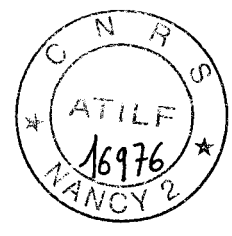


7111

CV03
11

Collection
recherches sur le français parlé 18

AUTOUR DU CORPUS DE RÉFÉRENCE DU FRANÇAIS PARLÉ



2004

PUBLICATIONS DE L'UNIVERSITÉ DE PROVENCE

© PUBLICATIONS DE L'UNIVERSITÉ DE PROVENCE

29, avenue Robert Schuman - F - 13621 Aix-en-Provence cedex 1
Tél. + 33 (0) 4 42 95 31 91 – Fax + 33 (0) 4 42 20 28 04
Courrier électronique pup@up.univ-aix.fr
Catalogue complet sur internet www.up.univ-mrs.fr/wpup

Sommaire

Présentation	7
<i>Paul CAPPEAU</i>	
Présentation du <i>Corpus de référence du français parlé</i>	11
<i>Équipe DELIC</i>	
Les compléments de lieu réalisés par <i>y</i> : description des usages.....	43
<i>Frédéric SABIO</i>	
Quelques données sur les adverbes en <i>ment</i> dans le corpus de référence de français parlé.....	63
<i>Mireille BILGER</i>	
✓ Fréquence des marqueurs discursifs en français parlé : quelques problèmes de méthodologie	83
<i>Catherine CHANET</i>	
Les formes disjointes des pronoms sujets	107
<i>Paul CAPPEAU</i>	
Quelques éclairages sur le sujet type <i>moi je</i> à l'oral	127
<i>Mylène BLASCO-DULBECCO</i>	
<i>Lui le propriétaire, le propriétaire, lui</i> : deux constructions bien distinctes	145
<i>Sandrine CADDÉO</i>	
Du bon usage des amorces dans la transcription des corpus	163
<i>Marie-Noëlle ROUBAUD</i>	
✓ Étude des interactions entre pauses silencieuses et pauses remplies en français parlé	185
<i>Estelle CAMPIONE</i>	
Amorces de mots et répétitions dans les énoncés oraux	201
<i>Sandrine HENRY & Berthille PALLAUD</i>	

Compte-rendu de l'ouvrage de Norbert Dittmar : <i>Transkription.</i> <i>Ein Leitfaden mit Aufgaben für Studenten, Forscher und Laien</i> 231 <i>françois MULLER</i>	231
Du bon usage des notations phonétiques dans les corpus aphasiques 241 <i>Marie-Noëlle ROUBAUD</i>	241
Résumés de thèses	255

Présentation

Paul CAPPEAU
Université de Poitiers

On retrouvera dans cette nouvelle livraison de *Recherches sur le français parlé* des préoccupations déjà anciennes sur les corpus oraux, les descriptions syntaxiques originales qu'ils permettent d'envisager ainsi qu'un traitement de certains faits de productions comme les hésitations ou les bribes. Mais ce numéro apporte aussi son lot de nouveautés : il présente un nouveau corpus, fruit d'un travail de plusieurs années (le *Corpus de référence du français parlé* désigné comme CRFP), il élargit les analyses à de nouveaux domaines et il amorce l'approche thématique des prochains numéros.

Dans ce numéro 18, les articles ont été regroupés autour de trois axes majeurs : la présentation des données utilisées dans la description, des études de syntaxe, des faits liés à la production orale.

1. Les données

Le premier article (qui doit beaucoup à Jean Véronis) fournit des informations sur la constitution du *Corpus de référence du français parlé*. Il rappelle succinctement l'historique du projet : recueillir sur une période de temps limitée des enregistrements dans les principales grandes villes de France en tenant compte de paramètres d'âge, de niveau scolaire et de situations. Ce nouveau corpus se compose de 134 enregistrements qui représentent environ 440 000 mots transcrits. Il était donc important, voire indispensable, de fournir une claire vision des subdivisions internes de cette banque de données en tenant compte, par exemple, du nombre de locuteurs par enregistrement, de l'âge des intervenants, des situations de parole, etc. Enfin, ce premier article indique au lecteur intéressé comment accéder à ces données et les interroger !

La plupart des articles de ce numéro s'appuient sur cette nouvelle banque de données dans leurs analyses et intègrent certains de ces paramètres dans leur description.

2. Quelques points d'analyse syntaxique

Les articles qui adoptent une orientation syntaxique sont parcourus par deux préoccupations. Certains présentent des descriptions sur des points nouveaux (le pronom *y*, les adverbes, etc.), d'autres prennent prétexte de ce corpus inédit pour revenir sur des analyses antérieures ou s'intéresser aux décalages que provoque le recours à des banques de données différentes.

L'article de Frédéric Sabio est consacré à l'étude du clitique *y* dans sa valeur locative. À partir d'emplois relevés dans les données orales et dans de l'écrit de presse, l'auteur montre l'importance de plusieurs facteurs tels que le statut valenciel ou non du clitique *y* et le caractère plus ou moins formel de la production. Il peut ainsi distinguer trois grands cas dans la distribution. Cette analyse illustre bien la différence entre le *possible de la langue* qu'indiquent les grammaires et les *possibles du discours* que révèlent les données attestées.

Un deuxième article (celui de Mireille Bilger) confronte données orales et données de presse écrite pour mieux faire ressortir la distribution du phénomène étudié (les adverbes en *-ment*). L'auteur présente un inventaire des adverbes en *-ment* qui comportent plus de dix occurrences dans le CRFP. Elle tient compte des trois grandes situations de parole retenues (privée, professionnelle et publique) pour rechercher les formes communes et spécifiques à chaque situation et envisager des rapprochements avec l'écrit. Ce travail sur corpus permet aussi de dégager les lignes de force du fonctionnement syntaxique des quatre adverbes les plus abondants.

L'article de Catherine Chanet constitue une belle illustration des différentes étapes par lesquelles doit passer le chercheur qui travaille sur des productions orales. L'auteur s'intéresse aux marqueurs discursifs, classe jusque-là assez négligée dans cette revue. Elle doit d'abord, parmi la diversité des désignations utilisées notamment dans les approches pragmatiques, dégager les caractéristiques de ces unités. Elle montre ensuite comment le recours à un concordancier pour obtenir des données produit un résidu abondant qu'il faut savoir écarter, sur la base des

contextes syntaxiques. Ce travail fournit enfin un inventaire quantifié des marqueurs discursifs qui offre une vision inédite de cette classe.

Trois articles apportent des éclairages complémentaires sur le fonctionnement des pronoms conjoints liés au sujet. Paul Cappeau s'intéresse à la distribution des pronoms sujets disjoints. Son analyse tient compte des environnements syntaxiques possibles et de l'importance des données obtenues. Les données recueillies conduisent à s'interroger sur la pertinence de certaines subdivisions, à remodeler la classe de ces pronoms et à la présenter en quatre séries.

L'article de Mylène Blasco-Dulbecco porte sur la séquence *moi ... je*. Le CRFP permet à l'auteur de revenir sur certaines distributions. Son analyse tient compte, entre autres, du caractère lié ou non des éléments *moi* et *je* et de la position du pronom *moi* par rapport au verbe. Elle montre aussi que le comportement syntaxique de ce pronom (parfois négligé aux dépens d'une approche plus pragmatique) apporte son lot d'enseignements.

Sandrine Caddéo revient sur cette concurrence entre approche pragmatique et syntaxique dans l'étude des pronoms disjoints détachés. Elle utilise des exemples oraux et écrits pour dégager la distribution de deux ordres distincts (selon que le pronom détaché précède ou suit la tête nominale du syntagme). L'article s'appuie sur des critères syntaxiques et sémantiques pour montrer une différence de fonctionnement importante que d'autres langues romanes permettent de conforter.

3. Les faits de production orale

Les trois derniers articles de ce numéro portent sur des points étroitement liés à la production orale et sont conduits à discuter les choix de transcription des corpus.

Le premier article de Marie-Noëlle Roubaud revient sur les difficultés de transcription. Elle montre (en prenant appui sur l'important travail de révision du CRFP) les différentes stratégies que le transcrip-teur met en œuvre lorsqu'il doit identifier des amorces. Il est utile de rappeler combien les transcriptions demandent un soin vigilant et les diverses solutions entre lesquelles le transcrip-teur doit naviguer.

Estelle Campione revient sur la distinction entre pauses silencieuses et pauses remplies. A partir d'une sélection d'enregistrements du CRFP, elle présente une description du fonctionnement et de l'emplacement des pauses remplies. Elle s'intéresse aussi aux conditions dans lesquelles les

pauses silencieuses peuvent jouer le rôle de marque d'hésitation ou de reformulation.

Enfin l'article de Sandrine Henry et de Berthille Pallaud propose deux regards complémentaires sur les amorces et les répétitions. Cette étude de ces faits d'achoppement s'appuie sur des données quantitatives nombreuses. Elle démontre leur caractère organisé et l'importance de facteurs syntaxiques qui éclairent la description de ces phénomènes.

Enfin, signalons que, dans le numéro 17, une erreur technique a rendu l'article de Marie-Noëlle Roubaud peu compréhensible. Sa republication dans ce numéro s'imposait !

Présentation du *Corpus de référence du français parlé*

Équipe DELIC

Résumé

Le *Corpus de référence du français parlé*, qui vient d'être réalisé par l'équipe DELIC, vise à mettre à la disposition de la communauté des linguistes, chercheurs et enseignants, un témoignage de la langue française parlée aujourd'hui dans les principales villes de l'hexagone. Le corpus comporte 440 000 mots, correspondant à plus de 36 heures de parole. Il est composé de 134 enregistrements échantillonnés en fonction de plusieurs situations de parole et de niveaux d'études des locuteurs. La totalité du corpus se présente sous une forme transcrite alignée avec le son.

1. Introduction

Le *Corpus de référence du français parlé* que nous présentons ici (et qui a été utilisé dans la plupart des articles de ce numéro) répond à la 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'équipe Corpus de l'Université de Provence, dirigée par Claire Blanche-Benveniste et associée au CNRS. A partir de 2000, le projet a été pris en charge par l'équipe DELIC (DEscription Linguistique Informatisée sur Corpus), dirigée par Jean Véronis.

L'objectif de ce corpus est de mettre à la disposition de la communauté des linguistes, chercheurs et enseignants, un témoignage de la langue française parlée aujourd'hui dans les principales villes de l'hexagone. Il s'agissait avant tout de recueillir des données représentatives d'un français parlé que nous pourrions qualifier d'« usage général et

Équipe DELIC

courant », ce qui nous a amenés à effectuer certains choix touchant aussi bien aux caractéristiques des locuteurs qu'aux situations de parole.

Ce corpus compte environ 440 000 mots et il est constitué de 134 enregistrements dont la partie transcrite correspond à une durée moyenne de 16 min. 48 s¹. La totalité représente 36 heures et 50 minutes de parole. Dans le projet initial, il était prévu que le corpus comporte 160 enregistrements, mais les problèmes de coordination et d'harmonisation, dus à la multiplicité des régions géographiques, des enquêteurs et des transcrip-teurs, avaient été sous-estimés. Nous n'avons pas voulu sacrifier la qualité du résultat, ni accroître exagérément les délais de réalisation, car ce projet représente une occasion unique de fournir à la communauté de recherche francophone un outil scientifique important, qui lui fait cruellement défaut, alors que des corpus oraux importants existent ou sont en cours de réalisation pour la plupart des langues européennes. Pour le français parlé, il offre une base de comparaison avec les corpus de français parlé hors hexagone, le corpus Valibel en Belgique et le corpus d'Ottawa-Hull au Canada.

2. Échantillonnage

Nous décrivons ici les principes d'échantillonnage qui ont présidé à la construction du corpus, et nous fournissons un certain nombre de données quantitatives qui permettent de se faire une idée de la diversité des données disponibles².

2.1. Répartition géographique

Les données ont été recueillies dans 37 villes de province (27 dans la zone nord et 20 dans la zone sud), et en région parisienne. Les villes de province sont des villes de dimension moyenne comme Perpignan ou Pau, ou beaucoup plus grandes comme Bordeaux ou Lyon. Seules deux villes ont moins de 10 000 habitants (Corte et Saint-Affrique) (Figure 1). La

¹ Les enregistrements étaient généralement plus longs mais seule une partie a été transcrite. Cette partie n'est pas nécessairement le début de l'enregistrement, de façon à éviter des fragments liés à la mise en place de l'interaction (« bon, alors, ça enregistre ? », « alors, de quoi je parle ? », etc.).

² Les critères d'échantillonnage adoptés pour la constitution de ce corpus ont été élaborés par Mireille Bilger.

Présentation du *Corpus de référence du français parlé*

région parisienne a été divisée en 5 zones : Paris-centre et les secteurs nord-ouest, nord-est, sud-ouest et sud-est.

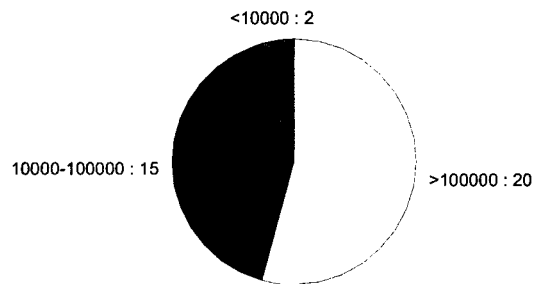


Figure 1. Répartition des villes de province en fonction de la population

Au total, 107 enregistrements ont été effectués en province (50 dans la zone nord, 57 dans la zone sud) et 27 enregistrements dans la région de Paris et sa banlieue (Figure 2). Nous obtenons une moyenne de 3 enregistrements par ville, sauf pour la région parisienne pour laquelle nous avons envisagé dès le début un plus grand nombre d'enquêtes ; chaque secteur compte entre 5 et 6 enregistrements (Figure 3).

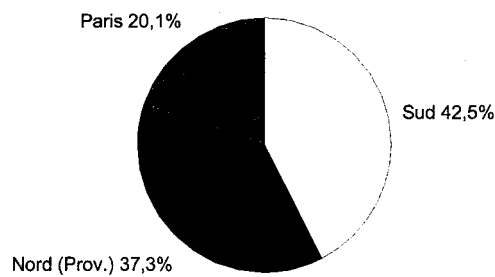


Figure 2. Répartition géographique des enregistrements

Équipe DELIC

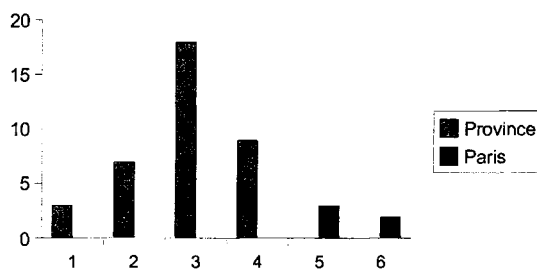


Figure 3. Nombre d'enregistrements par ville

2.2. Taille des transcriptions

Comme il a été dit en introduction, la durée moyenne des transcriptions est de 16 min. 48 s³. La plupart des transcriptions se situent autour de cette valeur, mais quelques transcriptions sont notablement plus courtes ou plus longues (de 6 min. 45 s. à 31 min. 16 s.). La Figure 4 donne la distribution des durées des transcriptions.

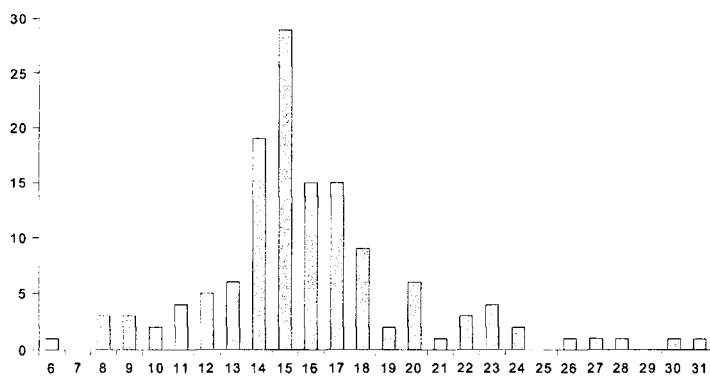


Figure 4. Distribution des durées des transcriptions

³ Les valeurs que nous donnons tiennent compte des interruptions du discours (musique, déplacement du locuteur, etc.), dont la durée a été déduite de la durée totale de l'enregistrement.

2.3. Répartition selon la situation de parole

Au critère géographique s'ajoute la prise en compte de la situation de parole. Nous avons défini trois situations d'enregistrement :

- **la parole privée** : sous la forme d'un entretien sollicité spécifiquement dans le cadre de l'enquête, cette situation de parole renvoie à deux types de productions : le récit de vie (dont le contenu peut varier : récit d'un voyage, d'une expérience, souvenirs d'enfance, etc.), ou la présentation d'un « savoir-faire » professionnel ou autre.
- **la parole professionnelle** : entretiens également sollicités spécifiquement, mais dans lesquels les locuteurs ont été enregistrés dans l'exercice de leur fonction ou quand ils parlent de leur profession sur leur lieu de travail.
- **la parole publique** : cette situation se distingue des deux autres par le fait que les intervenants s'expriment toujours en présence d'un public ; elle comporte une partie d'entretiens sollicités, le reste étant constitué d'émissions radiophoniques (actualités, interview, table ronde, tribune téléphonique, etc.), de cours et conférences, de réunions politiques ou associatives (conseil municipal, discussion syndicale, comité de quartier, etc.), et de quelques situations plus spécifiques (visite de musée, dégustation de vins, etc.).

Remarquons que la distinction entre les types « privé » et « professionnel » n'est pas subordonnée à une différence de contenu, puisque le fait pour un locuteur de parler d'un savoir-faire professionnel est représenté dans l'un comme dans l'autre type : simplement, on a tenu à considérer de manière indépendante les échanges qui se situent sur le lieu professionnel dans la mesure où la parole des locuteurs peut y prendre une forme plus « institutionnelle ».

La répartition sur l'ensemble du corpus est la suivante :

Type	Enregistrements		Mots	
<i>Privé</i>	84	62,7%	282857	64,5%
<i>Prof.</i>	22	16,4%	75001	17,1%
<i>Public</i>	28	20,9%	80601	18,4%
Total	134	100,0%	438378	100,0%

On voit que la parole « privée » représente environ les deux tiers du corpus, tandis que les types « professionnel » et « public » se partagent le tiers restant à parts à peu près égales (Figure 5).

Équipe DELIC

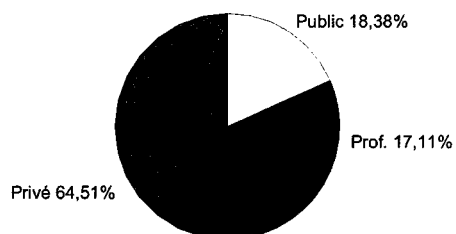


Figure 5. Répartition par type de parole (en nombre de mots)

La parole « publique » se décompose de la façon résumée par la Figure 6. On notera que les enregistrements de parole « publique » comportent trois entretiens-enquêtes réalisés en public. Ces entretiens s'ajoutent à ceux des sous-corpus de parole privée (84 entretiens) et professionnelle (22 entretiens) ; au total, le corpus comprend donc $84 + 22 + 3 = 109$ entretiens, ce qui représente 81% des enregistrements.

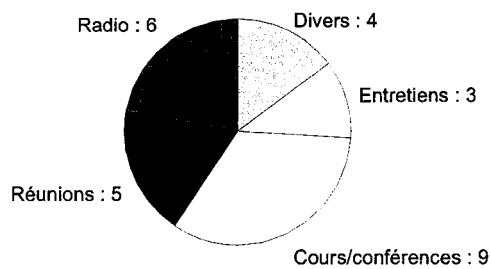


Figure 6. Situations de parole publique

2.4. Nombre de locuteurs

Le nombre de locuteurs⁴ s'échelonne entre 1 et 18, selon la répartition donnée par la Figure 7. Le nombre moyen de locuteurs est de 2,7 ; la situation la plus fréquente fait intervenir 2 locuteurs.

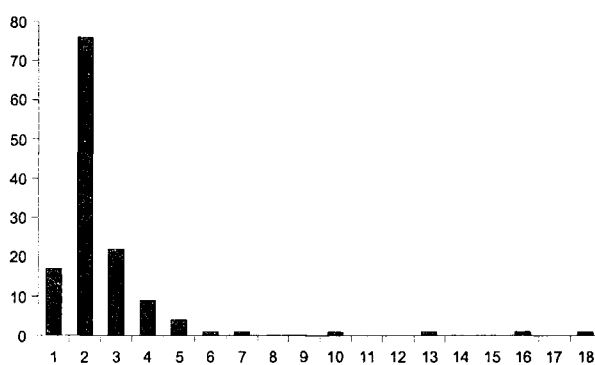


Figure 7. Répartition par nombre de locuteurs

Si ce paramètre permet d'apprécier la part respective de monologues (17%) ou de situations faisant intervenir des locuteurs nombreux (19% des transcriptions font intervenir 4 locuteurs et plus), il reflète mal la répartition réelle de la parole. Par exemple, une situation à deux locuteurs peut être une situation d'entretien dans laquelle l'un des locuteurs a un usage prépondérant de la parole, ou une situation de discussion, où les prises de parole sont plus équilibrées. De même, une situation peut faire intervenir trois locuteurs, mais l'un d'entre eux peut avoir un rôle extrêmement marginal (par exemple traverser la pièce et dire bonjour).

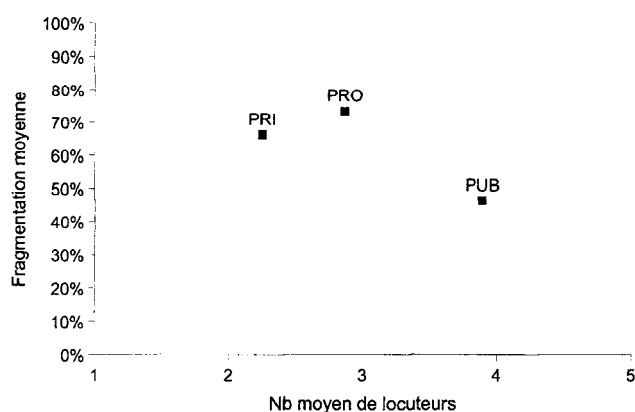
Nous avons donc défini un autre paramètre, de fragmentation de la parole, qui permet de mieux quantifier la répartition de la parole entre locuteurs. Le taux de fragmentation varie entre 0 (parole totalement monopolisée par un locuteur ou monologue) et 100% (parole fragmentée de

⁴ Dans la partie transcrite. Il peut arriver que d'autres locuteurs interviennent dans une autre partie de l'enregistrement. Ils n'ont pas été comptabilisés.

Équipe DELIC

façon maximale entre locuteurs, dans le cas théorique de prises de parole en nombre égal et de même longueur)⁵.

La Figure 8 représente le nombre moyen de locuteurs et le taux de fragmentation moyen dans chacun des sous-corpus. On notera le taux important de fragmentation des entretiens : on aurait pu attendre des situations asymétriques, où la parole revient de façon prépondérante au locuteur interviewé, mais on s'aperçoit que le taux de fragmentation est important, la plupart des entretiens se situant dans la zone de 75% à 100%, qui indique que l'enquêteur intervient largement dans la discussion⁶. En ce qui concerne le sous-corpus « public », un examen plus approfondi des données révèle une distribution bi-modale, à mettre en relation avec son caractère composite, comportant à la fois des situations de parole très fragmentées (réunions) et des situations proches du monologue (par exemple cours).



⁵ Ce paramètre se calcule de la façon suivante :

$$f = 1 - \frac{\sigma/m}{\sqrt{N-1}}$$

N étant le nombre de tours de parole, m le nombre moyen de mots par tour de parole et σ l'écart-type correspondant.

⁶ Le taux de fragmentation ne mesure que la part relative de la parole, d'un point de vue strictement quantitatif, sans préjuger du contenu de l'interaction. On notera par exemple, de ce point de vue, que la transcription la plus fragmentée (PRI-SAI-1) est un entretien, dans lequel les interventions de l'enquêteur sont presque exclusivement composées de « oui oui oui », « ah d'accord », etc.

Figure 8. Fragmentation de la parole

2.5. Age, sexe et niveau scolaire des locuteurs (entretiens)

Ces différents paramètres n'ont pu être contrôlés que pour les entretiens sollicités dans le cadre de l'enquête. Les indications suivantes portent donc sur les 109 enregistrements de cette catégorie, en ne considérant que le locuteur principal (l'interviewé).

Les tranches d'âge se répartissent ainsi (voir aussi Figure 9) :

Age	Transcriptions
18-30 ans	33
30-65 ans	55
+ 65 ans	21
Total	109

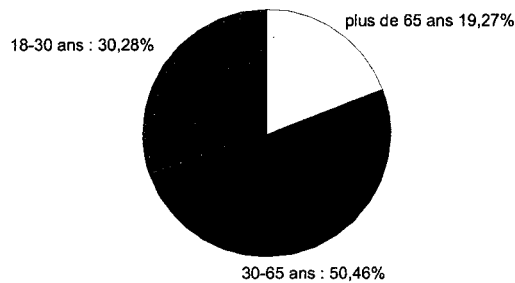


Figure 9. Répartition des entretiens par tranche d'âge

La répartition par sexe ne faisait pas partie des critères initiaux, mais elle a pu être contrôlée a posteriori. Les interviewés se répartissent en 44% de femmes et 56% d'hommes (Figure 10).

Équipe DELIC

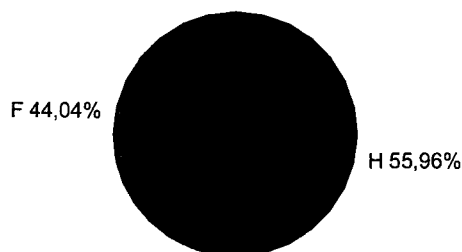


Figure 10. Répartition des entretiens par sexe

Trois tranches ont été définies en ce qui concerne le niveau scolaire des locuteurs : niveau primaire ou collège, niveau bac ou niveau enseignement supérieur (à partir de Bac+3). La répartition se présente ainsi (voir aussi Figure 11) :

Niveau	Transcriptions
Inconnu	6
Primaire-Collège	27
Bac	45
Supérieur	31
Total	109

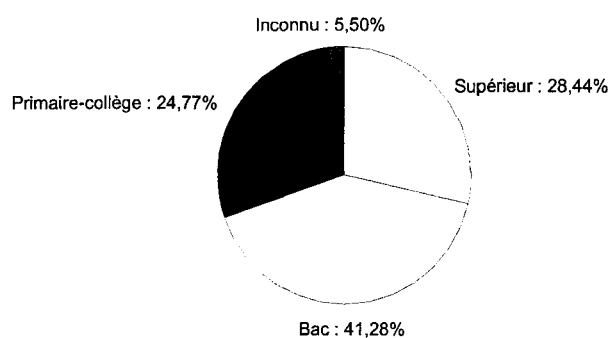


Figure 11. Répartition des entretiens par niveau scolaire

3. Recueil et transcription

3.1. Les enregistrements

Les enregistrements ont été effectués à l'aide de « baladeurs » mini-disques, sans faire appel à des microphones bidirectionnels ou à des micro-casques, ce qui a néanmoins permis une qualité sonore correcte dans la plupart des cas. Le but du corpus étant de recueillir du français non lu dans un cadre le moins contraint possible, la qualité sonore n'est pas forcément celle que peuvent exiger, par exemple, des études phonétiques précises. Les enregistrements ont été catégorisés en trois niveaux de qualité : A (excellent), B (bon) et C (passable). Le tableau ci-dessous montre la répartition dans le corpus :

Qualité	Enregistrements	
A	104	77,6%
B	16	11,9%
C	14	10,4%
Total	134	100,0%

Tableau 1. Répartition de la qualité des enregistrements du corpus

La numérisation, la délimitation des parties à transcrire ainsi que le réglage du volume ont été réalisés grâce au logiciel Sound Forge XP de la société Sonic Foundry⁷.

3.2. La fiche signalétique

Chaque transcription est accompagnée d'une fiche signalétique qui donne les informations suivantes :

- identificateur de la transcription ;
- titre ;
- lieu d'enregistrement ;
- responsable de l'enregistrement ;
- transcripteur ;
- réviseur(s) ;
- type de parole ;
- situation ;

⁷ <http://www.sonicfoundry.com/>

Équipe DELIC

- résumé ;
- nombre de locuteurs ;
- numéro du locuteur principal (interviewé), le cas échéant ;
- nombre de mots ;
- durée ;
- particularités.

De plus pour chaque locuteur, la fiche donne un certain nombre de renseignements :

- numéro du locuteur
- âge
- sexe
- niveau d'études
- profession
- rôle (enquêteur, interviewé, etc.)

L'Annexe 1 donne un exemple de fiche signalétique.

3.3. La transcription

Les conventions utilisées pour la transcription du *Corpus de référence du français parlé* sont largement inspirées de celles mises au point par le G.A.R.S. (cf. Blanche-Benveniste et al., 1991: 228-229). Ces conventions préconisent une représentation en orthographe standard, sans ponctuation, ni indications prosodiques, signes démarcatifs ou expressifs. Les pauses sont notées, ainsi que les amorces (mots inachevés). On tient également à conserver les traces de certaines difficultés de transcription liées à des problèmes d'interprétation. Le ou les transcrip-teurs peuvent percevoir des séquences sonores différentes ou proposer plusieurs versions orthographiques pour une seule et même séquence sonore. Un certain nombre d'aménagements ont été apportés aux conventions du G.A.R.S., principalement pour des raisons techniques, en particulier pour éviter des choix liés à des logiciels de traitement de textes particuliers, et permettre une plus grande robustesse et une meilleure portabilité. Par exemple, les chevauchements, qui étaient notés par un soulignement :

- L1 de me penser à penser à trouver un travail - et l'imprimerie
- L2 vous aviez à peu près quel âge

sont désormais notés par des chevrons ouvrants et fermants :

- L1 de me penser à penser à trouver un travail + < et l'imprimerie
- L2 vous aviez à peu près > quel âge

Pour la même raison, les transcriptions phonétiques ponctuelles (mots mal prononcés, etc.) ne sont plus notées à l'aide d'une police particulière, mais font appel à l'alphabet SAMPA (cf. Annexe 3).

D'autres choix visent à une amélioration ergonomique du travail de transcription, en évitant des confusions possibles de la part du transcripteur. Ainsi, le tiret servait précédemment aussi bien à marquer les amorces que les pauses :

alors nous le euh - nous avons a- accepté de - de les prendre avec nous

Toutefois, ces conventions étaient « fragiles » dans la mesure où une espace mal placée pouvait totalement transformer le codage. Les pauses sont désormais notées par le signe + (plus) :

alors nous le euh + nous avons a- accepté de + de les prendre avec nous

L'annexe 2 donne la liste des transcriptions révisées.

La transcription a été réalisée à l'aide du logiciel *Transcriber* (Barras et al. 1996)⁸, qui a également permis de réaliser un alignement de la totalité du corpus avec le son. La segmentation a été faite aux pauses les plus importantes (la durée moyenne des segments est de 3,1 s. pour 9,6 mots). L'exemple ci-dessous montre un exemple d'alignement (la durée est exprimée en secondes) :

- L1 0.0 voilà nous nous étions interrompu au moment où nous avons
rencontré les deux Anglaises dont une était euh +
- 7.3 noire +
- 8.6 et qui était d'ailleurs la seule parlant français alors nous le euh
+
- 12.3 nous avons a- accepté de +
- 14.1 de les prendre avec nous nous nous sommes mis d'accord
pour partager les frais d'essence euh +
- 18.9 parce que c'est vrai que ça faisait hum +
- 22.5 ça fait l'éch- l'essence était très chère +
- 25.1 et euh +
- 27.3 quand on en trouvait +
- 29.4 et nous voilà partis donc euh à cinq +
- 33.4 le Dogon les deux Anglaises mon épouse et moi euh et euh +

⁸ Logiciel téléchargeable gratuitement sur le site :
<http://www.etca.fr/CTA/gip/Projets/Transcriber/>

Équipe DELIC

38.7 et six avec le chauffeur même +
41.1 nous nous n'oublions pas ce +
44.0 le chauffeur +

Cet alignement très fin permet d'écouter à volonté les passages souhaités lors de l'examen des concordances. Il constitue un nouveau mode d'exploration des corpus et permet une vérification immédiate de chaque passage de la transcription.

3.4. Problèmes juridiques et anonymisation

Le recueil et la diffusion de corpus oraux pose des problèmes délicats du point de vue juridique, et l'équipe a pris deux mesures permettant d'éviter des difficultés ultérieures⁹ :

- chaque personne enregistrée a signé une autorisation d'enregistrement et de diffusion, à l'exception des enregistrements radiophoniques pour lesquels une négociation est en cours ;
- tous les noms de tierces personnes, de sociétés ou de lieux pouvant poser difficulté (propos péjoratifs, etc.) ont été anonymisés aussi bien dans les enregistrements (remplacés par un bip sonore) que dans les transcriptions (remplacés par un code du type *P* pour les patronymes, etc.).

4. Exploitation et diffusion des données

Le corpus aligné peut être exploité à l'aide du concordancier *Contextes*, développé par Jean Véronis¹⁰, qui possède de nombreuses fonctionnalités, et permet d'écouter les segments correspondant à chaque ligne de concordance.

Trois modes de diffusion sont envisagés à l'heure actuelle¹¹ :

- une version libre sur le World Wide Web, permettant une extraction limitée de concordances¹², sans accès au son ;

⁹ L'équipe est reconnaissante à Sandra Teston pour ses recherches sur les questions juridiques.

¹⁰ <http://www.up.univ-mrs.fr/veronis/logiciels/Contextes>

¹¹ Ces dispositions peuvent changer en fonction des problèmes juridiques ou techniques rencontrés.

Présentation du *Corpus de référence du français parlé*

- une version orientée vers les enseignants et chercheurs en linguistique, distribuée sur CD-ROM comportant l'intégralité du son compressé au format WMA (Windows Media Audio), et permettant l'extraction de concordances à l'aide du logiciel *Contextes* ;
- une version destinée à la recherche sur les technologies de la parole, avec son au format WAV, et comportant les transcriptions au format XML utilisé par le logiciel *Transcriber*.

5. Conclusion

Il va de soi que, malgré l'effort considérable qu'a représenté sa constitution, le *Corpus de référence du français parlé* demande à être étendu : bien des situations de parole sont encore absentes, trop d'usages de la langue ne sont pas représentés. Néanmoins, en l'état actuel, le corpus se présente déjà comme une ressource majeure pour l'étude et l'enseignement du français. Par sa taille (440 000 mots, 36 heures d'enregistrement), et son large échantillonnage, il constitue une base de données comportant une variété importante d'accents, d'expressions, de tournures, qui donne une image linguistique précise de la France urbaine contemporaine. La présentation, dans laquelle le texte et le son sont alignés, offre aux utilisateurs de nouvelles possibilités d'exploitation et d'exploration du corpus.

Références

- Barras, C., Geoffrois, E., Wu, Z. & Liberman, M. (2000). *Speech Communication* (Special issue on Speech Annotation and Corpus Tools). 33(1-2) : 5-22.
- Blanche-Benveniste, C., Bilger, M., Rouget, C. & van den Eynde, K. (1991). *Le français parlé. Études grammaticales*. Paris : CNRS Éditions.

¹² <http://www.up.univ-mrs.fr/delic/crfp>

Équipe DELIC

Annexe 1. Exemple de fiche signalétique

Identificateur	PRI-AIX-1
Titre	Afrique
Lieu d'enregistrement	Aix-en-Provence
Responsable de l'enregistrement	André Valli
Transcripteur	Christophe Rey
Réviseur	André Valli
Type de parole	Privée
Situation	Entretien
Résumé	Récit d'un voyage au Mali
Nombre de locuteurs	2
Locuteur principal	L1
Nombre de mots	2903
Durée	00:15:17
Particularités	Néant

Locuteurs

L1	Age	30-65
	Sexe	H
	Niveau d'études	Bac
	Profession	Retraité
L2	Rôle	Interviewé
	Age	30-65
	Sexe	H
	Niveau d'études	Supérieur
L2	Profession	Professeur des universités
	Rôle	Enquêteur

Annexe 2 – Conventions de transcription

1. Principes généraux

La transcription se fait en orthographe standard, sans ponctuation ni majuscule de début de phrase. Chaque tour de parole fait l'objet d'un paragraphe séparé, qui commence par l'identification du locuteur.

Attention, ne pas insérer une fin de paragraphe (retour chariot) à chaque bout de ligne d'écran, mais seulement aux fins de tours de parole.

Exemple :

L1 donc vous faites quoi dans la vie
L2 mais je travaille ici à la fac + euh je fais le ménage + je fais un peu
 dans le secrétariat le mercredi après-midi + à l'imprimerie quoi +
L1 et vous faites quoi dans le secrétariat
L2 ben je fais les photocopies {rire} + et je trie euh et j'agrafe +

Un certain nombre de symboles particuliers peuvent ou doivent être utilisés au cours de la transcription. Ils sont résumés dans le tableau ci-dessous, et leur utilisation est détaillée dans la suite du texte.

{...}	éléments métalinguistiques (commentaires, etc.)
[...]	prononciations particulières
< ... >	chevauchements de locuteurs
+	pauses
///	pause très longue (enregistrement non coupé)
###	partie non transcrite (enregistrement non coupé)
\$\$\$	coupure de l'enregistrement
=	liaison non-standard remarquable
#	absence remarquable de liaison
/..., .../	hésitations entre transcriptions
(...)	variantes graphiques indécidables
...-	amorces
*	syllabe incompréhensible
***	suite de syllabes incompréhensible
"..."	titres
...	nom propre ou suite de chiffres anonymisés

Tableau 1. Résumé des symboles de transcription

Équipe DELIC

2. Locuteurs et tours de parole

2.1. Identification des locuteurs

Les locuteurs sont désignés par L1, L2, etc. *dans l'ordre de prise parole dans la partie transcrite*. Ils apparaissent en début de chaque tour de parole, séparés du texte par une tabulation.

L1 est-ce que je peux vous demander votre âge
L2 vingt-deux ans

Attention : (1) ne pas mettre d'espace ou tabulation avant la marque du locuteur ; (2) ne pas mettre d'espace entre le L et le numéro ; (3) toujours utiliser un L majuscule.

Dans le cas où il n'est pas possible de décider quel est le locuteur qui parle parmi L1, L2, etc. on utilisera la notation LX (attention L et X majuscules) :

L1 en ligne ça veut dire sur Internet
LX on line

Attention : LX ne signifie pas locuteur inconnu. Tout locuteur, même s'il ne prononce qu'un mot doit avoir un identificateur propre du type L1, L2, etc.

2.2. Chevauchements de parole

Les chevauchements de parole sont notés au moyen de chevrons < >. Le chevron ouvrant marque le début du chevauchement, le chevron fermant en marque la fin. Exemple :

L1 de me penser à penser à trouver un travail + < et l'imprimerie
L2 vous aviez à peu près > quel âge

Les chevauchements peuvent mettre en jeu plus de deux locuteurs :

L1 tu veux < que je te chauffe un peu le café pupuce
L3 oui je veux bien
L2 tu vois un peu hein c'est > + donc il faut faut s'intéresser à
L1 je vais mettre dans mon bol et je le mets au < micro-onde
L3 oui
L2 au fonctionnement > au fonctionnement même

Laisser une espace avant et après les chevrons (sauf en début et fin de ligne).

3. Règles typographiques

Ce sont celles de l'écrit standard, à part l'absence de majuscule en début de phrase, cette notion n'ayant pas cours dans les transcriptions d'oral (voir manuels de typographie).

Il faut apporter beaucoup de soin aux détails typographiques. Les incorrections entraînent de grosses difficultés pour les traitements automatiques des corpus.

Important : en cas d'utilisation de logiciels de traitement de texte évolués (du type *MS Word*) ne pas utiliser de caractères spéciaux, d'espace insécable, de tirets particuliers (cadratin ou autre), de mise en exposant ou indice, d'enrichissement typographique (gras, italique, etc.) ou de mise en forme (changement de marges, etc.). Attention aux automatismes de ces logiciels qui ont tendance à rajouter des majuscules en début de phrase, changer le type des guillemets ou des tirets, etc. Désactiver toutes les options ou mieux, utiliser un logiciel moins « intelligent », du type *Wordpad*.

3.1. Majuscules

On utilise normalement les majuscules sur les noms propres, géographique, historique, les titres et sigles, et autres mots portant normalement une majuscule à l'écrit :

ils ont créé le tout de **A à Z**
on a une zone industrielle de l'**Auxerrois** qui est assez dynamique
d'autres Anglais qui appartenaient à l'**O.M.S.**

Attention: on met une majuscule sur le nom d'un peuple (le Français est gourmand), mais pas sur celui de la langue (j'apprends le français). De même, pas de majuscule sur les mois (en janvier et non en Janvier).

On utilisera les majuscules accentuées :

normalement "**L'Énéide**" est dans le programme de français
et alors j'ai appelé **Éléonore**

Équipe DELIC

3.2. Sigles et acronymes

Dans la transcription, les sigles sont ponctués quand les lettres ont été prononcées isolément (S.N.C.F.), non ponctués lorsqu'il s'agit d'un acronyme prononcé comme un mot ordinaire (CROUS).

il peut apporter des **C.D.** à la **FNAC**
j'ai fini mon **DEUG** ça s'est bien passé

Le sens des sigles est précisé si besoin en commentaire lors de la première apparition (sigles peu courants uniquement) :

L'A.E.E. {sigle = Agence Européenne de l'Environnement}
actuellement il y a des négociations avec la **SANEF** {sigle = Société des
Autoroutes du Nord Est de la France}

Les prononciations particulières doivent être notées (voir section Prononciation) :

une conférence de **l'I.E.E.E.** {sigle = Institute of Electrical and Electronics
Engineers} {pron = [itRwaz2]}

3.3. Abréviations

En règle générale ne pas utiliser d'abréviations (titres honorifiques, unités, etc.), sauf dans les cas où la forme développée n'est normalement pas utilisée à l'écrit (etc.) :

Monsieur Chirac
en 300 **avant Jésus-Christ**
40 **kilomètres à l'heure**
douze ou treize **pour cent**
35 **degrés Celsius**

mais :

on nous ressort l'**Inquisition** les **Dragonnades** etc.

3.4. Lettres

Les lettres utilisées en tant que telles sont écrites en majuscules :

ils ont créé le tout de **A à Z**
un rang euh catégorie **B**

C'est le cas des mots épelés :

- L1 fees {lang = anglais} c'est des c'est des comment est-ce qu'on dit c'est de la rémunération en fait + c'est de la parti- c'est c'est
- L2 c'est un mot anglais
- L1 oui c'est un mot anglais ouais **F E E S** +

3.5. Nombres et chiffres

Les nombres doivent respecter les normes habituelles de l'écrit. On écrit en lettres les nombres inférieurs à 10, les nombres employés comme substantifs, certaines expressions :

- il pleut depuis **trois** jours
- il y a **six** mois
- il a raté sa **quatrième**
- les années **cinquante**
- la guerre de **Trente** ans

Dans les autres cas, écrire en chiffres (on sépare les tranches de mille par des espaces, sauf dans les dates) :

- la lumière parcourt **300 000** kilomètres par seconde
- il est né en **1986**

Respecter le format habituel des numéros de téléphone :

- ce sera juste après le journal de dix-huit heures avec vos questions + au standard au **05 59 59 17 17**

3.6. Titres d'œuvres

Respecter l'écriture habituelle des titres d'œuvres littéraires, journaux, films etc., avec une majuscule sur le premier mot, et sur les noms propres internes. Isoler le titre du reste de la transcription par des guillemets droits.

- on est allé regarder "**Matrix**"
- le roman "**Vingt mille lieues sous les mers**"
- j'ai revu "**Les vacances de Monsieur Hulot**" de Tati
- je lis régulièrement "**Le Monde**"
- je suis abonné au "**Monde**"

Attention : utiliser impérativement les guillemets droits. Pas d'espace entre les guillemets et le titre qu'ils encadrent.

Équipe DELIC

3.7. Espaces

Ne jamais redoubler les espaces.

Ne pas laisser d'espace :

- entre l'amorce et le tiret (*un héli- hélicoptère*)
- entre la consonne de liaison remarquable et le signe = (*quatre =z= yeux*)
- entre les guillemets droits et le titre qu'il encadrent ("*Le Monde*")
- avant la virgule de multi-transcription (*/était, a été/*)
- à l'intérieur des parenthèses de variantes (*on (n') a pas*)

Toujours laisser une espace :

- après le tiret d'amorce (*un héli- hélicoptère*)
- avant et après le signe de pause + (*il répond + à*)
- avant et après les signes d'interruption ///, ###, \$\$\$
- après la virgule dans une double écoute (*/était, a été/*)
- avant et après les {...} et les [...]
- avant et après (n') (*on (n') a pas*)
- avant et après les chevrons de chevauchement (*L1 tu la trouves < bien, L2 très belle > oui*)
- avant et après # (de bons # amis)
- avant et après les *, ***, *P*, etc.

4. Orthographe

4.1. Accords non standard

Ne pas corriger, mais ajouter l'indication {sic}. Cette notation n'a aucun caractère de jugement normatif, mais indique simplement qu'il ne s'agit pas d'une erreur du transcripteur :

les conseils **nationals** {sic}
on gagne pas des sommes **exorbitants** {sic}

4.2. Élisions non réalisées

Ne pas corriger :

nous savions déjà par le "Le Guide du Routard" **que il** y avait quelques problèmes avec les enfants de de Djenné
elle est passée de quinze ouvrières à deux **parce que elle** pouvait plus euh les payer

4.3. Variantes morphologiques indécidables

En cas de variantes morphologiques non réalisées à l'oral, la transcription est indécidable et les alternatives sont notées entre parenthèses :

l'avantage de de l'I.U.T. euh c'est qu'il y a il y a pas de **spécialité(s)** euh **définie(s)**
quelques-unes à ne pas mettre entre toutes les mains comme **il(s) disai(en)t**
on (n') est pas quitte avec les gens

Sauf indication explicite du contraire, les adjectifs et participes se rapportant au pronom *on* restent au singulier, sauf réalisation phonique du pluriel. Exemple :

on est **parti** en Afrique de l'Ouest
on est **parti** tous les deux

mais :

on s'est **mises** dans de beaux draps

4.4. Onomatopées

Un certain nombre d'onomatopées sont codifiées et doivent être transcrites selon l'orthographe fournie :

ah, aïe, areu, atchoum, badaboum, baf, bah, bam, bang, bé, béeê, beurk, bien, bing, boum, broum, cataclap, clap clap, coa coa, cocorico, coin coin, crac, croa croa, cuicui, ding, ding deng dong, ding dong, dring, eh, eh ben, eh bien, euh, flic flac, flip flop, frou frou, glouglou, glou glou, groin groin, grr, hé, hep, hi han, hip hip hip hourra, houla, hourra, hum, mêêê, meuh, mh, miam, miam miam, miaou, oh, O.K., ouah, ouah ouah, ouais, ouf, ouh, paf, pan, patatras, pchhh, pchit, pff, pif-paf, pin pon, pioupiou, plouf, pof, pouet, pouet pouet, pouf, psst, ron ron, schlaf, snif,

Équipe DELIC

splaf, splatch, sss, tacatac, tagada, tchac, teuf teuf, tic tac, toc, tut tut,
vlan, vroum, vrrr, wouah, zip.

Exemples:

hum hum je n'en suis pas si sûr
et alors **crac** le truc s'est cassé
et d'un coup j'entends **miaou miaou** au dessus de ma tête

Pour les onomatopées qui ne figurent pas dans cette liste, l'orthographe est libre.

4.5. Orthographe de mots inconnus

Lorsque l'orthographe est incertaine (cas de certains noms de marques, toponymes, etc.) on pourra utiliser une orthographe approximative *si elle est plausible*, en la faisant suivre de la mention {approx} (cas d'un mot unique) ou en l'englobant dans la séquence {début approx} ... {fin approx} (plusieurs mots) :

un certain **Dupont** {approx}
le lieu-dit {début approx} **Sous la Voivre** {fin approx}

Si aucune orthographe n'est plausible, on utilisera une marque de séquence incompréhensible *, *** ou une transcription phonétique entre [...] (voir plus bas).

Attention : il faut utiliser cette possibilité avec précaution, et ne pas inventer d'orthographe farfelues. S'il y a réelle difficulté de transcription, on utilisera la notation * ou ***, en ajoutant la prononciation en commentaire uniquement si elle a été perçue de façon certaine (voir plus bas, *Difficultés de transcription*).

4.6. Mots étrangers

Les mots étrangers reconnus par le transcripateur sont orthographiés conformément à la norme de la langue d'origine (majuscule aux noms allemands, etc.) ou dans un système de translittération standard (pour le russe, chinois, etc.). Lorsque le mot n'est pas largement adopté en français, ajouter une indication en commentaire du type {lang = ...} :

comment on dit **maze** {lang = anglais}

mais :

ils se sont garés sur le **parking** du supermarché
on devrait normalement avoir une + une séance de ce qu'ils appellent
debriefing enfin

Lorsque plusieurs mots consécutifs sont dans une langue étrangère,
utiliser la notation {début lang = ...} ... {fin langue}

eh ben peu à peu il devient {début lang = allemand} **eine Sache** {fin
langue}

Si le mot n'est pas reconnu, on utilisera la notation * ou *** (voir plus
bas).

5. Prononciation

Les prononciations particulières sont notées entre crochets carrés [...] à l'aide de l'alphabet SAMPA (Annexe 3).

5.1. Prononciations déviantes

Le mot est parfaitement reconnaissable, mais sa prononciation est déviante (lapsus, prononciations non normatives, etc.). Dans ce cas, le mot est transcrit dans son orthographe habituelle et la prononciation est donnée en commentaire :

il se **passé** {pron = [plas]} plein de choses
les les **gens** {pron = [Sa~]}
pneumonie {pron = [pl2monij]}
aéroport {pron = [aReopOR]}

Si une prononciation s'applique à plusieurs mots, elle doit être encadrée par une balise ouvrante et une balise fermante {début pron = [...]} ... {fin pron}.

on est passé dans la boutique {début pron = [tytifRi]} **duty free** {fin pron}

5.2. Liaisons

On ne notera que les liaisons ou absences de liaisons réellement remarquables. Il ne faut pas noter les cas banals (*ils sont à Paris*).

Équipe DELIC

On notera les liaisons remarquables en marquant la consonne entre deux signes = :

quatre =z= amis
donne-moi =z= en
on va écouter plutôt les vieux qui =z= ont vécu
le texte sur la vie financière + qui a été voté à mon avis a été =t= une sorte
je sais pas quand =t= euh ben
mon grand-père était =t= euh cordonnier
ça c'est le travail de certains euh =z= employés

L'absence d'une liaison obligatoire se notera avec le signe # :

c'étaient trois # amis

5.3. Prononciations courantes

Il ne faut pas indiquer la prononciation élidée (sans schwa) pour les clitiques (*je dis, tu le sais, etc.* prononcés [Zdi]). De même, il est inutile de noter les prononciations courantes de type *il y a* prononcé [ja], *tu sais* [tze], *tu as* [ta], *puis* [pi], *peut-être* [ptEt].

Lorsqu'une prononciation est à la fois remarquable et récurrente pour un locuteur, on ne la reportera pas dans la transcription, mais on la fera figurer en commentaire dans la fiche signalétique.

6. Phénomènes propres à l'oral

6.1. Amorces (mots inachevés)

Les amorces de mots sont notées par un tiret final (sans espace *avant* le tiret) :

une des rares **rou-** routes goudronnées
oui mais par **con-** bon au niveau euh euh au niveau social

Les notations *j'-*, *l'-*, etc. sont exclues. Utiliser *j-*, *l-* :

je me suis dit **j-** j'ai dit
l- l'association

6.2. Pauses et interruptions

Toutes les pauses devront être marquées à l'aide du signe +, même brèves.

mais + un jour de septembre 1937 + il me convoqua chez lui + il me dit tu
sais + je suis gravement malade +

Les interruptions longues du discours (plusieurs secondes), le plus souvent liées à un événement particulier (le locuteur écrit au tableau, lancement d'un jingle à la radio, etc.), sont notées par une triple barre oblique ///. Une explication doit être fournie en commentaire :

/// {L1 se lève et va fermer une porte}
/// {musique}

Un certain nombre de mot-clés sont prédéfinis :

indicatif
jingle
musique
applaudissements

7. Difficultés de transcription

7.1. Mots incompréhensibles ou d'orthographe inconnue

Un mot ou une suite de mots incompréhensibles (inaudibles, dans une langue inconnue, etc.) seront notés * s'il s'agit d'une seule syllabe, *** s'il s'agit d'une suite de syllabes. Le cas échéant une transcription phonétique peut être fournie en commentaire, si elle est discernable (patronymes, toponymes ou mots étrangers, par exemple) :

le chemin était vraiment très long et que nous * la nuit a- arrivait
avec départ vers le puissant euh trois quart aile Frédéric * {pron = [gaz]}
je parle de Joseph *** je parle pas de de Jean-Pierre
vraiment dans la dans le coeur de l'Afrique où c'est euh à à *** {pron =
[basje]}

Si la langue est inconnue, utiliser ??? pour le nom de la langue :

il lui a dit *** {lang = ???}

Équipe DELIC

7.2. Multi-transcription

L'hésitation entre plusieurs séquences est notée entre barres obliques (sans espace après la barre ouvrante ni avant la barre fermante) :

après on a décoloré avec /des, les/ acides
alors /j'ai cherché, je cherchais/

Lorsqu'il y a hésitation sur la présence d'une séquence sonore, on note une alternance avec 0 (attention, le chiffre zéro, pas la lettre O, ni le signe ensemble vide Ø) :

vous croyez qu'on a /l/, 0/ envie de ri- envie de rigoler

Il peut y avoir hésitation entre un morphème reconnu et une séquence incompréhensible. Dans ce cas on notera une alternance avec * ou *** :

bon ben de toute façon tu /dois, */ avoir quand même l'expérience d'autres choses

7.3. Événements non linguistiques

Les rires, les bruits, sont signalés dans le texte entre accolades à l'endroit où ils se produisent s'ils sont ponctuels :

donc vous allez me parler de votre travail {porte qui claque}

ou bien sous forme de balises encadrantes si l'événement a une durée notable :

{début conversation de fond} ... {fin conversation de fond}

Si des événements sont récurrents dans un enregistrement (bruits, etc.), on ne les notera pas dans la transcription, mais en commentaire dans la fiche signalétique.

Un certain nombre de mot-clés sont prédéfinis¹ (les parenthèses indiquent des abréviations possibles) :

¹ Ces mots-clés sont les valeurs par défaut du logiciel *Transcriber*.

Présentation du *Corpus de référence du français parlé*

Locuteur :

respiration (r)	bruit de bouche (bb)	rire
inspiration (i)	bruit de gorge (bg)	sifflement (sif)
expiration (e)	toux (tx)	
reniflement (n)	râclément	
souffle (pf)	éternuement	

Auditoire ou autres locuteurs :

rires
rire(s) en fond
toux en fond
applaudissements
conversation de fond

Bruits divers :

bruit indéterminé (b)
froissement de papiers (pap)
souffle électrique (shh)
bruits micro (mic)
indicatif
jingle
musique

8. Anonymisation

Dans les cas où les noms propres doivent être supprimés on respectera le codage suivant : *P* = patronyme ; *T* = toponyme ; *S* = marque commerciale, nom de société.

il me dit Monsieur *P* il y a pas de problème
et cette entreprise donc euh germano-française *S*

S'il y a plusieurs noms propres de la même catégorie dans la même transcription, chacun reçoit une numérotation : *P1*, *P2*, etc. :

des gens comme monsieur *P1* et monsieur *P2* madame *P2*
ce qui serait marrant c'est que tu ailles demander au propriétaire du *S1* là
tu vois le bar à vin

Équipe DELIC

Lorsque des suites de chiffres (numéros de téléphone, numéros de rue, numéro de sécurité sociale, etc.) doivent être anonymisées, on utilisera le codage *C* :

rappelez-moi au *C*
il habitait au *C* des Champs-Élysées

9. Relation avec l'enregistrement

9.1. Parties non transcrites

Une partie comportant de la parole non transcrite (mais non coupée sur l'enregistrement) doit être notée par ###. La raison doit être donnée en commentaire :

{longue déclaration en basque}

Un certain nombre de mot-clés sont prédéfinis :

inintelligible
faible
très faible
voix superposées

9.2. Coupures de l'enregistrement

Une coupure de l'enregistrement doit être notée par \$\$\$\$. Un commentaire peut être ajouté si l'on connaît la raison de la coupure, et/ou sa durée :

\$\$\$ {l'invité joue du piano}

\$\$\$ {longue séquence musicale, durée = 13 min.}

Un certain nombre de mot-clés sont prédéfinis :

indicatif
jingle
musique
applaudissements

Annexe 3. L'alphabet SAMPA²

Consonnes

	Symbole	Exemple	Transcription
Plosives	p	pont	po~
	b	bon	bo~
	t	temps	ta~
	d	dans	da~
	k	quand	ka~
Fricatives	g	gant	ga~
	f	femme	fam
	v	vent	va~
	s	sans	sa~
	z	zone	zon
Nasales	S	champ	Sa~
	Z	gens	Za~
	m	mont	mo~
Liquides	n	nom	no~
	J	oignon	oJo~
	N	camping	ka~piN
	l	long	lo~
Semi-consonnes	R	rond	Ro~
	w	coin	kwe~
	H	juin	ZHe~
	j	pierre	pjER

² <http://www.phon.ucl.ac.uk/home/sampa/home.htm>

Équipe DELIC

Voyelles

	Symbole	Exemple	Transcription
Orales	i	si	si
	e	ses	se
	E	seize	sEz
	a	patte	pat
	A	pâte	pAt
	O	comme	kOm
	o	gros	gRo
	u	doux	du
	y	du	dy
	2	deux	d2
	9	neuf	n9f
	@	justement	Zyst@ma~
	Nasales	e~	vin
a~		vent	va~
o~		bon	bo~
9~		brun	bR9~
E/		= e ou E	
Indéterminées	A/	= a ou A	
	&/	= 2 ou 9	
	O/	= o ou O	
	U~/	= e~ ou 9~	

Les compléments de lieu réalisés par *y* : description des usages

Frédéric SABIO
Université de Provence

Résumé

Cette étude concerne le pronom *y* utilisé en tant que complément locatif, comme dans les énoncés suivants qui en présentent plusieurs occurrences successives :

j'y retourne de- j'y retourne souvent + mais euh de de là à *y* vivre je sais pas peut-être que *y* passer quelque temps mais m- enfin moi je suis assez casanier (corpus oral)

Je peux m'identifier à chaque larme qui est versée là-bas, à chaque enfant qui *y* est resté, à chaque souffrance qui *y* est éprouvée. Mais je ne peux m'identifier à aucun parti qui s'y déchire, à aucune des conceptions politiques qui essayent de s'y imposer. (corpus écrit)

S'il est bien établi que *y* possède entre autres fonctions celle de réaliser la rection locative des verbes, il reste utile de s'interroger, données à l'appui, sur la manière dont les locuteurs en font usage dans leurs productions linguistiques, afin d'en évaluer la plus ou moins grande disponibilité en fonction des genres de textes ou des propriétés syntaxiques et sémantiques des verbes constructeurs auprès desquels il apparaît. Cela permettra d'observer que, loin d'être une proforme locative « passe-partout », *y* fait l'objet d'un usage plutôt parcimonieux, tout particulièrement dans les productions orales informelles.

Frédéric SABIO

1. Présentation des données

1.1. Délimitation grammaticale

Nous voulions limiter cette étude aux seules occurrences du *y* rectionnel à valeur locative. À l'issue d'une sélection parfois délicate, nous avons décidé d'exclure les occurrences de *y* qui n'ont pas de statut pronominal et celles qui réalisent certaines rections prépositionnelles sans lien avec l'expression du lieu.

1.1.1. Le *y* non pronominal

Nous considérons que *y* est dépourvu de valeur proprement pronominale¹ dès lors qu'il ne supporte aucune lexicalisation ou accepte mal d'être remplacé par d'autres proformes à sémantisme locatif comme *où*, *ici*, *là*, *là-dedans*.

Ceci a permis d'écarter les fort nombreuses structures présentatives en *il y a*, dans lesquelles *y*, totalement grammaticalisé, ne commute jamais avec du lexique.

On a délaissé également les *y* qui apparaissent dans les structures ressenties comme locutionnelles ; ici se pose le problème du figement, qui peut selon les constructions paraître plus ou moins « avancé ».

Le degré maximal du figement semble atteint dans des énoncés comme :

il y est allé de son discours (oral)
Il n'y va pas de main morte. (écrit)
marche, polka, farandole, triple galop, tous les tempo y passent (écrit)
(...) plus éclectique qu'il n'y paraît (écrit)

dans lesquels l'élément *y* ne commute avec rien (si ce n'est avec « zéro » dans le dernier énoncé), pas même avec un autre pronom locatif :

*où est-il allé de son discours ?
*il est allé là-dedans de son discours
*il ne va pas là-bas de main morte
*tous les tempo passent là, à cet endroit [structure possible mais sans lien avec l'énoncé d'origine]

¹ Voir Blanche-Benveniste et al. (1984).

Les compléments de lieu réalisés par y : description des usages

On rencontre cependant de nombreux énoncés pour lesquels le figement semble moins absolu, soit que d'autres pronoms locatifs puissent être utilisés, soit qu'une forme de lexicalisation soit envisageable :

je sais pas si le Ministère y trouvera son compte (crfp, prigap2)

→ s'il trouvera son compte là-dedans, dans ces négociations

à un moment donné mon époux m'ai- euh m'aidait + et euh + euh il y a laissé une ponceuse une perceuse euh enfin bon ça le mettait de très mauvaise humeur (crfp, pripse3)

→ il a laissé une perceuse dans l'aventure

Si leur seconde mission a échoué, les juifs des États-Unis y sont un peu pour quelque chose. (Le Monde)

→ ils sont pour quelque chose là-dedans, dans cet échec.

De telles données ont tout de même été exclues de nos comptages en raison de leur caractère quasiment figé.

1.1.2. Le y qui indique une rection prépositionnelle de type « à lui / à cela »

Ont été également écartées les valeurs pronominales non locatives de type « y... à (faire) cela » ou « y ... à lui », comme dans :

alors + pour accéder à la masse de ressources + disponibles sur Internet + on a des outils pour nous y aider qui vont diriger nos recherches (crfp, pubpsol)

→ pour nous aider à faire ça, à accéder à ces ressources

eux ils y trouvent quand même un intérêt (crfp, pripno4)

→ ils trouvent un intérêt à cela (produit par un enseignant qui évoque l'intérêt des élèves pour l'école)

Ici encore, quelques énoncés ont posé problème dans la mesure où la répartition entre lieu et objet n'est pas toujours transparente ; nous avons par exemple exclu l'énoncé suivant, où il pouvait s'agir tout autant de *s'exercer là-bas / dans le massif de Coquibus* que de *s'exercer à cela / au tir* :

Le massif de Coquibus est connu pour sa fréquentation discrète de truands à la recherche d'un coin tranquille pour s'entraîner au tir. Mesrine, dit-on, venait s'y exercer. (Le Monde)

Frédéric SABIO

Ont été également écartées du corpus les structures en *y venir* ou *y revenir* qui correspondent en fait à *venir/revenir à cela*, souvent utilisées dans un emploi méta-discursif² :

laissons de côté le *dramatis personae* + nous **y reviendrons** ultérieurement (crfp, pubpau1)

Il n'y avait que seize ateliers flexibles en fonctionnement et moins de 25 % des établissements possibles s'étaient convertis à la gestion de production automatisée. Cela devrait évoluer vite : quatre sur cinq des entreprises équipées en productique **y sont venues** depuis 1980. (Le Monde)

Les exemples suivants, qui nous semblent supporter les deux interprétations syntaxiques, ont été finalement conservés pour l'étude, mais on n'oubliera pas dans la suite de la discussion que leur interprétation comme « complément de lieu » est sujette à caution :

- *y ajouter cela* :

c'est comme une crêpe mais c'est un peu plus épais que la crêpe **on y ajoute des petits morceaux de jambon** à l'intérieur (crfp, prosai1)

- *y mettre de l'énergie* :

La LCR, à laquelle Gilles Perrault veille à rendre scrupuleusement la paternité de l'idée du 8 juillet, **y a mis toutes ses énergies militantes**. (Le Monde)

- *y associer cela* :

Le Quai d'Orsay est en train de créer des radios locales FM dans les grandes métropoles étrangères (Rome, Lisbonne, bientôt Londres), en faisant appel aux services d'une société privée, OFREDIA sans **y associer RFI**. (Le Monde)

- *y accéder* :

Les quatre nacelles de métal (...) permettront d'y **accéder** enfin. (Le Monde)

1.1.3. Ce que « locatif » veut dire : structure syntaxique et induction de sens

La décision d'écartier certains exemples doit, dans la recherche syntaxique, être établie autant que possible selon des critères formels. C'est

² On a gardé bien sûr les occurrences de *y venir* à sémantisme locatif :

On ne démolirait pas les postes frontalières. Les troupes **y viendraient** tous les jours. (corpus écrit)

ce que nous avons tenté de faire plus haut, par le biais du recours aux pronoms et des possibilités de commutation avec le lexique.

En revanche, il n'aurait pas été juste d'écarter certains énoncés au seul motif que le pronom *y* ne réfère pas directement à ce que le sens commun interpréterait comme un véritable « lieu », dans le sens le plus concret qu'on puisse donner à ce terme.

C'est ainsi que *y* peut reprendre un lexème comme *voie* employé non pas dans son sens concret de « chemin d'accès », mais de manière nettement métaphorique :

telle est bien la *voie* envisagée par les écoles qui s'interrogent sur leur fonction en cette période de mutations accélérées des techniques et des organisations. M. Esambert souhaite que Polytechnique s'y engage résolument. (Le Monde)

Le plus souvent, les emplois atypiques du *y* locatif concernent des syntagmes qui ne sont pas habituellement attachés à l'expression du lieu ; par exemple, ces deux exemples en *y trouver* donnent une valeur spatiale à des syntagmes qui désignent respectivement un espace de temps et une collectivité d'individus :

Un bustier parme, un ton de chartreuse, l'atmosphère de Boldini, les camaïeux de Whistler, l'envers d'une feuille de laurier... », dit Christian Lacroix, qui remue des sensations, chinant dans le passé pour *y trouver* sa vérité. (Le Monde)

80 % des habitants du monde rural sont à présent des non-agriculteurs. On *y* trouve des employés, des ouvriers, des entrepreneurs. (Le Monde)

Il en va de même des énoncés dans lesquels le pronom paraît « déborder » la seule indication du lieu, pour référer de façon plus globale à l'institution qui *y* est attachée : *y être admis, dans une grande école, y rentrer, dans une entreprise...*, ainsi, le locuteur de cet extrait qui parle d'aller *aux championnats de France* :

cette année on a été battu en demi-finale enfin on aurait pu faire mieux comme on dit pour *y aller* (crfp, prilil1).

Enfin, lorsqu'il est régi par des verbes comme *voir, reconnaître, être impliqué*, le pronom *y* acquiert souvent un sémantisme qui, sans perdre totalement son caractère locatif, tire nettement vers l'expression d'un processus : il se pronominalise aisément en « dans cela » et entre en relation avec un nom d'action (*être impliqué dans des affaires politiques, voir une provocation dans une opération militaire, se reconnaître dans l'action d'un syndicat*) ou avec une séquence verbale, comme dans l'énoncé :

Frédéric SABIO

Quand le vent se leva en milieu d'après-midi, la foule y vit comme un mauvais présage. (Le Monde)

1.2. Le corpus étudié

L'étude porte sur des données orales et écrites.

La description des données orales a été menée de manière systématique à partir d'environ 500 000 mots de parole transcrite. Deux corpus différents ont été utilisés : une sous-partie du corpus de français parlé Corpaix (département de linguistique française, Université d'Aix-en-Provence), qui a fait l'objet d'une publication sous forme d'ouvrage³, et qui totalise un peu plus de 66 000 mots ; le Corpus de référence de français parlé, élaboré par l'équipe DELIC, qui avoisine 440 000 mots.

Les données écrites proviennent du journal Le Monde, et totalisent 250 000 mots⁴.

On a enfin utilisé des énoncés rédigés par des enfants d'âge scolaire pour quelques illustrations ponctuelles (mais sans les intégrer dans les comptages).

2. Résultats

2.1. Occurrences de y locatif

On observe que l'ensemble du corpus oral présente 268 occurrences, ce qui fait une moyenne de un y locatif tous les 1866 mots.

Sur l'ensemble du corpus écrit, le total de 235 occurrences donne en moyenne un y locatif tous les 1163 mots.

Voici le relevé des verbes qui s'accompagnent d'un y locatif, donnés par ordre de fréquence décroissante.

³ Blanche-Benveniste et al. (2002).

⁴ Je remercie Jean Véronis, qui m'a permis d'accéder aux données journalistiques. Dans l'analyse, nous prendrons évidemment en compte le fait que les données orales sont deux fois plus nombreuses que les données écrites.

Les compléments de lieu réalisés par y : description des usages

Structure	Nombre d'occurrences
aller	119 (= 45%)
être	35 (= 13%)
retourner	14
rester	13
mettre	12
trouver	11
vivre	7
travailler	6
passer	5
faire	4
se passer, se plaire, revenir (analyse parfois incertaine)	3
s'arrêter, envoyer, lire, voir	2
accueillir, (adhérer), (ajouter), (apprendre à INF), arriver, avoir ("des amis"), chercher, se construire, déposer, (développer), enterrer ("elle est enterrée"), entrer, habiter, monter, naître, nommer, parler de N, partir, passer tant de temps, rentrer, retracer, se retrouver, se sentir bien, se trouver, se trouver bien	1
total	268

Tableau 1. Données orales (500 000 mots)

Structure	Nombre d'occurrences
être	20
trouver, voir	17
aller, avoir, installer	6
rester	5
établir, rencontrer	3
apprendre, cacher, construire, côtoyer, déceler, déposer, disposer, donner, faire, s'installer, lire, mettre, occuper, prendre, puiser, régner, retourner, siéger, souligner, subir, tenir, transporter, travailler, se trouver	2
aborder, accéder, accueillir, acheter, acquérir, admirer, affirmer, aller comme ça, apporter, apprécier, s'arrêter, s'attarder, attirer, avoir lieu, se battre, calmer le jeu, considérer (forme : être considéré ainsi), constituer, contempler, se déchirer, déclarer, découvrir, défendre, déléguer, dénombrer, se déplacer, se dérouler, détruire, développer, se développer, devenir, se dissimuler, se dire, dominer, doubler, dresser, édifier, égarer, s'embusquer, émerger, emmener, s'engager, s'engouffrer, entasser, entrer, éprouver, étudier, expliquer, faire + inf., fêter, gagner sa vie, héberger, se heurter, s'imposer, incarcérer, inclure, insuffler, investir, jouer un rôle, lapider (passif : y être lapidé), mettre en scène, négocier, ouvrir, parler, passer sa vie, se pendre, pénétrer, perdre, piocher, pointer son nez, posséder, se précipiter, prendre ses aises, se préparer, se presser, proclamer, produire, se loger, se maintenir, se rassembler, se reconnaître, récupérer, se refléter, reflleurir, renoncer, repérer, représenter, respecter, rétablir, rétorquer, retrouver, revenir, saisir (forme : être saisi par), scolariser (forme : être scolarisé), semer la terreur, se sentir ainsi, servir, totaliser, traiter, transiter, venir, vivre	1
total	235

Tableau 2. Données écrites (250 000 mots)

2.2. Description linguistique

2.2.1. Dispersion lexicale

Dans le corpus oral, le nombre de formes verbales différentes qui régissent un *y* locatif est au nombre de 41 (dont quelques-unes sont sujettes

Les compléments de lieu réalisés par *y* : description des usages

à caution, voir avertissement *supra*). Le corpus écrit, pourtant deux fois moins important en nombre de mots, voit ce nombre de lexèmes verbaux s'élever à 136.

On observe donc :

- Une « dispersion lexicale » importante du côté des sources écrites, qui se manifeste ici par le fait que la totalité des occurrences de *y* locatifs se répartit sur un grand nombre de lexèmes verbaux différents.
- Une dispersion lexicale nettement « resserrée » du côté de l'oral, qui est illustrée par le fait qu'un nombre très réduit de formes verbales différentes est à l'origine de la majorité des occurrences relevées ; c'est ainsi que les verbes *aller* et *être* arrivent nettement en tête en totalisant à eux deux 58% de toutes les occurrences de *y* locatif rencontrées à l'oral :
 - Le verbe *aller* compte à lui seul 119 occurrences, soit près de 45% du total ;
 - le verbe *être* est bien représenté, dans des structures assez diverses (*y avoir été, y être bien, y être trois...*) à hauteur de 13% du total.

La prééminence de ces deux verbes sur la somme totale de *y* locatifs produits à l'oral est à ce point décisive que, si l'on mettait *aller* et *être* de côté, les occurrences chuteraient à une moyenne de 1 tous les 3246 mots.

Pour faciliter certaines comparaisons, nous faisons figurer dans la liste ci-dessous les cinq verbes qui sont à l'oral les plus fréquemment accompagnés du *y* locatif, ainsi que le pourcentage d'occurrences relevées dans le corpus oral et écrit.

Remarquons que pour l'oral, la somme des *y* locatifs affectés à ces cinq verbes rend compte de 74,5% du total des occurrences relevées :

verbes par fréquence à l'oral	occurrences de <i>y</i> locatifs (%) oral	occurrences de <i>y</i> locatifs (%) écrit
1- aller	45,0	2,5
2- être	13,0	8,5
3- retourner	05,2	0,8
4- rester	04,8	2,0
5- mettre	04,4	0,8

Nous nous arrêtons à présent sur les facteurs syntaxiques qui paraissent favoriser l'occurrence de *y* locatif.

Frédéric SABIO

2.2.2. Occurrences de *y* et valence verbale

Nous nous demanderons en premier lieu si le caractère valenciens des rections locatives favorise l'occurrence du pronom *y*. Rappelons qu'on considère comme *valenciens* les compléments qui « caractérisent le sens et la construction minimale d'un verbe » (Blanche-Benveniste et al., 1990 : 292) ; ils correspondent à ce que l'école désigne sous les termes de « compléments essentiels » ou « obligatoires ». Par exemple, *y* est nettement valenciens dans des énoncés comme *il y est allé* ou *il les y a amenés*.

On sait à quel point il est délicat d'affirmer que telle rection locative relève de la valence. La question ne sera pas débattue ici : il a paru plus pratique de relever, parmi les formes attestées dans le corpus, les formes verbales dont le complément de lieu est perçu comme un élément de simple rection, c'est-à-dire comme un complément de type « circonstanciel », relativement « secondaire » tant au plan sémantique que syntaxique.

Avec toute la prudence qu'impose ce type de sélection, qui ne peut se passer d'une certaine part d'intuition, nous proposons de considérer comme de simples rectionnels les emplois de *y* attachés aux verbes suivants :

- Corpus oral : *faire, lire, apprendre, avoir, se construire, parler, retracer*.
- Corpus écrit : *être, voir, avoir, apprendre, côtoyer, déceler, faire, lire, occuper, souligner, subir, tenir, aborder, acheter, acquérir, admirer, affirmer, apprécier, se battre, calmer le jeu, se sentir ainsi, totaliser, traiter, éprouver, expliquer, fêter, insuffler, jouer un rôle, lapider, mettre en scène, négocier, ouvrir, parler, se pendre, posséder, prendre ses aises, se préparer, proclamer, produire, se reconnaître, reflourir, renoncer, représenter, considérer, constituer, contempler, se déchirer, déclarer, défendre, détruire, développer, devenir, se dire, dominer, doubler, dresser, respecter, rétablir, rétorquer, saisir, semer la terreur*.

Ramené au nombre total d'occurrences des pronoms locatifs à l'étude, le contraste entre les données écrites et les données orales est saisissant : à l'oral, les *y* locatifs non valenciens ne comptent que pour 4% des pronoms recensés ; pour l'écrit, ils atteignent près de 48%.

Les compléments de lieu réalisés par *y* : description des usages

<i>y</i> locatif de simple rection	
(en % de l'ensemble des occurrences)	
données orales	04,0
données écrites	47,6

Ces résultats illustrent que, dans les productions orales, la présence de *y* locatif paraît fort étroitement liée au caractère valenciel du complément.

On peut cependant noter qu'outre le caractère valenciel, le second facteur qui favorise de la manière la plus décisive l'occurrence de *y* locatif à l'oral est lié au caractère *obligatoire* de la réalisation du complément, typique de verbes (le plus souvent mono-valenciels) comme *aller* et *être*, auxquels on peut ajouter *vivre*, *naître*, *habiter*, *se trouver*. Ces verbes, lorsqu'ils sont employés en structure locative, semblent en effet exiger la réalisation de leur valence : **il va⁵*, **il est*, **tu as vécu*, **tu es né⁶*, **tu habites*, **tu te trouves*. À eux seuls, ces verbes totalisent 62% des occurrences de *y* locatif.

Dans le corpus écrit, les objets locatifs dont la réalisation peut être considérée comme obligatoire (qui concernent *être*, *aller*, *se trouver*, *accéder*, *passer sa vie*, *pénétrer* et *vivre*) totalisent 13,6% de toutes les formes pronominales recensées, soit nettement moins que dans le corpus oral.

<i>y</i> locatif valenciel à réalisation obligatoire	
(en % de l'ensemble des occurrences)	
données orales	62,0
données écrites	13,6

⁵ On pourrait en revanche trouver *tu vas ?* ou *ça va*, au sens de *aller bien* : mais il s'agit là d'un emploi non locatif. À noter que dans ces structures, l'adverbe *bien* est fréquemment non réalisé sans que le sémantisme positif de la construction disparaisse : *il s'est débrouillé*, *ils s'entendent*, *elle s'en sortira...* Lorsqu'on parle de valence locative à réalisation obligatoire, on doit bien sûr mentionner les exceptions bien connues et fort anciennes, dues à des facteurs phonotactiques : *y* n'apparaît pas lorsque le radical de *aller* prend la forme [i] : *il ira*, *vous iriez*.

⁶ *Naître* et *vivre* relèvent de ces verbes qui passent très mal en l'absence de tout complément, mais qui acceptent en revanche des valences diverses : le lieu mais aussi le temps (*il est né en 78*, *il a vécu au début du siècle*), la manière (*il est né complètement chauve*, *il a vécu sans argent*).

Frédéric SABIO

Toujours à propos de l'analyse en valences verbales, le verbe *travailler* (qu'on n'a pas signalé dans les listes ci-dessus) a un statut intéressant ; c'est le seul verbe auquel l'intuition n'accorderait pas un sémantisme foncièrement locatif, et qui pourtant apparaît assez régulièrement précédé de *y* : 6 occurrences à l'oral, 2 à l'écrit. Cela semble confirmer la remarque de Blanche-Benveniste et al. (1990 : 45) qui considèrent que le locatif appartient bien à la valence d'un verbe comme *travailler*.

3. Le pronom *y* et les usages du français

La zone de réalisation du *y* locatif apparaît donc bien différente selon le type de corpus décrit : elle est relativement vaste dans le type de productions écrites retenu pour l'étude, puisqu'elle touche divers types de compléments de lieu, qu'ils soient valenciels ou simples rectionnels. Une telle distribution est d'ailleurs conforme à l'intuition souvent exprimée par la plupart des syntacticiens, pour qui *y* peut instancier aussi bien un complément « essentiel » du verbe qu'un complément « circonstant ». Parmi ces derniers, Le Goffic (1993 : 231) donne des exemples comme *j'y ai rencontré quelques amis, j'y ai lu que P, j'y ai fait construire trois nouveaux immeubles, on y passe des films anciens*.

La description de données orales oblige cependant à nuancer ce point de vue : la zone de réalisation de *y* locatif *y* apparaît pratiquement restreinte aux compléments qui relèvent de la structure valencielle des verbes constructeurs, avec un avantage très net pour ceux dont la réalisation prend un caractère obligatoire. Si bien que les exemples donnés par Le Goffic, s'ils correspondent bien au « possible de la langue », doivent être soumis à une analyse plus précise qui s'interrogerait, elle, sur les « possibles de discours ».

Disons d'emblée que l'opposition illustrée par les données recueillies ne saurait être interprétée comme une différence stylistique binaire entre « la syntaxe de l'écrit » et « la syntaxe de l'oral ». Il est plus éclairant de retenir le fait que les données écrites prises en compte dans cette étude (articles du journal *Le Monde*) relèvent d'un registre nettement plus cérémonieux que les données orales, dont beaucoup ont un caractère assez informel : c'est ainsi que je distinguerai à titre d'hypothèse trois usages distincts du *y* locatif, dont chacun est lié à un niveau spécifique de « formalité » :

3.1. Type *j'y vais* : non formel

L'usage de *y* avec des verbes mono-valenciels comme *aller, être, naître, habiter...* est extrêmement fréquent, quel que soit le « genre » du texte produit.

j'y allais avec une de nos petites arpètes (Corpus Delic, fleuriste)

le boulot faut faut reconnaître on (n') *y va pas* par plaisir + on *y va* par obligation (crfp, pricle3)

on *y est* depuis euh bé depuis cinq ans maintenant (crfp, pribe1)

on *y a été* tout simplement parce qu'on a eu une une brochure dans la boîte aux lettres (crfp, prinar1)

oui + par le fait qu'il est qu'il *y est né* en plus (crfp, princy2)

L'usage de *y* dans ces exemples n'a aucune valeur particulièrement cérémonieuse. Le fait que ces verbes se voient invariablement accompagnés d'un complément indiquant le lieu⁷ paraît favoriser l'utilisation du pronom locatif dans les textes oraux informels.

3.2. Type *j'y reste* : formel

Les verbes dont la valence locative peut être instanciée par « zéro » apparaissent rarement accompagnés de *y*, dès lors que la situation langagière est informelle : c'est très net pour *entrer, rentrer, rester, s'arrêter, monter, arriver...*

quand il fallait monter au sixième avec un truc lourd ben c'était pas drôle hein ou même à l'ascenseur si c'était un trop gros + une trop grosse corbeille eh ben on pouvait pas **monter** avec l'ascenseur + fallait **monter** à pied hein (Corpus Delic, fleuriste)

j'allais en m'engouffrant euh dans un sas c'est-à-dire une porte avec un + derrière lequel se situait un bac avec de l'eau où il fallait essuyer ses bottes avant de **rentrer** (Corpus Delic, poulets)

on arrivait très tôt à l'usine euh déjà il y a pas mal de gens qui étaient **rentrés** avant parce que tous les toutes les personnes qui qui passaient à l'a- qui travaillaient à l'abattoir travaillaient au moins **étaient rentrés** au moins depuis une heure ou voire deux heures deux heures avant nous pour que nous puissions avoir les les poulets prêts quand nous **arrivions** + et bon nous on **arrivait** donc vers six heures six heures trente (Corpus Delic, poulets)

⁷ Il se peut que l'obligation soit moins absolue pour *retourner*.

Frédéric SABIO

Bien souvent d'ailleurs, la non réalisation de la rection locative ne crée aucun effet d'indétermination sémantique et ne provoque chez l'auditeur ou le lecteur aucune difficulté d'interprétation. Que l'on compare ces deux exemples, écrits par des enfants de 10 ans :

- (a) Nous entrâmes par un trou à peine voyant car des genêts avait poussés par dessus Antoine y entra le premier et déboucha sur une galerie.
- (b) À ce moment là la petite fille futée arriva chez la mère-grand, elle entra.

Ce qui est remarquable, c'est que même si au plan formel, seul le premier énoncé réalise la valence locative de *entrer*, il apparaît qu'en termes sémantiques, la relation qui unit le procès du verbe et l'indication du lieu semble aussi aisément interprétable dans le second énoncé que dans le premier. Autrement dit, que l'anaphore soit marquée par *y* ou qu'elle soit non marquée ne semble pas faire une grande différence au plan de la signification exprimée.

Il apparaît donc que, lorsque sa réalisation n'est pas obligatoire au plan formel, le rendement sémantique de *y* locatif soit assez faible. C'est très certainement ce phénomène qui contribue à faire du *y* « facultatif » un excellent marqueur de style cérémonieux. En voici d'autres exemples :

En marchant il trouva une maison, elle était abandonnée, et curieux il décida **d'y entrer** pour voir ce qu'il y avait dedans. (écrit d'élève : fin cycle trois)

de ce poste en fait je retourmai euh plus ou moins à mon poste précédent alors euh j'y **restai** là euh encore une ou deux semaines avant de passer à d'autres emplois (Corpus Delic, poulets)

peut-être que qu'on pourrait s'y **retrouver** (Corpus Delic, répondeur)

3.3. Type « il y a parlé de ceci » : très formel

Lorsque l'élément locatif est extérieur à la structure valencielle du verbe, l'occurrence de *y* prend une dimension extrêmement cérémonieuse, et n'est guère attestée dans les textes oraux informels : il semble donc que le degré de formalisme induit par le *y* locatif soit inversement proportionnel à son caractère valencielle.

Voici plusieurs exemples de ces *y* de simple rection :

- à l'oral :

on **y parle** de chaussée routière parking planté piste cyclable (*y* : dans ce projet) (crfp, pubami1)

Les compléments de lieu réalisés par y : description des usages

on a un carnet de liaison donc chacun + **y retrace** un peu + ce que le travail qui a été effectué ce qui ce qui reste à faire (crfp, pronan1)

vous parlez monsieur Mazerolles de du rapport sur la France de l'An Deux Mille vous **y verrez** que il est recommandé que les salaires croissent moins vite que la productivité (Corpus Delic, économiste)

je connaissais l'Espagne hein + **j'y avais** des amis déjà un petit peu (crfp, pricle1)

on (n') ouvre pas un dictionnaire pour **y chercher** le sens d'un mot (crfp, pubcle1)

- à l'écrit (corpus journalistique) :

ce royaume de cinquante-cinq millions d'habitants a récemment battu tous les records : pendant trois années consécutives, de 1988 à 1990, le taux d'expansion **y a été** supérieur à 10 %. (Le Monde)

Mme Ghislaine Toutain, dynamique parisienne « parachutée » dans la Marne en 1986, a été largement distancée par M. Reyssier. Elle **y a perdu** son siège de député. (Le Monde)

le Conservatoire national de musique **y a donné** récemment quelques cours publics. [= dans une salle de spectacle] (Le Monde)

« no comment », nous **y a-t-on rétorqué** [= au consulat américain]

La journée du mercredi 27 mai, même sans Klaus Barbie laissé de nouveau dans sa prison puisque tel est son bon plaisir, fut encore dense et éprouvante. On **y abordait** les témoignages de la rafle des enfants juifs d'Izieu, dans l'Ain, le 6 avril 1944. (Le Monde)

Un juif américain sur trois s'est rendu au moins une fois en Israël. Les plus fortunés **y achètent**, de préférence face aux murailles de Jérusalem, un « pied à terre sainte », où ils vivent quelques semaines par an. (Le Monde)

la galerie du Kunstler, à Munich, a ouvert ses portes durant l'automne 1986 sur son exposition « Images-Digita »". On pouvait **y admirer** des travaux de Wolfgang Blobel. (Le Monde)

- à l'écrit (textes d'élèves de l'école primaire) :

Il était une fois un loup qui avait faim et qui allât se promener pour **y trouver** à manger (CE2)

j'avansais vers l'eau et **y trempas** mes pieds (CM2)

Camille va tous les matins dans la forêt pour **y cueillir** des fleurs, des fruits, etc [...] Un jour, ses parents, sa sœur et elle allèrent dans la forêt voisine pour **y ramaser** des champignons (CM2)

Frédéric SABIO

3.4. Remarques sur quelques lexèmes verbaux

Certains verbes ont un comportement particulier, vis-à-vis de l'emploi de *y*.

3.4.1. *Y arriver, y parvenir*

Le corpus oral présente une seule occurrence de ce verbe précédé d'un *y* locatif :

je le vois passer de l'autre côté de l'Allier il y avait au moins soixante mètres pour **y arriver** (Corpus Delic, saumon)

La rareté du phénomène a été confirmée par la consultation de l'ensemble de la base Corpaix (1 million de mots) qui n'offre que deux exemples supplémentaires de *y arriver* locatifs, dont l'un, produit par un avocat lors de sa plaidoirie, est extrêmement cérémonieux :

il rentre donc directement chez lui et **y arrive** dans les cinq minutes qui suivent (Corpaix, avocat, 33, 14)

ils ont tracé une route pour que les cars puissent **y arriver** (Corpaix, Jérusalem, 5, 5 : *y* fait référence à « cet endroit »)

La rareté de ce type de structure doit, semble-t-il, être mise en relation avec le fait que ce verbe connaît une construction de type *arriver à ça / à faire ça* qui, elle, se pronominalise très naturellement au moyen de *y* :

j'y suis **arrivée** et ça ça a marché du tonnerre + j'étais ravie (crfp, pripse3).

Il semblerait donc que, pour ce verbe, l'usage de *y* soit largement réservé, dans le français parlé informel, à l'expression de la réaction prépositionnelle.

Y parvenir se comporte de la même manière. Sur les sept exemples que compte *Le Monde*, aucun ne serait interprété comme un locatif :

Sergei était douanier, il veut être avocat. En Amérique évidemment. Mais il a peu de chances d'y **parvenir**. (Le Monde)

3.4.2. *Y entrer, y rentrer*

Le corpus oral ne comporte qu'un exemple pour chaque verbe. En y ajoutant les données de la base Corpaix et celles du corpus journalistique, on parvient à un total de 10 exemples, dont les huit qui suivent méritent un commentaire :

c'est un milieu dans lequel il faut faire son trou c'est très dur d'y **entrer** (crfp, pribor3)

Les compléments de lieu réalisés par y : description des usages

là on ne peut **y entrer** qu'à dix-sept ans (Corpaix, Alsace, 72, 16 : il s'agit d'une école)

il **y entre** au compte-gouttes (Corpaix, Chrabi, 18, 8 : il est question d'un livre qui a du mal à entrer au Maroc)

je vais présenter + les Beaux-Arts certainement Cergy + Art-Déco mais je sens que je vais me faire jeter mais minablement quoi + parce que c'est un peu + un peu chaud d'y **rentrer** (crfp, pripr1)

et pourquoi tu n'as pas pu **y rentrer** par rapport à ton âge (Corpaix, vendeur, 7, 8 : il est question d'entrer dans une école)

Julie, la trentaine, a toujours aimé la police. Elle rêvait d'y **entrer**. (Le Monde)

Je ne l'ai donc pas découverte en **y entrant**. (Le Monde : il est question d'une entreprise)

Les autres pourraient soit **y entrer** par la suite, soit rester en marge avec des statuts particuliers. (Le Monde : il est question de l'entrée de certains pays au sein de la Communauté Européenne).

Dans ces énoncés, le pronom *y* prend une caractéristique sémantique particulière : il est utilisé moins pour décrire le passage concret d'un lieu à un autre que pour indiquer le fait d'être accepté dans une institution en y obtenant un « droit d'entrer » : (*r*)*entrer dans un milieu, une profession, une école, un pays, une communauté...* (On note d'ailleurs que dans les exemples oraux, l'entrée en question semble toujours subordonnée à des règles restrictives : « c'est très dur », « c'est un peu chaud », « tu n'as pas pu par rapport à ton âge », « qu'à dix-sept ans », « au compte-gouttes »).

En revanche certains exemples d'allure banale, comme :

j'ai vu un café ouvert et j'y **suis entré**

paraissent peu attestés dans les productions orales informelles.

3.4.3. *Y lire*

Dans les quatre occurrences que compte le corpus (2 pour l'oral, 2 pour l'écrit), le locatif ne désigne jamais le lieu où se situe le lecteur (*le jardin, la bibliothèque...*), mais toujours le support physique sur lequel s'effectue la lecture : un rapport écrit, un ouvrage, un journal.

on peut **y lire** donc sur ces produits les olives qui sont utilisées (crfp, monpri2)

alors moi j'y **ai lu** des choses qui euh qui m'ont enfin je dirais qui qui me laissent euh réfléchir (*y* : dans ce projet) (crfp, pubami1)

Frédéric SABIO

Le policier n'est pas seulement un régulateur de la vie sociale, peut-on y lire. (dans un rapport officiel)

dans le Journal de l'armée, tous les aspects de la guerre moderne sont décortiqués. On y lit comme un message au pouvoir.

Voir et parler présentent à plusieurs reprises un fonctionnement similaire :

vous parlez monsieur Mazerolles de du rapport sur la France de l'An Deux Mille vous y verrez que il est recommandé que les salaires croissent moins vite que la productivité (Corpus Delic, l'économiste)

On y voit un détective privé, Merlot, remonter la « filière » des MST (maladies sexuellement transmissibles) contractées par une actrice de cinéma, Greta Barbo. (y : dans un album) (Le Monde)

Tout un chacun a le droit de mourir de rire devant les vaudevilles de Feydeau, et même, comme tant d'exégètes, d'y voir des chefs-d'œuvre... (Le Monde)

on y parle de chaussée routière parking planté piste cyclable (y : dans ce projet) (crfp, pubami1)

3.4.4. *Y voir, y reconnaître, y être impliqué*

Avec ces verbes, le pronom locatif acquiert une valeur sémantique que l'on peut assimiler à l'expression d'un processus ; cela est particulièrement net pour *y voir*, qui peut servir à évaluer un fait décrit par une séquence verbale, comme dans :

véhiculer ces cadavres sud nord puis redescendre pour aller enfin + se faire incinérer + à près de Vichy à trois cent vingt kilomètres d'ici + où est la logique + nous nous n'y voyons qu'une logique du profit (crfp, pubval1)

le procureur général se demande de nouveau pourquoi ni Barbie, ni Oberg, ni Knochen ne furent poursuivis pour crimes contre l'humanité. « J'y vois, dit-il, les effets de la pauvreté de la réflexion juridique française à cette époque ». (Le Monde)

Les gestionnaires de fonds sont prêts à accepter des conditions d'autant plus justes que la taille d'une transaction est élevée, car ils y voient un facteur de négociabilité des titres. (Le Monde)

Quand le procès n'est pas proprement verbal, l'élément lexical proportionnel à *y* a fréquemment le statut d'un nom d'action : *y voir cela, dans la politique menée, dans les décisions prises, dans une opération militaire envisagée...* :

Les compléments de lieu réalisés par *y* : description des usages

D'autres *y voient* plus de machiavélisme en interprétant les projets du Thatcherisme comme un moyen de museler une voix trop critique. (Le Monde)

Bien sûr, la fusion du groupe tricolore avec son homologue britannique explique, selon son PDG Jean-Marie Descarpentrie, cette décision de nature stratégique. Nombre d'observateurs *y voient* cependant un signe alarmant. (Le Monde)

Le Pentagone aurait fait valoir les difficultés techniques de l'opération. Les commandants de la FORPRONU (la force de l'ONU déjà sur place) *y* sont hostiles : ils craignent que les Serbes n'*y voient* une provocation. (Le Monde)

La retransmission télévisée des débats du Congrès a été une grande première pour le public soviétique. Les journalistes n'*ont-ils pas vu* une sorte de feu vert pour enquêter en profondeur ? (Le Monde)

Les éléments *y être impliqué, y reconnaître ceci* ont parfois un fonctionnement similaire :

Lors de notre congrès, 69 % d'entre eux viennent d'approuver l'action du SNES, simplement parce qu'ils *s'y reconnaissent*. (Le Monde)

La même complaisance n'est plus de mise lorsque des hommes proches de la gauche *y sont impliqués*. (il s'agit d'être impliqué dans des affaires politico-financières) (Le Monde)

On remarque que les séquences locatives de tous ces énoncés se pronominalisent en *là-dedans* : *ils sont impliqués là-dedans, ils se reconnaissent là-dedans, ils voient une provocation là-dedans*. La proforme interrogative la plus adéquate est *dans quoi* (plutôt que *où*) : *dans quoi sont-ils impliqués ? dans quoi se reconnaissent-ils ?*

On opposera donc deux usages de *y voir* : l'un, assez répandu, contribue à exprimer l'évaluation d'un processus (*j'y vois un signe encourageant, là-dedans, dans le fait qu'il ait accepté de me parler*) ; l'autre, quasiment absent du corpus, conserverait au pronom sa valeur de « lieu dans lequel s'inscrit le procès verbal » (*j'y ai vu une très bonne pièce, dans ce théâtre*).

4. Conclusion

Cette étude a montré que l'emploi du pronom de lieu *y* était lié à certaines restrictions.

Restrictions grammaticales d'abord : les compléments locatifs de simple réaction (c'est à dire « non essentiels ») ne sont

Frédéric SABIO

qu'exceptionnellement réalisés par *y* dans les productions informelles. À un degré moindre, les *y* valenciens dont la réalisation n'est pas obligatoire sont également peu nombreux dans les échanges oraux quotidiens. Le caractère cérémonieux, pour ne pas dire un peu artificiel, paraît dans ces deux cas si prononcé que les locuteurs semblent assez souvent développer d'habiles stratégies pour ne pas avoir à en faire usage : ce peut être en laissant l'information locative non exprimée (dire *je suis entré* plutôt que *j'y suis entré*) ; ou encore en convoquant d'autres moyens linguistiques (un autre pronom par exemple). L'anecdote suivante est à cet égard très instructive : j'ai entendu à l'issue d'un voyage en train la diffusion d'un avertissement à l'intention des passagers, exprimé une première fois par « Avant de quitter votre voiture, vérifiez que vous n'avez rien oublié *dans celle-ci* », puis « vérifiez que vous n'avez rien oublié *à l'intérieur* ». En revanche, il ne m'est jamais arrivé d'entendre « Avant de quitter votre voiture, vérifiez que vous n'y avez rien oublié ».

L'usage de *y* connaît aussi certaines restrictions d'ordre lexicosémantique : employé avec certains lexèmes verbaux, le pronom locatif paraît assez régulièrement induire une orientation sémantique particulière, comme on l'a illustré *supra* pour des tournures comme *y entrer*, *y lire*, *y voir quelque chose*, ... Avec ces structures, il apparaît que *y* ne se comporte pas comme un pronom locatif aussi polyvalent qu'on pourrait le penser, mais qu'il est comme « spécialisé » dans l'expression d'une sous-partie des informations locatives possibles.

Références

- Blanche-Benveniste, C., Deulofeu, J., Stéfani, J., & Van den Eynde K. (1984). *Pronom et syntaxe, l'approche pronominale et son application au français*. Paris : SELAF.
- Blanche-Benveniste, C., Bilger, M., Rouget, C., & Van den Eynde K. (1990). *Le français parlé, études grammaticales*. Paris : Éditions du CNRS.
- Blanche-Benveniste, C., Rouget, C., & Sabio, F. (2002). *Choix de textes de français parlé, 36 extraits*, Paris : Honoré Champion.
- Le Goffic, P. (1993). *Grammaire de la phrase française*. Paris : Hachette Éducation.

Quelques données sur les adverbes en *ment* dans le corpus de référence de français parlé

Mireille BILGER
Université de Perpignan

Résumé

L'objectif principal de cet article est de présenter un relevé des adverbes en « ment » rencontrés dans le *Corpus de référence de français parlé*. Nous avons supposé que ces formes devaient se distribuer de manière différenciée selon les types de corpus définis par les termes¹ de « Parole Privée », « Parole Professionnelle », « Parole Publique », et cela semble se vérifier. La confrontation avec les résultats obtenus par l'étude d'un corpus de presse écrite, nous a permis de noter « là aussi » une répartition différente. Dans cet article, nous ne chercherons pas à revenir sur les problèmes généraux de classification qui ont donné lieu à de nombreuses études². En revanche, nous compléterons cette étude quantitative par la présentation des principaux types de fonctionnement pour les formes adverbiales les plus représentées dans le corpus.

¹ Ces différents types de parole sont définis dans la présentation générale du *Corpus de référence de français parlé* qui est donnée dans ce numéro, mais nous en rappelons ici les grands traits : « la parole privée » correspond à des entretiens dans lesquels les locuteurs racontent un événement de leur vie ou présentent un savoir-faire professionnel ou autre. « La parole professionnelle » a été obtenue lors d'enregistrements des locuteurs dans l'exercice de leur fonction ou quand ces derniers parlent de leur profession sur le lieu même de leur travail. « La parole publique » correspond à des interventions d'un ou de plusieurs locuteurs dans le cadre d'un échange public : émission radio, réunion politique ou associative, etc.

² cf. entre autres les travaux de Nølke (1990), Molinier (1990), Guimier (1996), Nøjgaard (1992) ou encore Molinier & Levrier (2000).

Mireille BILGER

1. Résultats du dépouillement

Nous relevons pour l'ensemble du « Corpus de Référence » 243 formes différentes d'adverbes en « ment ». Mais pour 75% d'entre elles, on compte moins de 10 occurrences, c'est le cas, par exemple, de :

- *affreusement* 1 occurrence,
- *brusquement* 3 occurrences,
- *clairement* 9 occurrences.

La variété lexicale des formes adverbiales est différente selon les types de corpus. C'est le corpus de « Parole Publique » qui présente proportionnellement la plus grande variété, suivi de près par le corpus de « Parole Professionnelle », loin derrière se situe le corpus de « parole privée ».

	Publique	Professionnelle	Privée
<i>Nombre de mots</i>	66.300	68.700	247.000
<i>Nombre d'adverbes différents</i>	134	104	183
<i>Calcul sur la variété</i>	0,20%	0,15%	0,07%

Afin de rendre les fréquences comparables et de juger de la pertinence des différences rencontrées, nous allons poursuivre notre étude, non pas sur les effectifs observés dans les trois sous-corpus, mais sur ceux ramenés à 1 pour 100 000³. L'analyse des trois corpus donne les résultats suivants :

	Privée	Professionnelle	Publique
<i>Nombre total des occurrences</i>	826	815	1181
<i>Proportion par rapport au nombre de mots</i>	0,34%	1,18%	1,78%

Proportionnellement, le corpus « parole privée » compte 5 fois moins de formes adverbiales que le corpus « Parole Publique », et près de 3,5 fois moins que le corpus « Parole Professionnelle ». Ces écarts expliquent sans doute la différence en ce qui concerne la variété lexicale notée précédemment.

Si nous poursuivons l'étude, toujours sur les données ramenées à 1 pour 100 000, mais sur les seules formes présentant plus de 10 occurrences, nous obtenons :

³ Nous remercions grandement Jean Véronis pour nous avoir fourni l'ensemble des données sur lesquelles cette étude s'appuie.

Quelques données sur les adverbes en *ment*...

	Privée	Professionnelle	Publique
Nombre total des formes ayant plus de 10 occurrences et pourcentage par rapport au nombre total des occurrences (826, 815, 1181)	618 75%	587 72%	874 74%

Dans les trois types de corpus, on relève qu'environ un quart des adverbes présente un nombre d'occurrences ne dépassant pas 10. En revanche, les 3/4 qui restent révèlent une distribution différenciée, selon les types de corpus :

	Privée	Professionnelle	Publique
Total des occurrences des 10 adverbos les plus fréquents et pourcentage par rapport au nombre total des occurrences : 826/815/1181/	438 53%	400 49%	508 43%
Nombre d'occurrences de l'adverbe le plus fréquent et pourcentage par rapport au nombre total des occurrences : 826/815/1181/	158 19%	106 13%	92 7,80%
Nombre d'occurrences des 4 adverbos suivants et pourcentage par rapport au nombre total des occurrences : 826/815/1181	152 18,40%	172 21%	267 22,60%
Nombre d'adverbos de plus de 10 occurrences	23	23	33

Comme cela se vérifie souvent dans les distributions de l'oral, la plupart des emplois se concentrent sur un nombre réduit de formes. Ainsi, dans le corpus de « Parole Privée » 53% des emplois reposent essentiellement sur 10 adverbos ; cela se vérifie aussi pour le corpus de « Parole Professionnelle » (49%), et dans une moindre mesure, en ce qui concerne le corpus de « Parole Publique » (43%).

De fait, le corpus de « Parole Publique » se caractérise par une variété lexicale plus grande (33 adverbos différents) mais aussi par une répartition plus homogène des formes adverbales. L'adverbe le plus fréquent de ce corpus ne représente que 7,80% des emplois, alors que la forme la plus fréquente du corpus « Parole Privée » en totalise à elle seule 19%. Le corpus de « Parole Professionnelle » se situe, de nouveau, entre ces deux extrêmes (13%).

Mireille BILGER

1.1. Les formes du corpus « Parole Privée »

Dans ce corpus, on compte 23 adverbes totalisant plus de 10 occurrences⁴ :

N°	Forme	Occurrences
1	vraiment	158
2	justement	52
3	évidemment	37
4	tellement	34
5	finaleme nt	29
6	énormément	28
7	forcément	27
8	complètement	26
9	autrement	24
10	absolument	23
11	exactement	21
12	également	20
13	simpleme nt	16
14	malheureusement	15
15	pratiquement	15
16	effectivement	14
17	seulement	13
18	facileme nt	12
19	uniquement	12
20	vachement	11
21	généralement	11
22	actuellement	10
23	certainement	10

Si on compare ces différents taux de fréquence par rapport au nombre total des formes ayant plus de 10 occurrences dans ce corpus, à savoir 618, on note que c'est l'adverbe *vraiment* qui constitue plus de 25% des emplois adverbiaux recensés. La seconde forme la plus fréquente *justement* n'est présente que pour un peu plus de 8%. Les trois suivantes *évidemment*, *tellement*, *finaleme nt* ont des pourcentages qui se situent entre 6% à 4,5%.

⁴ Les indications chiffrées correspondent au nombre d'occurrences ramenées à 1/100000.

Quelques données sur les adverbes en *ment*...

1.2. Les formes du corpus « Parole Professionnelle »

Les adverbes les plus fréquents de ce sous-corpus sont là encore au nombre de 23 :

N°	Formes	Occurrences
1	vraiment	106
2	forcément	59
3	énormément	42
4	effectivement	39
5	justement	32
6	également	31
7	complètement	25
8	essentiellement	23
9	simplement	23
10	exactement	20
11	uniquement	20
12	absolument	17
13	normalement	17
14	tellement	16
15	autrement	15
16	directement	15
17	seulement	15
18	évidemment	13
19	pratiquement	13
20	carrément	12
21	facilement	12
22	relativement	12
23	malheureusement	10

Par rapport au total des occurrences qui s'élèvent cette fois à 587, l'adverbe *vraiment* est de nouveau la forme la plus fréquente mais dans une proportion moindre ; elle ne représente que 18% des emplois. La seconde forme *forcément* est utilisée dans 10% des cas ; et les trois suivantes *énormément*, *effectivement* et *justement* donnent des pourcentages allant de 7% à 5,45%.

1.3. Les formes du corpus « Parole Publique »

Dans ce corpus, les adverbes qui comptent 10 occurrences sont plus nombreux, on en relève 33 :

Mireille BILGER

N°	Formes	Occurrences
1	évidemment	92
2	vraiment	77
3	effectivement	75
4	également	62
5	justement	53
6	notamment	39
7	forcément	29
8	finalement	27
9	simplement	27
10	autrement	27
11	seulement	26
12	absolument	23
13	normalement	21
14	directement	21
15	exactement	21
16	éventuellement	20
17	complètement	20
18	extrêmement	18
19	sûrement	18
20	obligatoirement	18
21	particulièrement	18
22	essentiellement	17
23	actuellement	15
24	pratiquement	14
25	suffisamment	14
26	relativement	14
27	clairement	12
28	personnellement	12
29	rapidement	12
30	apparemment	11
31	certainement	11
32	probablement	11
33	totalement	11

Les formes adverbiales de ce type de corpus sont donc plus variées et elles se répartissent de manière plus homogène. Sur un total de 874 occurrences, l'adverbe le plus fréquemment employé est *évidemment*, mais pour seulement un peu plus de 10%, ce qui ne le différencie que de très peu par rapport à la seconde forme *vraiment* (près de 9%) et des suivantes *effectivement* (8,68%), *également* (7%) et *justement* (6%).

Quelques données sur les adverbes en *ment*...

Le tableau suivant donne le récapitulatif des 5 adverbes les plus fréquents dans chacun des sous corpus :

	Privée (618)	Professionnelle(587)	Publique (874)
vraiment	25%	18%	9%
justement	8%	5,45%	6%
évidemment	6%		10,5%
effectivement		6,5%	8,7%
tellement	5%		
finale ^{ment}	4,5%		
forcément		10%	
énormément		7%	
également			7%

2. Comparaison entre les sous-corpus

En tenant compte seulement des adverbes dont le nombre d'occurrences est égal ou supérieur à 10, l'influence du « genre » se fait nettement sentir : certaines formes sont peu représentées, voire absentes, dans tel ou tel corpus.

2.1. Corpus « Parole Privée » opposé aux deux autres

Comme on peut s'y attendre, l'adverbe *vachement* est peu employé dans le corpus « Parole Professionnelle » (4 occurrences) et inexistant dans le corpus « Parole Publique ». Mais cette différence de distribution concerne aussi un autre adverbe qui n'est pourtant pas marqué comme familier, c'est le cas de *généralement* (corpus « Parole Professionnelle » 7 occurrences ; corpus « Parole Publique » 5 occurrences).

D'autres adverbes, comme *tellement*, *énormément*, *malheureusement*, *facilement*, *uniquement* se retrouvent à des degrés divers dans les deux corpus « Parole Privée » et « Parole Professionnelle », mais sont peu présents dans le corpus « Parole Publique » ; le nombre de leurs occurrences se situant seulement entre 8 et 9.

En revanche, l'adverbe *finale^{ment}* qui est la cinquième forme la plus fréquente du corpus « Parole Privée » et la 8^{ième} du corpus « Parole Publique » est peu représenté dans le corpus « Parole Professionnelle » puisqu'on n'en relève que 6 occurrences.

Mireille BILGER

2.2. Corpus « Parole Professionnelle » opposé aux deux autres

L'adverbe *carrément*, 12 occurrences dans ce sous-corpus, est peu utilisé dans le corpus « Parole Privée », 7 occurrences seulement, et n'apparaît pas dans le corpus de « Parole Publique ».

Les formes *essentiellement* et *relativement* sont présentes dans les deux corpus « Parole Professionnelle » et « Parole Publique » mais rares dans le corpus « Parole Privée », le nombre des occurrences allant de 7 à 9.

2.3. Corpus « Parole Publique » opposé aux deux autres

L'adverbe *notamment*, la 6^{ième} forme la plus fréquente dans ce corpus, est peu représenté dans les deux autres corpus, le nombre d'occurrence est de 10 pour le corpus « Parole Privée » et de 9 pour le corpus « Parole Professionnelle »

Vient ensuite toute une série d'adverbes, comme :

- éventuellement
- sûrement
- obligatoirement
- particulièrement
- suffisamment
- personnellement
- probablement
- totalement
- rapidement
- apparemment

qui sont également peu exploités dans ces deux corpus. Par exemple, l'adverbe *personnellement* n'a qu'une seule occurrence dans le corpus « parole Professionnelle » et seulement 4 dans le corpus de « Parole Privée ».

Une forme est commune aux deux corpus « Parole Publique » et « Parole Professionnelle », l'adverbe *normalement* mais celui-ci est de moindre fréquence dans le corpus « Parole Privée », seulement 8 occurrences.

En revanche, l'adverbe *certainement* de fréquence quasi identique dans les corpus « Parole Publique » et « Parole Privée » est absent dans le corpus « Parole Professionnelle ».

Quelques données sur les adverbes en *ment*...

2.4. Récapitulatif

Formes fréquentes (10 occurrences ou plus), peu exploitées (moins de 10 occurrences) ou absentes dans les sous-corpus :

	Privée	Professionnelle	Publique
actuellement	+	-	+
apparemment	-	-	+
carrément	-	+	absente
certainement	+	absente	+
énormément	+	+	-
essentiellement	-	+	+
éventuellement	-	-	+
facilement	+	+	-
généralement	+	-	-
finalelement	+	-	+
malheureusement	+	+	-
normalement	-	+	+
notamment	+	-	+
obligatoirement	-	-	+
particulièrement	-	-	+
personnellement	-	-	+
probablement	-	-	+
rapidement	-	-	+
relativement	-	+	+
suffisamment	-	-	+
sûrement	-	-	+
tellement	+	+	-
totalelement	-	-	+
uniquement	+	+	-
vachement	+	-	absente

3. Comparaison corpus oral-corpus écrit

Après dépouillement des données globales, ramenées toujours à 1 pour 100.000, du corpus de français parlé, on relève que les 10 adverbes les plus fréquents sont :

- vraiment
- justement
- évidemment
- forcément
- effectivement

Mireille BILGER

- également
- énormément
- tellement
- complètement
- finalement

vraiment étant près de trois fois plus fréquent que la seconde forme *justement*, suivie de près par la troisième forme *évidemment*.

Dans le corpus de presse écrite, constitué essentiellement d'extraits du quotidien « Le Monde », et dont les données ont été également ramenées à 1 pour 100.000, les adverbess les plus fréquents sont :

- seulement
- notamment
- également
- actuellement
- vraiment
- finalement
- largement
- particulièrement
- simplement
- rapidement

Les trois premiers adverbess ont des taux d'occurrence qui sont respectivement les suivants :

- seulement (41,5%)
- notamment (37,5%)
- également (28,5%)

et ils sont eux aussi près de trois fois plus fréquents que les suivants. L'adverbe *actuellement* bien représenté (4^{ième} position, 15%) doit sans doute ce taux de fréquence à la nature même du corpus.

Si on compare les résultats, seuls trois adverbess sont communs aux deux listes :

- également
- finalement
- vraiment

les autres adverbess que l'on trouve dans la liste du corpus oral ont des rangs différents à l'écrit :

- *évidemment* arrive en 12^{ième} position ;
- *forcément*, *effectivement*, *tellement* et *complètement* se regroupent respectivement aux 33^{ième}, 34^{ième}, 35^{ième} et 36^{ième} positions ;

Quelques données sur les adverbes en *ment*...

- *justement* est loin dans la liste et se situe au 56^{ième} rang.

Enfin, l'adverbe *énormément* n'apparaît pas dans la liste des 100 adverbes les plus fréquents de ce corpus écrit.

En revanche, tous les adverbes relevés dans la liste du corpus écrit sont inclus dans la liste des 100 adverbes les plus fréquents du corpus oral, mais bien entendu ils apparaissent à des rangs différents :

- les adverbes *simplement*, *seulement* et *notamment* se regroupent aux places 14, 15 et 16 ;
- *actuellement*, *rapidement* et *particulièrement* occupent respectivement les rangs 24, 27 et 31 ;
- seul l'adverbe *largement* occupe une place éloignée dans la liste, la 58.

	Corpus oral	Corpus écrit
vraiment	1 ^{er} rang	5 ^{ième} rang
justement	2 ^{ième} rang	56 ^{ième} rang
évidemment	3 ^{ième} rang	12 ^{ième} rang
forcément	4 ^{ième} rang	33 ^{ième} rang
effectivement	5 ^{ième} rang	34 ^{ième} rang
également	6 ^{ième} rang	3 ^{ième} rang
énormément	7 ^{ième} rang	Absent de la liste des 100 premiers adverbes
tellement	8 ^{ième} rang	35 ^{ième} rang
complètement	9 ^{ième} rang	36 ^{ième} rang
finalement	10 ^{ième} rang	6 ^{ième} rang

Distribution dans le corpus écrit de la liste des 10 adverbes les plus fréquents relevés à l'oral.

	Corpus écrit	Corpus oral
seulement	1 ^{er} rang	15 ^{ième} rang
notamment	2 ^{ième} rang	16 ^{ième} rang
également	3 ^{ième} rang	6 ^{ième} rang
actuellement	4 ^{ième} rang	24 ^{ième} rang
vraiment	5 ^{ième} rang	1 ^{er} rang
finalement	6 ^{ième} rang	10 ^{ième} rang
largement	7 ^{ième} rang	58 ^{ième} rang
particulièrement	8 ^{ième} rang	31 ^{ième} rang
simplement	9 ^{ième} rang	14 ^{ième} rang
rapidement	10 ^{ième} rang	27 ^{ième} rang

Distribution dans le corpus oral de la liste des 10 adverbes les plus fréquents relevés à l'écrit.

Mireille BILGER

Pour conclure cette étude comparative, il est intéressant de noter que si on reprend la liste des 10 adverbes les plus fréquents dans chacun des sous-corpus oraux et qu'on la compare à celle du corpus de presse écrite, c'est le corpus « Parole Publique » qui s'en rapproche le plus, puisque nous y retrouvons 5 formes communes : *vraiment*, *également*, *notamment*, *finalement* et *simplement*. Ce résultat n'est pas pour nous surprendre⁵.

Le corpus « Parole Professionnelle » n'en partage que trois : *vraiment*, *également* et *simplement*.

Le corpus « Parole Privée » n'en fait apparaître que deux : *vraiment* et *finalement*.

4. Quelques observations sur le fonctionnement des adverbes les plus fréquents à l'oral

Si les adverbes en « -ment » semblent former une classe homogène en ce qui concerne la morphologie, il en est bien autrement du point de vue sémantique et syntaxique, et les propositions de classification varient souvent en fonction de la nature et des critères qui sont utilisés. Dans cet article, nous ne reviendrons pas sur ces classements, nous allons simplement rendre compte de quelques fonctionnements « remarquables » observés dans le corpus pour les 4 adverbes suivants :

- vraiment
- justement
- évidemment
- effectivement

en tenant compte essentiellement de leur mode d'apparition. Les réflexions sur la portée (référence sémantique) et l'incidence (référence à un support) de ces formes ne seront pas développées de manière systématique.

⁵ D'autres études ont montré que la parole « Parole Publique » partage effectivement des traits avec les productions écrites ; le nombre des formes sujets lexicales, du pronom sujet *nous*, du pronom relatif *dont*, des *ne* de négation, etc. est généralement beaucoup plus élevé que dans les productions de « Parole privée ».

4.1. L'adverbe *vraiment*

Tous les travaux semblent s'accorder pour donner à l'adverbe *vraiment* deux grandes significations : il est adverbe de « modalité » quand sa portée concerne l'ensemble de la construction, comme dans⁶ :

c'était pas mal l'Angleterre vraiment
vraiment il a l'impression qu'on s'occupe de lui
cet atelier vraiment je m'éclate

mais il peut aussi être interprété comme adverbe de « degrés » ou « caractérisant quantifiant » (cf. Wilmet, 1998) quand sa portée ne semble concerner que le terme qu'il précède, c'est le cas notamment quand celui-ci précède une forme adjectivale ou un participe passé :

l'année dernière c'était vraiment lourd
ce qui va être vraiment abordé ou discuté
c'était ravissant vraiment délicieux

Cette dernière interprétation d'adverbe de degrés avec une portée de proximité pourrait, semble-t-il, s'appliquer à des formes autres que les formes adjectivales et assimilées. On pourrait avoir cette même interprétation, ou aboutir à une interprétation de type « épistémique » dans :

j'ai eu vraiment des contradictions
il y a vraiment des problèmes sociaux
j'avais vraiment du mal à visualiser
on s'était retrouvé vraiment au cœur de l'Afrique
moi je fais la fête vraiment différemment

Dans ce cas, si on fait le test de l'extraction en *c'est ... que*, on note que l'adverbe peut accompagner l'élément extrait au même titre que n'importe quelle autre quantification :

c'est vraiment des contradictions que j'ai eues
c'est vraiment différemment que je fais la fête
c'est vraiment au cœur de l'Afrique qu'on s'était retrouvé

sans que la signification initiale change.

À noter enfin que *vraiment* a aussi des emplois de modifieur de la négation :

je suis vraiment pas en forme
j'ai pas du tout aimé vraiment pas
cette fille était allongée mais vraiment pas bien

⁶ Tous les exemples cités dans cet article sont tirés du corpus étudié.

Mireille BILGER

En définitive, comme le signale Abeillé (2002), l'adverbe *vraiment* est très ambigu et peut modifier toutes les catégories.

La place de l'adverbe est un des critères essentiels pour la classification syntaxique que l'on peut donner à la forme en question, cf. Wilmet (1998) et Le Goffic (1993). Trois places sont présentées comme pertinentes : au début de l'énoncé, à la fin ou à l'intérieur de la construction verbale⁷.

En ce qui concerne *vraiment*, les auteurs cités précédemment signalent qu'on le rencontre rarement à l'initiale. Cette distribution se vérifie dans notre corpus⁸. Peu nombreux sont les exemples du type :

vraiment on a passé une soirée extraordinaire
vraiment il y a pas un niveau hyper pointu

On en relève 18 exemples dans le corpus « Parole Privée », ce qui correspond seulement à 5,6% des occurrences, 4% dans le corpus « Parole Professionnelle » et 2% dans celui de « Parole Publique ».

Encore plus rares sont les exemples où l'on trouve *vraiment* en fin de construction :

j'ai eu une super grossesse vraiment
j'ai honte vraiment
c'était la meilleure année de ma vie sans conteste vraiment

Dans l'ensemble du corpus, nous n'en relevons que 5 exemples. À noter cependant que dans les transcriptions des productions orales, en dehors de toute écoute qui permettrait peut-être de lever l'ambiguïté, il n'est pas toujours aisé de savoir si *vraiment* se situe à la fin ou au début de la construction. C'est le cas pour :

c'était extraordinaire vraiment c'était une journée super

En fait, la seule distribution qui semble remarquable dans les 3 sous-corpus étudiés est celle qui concerne cet adverbe avec le verbe *c'est*. En effet, on note que dans le corpus « Parole Privée », les exemples du type :

⁷ La notion de place « initiale », « finale » et « interne » est à prendre dans le sens le plus ordinaire du terme. À noter cependant que nous avons comptabilisé comme « adverbe » de place initiale, ceux qui pouvaient être introduits par des connecteurs du type : *et, puis, enfin, alors, donc, bon*, etc.

⁸ Dans un corpus particulier constitué uniquement d'entrevues d'hommes politiques (édité par Paul Cappeau) cette place initiale est nettement plus représentée puisqu'elle correspond à près de 15% des emplois.

c'est vraiment particulier
c'est vraiment là
c'était vraiment la fête
c'était vraiment les débuts

représentent 25% des emplois de *vraiment*. Et cela représente aussi près de 21% des emplois en ce qui concerne le corpus « Parole Professionnelle », et plus de 15% en ce qui concerne le corpus « Parole Publique ». Il n'est pas facile de donner une explication à ce phénomène. Cependant, nous pouvons avancer une raison qui tient compte à la fois des caractéristiques de « c'est » et de *vraiment*.

Comme le dit Le Goffic (1993), le verbe *c'est* est « le marqueur type d'identification en français et il peut jouer ce rôle devant n'importe quel type d'élément ».

De même, nous avons vu que l'adverbe *vraiment* semble être une forme permettant de « modifier » de manière identique tout ce qu'elle précède, quelle que soit la catégorie morphologique du terme en question, contrairement aux adverbes tels que : *beaucoup*, *très*, *bien*, etc. Cette particularité de non contrainte partagée par les deux formes serait peut-être l'élément explicatif de cette « collocation » fréquente.

4.2. L'adverbe *justement*

Cette forme peut avoir deux statuts syntaxiques : être sélectionnée par le verbe ou être un adverbe de « modalité » qui porte sur l'énoncé global.

Dans les corpus que nous avons étudiés, rares sont les exemples où *justement* s'interprète comme réactionnel, et où il peut être glosé par *avec raison* ou par *avec justesse*, comme dans⁹ :

comme le disait très justement Henri Lachman
vous avez très justement remarqué
on s'insurge justement dans ce projet

En fait, comme cela a été souvent signalé, *justement* semble avoir pour fonction de souligner la pertinence de la prise de parole. Dans cet emploi, l'adverbe apparaît au début de l'énoncé :

L1 justement c'est l'éternel débat
L1 là justement dans ce cas-là l'anglais va avoir une fonction

⁹ Il est intéressant de noter que dans cet emploi *justement* peut être modifié par *très* ce qui ne semble pas être le cas quand il porte sur l'ensemble de l'énoncé. Il semble difficile d'avoir, par exemple : *j'étais très justement en stage* ou *très justement c'est l'éternel débat*.

Mireille BILGER

Cet emploi lié à la prise de parole est d'ailleurs bien représenté dans les corpus. Cela correspond à près de 30% des emplois de *justement* dans le corpus « Parole Publique »¹⁰, à 21% des emplois dans le corpus « Parole Professionnelle ». Ce pourcentage tombe à 11% pour le corpus « Parole Privée » mais cela peut s'expliquer par la grande proportion de monologues.

En dehors de cette distribution, *justement* se trouve réalisé à l'intérieur de la construction verbale :

la création d'O.G.M peut justement conduire à ...
ce péage ça handicape justement la circulation
j'étais justement en stage
c'est quelque chose qui est justement fait pour...

et quasiment jamais en fin de construction.

4.3. L'adverbe *évidemment*

Cet adverbe est également un adverbe de « modalité » qui porte sur l'énonciation. Dans les corpus oraux, la place initiale est bien représentée. Cela correspond à près de 40% des emplois du corpus « Parole Privée » et à près de 35% des emplois du corpus « Parole Publique » :

évidemment c'est déjà une ville pavillonnaire
évidemment c'est moi à chaque fois

La place finale est également une distribution récurrente mais de moindre importance, près de 17% pour le corpus « Parole Privée » et 11% pour le corpus « Parole Publique » :

je redevais le repasser évidemment
bon il y a le communiqué de l'E.T.A évidemment

Il est intéressant de noter également que, dans le corpus « Parole Publique », l'adverbe *évidemment* est dans près de 30% des cas accompagné de *bien* :

l'entreprise se doit bien évidemment de connaître
ce qui ne se retrouve pas dans les autres corpus.

¹⁰ On retrouve ce même pourcentage, près de 30%, dans le corpus des hommes politiques.

4.4. L'adverbe *effectivement*

Dans les deux sous-corpus où cet adverbe est bien représenté, on note que là encore la place initiale est celle qui est privilégiée ; on relève près de 62% d'emplois initiaux dans le corpus « Parole Professionnelle » et 45% dans celui de « Parole Publique » :

effectivement c'est prédéfini
effectivement il faut savoir reconnaître des cultures

la place finale est en revanche sous exploitée, si ce n'est dans le corpus « Parole Publique » où l'on relève 7% d'exemples de ce type :

ça peut être apporté par des gens extérieurs effectivement
je pense que tout le monde aura satisfaction effectivement

4.5. Récapitulatif

Ces quatre adverbes ont des distributions qui ne sont pas similaires. L'adverbe *vraiment* apparaît de manière privilégiée à l'intérieur de la construction verbale. *Justement* a un emploi qui favorise la place initiale, mais cet emploi n'est pas majoritaire et semble dépendre d'une situation de parole de type dialogique, sinon on le retrouve également plutôt inséré dans la construction. *Évidemment* présente une distribution nettement plus diversifiée, les emplois internes sont à peine majoritaires, la place initiale est bien représentée et, pour une fois, la place finale loin d'être négligeable. L'adverbe *effectivement* en revanche semble avoir comme position nettement marquée celle du début de construction.

	Position initiale	Position interne	Position finale
vraiment	-	+	-
justement	+	+/-	-
évidemment	+	+/-	+/-
effectivement	+	-	-

Tableau récapitulatif des tendances en ce qui concerne les positions

5. Conclusion

Cette étude quantitative des adverbes en « ment » montre l'intérêt qu'il y a à travailler sur des corpus de types différents ; la simple opposition « écrit-oral » ne peut être qu'en partie satisfaisante, tant il est vrai que ces

Mireille BILGER

formes se distribuent de manière quelque peu différenciée selon la situation de parole. Comme nous l'avons vu, la parole publique orale donne des résultats proches de ceux rencontrés dans l'écrit journalistique, à savoir : une variété lexicale plus grande ainsi qu'une distribution des formes plus homogène, ce qui ne se vérifie pas dans les deux autres types d'oraux.

Ce travail gagnerait, bien entendu, à être complété par l'étude d'un corpus plus vaste, mais aussi plus diversifié, notamment, en ce qui concerne l'écrit. On peut supposer, par exemple, qu'un écrit de type littéraire donnerait des résultats différents.

La description du fonctionnement des adverbes, telle que nous l'avons présentée, aurait mérité également d'être plus détaillée. Nous reconnaissons en avoir seulement esquissé les grandes lignes. En fait, chaque forme adverbiale pourrait supporter une étude monographique. D'autre part, il serait sans doute intéressant de vérifier s'il n'existe pas d'autres cas de collocations, moins évidents que ceux que nous avons signalés (*c'est vraiment ... et bien évidemment*) qui nous auraient échappé.

Ce travail, malgré toutes ces lacunes, a permis de montrer que cette catégorie grammaticale (les adverbes en « -ment »), est effectivement sensible aux « genres ». D'autre part, les données du corpus ont révélé que contrairement à ce que suggèrent les grammaires, ces adverbes dits « de manière » sont loin d'être par vocation des adverbes réactionnels modifiant le sens du verbe. Les formes les plus fréquentes apparaissent essentiellement dans des emplois d'adverbes de « phrase » ou « d'énonciation ».

Références

- Abeillé, A. (2002). *Une grammaire électronique du français*. Paris : Édition du CNRS.
- Guimier, Cl. (1996). *Les adverbes du français, le cas des adverbes en -ment*. Paris : Ophrys.
- Gross, M. (1990). *Grammaire transformationnelle du français, 3, syntaxe de l'adverbe*. Paris : Asstril.
- Le Goffic, P. (1993). *Grammaire de la phrase française*. Paris : Hachette.
- Molinier, Ch. (1990). Une classification des adverbes en -ment. *Langue Française*, 88, 28-40.
- Molinier, Ch., & Levrier, F. (2000). *Grammaire des adverbes. Description des formes en -ment*. Genève : Droz.
- Nølke, H. (1983). Les adverbiaux paradigmatiques : fonction et analyse. *Revue Romane*, 23, numéro spécial, 150.
- Nølke, H. (1990). Les adverbiaux contextuels : problèmes de classification. *Langue Française* 88, 12-27.

Quelques données sur les adverbes en *ment*...

- Nøjgaard, M. (1992). *Les adverbes français. Essai de description fonctionnelle*. Munksgaard : Copenhague.
- Wilmet, M. (1998). *Grammaire critique du français*. Duculot, Hachette 2^{ème} édition.

Fréquence des marqueurs discursifs en français parlé : quelques problèmes de méthodologie

Catherine CHANET
Université de Provence

Résumé

Cet article a pour double objectif de recenser les problèmes spécifiques posés par l'étude des marqueurs discursifs du français parlé dans les gros corpus, et de donner les fréquences des marqueurs discursifs les plus utilisés à l'oral. Après avoir tenté de définir ce qu'on entend généralement en pragmatique par « marqueurs discursifs », l'article recense les problèmes liés à la catégorisation de ces marqueurs en discours. Il donne alors les fréquences des 85 formes les plus utilisées comme marqueurs discursifs en français parlé, à partir d'un corpus de 1 500 000 mots. Une étude des différents statuts (marqueur ou non) que peuvent revêtir ces formes est ensuite conduite, pour détailler les statuts des quatre formes ambiguës les plus utilisées : *bon*, *bien*, *quoi*, et *enfin*, pour lesquelles les fréquences d'occurrences en tant que particules sont calculées sur le *Corpus de référence du français parlé* (440 000 mots), détaillé dans ce numéro.

1. Objectifs et position du problème

1.1. Objectifs

L'objectif de cet article est de recenser les problèmes méthodologiques que pose une description quantitative des marqueurs discursifs du français parlé sur gros corpus. Deux corpus d'interactions orales seront utilisés ici : le corpus Corpaix recueilli entre 1977 et 1999 par le GARS, qui comporte 1 050 000 mots, et le *Corpus de référence du français parlé* recueilli par

Catherine CHANET

l'équipe DELIC et décrit dans ce numéro (pp. 11-42), qui en comporte 440 000. Sauf indication contraire, les exemples cités seront tous extraits du *Corpus de référence du français parlé*.

L'étude des marqueurs discursifs sur ce type de données soulève des questions intéressantes qui n'apparaissent pas nécessairement lorsque les corpus d'étude sont de taille moindre et ne requièrent pas de procédures automatiques d'analyse. Ces questions sont de trois types :

- on rencontre d'abord des problèmes de reconnaissance, et donc de catégorisation, de ces marqueurs en discours : la taille des corpus exclut un relevé manuel des occurrences recherchées, et l'on ne dispose pas encore d'outils permettant d'extraire celles-ci automatiquement de façon satisfaisante ;
- en second lieu, la taille des corpus permet quelques constats d'ordre quantitatif sur la fréquence de ces marqueurs, qui vont parfois à l'encontre des idées reçues en linguistique, et qui, en tout cas, permettent de dégager des pistes de recherche concernant le fonctionnement de ces marqueurs à l'oral ;
- enfin, les emplois relevés apparaissent suffisamment variés pour envisager une description qualitative des fonctions de ces « marqueurs », en se posant la question des méthodologies adaptées à un tel objectif.

Je traiterai ici des deux premières questions, laissant le problème d'une description qualitative de ces marqueurs à des études ultérieures.

1.2. Qu'est-ce qu'un « marqueur discursif » ?

Il est difficile de donner d'emblée une définition de ce qui est entendu ici par « marqueurs discursifs » : c'est précisément l'établissement d'une telle définition qui soulève des questions de méthodologie. Pour néanmoins se faire une idée plus précise de cet objet d'étude, on conviendra que le terme « marqueurs discursifs » (désormais : « MD ») utilisé dans ces pages possède les deux caractéristiques suivantes :

- le terme « marqueurs discursifs », tel qu'il est employé ici, **ne recouvre pas** ce que l'analyse conversationnelle appelle traditionnellement « phatiques » et « régulateurs », lesquels constituent des signaux de pilotage de l'interaction destinés à maintenir/orienter l'échange. On distingue généralement (Cosnier, 1988 ; De Gaulmyn, 1987) les éléments « phatiques » (*tu sais, tu vois, hein, ...*), émis par le locuteur,

des éléments « régulateurs » (*hm hm, oui, d'accord, je vois, ...*) émis par l'allocutaire et assurant un rôle de feed-back interactionnel. Ces signaux de régulation ne sont pas seulement verbaux : les gestes et les mimiques (regards, postures, hochements de tête, sourires) jouent un rôle très important dans la synchronisation de l'interaction.

- le terme « marqueurs discursifs » **recouvre** à la fois ce que la littérature pragmatique appelle habituellement « connecteurs », et ce qu'elle appelle « particules » (au sens de Fernandez, 1994). Par exemple, *mais, alors, donc* (habituellement considérés comme connecteurs) font partie des MD, et *ben, voilà, ou en fait* (habituellement considérés comme particules), également.

Après avoir longtemps été définis comme reliant des propositions, des énoncés, ou des actes de langage, les connecteurs semblent désormais décrits par un certain nombre de linguistes (dont Berrendonner, 1983 : 237 ; Mosegaard-Hansen, 1998 : 73) comme des items reliant une information produite par le discours à l'ensemble des représentations mentales antérieurement construites par ce même discours. Quant aux particules, elles sont appréhendées (Fernandez, 1994 : 3 *sqq*) comme des items n'assurant pas nécessairement ce rôle de « liaison », mais donnant des informations sur les opérations conduites par le locuteur dans la construction de son discours.

Connecteurs et particules ont donc en commun le fait de constituer des unités non référentielles (n'ayant pas un signifié dénotatif mais plutôt instructionnel), et le fait d'agir sur les représentations cognitives construites par le discours, et dans la construction de ces représentations. Ces « marqueurs » donneraient des instructions sur la manière dont les interactants peuvent co-construire des représentations, les modifier, et les ajuster les unes aux autres (Chanet, 2001a). Autrement dit, les marqueurs discursifs n'interviendraient pas nécessairement dans la construction discursive d'un univers de référence, mais pourraient donner des indications sur la manière de construire cet univers, et, de façon plus générale, sur les opérations cognitives à conduire par les interactants dans l'activité discursive pour optimiser la communication.

Dans la pratique, la distinction connecteurs vs particules n'est pas si claire : d'une part, certains auteurs emploient le terme « particules » là où d'autres emploient celui de « marqueur discursif », et inversement ; d'autre part, il apparaît que, dans certains de leurs emplois, *mais, donc, ou alors* ont des fonctionnements pragmatiques similaires à ceux des particules, tandis que des formes comme *cependant, car, ou puisque* sont toujours considérées comme des connecteurs dans la littérature. On n'a donc pas affaire à des

Catherine CHANET

catégories disjointes entre connecteurs d'une part et particules d'autre part, ce qui justifie la propension de certains linguistes à les regrouper à l'intérieur d'une classe unique « marqueurs discursifs ».

1.3. Position du problème

Le problème que posent des unités de ce type est de déterminer en quoi, ou de quel point de vue, elles constituent une classe. La terminologie employée dans la littérature est à cet égard révélatrice : selon les observables choisis, les objectifs des auteurs et leurs théories sous-jacentes, on rencontre une ou plusieurs des dénominations suivantes : « mots du discours » (Ducrot et al., 1980), « connecteurs » (pragmatiques et/ou argumentatifs) (Roulet et al., 1985, et bien d'autres), « organisateurs textuels » (que Adam, 1990, oppose aux connecteurs), « opérateurs » (généralement opposés aux connecteurs : Anscombe & Ducrot, 1983, et aussi, dans un autre sens, Rossari, 1989), « marqueurs de structuration de la conversation » (Auchlin, 1981a, 1981b), « marqueurs de reformulation paraphrastique » (Gülich & Kotchi, 1983), « ponctuants » (Vincent, 1993), « particules » (énonciatives ou discursives) (Fernandez, 1994, Mosegaard-Hansen, 1998, et d'autres), « marqueurs discursifs » (Shiffrin, 1987), « marqueurs méta discursifs » (Mosegaard-Hansen, 1995), « marqueurs métalinguistiques » (Cadiot et al. 1985), voire « petits mots de l'oral » (Bruxelles & Traverso, 2001) ...

Dans le domaine de la littérature américaine, Fraser (1999) recense encore les appellations suivantes¹ : « cue phrases », « discourse connectives », « discourse particles », « discourse signalling devices », « phatic connectives », « pragmatic connectives », « pragmatic expressions », « pragmatic formatives », « pragmatic markers », « pragmatic operators », « pragmatic particles », « semantic conjuncts », « sentence connectives ».

On le voit, se repérer dans cette multitude d'appellations demande de définir précisément l'extension de chacun de ces termes, et la façon dont ils se distinguent par ailleurs des termes de « modalisateurs », « régulateurs », et « phatiques » qui recouvrent des objets plus étudiés par les conversationnalistes.

Cette fastidieuse liste non exhaustive permet de dégager deux propriétés attribuées par la littérature aux « marqueurs discursifs » :

¹ Je renvoie à l'article de Fraser (1999) pour les nombreuses références correspondantes.

- d'une part, les unités en question sont considérées comme relevant de l'étude du discours, et donc (pour les études) de la pragmatique principalement. Très peu de travaux traitent du lien entre rôle pragmatique des marqueurs discursifs et syntaxe, ou entre marqueurs discursifs et prosodie, par exemple, du moins en ce qui concerne le français². Aucune étude, à ma connaissance, n'envisage ces marqueurs en liaison avec l'ensemble des différents « niveaux » traditionnellement distingués en linguistique. Il semble donc généralement admis que le fonctionnement pragmatique d'unités de ce type puisse se décrire à l'intérieur même d'un seul « niveau » pragmatique, ce qui constitue une prise de position épistémologique qui ne va pas de soi. Si l'on considère, à l'inverse, que la pragmatique doit rendre compte d'une construction de l'interprétation qui fait intervenir tous les niveaux traditionnellement distingués dans les sciences du langage, on ne voit guère pourquoi les phénomènes d'ordre discursif n'intéresseraient pas également la syntaxe, la sémantique et la prosodie ;
- d'autre part, dans ce cadre pragmatique, les études adoptent des démarches hétérogènes, et peuvent s'attacher à caractériser des fonctions, des formes, ou les deux à la fois, ce qui donne lieu à trois types d'approches privilégiées :
 - une approche « descendante », qui étudie certaines opérations discursives (opérations de reformulation, ou opérations de structuration de séquences d'énoncés) à travers leurs traces en discours, parmi lesquelles figurent des connecteurs et/ou des particules (Gülich & Kotchi, 1983 ; Rossari, 1989, 1997; Auchlin, 1981a et 1981b ; Roulet et al., 1985) ;
 - une approche « ascendante », qui cherche à caractériser le fonctionnement pragmatique d'une unité donnée, à travers ses emplois en contexte (étude de *là*, Forget, 1989, de *enfin*, Cadiot et al., 1985, de *ben*, Bruxelles & Traverso, 2001, etc.) ;
 - une approche portant sur une relation entre un ensemble de formes et une fonction considérée comme sémantique (marqueurs de « conséquence », Hybertie, 1996, de « concession », Morel, 1996, de « cause », Nazarenko, 2000).

² On peut tout de même citer l'étude de Östman (1990) sur le rapport entre particules et prosodie. Plus récemment, Simon et Grobet (2002) ont cherché à mettre en relation emplois de *mais* et de *parce que* et prosodie.

Catherine CHANET

La question qui se pose alors est la suivante : le terme « marqueurs discursifs » désigne-t-il préférentiellement des formes, ou des fonctions résultant d'une interprétation ? La difficulté que recouvre cette question tient précisément au fait que la relation entre une forme et une « fonction » donnée n'est jamais bi-univoque, et que la littérature sur la question ne traite pas de catégories établies formellement indépendamment de toute considération pragmatique. Un problème immédiat se pose donc sur le plan du repérage même des MD, qui ne peuvent être recensés par leur seule forme.

Il apparaît en effet que bon nombre de formes ne constituent pas des MD dans tous leurs emplois, soit parce qu'elles subsument des fonctionnements morphologiques et/ou syntaxiques très différents (le *bon* adjectif de *c'est bon* et le *bon* particule de *bon enfin bref*), soit parce que leurs fonctions discursives sont très différentes (comparer les formes *enfin* dans *il est enfin arrivé* et dans *il a réussi ses examens, enfin je crois*). La classe des MD peut donc difficilement être décrite en termes extensionnels, ce qui pose problème pour la reconnaissance des MD dans les gros corpus.

2. Les formes susceptibles de fonctionner comme MD

Les outils dont on dispose actuellement pour repérer les MD dans les gros corpus se ramènent à deux grands types :

- des concordanciers permettant d'extraire des formes ou des lemmes ;
- des outils d'annotation de corpus aux niveaux morphologique et syntaxique.

L'application de concordanciers permet de recueillir les occurrences de formes susceptibles de fonctionner comme des marqueurs en discours. Étant donné qu'une forme donnée n'est pas toujours un marqueur discursif dans tous ses emplois, elle produit donc du bruit. L'idéal serait alors de disposer d'un moyen d'annoter les corpus en amont, en attribuant une catégorie spécifique aux MD dès le repérage des formes.

2.1. Catégorisations automatiques

Au niveau de l'annotation de ces formes, un problème se pose donc, qui est celui de leur éventuelle catégorisation, et de la pertinence de cette catégorisation. On sait que les « connecteurs » et « particules » sont

considérés par les grammaires comme relevant de différentes catégories grammaticales : conjonctions de subordination, conjonction de coordination, prépositions, adverbes, voire interjections. Les analyseurs morphologiques, élaborés pour fonctionner sur des textes écrits pour lesquels la catégorisation de ces types d'unités se pose moins crucialement, sont démunis face à des corpus oraux. À titre d'illustration, le tableau ci-dessous indique les catégories morphologiques qui ont été attribuées par Cordial 7 à certains items issus d'une interaction du *Corpus de référence du français parlé* :

ah bon	I I	Interjection + Interjection
alors	Rgp	Adverbe (autre que comparatif ou de négation)
aussi	Rgp	Adverbe
ben	Rgp	Adverbe
	Ncm	Nom commun (masc, nombre indéterminé)
bon	Rgp	Adverbe
	Afpms	Adjectif qualificatif non comparatif (masc sing)
c'est-à-dire	Rgp	Adverbe
d'ailleurs	Sp Rgp	Préposition + Adverbe
de toute façon	Sp Dt-fs-Ncfs	Préposition + Adjectif indéfini + Nom commun
donc	Cc	Conjonction de coordination
	Da-ms-d	Article défini (masc sing) +
du coup	Ncms	Nom commun (masc sing)
en fait	Sp Ncms	Préposition + Nom commun (masc sing)
enfin	Rgp	Adverbe
là	Rgp	Adverbe
mais	Cc	Conjonction de coordination
même	Rgp	Adverbe
par contre	Sp Ncms	Préposition + Nom commun (masc sing)
	Sp Sp	Préposition + Préposition
par exemple	Sp Ncms	Préposition + Nom commun (masc sing)
parce que	Cs	Conjonction de subordination
puis	Rgp	Adverbe
putain	Ncfs	Nom commun (féminin singulier)
quoi	Pr-.-	Pronom relatif (genre et nombre indéterminés)
	Pr-..a	Pronom relatif objet direct
	Pr-..d	Pronom relatif objet indirect
sinon	Cs	Conjonction de subordination
surtout	Rgp	Adverbe
voilà	Sp	Préposition
vraiment	Rgp	Adverbe

Tableau 1. Étiquetage des marqueurs discursifs de l'oral par Cordial 7 (crfp, pripcr1).

2.2. Problèmes

Ces catégorisations sont problématiques à plusieurs titres : elles peuvent ne pas tenir compte du fait que certains MD sont graphiquement complexes, ou distinguer des catégories pour un fonctionnement unique, ou encore, à l'inverse, ranger sous une même étiquette des fonctionnements très différents.

2.2.1. MD graphiquement complexes

Certains marqueurs discursifs dont la forme s'étale sur plusieurs mots graphiques sont traités comme des successions d'items. Si ce choix pourrait à la rigueur être pertinent pour la combinaison *ah bon*, il l'est beaucoup moins pour *d'ailleurs*, *de toute façon*, *du coup*, *en fait*, *par contre* et *par exemple*, qui se voient attribuer des chaînes de catégories morphologiques, et non pas une catégorie unique.

2.2.2. Cas des multi-catégorisations non pertinentes

Dans le cas où une même forme peut se voir attribuer plusieurs catégories exclusives les unes des autres (i.e. dans le cas où la forme est considérée comme ambiguë sur le plan morphologique), cette multi-catégorisation n'est pas toujours cohérente, dans la mesure où elle ne rend pas toujours compte d'une ambiguïté homonymique. Ainsi, *bon*, en dehors de son statut d'adjectif, est considéré comme un adverbe, mais il est étiqueté « interjection » dans la combinaison *ah bon*, sans qu'on sache ici quels arguments pourraient permettre de distinguer adverbe et interjection. Par ailleurs, la forme *ben* est étiquetée « adverbe » lorsqu'elle se situe en début d'énoncé :

- (1) L2 je sais pas combien ça fait mais sinon qu'est-ce que enfin + toi tu
sais un petit peu ce que tu aurais tu sais déjà ce que tu vas avoir
envie de faire + après plus tard avec + donc passer passer un
concours déjà tu sais quelle école tu veux avoir
L1 **ben** je + je vais présenter + les Beaux-Arts certainement Cergy
+ Art-Déco mais je sens que je vais me faire jeter mais
minablement quoi -

et (plus curieusement) « nom commun » ailleurs, comme dans les exemples suivants³ :

- (2) L2 et tu as déjà créé un monté un petit film là
L1 ouais ouais je suis en train là < je suis en train de le finir
L2 c'est vrai > + et vous étiez en équipe
L1 euh **ben** moi j'ai j'ai tourné tout seul + mais j'ai participé au film d'une copine + euh pour le le tournage quoi + euh sinon moi j'ai j'ai fait ça tout seul avec des enfin j'ai c'est moi qui filmais
- (3) L2 et vous y allez en rollers
L1 ah non il y a il y a ça ça par contre c'est tendu hein + parce que il y a de l'eau
L2 d'accord
L1 et + ah **ben** non c'est impossible de descendre en rollers + il y a des mecs qui descendent avec leur vélo mais je sais pas comment ils font quoi

2.2.3. Attribution non pertinente d'une catégorie

Enfin, le problème le plus ardu posé par cet étiquetage réside dans le fait que l'attribution d'une catégorie donnée peut être non pertinente en discours ou sujette à caution. Ce peut être le cas notamment pour *bon*, *enfin*, *là*, et *quoi*, pour ne citer que ceux-là.

2.2.3.1. Le cas de *bon*

Bon peut en effet être étiqueté « adjectif » ou « adverbe » selon le contexte, et il se trouve que la valeur d'adjectif n'est pas toujours justifiée, comme l'atteste (4), où *bon* se voit attribuer la valeur "Afpms" (adjectif qualificatif non comparatif), alors qu'il ne se comporte plus ici comme un adjectif :

- (4) L2 c'était quoi c'était un documentaire une petite fiction
L1 bah on avait un sujet c'était faire faire faire + c'est-à-dire filmer une action ou euh + c'était enfin c'est assez compliqué comme sujet et moi mon action c'était jeter la pierre et en fait je suis allé + j'ai fait de la pratique et + je suis allé jeter la pierre et je suis allé en parler avec enfin j'ai parlé avec des gens * j'ai interrogé des gens + et euh

³ Le Littré (1873) répertorie effectivement un « ben » substantif masculin, dont la définition est la suivante : « Nom donné au moringa oléifère de Lamarck, arbre dont le fruit, appelé noix de ben, fournit une huile employée principalement dans la parfumerie ». Le Robert (1989) ne le répertorie pas. Que ce terme soit toujours en usage ou pas, on peut se demander s'il est pertinent de le prévoir dans des étiqueteurs morphologiques devant s'appliquer à des textes autres que de spécialité.

Catherine CHANET

- L2 je- jeter la pierre mais parce que **bon** enfin tu tu le prends dans < dans le sens
- L1 dans le sens euh >
- L2 du l- du langage aussi < jeter la pierre à quelqu'un là
- L1 ouais ouais > ouais ouais jeter la pierre dans le sens quand on + on met la faute sur quelqu'un d'autre aussi quoi

2.2.3.2. Le cas de *enfin*

Enfin est systématiquement étiqueté « adverbe ». Si elle peut être justifiée sur le plan morphologique (notamment par l'invariabilité de la forme), l'attribution automatique de cette catégorie unique ne permet pas de différencier des comportements aussi distincts que ceux de (5), (6) et (7) :

- (5) il a **enfin** terminé sa thèse (exemple construit)
- (6) L1 ça c'est + et puis ensuite il y a l'intersaison + où se mélangent cette clientèle de l'hiver plus une clientèle touristique + et **enfin** l'été on peut dire les mois de juillet et août + où là il n'y a pratiquement qu'une clientèle touristique (prosail)
- (7) L2 je connais pas moi Rome
L1 ah c'est pff + je sais pas c'est + c'est indescriptible quoi **enfin** c'est pas ça ressemble à Paris et en même temps c'est complètement différent

Deux problèmes se posent ici :

- d'une part, la catégorie « adverbe », comme on le sait, regroupe des unités aux fonctionnements syntaxiques très différents selon leurs emplois. Ainsi, en (5), *enfin* est intégré au constituant verbal et joue le rôle de modifieur du verbe, alors qu'en (6) et (7), il échappe aux réseaux de dépendances syntaxiques ;
- d'autre part, on a ici affaire à des emplois discursifs différents, dans la mesure où seul le *enfin* présent en (7) serait considéré comme une particule (cf. 3.3.5).

2.2.3.3. Le cas de *là*

Le cas de *là*, également catégorisé « adverbe » de manière systématique par Cordial, est plus complexe, dans la mesure où il est assez difficile de trouver des critères qui permettraient de poser des catégories distinctes pour cette forme.

Forget (1989) souligne d'ailleurs une certaine « continuité » entre un *là* purement déictique spatial ou temporel (illustré par (8) et (9)), et un *là* (très

fréquent en québécois) qu'elle appelle « discursif », et qu'elle considère comme une particule (10) :

- (8) mets ça là, sur la table. (Forget, 1989 : 58)
- (9) - Quand est-ce qu'on y va ?
- On y va là, tout de suite. Prépare-toi. (Forget, 1989 : 58)
- (10) une fois que t'as ajouté tes trois œufs là, bon ben tu mélanges pendant cinq minutes (Forget, 1989 : 65)

Elle signale cependant que les deux *là* se différencieraient sur le plan prosodique.

Or, un logiciel comme Cordial étiquette *là* « adverbe » dans tous les cas. Ainsi, il met sur le même plan les emplois de *là* dans (11) et (12) :

- (11) L2 mais tu fais les circuits + enfin très très cachés quoi parce que normalement < c'est interdit quoi
L1 ah ouais ouais je suis je suis jamais allé > par Denfert-Rochereau là où les touristes ils rentrent quoi je sais même pas quelle tronche ça a par là quoi
- (12) L1 j'ai fait de la pratique et + je suis allé jeter la pierre et je suis allé en parler avec enfin j'ai parlé avec des gens * j'ai interrogé des gens + et euh
L2 je- jeter la pierre mais parce que bon enfin tu tu le prends dans < dans le sens
L1 dans le sens euh >
L2 du l- du langage aussi < jeter la pierre à quelqu'un là
L1 ouais ouais > ouais ouais jeter la pierre dans le sens quand on + on met la faute sur quelqu'un d'autre aussi quoi

2.2.3.4. Le cas de *quoi*

Enfin, la forme *quoi* est systématiquement catégorisée pronom relatif dans l'interaction passée au crible de Cordial, ce qui ne manque pas de poser un certain nombre de problèmes. Ainsi, *quoi* est considéré comme un pronom relatif par le logiciel, sans autre choix possible, dans les emplois suivants :

- (13) il y a tellement de trucs en fait c'est c'est c'est c'est c'est infini **quoi** les possibilités de cette ville
- (14) L2 il y a des carrières c'est impressionnant **quoi** + c'est + c'est genre + ah ouais c'est c'est c'est c'est il y a quinze mètres de plafond c'est des cathédrales sous terre **quoi** c'est
L1 ah ouais
L2 ouais c'est un truc de fou **quoi** + sérieusement **quoi**

Catherine CHANET

Or, il est évident qu'ici, *quoi* ne possède pas les propriétés d'un pronom, dans la mesure où il ne commute pas avec un autre pronom.

2.2.4. Bilan

Il semble que la recherche des MD à partir de leurs formes, et donc à partir des catégorisations morphologiques de ces formes, pose deux types de problèmes.

- Il existe tout d'abord des cas où le marqueur discursif ne peut plus être considéré comme relevant de la catégorie morphologique qui lui est attribuée : c'est le cas notamment pour *quoi*, qui perd ses propriétés de pronom lorsqu'il devient particule. Dans des cas de ce type, la catégorie morphologique de la forme lorsqu'elle ne constitue pas un MD est en quelque sorte incompatible avec l'emploi de cette forme comme MD : si *quoi* est pronom, alors il n'est pas un MD, et inversement. Il pourrait être utile, pour des cas de ce type, de prévoir dès l'analyse morphologique, le fait que la forme puisse être employée en discours avec un statut tel qu'on ne peut plus lui attribuer les catégories grammaticales habituellement prévues.
- Dans d'autres cas, la catégorie morphologique n'est pas en elle-même problématique : elle occulte simplement le fait que la forme peut également, tout en relevant toujours de la classe qui lui est attribuée, fonctionner comme un MD en discours. L'exemple de *enfin* illustre ce type de cas : il n'est pas incompatible que *enfin* soit catégorisé comme un adverbe sur le plan morphologique et soit employé en tant que particule dans le discours. Le seul problème qui se pose est de savoir dans quelles proportions, ou avec quelle fréquence, l'adverbe en question va fonctionner comme une particule.

La catégorie « marqueurs discursifs » s'avère donc être une catégorie discursive, dont la constitution résulte de l'interprétation en discours de certaines formes. Si l'on veut continuer à parler des MD en terme de « classe », il semble que l'on ait tout intérêt à différencier des **niveaux** de catégorisation ou d'étiquetage, entre lesquels on n'observe pas nécessairement d'isomorphisme : le terme « marqueurs discursifs » peut s'appliquer à certains éléments seulement de la catégorie habituelle des adverbes, des conjonctions, voire des pronoms, dans certains seulement de leurs emplois, où les catégories morphologiques « habituelles » de ces formes peuvent ne plus s'avérer pertinentes.

Je parlerai, dans ce qui suit, de **statut** d'une occurrence pour désigner la catégorie, quel que soit son « niveau » (morphologique / discursif), qui s'avère pertinente pour rendre compte du fonctionnement de cette occurrence donnée.

En l'état actuel des connaissances sur les MD, il n'existe pas d'autre moyen, pour analyser ceux-ci, que d'effectuer un repérage forme par forme, pour ensuite inventorier et classer les statuts de ces formes en discours, et ne retenir que ceux qui relèvent du marquage textuel des activités discursives.

3. Formes et statuts : aspects quantitatifs

3.1. Fréquence des formes

Un premier repérage sur gros corpus des formes susceptibles de fonctionner comme marqueurs discursifs permet de dégager des phénomènes intéressants concernant la fréquence de ces formes en français parlé. Il apparaît en effet que certaines formes pourtant très fréquentes ont été très peu étudiées dans le champ linguistique (comme *bon*, *là*, *quoi*, ou *ben*), alors que d'autres, qui apparaissent plus rarement à l'oral, font l'objet d'une littérature abondante (*car*, *à cause de*, *puisque*, par exemple).

Le tableau ci-dessous fait apparaître le nombre d'occurrences obtenues sur deux corpus de français parlé : Corpaix, (1 050 000 mots), et le *Corpus de référence du français parlé* (CRFP, 440 000 mots). L'ensemble des deux corpus totalise donc 1 500 000 mots. La somme des occurrences pour chaque corpus permet de classer les formes par ordre de fréquence. En se fondant sur un débit moyen de 200 mots par minute, on donne, dans la dernière colonne, la fréquence avec laquelle ces formes apparaissent dans la parole, en jours, heures, minutes et secondes. Il est bien évident que ces chiffres indiquent des moyennes qui supposent que l'on parle 24 heures sur 24 et qui ne tiennent absolument pas compte de variations sociolinguistiques ou idiolectales⁴.

⁴ Ces fréquences ne tiennent pas non plus compte de la répartition des occurrences au sein d'un même discours, selon les tâches (narration, argumentation, ...) qui y sont effectuées.

Catherine CHANET

Forme	CORPAIX CRFP TOTAL			Soit 1 tou(te)s les
1 mais	7520	3075	10595	42s
2 donc	4268	3807	8075	55s
3 alors	4476	1683	6159	1mn 13s
4 bon	3564	2144	5708	1mn 18s
5 là	3585	1536	5121	1mn 27s
6 bien	3337	1550	4887	1mn 32s
7 parce que	3155	1568	4723	1mn 35s
8 quoi	2366	1497	3863	1mn 56s
9 ben	2754	1001	3755	1mn 59s
10 puis	2542	1153	3695	2mn 01s
11 enfin	2429	1145	3574	2mn 05s
12 aussi	1888	1045	2933	2mn 33s
13 voilà	1674	928	2602	2mn 52s
14 après	1525	755	2280	3mn 16s
15 quand même	1254	702	1956	3mn 48s
16 en fait	1035	719	1754	4mn 13s
17 par exemple	943	425	1368	5mn 27s
18 c'est-à-dire	584	424	1008	7mn 22s
19 puisque	474	274	748	9mn 58s
20 surtout	488	168	656	11mn 23s
21 justement	438	186	624	11mn 57s
22 plutôt	372	142	514	14mn 31s
23 d'ailleurs	304	200	504	14mn 48s
24 ensuite	339	165	504	14mn 48s
25 disons	298	130	428	17mn 26s
26 sinon	263	128	391	19mn 05s
27 en plus	238	142	380	19mn 38s
28 finalement	203	93	296	25mn 13s
29 par contre	180	95	275	27mn 08s
30 seulement	209	60	269	27mn 44s
31 de toute façon	164	62	226	33mn 01s
32 car	160	41	201	37mn 08s
33 pourtant	112	33	145	51mn 28s
34 or	61	81	142	52mn 33s
35 de plus	93	37	130	57mn 24s
36 en tout cas	95	35	130	57mn 24s
37 notamment	60	57	117	1h 04mn
38 tandis que	72	42	114	1h 5mn 30s
39 malgré	46	44	90	1h 23mn
40 au fond	66	23	89	1h 24mn
41 ainsi	52	26	78	1h 46mn
42 à cause de	60	14	74	1h 41mn

Fréquence des marqueurs discursifs en français parlé ...

43	en effet	45	20	65	1h 55 mn
44	à propos	47	9	56	2h 13mn
45	du moins	34	20	54	2h 18mn
46	en ce qui concerne	40	13	53	2h 21mn
47	du coup	19	29	48	2h 35mn
48	bref	31	14	45	2h 46mn
49	malgré tout	20	24	44	2h 50mn
50	précisément	30	9	39	3h 11 mn
51	étant donné que	25	10	35	3h 33mn
52	en fin de compte	18	15	33	3h 46mn
53	après tout	23	10	33	3h 46mn
54	par conséquent	27	6	33	3h 46mn
55	tout de même	28	4	32	3h 53mn
56	par ailleurs	16	12	28	4h 46mn
57	concernant	17	7	24	5h 11mn
58	au contraire	0	23	23	5h 24mn
59	quant à	15	7	22	5h 39mn
60	si bien que	14	7	21	5h 55mn
61	en l'occurrence	8	12	20	6h 13mn
62	autrement dit	6	13	19	6h 33mn
63	au fait	11	7	18	6h 55mn
64	en réalité	14	4	18	6h 55mn
65	cependant	15	3	18	6h 55mn
66	de toute manière	8	9	17	7h 19mn
67	certes	8	8	16	7h 46mn
68	en revanche	13	3	16	7h 46mn
69	en tous les cas	11	4	15	8h 17mn
70	néanmoins	6	7	13	9h 34mn
71	inversement	11	1	12	10h 22mn
72	au total	6	5	11	11h 18mn
73	en somme	10	1	11	11h 18mn
74	n'empêche	7	3	10	12h 26mn
75	de ce fait	4	4	8	15h 33mn
76	au bout du compte	3	3	6	20h 44mn
77	en définitive	6	0	6	20h 44mn
78	de fait	6	0	6	20h 44mn
79	à l'inverse	0	5	5	1j 52mn
80	somme toute	1	3	4	1j 7h 06mn
81	de sorte que	1	2	3	1j 17h 27mn
82	à tel point que		2	2	2j 14h 11mn
83	toutefois	1	1	2	2j 14h 11mn
84	en outre	1	1	2	2j 14h 11mn
85	tout compte fait	1	0	1	5j 4h 22mn

Catherine CHANET

3.2. Quelques commentaires à propos de ces fréquences

- La liste des marqueurs recherchés n'est pas exhaustive : elle a été établie a priori en recueillant les formes examinées par la littérature, y compris dans des dictionnaires pour le Français Langue Étrangère comme celui de Grieve (1996). Elle a ensuite été enrichie en fonction des corpus.
- Les formes les plus fréquentes (apparaissant au moins toutes les 2mn dans la conversation) sont essentiellement monosyllabiques, comme on le vérifiera aisément sur les dix premiers items qui, si l'on excepte *alors* et *parce que*, ne comportent qu'une seule syllabe. Notons d'ailleurs que *alors*, *parce que* et aussi *enfin*, qui intervient en 11^{ième} position, peuvent se réaliser également sous les formes /lOR/, /pask/ et /fe~/ respectivement⁵, c'est-à-dire comme monosyllabiques. Les caractéristiques phonologiques des items ne suffisent pas, bien sûr, pour expliquer leurs fréquences : *car*, *or*, et *bref* sont monosyllabiques et pourtant plus rares.
- Parmi les formes les plus fréquentes figurent des items généralement considérés comme connecteurs (*mais*, *donc*, *alors*, *parce que*), mais aussi, de façon peut-être moins attendue, des formes à fonctionnement de particule (*bon*, *là*, *quoi*, *ben*). Or, la littérature sur l'enseignement du français (langue maternelle ou langue seconde) met exclusivement l'accent sur les connecteurs : les particules ne sont jamais traitées et sont généralement catégorisées comme des « disfluences » en quelque sorte parasites, alors qu'elles sont manifestement très employées.

Ces données brutes ne permettent cependant pas d'évaluer l'importance quantitative de l'emploi des MD à l'oral, dans la mesure où, dans beaucoup de cas, une même forme peut recouvrir différents statuts. Il convient donc, pour chaque forme, de répertorier les différents statuts que celle-ci peut revêtir en discours.

⁵ Notation phonétique SAMPA (voir Annexe 3 de l'article de présentation du *Corpus de référence du français parlé*, pp. 11-42).

3.3. Différencier des statuts possibles pour les formes

3.3.1. Objectifs

Parmi les occurrences repérées dans les transcriptions, toutes ne sont pas exploitables : certaines ont en effet été perçues avec plus ou moins de certitude par le(s) transcripteur(s) et figurent dans ce que nous appelons une multi-transcription, du type de celle mise en gras en (15) :

(15) non parce que ça ça tu dis trois cents kilomètres mais du coup /**enfin**,
ben/ il y a des rats

Par mesure de précaution, ces occurrences ont donc été écartées de celles à analyser.

La recherche de différents « statuts » potentiels pour une même forme a été effectuée sur toutes les occurrences exploitables des 11 items les plus fréquents dans le *Corpus de référence du français parlé* susceptibles de constituer des MD, de manière manuelle et totalement intuitive. L'objectif à terme de ce premier classement est d'essayer de dégager des critères rigoureux qui permettraient d'établir (ou de modifier, voire de contester) ces catégories intuitives. Seules les transcriptions ont été examinées : les statuts ont donc été établis en l'absence de toute indication prosodique.

Parmi ces 11 formes, 5 peuvent être ambiguës concernant leur statut, c'est-à-dire recevoir d'autres statuts en discours que celui de connecteur et/ou celui de particule : *bon*, *là*, *bien*, *quoi*, et *enfin*. Le cas de *là* ne sera pas examiné ci-dessous, en raison du fait que la prosodie est la plupart du temps nécessaire pour décider si cette forme a un statut de MD ou pas.

3.3.2. Le cas de *bon*

La forme *bon* peut a priori revêtir trois statuts différents au moins dans le discours : adjectif (*c'est un bon élève*), particule (*bon ben voilà*), ou membre d'une locution. Les cas qui ont été considérés comme « membre d'une locution » concernent *ah bon*, *pour de bon*, à *quoi bon*. La grammaire traditionnelle répertorie également un statut supplémentaire d'adverbe, dans des constructions du type *ça sent bon*, ou *il fait bon*, où *bon* peut en fait être vu comme un modifieur du verbe.

Si l'on excepte les occurrences ayant fait l'objet d'une multi-transcription (qui concernent 36 cas sur 2144, et qui laissent à 2108 le nombre d'occurrences exploitables), les statuts de la forme *bon* dans le *Corpus de référence du français parlé* se répartissent de la façon suivante :

Catherine CHANET

Occurrences exploitables	Particules	Adjectif	Dans locution	Ambigu	Adverbe
2108	1824	137	105	40	2
100%	86,54%	6,50%	4,98%	1,89%	0,09%

On constate donc que le statut d' « adverbe » est quasiment négligeable dans le corpus, alors que celui de particule y est très nettement prédominant (plus de 86%).

3.3.3. Le cas de *bien*

La forme *bien* est sans doute l'une des plus polyvalentes, dans la mesure où elle revêt des statuts multiples dans les corpus oraux :

- elle peut d'abord faire partie de ce que les grammairres considèrent comme une conjonction de subordination, telle que *bien que* ou *si bien que* ;
- elle peut également, dans *ou bien*, être membre d'une « conjonction de coordination » ;
- elle peut être un substantif (*avoir des biens, faire du bien*) ;
- elle entre ensuite dans des locutions plus ou moins figées telles que *bien sûr, bien entendu, voire bien évidemment* ;
- elle peut jouer un rôle d'adverbe et/ou de quantificateur, comme en (16) et (17) :
 - (16) L2 [...] je n'ai rien contre les chiens parce que moi j'ai eu des chiens hein et je les aime **bien** mais
 - (17) alors beaucoup d'entre vous ne savent pas très **bien** dessiner
- elle peut faire partie de la particule *eh bien* ;
- elle peut constituer une particule à part entière, comme en (18) :
 - (18) je voudrais vous dire donc en français + aujourd'hui quelques mots + relatifs + au thème + de la femme + dans le théâtre lopesque + ce qui nous conduit à tout naturellement bien sûr + à étudier ultérieurement + le personnage féminin de notre comedia{mot étranger} je parle ça va je parle pas trop fort + **bien** + donc si vous voulez + euh aujourd'hui laissons de côté le dramatis personae{mots étrangers} + nous y reviendrons ultérieurement + je vais donc vous parler + du thème de la femme + dans la production euh dramatique lopesque

Ces différents statuts sont ceux qui ont été observés sur la totalité des occurrences dans le *Corpus de référence du français parlé*, après avoir éliminé dix cas de multi-transcriptions. Ils se répartissent comme suit :

Occ expl	Adverbe / quantif	Dans locution	eh bien	Conj de subord	Conj de coord	Particule	Ambigu	Nom
1550	1019	300	138	26	20	19	15	3
100%	66%	19%	9%	1,7%	1,3%	1,2%	0,9%	0,2%

À l'inverse de ce qui se passe pour les autres formes susceptibles de constituer des MD, le statut de particule est très minoritaire à l'oral : si l'on additionne les occurrences de *bien* particule et de *eh bien*, on obtient un total de 10,2% des emplois où *bien* fonctionne comme une particule.

Ce faible score pourrait sans doute s'expliquer par le fait que *eh bien* est concurrencé à l'oral par *eh ben*, et que les cas où *bien* constitue à lui seul une particule se rencontrent surtout dans les discours pré-planifiés : la moitié des *bien* particules du corpus figurent dans des interactions publiques.

3.3.4. Le cas de *quoi*

Une précédente étude sur Corpaix (Chanet, 2001) avait montré que *quoi* se rencontre sous trois statuts différents dans les corpus de français parlé :

- proforme (ou pronom) ;
- membre d'une locution plus ou moins figée : *n'importe quoi, comme quoi* (sous certaines réserves) ;
- particule énonciative.

Après avoir éliminé 11 cas de multi-transcriptions (sur 1497 formes), voici comment se répartissent ces différents statuts dans le *Corpus de référence du français parlé* :

Occurrences exploitables	Particules	Pronom	Dans locution	Ambigu
1486	1134	246	70	36
100%	76,31%	16,56%	4,71%	2,42%

Il est intéressant de constater que la proportion de particules est quasiment similaire à celle qui avait été établie pour Corpaix (70,5%). On peut donc considérer qu'à l'oral, les trois quarts environ des *quoi* prononcés sont des particules.

Catherine CHANET

3.3.5. Le cas de *enfin*

À ma connaissance (Chanet, 2003), *enfin* est considéré dans la littérature comme pouvant recevoir quatre statuts différents :

- adverbe ne jouant pas le rôle d'un MD, considéré comme de sémantisme « temporel », et comme véhiculant des connotations de « soulagement » du locuteur :

il a **enfin** terminé sa thèse

- connecteur temporel, reliant chronologiquement des événements auxquels il est fait référence (c'est-à-dire se situant dans le monde construit par le discours) :

il a d'abord fait ça, puis ça, et **enfin** ça

- organisateur du texte (Adam, 1990), reliant chronologiquement des opérations discursives effectuées par le locuteur (c'est-à-dire se situant au niveau de l'activité discursive même), souvent dans une activité d'énumération :

à la base, d'abord, c'était un carré de coton bleu [...] ; puis se tenait au second étage un donjon de gâteau de Savoie [...] ; et **enfin**, sur la plateforme supérieure, on voyait un petit Amour [...] (Adam 1990 : 154)

- particule énonciative (Fernandez 1994)

il me semble que les gens au fond éprouvent ce genre de chose s'ils sont un peu an- peut-être un peu **enfin** anormaux en quelque sorte mais bon (Fernandez 1994 : 184)

Il se trouve que ces quatre statuts ne sont pas représentés à l'oral : on ne trouve pas de « connecteur temporel » dans les conversations en face à face, du moins dans le *Corpus de référence du français parlé*. Par ailleurs, comme le montre le tableau ci-dessous, il est assez frappant de constater que l'oral comporte presque exclusivement des particules :

Occurrences exploitables	Particules	Ambigu	Organisateur	Adverbe régi	Connecteur
1131	1109	19	2	1	0
100%	98,06%	1,68%	0,18%	0,08%	/

Ces chiffres laissent imaginer la gymnastique mentale que doivent effectuer les enfants passant d'une perception orale de la langue à un apprentissage de l'écrit, où *enfin* n'est particule que dans les simulations

d'oralité, et où la majorité des emplois sont ceux d'organisateur du texte, et d'adverbe ne jouant pas un rôle de MD⁶.

Par ailleurs, les organisateurs textuels sont également des marqueurs discursifs. Au total, il n'existe qu'une seule occurrence de *enfin* (sur 1131), dans le *Corpus de référence du français parlé*, dont on est sûr qu'elle n'a pas un statut de MD.

3.4. Fréquences des occurrences de MD

Les chiffres précédents montrent, s'il en était besoin, qu'il ne suffit pas de repérer une forme donnée pour repérer un MD. Bien plus, la fréquence du statut de MD pour certaines formes semble dépendre des genres discursifs et donc, à l'oral, des types d'interaction (cf. le cas de *bien*).

En ce qui concerne les fréquences de MD tous types d'interactions confondus, le tableau donné en 3.1. peut donc être revu, en éliminant les cas de multi-transcriptions et en ne retenant que les statuts de MD des formes.

Les fréquences obtenues dans le *Corpus de référence du français parlé* pour le statut de MD de *bon*, *bien*, *quoi* et *enfin* sont résumées ci-dessous :

Forme	Nb formes	Occurrences exploitables	Formes à statut de MD	Soit, en % des occurrences	Soit 1 tou(te)s les
bon	2144	2108	1824	86,54%	1mn 13s
bien	1550	1550	157	10,2%	14mn 10s
quoi	1497	1486	1134	76,31%	1mn 57s
enfin	1145	1131	1011	98,24%	2mn 12s

On le voit, les fréquences restent élevées et ne sont pratiquement pas modifiées, excepté pour *bien*. On peut donc en conclure que l'étude des marqueurs discursifs de l'oral reste une priorité dans la description du français parlé et pour l'enseignement du Français Langue Étrangère, étant donné la fréquence d'occurrence des MD à l'oral.

⁶ Un repérage dans la presse (*Le Monde diplomatique*) sur 150 occurrences avait permis d'établir que dans 50% des cas, *enfin* n'était pas un MD, et que dans 40% des cas, il fonctionnait comme organisateur textuel (Chanet, 2003).

4. Conclusions

Les fréquences des MD en français parlé montrent que l'on a tout intérêt, si l'on veut traiter de grands corpus, à pouvoir repérer ceux-ci de la manière la plus précoce possible. Les informations quantitatives, qui permettent de prédire la probabilité qu'une forme donnée soit employée comme un MD, peuvent être importantes : il est par exemple intéressant de constater que dans le *Corpus de référence du français parlé*, plus de 98% des *enfin* sont employés comme particules. Mais ces chiffres demandent d'une part à être confirmés par l'étude d'autres corpus, et d'autre part, à être établis sur des bases autres qu'intuitives.

Le travail qui reste à établir dans l'optique d'un éventuel repérage automatique des MD est considérable, car il nécessite un inventaire minutieux des propriétés « distributionnelles » des MD. Il apparaît en effet que ces propriétés sont hétérogènes, et que les particules, à la différence des connecteurs, sont toujours hors des dépendances syntaxiques, ce qui rend leur approche assez complexe. Pourtant, les particules n'apparaissent pas toutes aux mêmes endroits dans la chaîne parlée, et certainement pas de manière anarchique (Teston, 2002). On a affaire ici à des contraintes de positionnement qui dépassent (dans le cas des particules au moins) celles de la syntaxe, et qui restent à mettre au jour de manière systématique. En définitive, la description des marqueurs discursifs dans de grands corpus conduit à étudier ces unités dans tous leurs aspects (morphologiques, micro-syntaxiques, macro-syntaxiques, et discursifs) et, en constituant un lieu d'étude privilégié des pratiques de l'oral, questionne les méthodologies usuelles en pragmatique.

Références

- Adam, J.-M. (1990). *Éléments de linguistique textuelle*. Bruxelles : Mardaga.
- Anscombe, J.C., & Ducrot, O. (1983). *L'argumentation dans la langue*. Bruxelles : Mardaga.
- Auchlin, A. (1981a). *Mais hein, pis bon, ben alors, voilà, quoi !* Marqueurs de structuration et complétude. *Cahiers de Linguistique Française* 2, 141-159.
- Auchlin, A. (1981b). Réflexions sur les marqueurs de structuration de la conversation. *Études de Linguistique Appliquée* 44, 88-103.
- Bruxelles, S., & Traverso V. (2001). *Ben* : apports de la description d'un « petit mot », du discours à l'étude des polylogues. Revue en ligne *Marges Linguistiques* 2, 38-55. (<http://www.marges-linguistiques.com>).
- Cadiot, A., & Ducrot, O., Fradin, B., & Nguyen, T.B. (1985). *Enfin*, marqueur métalinguistique. *Journal of Pragmatics* 9, 199-239.
- Chanet, C. (2001a). Connecteurs, particules, et représentations cognitives de la planification discursive. In E.T. Nemeth (Ed.), *Cognition in language use : Selected papers from the 7th International Pragmatics Conference, Vol.1* (pp.44-55). Antwerp : International Pragmatics Association.
- Chanet, C. (2001b). 1700 occurrences de la particule *quoi* en français parlé contemporain : approche de la « distribution » et des fonctions en discours. Revue en ligne *Marges Linguistiques* 2, 56-80. (<http://www.marges-linguistiques.com>).
- Chanet, C. (2003). La forme *enfin* en français parlé contemporain : vers une typologie des statuts et des emplois. *8 Simposio de Comunicación social (Santiago de Cuba, 20-24 janvier 2003), Actas I*, Santiago de Cuba : Centro de Linguística Aplicada, 394-399.
- Cosnier, J. (1988). Grands tours et petits tours. In J. Cosnier., N. Gelas, & C. Kerbrat-Orecchioni (Eds.), *Échanges sur la conversation* (pp.175-184). Lyon : Éditions du CNRS.
- De Gaulmyn, M.-M. (1987). Les régulateurs verbaux : le contrôle des récepteurs. In J. Cosnier, & C. Kerbrat-Orecchioni (Eds.), *Décrire la Conversation* (pp. 203-223). Lyon : Presses Universitaires de Lyon.
- Ducrot, O. (1980). *Les mots du discours*. Paris : Minuit.
- Fernandez, M. M. J. (1994). *Les particules énonciatives*. Paris : Presses Universitaires de France.
- Forget, D. (1989). *Là*. Un marqueur de pertinence discursive. *Revue Québécoise de Linguistique* 18-1, 57-83.
- Fraser, B. (1999). What are discourse markers ? *Journal of Pragmatics* 31, 931-952.
- Grieve, J. (1996). *Dictionary of contemporary french connectors*. London / New-York : Routledge.
- Gülich, E., & Kotschi, T. (1983). Les marqueurs de reformulation paraphrastique. *Cahiers de Linguistique Française* 5, 305-346.
- Hybertie, C. (1996). *La conséquence en français*. Paris/Gap : Ophrys.
- Morel, M.-A. (1996). *La concession en français*. Paris/Gap : Ophrys.
- Mosegaard-Hansen, M.-B. (1995). Marqueurs métadiscursifs en français parlé : l'exemple de *bon* et de *ben*. *Le français moderne vol. 63-1*, 20-41.

Catherine CHANET

- Mosegaard-Hansen, M.-B. (1998). *The functions of discourse particles*. Amsterdam/Philadelphia : John Benjamins Publishing Company.
- Nazarenko, A. (2000). *La cause et son expression en français*. Paris/Gap : Ophrys.
- Östman, J.O. (1990). Particles and Prosody. On the language internal interactions of prosody and pragmatic particles. In J. Verschueren, (Ed.), *Levels of linguistic adaptation* (pp. 5-35). Amsterdam : John Benjamins Publishing Company.
- Rossari, C. (1989). Analyse contrastive en français et en italien de l'adverbial *enfin*. In Ch. Rubattel (Ed.), *Modèles du discours en Suisse romande* (pp. 215-223) Berne : Peter Lang.
- Rossari, C. (1997). *Les opérations de reformulation*. Berne : Peter Lang.
- Roulet, E., Auchlin, R., Schelling, M., Moeschler, J., & Rubattel, C. (1985). *L'articulation du discours en français contemporain*. Berne : Peter Lang.
- Schiffirin, D. (1987). *Discourse markers*. Cambridge : Cambridge University Press.
- Simon, A.-C., & Grobet, A. (2002). Intégration ou autonomisation prosodique des connecteurs. *Proceedings of the 1st International Conference on Speech Prosody, Aix-en-Provence (Speech Prosody 2002), 11-13 April 2002*, 647-650.
- Teston, S. (2002). *Éclairée par les apports de la pragmatique et de la prosodie, une analyse syntaxique des particules discursives*. Mémoire de D.E.A. de Linguistique Française. Université de Provence.
- Vincent, D. (1993). *Les ponctuants de la langue et autres mots du discours*. Québec : Nuit Blanche Éditeur.

Les formes disjointes des pronoms sujets

Paul CAPPEAU
Université de Poitiers

Résumé

La classe des pronoms personnels sujets disjointes¹ mérite d'être regardée de plus près². On connaît déjà bien les propriétés qui les différencient des pronoms conjoints (point 1). Pour autant, peut-on considérer qu'il existe une classe homogène des pronoms sujets disjointes ? Certes, il existe bien une seule catégorie morphologique mais peut-on soutenir que tous les éléments possèdent un même fonctionnement syntaxique ? Le travail sur corpus oraux présente l'intérêt d'apporter des précisions irremplaçables tant sur la répartition quantifiée de ces différentes formes que sur leurs contextes d'apparition. Cette description permet aussi de s'intéresser aux conditions externes qui semblent jouer un rôle dans la réalisation de ces pronoms et qui permettent d'expliquer les différences observées entre les banques de données utilisées (point 2). Au terme de cet examen, on peut espérer voir émerger un classement inédit de ces éléments.

1. L'opposition formes conjoints / formes disjointes

Les pronoms clitiques (ou conjoints) occupent une position qui est généralement présentée comme étant celle des SN. La correspondance que l'on serait tenté d'établir entre les formes sujets *ils*, *eux* et un syntagme

¹ Il s'agit des formes suivantes : *moi*, *toi*, *lui*, *elle*, *nous*, *vous*, *eux*, *elles* qu'on appelle parfois pronoms « forts » ou pronoms « toniques ». On utilisera « classe » pour désigner l'ensemble des éléments et « séries » pour renvoyer à des sous-ensembles particuliers que révèle cette étude.

² Deux articles de ce numéro viennent compléter la distribution des formes toniques : celui de Mylène Blasco-Dulbecco qui présente une étude détaillée de la séquence *moi je* et celui de Sandrine Caddeo qui porte sur certains cas de détachement.

Paul CAPPEAU

comme *les clients* s'appuie sur une sorte d'équivalence paradigmatique illustrée par les exemples suivants :

ils étaient là
eux étaient là
les clients étaient là

Creissels (1995) rappelle le caractère factice de cette présentation. Il pose deux emplacements distincts qui permettent de mieux rendre compte des spécificités de chaque catégorie. Il précise que :

« la commutation n'est qu'apparente : il y a là deux positions structurelles différentes, l'apparence d'une commutation tenant au fait que l'une des deux positions peut rester vide à condition que l'autre soit garnie » (p. 26).

Il distingue ainsi la position « du constituant nominal sujet (qui admet *moi, toi* ou *lui* mais pas *je, tu* ou *il*), et celle de l'indice de sujet (ici : *je, tu, il*) » (p. 26). Ces deux positions reçoivent la représentation suivante dans Creissels (1991 : 194) :

Constituant nominal	Indice de sujet	
sujet		
Michel	Ø	viendra
lui	Ø	viendra
Ø	il	viendra
(moi)	je	viendrai
(toi)	tu	viendras

Les pronoms disjoints, dans ce cadre-là, sont donc des représentants intéressants pour caractériser une position particulière du sujet. L'approche en termes de « zones » au sens de Skarup (1975) permet d'affiner la description et, notamment, de proposer un traitement amélioré des cas avec cumul de deux formes comme *lui + il*. En fait, pour rendre compte de la distribution des divers pronoms, il est nécessaire de disposer de trois emplacements distincts (au lieu des deux précédemment retenus) :

Extrapolation	Zone préverbale	Zone verbale	Zone postverbale	Extrapolation
(a)		il aurait gagné		
(b)	lui	aurait gagné		
(c) lui		il aurait gagné		

Les séquences réalisées dans les différentes zones n'ont pas les mêmes propriétés. Le sujet réalisé en zone verbale (a) est généralement un clitique.

Il se différencie du sujet réalisé en zone préverbale (b) par l'impossibilité d'être séparé du verbe par une insertion de type parenthèse, adverbe, etc., comme l'illustrent les deux exemples qui suivent :

* il en tout cas me parle d'une manière déplaisante
lui en tout cas me parle d'une manière enfin + (CA)

Le découpage en zones permet d'assigner un statut particulier au pronom disjoint lorsqu'il est cumulé avec un pronom conjoint. Dans un exemple tel que :

lui il a une formation euh une bonne formation + (CRFP)

on considèrera que la position sujet est « saturée » par la réalisation du pronom conjoint *il*. Le pronom disjoint *lui* est alors associé au clitique et sort du cadre des relations de dépendance syntaxique³. Dans ce qui suit, seule la forme *lui* réalisée en zone préverbale nous intéressera.

2. Paramètres d'observation

2.1. Présentation des corpus retenus

L'observation des exemples attestés s'appuiera, selon les cas, sur deux ou trois banques de données⁴ qui seront rapidement présentées :

- CA (Corpaix) est constitué de corpus très variés et comporte un million de mots⁵ ;
- CRFP est la nouvelle banque de données recueillie par l'équipe DELIC et présentée dans ce numéro de *Recherche sur le français parlé*. Il s'agit du *Corpus de référence du français parlé* qui comprend près de 440 000 mots. Les situations ainsi que les classes d'âge ont été orientées ;
- HP est un corpus de productions politiques réalisées dans les médias. Il comporte 300 000 mots. Il est peu utilisé ici et sert simplement de contrepoint pour illustrer certaines hypothèses.

³ On reviendra en 2.2. sur ce point.

⁴ Les deux premières banques de données (CA et CRFP) ont pu être dépouillées de façon exhaustive grâce aux concordanciers utilisés : respectivement, *Xcor* (développé par Jean-Pierre Adam) et *Contextes* (développé par Jean Véronis).

⁵ On renvoie à Blanche-Benveniste (2000) pour plus de précisions sur « l'histoire » de ce corpus.

Paul CAPPEAU

2.2. Les différents contextes à retenir

Les formes disjointes se réalisent dans de nombreux contextes syntaxiques. On présentera rapidement les emplois en relation avec la position sujet.

2.2.1. La position directement sujet

Dans ce cas, le pronom remplit la position sujet sans l'intermédiaire d'un quelconque support morphologique :

justement **lui** avait ce côté distant (CA)

2.2.2. Encadré par *c'est ... qui*

Cet environnement correspond notamment à la construction clivée, dans laquelle le pronom peut être considéré comme sujet et la séquence *c'est ... qui* comme une sorte de béquille morphologique (dans le sens où elle n'interfère pas avec les relations syntaxiques qui relient les différents constituants) :

ben oui c'est **moi** qui ai téléphoné (CRFP)

On verra que les données recueillies conduisent à reprendre cette présentation initiale et à identifier plusieurs structures différentes sous une seule trame morphologique.

2.2.3. Antécédent d'un relatif

Dans ce cas, le pronom disjoint n'est pas réellement le sujet et c'est le relatif qui assure cette fonction. On a toutefois pensé utile de retenir cet environnement pour deux raisons : les pronoms conjoints en sont exclus, il peut être intéressant de comparer cet environnement au cas précédent⁶.

et **moi** qui déj- qui aimais bien le cinéma j'avais donc le côté euh financier (CA)

⁶ L'analyse en terme de dispositif de la relative (Blanche-Benveniste et al., 1990) plaide aussi en faveur de ce choix.

2.2.4. La restriction

C'est un contexte qui présente une certaine parenté avec l'extraction. Le pronom disjoint, là encore, y remplit la fonction sujet grâce à l'appui d'un support morphologique :

il n'y a pas que **moi** qui crie (CA)

2.2.5. Les associés au sujet

À l'inverse, on n'a pas retenu les cas d'association dans lesquels le pronom disjoint est couplé à une forme clitique qui occupe la fonction sujet :

donc enfin **toi tu** es de Caen je crois (CRFP)

L'analyse en zones fournit un premier argument pour rejeter les pronoms disjointes réalisés dans cet environnement : c'est le clitique seul qui remplit la fonction sujet.

Certains exemples pourraient laisser croire que la relation entre la forme disjointe et la forme conjointe est un simple phénomène de répétition :

donc **eux ils** travaillent directement dans la vigne (CRFP)

Cette fausse impression repose sur le fait que le disjoint semble pouvoir se substituer directement au clitique. En fait, il s'agit là d'un cas particulier qui n'est en rien obligatoire (Blasco-Dulbecco, 1999). Les relations entre les deux formes échappent au strict cadre syntaxique⁷. L'exemple suivant peut laisser croire à une relation transparente entre conjoint et disjoint :

et lui il les faisait en doigté réel (CRFP)

a) et lui les faisait en doigté réel

b) et il les faisait en doigté réel

Les exemples ci-après montrent que cette « opération » grammaticale se délite progressivement et ne peut être contrôlée :

d'après ce qu'on dit le prix de l- du pain est libre mais + nous on l'augmente

tous les deux ans (CRFP)

nous un salaire moyen par exemple c'est + cinq mille (CA)

nous ça se passait comme ça (CA)

Ce contexte est apparu comme trop instable pour pouvoir être retenu dans ce travail et l'on a préféré se centrer sur les pronoms disjointes qui

⁷ C'est d'ailleurs la raison pour laquelle a été créée la notion d'« associé ».

Paul CAPPEAU

entrent plus étroitement en relation avec la position sujet. Finalement seuls les quatre premiers contextes ont été retenus.

2.3. L'existence de différentes séries

Il n'est pas possible de travailler en prenant en bloc la classe des pronoms disjoints. Trois sous-groupes peuvent être posés si l'on veut tenir compte des possibilités de réalisation dans les divers contextes syntaxiques présentés précédemment.

- Une première subdivision doit séparer les formes *moi*, *toi* des autres éléments. Les deux premiers pronoms ne peuvent en effet directement assumer la fonction sujet comme l'illustre l'opposition suivante :

* moi ai réussi / * toi as réussi
lui a réussi / eux ont réussi

- Pour un certain nombre d'éléments, le pronom conjoint et le pronom disjoint se présentent sous une forme semblable. Les exemples deviennent alors ambigus. Ainsi, en position directement sujet, il s'avère difficile voire impossible (en l'absence d'indication prosodique réellement fiable) de déterminer à quelle catégorie il faut rattacher le pronom *elle* dans l'exemple :

elle a réussi

- Pour le pronom masculin, cette incertitude est levée puisque la différence de catégorie passe par deux formes différentes :

il a réussi
lui a réussi

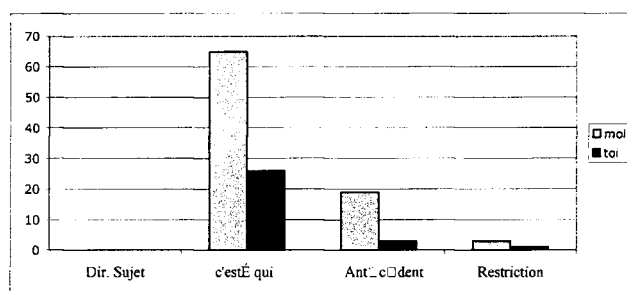
Par précaution, on ne travaillera que sur les formes non ambiguës (ce qui conduira à écarter des contextes pour certaines séries).

Sur la base de ces propriétés distributionnelles, trois séries se dégagent⁸ :

⁸ Voici la légende des tableaux : + possible et attesté, - impossible, (+) possible mais non attesté ou une seule occurrence rencontrée.

2.3.1. *moi* et *toi*

	directement sujet (a)	c'est PRO + qui (b)	PRO qui (c)	il n'y a que PRO qui (d)
moi	-	+	+	+
toi	-	(+)	+	(+)



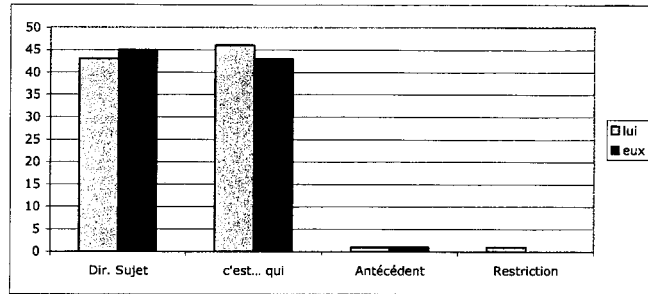
Graphique 1. Nombre d'occurrences des formes *moi* et *toi*

Ce graphique pourrait laisser croire que la différence entre *moi* et *toi* est une simple affaire de quantité. En fait, un examen plus précis montrera que les écarts touchent aussi la réalisation dans certains contextes.

2.3.2. *lui* et *eux*

	directement sujet (a)	c'est PRO + qui (b)	PRO qui (c)	il n'y a que PRO lui (d)
lui	+	(+)	+	(+)
eux	+	(+)	+	+

Paul CAPPEAU



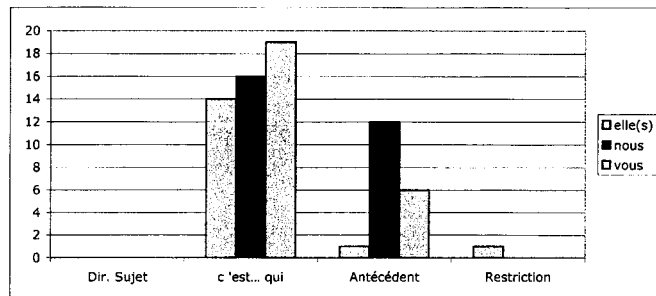
Graphique 2. Nombre d'occurrences des formes lui et eux

Dans cette série, on est frappé par les tendances quasi identiques que présentent les deux formes.

2.3.3. *elle(s), nous et vous*

	directement sujet (a)	c'est PRO + qui (b)	PRO qui (c)	il n'y a que PRO lui (d)
elle	non retenu	(+)	+	(+)
nous	non retenu	(+)	+	(+)
vous	non retenu	(+)	+	(+)
elles	non retenu	(+)	+	(+)

Pour le graphique on a regroupé les formes *elle* et *elles* que l'on rencontre peu dans les contextes examinés.



Graphique 3. Nombre d'occurrences des formes elle(s), nous et vous

Les formes disjointes des pronoms sujets

On peut apporter un autre regard sur les formes disjointes en les mettant en relation avec les formes sujets conjointes. On a limité les observations aux formes qui n'entrent pas en concurrence avec du lexique. Le tableau suivant effectué sur CA montre l'importance relative pour chaque personne des formes disjointes et, en particulier, le statut original de *nous*⁹ :

	1 ^{ère} pers. (<i>je & moi</i>)	2 ^{ème} pers. (<i>tu & toi</i>)	4 ^{ème} pers. (<i>nous</i>)	5 ^{ème} pers. (<i>vous</i>)
conjoint	11 009	5 950	849	3424
disjoint	56	19	19	15
%	0,51	0,32	2,24	0,44

3. Faits de distribution

Les diverses séries dégagées vont maintenant être présentées plus dans le détail. Deux aspects seront plus particulièrement commentés : les différences entre les contextes retenus et les écarts (plus ou moins sensibles) entre les banques de données utilisées.

3.1. La série *moi / toi*

Les données attestées pour la forme *moi* conduisent à préciser les contextes syntaxiques. L'importance de certains exemples incite notamment à affiner le découpage et à différencier, dans l'extraction, entre la polarité positive (*c'est moi qui*) et négative (*c'est pas moi qui*). Quatre contextes se dégagent¹⁰ :

	c'est moi qui		moi qui	il n'y a que moi qui	Total
	positif	négatif ¹¹			
CA	29	12	12	3	56
CRFP	23	1	7	0	31

⁹ Je remercie Berthille Pallaud qui a confirmé cette hypothèse à l'aide du test du Khi2.

¹⁰ Sauf indication contraire, tous les tableaux qui suivent comportent le nombre d'occurrences.

¹¹ On n'a pas tenu compte de la réalisation ou non du *ne* de négation.

Paul CAPPEAU

On fournit ci-dessous un exemple de chacune des rubriques :

c'est moi qui me la suis payée (CA)
c'est pas moi qui vais le chanter (CA)
moi qui aimais bien le cinéma j'avais... (CA)
il n'y a que moi qui peux les avoir (CA)

Les corpus CA et CRFP ne donnent pas exactement la même représentation de la langue. Les variations peuvent tenir à :

- une différence de taille qui fait que certains contextes rares vont « disparaître » quand la banque de données se rétrécit (ainsi le contexte *il n'y a que moi qui* se rencontre peu dans CA et est absent de CRFP) ;
- une différence dans la quantité de certaines productions argumentatives en particulier qui vont favoriser l'apparition de *c'est pas moi qui*. Cet emploi négatif est concentré dans des corpus plus formels de CA (un avocat qui expose son travail, une responsable de holding...).

À titre de vérification, on a cherché à voir la distribution de *moi* dans le corpus d'hommes politiques (HP) :

	c'est moi qui	c'est pas moi qui	moi qui	il n'y a que moi qui
HP	3	6	4	1

Ce tableau conforte l'hypothèse précédente (même si le nombre d'attestations doit rendre prudent...) : les caractéristiques des productions semblent bien influencer sur la réalisation de certains contextes. On peut supposer que le style délibératif, le caractère argumentatif contribuent à augmenter les occurrences de la tournure *c'est pas moi qui*.

Une différence se manifeste quand on élargit le contexte dans lequel apparaît la séquence *pas moi*. Dans les corpus CA, le terme positif qui s'oppose à *pas moi* est souvent réalisé dans le voisinage immédiat¹² :

c'est pas moi qui le dis c'est vous (CA)
c'est pas moi qui ai demandé à aller là-bas c'est eux (CA)

¹² Dans une étude plus générale portant sur les modalités, Sabio (2002) a remarqué que l'ordre de succession est généralement « partie négative + partie positive ».

Dans les productions HP, le terme de contraste reste le plus souvent implicite :

oh ben écoutez ce n'est pas moi qui à Essen ai dit je peux vous dire que monsieur (HP)
 L1 [...] ce n'est pas moi qui pense je lis et je regarde L2 [...] (HP)

Une observation plus fine des données permet aussi de dégager un point commun : la forte représentation du verbe *faire* qui se retrouve dans 20 % des exemples de CA et dans plus de 30 % des exemples de CRFP¹³ :

c'est moi qui le faisais à la maison (CA)
 c'est moi qui fais le dossier directement (CRFP)

Pour la forme *toi*, les deux banques de données donnent des résultats comparables, ce qui n'est pas surprenant : les caractéristiques externes des productions dans lesquelles apparaît la deuxième personne sont certainement plus semblables. La différence de taille entre les deux corpus peut expliquer pourquoi un contexte rare dans CA (une occurrence) n'est pas attesté dans CRFP :

	c'est toi qui	c'est pas toi qui	toi qui	il n'y a que toi qui	Total
CA	17	0	1	1	19
CRFP	9	0	2	0	11

et les textes c'est toi qui vas les chercher (CRFP)
 et et toi qui es qui es branché + astrologie + étoiles etc. (CRFP)
 il n'y a que toi qui étais prête à repartir probablement (CA)

Si l'on regarde de plus près leur distribution, les deux unités de cette série (*moi* et *toi*) se distinguent assez fortement. Sous une même apparence morphologique, il existe deux contextes syntaxiques :

- *c'est moi qui* correspond généralement à un cas d'extraction (ce que traduit bien la forte proportion de suites avec formes positives et négatives). Le clivage sert alors à singulariser une personne et à l'opposer à d'autres. Cette particularité d'emploi se retrouve à plusieurs reprises dans les exemples relevés :

c'est pas moi qui le dis c'est vous (CA)
 c'est pas moi qui ai demandé à aller là-bas c'est eux (CA)

¹³ À titre de comparaison, si l'on prend l'ensemble des verbes ayant pour sujet *je* ou *j'*, le verbe *faire* représente 4,4 % des cas.

Paul CAPPEAU

- *c'est toi qui* est une tournure principalement utilisée dans les questions (ce qui n'était pas le cas de la séquence *c'est moi qui*). On identifie ce contexte par le changement de locuteur qui suit :

- L1 c'est toi qui fixes ton emploi du temps L2 non non (CA)
- L3 c'est toi qui donnes qui donnes ce cours-là en fait L1 non (CA)
- L2 ben c'est toi qui l'as écrit eh
- L1 oui c'est moi qui l'ai écrit (CA)

Dans ce cas, il paraît exclu d'analyser ce tour comme une extraction : le pronom n'est pas réellement opposé à d'autres personnes. L'absence d'une modalité négative plaide dans le même sens. Cet emploi, a priori envisageable, n'est, de fait, pas présent dans nos relevés :

c'est pas toi qui devais y aller c'est moi (non attesté)

La tournure *c'est toi qui* est plutôt utilisée comme un mécanisme d'identification que de contraste. On retrouve cette même tendance pour la tournure *c'est vous qui*.

Un autre point différencie ces deux pronoms : la séquence en apposition au sujet clitique est assez productive avec *moi* :

moi qui fais de la danse j'arrivais pas à suivre (CA)
moi qui suis au au-delà du SMIC par exemple + je gagne douze mille francs (CRFP)

Avec la forme *toi* elle est exceptionnelle (un seul cas attesté dans nos données) :

et toi qui as vécu en Angleterre peux-tu nous dire (CA)

À partir de ces faits, il semble préférable de ne pas présenter *moi* et *toi* comme appartenant à une même série.

3.2. La série *lui / eux*

On rencontre les formes *lui* et *eux* dans les quatre contextes syntaxiques retenus, c'est-à-dire :

- directement en position sujet :
je tiens le magasin aussi et lui est magicien et jongleur (CRFP)
- dans l'extraction :
c'est lui qui gère toute la technique (CRFP)
- comme antécédent du relatif :
et il avait même fait lui qui n'était pas très bricoleur (CA)

Les formes disjointes des pronoms sujets

- dans la restriction :
il n'y a que lui qui peut trancher

Les tableaux ci-dessous indiquent les résultats obtenus avec les formes *lui* et *eux* :

	lui + V	c'est lui qui	lui qui	il n'y a que lui qui	Total
CA	23	37	1	0	61
CRFP	20	9	0	1	30

	eux + V	c'est eux qui	eux qui	il n'y a que eux qui	Total
CA	31	26	? 1	0	58
CRFP	14	21	0	0	35

Les deux banques de données présentent des tendances comparables :

- la place d'antécédent du relatif est soit absente, soit totalement marginale, de même que la restriction ;
- la forme *moi* avait orienté vers l'hypothèse d'un statut original de l'extraction négative. Cela n'est plus le cas pour les pronoms *lui* et *eux* ;
- deux contextes sont bien représentés (pronoms directement en position sujet, extraction positive).

Lorsque le pronom remplit directement la place sujet, il est assez rare qu'il soit séparé du verbe par une insertion¹⁴ :

et eux en fait suivent ce qu'on leur raconte (CA)

De même les formes modifiées ou prolongées des pronoms sont extrêmement rares. Seules les formes *eux aussi* et *lui aussi* ont été rencontrées :

mais euh d'un autre côté eux aussi auraient pu réagir (CA)
parce que lui aussi était euh assez nerveux (CA)

Le travail sur corpus permet de dégager les spécificités des deux emplois les plus fréquents et de repérer des fonctionnements très différents dès que l'on élargit le contexte d'apparition. Quand *lui* est directement

¹⁴ On ne peut donc envisager de retenir cette propriété (postulée d'après le découpage en zones) pour pouvoir distinguer, dans la série *nous, vous, elle(s)*, les formes conjointes et disjointes.



Paul CAPPEAU

sujet, il entre dans une configuration de contraste. On le rencontre souvent en opposition avec les formes de première personne :

lui a arrêté et **moi** je continue (CRFP)
moi je suis arrivé au collège + et **lui** est arrivé au collège deux ans après (CRFP)
j'écrivais les duos pour les chanteurs et **lui** euh X allait chanter chez Jacques Chancel (CA)

L'extraction se rencontre plutôt dans les cas de validation ou de confirmation d'un référent de troisième personne déjà indiqué. Ce n'est pas l'emploi contrastif, d'opposition entre deux formes, qui domine :

j'en ai fait plusieurs fois enfin mon oncle + il /il, 0/ sait bien en faire euh + c'est lui qui m'a appris au début (CA)
/c'est, 0/ bizarre d'ailleurs qu'il soit pas venu monsieur G. ce matin parce que c'est lui qui a insisté (CA)

Autre particularité, les emplois dans l'extraction sont souvent regroupés :

L2 ah oui c'est pas pareil mais n'importe comment le transport il est c'est **lui** qui l'a payé
L1 oui c'est **lui** qui le charriait (CA)
L1 lui il fait le salon + il fait la chambre à la gamine et puis + puis après il aspire + c'est **lui** qui aspire tous les tapis et tout ce qui s'ensuit la poussière
L2 c'est quand même autre chose hein
L1 et c'est **lui** + c'est **lui** qui balaie hein (CA)

3.3. La série *nous / vous / elle(s)*

Pour cette série, les données attestées sont peu nombreuses, ce qui rend délicat tout travail de description. On peut, à la rigueur, s'appuyer sur les formes déjà rencontrées pour voir si un certain parallélisme peut être établi.

Pour les formes *elle* et *elles* regroupées dans ce tableau, c'est l'emploi dans l'extraction qui est dominant :

Les formes disjointes des pronoms sujets

		elle(s) + V ¹⁵	c'est elle(s)	elle(s) qui	il n'y a que	elle(s) qui	Total
		qui					
CA	non retenu	10		1		1	12
CRFP	non retenu	4		0		0	4

Le plus souvent, l'extraction semble avoir une valeur de contraste (ce qui n'était pas forcément le cas pour la forme *lui*). On peut supposer que le système du masculin qui permet de jouer sur deux contextes distincts (la position directement sujet, l'extraction) sépare des fonctionnements plus imbriqués dans le cas du féminin. Ainsi, la valeur de validation d'un référent déjà mentionné se rencontre bien mais elle reste liée à un contraste :

après j'ai ma collègue qui arrive à deux heures donc c'est elle qui s'occupe de la blanchisserie moi je m'en vais à deux heures et demie (CRFP)

On retrouve cette valeur de contraste, d'opposition entre personnes, dans les cas d'extraction à modalité négative ou dans des constructions parallèles :

et si c'était pas elle qui le disait c'était ma marraine + ou ma grand-mère (CA)

si l'assistante sociale connaît pas du tout la famille c'est moi qui fais le dossier directement si l'a- l'assistante sociale de secteur connaît déjà la famille elle suit la famille + c'est elle qui me fait le dossier (CRFP)

Dans les contextes retenus, les pronoms *nous* et *vous* sont, eux aussi, particulièrement peu attestés, ce qui rend une nouvelle fois difficile d'indiquer des tendances.

		nous + V	c'est nous	nous qui	il n'y a que	Total
		qui				
		nous qui				
CA	non retenu	11		8	0	19
CRFP	non retenu	11		4	0	15

		vous + V	c'est vous qui	vous qui	il n'y a que	Total
		vous qui				
CA	non retenu	11		4	0	15
CRFP	non retenu	11		2	0	13

¹⁵ On n'a pas trouvé de pronoms avec prolongement (*elle-même, elle aussi ...*) directement en position sujet. Dans CRFP, la seule forme recueillie est en position associée :

alors elle-même elle avait hérité des tableaux de son père (CRFP)

Paul CAPPEAU

Les données orales conduisent à remettre en cause la présentation habituelle dans laquelle on considère que *nous* et *vous*, quand ils sont disjoints, pourraient être directement sujets. Il est exclu, à partir de transcriptions, de se prononcer sur le caractère tonique du pronom mais on dispose d'indices indirects :

- On ne trouve pas d'exemples avec les formes renforcées :
 - ?? nous aussi sommes partis
 - ?? vous-mêmes aviez été prévenus
- On ne dispose pas d'attestations dans lesquelles le sujet serait séparé du verbe par une insertion :
 - ?? nous cette année-là avons passé l'hiver chez eux
 - ?? vous à cette époque étiez malade
- Des emplois proches de ceux cités auparavant se rencontrent lorsque le sujet clitique est réalisé. La séquence qui comporte le pronom disjoint est alors associée :
 - alors **nous** comme un petit troupeau docile **nous** rentrions (CA)
 - nous autres nous** appartenions tellement à une classe ouvrière (CA)

Enfin, la présentation des pronoms est très sensible à l'approche utilisée. Les grammaires mettent en avant le comportement atypique de *moi* et *toi* qui ne peuvent être directement sujet alors que les formes disjointes *nous* et *vous* pourraient l'être. Les données de corpus font disparaître cette coupure : aucun pronom disjoint de première ou deuxième personne (*moi*, *toi*, *nous*, *vous*) ne se réalise directement en position sujet. Sur la base de ce critère, seules les formes de troisième personne ont un comportement original.

Pour les autres contextes, les résultats obtenus montrent des tendances comparables : on rencontre les formes disjointes, de façon majoritaire, dans l'environnement *c'est ... qui*. Toutefois, le nombre d'occurrences rencontrées dans les deux banques de données (CA et CRFP) soulève un problème : on aurait pu s'attendre à deux ou trois fois moins d'exemples dans CRFP. Le corpus de productions politiques contient 7 cas d'extraction avec *nous*, ce qui le place entre les deux banques de données si l'on rapporte ce résultat au nombre de mots total.

On peut pondérer cet écart en tenant compte du nombre de formes sujets *nous* (pronoms disjoints et conjoints) dans chaque cas. Dans ce cas, CA comprend environ deux fois moins de sujets *nous* disjoints que CRFP.

	<i>nous</i> sujets	<i>c'est nous</i> qui	%
CA	868	11	1,2
CRFP	523	11	2,1

Il se confirme bien que le corpus CRFP présente des caractéristiques différentes de celles de CA : la langue formelle, peu représentée dans le corpus initial (CA), a été augmentée (dans CRFP) ce qui retentit sur la fréquence de certains tours. Mais le nombre réduit d'exemples atténue grandement la portée de cette observation.

Un examen plus précis de la tournure *c'est ... qui* permet de retrouver la distinction déjà établie pour les formes *moi* et *toi* :

- *c'est nous qui* correspond à un phénomène de clivage qui permet de singulariser, d'opposer *nous* à une autre personne :

il ne parle jamais de fessée *c'est nous* qui parlons de la fessée chez
Rousseau (CA)
c'est nous qui organisons tout en fait (CRFP)

De nombreux exemples avec modalité négative ont pu être relevés (ce qui accentue le voisinage avec *moi*) :

c'est pas nous qui décidons en fait *c'est c'est le c'est le* metteur en
scène (CA)
c'est pas nous qui l'enseignons on s'en sert comme soutien (CA)

- *c'est vous qui* se rencontre dans des questions et ne peut être considéré comme une extraction :

L1 oui *c'est vous* qui le faites L2 bé ça dépend (CRFP)
L1 *c'est vous* qui allez en prison L2 oui (CA)
L1 *c'est vous* qui avez fait les grillages L2 ah (CA)

Les exemples avec modalité négative sont absents de nos relevés.

4. Conclusion

Plusieurs aspects se dégagent de cette courte observation. La présentation sous forme d'une classe (i.e. la liste des pronoms disjointes) masque ou écrase les propriétés originales et spécifiques des éléments. À partir des corpus oraux, on peut envisager de remodeler la description sur plusieurs points et de présenter les quatre séries suivantes :

Paul CAPPEAU

- *moi*

Le pronom *moi* possède une distribution originale (par le nombre d'attestations ainsi que par les contextes dans lesquels il se réalise), il convient donc de le présenter à part ;

- *toi et vous*

Les pronoms *toi* et *vous* peuvent être rapprochés. On les trouve fréquemment dans des questions et leur distribution est assez semblable ;

- *eux et lui*

Les pronoms *eux* et *lui* sont deux formes aux propriétés voisines et doivent être séparées de leurs « correspondants » féminins qui sont beaucoup moins répandus ;

- *nous, vous, elle(s)*

Les pronoms *nous*, *vous*, *elle(s)* se caractérisent par leur rareté : dans les exemples attestés, ils répugnent à occuper directement la position sujet (en tant que forme disjointe) et ils sont peu attestés dans les autres contextes.

D'autre part, les deux principales banques de données CA et CRFP utilisées pour ce travail, bien que constituées sur des bases différentes permettent de disposer de la plupart des contextes pertinents. CRFP présente quelques lacunes pour les exemples peu fréquents. Il semble que l'hétérogénéité plus grande de CA (composé sur une période de temps bien plus grande et sur des bases plus disparates) permet justement de rencontrer des emplois singuliers. Reste que certaines des différences qui apparaissent dans ces deux banques de données devront être mieux cernées et bien identifiées pour apprécier l'état de la langue qu'elles reflètent.

Références

- Creissels, D. (1991). *Description des langues négro-africaines et théorie syntaxique*. Grenoble : Ellug.
- Creissels, D. (1995). *Éléments de syntaxe générale*. Paris : PUF.
- Blanche-Benveniste, C., Bilger, M., Rouget, C., & Van den Eynde, K. (1990). *Le français parlé. Études grammaticales*. Paris : Éditions du CNRS.

Les formes disjointes des pronoms sujets

- Blanche-Benveniste, C. (2000). Corpus de français parlé, In M. Bilger, (Ed.), *Corpus, méthodologie et applications linguistiques* (pp. 15-25). Paris : Champion et les Presses Universitaires de Perpignan.
- Blasco-Dulbecco, M. (1999). *Les dislocations en français contemporain. Étude syntaxique*. Paris : Honoré Champion.
- Sabio, F. (2002). L'opposition de modalité en français parlé : étude macro-syntaxique. *Recherches sur le français parlé*, 17, 55-77.
- Skårup, P. (1975). Les Premières zones de la proposition en Ancien Français. Essai de syntaxe de position. *Études Romanes de l'Université de Copenhague, Revue Romane* 6.

Quelques éclairages sur le sujet de type *moi je* à l'oral

Mylène BLASCO-DULBECCO
Université Clermont-Ferrand II

Résumé

Cet article s'intéresse au statut du pronom disjoint *moi* dans des réalisations de sujet de type *moi ... je* et *je ... moi*. Il s'appuie sur un corpus important d'exemples attestés pour mettre à jour, en français parlé, des distributions syntaxiques régulières et imprévisibles du pronom *moi*. Nous essayons de montrer que les différents fonctionnements observés pour ce pronom obligent à nuancer les fonctions purement énonciatives que lui attribue souvent la littérature linguistique et à s'interroger sur sa fonction syntaxique dans la construction verbale.

1. Introduction

Ce travail s'inscrit dans une recherche approfondie sur le fonctionnement des dislocations en français parlé et représente une seconde étape dans une réflexion portant sur la cooccurrence *pronom disjoint* + *pronom conjoint* en français. Dans cet article, nous nous pencherons sur les dislocations du sujet du type *moi ... je* et *je ... moi* dans lesquelles un pronom disjoint est mis en emphase avec le pronom conjoint sujet correspondant. À partir de l'exploitation du « Corpus de référence de français parlé » nous nous demanderons quelles sont les occurrences à l'oral de ce type de sujet. Quelles distributions ressortent de l'examen de ces données attestées dans l'usage ? Quelles sont les particularités syntaxiques et sémantiques de ce type de dislocations ? Ces questions auront pour acquis que, dans les énoncés à dislocation, les pronoms disjoints ne se laissent pas décrire dans les mêmes termes que les éléments

Mylène BLASCO-DULBECCO

lexicaux et que le pronom *moi* présente des particularités syntaxiques et énonciatives propres en regard des autres pronoms.

2. Sur le traitement des pronoms disjoints

Dans la littérature grammaticale, l'emploi de ces pronoms est dit réservé aux constructions « détachées » de type dislocation ou extraction. En effet, en position détachée les pronoms disjoints peuvent être suivis d'une relative ou d'une apposition : *Toi qui sais tout ... – Lui, toujours si discret ...* ce qui renforce la capacité expressive qui leur est généralement attribuée et convient parfaitement aux emplois contrastifs. Le traitement de ces pronoms ne soulève pas de problème particulier ; leur description englobe tous les pronoms disjoints avec peut-être une réserve pour *lui, elles(s)* et *eux* qui ne se prêtent pas systématiquement à la dislocation.

« Les formes disjointes ont un comportement syntaxique analogue à celui d'un groupe nominal séparé du verbe (par une préposition, une pause, etc.). Aussi apparaissent-elles dans les constructions caractéristiques suivantes : [...] en position détachée, par dislocation ou extraction [...]. Comme sujets détachés, *lui, elle(s)* et *eux* ne sont pas obligatoirement repris par le pronom sujet conjoint correspondant, surtout en emploi contrastif. » (*Grammaire méthodique du français*, 1994 : 201)¹

¹ Pour Grevisse (1986), les formes disjointes de pronoms s'emploient comme sujet dans certains cas et le français parlé userait de manière systématique de la dislocation du sujet :

« Dans la langue ordinaire, le pronom est redondant par rapport au sujet se trouvant à sa place ordinaire : *Moi, je le ferai.* » (*Le Bon Usage*, parag. 636).

Aux XVII^e et XVIII^e siècles, d'un point de vue général, la redondance du sujet est affaire de rhétorique dès qu'elle concerne un pronom conjoint et un pronom disjoint ; la redondance entre *moi* et *je* semble légitime si elle provoque des figures d'insistance : Meigret (1550 : 49) évoque la possibilité de cumuler *je* et *moi* ou *moi-même* en fin de phrase :

« Ceci n'ajoute rien de plus à la signification du discours, mais l'affection est toutefois plus grande par cette répétition que par une simple locution ».

« À tout autre mode que l'impératif, « MOI » ne peut pas s'employer seul, il se construit avec « JE » lorsqu'il est le sujet de la proposition, lorsqu'il est l'objet ou le terme du verbe, il se construit avec « ME ».

À notre connaissance, le fonctionnement spécifique des pronoms disjoints n'a été jusque-là ni relevé ni retenu dans les analyses modernes de la dislocation, même si le *moi* du locuteur est presque toujours mentionné dans les analyses conversationnelles.

L'on pourrait s'en tenir à une présentation de ce que devraient être certains emplois des pronoms disjoints en français. Mais l'examen général sur un vaste corpus de données orales attestées met à jour des faits imprévisibles a priori sur leur fréquence et sur leur distribution.

D'une part, le pronom *moi* semble acquérir en position sujet disloqué divers statuts et fonctionnements autant syntaxiques que pragmatiques. D'autre part, les dislocations du type *moi je* et *je moi* se prêtent dans certains cas à des analyses différentes.

3. Les particularités des pronoms disjoints dans la dislocation

Dans un travail antérieur (Blasco-Dulbecco, 1999) nous montrions que la construction disloquée demande à être décrite pour elle-même et en elle-même, et non comme une dérivation ou comme une transformation. Cette exigence émanait de la description d'un corpus important d'exemples attestés.² L'utilisation *quasi* exclusive de données attestées permettait de découvrir des fréquences qui mettaient à leur juste place le phénomène de dislocation lui-même et la répartition de ses différentes formes : forte majorité de sujets grammaticaux disloqués, variété plus grande des positions syntaxiques à l'écrit, très forte majorité des syntagmes non prépositionnels en position initiale à l'oral.

Une des clés de l'étude des dislocations dans cette approche résolument descriptive de la langue était, de manière totalement imprévisible, la pertinence des catégories grammaticales dans l'analyse des dislocations et ce à plusieurs titres.

- En termes de catégories grammaticales :

Vous concevez que lorsqu'on joint à propos ces deux noms de la première personne, la phrase peut en avoir que plus d'énergie » (Condillac, 1789 : 197).

Il semble donc que, très tôt, les grammairiens ou observateurs de la langue classique accordent un traitement particulier au pronom de première personne *moi*.

² Le corpus du G.A.R.S. : corpus de productions spontanées ou préparées particulièrement varié dans les situations enregistrées d'un million de mots.

Si dans la dislocation avant le verbe et dans la dislocation après le verbe la répartition entre éléments lexicaux et pronoms disjoints est sensiblement la même à l'oral (alors que l'écrit contient plus de deux fois plus de dislocations de lexique !), la dislocation à gauche du pronom *moi* est dominante sur les autres pronoms³.

- Régularité dans l'emploi des prépositions :

Elle était dans l'étude des plus surprenantes : « le français oral utilise très rarement un énoncé du type *aux femmes je leur fais pas confiance*, mais de manière presque exclusive *les femmes je leur fais pas confiance* ». Ce critère de la préposition était retenu pour décrire le statut syntaxique de la dislocation.⁴ Mais curieusement, les pronoms disjoints dérogeaient à cette régularité distributionnelle des prépositions :

moi ils me font pas peur
à moi elle me fait peur la bête
ils me font peur moi
ils me font pas peur à moi

³ Seul le corpus garantit la justesse de la perception que l'on a de certains phénomènes dans la langue. Ainsi nos résultats diffèrent considérablement de ceux présentés par Le Querler (1999 : 267) qui dit que :

« La catégorie à laquelle appartient le plus souvent l'élément disloqué est celle du syntagme nominal [...]. Enfin, l'élément disloqué est assez souvent précédé d'un terme introducteur comme *question ...* ou *quant à ...* »

L'auteur s'appuie sur l'examen d'environ 2000 énoncés tirés de journaux, de romans, de bandes dessinées et de quelques enregistrements radiophoniques.

⁴ Déjà Damourette et Pichon (1911-1940 : § 1008) distinguaient les *semi-dislocatures* de type *Au roi, je lui ai parlé* et les *dislocatures vraies* de type *Le roi, je lui ai parlé*.

Notre étude des dislocations en français parlé nous a conduit à distinguer, selon divers critères, deux types de constructions : dans le premier type, le syntagme disloqué entre dans la recton du verbe - auquel cas il est un prolongement, même anticipé, de la fonction syntaxique représentée par le pronom clitique avec lequel il est coréférent - ce que nous avons appelé un « double marquage ». Dans le second type, le syntagme disloqué n'est pas régi par le verbe : il est alors soit adjoint à l'ensemble de la construction verbale, et n'entretient pas de lien référentiel avec le pronom clitique, soit il est adjoint au clitique seulement parce qu'il entretient avec lui un type de relation référentielle.

Quelques éclairages sur le sujet de type *moi je* à l'oral

- Mobilité et distribution :

Dans l'ensemble des contextes syntaxiques étudiés, le pronom disjoint se montrait très mobile dans et autour de la construction verbale et sa distribution n'obéissait pas à des contraintes de proximité avec le pronom clitique correspondant.

c'est dans ce but-là **moi** que **je** fais partie de l'association
c'est dans ce but-là que **moi je** fais partie de l'association
moi c'est dans ce but-là que **je** fais partie de l'association
c'est dans ce but-là que **je** fais partie **moi** de l'association
c'est dans ce but-là que **je** fais partie de l'association **moi**

- Difficulté d'insertion :

Enfin, nous avons identifié le *moi je* de l'énonciateur courant avec les verbes comme *dire, penser, croire* ; une sorte de morphème à deux éléments qui admettrait difficilement l'insertion d'autres éléments. D'ailleurs, certains locuteurs pourraient l'employer de façon presque systématique.

Pour toutes ces raisons, nous sommes en mesure de penser que, dans les énoncés à dislocation, les pronoms disjoints ont un fonctionnement différent de celui des éléments lexicaux et qu'ils méritent une analyse autre. Nous faisons l'hypothèse que lorsqu'ils ont une fonction syntaxique auprès du verbe recteur alors ils traduisent, d'un point de vue sémantique, une désignation, un contraste ou une insistance décelable par l'étude du contexte. En revanche, lorsqu'ils n'ont pas de fonction syntaxique sujet auprès du verbe alors ils donnent l'effet d'un point de vue ou d'un cadre de réflexion dans lequel vient s'inscrire la construction verbale.

C'est dans cette perspective que nous nous sommes demandé ce que pouvait apporter comme complément, précision ou même révision une nouvelle base de données dans l'étude du pronom disjoint *moi*.

Mylène BLASCO-DULBECCO

4. La distribution du pronom *moi* à l'oral

4.1. *Moi* plutôt qu'un autre

Seule une recherche sur une base de données conséquente, statistiquement interrogeable⁵ permettait de découvrir que *moi* est, de loin, le pronom disloqué le plus représenté dans la dislocation mais que, contrairement à toute attente, le nombre d'exemples – et plus particulièrement les occurrences *je moi* – est décevant en regard des préjugés que l'on porte habituellement sur le rapport entre la langue orale et les dislocations⁶.

Le *Corpus de référence de français parlé* fait ressortir les fréquences suivantes :

moi je	475	occurrences
je moi	66	occurrences
toi tu	43	occurrences
tu toi	15	occurrences
eux ils	74	occurrences
ils eux	22	occurrences

Moi, en position sujet, a donc un statut bien particulier et ces résultats sont confirmés par les occurrences rencontrées pour *moi* disloqué en position complément :

moi ... me	70	occurrences
me ... (à) moi	22	occurrences

⁵ Le corpus a été interrogé à l'aide du concordancier Contextes élaboré par Jean Véronis (<http://www.up.univ-mrs.fr/veronis/>).

⁶ Pour que nos résultats soient vraiment pertinents il faudrait pouvoir comparer les occurrences de *moi je* avec les occurrences de *je*.

Nous ne citons pas les occurrences de type *elle(s) elle(s)*, *nous nous*, *vous vous*. La recherche par concordancier ne s'avère pas fiable car il faut opérer une phase de nettoyage importante pour sortir, des exemples retenus, les cas de répétition, hésitation etc. De la même façon, nous ne pouvons pas nous fier aux résultats apparus pour *lui il* (81 occurrences) et *lui* (39 occurrences) à cause de certains emplois clitique de *lui*.

4.2. Les structures récurrentes avec *moi je*

Certaines convergences conduisent à proposer un classement des contextes dans lesquels se rencontre la dislocation *moi je*. Deux grands types ressortent :

4.2.1. *moi* et *je* semblent liés comme s'ils constituaient un seul et même morphème

On rencontre ce morphème sujet dans deux cas de figure.

- Dans un premier cas, *moi* entre dans un emploi nettement contrastif : deux constructions parallèles se succèdent dans lesquelles le pronom disjoint est mis en opposition avec un autre élément – majoritairement un autre pronom disjoint. Dans ces emplois, soit les verbes employés dans chacune des séquences diffèrent aussi, soit il y a reprise du lexème verbal avec un effet de parallélisme. Alors les séquences connaissent une variation qui peut affecter par exemple la modalité (positive / négative).

Le contraste apparaît aussi sur les éléments du groupe verbal :

- (1) ça va faire payant puisque **lui** a arrêté et **moi je** continue
- (2) et puis **eux ils** se sont arrêtés de jouer **moi je** continue à moi je continue eux toutes formes de sculpture
- (3) c'est **elle** qui s'occupe de la blanchisserie **moi je** m'en vais à deux heures et demie
- (4) **chacun** a sa façon de voir et chacun a sa méthode **moi je** pars du principe que euh quand on aborde

Séquences parallèles avec maintien du verbe :

- (5) le cap de la quarantaine **il** fallait l'accepter **moi je** l'ai accepté dans mon entourage
- (6) on a quand même un laps de de repos **moi je** suis du matin **ma collègue** est du soir euh elle a son samedi moi j'ai mon lundi on a notre re- nos repos
- (7) comprenez-vous **vous** avez l'air de rire mais **moi je** ne ris pas
- (8) mais bon **tu** connais le corps enseignant **moi je** le connais je l'ai pratiqué
- (9) **il y en a très peu** qui parlent le français et **moi je** parle pas un seul mot d'anglais

Le corpus ne donne que deux exemples dans lesquels s'insère entre *moi* et *je* un adverbe de type *aussi, personnellement, en tout cas* qui renforcerait l'effet d'insistance :

Mylène BLASCO-DULBECCO

- (10) comme ça en groupe on s'est orienté enfin **moi** personnellement
je me suis orienté vers les
- (11) je lui reprochais elle regardait les touches le problème **moi** aussi
je regarde les touches

On pourrait penser que, sémantiquement, nous touchons là à une sorte de stratégie dans l'échange conversationnel⁷ qui mériterait d'être étudiée dans le détail mais avec une manipulation très prudente des notions habituellement retenues de thème et de rhème pour traiter cette question.

- Dans un second cas, *moi je* est en cooccurrence avec des verbes de discours comme *dire, penser, trouver, voir*. Dans le corpus, aucun adverbe ne s'insère entre les deux éléments⁸:

- (12) on a une vie très difficile en ce moment enfin **moi je** vois avec les gens qu'on discute
- (13) la vie elle a ses imprévus elle a ses mais **moi je** trouve que j'ai j'ai eu beaucoup de chance
- (14) ça se passe bien enfin **moi je** trouve que ça se passe bien
- (15) [...] c'est incroyable **moi je** trouve que la route on est en république
- (16) il y avait les boudins qui étaient pleins de sciure alors **moi je** lui dis vous allez pas prendre ça
- (17) tout seul moi ça m'a toujours fait rire **moi je** sais pas j'aime bien ce côté dire un gars
- (18) il vaut mieux attendre de voir enfin **moi je** pense qu'il vaut mieux attendre de voir
- (19) [...] aussi oui euh **moi je** pense que c'est c'est toute une vie

Même si finalement ces exemples sont statistiquement peu révélateurs, il faut préciser qu'ils sont particulièrement rares avec une dislocation à droite du sujet *je ... moi*.

⁷ Le Querlers (1999) s'appuie sur ce type de corrélations pour repenser le caractère thématissant ou non des dislocations :

« Le deuxième cas de dislocation pour lequel il semble qu'il n'y ait pas de thématisation est celui d'énoncés où l'élément disloqué est représenté par un pronom, mais où il est souligné par aussi ou même. [...] Il semble aussi qu'à la lecture à voix haute de ces énoncés, l'intonation marquant la fin de l'élément disloqué soit la « conclusive majeure » que Mario Rossi désigne comme une marque obligatoire du rhème. » (Le Querler, 1999 : 271).

⁸ Contrairement à ce que nous pensions, les exemples sont peu nombreux : seulement 3% des dislocations en *moi je* sont de ce type. Il serait important de vérifier la proportion de ce type de construction verbale avec et sans dislocation du sujet comme *je trouve/moi je trouve*.

Le *moi je* dit de l'énonciateur est évoqué par l'ensemble des approches énonciatives (Morel & Danon-Boileau, 1999). Malheureusement, toutes spéculent, à notre avis de manière très contestable, sur ladite « volonté du locuteur » dans sa façon de participer au discours.⁹

S'il est raisonnablement difficile de contester le trait « énonciateur » de *moi*, nous pensons cependant qu'il convient de faire des sous-classements et d'isoler par exemple certains fonctionnements de *moi je*. Nous verrons plus loin que la dislocation du pronom *moi* dans certains contextes s'accompagne, de façon assez régulière, de l'antéposition d'autres éléments par rapport au verbe ; ce qui tend à lui donner un statut syntaxique particulier.

De manière générale, dans la littérature linguistique, c'est en fait le pronom *moi* qui est traité comme marque de l'énonciateur seul ou en reprise avec un autre élément comme *c'est*. D'ailleurs, Morel & Danon-Boileau (1999) le détachent catégoriquement de *je* dans leur analyse en *point de vue et modus dissocié*.

Mais nos données témoignent tout de même d'un fonctionnement très particulier du *moi je* et ce caractère soudé que nous lui attribuons se laisse illustrer par un comportement insolite que nous avons pu observer : celui des amorces.

- (20) ben écoutez **moi je** vous savez *je* notre association
- (21) euh ben disons que **moi je je** voudrais pas me prononcer là
- (22) y passer quelques temps mais enfin moi je suis assez casanier **moi je**
j'aime bien ma région
- (23) moi je suis désolé **moi je je** vois à la télévision

⁹ « Des éléments pronominaux comme *moi* indiquent la volonté du locuteur de participer au discours (« prise de parole ») sans en être forcément aussi le thème. » (Stark, 1999).

« Le marquage de l'énonciateur est une technique fréquente d'introduction des topics. De cette façon, ce qui est marqué est d'abord la prise de parole, l'affirmation de l'énonciateur, l'opération qui porte *je* à l'existence dans l'énonciation. En outre, le topic est ainsi légitimé par son inscription dans la sphère de l'énonciateur, ce qui le rend accessible. Le marquage de *moi* en X renvoie implicitement à un contraste possible avec l'interlocuteur, qui par ailleurs répond lui aussi par une thématization symétrique. La thématization, tout en soulignant un topic, marque en même temps la position singulière, différenciée, polémique, du locuteur par rapport à ce topic. » (Berthoud, 1999).

Mylène BLASCO-DULBECCO

Dans la plupart des cas, l'amorce de ce type de sujet se fait avec l'ensemble *moi je* et non pas sur *moi* à tel point que même une apposition sur l'ensemble *moi je* est possible :

- (24) là où était monsieur Bayrou le le ministre de l'Éducation nationale il y a quelques années comme prof comme jeune prof puis de l'École Normale **moi je** - *fils de prolétaire je* me disais /ça, c'est/ instituteur c'est euh c'est le pied

Dans cet exemple, le locuteur ne complète pas son amorce en reprenant le sujet depuis *moi* alors que l'on sait que, de façon générale, la reprise a lieu depuis le début du syntagme¹⁰. On peut voir là la confirmation que *moi et je* forment un groupe proportionnel à *je*.

4.2.2 *moi et je* sont séparés par un segment.

Dans ce deuxième fonctionnement de la dislocation, *moi* ne serait plus partie intégrante du morphème sujet : il aurait un statut d'adjectif à l'ensemble de la construction verbale et donnerait l'effet d'un point de vue ou d'un cadre de réflexion dans lequel vient s'inscrire la construction verbale. Au-delà des effets sémantiques, il y a des tendances surprenantes dans les contextes syntaxiques.

Divers types d'éléments peuvent s'insérer entre *moi et je* :

- *moi* est suivi d'un « complément » temporel et très rarement d'un complément locatif :

- (25) c'était pas l'époque encore du fax hein **moi** pendant les vacances
je gagnais ma vie
(26) mais bon euh enfin **moi** donc depuis depuis sa naissance **je** le
connais
(27) ils savent pas forcément comment y faire donc **moi** maintenant
je commence à avoir des compétences
(28) c'est ça hein ma formation de commerce bon **moi** à l'origine **je**
suis parti /dans, avec/ un C.A.P.

On dirait que l'antéposition du complément temporel est facilitée / sollicitée par la présence en tête de ce *moi* qui n'a pas de statut syntaxique auprès du verbe recteur dont il est éloigné. En effet, les exemples suivants que nous avons reconstruits nous paraissent moins naturels dès lors qu'il y a antéposition des compléments :

- (25') c'était pas l'époque encore du fax hein *pendant les vacances je*
gagnais ma vie

¹⁰ cf. Pallaud (2002 : 94).

Quelques éclairages sur le sujet de type *moi je* à l'oral

- (26') mais bon euh enfin donc depuis *depuis sa naissance je* le connais
- (27') ils savent pas forcément comment y faire donc *maintenant je* commence à avoir des compétences
- (28') c'est ça hein ma formation de commerce bon à *l'origine je* suis parti /dans, avec/ un C.A.P.

- En revanche, dans d'autres exemples, les éléments qui suivent *moi* n'ont pas de fonction syntaxique identifiable

- (29) remarquez qu'à Paris c'est pareil *moi ça je* l'ai vu à Paris un un homme même s'il
- (30) [...] pour fabriquer des merguez j'ai dit non *moi vos produits je* n'en veux pas
- (31) *moi* de ce côté-là voyez-vous *je* peux dire

Dans ces emplois *moi* serait à la tête d'un cadre énonciatif; il marquerait le début d'une zone hors construction verbale dans laquelle viendraient s'emboîter d'autres éléments non construits par le verbe : des éléments adjoints à l'ensemble de la construction verbale. Ainsi, tout type d'associé – y compris d'autres éléments disloqués à la construction verbale – peut s'y insérer :

D'ailleurs, dans de nombreux cas, l'éloignement entre les deux morphèmes résulte de la présence d'une construction verbale régie dans laquelle *je* occupe la fonction de sujet alors que *moi* est extérieur à la construction régie. De manière systématique, il s'agit là encore d'une temporelle :

- (32) *moi* depuis que *je* suis arrêté on a pris
- (33) *moi* quand *je* prends zéro trois que *je* multiplie
- (34) des choses comme ça et en fait quand on a enfin *moi* quand *je* suis arrivé là-haut ça m'a vraiment
- (35) tandis qu'ici vous n'avez rien et *moi* la mairie quand *je* pars si *je* suis bien
- (36) bon *moi* depuis que *je* suis là j'ai vécu euh au niveau

Ces constructions illustrent que *moi* est à l'extérieur de la zone verbale. Dans cette position, *moi* peut être associé à plusieurs places de construction comme le montre l'exemple (34) dans lequel *moi* est associé une première fois avec le sujet *je* et une deuxième fois avec le complément *me*. D'un point de vue syntaxique, dans Blasco-Dulbecco (1999 : 142-164), nous montrions longuement que lorsque l'élément disloqué dépassait la borne morphologique de la construction verbale, il sortait de la rection du verbe.

Mylène BLASCO-DULBECCO

Nous pouvons ajouter que, curieusement, *moi je* n'intervient jamais dans la construction verbale régie :

- (33') ? quand **moi je** prends zéro trois que je multiplie
- (36') ? bon depuis que **moi je** suis là j'ai vécu euh au niveau
- (35') ? tandis qu'ici vous n'avez rien et la mairie quand **moi je** pars si je suis bien

Pour la propriété sémantique des segments insérés entre *moi* et *je*, on pourrait retrouver ici un type de structure décrit par Jeanjean (1984). L'auteur a identifié une forme longue de type *Toi quand P : Je pense à toi quand tu étais petit*, dans laquelle le paradigme *quand P* exprime un « certain aspect » de *toi*.

Mais nous avons rencontré aussi d'autres types de constructions verbales entre *moi* et *je* dans lesquelles seul *je* est sujet ; il s'agit de constructions verbales clivées ou impersonnelles :

- (37) on marche beaucoup à l'émotion **moi** c'est pour ça que **je** marche beaucoup comme ça
- (38) alors **moi** ce que **je** proposerais au comité de quartier
- (39) et puis même **moi** ce que **je** fais c'est dans les trucs
- (40) en fait on me demande pas mon avis **moi** ce que **je** pense ça me regarde personnellement
- (41) les idées du parti politique **moi** ce que **je** pense on s'en fout
- (42) augmenter le coût du salaire ça veut dire que **moi** il va falloir que **je** répercute ce coût
- (43) il se dit ben moi ça m'intéresse pas **moi** il faut que **je** ramène il faut que je ramène

Au-delà des interprétations sémantiques que l'on peut attribuer à *moi* dans ce type d'énoncé, il nous paraît plus judicieux de pointer les spécificités des contextes syntaxiques dans lesquels il intervient pour identifier premièrement un *moi* autonome, à traiter séparément de *je*, et deuxièmement lui accorder un statut syntaxique différent des *moi je* observés précédemment.

D'ailleurs, certaines études, que ce soit dans une perspective « évolutionniste » ou encore informationnelle (Hopper et Traugott, 1992 ou Nølke, 1997), attribuent les constructions disloquées à un développement proprement syntaxique. Ainsi, le noyau de la phrase se serait fortement figé en français moderne sous la forme d'accumulation d'éléments clitiques – du type *je ne lui parle pas* –, la précision de la référence des arguments du verbe reviendrait aux éléments disloqués. Cette piste serait à discuter et à approfondir ; elle soutiendrait notre approche syntaxique du deuxième *moi* identifié comme borne d'une liste d'éléments possibles, soit construits

avant le verbe, soit dans une zone préverbale donc hors construction verbale.

4.3. Les fonctionnements de *je moi*

Il est d'usage de différencier la dislocation avant le verbe de la dislocation après le verbe en disant que la dislocation à gauche crée une attente à laquelle répond l'énoncé alors que la dislocation à droite ménage une surprise après coup, une fois l'énoncé entendu. Nos données de français parlé ont déjà montré qu'il convient de distinguer deux effets possibles de la dislocation après le verbe pour la catégorie des pronoms disjoints (Blasco, 1998).

Dans un premier cas, la dislocation à droite est un phénomène de double marquage, une sorte de prolongement syntaxique et sémantique du clitique.

Dans un second cas, plus qu'une réalisation semi-lexicale du pronom clitique, le pronom disjoint signale, d'un point de vue sémantique, une opposition avec un autre élément du contexte comme pourrait le faire une modalité négative ou restrictive :

je l'ai rencontré **lui** l'année dernière
je l'ai rencontré **lui** (et pas elle ; que lui ; lui seul) l'année dernière

Dans certains emplois, *moi* apporte (dans l'optique de cette deuxième interprétation) un développement pertinent pour la saisie référentielle du clitique. C'est comme un spécificateur sémantique proche en quelque sorte du *moi je* sujet, traité dans cet article (3.2). D'ailleurs, comme dans la dislocation de type *moi je*, cette propriété de spécification peut ou non être clairement identifiable par la présence d'indicateurs morpho-syntaxiques de type *même*, *en personne*, *seul* ou encore par la présence d'une modalité restrictive ou comparative. En l'absence de ces indicateurs, le contexte peut lui aussi éclairer l'interprétation que l'on doit donner au pronom : quelquefois, la présence d'une autre dislocation évoque cet effet de contraste ; d'autres fois, le pronom tonique est suivi d'une apposition ou d'une relative.

Ce qui différencie cependant le *moi je* de type contrastif du *je moi*, c'est l'absence d'exemples dans lesquels le contraste est mentionné dans le contexte syntaxique¹¹ :

¹¹ Pour illustrer ces propos, nous avons manipulé les exemples (5), (6) et (7) précédemment cités.

Mylène BLASCO-DULBECCO

- (44) le cap de la quarantaine il fallait l'accepter **je** l'ai accepté **moi** dans mon entourage
- (45) on a quand même un laps de de repos **je** suis du matin **moi ma collègue** est du soir euh elle a son samedi moi j'ai mon lundi on a notre re- nos repos
- (46) comprenez-vous **vous** avez l'air de rire mais **je** ne ris pas **moi**

Pour tenter un parallèle avec les résultats que nous proposons dans la partie précédente, nous préciserons que nous avons rencontré seulement 2 exemples de *moi* à valeur d'insistance en dislocation à droite :

- (47) à subir des peines d'emprisonnement eh bien **je** milite **moi aussi** pour que ces personnes
- (48) donc je vais essayer **moi aussi** sans le micro

seulement 3 exemples de *je moi* en cooccurrence avec des verbes de discours :

- (49) **je** trouve que ça marche mieux **moi**
- (50) alors moi après donc après l'I.U.T. comme **je** vous ai dit **moi**
- (51) c'est et **je** trouve que c'est contagieux **moi**

3 exemples de *moi* « prolongé » avec ce caractère obligatoire du *moi* afin que le sujet puisse supporter l'ajout d'un segment déterminatif¹² :

- (52) je réagirais pour ce qui concerne la France et **je** ne peux **moi que comme élu responsable**
- (53) alors qu'il faisait des horaires que **je** fais **moi patron** c'est-à-dire trente cinq heures

Donc les distributions présentées pour la dislocation *moi je* avant le verbe ne sont pas celles à retenir pour la dislocation *je moi* après le verbe.

En revanche, 15 exemples (soit près de 23% des emplois de ce type de dislocation à droite !) sont réservés à la tournure énonciative *je sais pas moi*, sorte « d'approximatif »¹³:

- (54) mais non **je sais pas moi** la campagne ça ça me plaît vraiment
- (55) ils arrivaient chez toi et avec quelque chose **je sais pas moi** un morceau un pain
- (56) quand on a été jouer là-bas bé ils étaient **je sais pas moi** plus de trois cents
- (57) c'est-à-dire que on euh on va demander aux gens **je sais pas moi** s'ils sont gais ou s'ils sont tristes
- (58) c'est-à-dire que nous pensons par exemple euh **je sais pas moi** la faim dans le monde par rapport

¹² cf. Blasco-Dulbecco & Caddéo (2000, 2002).

¹³ cf. Roubaud (2000).

Quelques éclairages sur le sujet de type *moi je* à l'oral

- (59) quand tu as la flemme ben enfin tu **je sais pas moi** c'est c'est comme pour tout hein
- (60) celui qui a besoin en fait d'être soutenu euh **je sais pas moi** dans un truc qu'on pourrait appeler
- (61) donc tu prends la dépêche A.F.P. sur **je sais pas moi** la guerre en au Liberia
- (62) on tourne en rond euh bon autrement euh **je sais pas moi** il y a il y a il y a

Tous les autres cas identifiés correspondent à des doubles-marquages du type :

- (63) et d'abord **je** suis déjà venue **moi** à la réunion
- (64) **je** suis née ici **moi** à Pompé
- (65) ce que je fais surtout depuis vingt-cinq ans **je** me soigne ici **moi** et ça ça **marche pas**
- (66) il faut demander simplement aux gens d'abord vous êtes qui **je** vous connais pas **moi** hein qui vous êtes vous
- (67) mais **je** ne vois pas **moi** d'autres façons de faire
- (68) **je** connais pas **moi** Rome
- (69) mais je sais pas du tout ce que **je** veux faire **moi** c'est clair que je veux pas travailler
- (70) ah ben **je** suis passé **moi** j'ai commencé j'ai commencé
- (71) pour se distraire **moi** j'emportais mon tricot **je** tricotais **moi**
- (72) les gens ils viennent même je fais pas **je** fais pas de vin rouge **moi** mais le Beaujolais
- (73) il y a des Italiens il y a de tout alors **je** suis un bâtard **moi** complètement mélangé
- (74) **je** pouvais pas m'empêcher de rire **moi**

Dans ces emplois, *moi* disloqué après le verbe est un prolongement syntaxique et sémantique de *je* sujet. Aucun contexte syntaxique, tels que ceux qui ont été dégagés pour les dislocations avant le verbe, ne conduit à envisager un autre statut.

5. Conclusion

Une fois de plus, la prise en compte de données orales attestées fait ressortir l'importance des catégories grammaticales dans le fonctionnement de la dislocation et ce tant d'un point de vue syntaxique que d'un point de vue sémantique.

Cette étude, que nous ne prétendons pas aboutie, permet tout de même d'affirmer qu'il ne convient pas d'isoler dans la description le *moi* pronom disjunctif et le *je* pronom conjoint correspondant. Nous avons

Mylène BLASCO-DULBECCO

identifié des comportements syntaxiques et sémantiques très particuliers, d'une part pour le morphème *moi je* sujet, et d'autre part pour le morphème *moi*. Travailler sur un corpus oral important permet ainsi de modifier la façon dont il faut aborder cette catégorie grammaticale dans la description et de faire ressortir les distributions telles qu'elles sont dans les usages attestés.

- Nous retiendrons pour la dislocation avant le verbe :
 - la présence d'un morphème sujet à deux éléments fortement liés *moi je*. Dans ce cas, *moi* ne se contente pas de marquer dans le discours des effets expressifs, il entre dans des constructions syntaxiques contrastives ou alors il marque la prise de parole avec des verbes spécifiques ;
 - quand *moi* permet de poser le cadre de l'énoncé, il est indépendant de *je* dont il est d'ailleurs éloigné ; il s'accompagne dans la majorité des cas de spécifications de type quand P ; il peut aussi représenter la borne d'une chaîne d'éléments associés avant la construction verbale.
- Nous retiendrons pour la dislocation après le verbe que :
 - c'est un *moi* en prolongement syntaxique et sémantique du pronom clitique qui domine dans la distribution ; dans cet emploi la présence du pronom disjoint peut être indispensable pour supporter un prolongement sémantique de type déterminatif ;
 - vient ensuite un emploi de marqueur approximatif fortement figé sous la forme *je sais pas moi*.

Ces tendances permettent de dire que *moi* est un pronom à part dans l'ensemble de la catégorie des pronoms disjoints, de nuancer les fonctions purement énonciatives qui lui sont toujours attribuées et de mettre à jour des comportements syntaxiques inattendus.

Références

- Bally, C. (1932). *Linguistique générale et linguistique française*. Berne : A. Francke S.A. Édition de 1944.
- Blanche-Benveniste, C., Bilger, M., Rouget, C., & Van den Eynde, K. (1990). *Le français parlé. Études grammaticales*. Paris : Éditions du CNRS.
- Blasco-Dulbecco, M., & Caddéo, S. (2002). Détachement et linéarité. *Recherches sur le français parlé*, 17, 41-54.

Quelques éclairages sur le sujet de type *moi je* à l'oral

- Blasco-Dulbecco, M., & Caddéo, S. (2001). Apposition et dislocation : la séquence pronom+lexique+clitique. *Recherches sur le français parlé*, 16, 125-150.
- Blasco-Dulbecco, M. (1999). *Les dislocations en français contemporain. Étude syntaxique*. Paris : Champion.
- Blasco, M. (1998). La séquence clitique + pronom tonique en français : un cas de prolongement pronominal. In M. Bilger, K. Van den Eynde, & F. Gadet (Éds.), *Analyse linguistique et approches de l'oral. Recueil d'études offert en hommage à Claire Blanche-Benveniste* (pp. 277-285). Leuven-Paris : Peeters.
- Berthoud, A.C. (1999). De la thématization des objets du discours. In C. Guimier (Éd.), *La thématization dans les langues. Actes du colloque de Caen, 9-11 octobre 1997* (pp. 379-398). New York : Peter Lang.
- Condillac, Abbé de, rééd. (1789). *Cours d'étude pour l'instruction du Prince de Parme, T.1, Grammaire*. Genève : François Dufart, Volland.
- Damourette, J., & Pichon, E. (1911-1940). *Essai de Grammaire de la Langue Française*. Paris : d'Artrey.
- Fournier, N. (1998). *Grammaire du français classique*. Paris : Belin.
- Hopper, P.J., & Traugott, C.E. (1992). *Grammaticalization*. Cambridge : Cambridge University Press.
- Jeanjean, C. (1984). Toi quand tu souris. Analyse sémantique et syntaxique d'une structure du français peu étudiée. *Recherches sur le français parlé*, 9, 131-165.
- Joly, A. (1992-1993). Thématization et focalisation : fondement d'une syntaxe énonciative. *Modèles linguistiques*, XIV-1, 87-98.
- Grevisse, M. (1986). *Le Bon Usage*. Gembloux-Paris : Duculot-Hatier. 13^{ième} édition.
- Le Querler, N. (1999). Dislocation et thématization en français. In C. Guimier (Éd.), *La thématization dans les langues. Actes du colloque de Caen, 9-11 octobre 1997* (pp. 262-275). New York : Peter Lang.
- Mertens, P. (1990). Intonation. In C. Blanche-Benveniste, M. Bilger, C. Rouget, & K. Van den Eynde (Éds.), *Le français parlé. Études grammaticales* (pp. 157-176). Paris : Éditions du CNRS.
- Morel, M.A., & Danon-Boileau, L. (1999). Thème, préambule et paragraphe dans l'oral spontané en français. In C. Guimier (Éd.), *La thématization dans les langues. Actes du colloque de Caen, 9-11 octobre 1997* : Peter Lang.
- Nölke, H. (1997). Note sur la dislocation du sujet : thématization ou focalisation ? In G. Kleiber, & M. Riegel (Éds.), *Les formes du sens. Études de linguistique française médiévale et générale offertes à Robert Martin* (pp. 281-294). Paris : Édition Duculot.
- Pallaud, B. (2002). Les amorces de mots comme faits autonymiques en langage oral. *Recherches sur le français parlé*, 17, 79-101.
- Riegel, M., Pellat, J.-C., & Rioul, R. (1994). *Grammaire méthodique du français*. Paris : PUF.
- Ronat, M. (1979). Pronoms topiques et pronoms distinctifs. *Langue française* 44, 106-128.
- Rossi, M. (1981). *L'intonation, de l'acoustique à la sémantique*. Paris : Klincksieck.
- Roubaud, M.-N. (2000). *Les constructions pseudo-clivées en français contemporain*. Paris : Champion.
- Sinclair, J. (1991). *Corpus, concordance and collocation*. Oxford : Oxford University Press.

Myiène BLASCO-DULBECCO

Stark, E. (1999). Antéposition et marquage du thème dans les dialogues spontanés. In C. Guimier (Éd.), *La thématization dans les langues. Actes du colloque de Caen, 9-11 octobre 1997* (pp. 333-358). New York : Peter Lang.

Lui, le propriétaire, le propriétaire, lui : deux constructions bien distinctes¹

Sandrine CADDÉO
Université de Provence

Résumé

Dans le domaine du détachement, la description des pronoms disjoints a souvent été traitée d'un point de vue pragmatique. L'élargissement de l'analyse aux cas où le pronom disjoint est suivi d'un syntagme nominal apposé du type *lui, le propriétaire*, ou est lui-même détaché d'un nom du type *le propriétaire, lui*, met en lumière une série de propriétés syntaxiques qui permettent d'identifier deux fonctionnements distincts des pronoms disjoints.

1. Introduction

Les différences de comportement des pronoms personnels disjoints² ont généralement été étudiées dans le domaine discursif, anaphorique ou emphatique³. Bien qu'apparemment très mobiles, les formes disjointes ne se laissent pas décrire dans les mêmes termes selon la position qu'elles occupent dans l'énoncé. Lorsque Nølke (1997) met en relation pronom et focalisation, il identifie deux comportements de ces éléments. Dans les exemples suivants que je lui emprunte, le pronom disjoint est dit

¹ Je remercie Claire Blanche-Benveniste et Mats Forsgren pour leur lecture attentive.

² Deux études viennent compléter la description des pronoms disjoints dans ce même numéro : l'une, de Paul Cappeau, s'intéresse à l'ensemble des formes dans les corpus de français parlé ; l'autre, de Mylène Blasco-Dulbecco, décrit la séquence « moi je ».

³ Kleiber (1994), Zribi-Hertz (1996), Nølke (1997).

Sandrine CADDÉO

successivement « élément focalisé » en (1) et « élément focalisateur » en (2) :

- (1) **Lui** n'a rien fait.
- (2) Pierre, **lui**, n'a rien fait.

Pour expliquer cette distinction, l'auteur fait un lien entre le rôle discursif et le rôle syntaxique des formes disjointes :

« les pronoms focalisés semblent toujours occuper des sites argumentaux, alors que les pronoms focalisateurs occupent des sites non argumentaux » (Nølke, 1997 : 58).

Cette conclusion paraît laisser de côté une série de caractéristiques fonctionnelles des pronoms disjointes que met en lumière l'analyse comparative des deux constructions. Une analyse plus poussée montre en effet que seuls les pronoms de la série *lui, elle, eux, elles* acceptent de jouer les deux rôles discursifs ; les pronoms *moi, toi, nous, vous* ne peuvent être que focalisés.

Autre contrainte intéressante à observer : les pronoms disjointes acceptent d'avoir dans leur dépendance un syntagme nominal apposé⁴, sauf s'ils focalisent eux-mêmes un autre syntagme nominal les précédant. En (3), *la photographe* est apposé à *elle*, mais pas en (4) :

- (3) Sans se douter qu'**elle, la photographe**, prenait date pour l'Histoire.
(Écrit)
- (4) ? Aline, **elle, la photographe**, n'a rien fait.

Blanche-Benveniste et al. (1987) avaient observé une autre contrainte liant, cette fois-ci, ordre des mots et compatibilités sémantiques. Quand le pronom est préposé au nom, celui-ci a obligatoirement le trait [+humain], alors que dans la suite *SN+Pro. disjoint*, aucune restriction ne s'impose. Il serait possible d'avoir (5) et (6), mais sûrement pas (7) :

- (5) **Lui, Pierre**, n'a rien fait.
- (6) Le tiroir, **lui**, est coincé.
- (7) ? **Lui, le tiroir**, est coincé

Au rôle discursif, se superpose un rôle grammatical qui a été moins décrit que le rôle pragmatique de ces éléments, sans doute à cause de leur faible charge lexicale.

J'étudierai le comportement des pronoms du type *moi, toi, lui, elle, nous, vous, eux, elles* dans deux positions. Quand ils sont suivis d'un ou de plusieurs groupes nominaux apposés, ce sera le type 1 (*Pro. disjoint+SN*) :

⁴ « Syntagme nominal » sera parfois abrégé SN.

Lui, le propriétaire, le propriétaire, lui : deux constructions bien distinctes

- (8) (...) que **lui, le prisonnier, il** serait le maître de cette main de bronze qui avait signé son écrou, et que de ce roi de boue il ne resterait que ce bras d'airain ! (Écrit)⁵

Quand ils suivent eux-mêmes un groupe nominal, ce sera le type 2 (*SN+Pro. disjoint*) :

- (9) ça peut des fois être un oui il y a des petits bruits que **le copain lui il** va entendre (Oral)

Les propriétés du pronom disjoint dans le deuxième type (cf. point 3) orientent le travail de comparaison des deux modèles vers des cas qui ne concernent que la position syntaxique de sujet.

Je passerai en revue quelques-unes des caractéristiques des pronoms qui justifient les regroupements que je propose pour les formes disjointes. Puis j'analyserai les propriétés syntaxiques des deux organisations linéaires mettant en jeu les pronoms personnels disjoints et les noms.

Ce travail dépasse le cadre même de l'analyse des pronoms personnels disjoints pour aboutir à une caractérisation des éléments détachés, dont ils illustrent bien certaines propriétés

2. Caractéristiques générales

Le tableau ci-dessous rappelle certaines des propriétés distinctives des pronoms personnels disjoints⁶.

⁵ Les exemples écrits sont tirés de la presse (Le Monde Diplomatique) et de « ABU » qui réunit une centaine d'œuvres littéraires du XIXème et XXème siècles. Les exemples oraux appartiennent en partie aux corpus du G.A.R.S. (Université de Provence) et à ma collection personnelle. Pour ma recherche, j'ai utilisé les concordanciers Contextes (Jean Véronis, <http://www.up.univ-mrs.fr/veronis>) et XCor (Jean-Pierre Adam).

⁶ J'ai laissé de côté l'étude de *ça* qui apparaît dans très peu d'exemples et qui semble présenter un comportement tout à fait original par rapport aux autres pronoms.

Sandrine CADDÉO

Liste des pronoms (Pro.)	Caractéristique 1	Caractéristique 2	Caractéristique 3
	Pro. + Nom [+humain]	Sujets directs	Nom [+/-humain] + Pro.
moi	+		
toi	+		
nous	+		
vous	+		
lui	+	+	+
elle	+	+	+
eux	+	+	+
elles	+	+	+

Pour développer mon analyse, j'ai sélectionné des exemples dans lesquels le pronom disjoint est sujet direct ou associé au sujet. Les observations n'en sont pas fondamentalement modifiées, mais il faut noter que seules les formes disjointes *lui*, *elle*, *eux*, *elles* acceptent d'être sujet direct d'un verbe (caractéristique 2)⁷ :

- (10) Sans se douter qu'**elle**, **la photographe**, prenait date pour l'Histoire.
(Écrit)
- (11) j'avais un pilote avec moi aussi **lui** a été éjecté mais sans gravité et tout s'est heureusement tout s'est très bien passé après (Oral)

Les formes disjointes *moi* et *toi* doivent obligatoirement être reprises par un clitique correspondant :

- (12) **moi** j'ai été surpris par l'attitude des Espagnols qui ont jamais essayé de rester ensemble (Oral)
? **moi** ai été surpris...
- (13) et mon père a dit **toi tu** feras comme tes frères tu te débrouilleras (Oral)
? **toi** feras comme tes frères...

Tous les pronoms acceptent d'être suivis d'un syntagme nominal apposé (caractéristique 1) :

⁷ Se pose le problème des pronoms *nous* et *vous* qui, en position sujet, devraient être considérés seulement comme des clitiques. En effet, ils ne peuvent guère être accentués : *nous viendrons*, mais **nous et pas vous viendrons* à côté de *lui et pas elle viendra* (José Deulofeu, communication personnelle).

Lui, le propriétaire, le propriétaire, lui : deux constructions bien distinctes

moi, la consommatrice	je	...
elle, la photographe,	elle	...
nous, les femmes	on	...
vous, les Occidentaux	vous	...
eux, les hommes de la délivrance...	ils	...

- (14) que **moi la consommatrice** j'en ai marre de consommer il faut le faire (Oral)
 (15) Sans se douter qu'**elle, la photographe**, prenait date pour l'Histoire. (Écrit)
 (16) et **nous les femmes on** est toutes sur ZOZO à Aix (Oral)
 (17) Pourquoi **vous, les Occidentaux, vous** ne faites rien (Oral)
 (18) Tandis que l'homme des préjugés et des servitudes, subitement transformé, rentrait dans l'humanité, **eux, les hommes de la délivrance et de l'affranchissement**, ils resteraient dans la guerre civile, dans la routine du sang, dans le fratricide ! (Écrit)

La détermination du nom peut être très variée. Les articles définis (exemples 14 à 18) et les possessifs (19) sont les plus fréquents ; indéfinis (20) et déterminants zéro (21) le sont beaucoup moins⁸ :

- (19) Capus me dit qu'il va mettre de l'ordre dans ses affaires, prendre une assurance pour sa femme, etc., que, d'ailleurs, il n'est pas inquiet, car, s'il venait à mourir, **nous, ses amis, nous** nous mettrions en quatre pour arranger son héritage de pièces. (Écrit)
 (20) **Lui, un homme d'honneur**, n'aurait jamais agi de la sorte. (exemple inventé)
 (21) Cela montre combien nous éprouverions de difficultés à trancher dans le vif si **nous, élus locaux**, nous nous attaquons à ces problèmes. (Écrit)

La détermination du nom joue un rôle primordial dans la description de l'apposition ; elle demanderait une étude à elle seule et ne sera pas développée dans cet article⁹.

Dans la suite *SN+Pro. disjoint*, nous n'avons pas à faire au même paradigme de pronoms (caractéristique 3). Cette construction n'est possible qu'avec les pronoms de troisième personne (*lui, elle, eux, elles*). Cette contrainte est imposée par l'élément lexical lui-même, « qui contient une indication de 3^{ième} personne » (Guillaume, 1988 : 207) :

- (22) **Les sympathisants, eux**, voteront Villiers. (Écrit)
 ? **Les sympathisants, vous**, voterez Villiers

⁸ Pour une étude détaillée, cf. Caddéo (2000).

⁹ Forsgren (1988), Picabia (1991, 2000), Tamine (1976).

Sandrine CADDÉO

(23) **Son autre fils de huit ans, lui**, a été passé à tabac quelques mois plus tard, pour les mêmes motifs. (Écrit)

? **Son autre fils de huit ans, moi**, a été passé à tabac quelques mois plus tard, pour les mêmes motifs¹⁰.

Au final, l'ensemble de ces observations place les pronoms personnels *moi* et *toi* à part. Ces pronoms ont un comportement très particulier dans l'apposition et les exemples sont peu fréquents. Les autres pronoms personnels disjoints ont une distribution assez similaire dans le type *Pro. disjoint+SN*, alors que c'est le comportement des pronoms disjoints *lui, elle, eux, elles* qui permettra de confronter les constructions du type *lui, le propriétaire* et *le propriétaire, lui*. La comparaison fera apparaître des contraintes syntaxiques (la possibilité ou non du pronom disjoint d'avoir une « extension nominale apposée ») et sémantiques (les compatibilités avec les noms [+/-humain]) très différentes qui seront approfondies.

3. Le type *lui, le propriétaire*

Les relations de dépendance qui se développent entre le pronom personnel disjoint et le syntagme nominal jouent sur les compatibilités lexicales et le critère de contiguïté.

3.1. Sélection sémantico-lexicale restreinte

Seuls les noms qui ont le trait [+humain] sont possibles dans cette position. Quel est l'élément qui impose cette contrainte ? le pronom personnel disjoint ? le verbe ? Avec des verbes qui peuvent accepter des sujets [+humain] ou [-humain], par exemple *englobait* en (24) et (25), les possibilités se réduisent dès que c'est une forme disjointe du pronom qui tient la position sujet. Seuls les noms ayant le trait [+humain] peuvent se postposer au pronom, cas (26) vs (27) :

(24) **Le conférencier, lui**, englobait tous les cas de figure dans son discours. (exemple inventé)

¹⁰ C'est peut-être possible avec une forte marque intonative qui rendrait une équivalence de type *le facteur, (en l'occurrence) moi, passera désormais à huit heures*. Dans ce cas, le pronom disjoint est proche d'une apposition. (José Deulofeu, communication personnelle).

Lui, le propriétaire, le propriétaire, lui : deux constructions bien distinctes

- (25) Mais un simple « défavorable » ne nous éclairerait pas sur cette modification qui jette la confusion alors que l'article 342, lui, englobait tous les cas de figure. C'est la raison pour laquelle nous l'avons repris dans notre amendement. (Écrit)
- (26) Lui, le conférencier, englobait tous les cas de figure dans son discours.
- (27) [?] Lui, l'article 342, englobait tous les cas de figure.

Les corpus ne contiennent aucun exemple de suite *lui, nom [-humain]* du type¹¹ :

- (28) [?] Lui, le commerce électronique, les priverait également de recettes fiscales, puisqu'un produit acheté à l'étranger et payé par carte de crédit, via un site de la Toile, puis acheminé directement à son acquéreur, ne donnerait pas lieu à l'acquittement de la TVA.
- (29) [?] Lui, le problème de fond, demeure, et il est toujours social.

Cette contrainte ne dépend pas du verbe, elle est donc imposée par le pronom disjoint dont la particularité avait été notée par Blanche-Benveniste et al. (1987) qui l'expliquaient ainsi :

(...) on a affaire à un élément quasi lexical, interprété comme « humain »; les rencontres avec le lexique pris dans la construction verbale sont de l'ordre des sélections lexicales plus ou moins bonnes. (pp. 49-50)

J'associerai la restriction de sélection que le pronom personnel disjoint impose au nom à la présence d'une relation forte de dépendance, que viennent appuyer les observations suivantes.

3.2. Insertion limitée

Le pronom et le syntagme nominal forment un groupe fortement soudé. Ils sont rarement séparés, sauf par des groupes restreints d'éléments : des adverbes temporels (*aujourd'hui*) ou des incises (*c'est vrai*) :

- (30) vous aujourd'hui syndicats vous pouvez répondre quoi (Oral)
- (31) nous c'est vrai purs Catalans comme je dis on préfère les produits catalans + mais les autres non (Oral)

Les autres éléments se placeront plutôt avant (32) ou après (33) la suite *Pro. disjoint+SN* :

¹¹ Les exemples (28) et (29), relevés dans *Le Monde Diplomatique*, ont été modifiés. Dans la version initiale, le pronom personnel disjoint est dans les deux cas immédiatement après le nom.

Sandrine CADDÉO

- (32) *Trop souvent, nous, les politiques, n'anticipons pas suffisamment sur les problèmes de société, et précisément sur ceux qui concernent les personnes âgées et les retraités.* (Écrit)
- (33) Les jeunes qui bénéficieront de ces nouveaux emplois seront employés sous contrat de droit privé à durée déterminée de cinq ans. Il conviendra donc que **nous, élus, dans le souci de maintenir un vrai dialogue social et un bon climat au sein de nos collectivités**, nous consultations les personnels sur la manière de régler un certain nombre de problèmes. (Écrit)

3.3. Le pronom personnel disjoint : une tête de syntagme

Leur « résistance » à la disjonction est l'indice de leur forte solidarité. Pour preuve, dans le dispositif clivé en *c'est ... que*, les deux éléments sont obligatoirement extraits ensemble :

- (34) Ce n'est pas **nous, laïcs et libéraux**, qui le lui aurions reproché : notre philosophie à nous, précisément, nous conduit à respecter l'opinion de chacun, même si les croyants, eux, ne respectent pas la nôtre. (Écrit)
- (35) Car c'est **lui, le dénonciateur**, qu'on fait, là aussi, passer pour un provocateur, un amateur de coupes médiatiques, condamné à dix-huit ans de prison pour trahison de secret d'État et actuellement jugé en appel dans l'indifférence totale des médias et de l'opinion publique. (Écrit)

L'ensemble de ces contraintes m'autorise à considérer que le groupe *Pro. disjoint+SN* forme un seul constituant dont le pronom personnel disjoint est la tête.

Lorsque le pronom personnel disjoint est postposé au nom, il ne véhicule pas les mêmes propriétés qui viennent d'être mises à jour.

4. Le type *le propriétaire, lui*

Lorsque le pronom personnel disjoint est postposé au nom, il a un comportement très différent de celui qui vient d'être décrit. Il apparaît comme un « pointeur » paradigmatissant qui marque le choix d'un élément parmi d'autres¹². Cette caractéristique avait également été mise en lumière par C. Blanche-Benveniste et al. (1987 : 49-50) :

¹² Cappeau (2000) a observé que le français parlé privilégiait la construction pour « singulariser » un élément parmi deux et marquer ainsi une opposition (p. 227).

Lui, le propriétaire, le propriétaire, lui : deux constructions bien distinctes

« Dans le cas d'un *lui* succédant à un élément lexical, on a affaire à une particule individualisante, et on est dans le domaine des inductions de sens : tout élément lexical auquel s'adjoint un *lui* sera affecté d'un sens individualisant. »

En (36), *dirigeants* est opposé à *élèves* ; en (37), dans le paradigme « genres romanesque ; [genres] poétique, théâtre », le pronom disjoint « individualise » *théâtre* des autres noms de la liste :

- (36) je parle des élèves + non pas des dirigeants **les dirigeants eux** ils sont au courant de bruits (Oral)¹³
- (37) Si les genres romanesque et poétique relèvent essentiellement d'une créativité individuelle, **le théâtre, lui**, a besoin pour s'épanouir d'un certain contexte. (Écrit)

À la différence du type précédent (*lui, le propriétaire*), le pronom disjoint perd son statut de tête de constituant, et le nom dont c'est le rôle impose de nouvelles contraintes.

4.1. Sélection sémantico-lexicale

Postposé au nom, le pronom disjoint ne privilégie pas de classe lexicale ; le nom peut avoir le trait sémantique [-humain], exemples (38) et (39) ou [+humain], exemples (40) et (41) :

- (38) Notre sergent instructeur nous disait : « Vous pouvez mourir, **votre fusil, lui**, subsistera. » (Écrit)
- (39) D'où le malaise des entrepreneurs : ils savent que le marché parfait des théoriciens libéraux n'existe pas, ou pas encore, dans la réalité, et que **le marché concret, lui**, a besoin de normes, de règles et même de règlements pour fonctionner. (Écrit)
- (40) **l'homme blanc lui il** ne se pose pas de questions (Oral)
- (41) **Les terroristes, eux, ils** ont des papiers ! (Écrit)

Il sera uniquement soumis aux phénomènes de l'accord en genre et nombre.

4.2. Restriction syntaxique

Le pronom disjoint ne porte que sur des éléments, nominaux ou pronominaux, en fonction sujet. Le plus ou moins grand éloignement entre le pronom détaché et le nom « cible » ne modifie pas cette restriction de sélection.

¹³ cf. Blasco-Dubelcco (1998).

Sandrine CADDÉO

Par exemple, lorsque le pronom disjoint suit un objet nominal du verbe susceptible d'être sous la portée de la focalisation, il n'y a pas de changement de « cible ». Dans l'exemple (42), on compte deux syntagmes nominaux : l'un est sujet (*mon beau-père*), l'autre est objet du verbe *parler* (*le patois*). Bien que le pronom disjoint soit contigu au second nom, il focalise le nom sujet sans aucune ambiguïté :

(42) **mon beau-père** parlait que le patois **lui** (Oral)

Pour nous assurer que ce n'est pas une question de sélection lexicale, analysons un exemple dans lequel se succèdent deux noms propres : *Max* est sujet, *Fabio* est le second élément de la comparaison. Le pronom disjoint focalise le sujet, *Max* :

(43) **Max**, moins artiste que Fabio, n'aimait, **lui**, que les entreprises difficiles, que les intrigues compliquées ; il cherchait des résistances à vaincre, des vertus à séduire, et conduisait l'amour comme une partie d'échecs, avec des coups médités longtemps, des effets suspendus, des surprises et des stratagèmes dignes de Polybe. (Écrit)

L'analyse des cas avec syntagmes nominaux sujets constitués d'un nom tête et de compléments restreint encore les observations : seul le nom tête peut être focalisé par le pronom disjoint. Dans l'exemple (44), le syntagme nominal présente la forme *N1 de N2* (*le vicaire de mon patelin Cossé-le-Vivien*), mais le pronom personnel disjoint ne prend pas pour « cible » *mon patelin Cossé-le-Vivien*, qui est le plus proche. Il individualise le nom tête *le vicaire* :

(44) et le vicaire de **mon patelin Cossé-le-Vivien** + dans la Mayenne **lui** euh m'a tout de suite mis le grappin + n'est-ce pas euh on allait au patronage avec l'abbé etc. et alors il a jugé que je pouvais faire un excellent prêtre (Oral)

Il existe une manière habile de focaliser un nom qui occupe une fonction autre que sujet ou qui n'est pas nom tête : la relative. Le nom peut être considéré comme l'élément focalisé s'il est repris par un pronom relatif sujet, comme le montre l'exemple suivant. L'objet prépositionnel « avec Des Bois » est suivi d'une relative. Le pronom détaché focalise le groupe formé par l'antécédent et le pronom relatif qui est sujet. La condition est remplie :

(45) Mais, au fait, c'est bien simple, dit-il tout à coup à voix basse, répondant à d'interrogantes pensées intérieures plus basses encore, Marchenoir est en fort bons termes avec **Des Bois**, **qui** est riche, **lui**. Je déciderai sans doute le docteur à faire quelque chose. (Écrit)

Lui, le propriétaire, le propriétaire, lui : deux constructions bien distinctes

Dans ce cas, le syntagme nominal (en gras dans les exemples) peut avoir n'importe quelle fonction dans l'énoncé : objet prépositionnel du verbe (46), complément de nom (47), etc. :

- (46) En fait, le patronat se débarrasse de la séparation, économiquement contre-productive, entre travail manuel et travail intellectuel pour faire naître un « cotravailleur » individualisé se substituant à l'ouvrier ou à **l'employé d'hier qui, lui**, était membre d'une communauté. (Écrit)
- (47) À la dépossession complète du malade que revendiquaient les médecins, le juge californien a répondu avec une grande sévérité : il est d'une « ironie grossière » que les médecins prétendent, eux, avoir un droit de propriété sur les tissus **d'un malade qui, lui**, n'aurait aucun droit. (Écrit)

C'est ce qui explique qu'un pronom détaché focalisateur n'est jamais placé avant le pronom relatif sujet¹⁴ :

- (48) on a euh dans l'ordre un assistant de clientèle qui s'occupe du guichet on dit qu'ensuite le niveau au-dessus c'est un conseiller de clientèle qui s'occupe des placements et des prêts ensuite on a **le chargé de clientèle qui, lui**, s'occupe de tout ce qui est prêts professionnels (Oral)
? on a le chargé de clientèle, **lui**, qui s'occupe de tout ce qui est prêts professionnels

Au final, les contraintes qui existent entre le nom et le pronom se font selon deux orientations : du nom vers le pronom disjoint pour ce qui est des accords en personnes, du pronom disjoint vers le nom (ou le clitique) pour ce qui est des sélections syntaxiques. Est-ce l'indice d'une interdépendance ? La position du pronom détaché dans une suite de constituants donnera un élément de réponse.

4.3. Disjonction

Au contraire du type 1 (*lui, le propriétaire*), le groupe formé par le nom et le pronom a tendance à se dissocier. Toutes les positions sont envisageables. Le pronom détaché est généralement avant le verbe :

- (49) **Son autre fils de huit ans, lui**, a été passé à tabac quelques mois plus tard, pour les mêmes motifs. (Écrit)

¹⁴ Sauf dans le cas où le pronom disjoint est l'antécédent d'une relative :

Un tel défi signifierait que les entreprises deviennent performantes : mais l'actuel gouvernement n'a-t-il pas renoncé à cet objectif, **lui qui compte exclusivement sur le pétrole et le gaz pour soutenir la croissance** ? (Écrit)

Sandrine CADDÉO

Il peut apparaître également dans la zone verbale, par exemple entre l'auxiliaire (*a*) et le verbe recteur (*choisi*) :

- (50) **L'Algérien Saïd Ould Khalifa a, lui**, choisi un sujet grave et original pour son premier film : *Ombres blanches*. (Écrit)

Il peut être postposé au verbe et le séparer de son objet. Dans l'exemple suivant, il est inséré entre le verbe *savoir* et la subordonnée complétive :

- (51) Et surtout mal informé : il ne soupçonnait pas un instant que **le bon peuple**, pour ignorant qu'il fût, sait bien, **lui**, qu'il existe de vrais privilégiés. (Écrit)

Le pronom détaché peut également suivre la réaction verbale, mais ce n'est pas une place privilégiée :

- (52) **mon beau-père** parlait que le patois **lui** (Oral)

Dans ce cas, le sujet est plus fréquemment un clitique (53) et le syntagme verbal se réduit souvent au verbe (*est élu* dans l'exemple 54) ou au verbe et un élément de réaction (*trouverait* et *quelque chose* dans l'exemple 53) :

- (53) La supériorité de Rostand, c'est qu'il nous accable et que nous ne trouvons rien à lui dire. Si nous avons fait son *Cyrano*, **il** nous trouverait quelque chose, **lui**. (Écrit)
(54) Veuillez m'excuser, monsieur le garde des sceaux, mais il y a une différence fondamentale : **le conseiller municipal** est élu, **lui**. (Écrit)

Cette exigence assez grande de contiguïté s'explique par le lien syntaxique que l'élément nominal et le pronom détaché développent.

4.4. Le pronom personnel détaché : une borne fermante

Lorsque la zone préverbale est occupée par plusieurs groupes d'éléments, le pronom détaché « passe » souvent après le verbe :

- (55) **Amos Tutuola**, ce « bricoleur inspiré », à l'univers verbal *exceptionnel*, réussit, **lui**, l'exploit de fonder le roman nigérian (en 1952) par le biais d'un « bricoleur inspiré », très yorubaïsé. (Écrit)
(56) **Mme Irène Herbulot**, « mère de famille de trente et un ans », demeurant à Caluire, s'indigne, **elle**, que l'on puisse faire parler des criminels nazis : (...) (Écrit)
(57) **Le troisième principe**, celui de l'adéquation des installations d'élimination, est, **lui**, contesté par les autorités américaines et japonaises. (Écrit)

Lui, le propriétaire, le propriétaire, lui : deux constructions bien distinctes

Mais lorsque le pronom détaché reste en zone préverbale, il paraît sensible aux types d'éléments qui se succèdent et sa position est contrainte. S'il s'agit de groupes d'éléments appartenant à la rection du verbe (constituants prépositionnels, compléments de type « quand »), le pronom détaché reste proche du syntagme nominal qu'il individualise :

(58) Tandis que l'écrivain, **lui**, jusqu'à nouvel ordre, reste un individu. (Écrit)

Si le nom sujet est suivi d'éléments apposés, le pronom détaché est souvent repoussé derrière ces éléments¹⁵ :

(59) **Les pays méditerranéens (Grèce, Italie, Espagne et Portugal), eux**, souhaitent retarder l'échéance le plus longtemps possible par peur d'être débordés par des transporteurs puissants comme ceux que possèdent la France et la RFA, qui disent, quant à elles, préférer une stratégie évolutive, qui les maintiendrait, de fait, à la première place en Europe. (Écrit)

Cette caractéristique ne s'applique pas si les éléments détachés sont des groupes adjectivaux (60) ou participiaux (61 et 62). Le pronom détaché demeure près du nom tête :

(60) Si l'abstentionnisme sévit surtout dans les milieux défavorisés, **les riches, eux, conscients de leurs intérêts**, votent, font voter et ne lésinent pas pour financer les campagnes électorales des candidats qui leur sont dévoués. (Écrit)

(61) **Leurs parents, eux, venus pour la plupart des campagnes entourant la grande cité ottomane, poussés par on ne sait quel mirage**, survivent comme ils peuvent sur la décharge industrielle où ils ont élu domicile, chemin des Ordures. (Écrit)

(62) Il y a une ironie de l'histoire dans le fait que les Russes s'apprentent à s'embarquer sur la « voie polonaise » précisément au moment où **les Polonais, eux, épuisés par la méthode Balcerowicz**, paraissent décidés à changer de cap. (Écrit)

Ces observations permettent de dégager une propriété principale. Le pronom semble rester à la limite externe du constituant, mais ce n'est pas seulement un élément détaché marqué par une intonation spécifique, il se comporte également comme une véritable « borne syntaxique ». Il reste à

¹⁵ Jean Véronis (communication personnelle) m'a fourni quelques exemples, appartenant tous aux homélies, dans lesquels le pronom disjoint est suivi d'un syntagme nominal apposé :

Jésus, Lui, le fils de Dieu est mis au rang des malfaiteurs, on ne mettait au supplice de la croix, que des criminels ou des bandits d'envergure. (Écrit)

Pierre, lui, le chef, cherche le coupable et s'il l'avait trouvé il l'aurait fracassé. (Écrit)

Sandrine CADDÉO

savoir quel rôle il joue. Est-il la borne finale du constituant fonctionnel ? de l'apposition ?

5. Le pronom disjoint et *autres, tous, aussi*

Un autre indice de la distinction du comportement du pronom personnel disjoint dans les deux types a trait aux cas où il est accompagné des éléments *autres, tous, aussi*¹⁶ dont la distribution est contrainte. Ces éléments ne sont pas compatibles avec les mêmes pronoms. *Autres* et *tous*, parce qu'ils sont pluriels, acceptent de suivre les formes disjointes de personne quatre et cinq pour le premier (*nous autres, vous autres*) et quatre, cinq et six pour le second (*nous tous, vous tous, eux tous, elles toutes*) ; *aussi* s'associe avec toutes les personnes.

La seconde restriction est étroitement liée au statut du pronom disjoint. Dans le type *lui, le propriétaire*, seuls *autres*¹⁷ et *tous* sont possibles. Ces éléments ne modifient pas les propriétés syntaxiques du pronom disjoint, mais lui donnent une valeur sémantique supplémentaire de type globalisant :

- (63) **Nous autres, Latino-Américains**, sommes intéressés par un dialogue avec les pays développés sur les grands problèmes économiques qui ont créé le dramatique clivage entre le Nord et le Sud. (Écrit)
- (64) Je suis sûr que **vous tous, membres de la représentation nationale**, serez porteurs d'idées et d'exigences neuves pour développer ce nouveau contrat moral. (Écrit)

Dans le type *le propriétaire, lui*, seul *aussi* est envisageable ; ce qui n'a aucune incidence sur l'analyse du pronom détaché, qui se charge simplement d'une valeur finale :

¹⁶ J'ai laissé de côté l'étude de *seul* et de *même* qui développent des comportements assez particuliers. Par exemple, *seul* est obligatoirement accompagné de la préposition *à* ou du connecteur *et* quand il est détaché du nom :

Les forces armées, **à elles seules**, absorbent entre 150 et 200 millions de dollars par an.
(Écrit)

La Cour de La Haye, **et elle seule**, peut en décider. (Écrit)

Et *lui-même* doit être considéré comme un autre pronom. Il ne répond en effet à aucune des propriétés observées avec *lui* (Caddéo, 2000).

¹⁷ Grevisse (1993, 13e éd. : 1084) mentionne *eux autres* comme un emploi régional (attesté au Québec) ou populaire ailleurs (notamment à Paris). Il ajoute que la forme est utilisée par Molière dans sa comédie *L'Étourdi* (1655).

Lui, le propriétaire, le propriétaire, lui : deux constructions bien distinctes

- (65) S'il a rencontré un grand écho en Italie, **ce mouvement, lui aussi**, est resté solitaire en Europe. (Écrit)

6. Conclusion

L'analyse a mis à jour des comportements du pronom personnel disjoint distinctifs selon sa position. Dans le cas de *lui, le propriétaire*, le pronom disjoint se comporte comme la tête d'une construction nominale qui accepte une « extension » apposée. En effet, le référent, non encore « saturé » par le pronom, peut se construire avec l'association d'un (ou plusieurs) groupe nominal apposé. Le pronom et le nom apposé sont l'un et l'autre solidaires (Forsgren, 1993). De plus, le pronom établit une sélection sémantique puisqu'il n'est compatible qu'avec des noms [+humain].

Dans le modèle *le propriétaire, lui*, le nom ne trouve auprès du pronom qu'un index à valeur discursive. Le pronom disjoint se comporte comme un élément paradigmatissant proche des éléments spécialisés dans la focalisation comme *pour sa part, par contre, au contraire* (cf. Cappeau, 2000 : 213). Il a pour effet d'« individualiser » le nom pris dans un paradigme de plusieurs éléments. Il est donc peu sensible au trait sémantique du nom, mais ne porte que sur des noms têtes de syntagmes nominaux sujets.

	Lui, le propriétaire	Le propriétaire, lui
Types de pronoms	tous les pronoms	« lui, elle, eux, elles »
Types de noms	nom [+humain]	nom [+/-humain]
Fonctions du nom	--	nom tête sujet
Solidarité	une unité	deux unités
Contraste	-	+

Dans d'autres langues, l'on rencontre des constructions semblables au type *lui, le propriétaire*. En portugais, le chapitre « Aposto » (« Apposition ») de la grammaire de Cunha & Cintra (2000, 16^e éd.) est illustré par un exemple :

Eles, **os pobres desesperados**, tinham uma euforia de fantoches. (Fernando Namora, DT, 237) (Eux, les pauvres désespérés, avaient une euphorie de fantoche)

Si, en portugais, aucune mention n'est faite du fonctionnement du pronom disjoint sur le modèle *le propriétaire, lui*, en italien et espagnol,

Sandrine CADDÉO

certaines constructions semblent s'en rapprocher tout en présentant d'autres spécificités.

Dans la « Grande grammatica italiana di consultazione », on lit :

« (...) il pronome funge da rafforzativo di un sintagma nominale soggetto, come negli esempi seguenti :

- a. Giorgio ha voluto cucinare lui oggi.
- b. I ragazzi hanno portato loro il gelato. » (p. 541)

Ces constructions sont rapprochées des sujets en position postverbale qui font un effet de mise en relief, mais nous ne savons pas si le pronom appelé « il pronom libero » peut se rencontrer immédiatement après le nom sujet. Dans la traduction française, nous serions même tenté de donner une équivalence avec le pronom réfléchi :

- a. Georges a voulu cuisiner lui-même aujourd'hui.
- b. Les enfants ont porté eux-mêmes leur glace.

En espagnol, la structure est très contrainte. Dans l'article consacré au pronom personnel de la « Gramática descriptiva de la lengua española » (1999), l'on trouve :

« En registros non formales, los pronombres tónicos de tercera persona pueden acompañar a ciertos adjetivos. Aparecen en estructuras apositivas, con entonación de coma, que modifican a nombres comunes, generalmente indefinidos :

- (100)a Un hombre, gordito él, con corbata de cuadros, ha venido a preguntar por ti. » (p. 1240)

Ici aussi, il est difficile de trouver une stricte équivalence en français. L'adjectif accepterait difficilement d'être suivi, et même précédé, d'un pronom disjoint, qu'il faut supprimer :

Un homme, tout gros, avec une cravate à carreaux, est venu te demander.

Chaque langue présente des spécificités dans ce domaine qu'il sera très intéressant de confronter.

Références

- Blanche-Benveniste, C., Deulofeu, J., Stefanini, J., & Van den Eynde, K. (1987). *Pronom et syntaxe. L'approche pronominale et son application à la langue française*. Paris : S.ELAF (2^{ème} Édition).
- Blanche-Benveniste, C., Bilger, M., Rouget, C., & Van den Eynde, K. (1990). *Le français parlé - Études grammaticales*. Paris : Éditions du CNRS.
- Blasco, M. (1998). La séquence clitique+pronom tonique en français : un cas de prolongement pronominal. In M. Bilger, K. van den Eynde, F. Gadet (Éds.),

Lui, le propriétaire, le propriétaire, lui : deux constructions bien distinctes

- Analyse linguistique et approches de l'oral. Recueil d'études offert en hommage à Claire Blanche-Benveniste* (pp. 277-285). Leuven/Paris : Peeters.
- Blasco-Dulbecco, M., & Caddeo, S. (2001). Apposition et dislocation : la séquence pronom+lexique+clitique. *Recherches sur le français parlé*, 16, 124-149.
- Bosque, I., & Demonte, V. (Eds.) (1999). *Gramática descriptiva de la lengua española*. España : Espasa.
- Cappeau, P. (2000). Sujets éloignés. Esquisse d'une caractérisation des sujets lexicaux séparés de leur verbe. *Recherches sur le français parlé*, 15, 199-231.
- Cunha, C., Cintra, L. (2000). *Nova gramática do português contemporâneo*. Lisboa : Edições João Sá da Costa (16^{ième} Édition).
- Diderichsen, P. (1972). *Essentials of Danish grammar*. Copenhagen : Akademisk forlag.
- Forsgren, M. (1988). Apposition nominale : déterminants et ordre des constituants. *Travaux de linguistique française*, 17, 137-157.
- Forsgren, M. (1993). Connecteurs argumentatifs et prédication seconde : observation sur même concessif. *Actes du XX^{ième} Congrès International de Linguistique et de Philologie Romanes*, (pp. 237-246), Zürich, Allemagne.
- Grevisse, M. (1993). *Le bon usage*. Paris : Duculot (13^{ième} Édition).
- Guillaume, G. (1988). *Leçons de linguistique de Gustave Guillaume - 1947-1948*. Laval/ Lille : CPU de Laval/ PU de Lille.
- Guillaume, G. (1989). *Leçons de linguistique de Gustave Guillaume - 1946-1947*. Laval/ Lille : CPU de Laval/ PU de Lille.
- Kleiber, G. (1994). *Anaphores et pronoms*. Louvain-la-Neuve : Duculot.
- Nölke, H. (1997). Anaphoricité et focalisation : le cas du pronom personnel disjoint. In W. Mulder De, L. Tasmowski-De Ryck, & C. Vetter (Eds.), *Relations anaphoriques et (in)cohérences* (pp. 55-67). Amsterdam : Rodopi.
- Picabia, L. (1991). Article zéro et structures apposées. *Langages*, 102, 88-102.
- Picabia, L. (2000). Appositions nominales et déterminant zéro : le cas des appositions frontales. *Langue française*, 125, 71-89.
- Renzi, L., Giampaolo, S., & Cardinaletti, A. (Eds.) (2001). *Grande grammatica italiana. Vol. I* (Nuova edizione). Bologna : Il Mulino.
- Skårup, P. (1975). Les premières zones de la proposition en ancien français. Essai de syntaxe des positions. *Études Romanes*, 6. Danemark : Akademisk Forlag.
- Tamine, J. (1976). Une discussion de méthode à propos de l'apposition. In J.-C. Chevalier, & M. Gros (Eds.), *Méthodes en grammaire française* (pp. 133-150). Paris : Klincksieck.
- Zribi-Hertz, A. (1996). *L'anaphore et les pronoms : une introduction à la syntaxe générative*. Lille : Éditions du Septentrion.

Du bon usage des amorces dans la transcription des corpus

Marie-Noëlle ROUBAUD
Université de Provence

Résumé

L'équipe du GARS (Groupe Aixois de Recherches en Syntaxe) relayée par l'équipe DELIC (DEscription Linguistique Informatisée sur Corpus) s'est penchée sur le problème de la transcription des amorces de mots (ou fragments de morphèmes). L'examen de dix-neuf corpus de français parlé montre que cette tâche est ardue. Des questions se posent au transcripteur : Comment identifier les amorces ? Comment les transcrire en restant fidèle à la parole du locuteur ? N'existe-t-il pas plusieurs interprétations possibles ? Cet article tente d'apporter des réponses et veut surtout montrer que la transcription des amorces passe par une analyse syntaxique des corpus.

1. Introduction

Depuis plus de vingt-cinq ans, l'équipe du GARS relayée par l'équipe DELIC collecte des corpus de français parlé en vue de recherches syntaxiques. Ce travail l'a conduite à transcrire les données orales en orthographe standard, sans ponctuation et sans marque intonative et à établir des conventions de transcription. Parmi celles-ci figure la transcription des amorces de mots (en gras dans l'extrait) dont Pallaud (2002 : 79) indique que leur présence se traduit par « une interruption de morphèmes en cours d'énonciation »¹ :

¹ Son étude quantitative révèle que les amorces « constituent plus de la moitié des achoppements de langage identifiés dans un corpus enregistré de français contemporain » (Pallaud, 1999, 2002).

Marie-Noëlle ROUBAUD

(1) voilà le jour de l'Ascension il fallait rien faire fallait pas trava- faut pas travailler le jour de l'Ascension c'est un jour particulier alors l'herbe de l'Ascension l'œuf de l'Ascension la veilleuse de l'Ascension tout ça ça permettait de se protéger contre les éléments les élé- les les éléments naturels et je ne sais pas si ça donnait des bons résultats mais nos vieux ils y croyaient et et le soleil il revenait et l'ar- l'arc-en-ciel on le voyait au loin on disait tu vois et maintenant ils vont rentrer et il n'y aura pas eu de + voilà ça ce sont des usages qui ont qui existent encore à l'heure actuelle (COR-PRI001)

L'observation et la correction de dix-neuf corpus de français parlé² (soit un peu plus de 46 000 mots) montre que la transcription des amorces n'est pas une tâche facile, contrairement à ce qu'on pourrait penser : elle met en cause une analyse syntaxique.

Cet article veut rendre compte de cette réflexion sur les amorces, reposant sur tout un travail collectif de transcrip-teurs³. Pour cela, nous reviendrons d'abord sur les principes de transcription du GARS pour ensuite considérer le cas des amorces. Puis nous tenterons de dresser une liste d'outils aidant à l'identification des amorces. Et enfin, nous reviendrons sur les problèmes rencontrés dans les corpus observés : nous nous interrogerons sur le bon usage des amorces, examinerons les cas difficiles, proposerons quelques solutions⁴.

2. Les principes de transcription

Quand nous lisons, nous avons l'habitude de reconnaître, sous le découpage en mots, des morphèmes, c'est-à-dire des unités de sens. L'orthographe est notre « déchiffreur » courant des morphèmes⁵. C'est pourquoi le GARS puis DELIC ont choisi de transcrire des morphèmes, en orthographe standard, orthographe donnée par les dictionnaires.

² Ces corpus ont été enregistrés dans différentes régions de France et dans différentes situations (conversations, relations d'évènements, discours de guide, cours de professeur, etc.), dans le cadre d'un projet de constitution d'un *Corpus de référence de français parlé*, conduit par l'équipe DELIC. La dernière phase de ce projet a consisté en la relecture et la correction de ces corpus.

³ Cette réflexion a été menée par l'équipe du G.A.R.S. et se poursuit au sein de l'équipe DELIC.

⁴ Les exemples qui ont été corrigés seront présentés sous deux versions : la version initiale (ex.) et celle corrigée (ex. corrigé).

⁵ cf. Blanche-Benveniste et Jeanjean (1986 : 124).

Cette précaution à transcrire tous les morphèmes justifie que dans les conventions de transcription apparaît un *n'*, hésitation orthographique qui rend compte de la présence possible du morphème *ne* de négation, impossible à entendre après une liaison comportant une nasale :

- (2) on (**n'**) a pas de moyens (crfp, prilar1)

La transcription ne comporte aucun « trucage orthographique »⁶ pour rendre compte des réductions de morphèmes absolument courantes, caractéristiques de la langue parlée, quels que soient les locuteurs. Que la voyelle du pronom *tu* ou la consonne /r/ de *parce que* soit élidée, l'orthographe du morphème est conservée :

- (3) et **tu** as **tu** as pas mal d'hôtels qui sont aménagés (crfp, prilem1)

et non pas :

et t'as t'as pas mal d'hôtels qui sont aménagés

- (4) mais à part les ruines il te reste vraiment pas grand-chose **parce que** tout a été détruit (crfp, prilem1)

et non pas :

mais à part les ruines il te reste vraiment pas grand-chose **pa'ce que** tout a été détruit

Il en est de même pour les nombreux pronoms *il(s)* ou *elle(s)* parsemant les corpus, dont la prononciation est réduite le plus souvent à l'émission d'une voyelle. La transcription rétablit l'orthographe du morphème :

- (5) par exemple **il** tu as remarqué le long de la rivière il y a des petites chapelles (crfp, pricor1)

et non pas :

par exemple **i-** tu as remarqué le long de la rivière

- (6) bon la vigne c'est sûr que **elle** + il y a **i-** + elle est plus souvent chez nous du moins elle est plus souvent défeuillée que feuillée (crfp, prigap1)

et non pas :

bon la vigne c'est sûr que **è-** + il y a **il** + elle est plus souvent chez nous ...

La particularité de prononciation peut être mentionnée ailleurs, par exemple sous forme d'une étude phonétique adjointe à la transcription ou en note de bas de page. Il est plus facile de partir de l'identification des

⁶ cf. Blanche-Benveniste et Jeanjean (1986 : 130).

Marie-Noëlle ROUBAUD

morphèmes, d'après une transcription orthographique, et de rajouter ensuite les particularités phoniques, que d'aller en sens inverse. En effet si nous considérons l'exemple de *il y a*, plusieurs transcriptions phonétiques sont possibles [ilija], [ilja], [ija], [ja] ou encore chez les enfants [ila], [la]. Toutes les prononciations évoquées représentent la même suite de morphèmes. Si on s'intéresse à la combinaison des morphèmes, et non au niveau inférieur de la prononciation des morphèmes, on ne note pas la prononciation.

C'est au nom de ce principe que nous ne descendons pas en dessous du morphème⁷. Ainsi dans l'exemple qui suit, nous n'indiquerons pas l'enchaînement entre le déterminant *une* et le nom *équipe* :

- (7) on peut tomber la première partie contre **une équipe** qui est qui est
can- cant- comme nous (crfp, prilil1)

et non pas :

on peut tomber la première partie contre **une /n/ équipe** qui est qui est
can- cant- comme nous

3. Le cas des amorces

Les amorces sont des fragments de morphèmes, ou d'hypothèses qu'on fait sur les fragments de morphèmes, elles sont transcrites avec un trait d'union, collé à l'amorce. Ce sont des unités de sens ainsi que des représentants de constituants syntaxiques. En effet, il nous suffit de lire les amorces, dans les trois exemples suivants, pour identifier des unités de sens et les syntagmes correspondants :

- (8) j'en connais qui + des élevages porcins la la preuve c'est quand même
des + des **polémi-** des polémiques au niveau de environnemental pour
le la création de nouveaux élevages et tout ça (crfp, prigap1)
- (9) c'est-à-dire que les les **f-** les **états fédérés** ressemblent à l'Allemagne
ressemblent à l'Italie ressemblent à l'Angleterre (crfp, prigap2)
- (10) et c'est le coq au vin euh est un de ces est un de ces plats + on l'**ap-** au
départ on faisait euh le le euh le macérer le le le poulet (crfp, pridij1)

Si pour l'exemple 8, l'amorce est évidente puisque elle est complétée aussitôt après par *des polémiques*, pour l'exemple 9, c'est le contexte qui permet d'interpréter l'amorce : amorce de *fédérations* ou anticipation du lexème *fédérés* produit ensuite. Quant à l'exemple 10, c'est la place syntaxique occupée par l'amorce (en position verbale) qui peut la faire

⁷ cf. Giovannoni & Savelli (1990 : 21).

Marie-Noëlle ROUBAUD

- recherche du nom dans le syntagme nominal *l'intitulé du diplôme* :

(12) alors que **l'in-** **l'intitu-** l'intitulé du du diplôme est le même hein (crfp, pubcor1)

G : alors que **l'in-**
l'intitu-
l'intitulé du diplôme est le même hein

- recherche du nom dans le syntagme nominal *un avenir certain* :

(13) euh hum l'Aligoté aussi a un vin qui + qui aura euh + **un un aven-** un **a-** un avenir certain dans dans des plats euh surtout dans des entrées ou des apéros (crfp, pubdij1)

G : qui aura **un**
un aven-
un a-
un avenir certain

- recherche du nom avec correction du lexème adjectival dans le syntagme nominal :

(14) quand il y avait une belle **nou-** une bonne nouvelle elle allait elle allait très vite même s'il n'y avait pas de téléphone (crfp, pricor1)

G : quand il y avait **une belle** **nou-**
une bonne **nouvelle**

- recherche du nom avec correction du nombre du déterminant dans le syntagme nominal :

(15) on a quand même un laps de de repos moi je suis du matin ma collègue est du soir + euh elle a son samedi moi j'ai mon lundi on a notre **rep-** nos repos + qui sont obligatoires (crfp, pricle3)

G : on a **notre rep-**
nos repos

- recherche du verbe dans le syntagme verbal :

(16) et puis lui il **d-** il désignait deux tireurs deux pointeurs (crfp, prilil1)

G : et puis lui **il d-**
il désignait deux tireurs
deux pointeurs

- recherche du verbe infinitif et abandon de l'amorce (probablement *faire*) pour un lexème plus spécifié *construire* :

(17) quand on a commencé à détruire le vieux couvent pour **f-** construire le groupe scolaire on a construit en même temps le clocher actuel de l'église (crfp, pricor1)

G : quand on a commencé à détruire le vieux couvent **pour f-**
construire ...

Dans les corpus, nous trouvons de nombreux monosyllabes concernés par les amorces, comme les déterminants (ex. 18), les pronoms (ex. 19), les prépositions (ex. 20) :

- (18) bon on va repartir sur **c-** après **cette** évocation de de de un peu de l'histoire du conflit entre les deux langues (crfp, prigap2)
- (19) je **v-** **vous** citerai pas tout (crfp, pridij1)
- (20) je leur souhaite à se donner les moyens + **d-** **dans** l'Europe c'est peut-être un peu plus difficile (cfp, prigap2)

Ces amorces subissent les déformations phonémiques et leur correction a lieu, toujours en revenant au départ du groupe syntagmatique. Dans l'exemple 21, le locuteur anticipe, à l'intérieur de l'amorce, sur le phonème /r/ qui vient trop tôt dans le mot *patrimoine*⁹ et dans l'exemple 22, il anticipe sur une voyelle du mot *appellation* :

- (21) il y a des des des sujets qui ont /a-, a/ trait au monde rural et des a- des sujets euh qui ont /a-, a/ trait au au **par-** au au **patrimoine** euh euh naturel (crfp, procor1)
- (22) il vient de Mercurey donc c'est en Saône-et-Loire + normalement bon c'est une euh région qui est assez connue euh pour ses blancs + ses blancs de Chardonnay + où les blancs sont souvent élevés en fûts donc on verra si ce vin est élevé en fût euh d- de cette **epp-** de cette **appellation** (crfp, pubdij1)

Comme nous le voyons au travers de ces exemples, ce sont, dans la majorité des cas, des amorces d'anticipation que le locuteur produit, témoignant de sa capacité à effectuer des allées et venues sur l'axe syntagmatique. Comme l'écrit Claire Blanche-Benveniste (1990 : 23) :

« 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. »

Dans les exemples précédents, cette mise en mémoire était brève car peu de syllabes séparaient l'amorce du morphème appartenant au même groupe syntagmatique. L'exemple 23 constitue une exception car sept syllabes séparent l'amorce *op-* du lexème *opération*, mais le syntagme *on peut considérer que*, qui s'intercale entre eux, agit comme une sorte de précaution oratoire, dans un discours qui paraît un peu embrouillé. C'est ce que visualise la mise en grille sous l'exemple :

- (23) bon maintenant on a quand même une population de près de cent cinquante cerfs bon /dont, donc/ une partie + hein euh **l'op-** on peut

⁹ Ce phénomène de déplacement de la consonne « r » d'une syllabe à l'autre est fréquent chez les enfants (Blanche-Benveniste & Pallaud, 2001 : 22).

Marie-Noëlle ROUBAUD

considérer que **l'opération** de réintroduction est en en voie d'aboutissement parce qu'on ne en en général hein une une opération de réintroduction n'est réussie que lorsque il y a des animaux hein qui sont passés par une phase /d/, 0/ élevage qui sont retournés au milieu naturel + (crfp, procor1)

G : l'op-
on peut considérer que **l'opération** de réintroduction est en en voie d'aboutissement parce que ...

Dans les corpus étudiés, on trouve peu d'exemples montrant un locuteur mettant en mémoire le syntagme dans son intégralité puis le reprenant d'une façon fragmentée. Dans l'exemple 24, c'est le syntagme *ça me déplairait* qui est mis en mémoire et dans l'exemple 25, *on avait un effectif* qui sera repris ensuite mais d'une façon incomplète, avec une amorce (signalée en gras) :

(24) mais ça me déplairait pas + ça me **dé-** d'autant plus que je suis assez bon en anglais (crfp, prigap2)

(25) on avait un effectif on a eu jusqu'à un **eff-** euh ici sur Moulin on a été jusqu'à soixante-quatorze (crfp, prilil1)

L'amorce porte sur les premiers éléments du morphème et très rarement sur les derniers éléments¹⁰ ; dans ce cas le trait d'union collé à l'amorce se trouve à gauche de l'élément. Dans l'exemple 26, c'est l'interlocuteur (noté L2) qui corrige la production du locuteur (noté L1) et dans les deux exemples suivants, c'est le locuteur qui se corrige lui-même. En 27, c'est un enfant de 10 ans qui rectifie le temps du verbe et en 28, c'est un adulte qui rétablit la consonne initiale :

(26) L1 ça pousse et ça ça fait des petites **stalactites**
L2 **-mites -mites** qui montent
L1 c'est ça **-mites** qui montent oui des stalagmites (Corpaix, 1995-
Champignons)

(27) bon alors je reviens à la question que tu m'avais posée tout à l'heure Isabelle + tu m'as demandé si je **préfèrais -rerais** chanter en groupe enfin à deux ou à ou toute seule (Corpaix, 1984-Puget1)

(28) des négociations ont d'ailleurs été entamées avec douze pays Chypre Malte et dix pays de l'Europe de l'Est dont la **-ologne** la **Pologne** pardon et la Hongrie (Corpus Delic, 2002-Poitiers)

En dehors d'une analyse syntaxique, des outils peuvent aider le transcripteur à identifier les amorces.

¹⁰ Nous n'en avons trouvé aucun exemple dans les 19 corpus étudiés. Nos exemples, fournis aimablement par Paul Cappeau, proviennent d'autres corpus de français parlé.

4. L'identification des amorces

4.1. Le dictionnaire

Le dictionnaire¹¹ peut nous renseigner sur l'existence du morphème, comme il le fait pour « ovis », indiquant que ce mot regroupe les moutons et les mouflons :

(29) vous savez que le le mouflon c'est un ovis c'est un ovin (crfp, prorol)

Mais il faut être prudent : les mots étrangers, le lexique spécialisé, etc. demandent une recherche complémentaire, non seulement dans les dictionnaires mais également auprès des locuteurs qui les emploient.

4.2. Le contexte

L'identification de l'amorce se fait dans le syntagme et en contexte. C'est ainsi que dans les exemples suivants, c'est le contexte qui suggère l'interprétation.

Dans l'exemple 30, c'est l'idée de *ramasser le bois* qui peut faire interpréter l'amorce comme une anticipation de *cueillette* :

(30) il y avait une période où les où les où les paysans avaient le droit d'aller dans les propriétés des forêts des seigneurs pour faire la *cue-* le bois pour ramasser le bois pour etc. (crfp, pricorl)

Dans l'exemple 31, l'amorce *vi-* occupe la place d'un nom et le contexte de *paysans* peut la faire interpréter comme une amorce de *village* :

(31) et la moitié du *vi-* des *pay-* de nos de de de nos *paysans* eh habitaient tantôt à Baste- à la plaine tantôt à la montagne (crfp, pricorl)

Mais le morphème peut apparaître dans le contexte antérieur et servir d'amorce pour un morphème qui sera corrigé ensuite. Ainsi dans les deux exemples suivants, les locuteurs prononcent la première syllabe d'un morphème produit précédemment, s'en avisent et corrigent le groupe syntagmatique en abandonnant l'amorce pour un autre morphème. Dans l'exemple 32, l'amorce *ci-*, renvoyant au nom *cimetière*, est corrigée par *clocher* et dans l'exemple 33, l'amorce *es-*, renvoyant au nom *escalier*, est corrigée par *chemins-là* :

¹¹ Nous conseillons l'usage du TLF (*Trésor de la Langue Française*) dont il existe une version informatisée.

Marie-Noëlle ROUBAUD

- (32) ce **cimetière** le premier le plus ancien a été inauguré en mille neuf cent trente + un peu à l'époque où on a construit le **ci-** le clo- le clocher de l'église (crfp, pricor1)
- (33) on a fait le chemin de l'Inca qui en fait qui est un **un escalier** plus tu descends tu montes qui est un escalier vraiment très tortueux abrupt + et quand on pense que les Incas se se se déplaçaient que par ces **es-** ces chemins-là chargés de de blé de grains de marchandises (crfp, prilem1)

4.3. La connaissance du lexique

L'identification de l'amorce suppose également de la part du transcripateur une connaissance du lexique : des fréquences et de son organisation (champs lexicaux, séries, lexèmes substituables).

4.3.1. Les fréquences

Si les amorces ont une forte probabilité d'apparition dans le syntagme, comme *même* (ex. 34), *plutôt* (ex. 35) ou *déjà* (ex. 36), le transcripateur les identifie facilement :

- (34) notre travail est aussi bien fait que si on avait constamment le patron + parce que **m-** en sachant qu'il y a quelqu'un derrière qui regarde ton travail + moi personnellement ça me s- ça m'énerve + c'est pas mon truc (crfp, pricle3)
- (35) il y a euh huit assistants de clientèle qui eux sont **plut-** plus spécialement chargés de du service clientèle aux guichets (crfp, progre1)
- (36) d'ailleurs certains qui avaient des exploitations d'arbres fruitiers se sont **déj-** reconvertis vers la vigne (crfp, prigap1)

4.3.2. Les champs lexicaux

Quand l'amorce est corrigée par un autre morphème dans le syntagme, le transcripateur peut établir un rapport de sens étroit entre les éléments syntagmatiques. Ainsi, dans l'exemple 37, le transcripateur peut identifier l'amorce *é-* comme un fragment de *édifices*, *établissements* ... en rapport de sens avec *monuments* et dans l'exemple 38, l'amorce *en-* comme un fragment de *entreprise* ... en rapport de sens avec *banque* :

- (37) bien sûr euh nous avons euh des monuments exceptionnels à voir en Bourgogne des **é-** des des **monuments** qui qui partent bien sûr euh du du roman (crfp, pridij1)
- (38) et euh + l'évolution à l'intérieur de la de l'**en-** de de la de la **banque** comment comment elle s'effectue (crfp, progre1)

4.3.3. Les séries

L'identification peut se faire grâce à la connaissance qu'a le transcripateur des séries lexicales, comme ici, celle des jours de la semaine :

- (39) et puis le **vendre-** le **samedi** quand je peux parce que avec Thomas à emmener au foot souvent je suis contrainte à jouer au taxi avec lui là je là c'est l'atelier (crfp, prilar1)

4.3.4. Les substitutions

Le transcripateur doit aussi connaître les lexèmes qui sont substituables, comme dans l'exemple suivant où *der-* est l'amorce probable de *dernière* car dans ce contexte, *dernière guerre* et *deuxième guerre* sont interchangeables. C'est au transcripateur que revient la charge d'interroger les locuteurs sur ces substitutions, afin de ne pas passer à côté de l'identification de certains morphèmes :

- (40) il produisait encore cinquante mille hectolitres à la fin de la **der-** de la **deuxième guerre mondiale** (crfp, progap1)

4.4. La connaissance des différentes perceptions

La perception du transcripateur n'est pas forcément celle des autres locuteurs. Ainsi un transcripateur du sud de la France qui a l'habitude de prononcer le *e* final, situé après consonne, des verbes du premier groupe, pourrait transcrire l'exemple suivant avec une amorce parce que le *e* n'est pas audible (ex. 41a) :

- (41a) et puis on le **veil-** on le veille jusqu'à minuit (crfp, pricor1)

Il est préférable de ne pas amputer la syllabe de son *e* muet et de conserver l'écriture complète du morphème (ex. 41b) :

- (41b) et puis on le **veille** on le veille jusqu'à minuit

4.5. La connaissance des prononciations

Le transcripateur doit aussi se méfier de sa prononciation. Nous savons combien il est difficile de savoir comment se prononcent certains noms terminés par *s*, *l*, *z* ou *x*. Doit-on dire pour *cassis* : /kasis/ ou /kasi/ ? Le dictionnaire indique que la consonne finale est prononcée alors que pour *persil*, elle ne l'est pas (/pErsi/). Que faire alors devant l'exemple suivant, transcrit ainsi ?

Marie-Noëlle ROUBAUD

(42a) euh hum l'Aligoté aussi a un vin qui + qui aura euh + un un aven- un a- un avenir certain dans dans des plats euh surtout dans des entrées ou des apéros euh comme le **cassi-** avec le **cassis** il s'accorde très bien avec le **cassis** (crfp, pubdij1)

Doit-on garder l'amorce (ex. 42a) puisque le locuteur prononce /kasis/ deux fois par la suite ou doit-on rétablir l'écriture du morphème, tout en indiquant la particularité de prononciation à un autre endroit (ex. 42b)¹² ?

(42b) euh hum l'Aligoté aussi a un vin qui + qui aura euh + un un aven- un a- un avenir certain dans dans des plats euh surtout dans des entrées ou des apéros euh comme le **cassis** avec le **cassis** il s'accorde très bien avec le **cassis**

La prise en compte de ces outils a permis de corriger certaines erreurs de transcription, comme nous allons le voir, mais a soulevé d'autres problèmes.

5. Les problèmes rencontrés dans les corpus concernant les amorces

5.1. Quelle orthographe ?

La prise en compte du syntagme dans lequel est inscrite l'amorce nous a amené à corriger l'orthographe de certaines amorces, la mise en grille visualisant l'amorce et le syntagme qui lui correspond :

(43) c'est sûr que il y a un impact sur le terrain bon l'**in-** l'**in-** l'impact ce sont les le le premier impact ce sont les rejets directs (crfp, procor1)
(ex. 43 corrigé) c'est sûr que il y a un impact sur le terrain bon l'**im-** l'**im-** l'impact ce sont les le le premier impact ce sont les rejets directs

G : bon l'**im-**
 l'**im-**
 l'impact ce sont ...

En 44, la prise en compte du contexte montre que l'amorce n'est qu'un fragment du nom propre qui suit, d'où la majuscule dans l'exemple corrigé :

¹² Il n'est pas rare qu'un même locuteur ait les deux prononciations : /kasi/ et /kasis/ ou /pErsi/ et /pErsil/.

Du bon usage des amorces dans la transcription des corpus

- (44) c'est-à-dire + aient une organisation très régionalisée comme l'a
l'Allemagne bien entendu depuis longtemps (crfp, prigap2)
(ex. 44 corrigé) comme l'A- l'Allemagne bien entendu
G : aient une organisation très régionalisée comme l'A-
l'Allemagne ...

Dans l'exemple qui suit, on peut penser que *con-* est l'amorce de *compétences*, occupant la même place syntaxique :

- (45) il faut qu'il sorte de l'univer- qu'il ait été formé ici c'est normal + euh
euh donc après et après qu'il ait eu une formation corse + sur les **con-**
à à compétences égales (crfp, pubcor1)
(ex. 45 corrigé) il faut qu'il sorte de l'univer- qu'il ait été formé ici c'est
normal + euh euh donc après et après qu'il ait eu une
formation corse + sur les **com-** à à compétences égales
G : et après qu'il ait eu une formation corse sur les **com-**
à
à compétences égales

Mais nous nous rendons compte, au travers de ce dernier exemple, qu'il n'y a pas toujours de certitude mais seulement une probabilité de solution reposant sur des critères formels.

5.2. Amorce ou morphème ?

L'analyse en syntagmes a permis de corriger des amorces qui n'en étaient pas. Ainsi dans l'exemple qui suit, il s'agit d'un cas banal de répétition du déterminant. Pallaud et Henry (dans ce numéro) montrent que les déterminants sont majoritairement concernés par la répétition (dans plus de 80% des cas) :

- (46) hein des + enfin parlons pas des grands des des grands crus mais +
c'est l'équivalent de + à quoi peut-on le comparer dans l'arrière pays de
Montpellier les les Costières les non pas les Costières le + zut le nom
m'échappe (...) enfin des + **dé- dé-** des produits sur ces coteaux qu'il y
a derrière Montpellier là (crfp, prigap1)
(ex. 46 corrigé) enfin des + **des des** des produits sur ces coteaux qu'il y a
derrière Montpellier là

Marie-Noëlle ROUBAUD

G: hein des
[...]
enfin des
des
des
des produits sur ces coteaux qu'il y a derrière Montpellier là

Le fait que le locuteur revienne, après la production de l'amorce, au départ du groupe syntagmatique aide à l'interprétation et à la correction des exemples qui suivent.

Dans l'exemple 47, le verbe *avoir* est employé comme verbe générique, occupant la place du lexème verbal qui sera produit ensuite¹³ (*qui a / qui existe*) :

(47) quand euh quelqu'un mourait euh au village ou à la plaine eh il fallait l'enterrer alors d'abord il y avait déjà la confrérie la confrérie qui **a** qui existe depuis des temps immémoriaux qui prenait en charge le mort (crfp, pricor1)

(ex. 47 corrigé) il y avait déjà la confrérie la confrérie qui **a** qui existe depuis des temps immémoriaux

G : il y avait déjà la confrérie la confrérie qui **a**
qui existe depuis des temps ...

Dans l'exemple 48, il est probable que le locuteur reprend le syntagme verbal en remplaçant *falloir* par un synonyme *il y a ... + infinitif* :

(48) ils mettent en place pour les enfants euh bon des activités randonnée(s) des activités escalade(s) bon et puis **fo-** il y a les aménagements à proprement dit hein à à surveiller et à à et à entretenir sur les sites même (crfp, procor1)

(ex. 48 corrigé) bon et puis **faut** il y a les aménagements à proprement dit hein à hein à à surveiller et à à et à entretenir sur les sites même

G : bon et puis **faut**
il y a les aménagements à proprement dit hein à
à surveiller
et à
à
et à entretenir
...

Dans l'exemple 49, le locuteur amorce une question en *qu'est-ce que* puis produit une sorte d'énoncé parenthétique avant de reprendre le syntagme en *qu'est-ce que*, comme le montre la mise en grille :

¹³ Ces verbes génériques ont été appelés « verbes d'approximation » (cf. Roubaud, 1987).

(49) et et dans le domaine du travail **què-** vous qui avez connu disons en tant que fils de vigneron les les la manière de travailler il y a cinquante ans et qui la connaissez maintenant qu'est-ce que vous + donneriez comme différence principale (crfp, prigap1)

(ex. 49 corrigé) et et dans le domaine du travail **qu'est-** vous qui avez connu disons en tant que fils de vigneron les les la manière de travailler il y a cinquante ans et qui la connaissez maintenant qu'est-ce que vous + donneriez comme différence principale

G : dans le domaine du travail **qu'est-**
[vous qui avez connu disons en tant que fils de vigneron les les la manière de travailler il y a cinquante ans et qui la connaissez maintenant]
qu'est-ce que vous donneriez comme ...

5.3. Cas de double-transcription

Mais il nous a aussi parfois été impossible de décider si nous nous trouvions face à un morphème ou une amorce parce que nous hésitions sur le départ du groupe syntagmatique. Dans ce cas nous avons opté pour une double-transcription : l'une avec un morphème complet et l'autre avec seulement l'amorce d'un morphème, les deux éléments étant placés entre barres obliques.

Tout compte de nombreux allomorphes, aussi peut-on se demander si dans l'exemple transcrit (ex. 50a), il s'agit bien de l'amorce de *toute*, lexème produit un peu plus loin ou s'il s'agit du morphème *tous*, réalisé dans le syntagme *tous les occitans* (ex. 50b) :

(50a) au cirque des corbeaux il y avait Marti il y avait **tou-** il y avait **toute** la + tous les occitans qui chantaient là il y avait vraiment la liaison sur le le Larzac (crfp, prigap2)

(50b) au cirque des corbeaux il y avait Marti il y avait **tous** il y avait toute la + **tous** les occitans

La solution, pour le transcripteur, est de présenter une double-transcription (ex. 50c) :

(50c) au cirque des corbeaux il y avait Marti il y avait /**tou-**, **tous**/ il y avait toute la + tous les occitans

Que décider pour l'exemple 51 ? Le locuteur produit-il le verbe *avoir* au présent, comme l'indique la transcription (ex. 51 a) pour se corriger ensuite en *avait* ou a-t-on une amorce de *avait* : *si on t'a-* / *si on t'avait* (ex. 51b) ?

Marie-Noëlle ROUBAUD

(51a) mais pourquoi et si si on t'a et si on t'a si on t'avait laissé le choix un jour qu'on t'avait dit si tu veux tu peux prendre français euh pour euh pour rehausser un peu tu aurais pris français (crfp, pubcor1)

(51b) mais pourquoi si si on t'a- et si on t'a- si on t'avait laissé le choix un jour qu'on t'avait dit ...

Encore une fois, la double-transcription laisse la place aux deux interprétations (ex. 51c) :

(51c) mais pourquoi si si on t/a, a-/ et si on t/a, a-/ si on t'avait laissé le choix un jour qu'on t'avait dit ...

Dans l'exemple (52), le transcripteur peut hésiter entre la production d'un adverbe dans la réaction du verbe *faire manger* ou l'amorce du syntagme en *il fallait*, les deux mises en grille (G1 et G2) visualiseront ces deux possibilités :

(52) et les hommes sont dans les pièces à côté ou chez les voisins s'il n'y a pas assez de place et alors ceux-là /t-, il/ on les fait manger comme si c'étaient des gens qui étaient arrivés de très loin et qu'il fallait nourrir pour qu'ils puissent euh passer la nuit (crfp, pricor1)

G1 alors ceux-là i- (immédiatement) on les fait manger comme si ...

ou

G2 alors ceux-là **il**
et qu' **il** fallait nourrir pour qu'ils puissent euh passer ...

Dans l'exemple 53, le transcripteur peut hésiter entre la production de l'indéfini *tout* comme élément de la liste des syntagmes infinitifs (G1) ou comme adverbe portant sur le verbe *répondre* (G2) :

(53) les six conseillers commerciaux sont là essentiellement pour provoquer les contacts avec les clients répondre à leurs attentes + /**tout, tou-**/ euh leur proposer des placements d'argent (crfp, progrep1)

G1 les six conseillers commerciaux sont là essentiellement
pour provoquer les contacts avec les clients
répondre à leurs attentes
tout
leur proposer des placements d'argent

ou

G2 les six conseillers commerciaux sont là essentiellement
pour provoquer les contacts avec les clients
répondre à leurs attentes **tou-** (toujours)
leur proposer des placements d'argent

Dans l'exemple 54, le transcripteur peut hésiter entre la retouche de la préposition dans le syntagme prépositionnel à *ce moment-là* (G1) ou envisager une mise en mémoire (plus longue que celle que nous avons vue

dans les exemples précédents) du lexème *octobre* produit un peu plus loin (G2) :

(54) les maladies cryptogamiques quoi + bon là c'est environ un traitement tous les quinze jours à peu près + pendant bon + surtout le mois de juin et juillet (...) c'est le plus important le plus important c'est /**au**, **o-**/ à ce moment là que ça craint le plus (...) et puis on arrive aux vendanges vers le mois de fin octobre (crfp, prigap1)

G1 le plus important c'est **au**
à ce moment là que ça craint le plus (...)

ou

G2 le plus important c'est **o-** à ce moment là que ça craint le plus (...)
et puis on arrive aux vendanges vers le mois de fin **octobre**

5.4. Cas où l'identification est plus difficile

Mais il arrive que le transcripateur ait beaucoup de mal à identifier l'amorce pour différentes raisons.

5.4.1. Amorce isolée

Dans certains cas, le syntagme contenant l'amorce n'est pas repris par le locuteur si bien qu'il est difficile d'interpréter quoi que ce soit, c'est le cas de *sans le so-* (ex. 55) :

(55) maintenant pour la langue bon il y a quelques cours par ci par là mais encore une fois + sans le **so-** c'est c'est in- c'est insuffisant (crfp, prigap2)

5.4.2. Discours difficile à percevoir

Si le discours est coupé par les interventions d'un interlocuteur (noté L1), par des chevauchements de paroles (indiquées entre chevrons>, la tâche du transcripateur est encore plus difficile, c'est ce qu'illustre la transcription de l'amorce *contè-* dans cet extrait :

(56) L2 j'ai quand mê- j'ai enregistré personnellement des contes j'ai
L1 ah < d'accord
L2 j'ai > au moins deux cassettes
L1 < ah bé oui
L2 de **contè-** > au- conteuses aujourd'hui /deux C.D., décédées/
L1 ah c'est bien
L2 tout à fait d'un d'un bon niveau sur les contes traditionnels en
terme de de ça il y en a

Marie-Noëlle ROUBAUD

L1 ça ça devrait être édité tu penses pas
L2 à ma connaissance il y en a peu de il y en a peu >
L1 l'éditer

Si l'on se réfère au contexte et à la place occupée par cette amorce, il est plus probable d'envisager deux morphèmes à la place de l'amorce :

(ex. 56 corrigé) j'ai enregistré personnellement des contes j'ai (...) j'ai au moins deux cassettes (...) de **contes et au-** conteuses aujourd'hui /deux C.D., décédées/

De la même façon, dans l'exemple qui suit, il nous a semblé possible d'envisager autre chose que ce qui était proposé, l'amorce *almal-* (ex. 57a) suggérant que le locuteur commence le syntagme *les âmes damnées* par une amorce anticipant sur la voyelle « A » et la consonne « L », comme le visualise la mise en grille :

(57a) derrière le vieux couvent il y avait + trois euh nécropoles si tu veux + euh euh il y avait une nef pour euh pour les + centrale pour les religieux + ceux qui étaient très proches de l'église une à l'est c'était pour les chrétiens et puis celle de gauche c'était pour les les **almal-** les âmes damnées les gens qui étaient morts euh en mal de religion etc. (crfp, pricor1)

G : et puis celle de gauche c'était pour les
les **almal-**
les âmes damnées
les gens qui étaient morts ...

Mais ne peut-on pas envisager la production de deux parties de syntagmes (ex. 57b), celui contenant le mot *âmes* avec anticipation d'un /l/ à la place d'un /m/ et celui contenant l'adverbe *mal*, anticipant sur le syntagme à venir *en mal de religion* ? Nous aurions alors, à la place d'une seule amorce, un fragment de morphème et un morphème. Cette possibilité d'interprétation sera visualisée dans la grille sous l'exemple :

(57b) et puis celle de gauche c'était pour les les **al-** **mal** les âmes damnées les gens qui étaient morts euh en mal de religion etc.

G : c'était pour les
les **al-** **mal**
les âmes **damnées**
les gens qui étaient morts euh en **mal** de religion

5.4.3. Discours spécialisé

Dans les corpus, l'identification des morphèmes et donc des amorces, peut être rendue difficile parce que le transcritteur se trouve face à un discours de spécialiste, qu'il ne connaît pas forcément et dont le lexique

spécialisé sera absent de son dictionnaire. Si le locuteur explique les termes qu'il emploie, le transcripteur n'a plus aucun problème pour identifier les morphèmes ; c'est le cas de *l'ouvrée* dans cet exemple, dont la définition est donnée dans le syntagme en *c'est-à-dire que* :

- (58) hein nous on compte à **l'ouvrée** + c'est-à-dire que dans un hectare il y a vingt-quatre ouvrées c'est quatre cent vingt-huit mètres carrés exactement l'ouvrée quand on vend des vignes en Bourgogne eh bien on parle de on parle de l'ouvrée + euh quatre cent vingt-huit mètres carrés donc à l'hectare au je pense que ça c'est facile à se rappeler (...) + on vend à l'ouvrée alors l'ouvrée eh bien pourquoi parce que c'est ce que faisait dans la journée euh il y a une centaine d'années ou peut-être même moins euh à la main + un un ouvrier qui travaille dans à la tâche dans dans la vigne alors voilà c'est pour ça c'est le travail d'une journée hein le l'ouvrée de l- de l- l'ouvrage de la de la de la journée + (crfp, pridij1)

Mais il arrive aussi que le corpus ne renseigne pas le transcripteur. En effet, à quoi correspond l'amorce *éti-* de ce dernier exemple ?

- (59a) L2 et puis vers la f- au quinze juin à peu près c'est ce qu'on appelle la floraison c'est-à-dire au ça se passe très rapidement cette floraison ça se passe très rapidement + il f- il faut qu'elle se passe très rapidement pour que chaque chaque grain devienne un raisin quoi si-
- L1 hum
- L2 sinon bon c'est ce qu'on appelle la coulure s'il pleut ou que c'est /dans, la/
- L1 oui
- L2 la pollinisation si le pollen est mouillé **éti-** + il vole pas de d'un grain à l'autre
- L1 oui c'est
- L2 c'est-à-dire qu'il part dans l'air et puis + ça ça pollinise pas quoi (crfp, prigap1)

La transcription (ex.59a) ne livre qu'une interprétation, difficile d'ailleurs pour des non-spécialistes de la vigne, le locuteur essayant d'expliquer que la coulure est un accident qui empêche la fécondation de la fleur et qui fait le plus souvent couler le pollen. Transcrite de cette manière, l'amorce peut s'interpréter comme étant celle d'un adjectif (comme *étiolé*), produit sur la même place syntaxique que *mouillé* dans le syntagme *si le pollen est mouillé*, avec une mise en grille de ce type :

G : si le pollen est mouillé
éti- il vole pas de d'un grain à l'autre

Mais ne peut-on pas envisager une autre transcription, qui ferait apparaître une coordination de deux adjectifs dans la valence de *être* : par

Marie-Noëlle ROUBAUD

exemple : *si le pollen est mouillé et tiède (ou mouillé et tigré ou mouillé et tiqueté ...)* ? La mise en grille serait alors différente ainsi que la transcription (ex. 59b) :

(59b) L2 la pollinisation si le pollen est mouillé et ti- + il vole pas de d'un grain à l'autre

G si le pollen est mouillé
et ti- il vole pas de d'un grain à l'autre

Il aurait pu s'agir également d'un *et qu'il* mal perçu par un locuteur du nord de la France (ex. 59c)¹⁴ et nous n'aurions alors aucune amorce :

(59c) L2 : la pollinisation si le pollen est mouillé et qu'il + il vole pas de d'un grain à l'autre

Comme l'écrivent Blanche-Benveniste & Jeanjean (1987 : 112) :

« Dans toutes les écoutes, aussi méticuleuses qu'elles soient, il y a toujours une part de flottement et d'approximation. »

6. Conclusion

Les amorces ne sont pas seulement des traces d'oralité, elles sont également riches d'informations car elles nous renseignent

- sur le plan syntaxique : sur la composition de la syntaxe avec les allées et venues qu'effectue le locuteur sur l'axe syntagmatique et les recherches paradigmatiques qu'il opère (quels lexèmes ? quelle place ? quelle correction ?) ;
- sur le plan sémantique : sur la progression du sens (le locuteur anticipant le plus souvent sur le morphème à venir), sur ce que veut dire le locuteur ;
- et sur le plan énonciatif : sur les phénomènes de mise en mémoire, de savoirs partagés ou non.

La tâche du transcripteur pour identifier les amorces est complexe. Il doit non seulement se référer à des données disponibles (dictionnaires, contexte) mais aussi se construire ses propres outils (connaissance du lexique, des prononciations, des perceptions) et se documenter dans le cas de morphèmes inconnus ; la connaissance d'un transcripteur isolé ne suffit pas. Transcrire un corpus nécessite d'en effectuer une analyse syntaxique.

¹⁴ Je remercie José Deulofeu pour cette remarque.

Une première transcription, même faite par un transcripteur averti, ne donne qu'une première analyse du texte en syntagmes, qu'une première interprétation. Coppieters (1997 : 35) indiquait déjà que l'interprétation d'un corpus demeurerait toujours incomplète et donc en attente d'interprétation. Le transcripteur doit ensuite affiner la transcription au niveau des syntagmes et des morphèmes en revenant sur les amorces signalées, en réécoutant les passages difficilement interprétables (ce qui augmente encore le temps qu'il faut pour transcrire une minute de parole)¹⁵.

Une bonne transcription ne peut se faire qu'à ce prix car nettoyer les amorces des textes serait perdre toutes ces précieuses informations et conduirait à produire un autre texte, qui risquerait de ressembler à un texte écrit mais qui, en fait, aurait peu de chose à voir avec le discours du locuteur.

Références

- Bilger, M., Blasco, M., Cappeau, P., Pallaud, B., Sabio, F., & Savelli, M.-J. (1997). Transcription de l'oral et interprétation ; illustration de quelques difficultés. *Recherches sur le français parlé*, 14, 57-86.
- Blanche-Benveniste, C. (1997). *Approches de la langue parlée en français*. Paris : Ophrys.
- Blanche-Benveniste, C., & Jeanjean, C. (1987). *Le français parlé. Transcription et édition*. Paris : Didier Érudition.
- Blanche-Benveniste, C., Bilger, M., Rouget, C., & van Den Eynde, K. (1990). *Le français parlé. Études grammaticales*. Paris : Édition du CNRS.
- Blanche-Benveniste, C., & Pallaud, B. (2001). Le recueil d'énoncés d'enfants : enregistrements et transcriptions. *Recherches sur le français parlé*, 16, 11-37.
- Cappeau, P. (1997). Données erronées : quelles erreurs commettent les transcripteurs ? *Recherches sur le français parlé*, 14, 117-126.
- Coppieters, R. (1997). Quelques réflexions sur la question des données : corpus et intuitions. *Recherches sur le français parlé*, 14, 21-41.
- Giovannoni, D.-C., & Savelli, M.-J. (1990) Transcrire et orthographier le français parlé. De l'impossible copie à la falsification des données orales. *Recherches sur le français parlé*, 10, 19-37.
- Jeanjean, C. (1984). Les ratés c'est fa- fabuleux. Étude syntaxique des amorces et répétitions. *Syntaxe et discours LINX*, 10, 171-177.
- Pallaud, B. (1999). Lapsus et phénomènes voisins dans la langue parlée : problèmes d'identification. *Recherches sur le français parlé*, 15, 1-33.
- Pallaud, B. (2002). Les amorces de mots comme faits autonymiques en langage oral. *Recherches sur le français parlé*, 17, 79-101.

¹⁵ L'écoute du corpus avec un casque facilite la perception des amorces.

Marie-Noëlle ROUBAUD

- Roubaud, M.-N.(1987). *Les mystères de l'approximation lexicale*. Mémoire de DEA. Sciences du Langage, Université de Provence.
- Roubaud, M.-N., & Loufrani, C. (1999). Éclairage des corpus de type aphasique par les grilles. *Recherches sur le français parlé*, 15, 41-57.
- Roubaud, M.-N. (2003). Du bon usage des notations phonétiques dans les corpus aphasiques. *Recherches sur le français parlé*, 18.

Étude des interactions entre pauses silencieuses et pauses remplies en français parlé

Estelle CAMPIONE
Université de Provence

Résumé

La terminologie classique regroupe sous le terme de *pauses* les *pauses silencieuses* et les *pauses remplies* (en français l'item quasi-lexical *euh*), tandis qu'elle met dans une catégorie séparée les allongements d'hésitation. Ces différents phénomènes sont très souvent associés, et l'étude de leur interaction a jusqu'ici fait l'objet de peu de travaux. Nous montrons ici, grâce à l'étude d'une heure d'enregistrements tirés du *Corpus de référence du français parlé*, qu'en fait pauses silencieuses et pauses remplies se distinguent totalement dans leur fonctionnement, et que la pause silencieuse, contrairement à l'idée reçue, n'a en elle-même aucun rôle de marque d'hésitation ou de travail de formulation. Elle n'intervient dans cette fonction qu'associée à d'autres marques, principalement allongement syllabique et item quasi-lexical *euh*. Ces deux marques d'hésitation ont d'ailleurs des propriétés acoustiques et articulatoires proches, et nous montrons que leur distribution et leur fonctionnement sont analogues.

1. Introduction

On distingue de façon traditionnelle deux sortes de pauses (cf. par exemple Goldman-Eisler, 1972 ; Duez, 1982) : les pauses *silencieuses*, dans lesquelles toute production vocale s'interrompt, à l'exception éventuellement de bruits respiratoires, et les pauses *remplies* (ou *sonores*), constituées d'un item quasi-lexical (*euh* en français, *er/erm* en anglais).

Regrouper ces phénomènes acoustiquement et articulatoirement très différents sous le même terme de *pause* revient à poser une hypothèse implicite quant à une identité de fonction. Or, rien n'est moins sûr qu'une telle identité. On a noté depuis bien longtemps (Boomer, 1965) que les pauses silencieuses ont un double rôle. Certaines pauses silencieuses sont *démarcatives*, et apparaissent à la jonction de segments du discours, à la structuration desquels elles participent. D'autres sont des pauses *d'hésitation*, et sont le reflet de difficultés que rencontre ponctuellement le locuteur dans ses opérations mentales de « recherche et d'encodage » (Barik, 1968) liées à la production du discours. À l'inverse, il semble que les pauses remplies soient cantonnées exclusivement à ce deuxième rôle : elles constituent en quelque sorte un signal conventionnel de la part du locuteur, lui permettant d'occuper le terrain de l'interaction et d'éviter d'être interrompu pendant le laps de temps nécessaire à la construction de la suite de son énoncé (cf. Clark & Clark, 1977 ; Levinson, 1983 ; Finegan, 1994 ; Stenstrom, 1994 ; Wennerstrom, 1994 ; Rose, 1998).

Il est intéressant de remarquer que le rôle de signal d'hésitation correspondant aux pauses remplies est également dévolu à certains allongements syllabiques (qui affectent généralement une voyelle en fin de mot), que nous appellerons *allongements d'hésitation*. On a déjà noté que leurs propriétés sont proches de celles des pauses remplies, au point que les mêmes algorithmes permettent de détecter les deux (cf. Goto, Itou, & Hayamizu, 1999). Morel & Danon-Boileau (1998) les regroupent sous l'appellation de « marques du travail de formulation ». Toutefois, ce nouveau terme concerne alors aussi, nécessairement, d'autres phénomènes liés à la difficulté du locuteur dans la production de son discours, tels que les répétitions ou les amorces (mots inachevés).

Il existe dans la littérature un grand nombre d'études sur les pauses silencieuses (voir par exemple l'état de l'art de Zellner, 1998). Par contre, comme le montre une thèse récente (Candéa, 2000), un nombre beaucoup plus restreint de travaux se sont intéressés aux marques du travail de formulation en général, aux pauses remplies et aux allongements en particulier (sans doute ceci est-il dû au fait, comme le font remarquer Cutler, 1998, Duez, 1998, etc., que la majorité des travaux phonétiques ont été consacrés à la « parole de laboratoire », au détriment de l'oral spontané). Il n'existe en tout cas presque pas de travaux sur l'interaction entre ces différents phénomènes. Or, comme nous le montrons ci-après, ils sont très souvent associés.

Nous proposons dans cet article de fournir une étude précise des interactions entre pauses silencieuses, pauses remplies et allongements

d'hésitation, à partir d'un échantillon du *Corpus de référence de français parlé* composé de 8 500 mots et 54 minutes de parole, faisant intervenir 10 locuteurs différents. Nous montrons que les pauses silencieuses, hormis les pauses très brèves, n'ont jamais à elles seules un rôle de pause d'hésitation, et que pour assurer une telle fonction, elles sont presque obligatoirement associées à une autre marque du travail de formulation, typiquement un allongement ou un *euh* (ou une combinaison de ces éléments).

La terminologie établie regroupe donc des phénomènes acoustiquement et fonctionnellement très différents sous le même terme de *pause*, alors qu'elle met dans des catégories séparées les allongements d'hésitation et les *euh* malgré leur similitude de propriétés et de fonctionnement. Nous regrouperons ci-après, comme le font Goto et al. (1999), allongements d'hésitation et *euh* sous le même terme de *pauses remplies*.

2. Corpus

Le corpus que nous avons utilisé est un sous-ensemble du *Corpus de référence du français parlé*. Nous avons sélectionné dix locuteurs (cinq femmes et cinq hommes) et pris pour chacun environ cinq minutes de parole au cours desquelles le locuteur principal n'était pas interrompu (Tableau 1).

Locuteur	Sexe	Mots	Durée (min)
<i>pricle1</i>	f	790	5,18
<i>pricle3</i>	f	1063	5,29
<i>prolyo1</i>	f	1003	5,23
<i>primon2</i>	f	1189	5,34
<i>pubstr1</i>	f	780	6,03
<i>pricle2</i>	h	650	5,20
<i>procor1</i>	h	883	5,19
<i>pridij1</i>	h	806	5,04
<i>pubpaul</i>	h	542	5,04
<i>propoil</i>	h	790	5,29
Total		8496	54,31

Tableau 1. Répartition des différents types de pauses silencieuses

Les conventions de transcription sont celles utilisées par l'équipe DELIC, c'est-à-dire une transcription orthographique sans marques de

Estelle CAMPIONE

punctuation, mais avec transcription des pauses silencieuses et de certaines marques du travail de formulation (*euh* d'hésitations, amorces et répétitions) (cf. Blanche-Benveniste & JeanJean, 1987, voir également l'article de présentation du crfp dans ce volume). Nous avons complété les conventions sur deux points pour cette étude :

- une catégorisation des pauses silencieuses en trois niveaux de longueur (représentés dans les corpus au moyen des symboles ^, +, ++);
- un marquage des allongements syllabiques d'hésitation (représenté par le symbole :).

3. Pauses silencieuses

3.1. Annotation

Nous avons remarqué, au cours d'une étude préliminaire portant sur des transcriptions produites par plusieurs annotateurs expérimentés de notre équipe, qu'environ 2/3 seulement des pauses silencieuses effectivement produites sont en fait reportées dans la transcription. C'est dire la difficulté extrême que présente cette tâche lorsqu'elle est accomplie de façon entièrement manuelle¹. Le système perceptif semble « gommer » les pauses après qu'elles ont rempli leur rôle éventuel dans l'interprétation de l'énoncé, particulièrement lorsqu'elles sont combinées à des phénomènes d'hésitation ou de répétition.

Dans une expérience de perception impliquant 30 auditeurs, Candéa (2000 : 123) remarque d'ailleurs que les pauses silencieuses sont perçues en moyenne dans 63% des cas seulement. Elle note également des différences importantes selon que les pauses apparaissent seules ou en combinaison avec des marques du travail de formulation (hésitation,

¹ Il s'agissait ici de transcriptions réalisées de façon classique, c'est-à-dire à l'aide d'un magnétophone et sans logiciel particulier. À la suite de notre étude, les enregistrements composant le *Corpus de référence du français parlé* ont été vérifiés et alignés avec la transcription à l'aide du logiciel *Transcriber* (Barras et al., 1998), et il est apparu que ce mode de transcription « assisté » permettait une bien meilleure retranscription des pauses. Le taux global de transcription des pauses dans le *Corpus de référence du français parlé* est donc certainement plus élevé que celui que nous avons relevé dans notre étude préliminaire.

amorces, etc.), puisque le taux chute de 72,5% à 47,4% dans ce deuxième cas. Elle détecte aussi 15% de cas de « pauses subjectives », c'est-à-dire de positions qui sont marquées comme des pauses par les auditeurs, sans qu'il y ait de pause réelle dans les enregistrements.

Nous n'avons trouvé, en ce qui nous concerne, aucun cas de pause subjective dans les corpus transcrits que nous avons analysés : à l'inverse des sujets utilisés par Candéa (étudiants non linguistes), les transcrip-teurs de nos corpus sont des « professionnels » de la transcription, et ont vraisemblablement acquis une capacité à rejeter les fausses pauses².

Nous avons détecté de façon automatique la totalité des pauses silencieuses, au moyen d'un programme qui calcule la fréquence fondamentale, et repère l'absence de voisement pendant une période donnée. Nous avons appliqué un seuil de 200 ms, en accord avec les études ultérieures (Candéa, 2000), sachant que les pauses plus courtes, dont l'existence et l'importance ont été soulignées dans la littérature (Hieke, Kowal & O'Connell, 1983), seraient ajoutées manuellement.

Nous avons corrigé la totalité des pauses silencieuses à l'aide de l'éditeur de signal MES (Espesser, 1996) : les pauses détectées à tort (correspondant généralement à des occlusives sourdes) ont été supprimées, celles qui n'ont pas été détectées ont été ajoutées (y compris en dessous du seuil initial de 200 ms), et les frontières de celles qui ont été bien repérées ont été ajustées si besoin. Le corpus comportait 1375 candidats-pauses détectés et 1163 pauses effectives après correction.

3.2. Catégorisation des pauses

Une étude statistique (Campione, 2001 : 240) nous ayant montré qu'il y avait de bonnes raisons de catégoriser les pauses silencieuses en pauses brèves (< 200 ms, notées ^), moyennes (200-1000 ms, notées +) et longues (> 1000 ms, notées ++), nous avons ajouté ce niveau de notation, qui n'apparaît pas normalement dans les conventions de l'équipe :

Je vais leur vendre un produit ^ c'est clair (primon1)
euh + vendre quelque chose c'est bien + surtout si c'est un produit de
qualité + mais le vendre av- + dans un climat de confiance c'est beaucoup
plus difficile (primon1)

² Maria Candéa utilise un seuil de 200 ms. Elle nous fait remarquer (communication personnelle) qu'il est possible que les pauses qu'elle considère comme « subjectives » soient en fait des pauses réelles, mais de durée inférieure à ce seuil.

Estelle CAMPIONE

c'était un monsieur *P* + nouveau promu à la faculté des Lettres de Clermont ++ que j'ai eu la chance de rencontrer plusieurs fois lorsque je revenais à Clermont (pricle2)

3.3. Pauses démarcatives ou non démarcatives

Il semble largement accepté par les psycholinguistes que la production de la parole relève d'un cycle planification-exécution, qui résulte en une succession d'unités relativement brèves (que nous appellerons ci-après *segments discursifs*), séparées par des pauses silencieuses (Fromkin, 1971 ; Goldman-Eisler, 1972 ; Garrett, 1975 ; Dell, 1988 ; Kempen & Hoenkamp, 1987 ; Levelt, 1989 ; etc.). Les pauses silencieuses sont nécessaires à la fois pour la planification des énoncés par les locuteurs, et le traitement par les auditeurs. Au contraire des pauses d'hésitation, ces pauses démarcatives ont donc un rôle important dans la structuration du discours.

Nous avons marqué dans le corpus toutes les pauses silencieuses de nature démarcative. Ces pauses sont assez faciles à repérer car de multiples indices convergent (intonation, qualité et durée vocalique, syntaxe, etc.), qui permettent de les différencier des pauses silencieuses d'hésitation. L'exemple ci-dessous montre le type de découpage obtenu (nous avons ajouté des flèches montantes ou descendantes qui indiquent les mouvements intonatifs majeurs, ainsi qu'un astérisque qui indique les accents) :

ben je travaille dans un pressing ↗
++
on fait pas que* le pressing ↗ on fait aussi la blanchisserie ↗
+
plus la blanchisserie d'ailleurs ↘
+
les draps les nappes la restauration ↗
++
(pricle3)

Le Tableau 2 donne la répartition des différents types de pauses silencieuses.

Étude des interactions entre pauses silencieuses et pauses remplies en français parlé

Longueur	Démarcatives	Non-démarcatives	Total
brèves	0	24	24
moyennes	663 (69%)	303 (31%)	966
longues	148 (86%)	25 (14%)	173
Total	811 (70%)	352 (30%)	1163

Tableau 2. Répartition des différents types de pauses silencieuses.

On constate que les pauses silencieuses sont démarcatives dans 70% des cas.

Les pauses brèves ne sont jamais démarcatives, ce qui semble justifier les seuils de l'ordre de 200 ms utilisés dans certaines études. Nous rejoignons la position de Morel & Danon-Boileau (1998) pour lesquels les pauses en dessous de ce seuil n'auraient pas de valeur iconique définie. Elles sont souvent de nature respiratoire et interviennent généralement en des points où la cohésion intonative est indiscutable :

donc c'est ^ la calandre qui travaille aussi (pricle3)
c'est même pas la peine ^ même de l'eau c'est pas bon (pricle3)
il y en a qui n'accepteront pas de le faire ^ parce qu'elles se sentent remet-
remises en question (prolyo1)
donc + euh les gens ^ travaillaient + euh sur le terrain (procor1)

4. Pauses remplies

4.1. Annotation

Allongements d'hésitation et *euh* se caractérisent par une voyelle continue de durée très supérieure à la normale, de qualité vocalique constante et sont associées à une fréquence fondamentale (F_0) plate ou très légèrement descendante (Guaitella, 1991) comme l'illustrent les Figures 1 et 2. Ces caractéristiques semblent être communes à de nombreuses langues (Quimbo, Kawahara & Doshita, 1998 ; Goto et al., 1999).

Estelle CAMPIONE

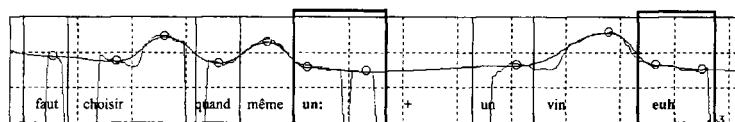


Figure 1. Patron mélodique des allongements syllabiques d'hésitation.

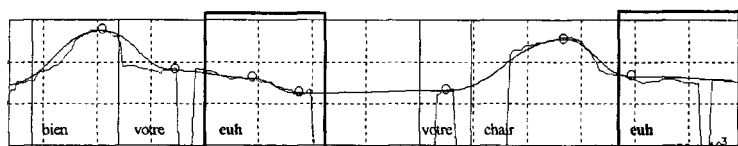


Figure 2. Patron mélodique d'un *euh*.

À part la longueur de la tâche, le marquage des pauses remplies dans les corpus présente donc assez peu de difficultés, sous réserve d'une écoute très soignée (on a tendance à ne pas les percevoir, comme pour les pauses silencieuses, voir Pallaud, 2003). Les allongements d'hésitation se distinguent clairement des allongements liés à la structure syntaxique, aux accents, etc., qui ont généralement, du moins en français, un contour intonatif ascendant ou descendant (cf. Vaissière, 1991, 1997). De plus, les allongements d'hésitation portent très souvent sur des mots-outils, ou se situent à des positions qui ne constituent pas des points de rupture syntaxique ou discursive. Les *euh* sont des items quasi-lexicaux facilement reconnaissables, et il n'y a pratiquement jamais désaccord entre annotateurs sur leur présence. Seuls quelques cas sont difficiles à trancher sur des mots terminant par un schwa, et pour lesquels on peut hésiter entre un allongement du schwa ou un passage progressif à un *euh*. La question semble totalement indécidable, aussi bien à l'écoute qu'au vu des courbes mélodiques et sonagrammes, et on peut se demander si cette « hésitation » sur l'hésitation n'est pas en fait révélatrice de l'identité de fonction de l'allongement et du *euh* (Figure 3).

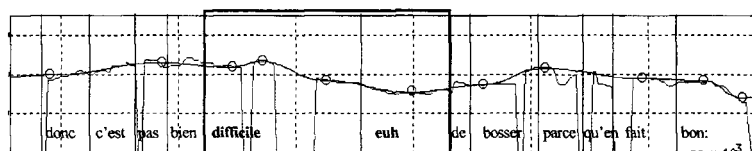


Figure 3. Incertitude entre allongement du schwa et *euh*.

Nous avons noté les hésitations sous forme d'allongement syllabique au moyen des deux points :

Étude des interactions entre pauses silencieuses et pauses remplies en français parlé

euh bon là c'était très bien hein j'étais au bord de la mer c'était super ça été un moment **de** : ++ où j'étais assez isolée mais aussi qui m'a été très profitable quoi **où** : + où il y a eu : enfin bon c'est : c'est c'était bien (pricle1) voilà alors hein ^ on **ne** : ++ il est il est difficile d'aborder **la** : + la question du métier sans parler un petit peu des origines et principalement des origines du service (procor1)

4.2. Séquences de pauses remplies

Dans 12% des cas, les pauses remplies se retrouvent sous une forme d'une séquence complexe, faisant intervenir plusieurs allongements, plusieurs *euh*, ou une combinaison mixte d'allongements et de *euh* (éventuellement entrecoupés par une ou plusieurs pauses silencieuses). L'extrait ci-dessous montre un exemple de combinaison particulièrement longue (on notera que la suite se continue par une amorce, suivie à nouveau d'une pause remplie) :

euh Beaune est une **euh la : la : la : euh le** : cé- le : cépage de : euh la ville de Beaune je veux dire (pridij1)

Au total, on observe 679 pauses remplies, qui constituent 594 séquences distinctes. Le Tableau 3 donne la répartition des différents types de pauses remplies, et la Figure 4 donne la répartition des séquences en fonction de leur longueur.

Estelle CAMPIONE

Type	N	sous-type	N
allongements	230	<i>simples</i>	216
		<i>complexes</i>	14
<i>eah</i>	323	<i>simples</i>	305
		<i>complexes</i>	18
mixtes	38		
Total	591		

Tableau 3. Répartition des différents types de pauses remplies.

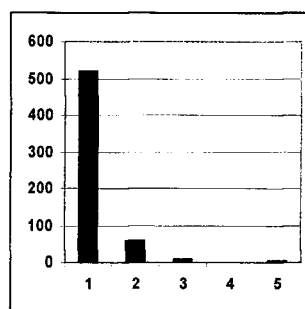


Figure 4. Répartition des séquences de pauses remplies en fonction de leur longueur

5. Étude des interactions

Pauses silencieuses et pauses remplies peuvent difficilement être étudiées de façon indépendante :

- 380 des 1163 pauses silencieuses de notre corpus interviennent au contact d'une pause remplie, c'est-à-dire 33%.
- Inversement, 344 des 591 séquences de pauses remplies sont au contact d'une pause silencieuse ou en contiennent une (58%).

C'est dire l'importance des interactions entre ces deux types de pauses.

5.1. Deux types de pauses remplies

Il nous semble important de distinguer du point de vue fonctionnel deux types de pauses remplies (ou de séquences). Les pauses remplies

Étude des interactions entre pauses silencieuses et pauses remplies en français parlé

internes à un segment constituent la majorité des cas (478 séquences sur 591, soit 81%). Elles marquent une interruption, suivie ou non d'une reprise et/ou d'une réparation :

on va parler **euh** + **euh** + de leur voyage en Italie par exemple (primon1)
euh Beaune / est une **euh la : la : la : euh le** : cé- le : cépage de : euh la
ville de Beaune / je veux dire / (pridij1)

Dans 113 cas sur 591 séquences, toutefois (19%), les pauses remplies interviennent au début d'un segment :

enfin bon voilà euh le dé*but de mon voyage / ça a été ça /
++
euh et après bon ben après les choses se sont mises en place / hein \
(pricle1)

Dans de tels cas, il ne s'agit pas d'une interruption liée à une difficulté de mise en place lexicale ou syntaxique de la fin du segment. Il s'agit sans doute d'« occuper le terrain » en attendant de trouver une suite au discours, et éviter ainsi que l'interlocuteur prenne la parole. Ce phénomène est très fréquent, puisqu'il affecte 113 segments sur 822, soit 14%. Ces 113 cas se répartissent en :

- 97 séquences commençant par euh ;
- 16 séquences commençant par un allongement.

Dans ce dernier cas, il s'agit (du moins dans notre corpus), uniquement de mots monosyllabiques et principalement des joncteurs, tels que *et*, *mais*, *puis* (11 cas sur 16). Les autres cas sont des mots-outils introducteurs de syntagmes (*de*, *où*, *le*).

5.2. Rôle des pauses silencieuses

Nous avons vu plus haut que les pauses silencieuses brèves ne sont jamais démarcatives. Parmi les 318 pauses silencieuses non démarcatives restantes, 289 (91%) sont associées à une pause remplie. Dans la grande majorité des cas (257 sur 289 soit 89%), la pause remplie précède la pause silencieuse :

on l'ap- au départ / on faisait euh le **le** : + euh le : ma*cérer le : le : le
poulet : + le poulet bien sûr qui est issu / aussi de la Bourgogne / puisque
euh + le poulet de Bresse / (pridij1)

Estelle CAMPIONE

Dans seulement 32 cas (11%), il y a interruption abrupte et la pause remplie intervient juste après :

et il a apprécié / + euh ce Corton blanc (pridij1)

et* / + euh on passe ^ donc on fait le choix de passer par l'école / (procor1)

Il conviendrait d'étudier plus en détail ces cas qui semblent, comme dans les exemples ci-dessus, intervenir après un mouvement mélodique majeur et/ou un accent.

Les 29 cas de pauses silencieuses non démarcatives qui ne sont pas associés à une pause remplie sont intéressants à analyser plus en détail. Leur répartition est la suivante :

1. Dans 18 cas (soit plus de la moitié), la pause silencieuse est en fait associée à une autre *marque du travail de formulation* :

- amorce (8 cas)

enfin bon mais disons que **c'est : ça t-** là c'est autre chose bon c'est plus tout à fait un voyage non plus (pricle1)
et puis là on vit **au jour le j-** + **au jour** le jour (pricle1)

- répétition (5 cas)

alors maintenant ++ euh maintenant euh + oui ben : je peux parler **d'un : de -** pas un de mes premiers voyages mais c'est pas vraiment (pricle1)
donc s- nous avons un rôle **de + de** soutien de marché (procor1)
je leur fais déguster un certain nombre de choses pour **qu'ils + ils** se ils sentent euh ils se rendent bien compte que c'est bon on fait déguster les huiles (primon1)

- allongement + onomatopée (*pff*, etc.) (2 cas) :

et en fait euh avec cette chienne je me suis sentie euh + là : + ben : **pff** + disons capable de bouger toute seule (pricle1)
alors que moi ça me dérange pas du tout au contraire : **pssff** + je trouve que chacun a ses limites et chacun a ses compétences et : voilà (prolyo1)

- marqueur (*ben*, *hein*, etc.) (3 cas) :

parce que bon tu en as qui apprennent sur le tas bon : euh + elle elle apprendra comme moi sur le tas **hein** + de toute façon (pricle3)
et puis malheureusement il reste des modules je dirais bâtards euh qu'on se partage **ben** + à regret parce qu'ils {rires} (prolyo1)
dans ce que j'essaye de : ++ proposer **hein** + aujourd'hui (pubstr1)

2. Dans 6 cas (20%), la pause silencieuse est associée à un *accent de focalisation*. Ce phénomène est assez intéressant et mériterait d'être étudié en détail sur des données plus volumineuses. Nous pouvons faire l'hypothèse qu'il permet à la fois de renforcer la perception de l'accent et au locuteur de refaire le plein de sa capacité pulmonaire avant, ou éventuellement après, l'émission de l'accent. Nous marquons l'accent par une étoile après la syllabe accentuée :

mais je veux dire euh j'avais + **te*llement** soif de cette liberté-là et
pour moi cette liberté elle passait par : (pricle1)
parce qu'il y a des clients qui + **tou*chent** la tache avant de venir
(pricle3)
c'est tout* un climat de confiance qui va s'installer + **de*** suite
(primon1)

On remarquera que, dans un cas, la pause silencieuse suit l'accent :

c'est la **deman*de** + du retour de la bonne heure (pubstr1)

3. Il reste seulement 5 cas qui ne soient associés à aucun autre indice :

va se sentir dans **un + endroit** de confiance donc* primon1)
enfin moi que j'ai besoin donc c'est vrai **que + avec** le client j'ai un
côté sincè :re - euh qui je pense est important (primon1)
euh par exemple de + que la : + euh d- euh cin- bon plus de cinquante
pour cent + de la : de la fl- de la flore (procor1)
en reconnaissance le **Corton + Charlemagne** (pridij1)
va être un **personnage + féminin** (pubpau1)

On remarquera que dans ces 5 exemples, il y a cohésion syntaxique du segment (il y a même liaison dans le premier exemple). Il n'y a, de plus, aucune discontinuité intonative au moment de la pause silencieuse. Il conviendrait de recueillir plus d'exemples et d'analyser très finement le phénomène au niveau acoustique et syntaxique, mais on peut dire qu'il n'y a pas hésitation dans les cas que nous avons observés. Ces pauses se comportent donc exactement comme les pauses brèves. Leur durée est d'ailleurs relativement peu élevée (200 à 470 ms).

6. Conclusion

Nous avons montré dans cette étude, à partir d'un sous-ensemble du *Corpus de référence du français parlé* d'environ une heure et faisant

Estelle CAMPIONE

intervenir 10 locuteurs (5 hommes et 5 femmes), que la pause silencieuse n'a en elle-même aucun rôle de marque d'hésitation ou de travail de formulation. Elle n'intervient dans cette fonction qu'associée à d'autres marques, principalement les pauses remplies (allongement syllabique et item quasi-lexical *eah*). D'autres indices apparaissent également en association avec les pauses silencieuses ou remplies, tels que les amorces, les répétitions, ou des items quasi-lexicaux comme *ben*, *hein*, *pf*, etc. Dans ce même volume, Sandrine Henry et Berthille Pallaud livrent une étude montrant l'interaction des amorces et des répétitions. Une étude globale de la totalité de ces phénomènes liés au travail de formulation reste à conduire.

Références

- Barik, H. C. (1968). On defining juncture pauses: a note on Boomer's « Hesitation and grammatical encoding ». *Language and Speech*, 11, 156-159.
- Barras, C., Geoffrois, E., Wu, Z., & Liberman, M. (1998). Transcribers: a Free Tool for Segmenting, Labeling and Transcribing Speech, *First International Conference on Language Resources and Evaluation (LREC)* (pp. 1373-1376).
- Bear, J., Dowding, J., & Shriberg, E. (1992). Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialogue, *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics* (pp. 56-63).
- Blanche-Benveniste, C., & JeanJean, C. (Eds.). (1987). *Le français parlé. Transcription et édition*. Paris : Didier Érudition.
- Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, 8, 148-158.
- Campione, E. (2001). *Étiquetage semi-automatique de la prosodie dans les corpus oraux : algorithmes et méthodologie*. Thèse de doctorat, Université de Provence, Aix-en-Provence.
- Candéa, M. (2000). *Contribution à l'étude des pauses silencieuses et des phénomènes dits « d'hésitation » en français oral spontané*. Thèse de doctorat nouveau régime, Paris III, Paris.
- Chafe, W. (1994). *Discourse, Consciousness and Time*. Chicago: University of Chicago Press.
- Clark, H. (1994). Managing problems in speaking. *Speech Communication*, 15, 243-250.
- Clark, H. & Clark, E. (1977). *Psychology and Language*. New York : Harcourt, Brace, Jovanovich.
- Cutler, A. (1998). The recognition of spoken words with variable representations, *Proceedings of the ESCA Workshop on the Sound Patterns of Spontaneous Speech* (pp. 83-92). Aix-en-Provence, France.

Étude des interactions entre pauses silencieuses et pauses remplies en français parlé

- Delattre, P. (1965). *Comparing the Phonetic Features of English, French, German and Spanish*. Heidelberg : Julius Groos Verlag.
- Dell, G. S. (1988). The retrieval of phonological forms in production : tests of predictions from connectionist model. *Journal of Memory and Language*, 27, 124-142.
- Duez, D. (1982). Salient pauses and non salient pauses in three speech style. *Language and Speech*, 25(7), 11-28.
- Duez, D. (19998). The aim of SPoSS, *Proceedings of the ESCA Workshop on the Sound Patterns of Spontaneous Speech* (pp.VII-IX). Aix-en-Provence, France.
- Esling, J. H. (1971). *Preliminary report on a research project in acoustic phonetics with comparisons of the vowel systems of several languages and a discussion of the neutral position*. PhD Thesis, University of Michigan, Michigan.
- Espesser, R. (1996). Mes : un environnement de traitement du signal. *21èmes Journées d'Études sur la Parole (JEP)* (pp. 447-451). Avignon.
- Finegan, E. (1994). *Language: its structure and use*, Second Edition. New York : Harcourt Brace.
- Fox, X. (1973). Tone sequences in English. *Archivum Linguisticum (New series)*, 4, 17-26.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27-52.
- Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 9). New York : Academic Press.
- Goldman-Eisler, F. (1972). Pauses, clauses and sentences. *Language and Speech*, 15, 103-113.
- Goto, M., Itou, K., & Hayamizu, S. (1999). A Real-time Filled Pause Detection System for Spontaneous Speech Recognition, *Proceedings of the 6th European Conference on Speech Communication and Technology (Eurospeech '99)* (pp. 227-230). Budapest.
- Grosz, B., & Hirschberg, J. (1992). Some Intonational Characteristics of Discourse Structure, *Proceedings of the 2nd International Conference on Spoken Language Processing (ICSLP)* (pp. 429-432).
- Guaitella, I. (1991). *Rythme et parole : comparaison critique du rythme de la lecture oralisée et de la parole spontanée*. Thèse de doctorat, Université de Provence, Aix-en-Provence.
- Hieke, Kowal, & O'Connell, D. C. (1983). The trouble with "articulatory" pauses. *Language and Speech*, 26, 203-214.
- Hirst, D. J., & Di Cristo, A. (1998). A Survey of intonation systems. In D. J. Hirst & A. Di Cristo (Eds.). *Intonation Systems. A Survey of Twenty Languages* (pp. 1-44). Cambridge : Cambridge University Press.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 201-258.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge MA : MIT Press.
- Levinson, S. (1983). *Pragmatics*. Cambridge : Cambridge University Press.
- Lickley, R. J. (1994). *Detecting Disfluency in Spontaneous Speech*. PhD Thesis, University of Edinburgh, Edinburgh.

Estelle CAMPIONE

- Morel, M. A., & Danon-Boileau, L. (1998). *Grammaire de l'intonation. L'exemple du français*. Paris : Ophrys.
- Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America (JASA)*, 95(3), 1603-1616.
- Nakatani, C. H., Grosz, B. J., Ahn, D. D., & Hirschberg, J. (1995). *Instructions for Annotating Discourses*. Cambridge, MA : Center for Research in Computing Technology, Harvard University.
- O'Shaughnessy, D. (1992). Recognition of hesitations in spontaneous speech, *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing* (pp. 521-524). San Francisco : CA. IEEE.
- Quimbo, F. C. M., Kawahara, T., & Doshita, S. (1998). Prosodic analysis of fillers and self-repair in Japanese speech, *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*. Sydney, Australia.
- Pallaud, B. (2003). Erreurs d'écoute dans la transcription de données orales. *Revue Parole*, 22-23-24, 267-294.
- Rose, R. L. (1998). *The Communicative Value of Filled Pauses in Spontaneous Speech*. Unpublished MA dissertation. University of Birmingham. [Online : <http://pubweb.northwestern.edu/~lro388/fprc/madissrt/> (accessed 12 dec. 2002)].
- Shriberg, E. (1994). *Preliminaries to a Theory of Speech Disfluencies*. PhD Thesis, University of California, Berkeley.
- Stenstrom, A. (1994). *An Introduction to Spoken Interaction*. London and New York : Longman.
- Vaissière, J. (1991). Rythm, accentuation and final lengthening in French. In J. Sundberg & L. Nord & R. Carlson (Eds.), *Music, Language, Speech and Brain*. Macmillan Press.
- Vaissière, J. (1997). Langues, prosodies et Syntaxe. *Traitement Automatique des Langues*, 38(1), 53-81.
- Wennerstrom, A. (1994). Intonational meaning in English discourse: a study of non-native speakers. *Applied Linguistics*, 15(4), 399-420.
- Zellner, B. (1998). *Caractérisation et prédiction du débit de parole en français. Une étude de cas*. Thèse, Université de Lausanne, Lausanne.

Amorces de mots et répétitions dans les énoncés oraux

Sandrine HENRY¹ & Berthille PALLAUD²
Université de Provence

Résumé

Cette étude se propose d'étudier, d'un double point de vue, deux sortes de phénomènes fréquents à l'oral : les amorces de mots et les répétitions. Elle permet d'une part de situer les amorces parmi les phénomènes de répétitions, d'autre part de décrire les répétitions dans les achoppements (notamment leurs fréquences et leur empan). Il est possible d'établir la catégorie (complétée, modifiée ou inachevée) de ces amorces répétées ainsi que leur situation syntaxique et leur catégorie grammaticale. Les phénomènes d'incises ne sont pas absents de cette catégorie d'amorces.

1. Introduction

Lorsque le linguiste refuse de se prononcer sur le caractère intentionnel ou non des énoncés produits par un locuteur (qu'ils soient oraux ou écrits, d'ailleurs), il délimite un champ d'où sont exclues les dimensions psychologiques. Ce faisant, il s'interdit, en particulier pour les énoncés oraux, de distinguer et d'écarter de ses études certains éléments des énoncés sous prétexte qu'ils sont (intentionnellement ou non) malformés, parcellaires, répétés, amorcés, jonchés d'hésitations de tous ordres.

¹ Équipe DELIC, Université de Provence, 29, av. R. Schuman, Aix-en-Provence, 13621 cedex 1

e-mail : sandrine.henry@up.univ-aix.fr

² UMR 6057 Parole et Langage, Université de Provence, 29, av. R. Schuman, Aix-en-Provence, 13621 cedex 1

e-mail : pallaud@newsup.univ-mrs.fr

Sandrine HENRY & Berthille PALLAUD

Cette position, découragée autrefois, connaît depuis un essor que les nouvelles technologies ont beaucoup soutenu et facilité (Blanche-Benveniste, 1997).

Notre étude s'intéresse à ces phénomènes caractéristiques de l'oral et souvent qualifiés de « scories » ou même de « ratés » : les amorces de mots et les répétitions. Outre que ces phénomènes réservés à l'oral étaient considérés comme polluant l'énoncé oral et, de ce fait, négligés par les linguistes, ils ne relevaient que d'une seule interprétation : l'embarras au sens étymologique du terme³. L'énonciation s'immobilise sur une place syntaxique ; l'élaboration de l'énoncé échoue plus ou moins momentanément à se poursuivre. Les amorces de mots comme les répétitions ont en commun cette suspension de l'énoncé sur une place syntaxique. Cependant, qu'ont-elles d'autre en commun ? Ces deux phénomènes, tous deux fréquents à l'oral, sont loin d'être des procédés⁴ énonciatifs, tant ils échappent, le plus souvent, à la perception même des locuteurs et de leurs « récepteurs » (Pallaud, 2003a ; Roubaud, ce volume). Pour autant, remplissent-ils les mêmes fonctions, touchent-ils les mêmes places syntaxiques et les mêmes catégories grammaticales ? Leurs effets en contexte sont-ils les mêmes ?

Ces phénomènes d'achoppements ont été au centre des interventions dans un colloque en 2001 à Edinburgh où la grande majorité des études sur les « dysfluences » à l'oral ont été abordées d'un point de vue psycholinguistique soucieux de préciser les fonctions et les causes de ces trébuchements dans les énoncés oraux :

« When a speaker detects trouble that is sufficiently alarming according to the speaker's current standards, the decision will be taken to interrupt speech and to initiate a repair » (Levelt, 1989 : 478).

Ce point de vue a conduit cet auteur à ne pas distinguer entre les amorces de mots et les amorces de syntagmes. En revanche, que le locuteur interrompe son énoncé après avoir prononcé un seul phonème lui semble suggérer que l'interruption doit avoir son origine avant que l'erreur ne soit patente, avis également partagé par Nootboom (2001, p. 24).

D'autres études comme celle de Lickley (2001) montrent une grande variabilité intra- et inter-individuelle quel que soit le type d'achoppement (pause remplie, répétition, amorce de mot, prolongation de mot, etc.). Nous

³ « *Obstacle au passage, à la circulation causé par la rencontre ou l'accumulation de plusieurs objets ou véhicules* » (Le Petit Robert, 1973).

⁴ Ce terme implique la conscience voire la maîtrise de l'élaboration chez le locuteur.

avons confirmé ce résultat en ce qui concerne les amorces de mots : dans notre recueil de français parlé, la fréquence de ces achoppements varie selon les locuteurs dans un rapport de 1 à 10 (Pallaud, 2002). D'autres sources possibles de variation pour ces « dysfluences » s'avèrent être le *cadre conversationnel* (selon la terminologie de Cosnier, 1987) et le thème (ou *acte de langage* chez Lickley, 2001) : fournir des instructions ou des explications suscite un fort taux de « dysfluences ».

Les amorces ou fragments de mots, tout comme les phénomènes de répétitions sont des événements langagiers d'une grande fréquence à l'oral. Elles sont facilement repérables dans les corpus oraux de toutes origines. On a vu (Pallaud, 1999) que les amorces de mots constituent plus de la moitié des achoppements de langage identifiés dans un corpus enregistré de français contemporain. Un recensement (Pallaud, 2002) du nombre total d'amorces (197 amorces) sur un recueil de corpus de 46 649 mots⁵ montre, toutefois, que ces phénomènes viennent loin derrière les pauses courtes silencieuses⁶ ou remplies⁷. Les amorces sont six fois moins nombreuses que ces phénomènes. En moyenne, on peut dire qu'on trouve environ 4 amorces tous les 1 000 mots soit pour un débit moyen de 200 mots par minute une amorce toutes les 75 secondes⁸.

Les amorces de mots ont parfois été interprétées comme des erreurs de langage de type lapsus. Il en est des amorces comme de toutes les « scories » constatées à l'oral, leur statut intentionnel ou non, leurs causes sont impossibles à identifier avec quelque rigueur (Pallaud, 2001).

La présente étude se situe à l'interface de deux analyses réalisées sur des corpus oraux de français parlé : les amorces de mots d'une part (Pallaud, 2002) et les répétitions d'autre part (Henry, 2002). À l'oral, les amorces de mots comme les répétitions sont constituées de phénomènes de catégories différentes ce qui permet de les caractériser comme un ensemble hétérogène obéissant à plusieurs types de contraintes ou de mécanismes. En

⁵ Cet ensemble est composé de 12 corpus recueillis par Pallaud dans le cadre du *Corpus de référence de français parlé* (Pallaud, 2001). 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/>).

⁶ Les pauses silencieuses courtes d'une durée de 200 ms environ. Cette appréciation n'est qu'approximative dans ces corpus : des intervalles mélodiques ont pu être identifiés comme des pauses.

⁷ Ces pauses s'accompagnent d'un *euh* d'« hésitation ».

⁸ Il reste que ces résultats sont dépendants de la taille du recueil de corpus. Sur notre corpus de travail d'environ un million de mots, la fréquence des amorces est alors de 6,1 pour 1 000 mots, soit une amorce toutes les 50 secondes.

Sandrine HENRY & Berthille PALLAUD

particulier, il a été établi qu'amorces et répétitions varient en fréquence et en caractéristiques selon leur place dans les grands constituants syntaxiques. Cela ne veut pas dire que ces deux types d'achoppements ne soient pas à distinguer. Des travaux neurophysiologiques⁹ sur la détection des répétitions et des interruptions syntaxiques dans les énoncés ont montré que la détection de ces deux sortes de « dysfluences » s'accompagne de potentiels évoqués différents (cf. la revue de question chez Mc Allister *et al.*, 2001).

1.1. Le point de vue des amorces

Les amorces recueillies s'avèrent être soit des phénomènes de listes (une recherche lexicale sur le même site syntaxique) soit des éléments qui introduisent une rupture syntaxique (le contexte qui suit l'amorce n'appartient pas à la même unité syntaxique que l'amorce). Trois grandes catégories d'amorces ont été distinguées :

1.1.1. Dans les phénomènes de listes :

- **amorces complétées** : le mot commencé et interrompu se trouve complété sur la même place syntaxique :

(1)¹⁰ c'est vrai que c'est pas **b- beau** d'associer les deux choses
(ARBORIGN, 5, 14)

- **amorces modifiées** : le mot amorcé ne se trouve pas complété mais est remplacé par un autre mot sur la même place syntaxique. Il s'agit des amorces où le locuteur ne complète pas ce qu'il avait commencé à dire mais poursuit son énoncé par un autre élément :

(2) dans une région où le mistral euh **su- souffle** euh bon train
(BIRNIE, 8, 10)

(3) on va + attaquer l'autre **b- morceau** l'autre moitié du dos
(ARBORIGN, 28, 1)

1.1.2. Dans le cas d'une rupture syntaxique :

- **amorces laissées inachevées** : le fragment de mot ne débute pas un phénomène de liste (Blanche-Benveniste, 1987b). On n'est donc pas

⁹ Études fondées sur les potentiels évoqués : la détection de répétitions ne s'accompagne pas du même potentiel évoqué que celle d'une interruption syntaxique. Cette corrélation linguistiquement fondée a été vérifiée plusieurs fois.

¹⁰ Référence du corpus, N° de page et N° de ligne.

en présence d'un piétinement sur la même place syntaxique. Ce qui suit l'amorce occupe une autre place syntaxique. Ces amorces ont parfois été considérées comme des lapsus non corrigés :

- (4) alors je vais euh faire un petite **diver-** on va diverger là pour expliquer ça euh au début (MICROBIO 28,15)

1.2. Le point de vue des répétitions

Bien qu'elles soient regroupées sous une dénomination commune, les répétitions, doivent être cependant scindées en deux grandes catégories. Dans un précédent travail (Henry, 2002b), nous avons établi cette distinction en opposant les répétitions « faits de langue » aux répétitions « faits de parole ». Les premières se retrouvent aussi bien à l'oral qu'à l'écrit. Elles sont prévues par la grammaire (elles ne font pas l'objet du présent travail), tandis que les secondes, présentes uniquement à l'oral, ne le sont pas. Ces dernières, celles qui « enfreignent » le « bon usage », ont été le plus souvent négligées. Nous pensons toutefois que ces répétitions représentent des événements langagiers fort intéressants car ils constituent de véritables indices dans la production des énoncés oraux. Nous avons par exemple constaté, en nous fondant sur un corpus d'environ un million de mots, que l'apparition de ces répétitions ne s'effectuait pas au hasard mais essentiellement à l'initiale de syntagme.

Exemples de répétitions faits de langue :

- (5) je trouve ça + **honteux + honteux + honteux** (Éditeur)
(6) je l'ai vuc **très très** perturbée mais en même temps **très très** forte (Femmes)

Exemples de répétitions faits de parole :

- (7) j'étais pas né mais enfin **les les** vieux me racontent + sur la Canebière ils avaient euh renversé **les les les** tramways (Navale)
(8) ils auront **leur propre leur propre** langage (Laurent)

Les répétitions d'amorces constituent un sous-ensemble des répétitions faits de parole :

- (9) non ça **su- su-** sur ce sujet-là je peux pas vous renseigner (FP257B)
(10) au bout de la rangée + **c'é- c'é-** c'était pas nécessairement parce qu'elles en avaient cueillies plus que les autres (Merl3)

Sandrine HENRY & Berthille PALLAUD

1.3. Méthode de recueil des répétitions d'amorces

La présente étude prend appui sur des données extraites d'un corpus informatisé *Corpaix*¹¹ auquel s'adjoignent des transcriptions inédites pour former un corpus de travail de 1 000 382 mots¹². Après un léger toilettage¹³ de nos résultats, nous recensons **150 répétitions d'amorces**. Le nombre total d'amorces dans ce recueil est de 6 094.

1.4. Fréquence des amorces répétées à l'oral

Les amorces répétées ne constituent qu'une faible partie des phénomènes d'amorces mais aussi de répétitions. Sur un total de **6 094 amorces**, nous n'avons trouvé que 150 amorces répétées, soit **2,4%**. Sur un total de **16 135 répétitions faits de parole** (répétitions d'amorces incluses)¹⁴, seules **0,93%** d'entre elles sont des répétitions d'amorces.

Si l'on considère que le phénomène d'amorce se produit dans une fréquence de 6,09 pour 1 000 mots (6 094 amorces dans un corpus de 1 000 382 mots), on peut estimer la fréquence des amorces répétées dans un corpus en confrontant ce chiffre au pourcentage d'amorces répétées dans notre recueil d'amorces, soit 2,4%. La fréquence des amorces répétées est ainsi beaucoup plus faible : 1,5 amorces répétées tous les 10 000 mots soit, avec un débit moyen de 200 mots/minute, une fréquence d'une amorce répétée toutes les 33 minutes. Cette fréquence est donc 40 fois moins importante que les amorces non répétées qui, elles, se produisent toutes les 50 secondes.

¹¹ Corpus constitué au cours de ces vingt-cinq dernières années par les membres du Groupe Aixois de Recherche en Syntaxe, actuellement géré par l'équipe DELIC.

¹² Afin d'assurer une fiabilité dans l'extraction des répétitions d'amorces, celles-ci sont détectées à l'aide d'un programme informatique : écriture d'un script en langage *Perl*, puis application sous *Linux*. Les résultats sont ensuite ouverts dans un tableur *Excel* pour en faciliter le traitement.

¹³ Nous avons supprimé les amorces *i-*, tellement nombreuses et diversement appréciées par les transcrip-teurs, et qui nous semblent correspondre plutôt à des facilités de prononciation du pronom *il* [il] en [i] (cf. Gadet, 1997 : 77).

¹⁴ Base de données « Répétitions Totales ».

2. Analyse de ces phénomènes du point de vue des amorces

2.1. Méthodologie

Trois types d'analyse seront conduits sur ces amorces répétées toujours comparativement aux analyses sur la totalité des amorces et à celles sur les phénomènes de répétitions. Il s'agit de repérer si les amorces répétées constituent une catégorie à part dans les répétitions ou les amorces. Le principe est de déterminer, pour chaque amorce répétée, à quelle catégorie d'amorce elle appartient, dans quel constituant syntaxique elle est située, et quelle est sa catégorie grammaticale. Rappelons avec Jeanjean (1984 : 171) que « pour assigner une valeur syntaxique à ce genre d'amorces, on n'a pas besoin de reconstituer un quelconque lexème ». Ce constat même lui permettait de conclure que les amorces ne constituaient pas des échecs dans l'énoncé : « Il n'y a pas lieu de parler de ratés puisqu'on peut rendre compte du statut grammatical de l'amorce » (*ibidem*, p. 174).

2.2. Résultats

Le recueil comprend 150 amorces répétées qui se répartissent en trois catégories selon que les amorces sont complétées, modifiées ou inachevées. Cette répartition est comparée à celle qui a été observée sur un échantillon de 948 amorces non répétées.

2.2.1. Amorces répétées et catégories d'amorces

Type d'amorce	Amorces répétées		Amorces non répétées	
	Effectifs	Pourcentage	Effectifs	Pourcentage
Complétées	123	82,0%	565	59,6%
Modifiées	7	4,7%	175	18,5%
Inachevées	20	13,3%	208	21,9%
TOTAL	150	100%	948	100%

Tableau 1. Place des amorces répétées parmi les catégories d'amorces

Dans une précédente étude (Pallaud, 2002), nous avons montré que les amorces, de façon générale, se trouvaient plus souvent complétées que modifiées ou inachevées.

Sandrine HENRY & Berthille PALLAUD

Lorsqu'il y a répétition, la répartition des amorces dans les trois catégories (inachevées, complétées et modifiées) est très significativement différente de ce que laissent prévoir les proportions obtenues dans les mêmes catégories sur les amorces non répétées ($\text{Khi}^2 = 29,86$; d.d.l. = 2 ; $p < 0,001$). Mais, c'est encore dans la catégorie des amorces complétées qu'a lieu, le plus souvent, la répétition (82%).

En revanche, les amorces modifiées ou inachevées sont moins souvent répétées. On a vu que ces amorces ne constituaient respectivement que 18,5% et 21,9% des amorces non répétées. Lorsqu'on ne considère que les amorces répétées, on constate que les répétitions sont alors significativement moins nombreuses dans ces deux catégories (4,7% et 13,3%)¹⁵. L'amorce répétée est donc moins souvent inachevée ou modifiée que ce que laisse prévoir la répartition dans les amorces non répétées.

Ces résultats attestent de la relation de dépendance qui existe entre la catégorie d'amorce, quelle qu'elle soit, et le phénomène de répétition. Cette relation se traduit par une augmentation de la proportion des répétitions dans la catégorie des amorces complétées. 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 plus souvent complété. Si tel n'était pas le cas, la répartition des amorces répétées (dans les trois catégories décrites) ne différerait pas de la répartition des amorces non répétées.

L'étude de Cappeau (1998 : 306) sur les phénomènes de bribes à l'oral (c'est-à-dire des amorces de syntagmes répétées) en relation avec le genre a montré que « très souvent la bribe ne s'accompagne pas d'un changement de genre » :

(11) donc la la présence de l'hôtesse est une garantie (Guides) (*ibidem*, p. 307)

En ce qui concerne les bribes avec changement de genre (donc amorces de syntagmes modifiées), Cappeau déclare qu'elles sont 5 à 8 fois moins nombreuses que les précédentes. Son étude souligne donc elle aussi la prépondérance des bribes complétées par rapport à celles qui seront modifiées. Ceci nous encourage à conclure que, selon nos résultats, ces

¹⁵ Pour les amorces inachevées répétées (par rapport aux amorces complétées + modifiées) : $\text{Khi}^2 = 5,83$; d.d.l. = 1 ; $p < 0,02$.
Pour les amorces modifiées (par rapport aux amorces complétées + inachevées) : $\text{Khi}^2 = 17,81$; d.d.l. = 1 ; $p < 0,001$.

phénomènes de bribes et d'amorces de mots sont très voisins et sont à aborder dans un ensemble plus vaste : la bribe dans le syntagme.

2.2.2. Amorces répétées et constituants syntaxiques

Tout comme les amorces en général, les répétitions d'amorces affectent des éléments en position Sujet, Verbe ou Objet¹⁶. Nous nous référons ici aux notions de Grands Constituants telles qu'elles ont été employées dans le cadre de l'équipe du GARS. Dans notre recueil d'amorces répétées, nous n'avons pas relevé de sujet postposé au verbe ; cela revient à distinguer les éléments qui se trouvent à « gauche » (constituant Sujet) ou à « droite » (constituant Objet) du verbe.

En voici quelques exemples :

- Sujet
(12) alors là l- l- le poisson frétille (PR6B)
- Verbe
(13) parce que ça p- p- ça peut poser des problèmes avec la population locale (VoyagesONG)
- Objet
(14) alors là nous avons contacté les a- les a- voilà (BusABCD)

Nous avons comparé la répartition, selon ces constituants syntaxiques, des éléments amorcés dans le cas de répétition, avec la répartition que nous avons observée sur les amorces non répétées.

Constituants syntaxiques	Amorces répétées		Total des amorces	
	Effectifs	Pourcentage	Effectifs	Pourcentage
Sujet	71	51,8%	121	12,9%
Verbe	26	19,0%	335	35,9%
Objet	40	29,2%	479	51,2%
TOTAL	137	100%	935	100%

Tableau 2. Amorces répétées et constituants syntaxiques

De façon générale (Pallaud, 2002), les amorces de mots, si l'on tient compte de la linéarité de la transcription, sont plus nombreuses (63%) à droite du verbe (c'est-à-dire en position d'Objet le plus souvent) qu'à sa

¹⁶ Nous avons préféré porter dans une catégorie « indéterminés » les éléments qui ne peuvent être classés dans les constituants Sujet, Verbe et Objet. Il n'est en effet pas toujours possible de les ranger pour autant dans la catégorie des éléments « associés » (non régis par le verbe) selon la terminologie de Blanche-Benveniste (1997 : 157).

gauche (en position Sujet, en général). Parmi les amorces de mots, celles qui concernent les éléments Sujet (en l'occurrence tous à la gauche du verbe) ne constituent que 12,9% des cas.

La répétition d'amorces a des effets même si elle ne se produit pas non plus de façon aléatoire selon les grands constituants syntaxiques (Sujet, Verbe, Objet). La répartition de ces amorces répétées, selon ces grands constituants, n'est pas la même que celle que nous avons trouvée dans les amorces non répétées. Elle est même très différente. Si les amorces répétées Sujet quadruplent de fréquence (52%) par rapport à ce qui avait été trouvé dans les amorces totales, cela se fait au détriment des amorces Objet et Verbe qui, elles, diminuent de moitié. La répétition d'amorces est donc observée de façon particulière dans la catégorie des constituants Sujet. Les amorces Sujet, qui ne représentaient que 12,9% des amorces en général, constituent, lorsqu'elles s'accompagnent de répétition, la moitié des amorces répétées. La présence des amorces est, donc, significativement dépendante de la variable constituant syntaxique (Sujet, Objet ou Verbe) et en interaction avec le phénomène de répétition¹⁷.

Ce résultat s'explique sans doute par le fait qu'il y a une relation de dépendance entre le type de constituant et le phénomène d'amorce : nous avons en effet trouvé que les amorces Sujet étaient dans les trois quarts des cas (76%) des amorces complétées (Pallaud, 2002). Cette proportion différait significativement¹⁸ de celle qui avait été observée dans les autres constituants syntaxiques (61%). Dans notre recueil d'amorces répétées, nous faisons la même constatation : la proportion d'amorces répétées Sujet qui seront complétées constituent la très grande majorité de ces amorces répétées (91,5%) et diffère significativement¹⁹ des amorces répétées complétées appartenant aux constituants Verbe et Objet (74,2%). Il est donc logique (puisque les amorces répétées sont le plus souvent des amorces complétées) de retrouver dans ce type de constituant syntaxique un taux de répartition des amorces répétées plus élevé que dans le reste des répétitions d'amorces.

Parmi les phénomènes d'amorces, lorsqu'il y a répétition, il s'agit plus souvent du constituant Sujet et d'amorces complétées. Les résultats obtenus sur les répétitions d'amorces contribuent à préciser la place particulière occupée par le constituant Sujet.

¹⁷ Pour les amorces Sujet : $\text{Khi}2 = 127,54$; d.d.l. = 1 ; $p < 0,001$.

Pour les amorces Objet : $\text{Khi}2 = 23,22$; d.d.l. = 1 ; $p < 0,001$.

Pour les amorces Verbe : $\text{Khi}2 = 14,13$; d.d.l. = 1 ; $p < 0,001$.

¹⁸ $\text{Khi}2 = 10,60$; d.d.l. = 1 ; $p = 0,001$.

¹⁹ $\text{Khi}2 = 7,33$; d.d.l. = 1 ; $p < 0,01$.

2.2.3. Place des amorces répétées en fonction des catégories d'amorces et des constituants syntaxiques

Dans le tableau suivant, nous présentons le nombre d'amorces répétées d'une part en fonction de la place syntaxique sur laquelle a lieu le phénomène d'achoppement, et d'autre part en fonction du type d'amorce.

	Complétées		Modifiées et Inachevées		Total des amorces répétées	
Sujet	65	52,8%	6	22,2%	71	47,3%
Verbe	17	13,8%	9	33,3%	26	17,3%
Objet	32	26,0%	8	29,6%	40	26,7%
Indéterminée	9	7,3%	4	14,8%	13	8,7%
TOTAL	123	100%	27	100%	150	100%

Tableau 3. Amorces répétées en fonction des catégories d'amorces et des constituants syntaxiques

La répartition des répétitions d'amorces dans les constituants syntaxiques (Sujet, Verbe, Objet) en fonction des catégories d'amorces permet de définir si ces deux variables sont en dépendance et en interaction.

En ce qui concerne les amorces complétées, on n'observe pas d'interaction. Les amorces complétées Sujet restent les plus nombreuses (la moitié des amorces répétées), les amorces sur les éléments Objet représentent le quart et les amorces sur les éléments Verbe représentent 14% seulement. Cette répartition ne diffère pas de celle qui est observée sur la totalité des amorces répétées ($Kh^2 = 0,93$; d.d.l. = 2 ; N.S.).

En ce qui concerne la catégorie des amorces modifiées (peu nombreuses) et des amorces inachevées (un peu plus nombreuses), la répartition selon les constituants semble être due au hasard : la proportion est à peu près la même que l'amorce soit Sujet, Objet ou Verbe. Toutefois, la faiblesse des effectifs empêche des conclusions fermes. Nous n'avons pas observé cette interaction dans le cas des amorces non répétées (Pallaud, 2002). Les amorces en position Sujet, Verbe ou Objet étaient les mêmes dans les amorces complétées, modifiées et inachevées : la moitié des amorces étaient en position Objet, 35% sur le constituant Verbe et le reste concernait des éléments Sujet.

Il semblerait, donc, qu'il y ait une interaction de la variable répétition dans le cas des amorces modifiées et inachevées. Dans ce cas-là seulement, les répétitions d'amorces ont lieu indifféremment sur des éléments en position Sujet, Verbe ou Objet. Pour les amorces complétées, au contraire, que l'élément amorcé et répété appartienne au constituant Sujet, Verbe ou

Sandrine HENRY & Berthille PALLAUD

Objet n'est pas indifférent. La répartition n'est pas aléatoire : les amorces
Sujet sont de loin les plus nombreuses ; viennent ensuite les amorces Objet,
puis les amorces Verbe.

3. Analyse du point de vue des répétitions

Cette étude s'articulera en deux volets :

- un premier volet s'intéressera uniquement à la structure des répétitions d'amorces, structure définie par un certain nombre de paramètres, de traits ;
- un second volet fournira une analyse morpho-syntaxique des répétitions d'amorces.

Nous proposons, pour certains points, une comparaison avec les résultats obtenus dans le cadre d'une étude quantitative des répétitions faits de parole (excluant les répétitions d'amorces, Henry, 2002a)²⁰ réalisée à partir du même corpus de travail).

3. 1. Analyse de certains traits de la répétition

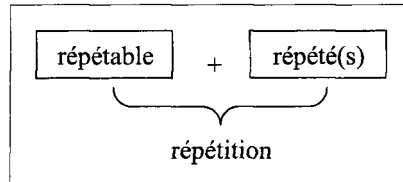
3.1.1. Empan de la répétition

Le premier point de comparaison porte sur l'empan des répétitions incluant des amorces, c'est-à-dire sur le nombre de *répétés* contenus dans la répétition. Nous employons ici le terme *répété* en accord avec la définition donnée par Candéa (2000 : 315) :

« Toute répétition forme un bloc dans la parole qui comporte au minimum deux éléments : un premier élément que nous appellerons le « répétable » et un deuxième élément, identique au premier, que nous appellerons le « répété ». Il va de soi qu'en théorie toute unité produite dans la parole est **en principe** un répétable et ce n'est que la présence d'un répété immédiatement après qui fait que ce répétable va entrer effectivement dans la composition d'un bloc que nous appelons a posteriori une « répétition ». »

C'est ce que traduit le schéma ci-après :

²⁰ Base de données « Répétitions ».



Nous avons distingué les répétitions d'amorces à **répété unique**, que nous nommons **répétitions simples** (15), de celles à répétés multiples. Si la répétition contient **deux répétés**, on parle alors de répétition **double** (16), si elle en contient **trois** (17), de répétition **triple** (ici, le maximum observé) :

- (15) il y a des moments /où, Ø/ il vaut mieux être **ho- ho-** honnête vis-à-vis des gens que de faire plaisir aux gens je crois (Brel)
- (16) mais **no- no- no-** notre base politique veut le que ouais que la que le peuple ait souvent son + son mot à dire (Genève)
- (17) non je sais pas trop **re- re- re- re-** reposez-moi la question (Yolande)

3.1.2. Longueur de la séquence du répétable

Par ailleurs, notons que le nombre d'éléments contenus dans le répétable (*NEI*) constitue un autre critère de classification des répétitions. En ce qui concerne nos répétitions d'amorces, ce nombre peut varier de 1 à 3. Les exemples (15) à (17) présentent des cas de répétitions dont le répétable contient un seul élément. Les exemples (18) et (19) présentent un *NEI* multiple :

- (18) et je j'**ai-** j'**ai-** j'aimais beaucoup énormément taper à la machine (Puget-VI)
- (19) s'**ha-** **oui** s'**ha-** **oui** s'habiller (Femmes)

Le croisement de ces deux paramètres – le type de répétition (simple, double ou triple) et le nombre d'éléments contenus dans le répétable (*NEI* = 1, 2 ou 3) – permet d'obtenir le tableau suivant.

	Simple		Double		Triple		TOTAL	
NEI = 1	110	73,3% ²¹	20	13,3%	2	1,3%	132	87,9%
NEI = 2	16	10,7%	1	0,7%	-	-	17	11,4%
NEI = 3	1	0,7%	-	-	-	-	1	0,7%
TOTAL	127	84,7%	21	14,0%	2	1,3%	150	100%

Tableau 4. Part des répétitions d'amorces selon le type de répétition et la longueur du répétable

²¹ Pourcentage calculé sur le nombre total de répétitions d'amorces.

Sandrine HENRY & Berthille PALLAUD

Que leur répétable soit constitué par un, deux ou trois éléments, les répétitions d'amorces simples (répétitions à répété unique) constituent la grande majorité des cas (84,7%). Dès que le répétable comporte plus d'un élément, la proportion de répétitions simples est, à une exception près, la seule concernée. De plus, la fréquence de ces cas diminue très fortement ; elle n'est plus que de 10,7% pour deux éléments et de 0,7% pour trois éléments.

Lorsque la répétition d'amorces est double ou triple, on peut dire que seule la catégorie de répétable à un élément est concernée et que, là aussi, la fréquence de ces cas d'amorces répétées chute drastiquement. Les répétitions d'amorces à un seul élément dans le répétable, qu'elles soient simples, doubles ou triples, forment la grande majorité (87,9%) des amorces répétées²².

Ainsi, la probabilité d'avoir une répétition d'amorces diminue pour un nombre multiple de répétés ou d'éléments contenus dans le répétable. Les répétitions d'amorces simples et qui présentent un seul élément dans le répétable sont prépondérantes (73,3% des répétitions d'amorces totales).

3.1.3. Comparaison des répétitions d'amorces et des répétitions en général selon l'empan

Nous avons vu que le phénomène de répétition affecte les amorces de façon différentielle : les proportions des catégories d'amorces (complétées, modifiées ou inachevées) ne sont pas les mêmes lorsqu'il y a répétition et lorsqu'il n'y en a pas. Il convenait donc de vérifier si le phénomène d'achoppement que constitue l'amorce de mot vient interférer sur les répétitions et se traduit notamment par une répartition différente selon les paramètres caractéristiques des répétitions, notamment l'empan (répétition simple ou multiple)²³.

²² Dans l'étude de Cappeau (1998), la différence est encore plus nette puisqu'il ne relève que 4% (5 cas sur 123) de bribes multiples.

²³ Ne disposant pas des données définitives concernant la longueur de la séquence du répétable dans le cas des répétitions en général, nous ne pouvons faire cette deuxième comparaison.

En retranchant aux 16 135 répétitions faits de parole les 150 répétitions d'amorces, nous obtenons 15 985 répétitions²⁴ pour lesquelles nous fournissons une répartition en fonction du type de répétition c'est-à-dire du nombre de répétés (ici *NR*) contenus dans la répétition. Les résultats sont présentés dans le tableau ci-dessous :

Type de répétition	Répétitions d'amorces		Répétitions ²⁵	
	Effectifs	Pourcentage	Effectifs	Pourcentage
Simple (<i>NR</i> = 1)	127	84,7%	14 988	93,8%
Double (<i>NR</i> = 2)	21	14,0%	854	5,3%
Triple (<i>NR</i> = 3)	2	1,3%	110	0,7%
<i>NR</i> ≥ 4	-	-	33	0,2%
TOTAL	150	100%	15 985	100%

Tableau 5. Répartition des répétitions d'amorces en fonction du nombre de répétés

La répartition des répétitions d'amorces en fonction du nombre de répétés s'avère significativement²⁶ différente de la répartition constatée dans le recueil « Répétitions ». On y trouve plus souvent des répétitions multiples.

Dans le cas des répétitions d'amorces comme dans celui des répétitions en général, la proportion des cas de répétitions simples est largement majoritaire (respectivement 84,7% et 93,8%). Par ailleurs, qu'il s'agisse de répétitions d'amorces ou de répétitions sur du lexique achevé, les répartitions suivent une même tendance puisque les effectifs diminuent au fur et à mesure que le nombre de répétés augmente.

Lorsque la répétition est simple (le nombre de répétés *NR* = 1), la part des répétitions d'amorces s'avère moins importante que celle obtenue pour les répétitions portant sur du lexique achevé. Cependant lorsque le nombre de répétés est supérieur ou égal à 2, le pourcentage de répétitions d'amorces s'avère plus élevé que le pourcentage de répétitions de mots non tronqués.

On constate, par ailleurs, une différence quant au nombre de types de répétitions : le nombre de répétés, limité à 3 pour les répétitions d'amorces, peut être supérieur à 4 pour les répétitions en général²⁷ (Henry, 2002).

²⁴ Base « Répétitions ».

²⁵ Répétitions de mots « achevés » par opposition aux répétitions d'amorces.

²⁶ $\text{Khi}^2 = 22,29$; d.d.l. = 1 ; $p < 0,001$.

²⁷ Notre base « Répétitions » recense un cas à 7 répétés.

3.1.4. Succession des séquences *répétable* et *répété(s)*

Nous avons constaté une certaine diversité dans la succession des séquences *répétable* et *répété(s)*. Dans certains cas, ces séquences sont produites en contiguïté (20), dans d'autres, un ou plusieurs éléments sont insérés, soit entre le *répétable* et le *répété* (21), soit à l'intérieur même de la séquence des *répétés* (22). La transition entre le *répétable* et le *répété* (avec incise ou non) correspond à la « phase d'édition » (*editing phase* dans la terminologie de Levelt, 1983), laquelle est marquée, dans le cas de séquences non contiguës, par la présence d'hésitations, ou bien de pauses, ou bien encore de commentaires « rédactionnels » (*editing terms*).

- Exemple de contiguïté :

(20) je me suis abonné **u- u-** une année à un magazine (Rose)

- Exemples de non contiguïté :

(21) tu es **t- franchement t-** tu as appris beaucoup de choses (Accent)

(22) oui moi ce que je veux **c- c- euh c-** c'est + XXX quand je me trouve dans des problèmes de pas les résoudre (Alcoolique)

Théoriquement, la possibilité de trouver une incise (quelle qu'elle soit) entre le *répétable* et le *répété* n'est pas affectée par le type de répétition (simple ou multiple). C'est ce que nous vérifions en comparant la répartition des incises en tenant compte de la proportion des types de répétitions²⁸.

Le tableau suivant propose une répartition des séquences contiguës et non contiguës des répétitions d'amorces en fonction du type de répétition :

	Séquences contiguës		Séquences non contiguës		Total Effectifs
	Effectifs	Pourcentage	Effectifs	Pourcentage	
Simple	118	86,1%	9	69,2%	127
Double et Triple	19	13,9%	4	30,8%	23
TOTAL	137	100%	13	100%	150

Tableau 6. Place des répétitions d'amorces contiguës et non contiguës en fonction du type de répétition

Les répétitions de séquences produites en contiguïté sont largement majoritaires : elles représentent 91,3% (137 sur 150) des 150 répétitions d'amorces.

Ceci veut dire que la répétition d'amorces s'accompagne rarement d'incises quelles qu'elles soient. Ce résultat diffère sensiblement de ce que

²⁸ La comparaison de la répartition des séquences contiguës à celle des séquences non contiguës est non significative (Khi2 = 2,61 ; d.d.l. = 1).

nous avons observé sur les amorces non répétées (Pallaud, 2002) puisque dans ce cas-là une fois sur cinq le phénomène d'amorce était escorté d'une incise. La répétition, dans le cas d'amorces, vient, pourrait-on dire, prendre la place d'une incise. Cette hypothèse nous semble confortée par l'analyse sur les amorces répétées à répétition multiple.

En effet, un autre résultat pour lequel nous ne fournissons pas de tableau concerne la répartition des répétitions d'amorces en fonction du nombre d'éléments contenus dans le répétable (*NET*) et de la contiguïté des séquences. Lorsque ce nombre est supérieur à 1, toutes les répétitions d'amorces sont produites en contiguïté, qu'il s'agisse de répétitions simples ou multiples.

Les répétitions d'amorces à répétition unique sont prépondérantes, et ce, quelle que soit la succession des séquences : 86,1% en contiguïté et 69,2% en non contiguïté. Mais lorsque $NR \geq 2$, la part des séquences non contiguës (23,1% pour $NR = 2$ et 7,7% pour $NR = 3$) est supérieure à celle observée pour les séquences contiguës (respectivement 30,8% et 13,9%). Étant donné le peu d'occurrences trouvées pour les séquences non contiguës à répétition multiples (4 cas), ce dernier résultat, validé dans le cadre de la présente étude, est à considérer avec précaution ; il serait certainement nécessaire de reconduire l'examen sur un recueil d'amorces répétées plus vaste afin de pallier cette faiblesse des effectifs.

3.1.5. Point de vue qualitatif sur les incises dans ces amorces répétées

Parmi les 13 répétitions d'amorces dont les séquences sont non contiguës, on constate que pour 10 d'entre elles l'élément inséré ne constitue ni une marque d'hésitation, ni une pause, ni encore un terme d'édition.

Dans le premier cas suivant, les éléments sont insérés²⁹ entre deux répétitions et suivront le dernier répété :

- (23) des trucs comme ça quoi + qui sont {**e- c- pour moi c-**} pour moi c'est pas malho- (Femmes)

Dans le deuxième cas proposé, l'élément inséré a précédé le répétable :

- (24) réserver à l'après-midi des activités euh + d'éveil ce qu'on appelle l'éveil ou euh {**d- ou d-**} d'E.P.S. (Instit)

Le programme informatique utilisé ici est fondé sur la détection puis l'extraction d'un motif récurrent contenant une amorce. Sont alors placées

²⁹ Éléments soulignés.

dans la colonne « occurrence » uniquement les répétitions d'amorces. Les autres marques de *réparation*³⁰ (par exemple, les autocorrections du genre *le la*) éventuellement présentes, et donc dans ce cas-là externes à l'« occurrence », sont renvoyées dans les colonnes « contextes gauche » et/ou « droit ».³¹

Il s'agit donc pour nous de reconstituer l'ensemble du syntagme au sein duquel se trouve une répétition d'amorces afin de laisser apparaître ces imbrications de marques de l'oral ; l'extrait ci-dessous constitue un exemple de syntagme à *réparations* multiples combinant autocorrection, répétitions et amorces :

- (25) il se passe quelque chose que tu n'as pas prévu et que qui te {de- qui de-} qui te déstabilise (Voyage)
que
qui te de-
qui de-
qui te déstabilise³²

Concernant les répétitions d'amorces, ces combinaisons des marques de l'oral ou « multi-réparations » ne sont évidemment pas réservées aux séquences non contiguës puisqu'on relève également des cas dans les séquences contiguës (26). On les retrouve aussi bien dans les répétitions d'amorces que dans les bribes portant sur du lexique achevé (27).

Dans l'exemple ci-dessous, la répétition d'amorces « p- p- » livrée en contiguïté, est encadrée par une autre répétition en « ça » :

- (26) parce que ça p- p- ça peut poser des problèmes avec la population locale (VoyagesONG)
parce que ça p-
p-
ça peut poser des problèmes avec la population locale

Dans cet autre exemple, la répétition de la préposition « dans » est combinée à une autocorrection du possessif « son » en « ses ». Nous considérons ici qu'il s'agit d'une répétition « béquille » car elle semble n'être présente que pour accompagner l'élément qui subit une autocorrection, ici la modification du nombre du possessif.

- (27) elle était très + très précise {dans son dans} ses projets (Femmes)

³⁰ Terme emprunté à Martinie (2000).

³¹ L'occurrence renvoyée par le programme informatique est placée entre accolades pour les exemples (23) à (25).

³² Cette disposition correspond à la « mise en grille » développée par Blanche-Benveniste (1987).

3. 2. Analyse morpho-syntaxique

3.2.1. Morphologie

Outre l'analyse en grands constituants syntaxiques réalisée (cf. supra point 2.2.2.), nous avons également procédé à une catégorisation morphologique des éléments amorcés répétés. Dans un souci de clarté et de simplification, nous désignons par répétitions d'amorces lexicales, les répétitions faisant intervenir des amorces de mots pleins³³, c'est-à-dire les mots ayant une charge lexicale forte, et par répétitions d'amorces grammaticales, les répétitions faisant intervenir des amorces de mots-outils (morphèmes grammaticaux³⁴, « petits mots »³⁵). Les amorces de noms, verbes, adjectifs et adverbes sont ainsi regroupées dans les amorces lexicales et les amorces de conjonctions, déterminants, pronoms, prépositions et verbes auxiliaires dans les amorces grammaticales.

3.2.1.1. Types d'unités lexicales

Les exemples (28) et (29) correspondent à des répétitions d'amorces lexicales :

- (28) et il faudra qu'on + qu'on remotive les mecs et que on le **re-re** revendique encore une fois sur le cahier (Navale)
- (29) parce que **gé- gé-** généralement on on dit qu'on est portugaise bon les gens ils s'en foutent (Portugal)

et les exemples (30) et (31) à des répétitions d'amorces grammaticales :

- (30) avis de dissolution de sociétés par voie de fusion-absorption + qui + paraîtra + obligatoirement dans l'un délai + euh prévu également **pa-pa-** par réglementation (LaHolding)
- (31) mais pendant la guerre **ce- ce-** cette fameuse de guerre on n'avait plus de maire (LaPoste)

La répartition des répétitions d'amorces en fonction du statut morphologique du répétable est proposée dans le tableau ci-dessous.

³³ Terme retenu dans Morel & Danon-Boileau (1998).

³⁴ Terme retenu dans Martinie (2000).

³⁵ Terme retenu dans Pallaud (2002).



	Répétitions d'amorces		Répétitions		Amorces	
	Effectifs	Pourcentage	Effectifs	Pourcentage	Effectifs	Pourcentage
Mots-outils	100	66,7%	14 594	91,3%	151	30,1%
Mots pleins	46	30,7%	1 391	8,7%	350	69,9%
Indéterminé	4	2,6%	-	-	-	-
TOTAL	150	100%	15 985	100%	501	100%

Tableau 7. Répartition des répétitions d'amorces en fonction du statut morphologique du répétable

- **Comparaison des amorces répétées par rapport aux répétitions en général** : les répétitions d'amorces touchent plus fréquemment les morphèmes grammaticaux (66,7%)³⁶ que le lexique (30,7%), qu'il s'agisse d'amorces répétées ou de répétitions en général. Dans ce dernier cas, la différence est encore bien supérieure : les éléments grammaticaux font massivement l'objet de bribe (90%). On vérifie d'ailleurs que la répartition des amorces répétées selon les catégories mots-outils et mots pleins est très significativement différente de celle qui est observée pour les répétitions en général ($\text{Khi}^2 = 92,72$; d.d.l. = 1 ; $p < 0,001$). Zellner (1992 : 3) avait déjà mis en évidence (très significativement) que, quel que soit le taux de dysfluences (amorces et répétitions), chez les locuteurs non bègues, « les mots fonction³⁷ monosyllabiques – qui permettent de structurer le discours – ont tendance à subir plus d'accidents relativement aux autres mots ». Si les mots pleins interviennent dans 30% des cas de répétitions d'amorces, ils ne représentent toutefois que 8,7% des répétitions en général. Les mots lexicaux sont donc plus sensibles dans la production de l'oral à une interaction de la répétition et du phénomène d'amorce. Cette hypothèse a pour corollaire que les morphèmes grammaticaux sont très massivement concernés par le phénomène de bribe et le sont moins lorsque lui est associé le phénomène d'amorce.
- **Comparaison des amorces répétées par rapport aux amorces en général** : la répartition des amorces dans ces deux catégories (lexique et morphèmes grammaticaux) est inverse de celle qui est observée

³⁶ Les statistiques fournies par le logiciel *Cordial* (Correcteur grammatical et analyseur de la langue française; *Synapse Développement*, version 7) montrent que la catégorie « Mots-outils » constitue 49,61% du total des mots. Dans les textes de genre théâtral, cette proportion est de 48,77%. Il nous semble donc légitime de dire que la catégorie de mots-outils est plus touchée par la répétition d'amorces.

³⁷ Mots fonction : déterminants, pronoms, conjonctions, prépositions, interjections (Zellner, 1992 : 3).

lorsqu'il y a phénomène de bribe. Dans les amorces de mots, les morphèmes grammaticaux sont moins concernés et ne représentent que le tiers des amorces (30%). C'est exactement le contraire lorsque se surajoute un phénomène de bribe. Ces résultats confirment la comparaison avec les répétitions en général : il y a une forte interaction entre ces deux types d'hésitation que représentent les amorces de mots et les bribes.

3.2.1.2. Catégories grammaticales

Les travaux de Cappeau (1998) sur les bribes à l'oral incitent à affiner notre analyse afin de déterminer les relations qui peuvent s'établir entre le phénomène de répétition d'amorces et la catégorisation grammaticale, notamment, ce que nous nommons les mots-outils (pronoms, déterminants, prépositions, conjonctions et locutions conjonctives) et les mots pleins (verbes, noms, adjectifs, adverbes et locutions adverbiales).

Les deux tableaux suivants proposent une répartition des répétitions d'amorces en fonction des catégories grammaticales :

- Mots-outils :
 - Pronoms
 - (32) mais c'est parce que j- j- j'avais un travail fixe (Alcoolique)
 - Déterminants
 - (33) parce que vous êtes u- u- une malade + vous pouvez crier (Alonso)
 - Prépositions
 - (34) Pierre Leroux qui était a- a- avec moi enfin dans dans l'expédition (Expédition)
 - Conjonctions
 - (35) mais va être ob- m- m- mais euh chaque pays (Olive)

Mots-outils Catégorie grammaticale	Répétitions d'amorces		Répétitions en général	
	Effectifs	Pourcentage	Effectifs	Pourcentage
Pronoms	61	61%	5 181	35,5%
Déterminants	22	22%	6 057	41,5%
Prépositions et loc. prép.	9	9%	1 897	13%
Conjonctions et loc. conj.	5	5%	1 021	7%
Verbes auxiliaires	3	3%	146	1%
Autres	-	-	292	2%
TOTAL	100	100%	14 594	100%

Tableau 8. Répartition par catégorie grammaticale des répétitions d'amorces grammaticales

Deux catégories grammaticales sont majoritairement concernées quelle que soit la « nature » de la répétition (répétition d'amorces *versus* répétitions de mots non tronqués) : il s'agit des pronoms et des déterminants (respectivement 83% et 77% pour ces deux types de répétitions). Viennent ensuite les prépositions et locutions prépositionnelles (respectivement 9% et 13%), ainsi que les conjonctions et locutions conjonctives (respectivement 5% et 7%). Des informations sur les fréquences de ces catégories grammaticales permettraient de savoir si elles sont sur-représentées dans nos recueils de répétitions et d'amorces. Nous n'en disposons pas actuellement. Le tableau montre également une répartition significativement³⁸ différente de ces catégories grammaticales selon la « nature » de la répétition. C'est particulièrement le cas pour les catégories pronoms et déterminants où la répartition s'inverse. Dans le cas des répétitions d'amorces, la proportion des pronoms est supérieure à celle des déterminants (61% contre 22%) alors que, dans les répétitions en général, le pourcentage des pronoms est inférieur à celui des déterminants (35,5% contre 41,5%). Nous avons observé (Pallaud, 2003) que lorsqu'il y avait amorce sur un élément d'un constituant Sujet, la catégorie grammaticale affectée dans la moitié des cas (53%) était celle des pronoms. Venait ensuite celle des déterminants (13,1%). Cette répartition est plus proche de celle que nous observons dans les répétitions d'amorces³⁹ que de celle des répétitions en général.

- Mots pleins

- Verbes

(36) non vous avez ré- ré- ré- répondu (Yolande)

³⁸ $\text{Chi}^2 = 24,86$; d.d.l. = 1 ; $p < 0,001$.

³⁹ Différence non significative : $\text{Chi}^2 = 1,16$; d.d.l. = 1.

Amorces de mots et répétitions dans les énoncés oraux

- Adverbes
(37) je travaillais bien c'était bien + s- s- surtout à l'époque (Alcoolique)
- Noms
(38) à l'intérieur euh de ces de ces de ce re- re- développement géographique on on a un on indique les choses d'une manière on regroupe les choses qui vont en- (Journal)
- Adjectifs
(39) c'était un monstre il était ho- ho- horri- horrible (Pingouin)

Mots pleins Catégorie grammaticale	Répétitions d'amorces		Répétitions en général	
	Effectifs	Pourcentage	Effectifs	Pourcentage
Verbes	26	56,5%	153	11%
Adverbes et loc. adv.	11	23,9%	723	52%
Noms	6	13,1%	167	12%
Adjectifs	3	6,5%	348	25%
TOTAL	46	100%	1 391	100%

Tableau 9. Répartition par catégorie grammaticale des répétitions d'amorces lexicales

Les effectifs étant faibles dans cette catégorie d'amorces répétées, nous n'indiquons que des tendances qui seraient à vérifier sur des exemples plus nombreux. Plus de la moitié des amorces lexicales répétées sont des amorces de verbes (56,5%) ; c'est loin d'être le cas pour les répétitions en général puisque ce pourcentage tombe à 11% et représente, pour cette « nature » de répétition, la catégorie grammaticale (pour les mots pleins) la moins représentée. La part des adverbes représente seulement un quart des répétitions d'amorces lexicales (23,9%) mais plus de la moitié des répétitions lexicales non amorcées (52%). Quant aux noms, qu'ils soient tronqués ou pas, lorsqu'ils sont pris dans une répétition, ils représentent un peu plus d'un dixième des effectifs. La proportion des adjectifs est très faible dans les amorces répétées (6,5%) alors que les adjectifs représentent un quart des répétitions en général. On constaterait donc une forte dépendance entre la « nature » de la répétition et la catégorie grammaticale des mots pleins.

3.2.2. Syntaxe

Si l'on couple à cette première distinction basée sur le type d'unité lexicale mis en jeu dans les répétitions d'amorces – lexicales *versus*

Sandrine HENRY & Berthille PALLAUD

grammaticales – une composante syntaxique en procédant à une Analyse en Grands Constituants, on obtient la répartition suivante :

	Répétitions d'amorces lexicales		Répétitions d'amorces grammaticales		Indéterminé	
Sujet	3	6,5%	68	68%	-	-
Verbe	21	45,7%	5	5%	-	-
Objet	16	34,8%	23	23%	1	25%
Indéterminée	6	13%	4	4%	3	75%
TOTAL	46	100%	100	100%	146	100%

Tableau 10. Répartition des répétitions d'amorces en fonction du type d'unité lexicale et de la place syntaxique

Le tableau 10 montre une forte dépendance entre le statut morphologique des amorces répétées et leur place syntaxique.

Lorsqu'il s'agit d'amorces lexicales, les répétitions se produisent de manière préférentielle sur les places syntaxiques Verbe et Objet, et ce, dans respectivement 45,7% et 34,8% des cas. Lorsqu'il s'agit d'amorces grammaticales, la répartition est tout à fait différente (5% pour le site syntaxique Verbe et 23% pour le site Objet). Le site Sujet, quant à lui, semble peu propice à la répétition d'amorces lexicales (6,5% des cas) alors que lorsqu'il s'agit de répétitions d'amorces grammaticales, ce site est majoritairement concerné (68%).

4. Discussion et conclusion

Les amorces répétées constituent un sous-groupe parmi les phénomènes d'achoppement à l'oral. Si elles ne constituent qu'un pour cent des répétitions, leur proportion dans les amorces est plus importante : 2,46%. Leur analyse tant du point de vue des amorces que des répétitions confirme l'intérêt de la distinction entre ces deux catégories de trébuchement. On retrouve dans ces répétitions d'amorces des caractéristiques propres soit aux amorces soit aux répétitions.

4.1. Le point de vue des amorces

Comme le laissaient prévoir les analyses sur les phénomènes d'amorces de mots, les trois catégories d'amorces répétées (complétées, modifiées ou inachevées) ne sont pas identiques. Les amorces complétées

demeurent les plus nombreuses. Cependant, lorsqu'il y a répétitions d'amorces, celles-ci sont beaucoup plus souvent complétées. Il existe une dépendance significative entre la répétition et le phénomène d'amorce. Les études de Cappeau (1998) avaient signalé, en ce qui concerne l'attribution du genre, la même prépondérance de la complétude (par rapport à la modification) dans le cas des amorces de syntagme (ou bribes). La répétition, loin d'être une anticipation d'erreur, annoncerait plutôt le choix correct de genre. Cette dépendance entre répétitions et amorces, qui se traduit par une proportion plus importante de la répétition des amorces complétées, suggère que ces répétitions d'amorces sont équivalentes à des pauses remplies qui constitueraient un piétinement sur une même place syntaxique suivi d'une poursuite de l'énoncé sur l'axe syntagmatique.

La répartition des répétitions d'amorces s'avère significativement dépendante de la variable constituant syntaxique (Sujet, Verbe ou Objet). Les amorces Sujet, qui ne représentaient que 12,9% des amorces en général, constituent, lorsqu'elles s'accompagnent de répétition, la moitié des amorces répétées. Ce renversement de proportion tient sans doute au fait que parmi les amorces Sujet, les deux tiers sont des achoppements qui seront complétés. L'examen de l'interaction entre le phénomène de répétition et celui d'amorce montre que la répétition vient accentuer la tendance à compléter le mot en position Sujet par rapport aux autres constituants ; par contre, les amorces répétées se répartissent, semble-t-il, au hasard lorsque l'amorce est modifiée ou inachevée, ce qui n'est pas le cas lorsque l'amorce n'est pas répétée.

4.2. Le point de vue des répétitions

En ce qui concerne la structure de la répétition, l'analyse de certains traits nous permet de dégager quelques tendances. Tout d'abord, à l'instar des répétitions d'éléments non tronqués, les répétitions d'amorces à répété unique et dont le répétable contient un seul élément sont majoritaires (7 sur 10), et la probabilité d'avoir une répétition diminue au fur et à mesure que le nombre de répétés augmente. Répétitions de lexique achevé et répétitions d'amorces diffèrent en ceci, que la part des amorces à répétition double est significativement plus importante lorsqu'il y a amorce de mot. Nous avons également pu constater, que les répétitions d'amorces sont données essentiellement en contiguïté (9 sur 10), c'est-à-dire sans incise. Lorsque le répétable est multiple, il n'y a aucune incise. Ce n'est pas le cas des amorces en général, où une fois sur cinq, l'amorce est escortée d'un énoncé parenthétique. Ce constat d'un effet inhibiteur d'incise par la répétition sur

l'interruption de mot vient également soutenir l'hypothèse que nous avons émise précédemment, à savoir que la répétition est proche d'une pause remplie.

Sur le plan morphologique, nous constatons une forte interaction entre ces deux types d'achoppement que constituent les amorces et les répétitions. Deux tiers des amorces répétées sont des amorces de mots-outils (mots grammaticaux). Cette prévalence des mots-outils sur les mots pleins (mots lexicaux) dans les répétitions d'amorces se retrouve également dans les répétitions en général, et ceci de façon significativement accentuée avec une proportion de 91,3%. En revanche, lorsqu'il s'agit d'amorces en général, cette tendance s'inverse parfaitement : deux tiers d'entre elles sont lexicales et seulement un tiers grammaticales. Il apparaît donc que la répétition ne touche qu'une partie des amorces : celle qui affecte les mots-outils. Les pronoms et les déterminants sont majoritairement concernés (plus de 75%, que ce soit les amorces répétées ou les répétitions en général).

Il existe également une importante différence de rang entre ces deux catégories grammaticales, selon que l'élément pris dans la répétition est amorcé ou pas. En effet, nous remarquons que les répétitions d'amorces grammaticales font intervenir majoritairement des pronoms (6 amorces répétées de mots-outils sur 10) par rapport aux déterminants. Cette proportion correspond à ce qui a été observé avec les amorces de mots Sujet, mais ne correspond pas du tout à ce qui est observé dans les répétitions en général où les répétitions grammaticales non amorcées portent essentiellement sur des déterminants (41,5%), et non sur des pronoms.

Les répétitions d'amorces lexicales sont peu nombreuses : 46 au total, aussi n'est-il pas possible de parler d'autre chose que de tendance. Ces répétitions font intervenir majoritairement des verbes (56,5%), tandis que les répétitions lexicales non amorcées portent essentiellement sur des adverbes (52%). De plus, à l'exception de la catégorie grammaticale « Noms » qui présente une proportion semblable quelle que soit la « nature » de la répétition (répétition d'amorces *versus* répétitions de mots non tronqués), toutes les autres catégories ne sont pas également concernées.

Les répétitions d'amorces lexicales se produisent de manière préférentielle sur les sites Verbe et Objet (respectivement 45,7% et 34,8%), tandis que pour les répétitions d'amorces grammaticales, c'est le site Sujet qui est plus fortement concerné (68%).

Si sur le plan structural, les répétitions d'amorces ne semblent pas présenter de différence majeure avec les répétitions d'éléments non tronqués, en revanche, sur le plan morphologique, selon la « nature » de la répétition, les catégories grammaticales ne sont absolument pas sollicitées dans les mêmes proportions.

Le signe linguistique a la capacité de renvoyer à lui-même ce qui est désigné sous le terme d'autonymie. Existe-t-il pour autant un métalangage repérable dans le langage et support de cette réflexivité ce qui permettrait de le décrire ? Bonnet et Tamine-Gardes (1984) et Rey-Debove (1988) revendiquent la prudence sur cette notion de métalangage en tant que fonctionnement repérable et caractérisable. L'autonymie n'est pas toujours facile à déterminer et présente le risque de « laisser entendre que la réflexivité du langage est un phénomène marginal dans la langue alors qu'elle en est une des propriétés essentielles » (Bonnet et Tamine-Gardes, 1984, p 136).

Les amorces de mots comme les répétitions soulèvent des questions sur leur statut linguistique. Doit-on les considérer comme des échecs de l'énonciation au regard d'une fluidité verbale au demeurant rarement (et peut-être jamais) constatée ? Ou bien peut-on leur accorder le statut d'autonymes ? C'est là le point de vue que nous adoptons. Accompagnés ou non de ces incises si précisément décrites par Authier-Revuz (1996), ces achoppements nous semblent faire partie de « ces mots qui ne vont pas de soi ».

Références

- Authier-Revuz, J. (1996). *Ces mots qui ne vont pas de soi. Boucles réflexives et non coïncidences du dire*. Paris : Larousse.
- Blanche-Benveniste, C., & Jeanjean, C. (1987). *Le français parlé. Transcription et édition*. Paris : Didier Érudition.
- Blanche-Benveniste, C. (1987). Syntaxe, choix du lexique et lieux de bafouillage. *DRLAV*, 36-37, 123-157.
- Blanche-Benveniste, C. (1997). Transcription et technologie. *Recherches sur le français parlé*, 14, 87-100.
- Bonnet, C., & Tamine-Gardes, J. (1984). *Quand l'enfant parle du langage*. Paris : Édition Mardaga.
- Candéa, M. (2000). *Contribution à l'étude des pauses silencieuses et des phénomènes dits « d'hésitation » en français oral spontané. Étude sur un corpus de récits en classe de français*. Thèse d'Etat, Université Paris III (Sorbonne Nouvelle).
- Cappeau, P. (1998). Quelques mots sur quelques bribes liées au genre. In M. Bilger, K. Van den Eynde & F. Gadet, (Éds), *Analyse linguistique et approches de l'oral*.

Sandrine HENRY & Berthille PALLAUD

- Recueil d'études offert en hommage à Claire Blanche-Benveniste* (pp. 301-311). Paris : Peeters, Leuven.
- Clark, H., & Wasow, T. (1998). Repeating words in spontaneous speech, *Cognitive Psychology*, 37, 201-242.
- Cosnier, J., & Kerbrat-Orecchioni, C. (Éds), (1987). *Décrire la conversation*. Lyon : Presses Universitaires de Lyon.
- Den, Y. (2001). Are word repetitions really intended by the speaker ? *In Proceedings of Disfluency In Spontaneous Speech (DISS'01)* (pp. 25-28). August 29-31, 2001, University of Edinburgh, Scotland.
- Gadet, F. (1997). *Le français ordinaire*. Paris : Armand Colin.
- Henry, S. (2002a). Étude des répétitions en français parlé spontané pour les technologies de la parole. *Actes de la 6^{ème} Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues. (RECITAL'02)* (pp. 467-476). Nancy, France.
- Henry, S. (2002b). Quelles répétitions à l'oral ? Esquisse d'une typologie. 2^{ème} Journées de linguistique de corpus, 12-14 septembre 2002. Lorient.
- Jeanjean, C. (1984). Les ratés c'est fa- fabuleux. Étude syntaxique des amorces et des répétitions. *Linx*, 10, 171-177.
- Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition* 14, 41-104.
- Levelt, W.J.M. (1989). *Speaking. From intention to articulation*. Cambridge : MIT Press.
- Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, 22, 1-75.
- Lickley, R.J. (2001). Dialogue moves and disfluency rates. *Disfluency In Spontaneous Speech (DISS'01)* (pp. 93-96). August 29-31, 2001, University of Edinburgh, Scotland.
- Martinie, B. (2000). Remarques sur la syntaxe des énoncés réparés en français parlé. *Recherches sur le français parlé*, 16, 189-206.
- Mc Allister, J., Cato-Symonds, S., & Johnson, B. (2001). Listeners'ERP responses to false starts and repetitions in spontaneous speech. *In Proceedings of Disfluency In Spontaneous Speech (DISS'01)* (pp. 65-68). August 29-31, 2001, University of Edinburgh, Scotland.
- Morel, M.-A., & Danon-Boileau, L. (1998). *Grammaire de l'intonation : l'exemple du français*. Paris : Ophrys.
- Nooteboom, S.G. (2001). Differences sources of lexical bias and overt self-corrections. *Disfluency In Spontaneous Speech (DISS'01)* (pp. 21-24). August 29-31, 2001, University of Edinburgh, Scotland.
- Pallaud, B. (1999). Lapsus et phénomènes voisins dans la langue parlée : problèmes d'identification. *Recherches sur le français parlé*, 15, 1-33.
- Pallaud, B. (2001). Les lapsus : des pierres dans le champ linguistique. In M. Arrivé & C. Normand (Éds). *Linguistique et psychanalyse*. (p.47-66). Colloque de Cerisy-la-Salle, 1-8 septembre 1998, In Presse.
- Pallaud, B. (2002). Les amorces de mots comme faits autonymiques en langage oral, *Recherche Sur le Français Parlé*, 17, 79-102.
- Pallaud, B. (2003a). Erreurs d'écoute dans la transcription de données orales. *Actes du colloque Transcription de la Parole normale et pathologique*. Tours, 8-9 décembre 2000. *Revue Parole*, 22-23-24, 267-294.

Amorces de mots et répétitions dans les énoncés oraux

- Pallaud, B. (2003b). Achoppements dans les énoncés de français oral et sujets syntaxiques. In J.-M. Merle (Éd), *Le Sujet* (pp. 91-104). Paris : Ophrys.
- Rey-Debove, J. (1988). À la recherche de la distinction oral/écrit. In N. Catach (Éd), *Pour une théorie de la langue écrite* (pp. 116-127). Paris : Éditions du CNRS.
- Roubaud, M.-N. (2003). Du bon usage des amorces dans la transcription. *Recherche sur le français parlé*, 18, 243-256.
- de Saussure, F. (1972). *Cours de linguistique générale*. Paris : Payot (1922).
- Zellner, B. (1992). Le bé- bégayage et euh ... l'hésitation en français spontané. *Actes des 19^{èmes} Journées d'Études sur la Parole* (pp. 481-487). Bruxelles.

Compte-rendu de l'ouvrage de
Norbert Dittmar :
*Transkription. Ein Leitfaden mit Aufgaben
für Studenten, Forscher und Laien*¹

François MULLER
Université de Paris

Norbert Dittmar est professeur de linguistique allemande et de linguistique générale au Département de Lettres et Sciences Humaines de l'Institut de Philologie allemande et néerlandaise de l'Université Libre (FU) de Berlin. Il a publié déjà plusieurs ouvrages de sociolinguistique (*Soziolinguistik*, 1973 ; *Grundlagen der Soziolinguistik* [Fondements de la sociolinguistique], 1997 ; *The discourse marker also in L2 acquisition*.)

Transkription est une présentation générale des systèmes de transcription en usage pour la transcription de la langue orale à des fins d'archivage et de recherche. La partie centrale de cette étude est constituée par la présentation détaillée des principaux systèmes de transcription actuellement utilisés en Allemagne par les sociolinguistes et plus généralement par les chercheurs en sciences humaines (ethnologie, didactique, apprentissage des langues, etc.) en vue de la représentation graphique de systèmes de signes oraux et non verbaux dans la communication de la vie quotidienne et professionnelle. Le souci constant de concilier un exposé théorique de haut niveau et un ensemble d'informations pratiques, notamment informatiques (CD, programmes et sites Internet) constitue sans conteste le mérite et la grande originalité de l'ouvrage, qui comprend aussi un glossaire des principaux termes techniques utilisés, ainsi qu'une bibliographie bien documentée. L'ouvrage

¹ Norbert Dittmar. *Transkription. Ein Leitfaden mit Aufgaben für Studenten, Forscher und Laien*. - Qualitative Sozialforschung 10 - Lesken + Budrich, Opladen, 2002. - 247 pages. ISBN 3-8100-2650-6 - 15,50 euros.

François Muller

certain nombres d'exercices pratiques dont les solutions se trouvent... sur la « home page » de l'auteur (http://www.userpage.fu-berlin.de/~nordit/HP_LINK:Transkriptionsbuch).

L'ouvrage débute par une réflexion sur le besoin – ou la nécessité – de vouloir ou de pouvoir disposer de traces écrites de ce qui a été dit, car les hommes savent depuis la plus haute antiquité que si les paroles volent, les écrits, eux, restent. Cette réflexion sur le primat du visuel sur l'auditif sert d'introduction générale à l'ensemble du texte ; elle porte sur la constitution de la mémoire collective, sur le haut degré d'authenticité dont bénéficie a priori l'écrit par rapport à l'oral et sur l'aspect contraignant, voire contractuel, du « document » écrit.

La transcription est le prérequis indispensable à toute étude scientifique de la communication orale, car elle fournit sur un support matériel (papier) ou virtuel (informatique) la matière première de la langue parlée dans ses multiples manifestations. L'auteur cherche ensuite à cerner de plus près les spécificités de la langue parlée en les opposant à celles de la langue écrite et esquisse à grands traits l'état actuel de la recherche en ce domaine.

Le chapitre suivant cherche à établir dans quelle mesure et dans quelles conditions une transcription peut être considérée comme une démarche scientifique. En d'autres termes, comment établir la validité de la pratique du transcripteur ? En effet, quelles que soient par ailleurs leurs orientations théoriques, toutes les disciplines des sciences humaines ont impérieusement besoin d'une théorie de la pratique sociale qui inclut les systèmes de transcription eux-mêmes. Mais comme une telle théorie ne peut se fonder que sur de vastes corpus de comportements communicatifs - au sens large - existant réellement, la constitution de cette documentation de processus communicatifs devient un problème épineux. Arrivé à ce point, on s'attend à ce que l'auteur évoque alors le problème bien connu des spécialistes d'herméneutique en général et de la critique des textes en particulier, c'est-à-dire ce que, à la suite de Schleiermacher, on a appelé la question du cercle herméneutique. En effet, depuis qu'Ehlich et Rehbein (1976) ont montré que toute transcription est aussi une interprétation, la problématique du cercle herméneutique s'imposait plus ou moins, ne serait-ce qu' en filigrane. Mais il n'en est rien, et c'est un peu dommage, car le rapprochement eût été intéressant, tant pour la problématique de la transcription elle-même que pour la contribution, même réduite à un éclairage nouveau, que cette problématique aurait pu apporter à la compréhension générale de la question du cercle herméneutique.

Le quatrième chapitre pourrait s'intituler : « Dis-moi comment tu parles et je te dirai qui tu es. » Dans l'impossibilité de pouvoir rendre compte de la charge affective, voire émotive – ce que l'auteur appelle « l'aura » – que recèlent certaines voix, on en est réduit à restituer leur couleur, c'est-à-dire leurs variétés sociologiques et/ou dialectales. C'est ainsi que Jane Austen (*Emma*, 1806), mais aussi Fontane, Tucholsky, et en France Balzac (on pourrait ajouter Maupassant) ont cherché à décrire le plus fidèlement possible les écarts essentiellement (mais pas uniquement) phonétiques par rapport à la langue standard de leur époque. Signalons que l'auteur s'est ici limité aux romans, omettant les parlers hauts en couleurs des servantes et valets de Molière, près d'un siècle et demi avant Jane Austen. Cette transcription peut être qualifiée de « littérale », car elle se contente d'utiliser les « lettres » de l'alphabet, sans recours à aucun signe diacritique. Klein et Schütte (*Transkriptionsrichtlinien für die Eingabe in DIDA*, 2000) considèrent ce système élémentaire de transcription comme une variante de la « translittération ». D'autres parlent ici de « dialecte visuel ». J.A. Edwards (*Transcription of Discourse*, 1992) est très critique pour ce sabir qu'elle appelle « an impressionistic extension of English spelling. » Sans aller aussi loin, N. Dittmar, tout en ayant conscience que la translittération est un pis-aller, c'est-à-dire un compromis acceptable entre la transcription orthographique classique et les transcriptions phonétiques dites « scientifiques », regrette tout de même que la translittération n'ait pas fait l'objet d'une uniformisation sur le plan international, comme ce fut le cas pour l'alphabet phonétique international de L'API (Association Phonétique Internationale), sur lequel l'auteur s'étend longuement. Le chapitre se termine par une brève présentation de deux autres systèmes de transcription conçus spécialement en vue du traitement informatique. Le premier, le PDL (*Pidgin- Deutsch-Lautschrift*), a été élaboré en 1974 par des chercheurs de l'université de Heidelberg. Le principe en est le suivant : pour ne pas avoir à recourir à la multitude de signes et de symboles de l'alphabet phonétique international, ceux-ci ont été remplacés par des lettres latines, ce qui facilite grandement l'implémentation. SAMPA (*Speech Assessment Methods Phonetic Alphabet*) a été mis au point dans les mêmes buts, mais par des chercheurs anglo-saxons (cf. Gibbon, Moore, Winski : *Handbook of Standards and Resources for Spoken Language Systems*, 1997). Pour le codage, SAMPA, à la différence de PDL, a recours au code ASCII.

Le coeur du livre, au dire même de l'auteur, est le grand chapitre cinq, qui compte une centaine de pages, soit près de la moitié de l'ouvrage et qui présente sept systèmes différents de transcription. En prolégomènes à cette

François Muller

présentation exhaustive, N. Dittmar cite les six maximes formulées par J.W. Du Bois (*Transcription design principles for spoken discourse research*, 1991, et J.W. Du Bois et alii : *Discourse transcription - Santa Barbara papers in linguistics*, vol. 4, 1992) :

1. Définis des catégories de transcription qui soient parfaitement adaptées aux buts de ta recherche.
2. Fais en sorte que ton système soit simple (c'est-à-dire aussi facilement déchiffrable que possible).
3. Choisis des caractères stables et bien définis.
4. Applique le principe d'économie dans l'inventaire des signes.
5. Module ton système afin qu'il puisse s'adapter à différentes fonctions.
6. Organise ton système en vue d'une application informatique la plus simple possible.

L'auteur, en accord avec A. Deppermann (*Gespräche analysieren. Eine Einführung in konversationsanalytische Methode*, 1999), insiste sur un point qui lui tient manifestement à coeur, car il y revient à plusieurs reprises, c'est-à-dire sur la première maxime : trop de chercheurs utilisent en effet des systèmes trop élaborés qui de ce fait compliquent inutilement le travail en aval. Il oppose ainsi une transcription élémentaire (*Basistranskript*) à une transcription fine (*Feintranskript*).

Avant de passer à la description proprement dite des sept systèmes retenus, l'auteur présente une grille de lecture qui lui servira à les évaluer. Pour ce faire, il a retenu sept critères :

1. Le « design » (c'est-à-dire l'architecture ou la configuration générale), qui comprend trois points :
 - 1.1. la présentation d'ensemble : Soumise à la linéarisation de l'énoncé, la transcription est réduite à choisir entre (a) la présentation sous une forme analogue à celle d'une partition musicale, (b) l'agencement séquentiel et (c) la présentation sous forme de colonnes. Aucune des trois façons de procéder n'est pleinement satisfaisante, bien que la première soit la plus aisée non seulement à mettre en oeuvre, mais aussi à exploiter.
 - 1.2. La translittération proprement dite : quel est l'inventaire des signes retenus ?

- 1.3. l'en-tête, dans laquelle sont données des informations générales concernant les conditions dans lesquelles a été faite la transcription.
2. Les contributions des divers participants, ce qui implique les données verbales, mais aussi communicationnelles et prosodiques (Identification des locuteurs, informations sur le début et la fin de la communication, les chevauchements, les enchaînements, etc.)
3. Unités énonciatives. Les unités minimales de communication sont : les mots, les limites de l'énoncé (début / fin), le fragment de mot (interruption), réactions de ou des interlocuteur(s). L'auteur rappelle qu'il n'est pas toujours aisé de segmenter le flux verbal, surtout chez des enfants ou des locuteurs qui maîtrisent mal la langue. Mais cette première étape n'en est pas moins décisive, car elle constitue un encodage minimal incontournable.
4. Phénomènes prosodiques : l'auteur en a retenu neuf : la structure intonative des mots, l'intonation des syntagmes, l'accent de mots, l'allongement des sons, l'intensité (*forte* VS *piano* ; *crescendo* VS *diminuendo*), le tempo (*allegro* VS *lento*) et l'accélération (*accelerando* VS *rallentando*), le rythme (combinaison de la hauteur, de l'intensité et de la durée : *staccato* ; *glissando*, *legato*), manifestations physiques qui accompagnent la parole (tousotements, rires, soupirs, etc.), et enfin les pauses (silences, hésitations, vocalisations).
5. Manifestations non verbales (proxémique, motorique, mimique, gestique, etc.) : elles ne sont pas toujours prises en compte dans la transcription.
6. Commentaires (point de vue du transcripteur) : l'auteur estime qu'il est parfois important de signaler tous les faits saillants qui peuvent surgir dans la communication et qui font sens, d'une manière ou d'une autre.
7. Signes spéciaux (« Extras ») : N. Dittmar en a sélectionné huit : la restitution phonétique, les noms propres, les traductions, les changements de langues, le signalement des passages non transcrits, les signes prévus pour autre chose que la transcription de la communication, les signes diacritiques optionnels, les signes conventionnels pour les juxtapositions de plusieurs signes particuliers.

François Muller

L'auteur entame ensuite l'examen des sept systèmes retenus : l'analyse de conversation (KA), HIAT, DIDA, le système de Du Bois, GAT, CHAT et CHILDES. Les diverses présentations suivent à peu près toutes le même canevas : historique, l'idée directrice, présentation des signes conventionnels, et se terminent toujours par un exemple de transcription.

1. L'analyse de conversation (KA)

L'auteur rappelle qu'à la fin des années soixante et au début des années soixante-dix, deux concepts étaient particulièrement répandus dans le milieu des chercheurs en sociologie: l'ethnométhodologie et « l'interaction sociale » (Goffman). Les sociologues de l'époque s'étaient fixé pour objectif d'étudier les pratiques sociales des individus dans leur vie quotidienne essentiellement sinon exclusivement sous l'angle de leurs interactions verbales. Pour ce faire, ces chercheurs ne travaillaient pratiquement que sur des transcriptions. L'accent était mis sur la structure séquentielle des contributions des différents participants. Une attention particulière était de ce fait accordée aux transitions d'une contribution à l'autre, ainsi qu'aux fonctions expressives de la parole (opposition : voix forte VS faible, mise en exergue d'un mot ou d'une expression, allongement des voyelles, aspiration, pauses, etc.). L'analyse de conversation a réussi à s'imposer en Europe essentiellement grâce à Levinson (1983). L'appréciation générale portée par l'auteur sur l'analyse de conversation est largement positive : les catégories sont bien définies, les signes sont pour la plupart alphabétiques, donc aisément utilisables. C'est ce qui explique que de nos jours encore, un grand nombre de spécialistes en communication, mais aussi de linguistes et de sociologues continuent à se servir des conventions établies par l'analyse de conversation. Une réserve cependant : l'implémentation du système tel qu'il se présente depuis le début des années quatre-vingt-dix se fait attendre.

2. HIAT

Avec « le tournant pragmatique » (Austin, Searle, Levinson) dans les années soixante-dix, l'activité langagière devint objet d'investigations

scientifiques. On se mit à archiver électroniquement les actes de paroles pour en étudier le fonctionnement. Les promoteurs de cette démarche nouvelle étaient en Allemagne Ehlich et Rehbein. Ils appréciaient le système de transcription de l'analyse de conversation, mais ils avaient rapidement discerné ses faiblesses ou tout simplement ses limites. C'est surtout l'agencement purement séquentiel qu'ils jugeaient peu opérationnel. C'est pourquoi ils optèrent pour la transcription sous forme de partition musicale, plus à même, selon eux, de rendre compte de l'interaction verbale quasi polyphonique des divers participants. Abandonnant la segmentation du discours en unités, ils préconisent des « marqueurs du flux verbal ». – Si l'on excepte DIDA (IdS, Mannheim), HIAT est sans conteste le système le plus utilisé en Allemagne, notamment par les pédagogues, les linguistes et les sociologues, ainsi que par les transpositeurs institutionnels (à l'école, à l'hôpital, au tribunal). Cinq raisons majeures expliquent, selon N. Dittmar, ce large succès :

- la présentation d'ensemble sous forme de partition offre une synopsis très claire, qui se laisse aisément déchiffrer. En effet, une portée étant attribuée à chaque participant, la synchronicité et la diachronicité de chaque contribution sont fidèlement documentées ;
- le système est très flexible : la transcription « primaire », qui fait abstraction de toute segmentation fine, peut être réalisée avec une grande économie de signes ; c'est la raison pour laquelle on l'appelle « semi-interprétative » ;
- les auteurs du système ont eu un souci constant, et ce dès la première version, d'apporter des solutions concrètes à tous les problèmes pratiques qui pouvaient se poser (par ex. pour les corrections) ;
- de même, les auteurs ont été très attentifs à l'implémentation de leur système, tant sur Macintosh que sur MS-DOS ;
- l'implémentation initiale, fondée sur des symboles essentiellement musicaux, a été progressivement abandonnée au profit d'un système plus convivial.

3. DIDA

DIDA (Diskurs-DAten-Bank) a été développé au début des années quatre-vingt-dix à l'Institut de la langue allemande (IdS) de Mannheim. Il s'agissait de mettre au point un système conforme aux dernières avancées

François Muller

de l'analyse du discours et de la conversation, afin d'exploiter par l'informatique les nombreux corpus que l'IdS avait accumulés au cours des dernières décennies. Au départ, ce système n'avait été conçu qu'à l'usage interne de l'IdS ; mais par la suite il rendit aussi d'appréciables services à d'autres chercheurs en sciences humaines. C'est ainsi par exemple que le corpus de la langue parlée dans la ville de Mannheim (cf. Kallmeyer : *Kommunikation in der Stadt*, 1994) a été recodé en DIDA. On notera que le système prend en compte des paramètres comme le tempo du flux verbal, ainsi que l'intensité de la voix. Parmi les nombreux mérites du système, N. Dittmar souligne essentiellement l'excellente iconicité des signes utilisés et l'existence d'une multiplicité de programmes d'application spécifiques, comme par exemple PRAAT pour l'étude des phénomènes prosodiques.

4. Le système de Du Bois

Le système de Du Bois (DT : Discourse Transcription) se situe dans le droit fil des réflexions initiées dès 1979 par E. Ochs. Par rapport à ses prédécesseurs et concurrents, DT présente essentiellement trois avantages :

- Du Bois et son équipe ont formulé des exigences précises en matière de lisibilité, d'iconicité et d'implémentation ;
- les auteurs prennent grand soin de distinguer la transcription proprement dite, qui est avant tout interprétative, de l'encodage, qui doit se faire en fonction du but recherché ;
- ils établissent ensuite une seconde distinction (toute scalaire) entre transcription « large », à caractère purement descriptif, et transcription « étroite », à visée plus scientifique.

5. GAT

GAT (*Gesprächsanalytisches Transkriptionssystem*), créé en 1997 par des linguistes (Selting, Auer), s'est fixé comme objectif de faire la synthèse des systèmes existants afin de mettre à la disposition des chercheurs de langue allemande un système unifié de transcription. On facilite ainsi les échanges de gros corpus non seulement au sein d'une même discipline, mais surtout entre les différentes disciplines des sciences humaines, ce qui était aussi l'un des buts des créateurs de GAT. Ce système a été conçu

avant tout pour l'étude de la langue courante dans le cadre d'une approche pragmatique de la communication. Le « design » présente les caractéristiques suivantes :

- notation séquentielle ;
- tout doit être écrit en minuscules (ce point n'est pas sans importance en allemand, où tous les noms, y compris les noms communs, s'écrivent avec une majuscule) ;
- dans l'en-tête (*Transkriptionskopf*) doivent figurer un certain nombre d'informations précises, notamment sur les conditions matérielles de la transcription.

GAT a développé de façon satisfaisante les conventions de l'analyse de conversation (KA). Il préconise par ailleurs de se fonder sur les unités prosodiques et non sur les unités énonciatives.

GAT est utilisé surtout par les linguistes, notamment par les spécialistes de l'analyse de la langue orale, qui utilisaient antérieurement KA.

6. CHAT et CHILDES

CHAT (*Codes for Human Analysis of Transcripts*) a été mis au point dans le cadre du projet CHILDES (Child Language Data Exchange System), sur lequel travaillent actuellement en Allemagne plus de quatre-vingt groupes de recherche. Le but de l'opération est de constituer une banque de données sur l'acquisition de la langue maternelle et de mettre à la disposition des chercheurs un système de transcription unifié. CHILDES comprend depuis 1991 trois éléments :

- un système de transcription et d'encodage (CHAT) ;
- une banque de données ;
- une série de programmes (CLAN) pour l'analyse des transcriptions.

CHAT présente les caractéristiques suivantes :

- Il ne se limite pas à la seule transcription des sons, mais note aussi les informations paraverbaux et non verbales ;
- toute nouvelle énonciation est notée sur une nouvelle ligne (même si le locuteur reste le même) ;

François Muller

- les informations complémentaires sont consignées dans un commentaire distinct de la transcription elle-même ;
- un système de notation très riche est mis à la disposition des transcrip-teurs pour qu'ils puissent, dans un souci d'authenticité maximale, noter toutes les particularités des différentes langues.

CHAT a le grand mérite, aux yeux de l'auteur, de bien séparer les tâches : ainsi l'observation (c'est-à-dire la transcription proprement dite) ne doit pas être confondue avec l'analyse (ou l'exploitation) scientifique des données.

Dans les trois derniers chapitres, beaucoup plus brefs, N. Dittmar traite successivement des informations non verbales (proxémique, cinésique, gestique, mimique et orientation du regard), ainsi que de l'insertion d'images dans les transcriptions. (chapitre 6). Ensuite il aborde les problèmes informatiques (logiciels, Internet comme source d'informations et d'accès aux grands corpus) (chapitre 7). Et il clôt son propos par des conseils très pratiques (par ex. sur le choix du système de transcription), conformément à l'esprit de l'ensemble de l'ouvrage, qui a privilégié tout au long de ses 247 pages une approche très conviviale de problèmes éminemment techniques, ce qui n'exclut évidemment pas une grande rigueur et une remarquable clarté dans la présentation de la matière elle-même.

Du bon usage des notations phonétiques dans les corpus aphasiques

Marie-Noëlle ROUBAUD
Université de Provence

Résumé

Les notations phonétiques sont plus nécessaires dans les corpus de locuteurs aphasiques que dans les corpus de locuteurs ordinaires. Elles traduisent ce que le transcripateur n'est pas en mesure de restituer sur un plan orthographique. Mais l'étude et l'analyse des corpus montrent qu'il faut faire un usage restreint de ces notations : tout ce qui peut l'être doit être orthographié. Mais qu'est-ce qu'un mot ou un fragment de mot orthographiable ? Il faut distinguer les jargons, qui seront notés phonétiquement, des amorces et déformations phonémiques, qui ont un traitement particulier. Le problème n'est pas simple : tout transcripateur, face à la recherche du mot par le locuteur aphasique, fait des hypothèses sur le morphème recherché ; que doit-il alors transcrire ? L'article expose ces difficultés et apporte des éléments de réponse pour l'usage des notations phonétiques.

1. Introduction

Le GARS¹ a une longue expérience de transcription de corpus qui l'a amené à établir des conventions de transcription. Ces conventions ont peu changé car les principes adoptés lors de leur élaboration restituent le texte écouté avec un maximum de fidélité, sans aucun toilettage.

Trois principes régissent les transcriptions effectuées par le GARS :

¹ Les travaux du GARS sont désormais poursuivis par l'équipe DELIC (DEscription Linguistique Informatisée des Corpus).

Marie-Noëlle ROUBAUD

1. la transcription se fait en orthographe standard, sans aucun trucage orthographique, les grands dictionnaires servant de référence ;
2. elle ne comporte aucun signe de ponctuation, aucune majuscule de début de phrase ;
3. elle comprend quelques notations phonétiques en note de bas de page.

2. Typologie

1.1. Les notations phonétiques correspondent à des morphèmes interprétables

C'est le cas des morphèmes prononcés rapidement et dont la déformation est banale :

- (1) enfin maintenant¹ je vais marquer ce qui a à mettre dans le cartable et puis vous allez sortir (corpus parodie enfants, Mor,IV,11,15)

¹prononcé [me~na~]²

Pour les mots de grande fréquence, comme *maintenant* (1), *voilà*, et *puis*, *peut-être*, qui subissent des raccourcissements caractéristiques ([me~na~], [vla], [epi], [ptEt]) quel que soit le locuteur qui les produit, le transcripteur a souvent intérêt à les signaler dans la présentation du corpus, supprimant ainsi les notes de bas de page. Mais la prononciation de certains morphèmes, à un certain moment du corpus, peut être utile à garder en note de bas de page :

- (2) il a fallu lui acheter un chauffage d'appoint à bain d'huile et également¹ une bouilloire (corpus adulte, Lav,38,15)

¹prononcé [egaZ@ma~]

Chez les enfants, la déformation d'un morphème peut s'expliquer parfois par la présence de la liquide /r/. Dans l'exemple (3), l'enfant produit le phonème /m/ au lieu de la suite consonantique /b/ + /r/ et dans l'exemple (4), il semble que le /r/, présent dans *réseaux* et dans *très*, parasite la

² Toutes les notations phonétiques sont en alphabet SAMPA : <http://www.phon.ucl.ac.uk/home/sampa/home.htm>.

production du morphème *tension*, que le locuteur adulte produira ensuite sans aucune erreur :

- (3) tout le nord a été ravagé les arbres ravagés euh euh les poteaux d'électricité tombés les toits des des maisons arrachés euh notre euh dix dix millions d'arbres¹ dix millions d'arbres sont abattus (corpus enfants, Ver, II, 21,6)

¹ prononcé [arm]

- (4) suite à la tempête alors c'est surtout des problèmes notamment de de réseaux de lignes très haute tension¹ euh très haute tension (corpus adulte, Hen,34,9)

¹ prononcé [tra~so~]

C'est le cas également des morphèmes qui en remplacent d'autres (cas des lapsus) :

- (5) la phrase est toute faute¹ + il y a qu'une se- il y a qu'un seul mot dans tout le + qui est juste (corpus parodie enfants, Mor,IV,5,11)

¹ pour « fausse »

C'est encore le cas des morphèmes que le transcripateur ne peut pas orthographier, parce qu'ils lui sont inconnus (noms propres, mots étrangers). Dans l'exemple (6), le transcripateur identifie un nom propre mais ne sait pas l'écrire, d'où la notation phonétique de bas de page :

- (6) un jour + il y avait un + il y avait quel- euh il y avait quelqu'un qui s'appelait + Jacques []¹ + qui fai- qui était hum + qui était turbulent comme lui (corpus parodie enfants, Mor,II,3,2)

¹ [satRER]

Dans l'exemple (7), le transcripateur identifie l'adjectif *provençal* mais du fait de sa prononciation, ne sait pas l'orthographier, d'où la notation phonétique de bas de page :

- (7) L1 c'est quoi la danse folklorique
L2 la farandole et tout ça ce qui est la Provence on nous déguise en []¹
et puis
L1 en quoi
L2 en []¹ (corpus enfants dyslexiques, Zab,50,3)

¹ [provEnsau], probablement « Prouvençau » [pruvEnsau] (cité dans le dictionnaire, au sens de « provençal »)

Marie-Noëlle ROUBAUD

Mais il n'est pas toujours nécessaire de recourir à une notation phonétique. Dans le cas des morphèmes dont l'écriture est codifiée (régionalismes, argots, jeux, onomatopées, emprunts), on peut rétablir l'orthographe dans le texte, comme pour *mama mia* (ex. 8) :

- (8) vous vous rappelez + quand le lendemain matin + elle a hurlé mama mia (corpus parodie enfants, Mor,I, 10,5)

2.2. Les notations phonétiques correspondent à des morphèmes non interprétables

Ces morphèmes que le transcripteur ne peut identifier comme des mots de la langue seront appelés « jargons ». Leur transcription ne peut pas se faire en orthographe standard et un espace vide entre crochets dans le corps du texte signale leur présence. Leur prononciation est restituée autant que possible en note de bas de page, entre crochets, en API. Dans les deux exemples qui suivent, le transcripteur n'a pas pu interpréter les morphèmes (9 et 10) :

- (9) eh bien quand vous + c'est quand je vous []¹ + c'est quand je vous + je vous redonne votre fils + c'est après les vacances regarde le calendrier (corpus parodie enfants, Mor, II,6,5)

¹ [lebe]

- (10) L1 que fait ta maman
L2 elle []¹ et
L1 elle fait du ménage et puis elle fait autre chose (corpus enfants, Romans)

¹ [demela~Z]

Ces notations phonétiques sont parfois nombreuses dans les transcriptions d'aphasiques : bon nombre de mots sont déformés³. L'étude des textes aphasiques⁴ a montré que c'est le manque du mot qui fait défaut au locuteur. La recherche envahit le texte : elle peut aboutir (le mot attendu est alors produit par le locuteur ou par le thérapeute) ou ne pas aboutir et

³ cf. bibliographie, études de Giorgini (1991) et Truong (1997) sur les déformations phonémiques des mots dans les corpus aphasiques.

⁴ Cette étude s'est faite en collaboration avec Claude Loufrani à l'Université de Provence, responsable d'un recueil de textes sur l'aphasie.

les corpus montrent alors le désarroi du patient qui n'arrive pas à dire ce mot⁵.

Pour un lecteur averti, les notations phonétiques de bas de page peuvent traduire cette recherche⁶ (ex. 11) :

(11) j'ai []¹ j'ai []² []³ []⁴ j'ai remis le téléphone et je suis resté sur le téléphone sur le lit (corpus aphasique, Tru97, 74,9)

¹[rakl@me]
²[rakl@]
³[rekoSe]
⁴[rarkle]

Dans cet exemple, le lecteur averti peut faire l'hypothèse que les notations phonétiques en bas de page sont des réalisations approximatives⁷ du morphème-cible *raccroché* (*j'ai raccroché le téléphone*) que le patient tente de produire par cinq essais successifs, dont deux essais sont identiques et correspondent au même numéro de note (note n°2). Cette recherche prend fin dès qu'un autre verbe *remis* (*j'ai remis le téléphone*) est produit par le patient.

Plusieurs questions se posent alors :

1. Quand utilise-t-on les notations phonétiques ?

La transcription de deux corpus de locuteurs aphasiques (que nous identifierons par I et II⁸) dans un mémoire de DEA a montré qu'il fallait poser des limites. Le transcripteur avait abusé des notations phonétiques qui masquaient des phénomènes tout à fait normaux : amorces de mots ou déformations phonémiques.

2. Où les place-t-on ?

⁵ Ce désarroi se lit au travers des paroles du patient dans les corpus : *je ne sais pas le dire, je ne trouve pas le mot ...* (cf. Roubaud & Loufrani, 1993)

⁶ Cette recherche se traduit aussi dans le débit du locuteur qui varie de 30 à 330 mots/mn, dans la durée et la fréquence des pauses (notées par : + pour une pause brève et -- pour une pause longue), dans la longueur des silences, dans les suites de syllabes inaudibles (notées par des X), dans le nombre de relances du thérapeute.

⁷ Pour la notion d'approximations lexicales, cf. Roubaud (1987, 1988).

⁸ Pour les deux corpus étudiés, le débit est en moyenne de 100 mots/mn pour le corpus I et de 110 mots/mn pour le corpus II (pour un locuteur ordinaire, il est de 200 mots/mn).

Marie-Noëlle ROUBAUD

Doit-on les rejeter en bas de page au risque de laisser apparaître un texte à trous ou doit-on les placer dans le corps du texte⁹ ?

3. Comment traiter les recherches lexicales ?

Doit-on différencier dans le texte les morphèmes qu'on ne peut vraiment pas interpréter (ex. 9) de ceux qui apparaissent au moins comme des recherches du mot (ex. 11) ?

C'est au travers d'exemples, extraits de corpus de locuteurs aphasiques, que nous essaierons de répondre à ces questions.

3. Les notations phonétiques et leur place dans les corpus aphasiques

3.1. Les déformations phonémiques

Comme dans les corpus de locuteurs ordinaires, certaines notations phonétiques correspondent à des morphèmes interprétables. On a intérêt à les écrire en orthographe dans le corps du texte et à restituer leur prononciation en note de bas de page (ex. 12 à 18).

C'est le cas des confusions de phonèmes :

- entre une consonne sourde /v/ et une sonore /f/

(12) j'ai j'étais là-bas en qua- en euh au mois de mars quarante-trois et à la fin de à la à la fin¹ la fin de la (Tru98,II,55,3)

¹ prononcé [ve~]

- entre deux liquides /r/ et /l/ (ex. 13) ou une liquide /r/ et une constrictive /Z/ (ex. 14)

(13) Thérapeute : il y a quelque chose qui va pas et puis
Patient : ils allaient¹ il m'a envoyé au docteur D- David à Lyon
(Tru98,I,18,14)

¹ prononcé [are]

⁹ Il est d'ailleurs précisé dans les conventions de transcription du G.A.R.S. (version mars 2000) que pour les textes « particuliers » (jeunes enfants, pathologies, étrangers), des conventions spécifiques sont établies, selon les besoins, pour placer des notations phonétiques dans le texte.

Du bon usage des notations phonétiques dans les corpus aphasiques

(14) Thérapeute : la guerre était terminée vous vous êtes engagé en trente-neuf

Patient : je je me suis je me suis ren- je me suis engagé¹ en- engagé pour la la la (Tru98,II,53,15)

¹ prononcé [a~gare]

- entre deux voyelles /o~/ et /a/

(15) et puis je vous dit avant elle elle venait chez sa chez sa sœur chaque année chaque année¹ et puis après elle allait passer trois euh trois ou quinze ou trois semaines chez chez son chez son fils (Tru98,II,66,7)

¹ prononcé [so~kane]

C'est aussi le cas des omissions de phonèmes, comme dans cet exemple où l'omission du /j/ par le locuteur peut laisser entrevoir la production d'un lapsus « marré » pour *marié* :

(16) et ensuite la guerre est finie je me suis marié¹ ma- marié (Tru98,II,52,6)

¹ prononcé [mare]

C'est encore le cas des ajouts de phonèmes. Dans l'exemple qui suit, le locuteur transforme un /t/ final en /sk/ :

(17) c'est un peu bête¹ (Nolot,I,52,7)

¹ prononcé [bEsk]

Le morphème peut aussi être identifié grâce à la connaissance culturelle que se partagent le locuteur et le transcripateur. Ainsi dans l'exemple 18, le patient imite la voix du général de Gaulle et lui emprunte la formule *Vive la France* qui terminait ses discours, en déformant par deux fois le mot *France* :

(18) Thérapeute : vous avez fait euh le général De Gaulle l'autre jour là
Patient : vive la France¹ vive la France² nous les Français mais je peux pas c'est pas la même voix (Tru98,II,60,7)

¹ prononcé [ka~s]

² prononcé [ka~]

Rétablir l'orthographe de ces mots dans le corps du texte, c'est alors les identifier comme des morphèmes de la langue. Il y a également un avantage à rassembler en bas de page les erreurs de prononciation : cela

Marie-Noëlle ROUBAUD

facilite une étude des déformations les plus fréquentes dans un corpus de locuteur aphasique.

3.2. Les jargons

Les mots « jargonnés » sont parfois nombreux :

- (19) Thérapeute : et qu'est-ce que vous avez dit
Patient : je sais pas rien c'est pour le les les les []¹ les XXX celle
qui était []² à Saint-Tropez (Tru98,II,56,3)

¹ [ziva~]
² [toloke]

La notation phonétique de ces mots dans le corps du texte rétablirait une bonne lisibilité du texte. En effet, un texte jargonné risque d'être moins lisible si le corps du texte est parsemé de « vides » (notés []), le lecteur devant sans cesse se reporter aux notes de bas de page au risque de perdre le sens du texte. Et l'écoute du texte aligné avec son signal sonore ne correspondra pas avec la transcription puisque aucune marque graphique ne correspondra à la suite de phonèmes produite. C'est pour cela que nous noterons dorénavant les notations phonétiques dans le corps du texte (ex. 19') :

- (19') je sais pas rien c'est pour le les les les [ziva~] les XXX celle
qui était [toloke] à Saint-Tropez

Afin de faire le tri entre ce qui était un jargon et ce qui ne l'était pas, nous avons relu systématiquement des transcriptions pour les améliorer. C'est une entreprise longue, indispensable, car sous l'apparence de jargons, nous avons découvert des morphèmes identifiables. Cela a été le cas :

- des noms propres non identifiés par le transcripteur

Dans l'exemple qui suit, Coblenze, prononcé en allemand Koblenz, a été transcrit phonétiquement, comme un jargon¹⁰ :

- (20) Thérapeute : vous m'avez dit bon euh votre votre épouse n'est
pas française
Patient : elle est de de près de de [k] [koble~s] sur sur le sur le
Rhin (Tru98,II,61,7)

¹⁰ Dans la fiche descriptive de ce corpus, le transcripteur aurait dû indiquer que certains des mots du locuteur étaient prononcés en allemand comme le nom de Coblenze.

alors qu'il aurait fallu transcrire :

(20') elle est de de près de K- Koblenz sur sur le sur le Rhin

- des fautes contre la norme

L'auxiliaire *avoir* est souvent employé là où il faudrait *être* selon la norme et cette erreur a été transcrite comme un jargon :

(21) mais mais je me sens bien de suite euh une amie de ma sœur est
[av@ny] à onze à dix heures du soir (Tru98,II,78,14)

Il aurait fallu transcrire :

(21') mais mais je me sens bien de suite euh une amie de ma sœur est
a venu à onze à dix heures du soir

- des amorces

Toute amorce de mot peut être transcrite dans le corps du texte, selon les conventions de transcription : signalée par un tiret collée au mot (par exemple : *m-*, amorce de *mais*). Tout locuteur, ordinaire ou aphasique, fait de nombreuses amorces et leur transcription dans les corpus participe à l'authenticité des textes. Mais une erreur serait de les noter phonétiquement, lorsqu'on peut faire « mieux », comme dans ces exemples où le transcripteur n'a pas réussi à les identifier :

(22) Thérapeute : est-ce que vous pourrez me citer des noms de journaux + locaux
Patient : [ko] le Nice matin (Tru98,I,41,4)

Il aurait fallu rétablir l'amorce de «*comme* :

(22') co- le Nice matin

(23) Patient : et puis j'étais je je mangeais avec chez mes mes mes cousins et puis tout d'un coup ce bras il répondait pas
Thérapeute : quel bras
Patient : celui-ci il ne répondait pas et et [s@kimoR] il tombait il tombait vous savez (Tru98,II,75,16)

On aurait pu rétablir l'amorce du syntagme verbal *ce qui m'aurait* :

(23') celui-ci il ne répondait pas et et ce qui m'aur- il tombait il tombait vous savez

Ces amorces peuvent subir des déformations, comme dans les exemples ci-dessous :

(24) mon père était pas [pi] prisonnier et et ma mère était là quoi (Tru98,II,54,12)

Marie-Noëlle ROUBAUD

On pourrait rétablir l'amorce de *prisonnier*, pour laquelle le /r/ de la première syllabe a été omis :

- (24') mon père était pas pi- prisonnier et et ma mère était là quoi
(25) mais et par moment et et et par moment je [bu] je doutais
(TRU98,II,84,14)

On pourrait rétablir l'amorce de *doutais*, pour laquelle la labiale /b/ a été remplacée par la dentale /d/ :

- (25') mais et par moment et et et par moment je bou- je doutais

Un nouveau problème se pose alors (ex. 24 et 25) : doit-on écrire en orthographe standard ce que l'on entend (ex. 24' et 25') ou doit-on mettre cette prononciation en note de bas de page (ex. 24'' et 25'') et rétablir l'orthographe de l'amorce ?

- (24'') mon père était pas pri-¹ prisonnier et et ma mère était là quoi

¹ prononcé [pi]

- (25'') mais et par moment et et et par moment je dou-¹ je doutais

¹ prononcé [bu]

Nous pensons que les amorces peuvent toutes être traitées au maximum de la même façon, qu'elles aient ou non subi une déformation phonémique, et c'est pour cette raison que nous préférons écrire en orthographe standard l'amorce de ce qui a été prononcé (ex. 24' et 25'). Nous réservons les notes de bas de page à la prononciation des morphèmes produits entièrement (ex. 12 à 18).

4. Les notations phonétiques et les recherches lexicales

Certaines notations phonétiques, même si le transcripateur ne peut les identifier comme des mots de la langue, sont des réalisations approximatives d'un morphème-cible. C'est le cas du verbe *raccrocher* de l'exemple 11 (les notations figurant désormais dans le corps du texte) :

- (11) j'ai [rakl@me] j'ai [rakl@] [rekoSe] [rarkle] [rarkle] j'ai remis le téléphone et je suis resté sur le téléphone sur le lit

Même pour un transcripateur averti, cette recherche n'est pas toujours sûre à cent pour cent, témoin cet exemple :

- (26) Thérapeute : ben racontez justement cette histoire là

Du bon usage des notations phonétiques dans les corpus aphasiques

Patient : oh il y a pas longtemps il y a il y a peut-être d- dou- douze ans de ça douze ans j'étais avec mon ami euh euh j'aimais bien aller sans [da-satale] le soir parce qu'il y avait un un où on on donnait tout en en mar- en en mangeant voyez-vous (Tru98,II,56,14)

Doit-on laisser la notation phonétique ou peut-on rétablir le syntagme *dans cette allée* ? Dans ce cas, il est possible que *sans* précédant le syntagme *dans cette allée* soit la préposition *dans*, prononcée [sa-], le locuteur répétant deux fois la préposition comme le ferait un locuteur ordinaire : *dans dans cette allée*. Mais n'est-on pas en train de trop interpréter le texte ?

Le cas inverse apparaît dans l'exemple qui suit : le transcripneur a rétabli l'orthographe standard d'un morphème à *pied* qui semble inadapté dans ce texte :

(27) on on (n') est pas divorcé elle m'aime toujours même l'autre jour elle m'a habité euh elle m'a à pied ap- [pa] [pa] elle m'a appelé euh non on se voit elle elle est venue chez chez ma soeur (Tru98,II,64,14)

Le locuteur aphasique semble rechercher le verbe *appeler* et le lecteur comprend difficilement la présence du lexique à *pied*, donné par le transcripneur. Il aurait fallu laisser la notation phonétique (ex. 27') :

(27') on on (n') est pas divorcé elle m'aime toujours même l'autre jour elle m'a habité euh elle m'a [apje] ap- [pa] [pa] elle m'a appelé euh non on se voit elle elle est venue chez chez ma soeur

Cette recherche est facilement repérable, si le mot produit possède tous les phonèmes du mot attendu mais dans le désordre (ex. 28) :

(28) euh je je je vous en donne je vous en- dans le dans le dans le euh euh [ynaZ] non nuage (Tru98,II,71,11)

Cette inversion dans l'ordre des phonèmes n'est pas caractéristique des locuteurs aphasiques, témoin cet exemple où c'est le thérapeute qui inverse les deux dernières syllabes alors que le patient les répète correctement :

(29) Thérapeute : [alize] et faites-le si vous pouvez
Patient : allez-y et allez-y et non alors pas pas très bien (Nolot,I,47,11)

Cette recherche est également repérable dans les différents essais, plus ou moins nombreux, produits par le patient en quête d'un mot (ex. 30 et 31) ou d'un syntagme (ex. 32) :

- recherche du mot *autorisation*

Marie-Noëlle ROUBAUD

- (30) je voulais faire la guerre mon père euh sans un mot il m'a écrit son [oriz] son [solitasjo~] ses autorisations que je pouvais y aller (Tru98,II,53,7)

- recherche du mot *contacts*

- (31) Thérapeute : vous ne vivez plus ensemble
Patient : non non non non j'ai été la voir en Allemagne il y a deux ans je suis resté dix jours je m- je mangeais chez elle avec avec sa soeur et j'a- j'allais le soir pour euh pour dormir à l'hôt- à l'hô- à l'hôtel voyez-vous nous sommes n- nous aurons de de bons euh de bons [ko~] [ko~] [ko~] [ko~katR] [ko~tR] [ko~katR] [ko~katR] merde [ko~] [ko~kat] [ko~kat] c'est bon comme ça non elle
Thérapeute : contacts (Tru98,II,65,8)

C'est l'orthophoniste qui clôt cette recherche en prononçant le mot *contacts* que le locuteur, par dix essais successifs, tente de prononcer. Pour ce dernier exemple, nous rétablirons l'orthographe standard des amorces *con-* du mot *contacts*, ce qui donne (ex. 31') :

- (31') n- nous aurons de de bons euh de bons con- con- con- [ko~katR] [ko~tR] [ko~katR] [ko~katR] merde con- [ko~kat] [ko~kat] c'est bon comme ça non elle

- recherche du syntagme *contremaître de fabrication*

- (32) Thérapeute : oui j'avais oublié de vous poser cette question là euh votre métier
Patient : j'é- j'étais chez Nestlé comme comme [komEt] [ko~mEtResa~] de [fRegyRasjo~] de + non attendez de [ko~tr] [ko~tR@mER] contremaître de de de [fa~] [fapRetasjo~] fabrication chez Nestlé (TRU98,II,70,1)

Le patient cherche à la fois le mot *contremaître* (4 essais) et le mot *fabrication* (3 essais). La recherche porte en fait sur tout le syntagme : *contremaître de fabrication*. Comme pour l'exemple précédent, nous rétablirons en orthographe standard les amorces *contr-* de *contremaître* et l'amorce déformée *fan-* de *fabrication* (ex. 32') :

- (32') j'é- j'étais chez Nestlé comme comme [komEt] [ko~mEtResa~] de [fRegyRasjo~] de + non attendez de contr- [ko~tR@mER] contremaître de de de fan- [fapRetasjo~] fabrication chez Nestlé

Nous nous sommes demandée s'il fallait indiquer au lecteur la présence de ces recherches lexicales. On pourrait envisager de placer, dans le texte, un signe graphique, comme dans l'exemple qui suit :

- j'é- j'étais chez Nestlé comme comme [komEt = contremaître] [ko~mEtResa ~ = contremaître] de [fregyrasjo ~ = fabrication] de + non attendez de contr-

[ko~tr@mER= contremaître] contremaître de de de fan- [fapRetasjo~ = fabrication] fabrication chez Nestlé

Mais, de toute évidence, cela ne ferait que surcharger le texte et nuire à sa lisibilité. Distinguer les recherches du mot dans les notations phonétiques sera la tâche du linguiste, travaillant sur le corpus. Le nombre d'essais, l'apparition de phonèmes appartenant au morphème-cible et les amorces seront des indices précieux pour identifier ces recherches.

5. Conclusion

La tâche du transcripneur, face à un corpus de locuteur aphasique, est multiple.

Dans un premier temps, il doit noter phonétiquement, dans le corps du texte, ce qui est prononcé par le locuteur et qu'il n'interprète pas immédiatement comme un mot de la langue.

Puis vient une étape plus difficile : faire le tri de ces notations phonétiques. En orthographe standard pourront être restitués les phénomènes ordinaires comme les amorces, les déformations phonémiques (dont la prononciation sera notée en bas de page) ...

Dans le corps du texte ne figureront que les notations phonétiques correspondant à des morphèmes non interprétables. Mais il appartiendra au linguiste de repérer, parmi ces morphèmes, ceux qui sont les indices d'une recherche lexicale. Cela suppose que le transcripneur donne du sens à la parole du locuteur aphasique, qu'il soit capable de repérer les réalisations approximatives d'un morphème-cible sans interprétations excessives.

Ce n'est que si le transcripneur a fait un bon usage des notations phonétiques que le travail sur le corpus pourra alors se faire.

Références

- Bernstein, B. (1975). *Langage et classes sociales. Codes socio-linguistiques et contrôle social*. Paris : Éditions de Minuit.
- Blanche-Benveniste, C., Bilger M., Rouget C. & van Den Eynde K (1990). *Le français parlé. Études grammaticales*. Paris : CNRS.
- Blanche-Benveniste, C. (1997). *Approches de la langue parlée en français*. Paris-Gap : Ophrys.
- Dorze, G. de (1985). *L'aphasie et le processus de lexicalisation*. Thèse de doctorat, Université de Montréal, Montréal.

Marie-Noëlle ROUBAUD

- Giorgini, M. (1991). *Étude d'un locuteur aphasique : corpus Y. L. : déformation phonémique et dénomination*. Mémoire de maîtrise de linguistique française. Université de Provence, Aix-Marseille 1.
- Loufrani, C. (1988). Un corpus de locuteur aphasique, originalité ou régularité ? *Recherches sur le français parlé*, 8, 59-78.
- Loufrani, C. (1990). *Analyse du discours de locuteurs classés comme aphasiques : de l'art de tourner autour du mot*. Thèse de doctorat de linguistique française. Université de Provence, Aix-Marseille 1.
- Loufrani, C. & Roubaud, M.-N. (1990). La notion d'approximation : langage ordinaire ou pathologique. *Recherches sur le Français parlé*, 10, 131-142.
- Nespoulous, J.-L. (1980). De deux comportements verbaux de base : référentiel et modalisateur. De leur dissociation dans le discours aphasique. *Cahiers de psychologie*, 23, 195-210.
- Nolot, V. (1991). Exploitation syntaxique et lexicale de corpus de locuteurs aphasiques. Mémoire de maîtrise de linguistique française. Université de Provence, Aix-Marseille 1.
- Roubaud, M.-N. (1987). *L'approximation lexicale*. Mémoire de maîtrise de linguistique française. Université de Provence, Aix-Marseille 1.
- Roubaud, M.-N. (1988). *Les mystères de l'approximation lexicale*. Mémoire de DEA de linguistique française. Université de Provence, Aix-Marseille 1.
- Roubaud, M.-N. & Loufrani, C. (1993). La syntaxe, c'est ce qui reste quand on a tout oublié : La syntaxe comme outil de description syntaxique. *Recherches sur le français parlé*, 12, 85-113.
- Roubaud, M.-N. & Loufrani, C. (1999). Éclairage des corpus de type aphasique par les grilles. *Recherches sur le français parlé*, 15, 41-57.
- Truong, K.-Y. (1997). *Déformations sémantiques et comptage des verbes chez deux locuteurs classés aphasiques*. Mémoire de maîtrise de linguistique française. Université de Provence, Aix-Marseille 1.
- Truong, K.-Y. (1998). *Classement et études sémantiques des verbes thématiques et de quelques verbes dits de grande fréquence présents chez deux locuteurs classés aphasiques*. Mémoire de D.E.A. de linguistique française. Université de Provence, Aix-Marseille 1.

Résumés de thèses

Méthodologies pour la description de quelques phénomènes syntaxiques d'un recueil de langue parlée. Application à l'hébreu moderne

II-II YATZIV- MALIBERT

Thèse dirigée par Claire Blanche-Benveniste

La thèse présente une étude syntaxique et lexicale d'un recueil de langue parlée. Le recueil est composé des transcriptions de douze monologues enregistrés de locuteurs israéliens natifs. La durée de chaque enregistrement varie de 10 minutes à 40 minutes. Les enregistrements ont été effectués par Madame Yatziv-Malibert.

C'est dans l'intention d'utiliser le corpus de langue parlée dans un but de description grammaticale de la langue hébraïque que cette recherche a été entamée. Il s'agit d'affiner la description de faits déjà connus, présents chez tous les locuteurs.

Pour mener cette recherche sur l'hébreu parlé, domaine rarement étudié dans la linguistique hébraïque, il a fallu s'inspirer de méthodes d'analyse appliquées à d'autres langues, notamment au français parlé (les méthodes pratiquées par le Groupe Aixois de Recherche en Syntaxe) et à l'anglais.

La première partie de la thèse aborde la question de savoir pourquoi l'hébreu parlé est si peu étudié. Le premier facteur d'explication est lié à l'histoire particulière de l'hébreu, et notamment à la longue rupture dans l'usage parlé de cette langue alors que le deuxième facteur regroupe des éléments explicatifs plus généraux qui concernent également d'autres langues à savoir la mise à l'écart de l'oral par rapport à la langue écrite et la perception de l'oral comme fautif. Cette partie rassemble des travaux antérieurs sur l'hébreu parlé et quelques recueils existants (Bar-Adon, Blanc et Shatil). De plus elle présente le cadre méthodologique ainsi que les principales unités d'analyse dégagées: les constructions verbales, les constructions non verbales, les divers dispositifs et notamment le dispositif pseudo-clivé étudié plus en détail.

Pour ce faire deux grandes lignes de recherche ont été croisées: nous nous sommes inspirés d'une part des travaux du Groupe Aixois de

Recherche en Syntaxe et d'autre part de la tradition linguistique hébraïque et particulièrement de la réflexion sur les schèmes de phrases de Haïm Rosen et de Maya Fruchtman.

La deuxième partie présente les phénomènes syntaxiques étudiés : le mécanisme paradigmatique, les insertions lexicales et syntaxiques qui perturbent le déroulement syntagmatique des énoncés et les répétitions du lexique.

Dans le chapitre traitant le mécanisme paradigmatique nous avons montré que des phénomènes appartenant à des domaines de nature différente : syntaxe, lexique ou énonciation et des effets tels que erreurs, répétitions intentionnelles ou non, recherche de dénomination ou anticipation peuvent être décrits à l'aide du même instrument de travail qu'est « le listing ». Nous avons pu trouver de fortes ressemblances avec la description de ces phénomènes dans d'autres langues.

Le procédé d'insertion consiste à intercaler une ou plusieurs constructions dans le cours d'une séquence d'énoncés. Sur le plan syntaxique, ce sont des « parasites » qui viennent couper des constructions grammaticales. Sur le plan énonciatif, ce sont des interventions métalinguistiques et notamment des éléments formant des commentaires sur l'énonciation. Ici également la description des faits trouvait son équivalent dans la description des insertions en français parlé.

Le chapitre sur les répétitions du lexique présente trois phénomènes : des reprises de termes à l'identique sans qu'elles provoquent de changement syntaxique, des glissements de lexique qui sont accompagnés de changement de fonction syntaxique et des reformulations de la **racine** autour de la distinction verbe et nom ainsi qu'autour de la notion de **schème**. C'est ce dernier phénomène qui est propre à l'hébreu parlé et probablement à d'autres langues sémitiques.

La troisième partie présente une synthèse des analyses sous forme de grilles, mode de présentation des discours oraux déjà appliqué à d'autres langues. Après avoir défini les unités syntaxiques d'analyse et les structures récurrentes apparaissant dans le corpus de l'hébreu oral, il est proposé de démontrer que, malgré les idées reçues sur le désordre de la langue parlée et les difficultés qui en découleraient pour l'étudier, le discours oral peut être organisé et structuré.

À l'examen des extraits de textes oraux sont apparues des configurations syntaxiques qui se présentent comme un ensemble structuré par plusieurs énoncés et délimité par des répétitions et des modifications de constructions. Ces configurations peuvent servir d'esquisses de classification de genres de discours parlé.

Étude comparative du fonctionnement des nombres singulier et pluriel dans les langues romanes : portugais, espagnol, catalan et français

Eulàlia VILAGINES SERRA

Thèse dirigée par Claire Blanche-Benveniste

Le phénomène de la « syllepse » a intéressé les grammairiens depuis l'Antiquité. Il s'agit d'un phénomène d'accord apparemment non régulier qui montre un décalage entre l'élément qui donne l'accord et celui qui le reçoit. L'accord de syllepse en nombre¹ est entre *un groupe d'hommes*, considéré comme syntagme nominal singulier, et le verbe *partent*, au pluriel. Diverses attitudes normatives se sont développées pour accepter ou refuser l'accord de syllepse.

On a souvent parlé d'« accord par le sens » plutôt que par la morphologie. Pour certains il s'agit d'un accord avec le complément de nom ou avec un « complément de nom non exprimé ». D'autres posent la notion de perspective selon qu'on prend en compte la forme ou le signifié du collectif. Des analyses plus récentes utilisent les notions de « collectif » et de « quantifieur » et définissent l'accord en fonction d'une analyse interne du syntagme : notion de tête nominale, de tête de syntagme, etc. Ces analyses intègrent des points de vue sémantique et grammatical. J'ai choisi de les utiliser dans une optique lexico-grammaticale pour l'étude de la syllepse.

Les langues romanes considérées, portugais, espagnol, catalan et français, diffèrent par leur lexique et leur grammaire mais elles présentent suffisamment de similitudes pour qu'on puisse les comparer. J'ai constitué un corpus de 4000 exemples attestés écrits et oraux qui se répartissent de la façon suivante : 741 en portugais, 923 en espagnol, 979 en catalan et 1342 en français.

Pour qu'il y ait possibilité de choix entre le singulier et le pluriel, il faut que le « syntagme nominal source » contienne un « quantifieur-

¹ La syllepse de genre, la syllepse de personne et la syllepse selon le distributif ne font pas l'objet de cette étude, qui se consacre uniquement à la syllepse de nombre.

collectif». J'ai réuni sous ce nom ce que la grammaire traditionnelle a nommé collectif (type *le personnel*) et quantifieur (type *la plupart*), mais j'ai aussi retenu certains lexèmes qui ne fonctionnent comme des « quantifieurs-collectifs » que dans certaines conditions (type *un chapelet de*).

Les données attestées montrent que ces lexèmes ont différentes réalisations lexico-grammaticales. Parfois, le terme n'accepte pas de complément désignant les parties constitutives qui serait tautologique (? *une pinède de pins*). Parfois le complément désignant les parties constitutives devient possible à condition d'être lui-même qualifié (*un jurat d'homes justos* (C), *una familia de parientes bienvenidos* (E)). Quelques-uns acceptent facilement la présence de ce complément désignant les parties constitutives (*un groupe d'amis, la majorité d'amis*). Il existe, enfin, des lexèmes qui ne peuvent se réaliser qu'accompagnés du complément désignant les parties constitutives (*un chapelet de bombes* (F), *un rio de curiosos* (E)). La présence ou absence de ce complément s'est révélée importante vis-à-vis du choix de l'accord.

D'autre part, en fonction de ces différents types d'emplois en tant que collectifs ou quantifieurs, j'ai proposé une typologie des termes. *Le corps médical* ou *le public* sont des collectifs qui ne deviennent jamais quantifieurs ; les expressions de pourcentage, *25%*, ou désignant des parties restent des expressions de quantité qui n'ont jamais un emploi de collectif ; mais il ne manque pas de lexèmes qui fonctionnent tantôt comme des collectifs (*une foule, une armée*), tantôt comme des quantifieurs (*une foule de métamorphoses, une armée d'engins*).

Par ailleurs, chaque terme manifeste des tendances d'accord. *Le reste des enfants*, en français, choisit nettement le singulier, alors que *la plupart des enfants* le pluriel. Des lexèmes comme *la majorité*, dans les quatre langues, ou *(l)a gent(e)*, en portugais, en espagnol et en catalan, naviguent entre les deux.

J'ai examiné les cibles d'accord dans différentes situations syntaxiques, en commençant par la construction directe sujet + verbe. Sur l'ensemble des lexèmes retenus pour faire la comparaison, lorsque le syntagme nominal source remplit la fonction sujet, au moins un tiers peuvent provoquer des accords de pluriel bien qu'ils aient une morphologie de singulier. Mais la proportion varie selon les langues : le portugais et l'espagnol ont seulement un tiers des exemples avec accord au pluriel ; le français 40%, et le catalan 50%. Le simple fait d'ajouter un « dispositif » du type *il y a ... qui* implique pour les quatre langues un changement des tendances d'accord.

Dans les relatives, source et cible se trouvent dans un même grand constituant (*Une génération qui ont perdu*). Pour les quatre langues, le pluriel est l'accord choisi dans 60% des cas.

La cible peut être un verbe « voisin », c'est-à-dire le verbe d'une nouvelle construction (*Elle voit un groupe. Ils s'arrêtent*) ou un verbe de la même construction (*Il regarde la famille quand ils mangent*). Le choix du pluriel s'impose alors, laissant peu de place à l'accord au singulier.

Il en va de même lorsqu'il s'agit d'un accord par inférence entre deux éléments nominaux : le singulier est très rare (*C'est une famille. Je les vis monter*).

Une partie importante de mon corpus atteste qu'un syntagme nominal source peut être concerné par plusieurs cibles, par exemple par un verbe recteur et par un verbe « voisin » (*Ce couple est très sympathique. Ils m'ont accueilli*).

Le paramètre du contexte syntaxique s'est révélé plus important que je ne le pensais au départ : j'ai pu observer un glissement du choix du singulier dans les constructions directes sujet + verbe vers le pluriel dans le cas des relatives, qui s'accroît quand il s'agit des accords avec les verbes « voisins » et les accords par inférence entre deux éléments nominaux.

Je propose une structure lexico-grammaticale des termes analysés, c'est-à-dire, l'ensemble des possibilités que chacun des termes possède. La structure sera établie à partir du lien entre lexique et grammaire. Par exemple, voici la structure typique de « foule » :

1. une foule va venir
2. une foule ? vont venir²
3. une foule de personnes va venir
4. une foule de personnes vont venir
5. une foule d'exemples va arriver
6. une foule d'exemples vont arriver
7. une foule ---- va venir³
8. une foule ---- vont venir

Faut-il en conclure que la syllepse est un cas de variation ? Il est vrai que dans certains cas oui, mais on observe que le choix du singulier ou du pluriel est significatif dans d'autres cas. D'autre part, il a fallu aussi tenir

² Le signe ? marque le doute sur l'existence d'une des possibilités.

³ Le trait ---- rappelle qu'il n'y a pas de tête lexicale.

compte de processus de grammaticalisation. Certains lexèmes, comme *la foule*, sont encore en évolution, d'autres, comme *la plupart*, sont arrivés à la fin de ce processus.

L'étude de la syllepse, enfin, oblige à raffiner la notion de nombre. Le système du nombre est plus complexe que l'opposition singulier-pluriel à laquelle nous sommes habitués, comme le montrent des études récentes. Dans le cas de l'accord de syllepse, il faut tenir compte du cryptotype « collectif ».

Le pronom anaphorique et l'architecture de l'oral en finnois et en français

Outi TIAINEN-DUVALLON

Thèse dirigée par Claire Blanche-Benveniste

Ce travail étudie l'emploi des pronoms anaphoriques du point de vue de la structuration syntaxique des textes oraux. L'objectif est de décrire, d'une part, l'émergence des constructions syntaxiques dans les productions orales et de montrer, d'autre part, comment l'emploi et l'interprétation des pronoms de troisième personne sont intimement liés à l'organisation syntaxique des textes. La base de données est un recueil d'enregistrements de conversations en finnois et en français.

L'étude se concentre sur les pronoms finnois *se* 'il/elle ~ ce' et *hän* 'il/elle' en position préverbale et sur les pronoms sujets français de type *il*. Ces pronoms sont présentés comme des expressions référentielles originales, non tributaires des autres. Au sein des systèmes pronominaux des deux langues, ils sont dotés de certaines propriétés inhérentes. Les pronoms français expriment le trait sémantique [\pm classifié] et les pronoms finnois le trait [\pm logophorique]. Ces traits font qu'il y a des divergences dans le choix des pronoms dans les contextes locaux, mais ils ne semblent pas peser sur l'emploi des pronoms comme des contraintes non locales. Dans les deux langues, les pronoms en question sont aptes à des emplois anaphoriques.

Les analyses syntaxiques s'inscrivent dans le cadre théorique de l'Approche Pronominale (Blanche-Benveniste *et al.*, 1987). La description est fondée sur la notion de verbe. Le verbe est considéré comme un élément recteur qui ouvre des places de construction autour de lui. L'analyse des données commence par l'examen des procédés par lesquels les constructions verbales sont mises en place dans les textes oraux des deux langues. L'attention est portée, d'une part, à l'émergence de l'axe syntagmatique des énoncés verbaux et, d'autre part, aux réalisations paradigmatiques des places de rection. Ainsi, l'étude intègre à la description syntaxique des phénomènes de « production » comme les réparations d'énoncés et différents types d'insertions qui dédoublent la

chaîne parlée ou qui suspendent momentanément le déroulement d'un énoncé cadre.

Les insertions parenthétiques sont conçues comme une troisième dimension textuelle, c'est-à-dire comme un site à partir duquel les locuteurs peuvent gérer les opérations en cours sur l'axe syntagmatique ou paradigmatique de l'énoncé cadre. L'examen des passages de texte plus longs met l'accent sur le phénomène de récursivité des moules syntaxiques. Les rééditions d'un verbe recteur et de sa construction forment dans les textes oraux des unités de structuration, appelées *configurations syntaxiques*. Pour mettre en valeur ces organisations structurelles, le travail applique aux extraits de corpus le mode de présentation « en grilles syntaxiques ». L'analyse des données permet de voir que les mécanismes de production de textes oraux se ressemblent dans les deux langues, qui possèdent pourtant des micro-syntaxes assez différentes.

Les pronoms sont examinés sous un double rôle : d'une part, ce sont des indicateurs syntaxiques, plus sûrs que les formes lexicales des noms, qui permettent d'identifier les places de rection autour d'un verbe, et, d'autre part, ce sont des expressions référentielles, dotées d'un minimum de charge descriptive. Le travail défend l'idée selon laquelle les pronoms de troisième personne sont des expressions référentielles non marquées qui, face à l'instabilité des désignations lexicales, permettent de saisir un référent d'une manière plus stable. Les exemples analysés montrent que les pronoms peuvent être utilisés non seulement là où les désignations lexicales sont déjà mises au point, mais aussi là où elles sont encore en cours, en train d'être négociées ou même momentanément indisponibles.

L'étude décrit les diverses façons dont l'ancrage référentiel des pronoms anaphoriques se détermine dans le contexte linguistique caractérisé par la complexité syntaxique. L'interprétation référentielle des pronoms est cruciallement guidée par les informations véhiculées par leur construction hôte et par la position de celle-ci dans l'organisation structurelle du texte. Pour les pronoms utilisés dans les séquences insérées, l'énoncé cadre constitue un domaine d'interprétation particulièrement saillant. Pour les pronoms dont la construction hôte forme une configuration syntaxique avec d'autres séquences dans le texte, c'est la configuration syntaxique qui leur fournit un précieux domaine d'interprétation.

Le travail souligne le fait que les domaines d'interprétation formés par les configurations syntaxiques ne sont pas toujours des espaces textuels linéaires. La distance entre le pronom et sa source d'interprétation dans le texte ou la simple présence dans le contexte linguistique de plusieurs

sources d'interprétation sémantiquement compatibles avec le pronom auraient moins d'importance sur l'emploi des pronoms anaphoriques que l'on ne l'a souvent pensé.

Composition et mise en page
Estelle CAMPIONE, Estelle.Campione@up.univ-aix.fr
Sandra TESTON, Sandra.Teston@up.univ-aix.fr

Maquette de couverture
PUP - Valérie JULIA

Imprimé en France
sur les presses d'OZAPLAN – Marseille

Dépôt légal 3^e trimestre 2004
ISBN 2-85399-571-2
ISSN 0224-0041