

# Nouvelles technologies et standards méthodologiques en linguistique

Édité par

Marianne KILANI-SCHOCH, Christian SURCOUF et Aris XANTHOS

Cahiers de l'ILSL, n° 45, 2016



UNIL | Université de Lausanne

Illustration de couverture basée sur un design de stALLio!  
<https://www.flickr.com/photos/stallio/3149911976>

Les Cahiers de l'ILSL  
(ISSN 1019-9446)  
sont une publication du Centre de Linguistique et  
des Sciences du Langage de l'Université de  
Lausanne (Suisse)

Centre de Linguistique et des Sciences du Langage  
Quartier UNIL-Dorigny, Bâtiment Anthropole  
CH-1015 Lausanne

## **L'utilisation des corpus oraux pour la recherche en (psycho)linguistique<sup>1</sup>**

Mirjam ERNESTUS

Radboud University Nijmegen & Max Planck Institute for Psycholinguistics (NL)

m.ernestus@let.ru.nl

### **1. Introduction**

Ma présentation abordera l'usage des corpus de langue dans la recherche en (psycho)linguistique. Elle se divise en deux parties. La première concerne la question : quelle est la structure phonétique des mots dans le flux de parole ? En général, ce sont des « linguistes de bureau », qui, en songeant à leur propre manière de produire des énoncés, tentent d'y répondre. Mon second questionnement portera sur la manière dont les locuteurs produisent de l'oral, point habituellement traité à l'aide d'expériences en laboratoire. Cependant, j'évoquerai pour ma part l'apport des corpus oraux à ces deux questions. Le problème avec « la linguistique de bureau » ou les expériences de production en laboratoire, c'est qu'elles ne montrent pas réellement comment les individus se comportent lors d'une conversation naturelle. Or, je pense que les corpus de conversations spontanées peuvent contribuer à répondre à ces deux questions, et j'espère pouvoir vous en convaincre. Je commencerai par discuter de l'utilité de ces corpus en évoquant succinctement les types de questions qu'ils permettent de résoudre. J'évoquerai également les problèmes auxquels nous avons été confrontés lorsque que nous avons utilisé des corpus pour répondre à nos questions.

---

<sup>1</sup> *The use of speech corpora for (psycho)linguistic research*. Transcription, traduction et adaptation par Guillaume Feigenwinter et Christian Surcouf.

## 2. Quelques questions autour de la structure phonétique des mots dans le flux de parole

Commençons par la première question : quelle est la structure phonétique des mots dans le flux de parole ? On trouve évidemment des réponses dans les écrits spécialisés, et j'aimerais poser à ce sujet deux questions, que je vais accompagner chaque fois d'un exemple : 1) « les faits qu'on trouve dans la littérature spécialisée sont-ils exacts ? », 2) : « la littérature spécialisée passe-t-elle à côté de faits intéressants ? ». Mon attention se portera sur une étude qu'on a réalisée il y a déjà plusieurs années, au sujet de l'assimilation régressive de voisement en néerlandais (ERNESTUS *et al.* 2006).

### 2.1. L'assimilation régressive de voisement en néerlandais

L'assimilation régressive de voisement en néerlandais est importante d'un point de vue théorique parce que cette langue se comporte apparemment de manière étrange, et que ce phénomène a un impact dans toutes sortes de théories concernant le voisement. Dès lors, nos données doivent être correctes. Je vais d'abord présenter ce que dit la littérature spécialisée sur le néerlandais, pour en venir à ce que nous avons trouvé en nous basant sur le Corpus Oral du Néerlandais<sup>2</sup>, également évoqué par Steven GILLIS (page 95).

On a les explications traditionnelles, qu'on trouve partout dans la littérature spécialisée, y compris dans celle que j'estime, et donc pas seulement dans la « mauvaise » littérature, mais bien dans des textes plutôt sérieux, écrits par des linguistes reconnus (voir par exemple BOOIJ 1999). Il y est rapporté qu'en néerlandais, si une syllabe se termine par une constrictive sourde, ce qui est systématique en raison du dévoisement en position finale – comme en allemand –, et qu'elle est suivie d'un /d/ ou d'un /b/, alors on a une

---

<sup>2</sup> Corpus Gesproken Nederlands (voir OOSTDIJK 2002).

assimilation régressive de voisement, c'est-à-dire que l'obstruante sourde devient sonore. Par exemple, si on a un /p/ suivi d'un /d/, alors ce /p/ devient /b/. En revanche, le contraire ne pourrait jamais se produire en néerlandais, c'est-à-dire que, dans une séquence /p/ – /d/, la seconde consonne soit dévoisée. Donc on n'aurait jamais d'assimilation progressive en néerlandais. C'est très important pour de nombreuses théories sur le voisement.

Nous avons procédé à l'étude de ce phénomène en recourant au Corpus Oral du Néerlandais. Étant donné la complexité de la tâche, nous avons simplifié la recherche en nous concentrant sur la partie du corpus la plus simple à transcrire : les histoires lues aux aveugles. Ces textes sont racontés de manière vivante, mais restent de l'écrit oralisé, ce qui toutefois vaut mieux que d'essayer d'imaginer dans l'abstrait comment les mots sont prononcés. Tous les mots dans lesquels une obstruante sourde précédait un /b/ ou un /d/ ont été sélectionnés et retranscrits par trois personnes. Le tableau 1 présente le nombre d'occurrences dans lesquelles nous avons trouvé une assimilation régressive de voisement, censée être obligatoire dans ces groupes consonantiques. Or on n'en trouve pas dans plus de la moitié des cas ! D'où notre surprise, puisque ce phénomène est prétendument obligatoire.

CATÉGORISATION	ASSIMILATION	N	POURCENTAGE
+sonore +sonore	Régressive	261	42,9%
-sonore +sonore	Non	121	19,9%
-sonore -sonore	Progressive	151	24,8%
absent présent		57	9,4%
autre		19	3,1%

**Tableau 1 – L'assimilation dans les histoires lues aux aveugles issues du Corpus Oral du Néerlandais (ERNESTUS *et al.* 2006 : 1042)**

Nous avons ensuite étudié le nombre de groupes consonantiques dans lesquels on ne trouvait aucun voisement, soit presque 20% des cas, ce à quoi on pouvait

encore s'attendre. Mais ce qui nous a particulièrement étonnés, c'est que 25% des cas relèvent d'une assimilation *progressive* de voisement, c'est-à-dire un phénomène censé *ne pas exister* en néerlandais, mais se produisant en réalité dans un cas sur quatre (voir tableau 1 page précédente). En somme, il ne s'agit pas d'un ou deux cas isolés, mais bien d'un nombre élevé d'occurrences. Nous avons également trouvé d'autres cas, sur lesquels je ne m'attarderai pas, où parfois la première obstruante était absente, et d'autres phénomènes encore. Mais comme le démontre cette analyse, il est indispensable d'étudier ce que les gens produisent *effectivement* dans leur quotidien pour savoir quelles configurations se trouvent dans la langue. En l'absence de données issues de corpus oraux, le simple fait d'affirmer que l'assimilation régressive de voisement est obligatoire en néerlandais ne suffit pas à décrire ce que font les locuteurs dans la réalité. Cela montre peut-être ce qu'ils « devraient » faire, mais pas ce qu'ils font. En définitive, les faits décrits dans la littérature phonologique ne sont pas toujours exacts.

## 2.2. Le recours aux corpus oraux et ses difficultés

À présent, j'aimerais parler des difficultés rencontrées lors de notre analyse. Notre étude portait sur tous ces groupes d'obstruantes, et nous avons trois transcrip-teurs – en l'occurrence des personnes qualifiées pour ce travail. Cependant – c'est un fait reconnu –, même les transcrip-teurs hautement qualifiés font beaucoup d'erreurs. Si on leur demande une première fois de transcrire un certain nombre d'obstruantes, et qu'on leur demande d'effectuer la même tâche une seconde fois, on observe souvent des différences. C'est qu'il est très difficile de faire ces transcriptions, en particulier quand la langue n'est pas articulée très clairement. Comme ce travail demande beaucoup de concentration, les transcriptions contiennent des erreurs si elles sont réalisées par des transcrip-teurs humains. Un des problèmes provient

notamment des attentes des transcrip-teurs face au phénomène analysé : s'ils croient qu'un schwa sera présent, il est probable qu'ils en entendent effectivement un. S'ils pensent que l'assimilation régressive de voisement se produit en néerlandais, ils auront tendance à transcrire ces groupes consonantiques avec des assimilations régressives de voisement. Le transcrip-teur humain, quel que soit son niveau de compétence, tend à se faire influencer par ses attentes. Voilà pourquoi on trouve des différences entre les participants.

J'ai moi-même effectué une expérimentation assez ingrate, il y a longtemps, dans laquelle j'étudiais la prononciation du mot néerlandais « *natuurlijk* » (« bien entendu », « naturellement ») (voir ERNESTUS 2000). Pour l'occasion j'ai demandé à deux excellents transcrip-teurs, de transcrire 274 occurrences de ce mot. Dans 58% des cas, il y avait désaccord sur la présence ou l'absence d'un schwa. J'étais quelque peu désemparée et j'ai décidé de ne pas étudier ces données... D'autres fois, ça s'est mieux déroulé. Par exemple, lorsqu'il a fallu transcrire le voisement des occlusives intervocaliques parmi deux mille occurrences, nous avons obtenu seulement 15% de désaccord, ce qui est beaucoup mieux. Dans cette étude-ci, nous avons trois transcrip-teurs pour travailler sur le voisement des groupes consonantiques d'obstruantes, et dans un cas sur trois, ils n'ont pas pu se mettre d'accord. C'est donc un véritable problème dans ce genre d'études. Alors que faire ? Comment gérer ces nombreux points de vue conflictuels ? En cas de désaccords entre transcriptions, une des solutions consisterait à demander à l'un ou aux deux transcrip-teurs s'ils seraient d'accord de changer leur transcription. C'est un peu risqué. Ce qui arrive en général, c'est que le plus jeune des transcrip-teurs aura davantage tendance à modifier son travail. Donc vous obtiendrez les transcriptions des personnes les plus sûres d'elles ; ce qui ne garantit pas pour

autant la qualité de la transcription. J'écarterais donc cette option. L'autre possibilité consisterait à éliminer toutes les transcriptions conflictuelles, ce qui par exemple dans mon étude, où 58% des jugements étaient problématiques – reviendrait à éliminer de grandes quantités de données, voire la plus grande partie. Et on connaît les problèmes soulevés par les travaux sur des données trop restreintes. Ceci constituait un premier problème.

Un autre problème vient de ce qu'en mettant de côté les résultats problématiques, on risque d'écarter des occurrences présentant en fait un grand intérêt. En effet, si les transcrip-teurs ne sont pas d'accord, c'est peut-être pour une bonne raison. Par exemple, si nous revenons à notre étude sur le voisement des obstruantes, il se pourrait que les transcrip-teurs n'aient pas été d'accord parce qu'il est relativement difficile de déterminer avec certitude si une occurrence est sourde ou sonore. Il est possible que le locuteur ait produit un son intermédiaire, ne permettant pas aux transcrip-teurs de trancher pour l'une ou l'autre des options, conduisant dès lors à des jugements divergents. Voilà pourquoi ces occurrences pourraient en réalité s'avérer vraiment intéressantes. Aussi me paraît-il préférable de les conserver pour les analyser, surtout si j'observe que la plupart des conflits surviennent dans un contexte particulier. En somme, si on écarte tous les jugements conflictuels, on élimine non seulement une des conditions que l'on entendait tester à l'origine, mais aussi un phénomène potentiellement digne d'intérêt pour la recherche. Exclure de telles données n'est donc, à mon avis, pas un bon choix. Alors que peut-on faire d'autre ?

Il reste en somme deux solutions. La première est de prendre la transcription de la majorité. Nous aurions pu choisir cette option et opter pour la version qui mettait d'accord au moins deux des trois transcrip-teurs. C'est envisageable, mais on perd cependant de l'information



puisque pour certaines occurrences, les trois transcrip-teurs étaient d'accord alors que pour d'autres ils ne l'étaient pas, ce qui en soi peut justement s'avérer intéressant. Au final, pour cette étude, nous avons décidé de garder toutes les transcriptions et de les analyser statistiquement. De cette manière, nous savions quelles occurrences posaient problème. Je n'entrerai pas ici dans les détails statistiques, mais je tiens à insister sur l'intérêt de conserver toutes les versions, même si elles sont conflictuelles, parce que, en définitive, ces informations sont pertinentes.

On pourrait se demander pourquoi pour cette étude sur le voisement des obstruantes, nous n'avons pas tout simplement recouru à des mesures acoustiques, d'autant plus qu'un tel traitement peut en grande partie s'effectuer de manière automatique et par exemple fournir le pourcentage d'obstruantes produites avec vibration des cordes vocales, permettant ainsi de discriminer les groupes consonantiques sourds et sonores. En fait, si nous n'avons pas recouru à cette méthode, c'est qu'elle est en réalité impossible. En effet, un trait tel que le voisement n'est pas seulement déterminé par la vibration des cordes vocales, mais aussi par la durée de la voyelle qui précède, la durée de l'obstruante, l'intensité de la friction, etc. Plusieurs dimensions seraient donc à prendre en compte, et mesurer n'est qu'une première étape. Il faudrait ensuite entrer les résultats dans des catégories spécifiques. Or nous n'avons pas la moindre idée de comment procéder à une telle catégorisation. En somme, les mesures acoustiques n'étaient pas possibles non plus, nous contraignant donc à recourir à des transcrip-teurs humains.

Pour conclure cette première partie, comme le montrent les résultats de nos recherches (ERNESTUS, LAHEY, VERHEES & BAAYEN 2006), il faut prendre avec beaucoup de prudence les descriptions fournies par la littérature spécialisée.

### 3. L'intérêt des corpus oraux

J'aimerais maintenant évoquer l'un des intérêts des corpus oraux, qui permettent aux chercheurs de découvrir des phénomènes dont ils ignoraient complètement l'existence auparavant. Je montrerai quelques-uns de ces aspects, à propos de la réduction, sur laquelle je travaille depuis plusieurs années (ERNESTUS 2000; ERNESTUS & WARNER 2011) (ERNESTUS 2014). Prenons l'exemple en français du syntagme « le ministre de la culture », extrait d'un des corpus, que nous avons compilés. Chacun sait comment prononcer le mot « ministre ». Or dans notre corpus, sa prononciation la plus fréquente est [mis], c'est tout ! (voir BRAND & ERNESTUS 2015). Un tel phénomène pourrait laisser perplexe, pourtant si on entend cette même séquence dans son contexte d'origine, on la trouve alors naturelle, même si une fois averti, on entend bien que le mot est réduit. Maintenant, ce qui est intrigant et à mon sens très intéressant, c'est que les francophones natifs n'ont pas conscience de tout le temps parler de cette manière, et je pense vraiment que les natifs le font en permanence sans en être conscients, ni en tant que locuteur ni en tant qu'auditeur. La seule façon de trouver la prononciation de mots comme celui-ci est d'étudier les corpus d'oral spontané. Prenons en anglais cette fois-ci l'exemple de *probably*. Une des prononciations possibles, et même la plus courante ressemble à [pʁɒbli]. On a deux syllabes au lieu de trois, et c'est la prononciation la plus fréquente. Ce que l'on constate, c'est que dans la conversation, les mots ont tendance à subir des réductions. Des phonèmes, voire des syllabes complètes peuvent disparaître. Ce n'est pas limité au français ou à l'anglais. On a trouvé des phénomènes identiques dans de nombreuses langues, comme en néerlandais, où *wedstrijd* (« jeu ») ressemble parfois à /ues/ au lieu de /uɛtstrɛit/. En allemand, on peut dire [va:ŋ] au lieu de [va:gan] *Wagen* (« véhicule »). En français, *fenêtre* peut se prononcer [fne:tʁ].

En espagnol, où on est censé dire [entonθes] *entonces* (*alors*), ça ressemble souvent plus à [entons] (« alors »). En tchèque, on remplace souvent la prononciation formelle [jestli/ (s)] par [ɛs] (ERNESTUS & WARNER 2011). J'insiste sur le fait qu'il ne s'agit pas d'exemples isolés.

De temps à autre, je remarque que des chercheurs utilisent les corpus oraux pour y trouver à tout prix une confirmation de leur théorie. Ainsi, partant de l'hypothèse qu'une prononciation donnée devrait exister, leur théorie se verrait confirmée s'ils en trouvent ne serait-ce qu'une seule occurrence. Je pense qu'une telle démarche est inacceptable. En effet, une occurrence unique peut tout à fait résulter d'une erreur ou d'une idiosyncrasie du locuteur. Bien qu'il s'agisse là d'une évidence, je tiens à insister sur ce point : observer une seule occurrence ne suffit pas, il en faut au contraire une grande quantité avant de pouvoir se prononcer sur les phénomènes se produisant *effectivement* dans une langue. Telle est en définitive la finalité de notre recherche. Aux États-Unis, JOHNSON (2004) a cherché à savoir avec quelle fréquence les mots se trouvaient réduits en anglais, et a établi qu'il manquait au moins un son dans 25% des mots, et au moins une syllabe dans 6% des mots. En somme, un tel phénomène s'avère relativement fréquent. Nous avons mené le même type de recherche sur le français et le néerlandais, langues pour lesquelles nous obtenons des résultats similaires (ADDA-DECKER *et al.* 2005 ; SCHUPPLER *et al.* 2011). Il y a donc toutes ces séries de données, auxquelles on peut s'intéresser pour étudier la réduction des mots, et on pense immédiatement à toutes les transcriptions déjà faites, et au temps considérable qu'elles ont demandé (voir GILLIS, page 95).

#### 4. Les défis de la transcription

Si on veut connaître la fréquence des réductions dans une langue, il faut de grandes quantités de données orales, et on

se retrouve confronté au cout élevé de la transcription, aux erreurs de transcription des transcripateurs humains, etc. Pour ces raisons, la transcription humaine nous a paru impossible, il fallait procéder autrement. Nous avons donc essayé de la faire de manière automatique à l'aide d'un dispositif de reconnaissance vocale (« Automatic Speech Recognizer » en anglais). À cet effet, nous avons utilisé le logiciel HTK (« Hidden Markov Model Toolkit »), téléchargeable gratuitement et facile à manipuler en raison de la qualité de son manuel d'utilisation. Il se base sur les modèles cachés de Markov. On l'alimente avec un fichier .wav constituant le signal de parole. Mais il faut là encore une transcription orthographique, que l'on doit faire faire à des humains s'il s'agit d'oral conversationnel. Le dispositif nécessite deux autres apports :

- 1) un dictionnaire de prononciation, indiquant quels mots peuvent se trouver dans la transcription orthographique et comment ils peuvent être prononcés. Ce qui signifie que le chercheur doit d'abord réfléchir à toutes les prononciations possibles. Par exemple, on pourrait imaginer les variantes [fənɛtʁə], [fənɛtʁ], [fnɛtʁ], [fnɛt] pour *fenêtre*, ou [ministʁə], [ministʁ], [minisʁə], [minisʁ], [minis], [mis] pour *ministre* ;
- 2) des modèles phoniques, c'est-à-dire des symboles, des lettres, qui doivent être associés au signal. En d'autres termes, les modèles phoniques sont une sorte de moyenne, une manière de montrer ce à quoi ressemblerait normalement un segment donné du signal acoustique. Donc il faut traduire ces symboles en format .wav.

Une fois toutes ces données rentrées dans le dispositif, on obtient un résultat comme celui de la figure 1 ci-après :

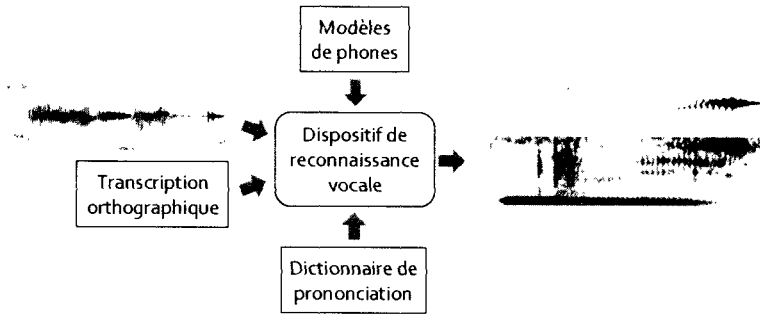


Figure 1 – Le dispositif de reconnaissance vocale automatique

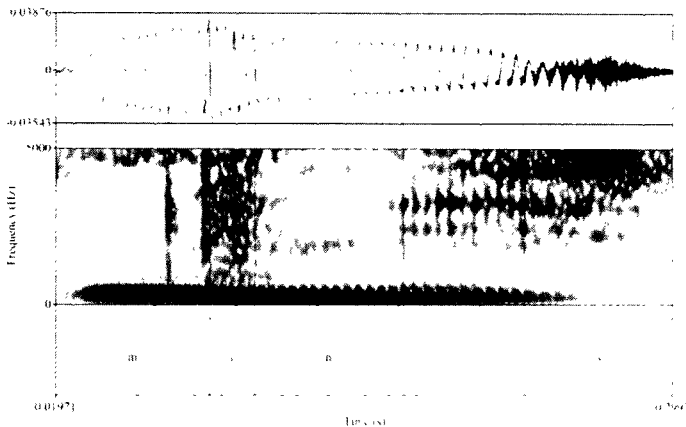


Figure 2 – Le spectrogramme d'une prononciation de *ministre*

La question est évidemment de savoir si la reconnaissance vocale automatique offre de bons résultats, ce qui dépend de plusieurs facteurs, dont deux très importants. Premièrement, la qualité des modèles phoniques, que l'on crée à partir de données orales déjà alignées au niveau des phones, par un transcripneur humain qui a procédé à ce travail d'alignement en décidant que tel son était un /s/, tel autre un /b/, etc. Plus on a de données, meilleurs seront les modèles de phones. Bien entendu, les erreurs dans la transcription phonétique –

et il y en a parce qu'elle est effectuée par des humains – se répercutent dans les modèles phoniques. Prenons par exemple le cas du néerlandais, où, pour certaines fricatives, nous confondons souvent la sourde et la sonore, par exemple /ɣ/ et /x/ ou /v/ et /f/. Donc quand on transcrit ces phonèmes, on voit apparaître beaucoup d'erreurs au niveau du voisement rendant les modèles phoniques et la transcription phonétique finale moins fiables à cet égard. Il faut par conséquent de bons modèles phoniques, dont on peut par ailleurs évaluer les limites.

Le second point important est évidemment la qualité du lexique. Le problème avec cette procédure, c'est qu'il est impossible de trouver des variantes de prononciation qui n'ont pas été introduites au préalable. Par exemple, si on n'a pas envisagé que les Français prononcent [miz] au lieu de [ministɐ] (*ministre*), alors la transcription phonétique finale ne montrera pas les occurrences de cette prononciation. En somme, on est vraiment limité par ce qu'on entre soi-même dans le lexique. Il paraît alors judicieux de mettre le plus de variantes possible, et c'est ce que nous avons fait. Mais il faut être prudent et ne pas entrer absolument tout ce qui nous passe par la tête non plus, parce que plus le dispositif de reconnaissance vocale a de choix, plus il risque de se tromper. Par exemple, si l'on rentre dans le lexique une prononciation qui ne se produit jamais, le programme va quand même la considérer comme une option valable et, de temps en temps, choisir cette option au lieu de la prononciation réelle. On voit donc toute la difficulté de la tâche. Mais au final, si l'on dispose d'une bonne reconnaissance vocale, je suis convaincue qu'il est possible d'obtenir une très bonne transcription, suffisante en tout cas pour travailler.

Nous avons comparé les transcriptions effectuées automatiquement par le dispositif de reconnaissance vocale avec celles des humains. En fait, il ne faudrait pas qu'il y ait 100% de correspondance, puisque, comme on le sait, les

humains font des erreurs. Cependant une certaine correspondance s'avère évidemment souhaitable. Dans une comparaison où nous avons regardé toutes les séquences, nous avons trouvé des désaccords avec les transcrip-teurs humains dans environ 14% des cas (SCHUPPLER, ERNESTUS, SCHARENBOG & BOVES 2011) ; ce qui paraît acceptable. Dans une autre étude, nous avons étudié la présence ou l'absence de schwa (HANIQUE *et al.* 2013), qui, comme je l'ai évoqué plus haut, est une voyelle particulièrement difficile à transcrire. Nos résultats montrent que le dispositif de reconnaissance vocale est en désaccord dans 26% et 23% des cas, suivant le transcrip-teur. Les transcrip-teurs humains eux-mêmes étaient en désaccord dans 18% des cas. La différence entre ces trois résultats n'est pas significative. Donc il semble que le transcrip-teur automatique se comporte comme un humain.

On s'est ensuite interrogé sur la portée de ces constats en ce qui concerne la durée des séquences. Les durées trouvées par le transcrip-teur automatique sont-elles comparables à celles établies par les humains ? Nous pensons que oui. Dans au moins 95% des cas, nous avons une différence inférieure à vingt millisecondes, un résultat assez similaire à ce qui est décrit dans la littérature sur les différences entre transcrip-teurs humains et automatiques.

Comme il est très difficile de faire travailler pendant de nombreuses heures des transcrip-teurs humains sur de l'oral spontané, il nous reste cette option et je pense que la qualité est suffisante s'il s'agit d'effectuer des recherches quantitatives et d'étudier, par exemple, la fréquence à laquelle un segment de mot disparaît. Toutefois pour des recherches phonétiques plus précises, les transcriptions humaines s'avèrent nécessaires, mais il faut alors limiter la quantité de données.

Pour clore cette partie sur les transcriptions humaine et automatique, j'aimerais, en rapport avec l'exposé de Steven

GILLIS (voir page 95), faire une remarque sur la fusion des corpus. Steven GILLIS suggère qu'en réunissant nos corpus – pas tous bien sûr –, on bénéficierait d'un potentiel accru, avec des données livrant davantage d'informations. Il faut toutefois rester prudent à cet égard. Des différences existent entre les corpus parce que les situations de production ne sont pas exactement identiques même s'ils ont été enregistrés et transcrits de manière similaire.

Le Corpus Oral du Néerlandais propose divers contextes de production : conversations spontanées en face-à-face, entretiens avec des enseignants de néerlandais, conversations téléphoniques, etc., bref, toutes sortes de contextes qui peuvent paraître similaires. Dans une de nos études, nous avons choisi dix mots néerlandais ayant une forte tendance à se réduire, et nous avons cherché à établir à quelle fréquence, en fonction des situations de production, ils se voyaient réduits de manière extrême, c'est-à-dire d'au moins une, voire deux syllabes. Comme l'illustre la figure 3, le pourcentage d'occurrences de mots fortement réduits varie de 5% à 40%.

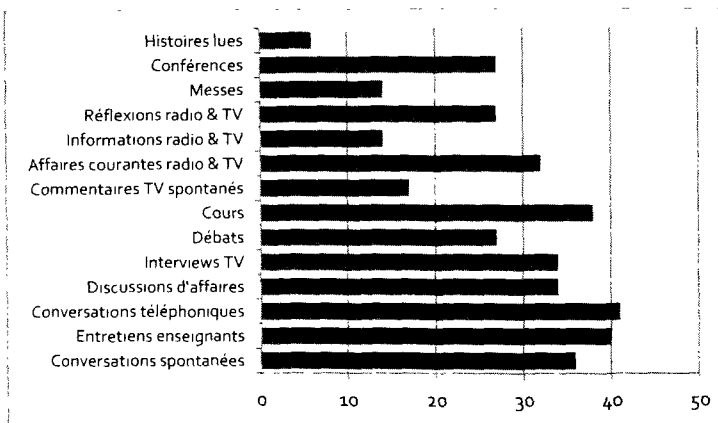


Figure 3 – Variation d'apparition de dix mots néerlandais très fréquents et extrêmement réduits selon la situation de production (ERNESTUS *et al.* 2015)



Les histoires lues aux aveugles font apparaître un taux de réduction de 5,5%, ce qui paraît conforme à l'intuition. Dans le cas des entretiens avec les enseignants de néerlandais et des conversations téléphoniques, ce taux se situe aux alentours de 40%. On pourrait alors dire que ce sont des registres de langue très différents, expliquant ainsi la variation. Cependant on observe également des différences parmi les conversations spontanées. Le pourcentage est plus élevé dans le cas des entretiens avec les enseignants (40%) que pour les conversations spontanées (36%). J'ignore la raison pour laquelle un tel écart existe. Quoi qu'il en soit, même de légères différences entre situations d'enregistrement peuvent déboucher sur des résultats différents. En somme, il faut être prudent si on compare ou rassemble des corpus, car il suffit d'une simple différence de situation pour faire varier les résultats. Si on veut comparer plusieurs langues, le choix de corpus similaires s'impose.

Pour ce faire, nous avons créé une série de trois corpus d'oral familier, enregistrés de la même manière : un en français (TORREIRA *et al.* 2010), un en espagnol (TORREIRA & ERNESTUS 2010) et un en tchèque (ERNESTUS *et al.* 2014). Dès lors, quand on utilise les données de ces corpus, on peut vraiment comparer les langues, comme par exemple le débit de parole. Et comme on sait qu'il existe des différences entre les corpus, il est préférable d'en utiliser beaucoup afin d'éviter que les pourcentages calculés soient uniquement basés sur la nature du corpus retenu. En somme, d'un côté, il faut être prudent et n'utiliser que des corpus similaires, et de l'autre, ces résultats montrent qu'on doit utiliser différents types de corpus.

J'ai consacré la première partie de mon exposé à insister sur la manière dont les mots sont réellement prononcés, ce qui constitue une question plutôt linguistique. Maintenant, j'aimerais passer à des considérations plus psycholinguistiques en essayant de répondre à la question de

savoir comment les locuteurs réduisent leur discours oral, ce qui conduira également à interroger les raisons pour lesquelles les locuteurs réduisent certains segments. Il s'agira ensuite de déterminer si l'absence d'un segment est catégorielle ou graduelle.

#### 5. L'apport des corpus oraux en psycholinguistique

Dans notre recherche, nous avons utilisé divers corpus pour voir quand les réductions apparaissent (ERNESTUS *et al.* 2014 ; TORREIRA *et al.* 2010 ; TORREIRA & ERNESTUS 2010). Nous avons trouvé, sans surprise, que les réductions sont plus fréquentes quand le débit de parole est plus rapide, quand le mot se trouve au milieu d'un syntagme et qu'il n'est pas accentué (HANIQUE *et al.* 2013 ; PLUYMAEKERS *et al.* 2005a ; b). Dans les trois cas, les mots sont prononcés à une vitesse relativement élevée. Mais le débit n'est pas le seul facteur déterminant. La nature du mot suivant est un autre bon prédicteur. Il y a plus de réduction si le locuteur n'a pas de problème à articuler le mot suivant. Si un mot est suivi d'une hésitation, il n'y aura pas de réduction, alors qu'il y en aura plus si le mot est suivi d'un mot fortement probable. Par exemple, pour les phrases « children swim » (*Les enfants nagent*) et « children smoke » (*Les enfants fument*), le mot « children » tend à être plus réduit dans la première que dans la seconde, simplement parce que « children smoke » est une phrase qu'on ne produit pas souvent ; elle est peu probable, ce qui la rendrait plus difficile à articuler pour les locuteurs. En somme, le locuteur réduirait quand il n'est pas obligé de ralentir, et qu'il n'a pas besoin de temps pour préparer le mot suivant. En définitive, il semblerait que les mots réduits constituent la norme, les mots non-réduits survenant uniquement lorsque le locuteur rencontre un problème et doit gagner du temps. Que nous apprennent ces résultats sur les raisons de la réduction ? D'après les données, bien qu'il s'agisse là de spéculation, nous pensons que les locuteurs

réduisent afin de diminuer leur effort articulatoire. Quant à la question de savoir à quel point le locuteur prend en compte son interlocuteur, il nous est impossible d'y répondre sur la base de nos études sur corpus.

Si nous savons désormais *quand* les mots sont réduits, comment avons-nous déterminé les facteurs susceptibles d'influencer le degré de réduction? Nos corpus sont constitués d'oral spontané. Aussi, conformément à notre volonté de recueillir de l'oral aussi spontané que possible, aucune indication n'était fournie quant au sujet de conversation durant les séances d'enregistrement. On ne savait jamais ce qu'on allait obtenir, et la variété des sujets de conversation s'avère très importante. Dès lors, les mots recueillis dans le corpus se révèlent très différents les uns des autres. Ils sont d'ailleurs produits par des locuteurs eux aussi différents. Chaque mot présente sa propre identité. Certains mots sont en première position dans la phrase, d'autres à la fin et d'autres encore au milieu. Les uns sont accentués, les autres non. Leur fréquence peut varier ainsi que la fréquence du mot qui précède et du mot qui suit. Ils diffèrent donc sur de nombreux points. Il n'en reste pas moins que nous aimerions identifier les facteurs responsables de leur réduction, et pour y parvenir, il faut accumuler beaucoup de données. Il ne s'agit pas de recueillir seulement cent ou deux cents occurrences, mais plusieurs milliers. C'est la raison pour laquelle un transcritteur automatique s'avère indispensable. À défaut, la recherche devient irréalisable.

## 6. L'utilisation des corpus oraux et le modèle statistique

Maintenant évoquons brièvement la dimension statistique, et plus particulièrement les modèles linéaires à effets mixtes. On utilise un modèle de régression, et on prédit la réduction ou non d'un mot ou bien la durée du segment, et on entre chaque prédicteur pour voir s'il améliore les résultats. À titre d'illustration, imaginons que

l'effet du débit de parole nous intéresse. On a alors le « débit de parole » comme prédicteur. On doit cependant gérer beaucoup d'autres variables, parce que si on n'entre que le prédicteur de débit de parole, on ne va probablement rien trouver. Il faut en effet prendre en compte tous les autres facteurs susceptibles de jouer un rôle. On ajoute donc la « fréquence du mot », sa « position prosodique », etc. Toutes ces variables de contrôle sont indispensables, sinon on n'obtient aucun résultat, et il faut par ailleurs recourir à des outils statistiques complexes. Notre expérience avec ce modèle n'est pas récente, et nous avons essayé beaucoup d'approches différentes. Dans la littérature scientifique, paraît tous les six mois un nouvel article sur la bonne manière d'utiliser ce modèle, mais les opinions continuent de diverger à cet égard. Nous avons étudié ces propositions, et nous sommes désormais assez conscients des écueils. Deux points fondamentaux ressortent.

Premièrement, il faut être attentif aux prédicteurs hautement corrélés. Un article de WURM & FISICARO (2014) évoquait la possibilité de rentrer pratiquement tous les prédicteurs, même s'ils sont corrélés, mais pas s'ils sont *hautement* corrélés. Par exemple, si on s'intéresse à l'impact de la fréquence d'occurrences d'un mot, et qu'on rentre parallèlement comme prédicteur le nombre de phonèmes – ce qui représente un exemple de deux prédicteurs hautement corrélés –, le modèle ne peut pas vraiment déterminer lequel des deux a le plus d'impact. Il est donc impossible de rentrer plusieurs variables hautement corrélées. On pourrait imaginer que pour résoudre le problème, il suffirait d'éliminer par exemple le prédicteur du 'nombre de phonèmes' et de conserver celui de la 'fréquence d'occurrence'. Toutefois, si une telle solution nous permet d'observer un impact du prédicteur retenu, on ne peut pas garantir qu'il résulte uniquement de la fréquence d'occurrence et non pas de

l'absence du prédicteur 'nombre de phonèmes', préalablement écarté.

Deuxièmement, il ne faut pas entrer tous les prédicteurs auxquels on pourrait à priori penser. En effet, il est parfois tentant d'essayer d'en intégrer le maximum – sans justification théorique – juste « pour voir » et observer les interactions possibles. Cependant, une telle stratégie est insatisfaisante. On obtient parfois un modèle à priori très performant, qui prédit bien les données, mais dès qu'on le teste sur un autre ensemble de données les prédicteurs spécifiques ne fonctionnent plus. En somme, on se retrouve avec différents ensembles de prédicteurs pour différents ensembles de données, les premiers résultats obtenus se révélant non-généralisables à de nouveaux ensembles de données. Évidemment, ce n'est pas ce qu'on veut. Voilà pourquoi j'insiste : il faut faire attention au nombre de prédicteurs qu'on utilise et aux interactions. Certains soutiennent que pour un prédicteur, il faut au moins trente points de données. Je pense que c'est insuffisant et qu'il en faut beaucoup plus<sup>3</sup>.

#### 7. L'apport des corpus oraux dans la détermination de la nature de la réduction des mots

Abordons désormais ma dernière question – typiquement psycholinguistique – sur la nature de la réduction des mots : est-elle graduelle ou catégorique ? Dans le premier cas, on pourrait alors la considérer comme le résultat d'un processus phonétique ou articulatoire. Imaginons un mot comportant un schwa au départ dont on réduirait de plus en plus la durée. Ce schwa finira par complètement disparaître, et uniquement en raison de cette réduction progressive, soit un processus phonétique ou articulatoire, qu'on peut qualifier de

---

<sup>3</sup> Il existe d'autres manières de résoudre ce problème que je n'évoquerai pas ici.

« graduel ». À l'inverse, il se pourrait que le schwa soit absent en raison d'une règle phonologique, qui, par exemple dans un mot comme *fenêtre*, impliquerait la chute du /ə/ et aboutirait à /fnetʁ/. On pourrait également envisager que le locuteur possède deux variantes de prononciation dans son lexique mental – /fənetʁ/ et /fnetʁ/ – entre lesquelles il peut choisir. Il est en fait impossible d'établir une distinction entre règle phonologique et variantes intégrées dans le lexique mental, mais une différence claire existe entre réduction graduelle et catégorique. La question est alors de savoir si les corpus peuvent nous apporter des éléments de réponse.

On peut bien entendu étudier la réduction en laboratoire en faisant des expérimentations, mais on n'aura plus affaire à de l'oral spontané. Il serait donc intéressant de tirer parti des corpus oraux pour obtenir des informations supplémentaires. À cet effet, nous avons étudié à quel moment surviennent les réductions, en partant de l'hypothèse que si, dans les mêmes conditions, un segment (par exemple un schwa, un /t/) est absent et plus réduit, l'absence et la réduction sont soumises aux mêmes prédicteurs. En d'autres termes, celles-ci résultent d'un processus identique, en l'occurrence nécessairement phonétique, puisque la réduction est un processus phonétique. Nous constatons que c'est effectivement le cas pour la chute du /t/ en position finale en néerlandais (HANIQUÉ *et al.* 2013).

PRÉDICTEUR DE PRÉSENCE/ABSENCE	PRÉDICTEUR DE DURÉE
débit	débit
registre	registre
segment précédent	segment précédent
	dernier mot du segment
	dernier mot du segment & débit

**Tableau 2 – Prédicteurs de l'absence du /t/ final ou de sa réduction en néerlandais**

Comme l'indique la colonne de gauche du tableau 2, le /t/ est plus souvent absent à débit rapide, dans un registre de langue plus familier, et lorsque le segment précédent est un

/s/ ou un /m/ plutôt qu'une autre consonne. La durée du /t/ obéit également à ces mêmes prédicteurs. En d'autres termes, les prédicteurs pertinents pour l'absence du /t/ le sont aussi pour sa durée, ce qui laisse à penser que l'absence et la durée du /t/ sont déterminées par le même processus, qui doit donc être *graduel, phonétique, articulatoire*. Toutefois, la colonne de droite fait apparaître deux autres prédicteurs, en l'occurrence la position du mot en fin de segment, ainsi qu'une interaction avec le débit de parole (signalée par '&'). On pourrait à priori en conclure que les prédicteurs de l'absence de /t/ diffèrent de ceux de sa durée. Certes, mais ces ensembles de données montrent malgré tout que les mêmes prédicteurs ont une influence et sur l'absence et sur la durée du /t/, et ce en raison du fait que dans le cas de la durée, on a juste une puissance statistique plus grande. Dès lors, il est bien plus probable de trouver des prédicteurs additionnels si la durée constitue la variable dépendante. En définitive ceci prouverait que l'absence du /t/ final en néerlandais résulte d'un processus phonétique.

L'autre possibilité concerne les cas dans lesquels les prédicteurs déterminant l'absence ou la durée d'un segment sont très différents. Ainsi en est-il du schwa en position médiane en français (par exemple [fənɛtʁ], *fenêtre*), comme l'illustre le tableau 3 (BÜRKI *et al.* 2011).

PRÉDICTEUR DE PRÉSENCE/ABSENCE	PRÉDICTEUR DE DURÉE
débit	débit
position du schwa dans le mot	position du schwa dans le mot
position du mot dans la phrase	voisement de la consonne suivante
nombre de consonnes de la séquence	voisement de la consonne précédente
respect du principe de sonorité <sup>4</sup>	probabilité d'apparition du mot avec le mot précédent

Tableau 3 – Prédicteurs de l'absence du schwa médian ou de sa réduction en français

<sup>4</sup> NdE : Correspondant à l'échelle : « occlusive sourde < occlusive sonore < fricative sourde < fricative sonore < nasale < liquide » (BÜRKI *et al.* 2011 : 3983).

À débit rapide, le schwa disparaît plus souvent. Déterminante s'avère également la position du mot dans l'énoncé, et, dans le cas où la disparition du schwa donne lieu à un groupe consonantique, il faut regarder combien de consonnes le composent et si le principe de sonorité y est respecté. Pour la durée du schwa, on observe que le débit et la position du segment dans le mot sont importants. Trois nouveaux prédicteurs interviennent également : le voisement de la consonne suivante, celui de la consonne précédente, et la probabilité d'apparition du mot avec le mot précédent. Nous sommes donc en présence de deux ensembles de prédicteurs différents, ce qui prouverait qu'en français, l'absence du schwa peut résulter d'un processus catégorique, comme une règle phonologique ou encore d'un stockage du mot sans schwa dans le lexique mental. Ces résultats confirment les données en production que nous avons collectées en laboratoire (BÜRKI *et al.* 2010).

À ce stade, il faudrait encore répondre à la question de savoir à partir de quel moment on peut dire que deux ensembles de prédicteurs diffèrent. Au cours de cet exposé, j'ai montré deux exemples, et il me semble que mes arguments suffisent pour affirmer que, dans un cas, il y avait un ensemble commun de prédicteurs, et dans l'autre, deux ensembles distincts. On peut cependant imaginer qu'il est parfois impossible de trancher, d'autant plus que la méthodologie n'est pas toujours évidente.

Pour conclure, j'insisterai sur le fait que les corpus de langue, et plus particulièrement les corpus d'oral spontané ont une réelle importance pour notre recherche. Ils nous révèlent en effet les processus de production de la parole en conversation naturelle, et la véritable prononciation des mots. J'ai montré qu'ils pouvaient s'avérer utiles tant pour des questions linguistiques, que psycholinguistiques, même si de nombreuses interrogations subsistent, auxquelles les corpus ne permettent pas de répondre.



## 8. Conclusion

Dans cet exposé, je n'ai pas du tout abordé la question de la compréhension, pour laquelle, on peut en partie recourir aux corpus. Par ailleurs, de nombreux points précis de psycholinguistique restent difficiles à traiter à l'aide d'un tel outil. Aussi paraît-il souhaitable, dans la mesure du possible, de recourir de manière complémentaire aux corpus et aux expérimentations en laboratoire pour voir si l'on obtient des résultats convergents.

Toutefois, comme je l'ai montré, il faut garder à l'esprit que l'utilisation des corpus n'est pas exempte de problèmes. Un grand nombre de conversations s'avèrent nécessaires, si possible transcrites automatiquement, entraînant dès lors un traitement statistique complexe, parsemé de nombreux écueils, dont il faut savoir se préserver. De plus, la comparaison des corpus est souvent malaisée, en raison de la différence des niveaux de langue. Enfin, il peut s'avérer difficile de déterminer si deux ensembles de prédicteurs sont identiques ou pas.

En définitive, la recherche psycholinguistique ne peut pas se passer des corpus d'oral spontané, car sans eux, on aurait seulement accès à ce que les locuteurs produisent lors d'expérimentations en laboratoire, ce qui manifestement ne correspond pas à ce qu'ils font réellement dans leur quotidien.

## Questions

Légende : « Q » pour « Question », « ME » pour « Mirjam ERNESTUS »

**Q :** Vous n'avez pas parlé des dysfluences, alors que je suppose que c'est un problème constant quand on utilise des corpus oraux. Comment les gérez-vous, par exemple, quand vous effectuez une transcription automatique ?

**ME :** Avant d'effectuer la transcription automatique, on filtre en écartant tous les éléments perturbateurs, comme les rires et les silences. En fait, on ne donne pas l'entièreté du signal au

transcripteur automatique. Au lieu d'un seul bloc d'une heure et demie, on découpe nos fichiers en segments d'environ deux secondes chacun, ce qui nous permet d'isoler les dysfluences. Donc il n'y en a pas vraiment, seulement dans les parties non-filtrées de la transcription.

**Q :** Mais... ici, une question méthodologique se pose : vous ignorez les dysfluences, alors que certaines personnes les considèrent comme une part essentielle du langage.

**ME :** Absolument. Mais on a aussi montré – pas seulement nous – que les dysfluences ont une influence sur le degré des réductions. Donc on ne les ignore pas ; ce n'est que pour la transcription automatique qu'on les filtre, parce qu'il est impossible de les transcrire. Sinon, le lexique des prononciations contiendrait un nombre gigantesque de transcriptions pour chaque mot. C'est pourquoi on les ignore dans nos transcriptions automatiques. Mais de toute évidence, elles sont très importantes pour la recherche.

**Q :** Puisqu'on parle de dysfluences, que se passe-t-il quand on hésite ? En français par exemple, si on dit « euh » avant « fenêtre » : est-ce que vous savez si les gens ont tendance à prononcer plus souvent le schwa dans cette condition ?

**ME :** À ma connaissance, ça n'a jamais été démontré. Mais le contraire, oui : les locuteurs prononceront « fenêtre » en entier si une dysfluence vient juste après, mais pas dans l'autre sens.

**Q :** Quand vous parlez de transcription humaine, dans un sens, c'est une manière de « discrétiser » un phénomène continu. A-t-on des moyens pour coder cette continuité ? Parce que, en cas de conflit entre transpositeurs, on peut penser que le phénomène en soi n'est en réalité pas catégorique. Est-ce qu'on arrive à coder, à rendre compte du flou des données ?

**ME :** On y a pensé, mais on ne l'a pas fait. Une des choses qu'on a songé à faire est qu'au lieu d'avoir des prédicteurs binaires en 0 ou 1 – sourd ou sonore, par exemple – on pourrait utiliser des valeurs intermédiaires, comme  $\frac{2}{3}$  ou  $\frac{1}{3}$  pour refléter le désaccord des transpositeurs.

**Q :** Avez-vous une idée de comment on pourrait le visualiser ? Parce qu'on aimerait pouvoir lire ces transcriptions.

**ME :** Je n'ai pas de solution qui me vienne à l'esprit. C'est un problème qui nous a préoccupés, mais on n'a pas encore essayé de trouver une solution. Normalement, on fait toutes les mesures acoustiques, car je pense qu'elles représentent ce continuum. Après, bien sûr, il y a le problème de la multiplicité des propriétés acoustiques d'un phénomène à prendre en compte, particulièrement si vous vous intéressez au voisement, par exemple. Mais les mesures acoustiques rendent compte du continuum. Voilà ce qu'on a fait pour le moment.

**Q :** Une idée pourrait être de superposer à la transcription une sorte d'indication visuelle de la confiance que vous accordez à la transcription, ou au désaccord des transcrip-teurs.

**Q :** Votre exemple des phrases « children swim » et « children smoke » me rappellent un des travaux de PIANTADOSI<sup>5</sup>, qui parlait d'efficacité dans la communication. Sa théorie était qu'on investit plus d'effort dans l'articulation lorsqu'on s'attend à ce que l'auditeur ait des difficultés à comprendre. Il a une théorie complète et un modèle statistique là-dessus.

**ME :** Oui, c'est une grande question en psycholinguistique : à quel point est-ce que vous ajustez votre production verbale en fonction de l'auditeur, à quel point est-ce que vous adaptez votre discours par rapport à lui ? Il y a des gens qui pensent que ces réductions résulteraient de la prise en compte de l'auditeur dans la mesure où, en tant que locuteur, vous utiliseriez ce qui est préférable pour votre auditeur, et que vous réduiriez, parce que vous savez que l'auditeur a entendu ce mot il y a peu de temps, et qu'il n'est pas important, alors que ne pas le réduire reviendrait au contraire à signaler à votre auditeur que ce mot est important. À vrai dire, je ne crois pas du tout à cette théorie, et ce pour deux raisons : premièrement, je ne peux tout simplement pas croire qu'en tant que locuteur, on soit capable de savoir ce qui est préférable pour l'auditeur. Deuxièmement, de nombreuses études montrent que ça ne fonctionne pas comme ça. Il y a une étude – je suis d'ailleurs jalouse de son auteur, j'aurais vraiment adoré la faire moi-même – faite par Ellen BARD de l'Université d'Édimbourg (BARD *et al.* 2000). On sait qu'en ce qui concerne la réduction, plus on

---

<sup>5</sup> NdE : voir PIANTADOSI *et al.* (2011).

répète un mot dans une conversation, et plus il est réduit. On pourrait dire que vous signalez à votre auditeur qu'il a déjà entendu ce mot avant. BARD a conçu une expérience dans laquelle deux personnes avaient devant elles deux cartes semblables mais pas identiques : dans l'une, un itinéraire était indiqué, et le participant qui l'avait devant les yeux devait l'expliquer à l'autre personne. Certains éléments revenaient périodiquement dans la conversation, comme par exemple le mot « cathédrale », dans des phrases comme « Tourne à droite après la cathédrale ! » auxquelles l'interlocuteur répondrait par « Je ne vois pas de cathédrale », etc. Donc vous imaginez bien que ces mots sont de plus en plus réduits, au fur et à mesure qu'ils sont employés. Ensuite, les chercheurs ont gardé le « locuteur-guide » mais lui ont présenté un autre auditeur, en donnant à chacun deux nouvelles cartes, la grande question étant « Qu'est-ce qui va se passer pour le mot "cathédrale" ? ». Le locuteur-guide va-t-il continuer à réduire ou au contraire repartir à la case zéro et réduire à partir de là ? On constate que le locuteur-guide continue tout simplement à réduire, sans se préoccuper du changement d'auditeur, ce qui indique clairement qu'on ne tient pas compte à ce point de l'auditeur.

**Q :** Comment était organisée l'expérience ? Ils avaient un retour ?

**ME :** Oui, tout le temps.

**Q :** Un retour visuel ?

**ME :** Non, mais ils pouvaient se parler.

**Q :** Mais pas de retour visuel.

**ME :** Il y avait plusieurs situations différentes pour cette expérience : parfois, il y avait un contact visuel, parfois pas.

**Q :** Parce que, quand vous dites qu'on ne s'adapte pas à ce niveau de précision, je suis tout à fait d'accord. Mais on s'adapte quand même plus précisément lorsqu'on a un retour visuel.

**ME :** Certes, mais ça ne signifie pas que si vous réduisez, le mot va devenir incompréhensible. Ça ne signifie pas que si vous réduisez plus qu'on ne s'y attend, l'auditeur va tout de suite signaler qu'il ne comprend pas le mot.

**Q :** J'ai une question au sujet de la reconnaissance vocale. Je suppose que la réponse est la même que pour les dysfluences, mais que faites-vous du bruit, des chevauchements dans une conversation ?

**ME :** Pour les chevauchements, c'est la même réponse que dans le cas des dysfluences : on découpe le signal en segments d'environ deux secondes et on vérifie que les limites entre ces segments se situent à des endroits logiques, de préférence entre des silences, de façon à perdre le moins possible de données. Mais on laisse les chevauchements de côté, parce que c'est impossible à gérer pour la reconnaissance vocale. Si on a trop de bruit, c'est pareil. Là encore, ça veut dire qu'on n'a pas de transcription automatique de ces segments, mais on s'y intéresse quand même. Évidemment que ce qui se passe pendant les chevauchements nous intéresse, mais on ne peut pas traiter cette question d'après des transcriptions automatiques.

### Références

- ADDA-DECKER Martine, BOULA DE MAREÛIL Philippe, ADDA Gilles & LAMEL Lori. (2005). Investigating Syllabic Structures and their Variation in Spontaneous French. *Speech Communication* 46-2, 119-139.
- BARD Ellen Gurman, ANDERSON Anne H., SOTILLO Catherine, AYLETT Matthew, DOHERTY-SNEDDON Gwyneth & NEWLANDS Alison. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language* 42-1, 1-22.
- BOOIJ Geert. (1999). *The phonology of Dutch*. Clarendon: Oxford University Press.
- BRAND Sophie & ERNESTUS Mirjam. (2015). Reduction of Obstruent-Liquid-Schwa Clusters in Casual French. *Proceeding of the 18<sup>th</sup> International Congress of Phonetic Sciences, ICPHS0251.1-5*.
- BÜRKI Audrey, ERNESTUS Mirjam & FRAUENFELDER Ulrich H. (2010). Is There Only one "Fenêtre" in the Production Lexicon? On-Line Evidence on the Nature of Phonological Representations of

- Pronunciation Variants for French Schwa Words. *Journal of Memory and Language* 62-4, 421-437.
- BÜRKI Audrey, ERNESTUS Mirjam, GENDROT Cédric, FOUGERON Cécile & FRAUENFELDER Ulrich Hans. (2011). What Affects the Presence versus Absence of Schwa and its Duration: A Corpus Analysis of French Connected Speech. *Journal of the Acoustical Society of America* 130-6, 3980-3991.
- ERNESTUS Mirjam. (2000). *Voice Assimilation and Segment Reduction in Casual Dutch*. Utrecht: LOT (Landelijke Onderzoeksschool Taalwetenschap).
- ERNESTUS Mirjam, LAHEY Mybeth, VERHEES Femke & BAAYEN R Harald. (2006). Lexical Frequency and Voice Assimilation. *The Journal of the Acoustical Society of America* 120-2, 1040-1051.
- ERNESTUS Mirjam & WARNER Natasha. (2011). An Introduction to Reduced Pronunciation Variants. *Journal of phonetics* 39-3, 253-260.
- ERNESTUS Mirjam. (2014). Acoustic Reduction and the Roles of Abstractions and Exemplars in Speech Processing. *Lingua* 142, 27-41.
- ERNESTUS Mirjam, KOČKOVÁ-AMORTOVÁ Lucie & POLLAK Petr. (2014). The Nijmegen Corpus of Casual Czech. *LREC 2014: 9th International Conference on Language Resources and Evaluation*, 365-370.
- ERNESTUS Mirjam, HANIQUE Iris & VERBOOM Erik. (2015). The Effect of Speech Situation on the Occurrence of Reduced Word Pronunciation Variants. *Journal of phonetics* 48, 60-75.
- HANIQUE Iris, ERNESTUS Mirjam & SCHUPPLER Barbara. (2013). Informal Speech Processes can be Categorical in Nature, even if they Affect many Different Words. *The Journal of the Acoustical Society of America* 133-3, 1644-1655.
- JOHNSON Keith. (2004). Massive Reduction in Conversational American English. In YONEYAMA Kiyoko & MAEKAWA Kikuo (Eds), *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10<sup>th</sup> International Symposium*, Tokyo: The National International Institute for Japanese Language, 29-54.

- OOSTDIJK Nelleke. (2002). The Design of the Spoken Dutch Corpus. In SMITH Adam, COLLINS Peter & PETERS Pam (Eds), *New Frontiers of Corpus Research*, Amsterdam: Rodopi, 105-112.
- PIANTADOSI Steven T., TILY Harry & GIBSON Edward. (2011). Word Lengths are Optimized for Efficient Communication. *Proceedings of the National Academy of Sciences* 108-9, 3526-3529.
- PLUYMAEKERS Mark, ERNESTUS Mirjam & BAAYEN R. Harald. (2005a). Articulatory Planning is Continuous and Sensitive to Informational Redundancy. *Phonetica* 62-2/4, 146-159.
- PLUYMAEKERS Mark, ERNESTUS Mirjam & BAAYEN R. Harald. (2005b). Lexical Frequency and Acoustic Reduction in Spoken Dutch. *The Journal of the Acoustical Society of America* 118-4, 2561-2569.
- SCHUPPLER Barbara, ERNESTUS Mirjam, SCHARENBERG Odette & BOVES Lou. (2011). Acoustic Reduction in Conversational Dutch: A Quantitative Analysis Based on automatically Generated Segmental Transcriptions. *Journal of phonetics* 39-1, 96-109.
- TORREIRA Francisco, ADDA-DECKER Martine & ERNESTUS Mirjam. (2010). The Nijmegen Corpus of Casual French. *Speech Communication* 52-3, 201-212.
- TORREIRA Francisco & ERNESTUS Mirjam. (2010). The Nijmegen Corpus of Casual Spanish. *Seventh Conference on International Language Resources and Evaluation (LREC'10)*, 2981-2985.
- WURM Lee H. & FISICARO Sebastiano A. (2014). What Residualizing Predictors in Regression Analyses does (and what it does not do). *Journal of Memory and Language* 72, 37-48.