Nous avons distingué trois sortes d'insertions pouvant avoir lieu à l'occasion de ces achoppements : les insertions "énonciatives", les insertions qui "modulent" le texte et les incises parenthétiques. Les insertions "énonciatives" sont les plus nombreuses (62% ; 183/295) :

DESSININ 15,2 alors là on voit le l'aspect euh de la *com- enfin* l'organe de *commande*

viennent ensuite les insertions qui modulent (28,1%) :

PORTUGA4 10,3 c'est vraiment de voir les choses *très différen-* **d'une façon très différente**

et les incisives parenthétiques (9,8%) :

BARBARA 1,15 tu allais à la boucherie *si tu av-* **certaines boucheries** *si tu avais* de l'argent

L'analyse de la répartition de ces trois types d'insertions selon les types d'amorces montre des variations très nettes. Les insertions "énonciatives" constituent presque la moitié des insertions lors d'amorces complétées (47,6% ; 100/210).

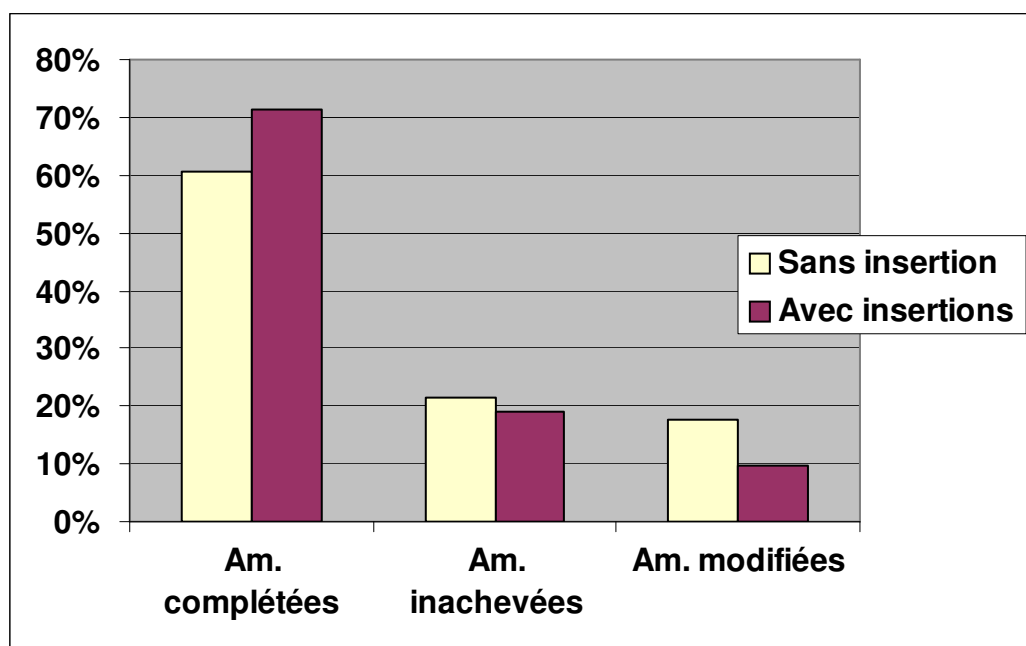


Figure 2. Répartitions des insertions d'éléments en fonction des types d'amorces.

Elle sont également les seuls ajouts pour les amorces inachevées ainsi que la quasi-totalité pour les amorces modifiées. La catégorie d'insertions non « énonciatives » (simples ajouts et incisives) semble donc réservée aux amorces complétées.

2.2.3. Le cas des amorces répétées

Dans une étude récente (Henry et Pallaud, 2003d), nous avons montré que les amorces répétées constituaient sur certains points une catégorie particulière à la fois d'amorces et de répétitions :

NAVALE 14,11 et que on le *re- re-* revendique encore une fois sur le cahier mais euh

Cette approche contrastive (au regard des amorces et des répétitions) permet de faire des hypothèses notamment sur une fonction probable de la répétition. En effet, on retrouve dans les amorces répétées les trois types d'amorces (complétées, inachevées et modifiées). La répartition de ces amorces est affectée significativement par la répétition ($\text{Khi}^2 = 29,85$; d.d.l. = 2 ; $p < .001$) : en effet, les amorces répétées complétées sont de loin les plus nombreuses (82%) et ce pourcentage est significativement plus élevé que dans le cas des amorces en général. Par ailleurs, les amorces inachevées sont deux fois moins nombreuses (13,3%) et les amorces modifiées sont quatre fois moins fréquentes (4,7%).

Ces résultats attestent de la relation de dépendance qui existe entre la catégorie d'amorce, quelle qu'elle soit, et le phénomène de répétition. Cela suggère une fonction de la répétition dans cet achoppement que constitue le phénomène d'amorce : elle soutient la recherche lexicale manifestée par l'amorce et contribue à l'aboutissement de cette recherche puisque, finalement, le mot interrompu est, alors, plus souvent complété et beaucoup moins souvent modifié. Cette dépendance entre répétitions et amorces suggère également que ces répétitions d'amorces pourraient être équivalentes à des pauses remplies ou silencieuses.

3. Place de l'amorce et de la répétition dans le syntagme

Blanche-Benveniste (2003) soutient que la répétition s'attache plus souvent au début du syntagme. Sandrine Henry (2002) confirme cette hypothèse. Qu'en est-il pour l'amorce ? Intervient-elle plus tardivement ou non que la répétition dans le syntagme ? Par ailleurs, les types d'amorces touchent-ils de façon semblable la zone de texte avant ou après le verbe ?

3.1. Amorces, répétitions et zones de texte.

Même dans les cas où on ne peut interpréter avec certitude toutes les amorces, leur place syntaxique est connue plus sûrement. Comme le soulignait Jeanjean (1984, p.171) qui conteste le terme de *ratés* pour ces phénomènes : "*Pour assigner une valeur syntaxique à ce genre d'amorces, on n'a pas besoin de reconstituer un quelconque lexème*". Nous nous sommes intéressées à la façon dont amorces et répétitions se situent dans le texte de l'énoncé : sur le verbe ou dans les zones pré- ou post-verbales (Figure 3).

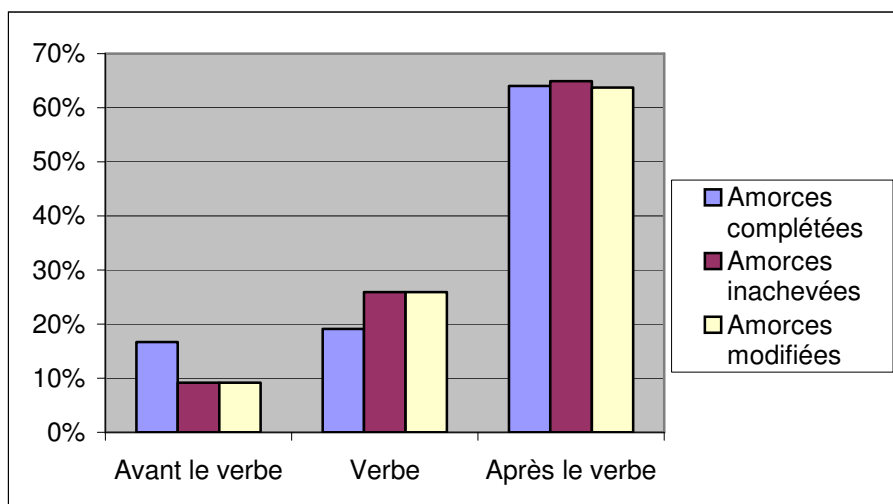


Figure 3. Types d'amorces et zones de texte (verbe, pré- et post-verbale)

La comparaison des localisations occupées par les amorces de mots (Pallaud, 2002a) et les répétitions⁵ a suggéré une dépendance⁶ entre le type d'achoppement et la zone de texte où les hésitations se situent. Si l'on admet que les zones-textes pré- et post-verbales sont de taille semblable, ces achoppements apparaissent tous deux plus fréquemment dans la zone post-verbale (64% pour les amorces et 54% pour les répétitions). En revanche, dans la zone verbale, les amorces sont plus fréquentes que les répétitions (21,9% vs 11,3%) ; dans la zone pré-verbale, les fréquences sont inverses : 13,9% pour les amorces et 33,8% pour les répétitions. Ce résultat suggère que, dans la zone pré-verbale, les répétitions sont plus

⁵ Corpus 2000-EDF : 15 751 mots (393 hésitations réparties : 8 sur place "associée").

⁶ $\text{Khi}^2 = 74,86$; d.d.l. = 2 ; $p < .001$; il reste pour être affirmatif à tenir compte de la taille des zones-textes avant et après le verbe, ce qui n'a pas été encore fait.

fréquentes que les amorces. Comparativement aux amorces, les répétitions qui se produisent plus tôt dans le syntagme, sont observées, également, plus tôt dans l'énoncé.

Une étude a été entreprise pour évaluer si les trois types d'amorces touchaient de façon similaire les zones verbales, pré-verbales et post-verbales (Figure 3). Qu'ils soient complétés, inachevés ou modifiés, lorsque les éléments amorcés sont à la droite du verbe, ils représentent un pourcentage semblable : 63%. Sur les zones verbale et pré-verbale, il existe une différence significative ($\text{Khi}^2 = 14,25$; d.d.l. = 2 ; $p < .001$.) entre les amorces complétées et celles qui ne le sont pas. Une amorce sur quatre touche le verbe et sera modifiée ou restera inachevée alors que l'amorce complétée n'affecte un verbe que dans un cas sur cinq. En revanche, celle-ci concerne la zone avant le verbe dans deux fois plus de cas que l'amorce inachevée ou modifiée. Ainsi, parmi les amorces, celles qui sont complétées sont plus fréquemment situées sur la zone pré-verbale, et moins nombreuses en zone verbale. Lorsque les amorces sont en zone post-verbale, les proportions de chacune des catégories sont semblables.

3.2. Les catégories grammaticales touchées par les amorces

Notre étude (Pallaud et Henry, 2003c) a montré que les unités lexicales ne sont pas affectées de la même manière par les amorces et les répétitions. Il apparaît utile de distinguer les mots qui ont une charge lexicale pleine, les *mots pleins*, de ceux qui participent de la structuration de la langue : les *mots-outils*. Les résultats obtenus pour les répétitions (Corpus *Corpaix* : 15 985 répétitions) et les amorces (999 amorces analysées) ont révélé une différence très nette. Les répétitions touchent plus souvent (91,3%) les mots-outils que les amorces qui affectent plus fréquemment les mots pleins (67,9%).

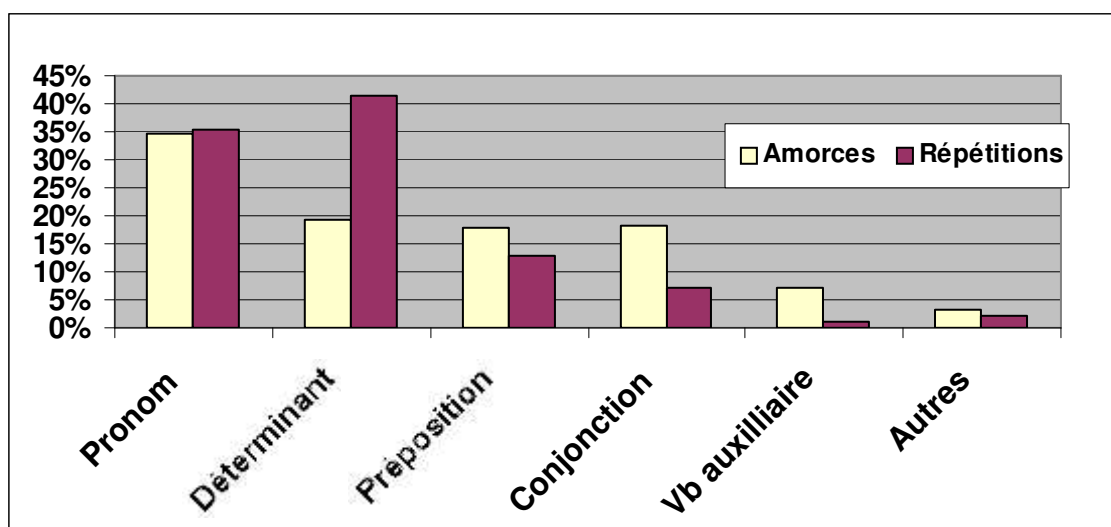


Figure 4. Types de mots-outils dans les amorces de mots et les répétitions.

Sur la totalité des amorces, celles qui concernent des "petits mots" ou des pronoms sont de loin les moins nombreuses (22,1%). Que cela soit dû à un lien entre la taille du mot et le phénomène d'amorce resterait à vérifier. Dans les syntagmes nominal et verbal, les mots pleins ne sont le plus souvent qu'en deuxième position sur l'axe syntagmatique. Les pronoms ou les déterminants les précèdent dans la très grande majorité des cas. Les répétitions affecteraient donc, plus tôt que les amorces, le syntagme nominal ou verbal.

Parmi les mots-outils, deux catégories grammaticales sont majoritairement concernées lors de la répétition : il s'agit des déterminants et des pronoms (respectivement 41,5% et 35,5%). Viennent ensuite les prépositions et locutions prépositionnelles (13%), ainsi que les

conjonctions et locutions conjonctives (7%). En ce qui concerne les amorces, la répartition est différente ($\text{Khi}^2 = 25,46$; d.d.l. = 2 ; $p < .001$.): la catégorie des pronoms est la plus touchée avec 34,8%. Les amorces affectent deux fois moins les déterminants mais deux fois plus les conjonctions et locutions conjonctives. Henry (2002a) a trouvé sur les formes *le*, *la* *les* et *leur* que les déterminants font nettement plus souvent l'objet d'une répétition que les pronoms clitiques compléments. Ces résultats et les précédents encouragent à conclure que la répétition s'attache plus souvent au début du syntagme alors que l'amorce qui touche plus fréquemment le lexique affecte le syntagme en un point plus avancé.

Conclusion

Les amorces de mots comme les répétitions ont en commun une suspension de l'énoncé sur une place syntaxique. Ces deux phénomènes, tous deux fréquents à l'oral, sont loin d'être des procédés énonciatifs, tant ils échappent, le plus souvent, à la perception même des locuteurs et de leurs « récepteurs » (Pallaud, 2003a, Roubaud, 2003). Leur fréquence comme leur catégorie peut varier d'un locuteur à l'autre. L'exploration des effets en contexte des phénomènes d'amorces montre que ces trébuchements sur un morphème ne constituent pas « l'erreur » elle-même mais l'indice qu'une modification va intervenir sur l'énoncé déjà prononcé. Dans moins d'un cas sur cinq, le morphème interrompu fera l'objet d'une modification. Tous les autres cas montrent que le morphème est repris pour être complété (60%) ou abandonné (22%).

Par ailleurs, dans un tiers des cas, toutes catégories d'amorces confondues, ces achoppements s'accompagnent d'une insertion ou d'une reprise de l'énoncé antérieur qui sera alors modifié. Le type d'amorce très majoritairement touché est l'amorce qui sera finalement complétée (74,9%). Outre que le cas de l'amorce complétée est le plus fréquent, il semble, au regard des insertions constituer une catégorie particulière : il est le seul (à deux exceptions près) à s'accompagner d'insertions qui modulent ou d'incises parenthétiques. Ces résultats suggèrent que l'usage du terme de *reparandum* (Shriberg, 1999) pour qualifier le morphème interrompu ou le terme qui va être répété est sans doute inapproprié. Dans la plupart des cas, ce n'est pas sur cet élément que va avoir lieu une modification.

Si on examine la localisation des achoppements sur les zones verbales, pré- et post-verbales, on observe des différences entre amorces et répétitions mais également entre catégories d'amorces. Comparativement aux amorces, les répétitions qui se produisent plus tôt dans le syntagme, sont observées, également, plus tôt dans l'énoncé. Parmi les amorces, celles qui sont complétées sont plus fréquemment situées sur la zone pré-verbale, et moins nombreuses en zone verbale. Lorsque les amorces sont en zone post-verbale, les proportions de chacune des catégories sont semblables.

Les répétitions touchent, bien plus souvent que les amorces, les mots-outils tandis que les amorces affectent, plus fréquemment que les répétitions, les mots pleins. L'examen des catégories grammaticales affectées par les répétitions et les amorces nous semble permettre de conclure que la répétition s'attache plus souvent au début du syntagme alors que l'amorce qui touche plus fréquemment le lexique affecte le syntagme en un point plus avancé.

Références

Blanche-Benveniste, C., 2003, La naissance des syntagmes dans les hésitations et les répétitions du parler. IN : Araoui J.L. (Ed.), *Le sens et la mesure. Hommages à Benoît de Cornulier.* , Editions Honoré Champion, p 40-55.

- Cappeau P., 1998, Quelques mots sur quelques bribes liées au genre. In: Bilger M, Van den Eynde & Gadet F, (Eds) *Analyse linguistique et approches de l'oral. Recueil d'études offert en hommage à Claire Blanche-Benveniste*. Peeters, Leuven, Paris, 301-311.
- Duez D., 2001 b, Signification des hésitations dans la production et la perception de la parole spontanée., *Revue Parole*, 17-18-19, p 113-138.
- Eklund R., 2003 *Disfluency in Spontaneous Speech Workshop*, Proceedings of DiSS'03, 5-8 September 2003, Göteborg University, Sweden.
- Henry S., 2002, Étude des répétitions en français parlé spontané pour les technologies de la parole, *Actes de la 6ème Rencontre (RECITAL'02)*. Nancy (France) : ATALA, 467-476.
- Henry S. & Pallaud B., 2003c, Amorces de mots et répétitions dans les énoncés oraux, *Recherche Sur le Français Parlé*, 18, à paraître
- Henry S. & Pallaud B., 2003d, Word fragments and repeats in spontaneous spoken French. In: Eklund R., *Disfluency in Spontaneous Speech Workshop*, Proceedings of DiSS'03, 5-8 September 2003, Göteborg University, Sweden, p 77-80.
- Jeanjean C., 1984, Les ratés c'est fa- fabuleux. Étude syntaxique des amorces et des répétitions., *Linx*, 10, 171-177.
- Martinie B., 2000, Remarques sur la syntaxe des énoncés réparés en français parlé. *Recherches sur le Français parlé*, 16, 189-206.
- Pallaud B., 2002a, Les amorces de mots comme faits autonymiques en langage oral, *RSFP*, 17, 79-102.
- Pallaud B., 2003 a, Erreurs d'écoute dans la transcription de données orales. Actes du colloque à Tours, 8-9 décembre 2000, *Transcription de la Parole normale et pathologique.*, revue *Parole*, 22-23-24, p 267-294.
- Roubaud M-N, 2003, Du bon usage des amorces dans la transcription de corpus. *RSFP*, 18 (à paraître).
- Schegloff E., Jefferson H. and Sachs H., 1977, The preference for self-correction in the organization of repair in conversation., *Language*, 53, 2, p 351-382.
- Shriberg E. (1999). Phonetic Consequences of Speech Disfluency. Symposium on The Phonetics of Spontaneous Speech (S. Greenberg and P. Keating, organizers), *Proc. International Congress of Phonetic Sciences*, vol. 1, pp. 619-622, San Francisco.