

La transcription synchronisée des corpus oraux.

Un aller-retour entre théorie, méthodologie et traitement informatisé¹

Anne DISTER et Anne Catherine SIMON

Centre de recherche VALIBEL – UCLouvain

{anne.dister,anne-catherine.simon}@uclouvain.be

Résumé

Les corpus oraux ont pour particularité qu'aux données primaires (les enregistrements) s'ajoutent des données secondaires (les transcriptions) nécessaires à leur exploitation. Cet article examine

- les conséquences, sur les conventions de transcription à adopter, de l'évolution des outils informatisés pour l'exploitation des corpus oraux ;
- l'implication qu'ont les conventions de transcription sur les types de requêtes et d'analyses qu'on peut mener sur les données ;
- comment la transcription synchronisée, en alignant le texte et le son, facilite l'accès au son mais introduit un artefact dû au choix d'une unité temporelle d'alignement.

De manière plus particulière, cet article examine l'évolution des conventions de transcription et d'alignement² utilisées dans la banque de données VALIBEL et analyse les problèmes et les conséquences de ces conventions pour l'annotation et la recherche linguistique sur corpus.

1. Grands principes des conventions de transcription

Toute analyse linguistique de productions orales est impossible à partir de la seule source sonore. En effet, le chercheur aura beau écouter et réécouter encore les enregistrements, il ne peut les appréhender uniquement par le biais du son. Ceux-ci ne pourront devenir objets d'étude à part entière qu'à partir de leur mise en/par écrit. La parole reste fluide, essentiellement fugace, même après avoir été captée sur bande sonore :

On ne peut pas étudier l'oral par l'oral, en se fiant à la mémoire qu'on en garde. On ne peut pas, sans le secours de la représentation visuelle, parcourir l'oral en tous sens et en comparer les morceaux. (Blanche-Benveniste 2000 : 24)

¹ Cet article a bénéficié des commentaires éclairants de Michel Francard et de Philippe Hambye. Toute erreur ou imprécision y subsistant est à attribuer à la seule responsabilité des auteures.

² Avec le logiciel Praat (Boersma et Weenink 2007).

On est donc là devant un paradoxe, mis en évidence par de nombreux chercheurs (Blanche-Benveniste et Jeanjean 1987, Blanche-Benveniste 1997 et 2000, Gadet 2003, Raingeard et Lorscheider 1977) : pour approcher l'oral, on doit « en passer » par l'écrit.

Dès ses débuts en 1989, le centre de recherche VALIBEL (acronyme pour *Variétés Linguistiques du français en Belgique*), qui constitue et exploite de grands corpus oraux³, a réfléchi à ses pratiques de transcription et établi des conventions de transcription explicites.

Celles-ci ont d'emblée suivi quatre grands principes⁴ : adoption de l'orthographe standard, non-recours à la ponctuation de l'écrit, valorisation de l'oralité des corpus et compatibilité avec un traitement informatisé.

1.1. Adoption de l'orthographe standard

Pour les recherches linguistiques autres que celles concernant directement la forme phonique (phonétique, phonologie, prosodie), VALIBEL transcrit en utilisant l'orthographe traditionnelle. Ceci est loin d'être le cas dans toutes les pratiques de transcription, où l'orthographe traditionnelle est souvent adaptée, dans la volonté de marquer des particularités de prononciation. Ces particularités apparaissent à travers la graphie et non dans une notation phonétique. Ainsi, le *Guide de présentation de mémoires et de thèses en linguistique et conventions pour la transcription de conversations* (Losier et al. 2002) détaille les conventions que doivent adopter les étudiants pour la transcription de corpus acadiens. Voici un extrait de ces conventions :

- **chus** pour *je suis*.
Il importe ici de garder le *s*, puisqu'il s'agit du verbe à la première personne. Il serait possible de transcrire *j'suis*, mais ce ne serait pas exact à cause de l'élision du *i*.
- **j'sais, j'savais, j'chante**, etc.
Il y a élision du *je* devant une consonne.
- **j'sais pas** pour *je ne sais pas*.
Même si le *sais* est prononcé [ʃe], il est préférable d'adopter cette transcription plutôt que *che pas*.
- **i** pour *il* et *ils*, **a** ou **alle** pour *elle*.
- **celle-cite** pour *celle-ci*
- **à c't'heure** pour *maintenant*.
- **qu** et **ti** prononcés [tʃ] sont écrits *tch*, comme dans **tchequ'un** (quelqu'un), **tchèque** part (quelque part), **moitché** (moitié), **pitché** (pitié), **tcheu** fou (quel fou).

³ Ces corpus forment aujourd'hui le plus grand corpus informatisé de données textuelles orales en francophonie : une banque de données de près de 4 millions de mots.

⁴ Voir Francard et Péronnet (1989) pour une première formulation.

Ces déformations graphiques affectent aussi bien des pronoms, des suites pronom+verbe que des lexèmes. Le but est de calquer, dans la graphie, les variantes de prononciation répandues dans le français acadien.

Ces aménagements graphiques sont clairement liés, dans les conventions citées ci-dessus, à une conception des unités linguistiques du système observé. L'idéologie sous-jacente est que *chus* est une unité à part entière du système acadien, et non pas une variante marquée de l'oral. Cet exemple illustre en quoi une transcription est une analyse.

Ces modifications graphiques participent de ce que Blanche-Benveniste et Jeanjean (1987) appellent des *trucages orthographiques*⁵. Trucages, parce que ces formes ne correspondent pas à des graphies connues d'un scripteur/lecteur de la langue, telles que répertoriées dans les ouvrages de référence par exemple, tout en y renvoyant néanmoins. Dans de telles transcriptions, on est entre les deux : ni tout à fait de l'oral, ni tout à fait de l'écrit. En outre, une représentation de la phonétique mêlée à l'orthographe est très imparfaite et est source d'erreurs puisqu'à chaque occurrence d'une forme (par exemple, les pronoms *il* ou *ils*), il faut prêter attention à sa prononciation exacte et la traduire par une graphie non standard (par exemple, *y* dans le cas d'une prononciation [i]).

On ne peut nier l'effet de stigmatisation qu'engendre la lecture d'une transcription ainsi aménagée sur les propos de la personne et, par là, sur la personne elle-même (Gumperz et Berenz 1993 : 97). Ces altérations graphiques (notamment l'élision de voyelles) sont d'ailleurs fréquemment utilisées par les écrivains, depuis le XIX^e siècle, pour stigmatiser le parlé de personnes issues de classes populaires.

Dans les publications scientifiques, l'utilisation de telles adaptations graphiques a indéniablement des implications sur l'analyse. Considérons le texte suivant, transcription d'une conversation entre élèves de CE2⁶.

Olivia : Oh ! le hamster ?

⁵ Giovannoni et Savelli (1990) parlent quant à elles de « bricolage orthographique », Gadet (2003) d'« aménagement graphique », Raingeard et Lorscheider (1977) de « bâtards phonético-orthographiques ». Dans le corpus édité par ces dernières, on est donc étonnées de trouver *main'ant* pour *maintenant*, *pa'ce* que pour *parce que*, *c'qui faut* pour *ce qu'il faut*, et *'pis* pour *et puis*, *i'* veulent pour *ils veulent*, *i' faut* pour *il faut*, *c't-à dire* pour *c'est-à-dire*, etc.

⁶ Cette transcription est extraite de Verstraete (2000 : 154-156). Nous avons respecté strictement la typographie et la mise en page originales. Le gras note que l'élève lit ; les majuscules marquent l'accentuation ; les doubles barres l'hésitation entre deux interprétations /« beuh », peur/ ; E indique une indétermination du locuteur. Ponctuation et didascalies ont été ajoutées par Verstraete en vue de sa publication.

Notre seule intervention concerne la graphie de *maitresse*, qui suit, comme tout notre texte, les recommandations orthographiques de 1990.

Sélalé: Hein !! Eh elle sait même pas c' que c'est qu'un hamster !!!

Lévy : C'est une bête hé !

Sélalé : C'est un p' tit truc : ça ressemble un peu à une souris !!

Olivia : (*très sûre d'elle*) Mais NON ! un HAMSTER ! Pas un animal !

Lévy : Un hamster c'est un animal.

Sélalé : Ben oui c'est un p'tit animal qui ressemble un peu à une souris...

Olivia : (*rires moqueurs*) C'est un animal ? un hamster ! ...Eh r' garde (*parle au magnétophone, en articulant bien*) Un hamster est un animal ! (*rires moqueurs*) ... (*puis se rangeant brutalement à l'avis des autres*) Ouais, j'en ai un chez moi.

Kévin : (*plus vite*)...**que faisaient les Gaulois pour effrayer les esprits ?** (*lit la question 3*)

Tarik : arrête Vincent !

Vincent : Y f'saient /« Beuh », « peur »/ !

EEE : *Rires*

Vincent : Ou bien y cassaient leurs couilles

Rire de Karen

Maitresse : chut

Vincent : y faisaient /la mort/

Tarik : Attends fais voir...

Kévin : Ils tuaient les gens...

Karen : Non !

Vincent : Ben ouais p'têt ça...

Karen : Non quand même pas ! ...Non les Gaulois y faisaient pas ça ! ... Je sais qu'ils ont tué... mais...

On voit les trucages utilisés ici : élision marquée au moyen de l'apostrophe (*p'tit, c'que, p'têt, f'saient, r'garde*), ce qui produit un effet d'ailleurs assez étrange pour *r'garde*, dont la mise en page laisse le *r'* seul en fin de ligne ; utilisation de la graphie *y* pour noter la prononciation du pronom *ils*⁷.

Par ailleurs, informant le lecteur du cadre sociologique dans lequel s'est effectué son travail, l'auteure rapporte que l'école « est située dans une zone urbaine sensible », que six élèves (sur vingt) ont déjà doublé une fois, que « la plupart des élèves ont un rapport au savoir qui n'est pas toujours direct », que certains ont des difficultés psychologiques et sont suivis par un psychologue (2000 : 153-154). Elle précise également que les paroles des enfants n'ont

⁷ Risquant par là de rendre la transcription ambiguë, y étant dans nos habitudes de lecture le pronom ou la lettre de l'alphabet.

pas été censurées : « elles sont retranscrites telles quelles dans la crudité du langage des élèves...⁸ ».

Inévitablement, les choix de transcription et le profil des élèves présentés ne font que cumuler leurs effets négatifs et risquent de faire passer pour particulier, pour spécifique ce qui ne l'est pas. La stigmatisation déjà possible dans la description du profil des élèves se voit renforcée par les choix de transcription. En fait, comme le dit Gadet (2003 : 30), « seule une graphie sans aménagement ni réécriture limite le risque de stigmatiser un énoncé ».

Concernant le lexique, les variantes lexématiques non standard sont transcrites suivant les principes de l'orthographe française de référence (par exemple, nous transcrivons l'emprunt au wallon *péquet* 'eau-de-vie de genièvre', et non *pékèt*, qui serait la graphie adoptée en wallon). Dans ce domaine, VALIBEL a été la première équipe francophone à prendre en compte la variation diatopique (géographique) dans ses corpus informatisés.

Si le choix de l'orthographe standard nous semble plus neutre que l'adoption de trucages orthographiques, il comporte néanmoins une part d'analyse, puisqu'on rapporte une série de variantes de prononciation à une forme graphique unique. La mise en mots graphiques de productions orales implique un travail est à la fois analytique et interprétatif, qui a parfois été qualifié de « traduction » (Cheepen 1995), de « représentation heuristique » (Mondada 2000) ou encore de « déformation » (Bally 1935).

En conclusion, outre leur effet de stigmatisation, les aménagements graphiques qui concernent la morphologie ou certaines variantes de prononciation ont selon nous quatre inconvénients majeurs : 1) il est difficile de les systématiser ; 2) il est parfois difficile de distinguer des variantes de prononciation à la seule écoute du signal sonore ; 3) ces déformations perturbent nos habitudes de lecture en compliquant le système graphique, et obligent le lecteur à oraliser pour décoder ; 4) ce mode de transcription complique la tâche du transcripateur, surchargeant son travail par l'attention qu'il doit porter à ces phénomènes, dont la fréquence est souvent très élevée.

Néanmoins, lorsque la prononciation effective est très éloignée de la prononciation standard, la possibilité est laissée au transcripateur d'ajouter à la transcription orthographique une transcription phonétique entre crochets en utilisant l'alphabet Sampa⁹.

⁸ Les points de suspension sont le fait de Verstraete.

⁹ Speech Assessment Methods Phonetic Alphabet, <http://www.phon.ucl.ac.uk/home/sampa/french.htm>

1.2. Non-recours à la ponctuation de l'écrit

Les corpus de langue parlée transcrits à des fins de recherche linguistique ne sont en général pas ponctués. En effet, il n'y a pas de correspondance stricte entre phénomènes prosodiques¹⁰ et ponctuation écrite. À une petite pause ne correspond pas nécessairement une virgule à l'écrit, pas plus qu'à une pause plus grande, une ponctuation forte. Il n'y a pas de relation bi-univoque entre les deux. De plus si certains énoncés se laissent enfermer relativement facilement dans le moule de la ponctuation graphique, comment s'en sortir avec des productions dans lesquelles foisonnent répétitions de mots, inachèvements et autres phénomènes propres à l'oral qui se construit (cf. 1.3) ?

Blanche-Benveniste et Jeanjean (1987 : 139) plaident pour des transcriptions non ponctuées : « La ponctuation, si on la met trop tôt, préjuge de l'analyse syntaxique et impose un découpage sur lequel il est difficile de revenir. » Selon elles, en ponctuant, le transcripteur « suggère[e] une analyse avant de l'avoir faite » (1987: 142).

En fait, l'absence de ponctuation va de pair avec la remise en cause de la notion de phrase à l'oral, ainsi que le dit Gadet :

Comme pour toutes les études de phénomènes oraux, la séquence fondamentale ne correspond généralement pas à ce que l'on entend par « phrase » à l'écrit. Il faut donc se passer de cette catégorie. (1992 : 69)

C'est ainsi que la notion de phrase, dans les études de productions orales, a été abandonnée et que s'est posée la question de l'unité minimale pertinente à prendre en considération¹¹.

Cependant, une transcription sans aucune autre marque de segmentation que les espaces entre les mots et les changements de locuteurs est très difficilement lisible. C'est pourquoi la plupart des systèmes de transcription s'accordent à reconnaître deux ou trois degrés de « pauses », étant entendu qu'il ne s'agit pas nécessairement de pauses silence (rendues par un silence acoustique mesurable) mais bien d'une perception subjective de segmentation, laquelle peut être rendue par une pause, par un allongement vocalique, par un contour intonatif, voire par la présence d'une frontière syntaxique qui influence le découpage du discours.

¹⁰ Que ce soit les pauses (silencieuses ou pleines) ou les contours intonatifs (montants, descendants, ou complexes). Nous parlons ici d'oral non planifié et non d'oral obtenu à partir d'un texte lu.

¹¹ Voir l'ouvrage collectif dirigé par Berthoud et Mondada (2000) pour la position des différents auteurs sur le sujet ; voir aussi Béguelin (2000 et 2002), Simon (2001).

1.3. Valorisation de l'oralité des corpus

1.3.1. « Disfluences »

Nous conservons dans la transcription les traces de l'élaboration de l'énoncé, traces souvent appelées dans la littérature *disfluences*. Nous notons ainsi

- les pauses pleines ou ponctuants (*euh, ben, bon, etc.*) ;
- les répétitions de mots ou de suite de mots ;
- les amorces de morphèmes.

Ce travail nécessite une grande attention de la part du transcripateur, pour noter des phénomènes qui sont habituellement gommés dans une écoute ordinaire. En effet, notre oreille semble ignorer totalement ces marques qui passent donc la plupart du temps inaperçues¹² tant elles sont communes dans l'oral spontané.

1.3.2. Tour de parole

Comme nous l'avons mentionné ci-dessus pour justifier l'absence de ponctuation des corpus, la question de l'unité pertinente à l'oral est loin d'être résolue (cf. 4.2) et une solution qui évite de faire intervenir un modèle théorique dès la transcription est de considérer l'unité, envisagée comme relativement neutre, du tour de parole. On donne alors au tour de parole un statut non pas théorique mais purement graphique, qui devient dans la transcription une unité visuelle par défaut, une unité pratique dotée de caractéristiques objectivables qui ne seraient pas définies par des partis-pris théoriques¹³. Définir le tour de parole de façon « neutre » revient à l'envisager de la sorte : un tour de parole correspond à l'occupation matérielle du canal de parole par un locuteur ; le tour de parole s'achève lorsqu'un nouveau locuteur prend la parole à son tour (cf. ci-dessous 1.3.3 pour la parole superposée).

Dans nos transcriptions, la succession des tours de parole se présente de manière horizontale : les paroles des locuteurs se succèdent, de haut en bas sur la « page » (ou à l'écran) et chaque paragraphe¹⁴ représente l'intervention d'un locuteur. Cette disposition, que Edwards (1995) appelle le format vertical, est cohérente avec nos habitudes de lecture : on

¹² Sauf dans des exercices spécifiques mis en place lors de tests de perception comme ceux de Candéa (2000) ou de Duez (1997) par exemple, où l'attention du sujet est explicitement attirée sur ces phénomènes.

En ce qui concerne notre œil, le correcteur orthographique de Microsoft attire d'ailleurs notre attention sur les formes répétées contiguës en les soulignant en rouge.

¹³ Marty Laforest (1992 : 31) note, à propos du tour de parole : « Cette notion est toujours donnée, jamais construite ».

Notre « tour de parole » ne doit donc pas recevoir le sens qu'il reçoit dans les analyses de la linguistique interactionnelle. Il s'agit bien pour nous d'une unité de transcription.

¹⁴ On parle ici de paragraphe au niveau graphique, car du point de vue informatique, il ne s'agit que d'une seule ligne.

commence par lire en haut de la page, et ce qu'on lit d'abord se déroule dans le temps avant ce qu'on lit ensuite, plus bas sur la page. Les textes des pièces de théâtre adoptent depuis longtemps ce format, qui fait partie de notre culture pour représenter la succession des paroles des protagonistes de la conversation¹⁵.

1.3.3. Chevauchement de parole

Dans les conversations, il est très fréquent que deux (ou plusieurs) locuteurs parlent en même temps. On a alors des séquences où la parole est superposée. Dans des études interactionnelles, il est important de garder une trace de ces chevauchements, et pas uniquement des paroles prononcées.

L'équipe du GARS (Groupe Aixois de Recherche en Syntaxe) utilise le soulignement pour noter les chevauchements de parole, comme ci-dessous :

transcription du GARS¹⁶

L2 on le tirait par la queue

L1 il y avait il y a avait on le tirait par la queue voilà eh c'est

Le même échange dans les conventions VALIBEL prend la forme suivante, où la barre et le tiret délimitent la portion de texte chevauchée (tiret à droite de la barre pour le début du chevauchement, tiret à gauche de la barre pour la fin du chevauchement) :

transcription de VALIBEL

L2 on le tirait |- par la queue

L1 il y avait il y a avait -| on le tirait par la queue voilà eh c'est

Dans ce cas (L1 qui avait la parole cesse de parler après le chevauchement et L2 poursuit), aucune différence n'apparaît à la lecture des deux formats de transcription. Par contre, les deux types de conventions impliquent une disposition spatiale différente lorsque, après le chevauchement de parole, le locuteur qui avait la parole la conserve :

transcription du GARS¹⁷

L1 ces choses-là j'en

L2 plus ou moins

L1 parle -je ne raconte pas une je vais pas raconter tout ça comme ça tu vois

transcription de VALIBEL

¹⁵ Edwards oppose cette disposition au format en colonnes (« columnar » format), où une colonne est dévolue à un locuteur. Pour une implication théorique de ces choix de disposition spatio-visuelle, voir Edwards (1995) et l'article fondateur de Ochs (1979).

¹⁶ Cette transcription est extraite de Blanche-Benveniste *et al.* (2002 : 47).

¹⁷ La pause brève est symbolisée par un tiret au GARS, par une barre oblique entourée de blanc chez VALIBEL. Cet extrait est issu de Blanche-Benveniste *et al.* (2002 : 152).

L1 ces |- choses-là j'en <L2> plus ou moins -| parle / je ne raconte pas une je vais pas raconter tout ça comme ça tu vois

Dans les transcriptions du GARS, le chevauchement de L2 apparaît visuellement dans un tour de parole séparé, alors que dans les conventions VALIBEL il est inséré dans la ligne d'un autre locuteur : il est de ce fait moins visible et risque de ne pas être interprété comme un tour de parole à part entière¹⁸. La différence la plus flagrante concerne sans doute le sort réservé au tour de parole de L1 : le découpage opéré par le GARS en fait deux tours de parole, là où VALIBEL en conserve un seul.

Ainsi, la représentation visuelle d'un même phénomène a des implications sur les unités de segmentation du discours, alors même que celles-ci avaient reçu une définition ne semblant pas devoir prendre en compte ce type de variation. Cette disposition différente a des implications directes sur le comptage des tours de parole : pour cette courte séquence, trois au GARS, contre un seul chez VALIBEL.

En fait, comme nous ne donnons pas de statut théorique à l'unité « tour de parole », c'est véritablement l'occupation du canal qui guide la transcription, indépendamment des paroles prononcées. Ainsi, le cas des régulateurs verbaux de type *mm* ou *oui*, souvent appelés « pseudo-tours » dans l'analyse interactionnelle : si la séquence régulatrice est prononcée en même temps que les paroles du locuteur qui occupe le canal, elle figure dans le tour de parole de celui-ci ; si elle n'est pas prononcée de façon superposée (pendant une pause de l'autre locuteur par exemple), elle apparaît dans un tour séparé. La seule analyse du transcripteur concerne donc bien le chevauchement ou non des paroles, et non le statut à leur accorder.

1.4. Compatibilité avec l'informatisation des données

D'emblée, les transcriptions ont été conçues en vue d'une utilisation informatique des données. Les choix de transcription ont évolué en fonction du développement des outils informatiques, lesquels offrent de nouvelles perspectives de recherche (cf. 3).

2. Conventions de transcription VALIBEL

Les quatre grands principes rappelés au point précédent ont guidé les pratiques de transcription du Centre VALIBEL même si, au cours des années, ces pratiques ont évolué parallèlement au développement de nouveaux outils ou en fonction des caractéristiques des données recueillies (conversations bilingues ou multi-locuteurs, etc.). Ces conventions de

¹⁸ La définition de ce chevauchement par VALIBEL va de pair avec sa disposition spatio-visuelle : « chevauchement interne au tour de parole du locuteur ».

transcription sont appliquées à l'ensemble des textes qui entrent dans sa banque de données¹⁹. Elles sont explicitées dans un manuel à destination des transcrip-teurs (Dister *et al.*, 2006). Nous les résumons brièvement.

2.1.1. Marques

marque	signifié
/	pause brève ²⁰
//	pause longue
(silence)	silence
(toux), (rire), (chuchotement)	commentaires para-verbaux de type : toux, rire, chuchotement, etc.
(x)	passage incompréhensible d'une syllabe
(xx)	passage incompréhensible de plusieurs syllabes
(xxx)	passage incompréhensible relativement long
cou/	amorce de morphème
cou/ -pure	amorce achevée sans reprise antérieure
?	question de forme déclarative à contour intonatif montant
-	début de chevauchement
-	fin de chevauchement
--	début de chevauchement impliquant plus de 2 locuteurs
--	fin de chevauchement impliquant plus de 2 locuteurs
§	début de conversations en parallèle (séquences simultanées)
§	fin de conversations en parallèle (séquences simultanées)
§	sépare 2 conversations en parallèle
{choix1 , choix2}	multitranscriptions
{incertain}	transcription incertaine

¹⁹ Nos conventions s'appliquent au format des données telles que nous les stockons. En ce qui concerne la publication scientifique par exemple, on peut modifier certaines conventions ou la mise en forme du texte pour des questions de meilleure lisibilité ou de mise en évidence du phénomène étudié. Cette remarque se fonde sur le principe que le format de stockage et d'exploitation (format textgrid ou texte) ne coïncide pas nécessairement avec le format de publication (dont on peut obtenir automatiquement divers types, à l'aide par exemple du logiciel Transformer, cf. Ehmer 2006).

²⁰ Les 3 degrés de pauses sont jugés intuitivement par le transcrip-teur, c'est-à-dire non corrélés à des seuils de mesures acoustiques.

[sampa]	indications de phonétique en Sampa
---------	------------------------------------

2.1.2. Locuteurs

Les locuteurs reçoivent chacun un code unique composé de 5 lettres et 1 chiffre : 3 lettres minuscules pour le nom du sous-corpus auquel l'enregistrement est rattaché, 2 lettres en majuscules pour les initiales du locuteur, un chiffre pour différencier les éventuels homonymes. Le chiffre 0 est réservé au locuteur qui a le rôle d'intervieweur, si un tel rôle est dévolu dans l'enregistrement.

Ce code est noté devant chaque tour de parole. Lorsque le locuteur prend la parole dans un chevauchement interne, ce code est indiqué entre balises : <mmmNN1>.

2.1.3. Enregistrement

Chaque texte est balisé au début et à la fin de la transcription. Il reçoit le code du locuteur principal de l'enregistrement, et l'extension *r* (pour *recording*). Les balises de début et de fin de texte sont respectivement : <deb id="mmmNN1r"> et <fin id="mmmNN1r">. Ces indications sont utiles afin de séparer les textes lorsque ceux-ci sont compilés automatiquement dans un seul fichier.

3. Évolution des outils informatiques

Ces dernières années, les outils informatiques ont fortement évolué : plus performants, plus rapides, ils permettent des exploitations autrefois inconcevables. Cette évolution est également liée à la capacité de stockage, dont les limites sont toujours repoussées plus loin. Elle a eu des répercussions sur les transcriptions elle-mêmes et sur les exploitations que l'on peut en faire.

3.1. Conséquences de l'alignement du texte et du son sur le degré de précision de la transcription

Notre choix d'utiliser l'orthographe standard pour transcrire les données orales, y compris celles qui relèvent d'un style très informel, vise à ne pas stigmatiser a priori les productions verbales, mais aussi à éviter toute préanalyse (cf. 1.1). Ce choix a pour conséquence de perdre au niveau de la transcription une partie de la variation observable dans les données sonores (élision des pronoms, chute de schwa, etc.). Dans notre optique, cet inconvénient est tout relatif : une transcription qui utilise des trucages orthographiques est plus couteuse en temps (pour chaque prononciation non standard, le transcripteur doit écouter et analyser exactement ce qui a été produit) et est rarement homogène (si l'on rajoute de l'interprétation au niveau de

la transcription, il est plus difficile de maintenir une pratique cohérente entre différents transcrip-teurs) (cf. les inconvénients décrits au point 1.1).

En outre, nous pensons que si un chercheur veut analyser, par exemple, les réalisations du marqueur de discours *enfin*, il devra de toute manière réécouter chaque occurrence et la coder selon ses propres critères de classification (réduction à la syllabe [fɛ º], prononciation [mãfɛ]²¹, etc.). Le maintien de l'orthographe standard permet de retrouver aisément toutes les occurrences ; l'alignement du son sur le texte permet de réécouter chaque occurrence ; la fonction d'étiquetage (cf. 4.2) permet de les coder. Le tout sans intervenir sur la transcription orthographique qui sert de guide pour naviguer dans l'enregistrement. La synchronisation du texte au son justifie aussi qu'on ait de moins en moins besoin de recourir à des inserts de transcription phonétique dans la transcription orthographique.

3.2. Conventions typographiques et concordanciers

Au début des recherches, VALIBEL utilisait un concordancier qui avait des répercussions concrètes sur le format des transcriptions. Ainsi, une espace devait être insérée après l'apostrophe, afin de ne pas reconnaître la forme élidée et la forme qui la suit comme un seul mot²² : on écrivait *l' amoureux* (avec espace) et non *l'amoureux*, pour pouvoir effectuer des recherches sur la forme *amoureux* et obtenir les occurrences dans une concordance. Dans un mouvement inverse, des traits d'union étaient insérés entre les termes d'une forme composée n'ayant pas de trait d'union dans l'orthographe standard : *parce-que, tandis-que, etc.*

L'utilisation d'un concordancier performant (Unitex, cf. Paumier 2006) n'oblige plus à ce type d'adaptations graphiques, qui nécessitaient deux versions des transcriptions : l'une consultable pour la lecture, l'autre pour la recherche d'occurrences.

Une autre modification qui a son importance pour les transcriptions est le format du texte. Au départ, nous travaillions sur des transcriptions au format .doc (format utilisé par Microsoft Word) qui permet une mise en forme du texte. Certaines conventions étaient ainsi notées en italique : c'était notamment le cas lorsque le locuteur utilise une autre langue que le français, en général pour des passages en langue régionale (wallon, luxembourgeois) dans nos corpus. Actuellement, nos fichiers sont au format texte simple. En effet, Unitex utilise des fichiers en

²¹ Le cas de [mãfɛ] illustre l'alternative entre la graphie *mais enfin* qui ramène la forme prononcée à une forme standard et la graphie *m'enfin* qui représenterait une lexicalisation du marqueur, et non un trucage orthographique. Quelle que soit la forme graphique utilisée, elle permet cependant de retrouver toutes les occurrences à partir d'une recherche sur la forme *enfin*.

²² Le logiciel ne traitait pas l'apostrophe comme un séparateur.

unicode, et les textes alignés dans Praat (cf. 4.1) sont sauvés sous un format texte. Concrètement, cela signifie que toute mise en forme particulière des caractères est perdue (italique, soulignement²³, etc.). Nous avons donc dû adapter nos conventions ; les passages dans une langue autre que le français sont maintenant encadrés par des balises de type XML.

Notons que nous n'avons jamais utilisé dans nos transcriptions les notes de bas de page. Nous n'avons jamais non plus joué sur la disposition graphique, comme le font certaines équipes pour, par exemple, aligner l'un en dessous de l'autre les chevauchements de parole. Si cette présentation a des avantages incontestables de lisibilité, elle est étroitement contrainte par la police et la taille des caractères. Modifier l'une ou l'autre nécessite de la part du transcripateur un ajustement manuel de l'alignement, qui peut se faire au cas par cas pour une publication mais est inenvisageable pour un format de stockage et de diffusion des données²⁴. De toute façon, ces indications sont inutilisables en format texte.

3.3. Types de recherches dans les transcriptions

L'utilisation d'un concordancier évolué comme Unitex permet de traiter des phénomènes liés à la temporalité des données (comme les pauses) dans les recherches effectuées sur le corpus. Les transcriptions comprennent 3 degrés de pauses, notées intuitivement dans le texte. La pause pouvant a priori apparaître n'importe où dans un énoncé – bien que l'on constate de fortes régularités –, une recherche sur une séquence de mots devait prévoir la possibilité d'insertion d'une pause entre chaque mot. Grâce aux modifications introduites dans les programmes d'Unitex, on peut retrouver une séquence donnée, même si celle-ci est interrompue par une pause.

De plus, le projet d'étiquetage morphosyntaxique actuellement en cours (Dister 2007) permettra d'envisager des recherches non plus sur des formes particulières telles qu'elles se rencontrent dans les textes mais bien sur des lemmes (formes canoniques) ou encore sur des catégories grammaticales. Le corpus ainsi étiqueté offre de nouvelles perspectives de recherche, inenvisageables jusqu'ici. La flexibilité d'Unitex permet également de combiner les requêtes de type syntaxique avec des contraintes proprement conversationnelles, par

²³ Utilisé au GARS pour la parole superposée.

²⁴ Cette exigence de conventions qui ne varient pas d'un traitement de texte à l'autre, d'une mise en page à l'autre, nous semble essentielle en vue du partage des données. Quand on sait le temps que prennent la récolte des données orales et leur transcription minutieuse, la conservation des données dans chaque centre de recherche apparaît comme une aberration. L'heure n'est plus (ou ne devrait plus) être à ce que chacun garde jalousement ses corpus. Se posent, dans cette optique de diffusion, de nouvelles contraintes liées notamment à l'anonymisation.

exemple pour formuler une recherche d'un certain « type » (par exemple la particule *enfin*) dans une certaine position (un segment de discours chevauché ou chevauchant), en combinant l'étiquetage morphosyntaxique avec les symboles temporels (comme les symboles | - et -| d'ouverture et de fermeture de chevauchement).

4. Transcription alignée

On n'utilise plus aujourd'hui, pour transcrire, un lecteur de cassettes audio muni d'une pédale permettant de revenir en arrière et de réécouter plusieurs fois le même segment. D'une part parce que la numérisation des données audio permet de les écouter sur un ordinateur, d'autre part parce qu'on transcrit en synchronisant le texte avec le son correspondant. Nous n'allons pas aborder les contraintes techniques de numérisation du son en lien avec l'analyse acoustique des données (cf. Espesser 2003) mais envisager les particularités d'une transcription orthographique alignée, réalisée avec le logiciel Praat.

4.1. Le logiciel Praat pour segmenter et transcrire

Il y a plusieurs raisons, externes et internes, d'adopter le logiciel Praat pour réaliser des transcriptions alignées (pour autant qu'on travaille sur des données audio, et non vidéo) :

- le logiciel Praat est distribué librement par ses concepteurs, Paul Boersma et Daniel Weenink (voir www.praat.org), avec pour conséquence qu'il est largement utilisé dans la communauté scientifique, pas seulement celle des phonéticiens, initialement visés par ce programme d'analyse phonétique, mais par beaucoup de chercheurs en linguistique sur corpus oraux ;
- de nombreux manuels présentent les fonctions de segmentation et d'étiquetage (*labelling and segmentation*) qui sont utilisées pour la transcription synchronisée ;
- Praat permet de traiter des sons très longs : il accepte les fichiers d'une taille maximale de 2 giga bytes, ce qui correspond à 3 heures de son stéréo de qualité CD ou à 12 heures de son mono échantillonné à 22050 Hz (16-bit) ;
- le fichier de sortie d'une transcription alignée sous Praat (textgrid) est un format ouvert²⁵, qui prend la forme d'un fichier texte structuré de telle manière que chaque segment de transcription correspond à un intervalle temporel du fichier son ; le fichier texte résultant est peut lisible comme tel, mais peut être transformé à l'aide d'un script

²⁵ Un fichier en format propriétaire ne peut être lu et modifié qu'à partir du logiciel spécifique utilisé pour le créer. À l'inverse, un format ouvert permet de lire le fichier avec n'importe quel logiciel conçu à cet effet. Cette distinction entre format propriétaire vs ouvert devient pertinente dès qu'on souhaite échanger ou diffuser des fichiers.

ou d'un logiciel²⁶ en un fichier texte qui adopte le format (vertical, en colonnes) choisi par l'utilisateur ;

- le format textgrid permet d'attribuer une couche d'annotation (*tier*) par locuteur, afin de noter les phénomènes spécifiques à l'oral (chevauchements de parole, silences, etc.).

Jusqu'à présent, nous avons expérimenté les inconvénients suivants lors de l'utilisation de Praat pour transcrire :

- le mode de visualisation du texte dans la fenêtre Praat (cf. figure 1) est relativement inconfortable et peut provoquer certaines erreurs (coquilles) qui seraient évitées dans un traitement de texte traditionnel ; le fait que le texte soit coupé n'importe où, selon la taille de la fenêtre, rend difficilement visible l'omission d'une espace, le fait qu'un symbole (par exemple de pause) est collé au mot qui précède (au lieu d'en être séparé), etc. ; par conséquent, toute transcription réalisée sous Praat doit être relue en utilisant un autre logiciel (par exemple Transformer, cf. Ehmer 2006) ;
- certains types d'interaction, qui font intervenir un grand nombre de locuteurs (comme une interaction en classe), compliquent la manipulation de l'interface car les paroles de chaque locuteur sont inscrites dans une couche d'annotation séparée (on a une idée du résultat si l'on imagine que le textgrid présenté à la figure 1 contient non pas 3, mais 20 couches empilées) ; un logiciel comme Transana (pour la transcription de vidéos) ne présente pas ce problème, puisque le transcripteur insère manuellement le nom du locuteur au début de chaque ligne ;

²⁶ Plusieurs scripts prennent un fichier textgrid en entrée pour le transformer en un fichier texte sans repères temporels (scripts développés par VALIBEL, entre autres). Des logiciels comme Winpitch (logiciel gratuit sur demande, développé par Philippe Martin, www.winpitch.com) ou Transformer (logiciel payant développé par Oliver Ehmer, <http://www.oliverehmer.de/transformer/>) proposent des fonctions similaires avec un large choix de formats de sortie (doc, txt, xml, etc.).

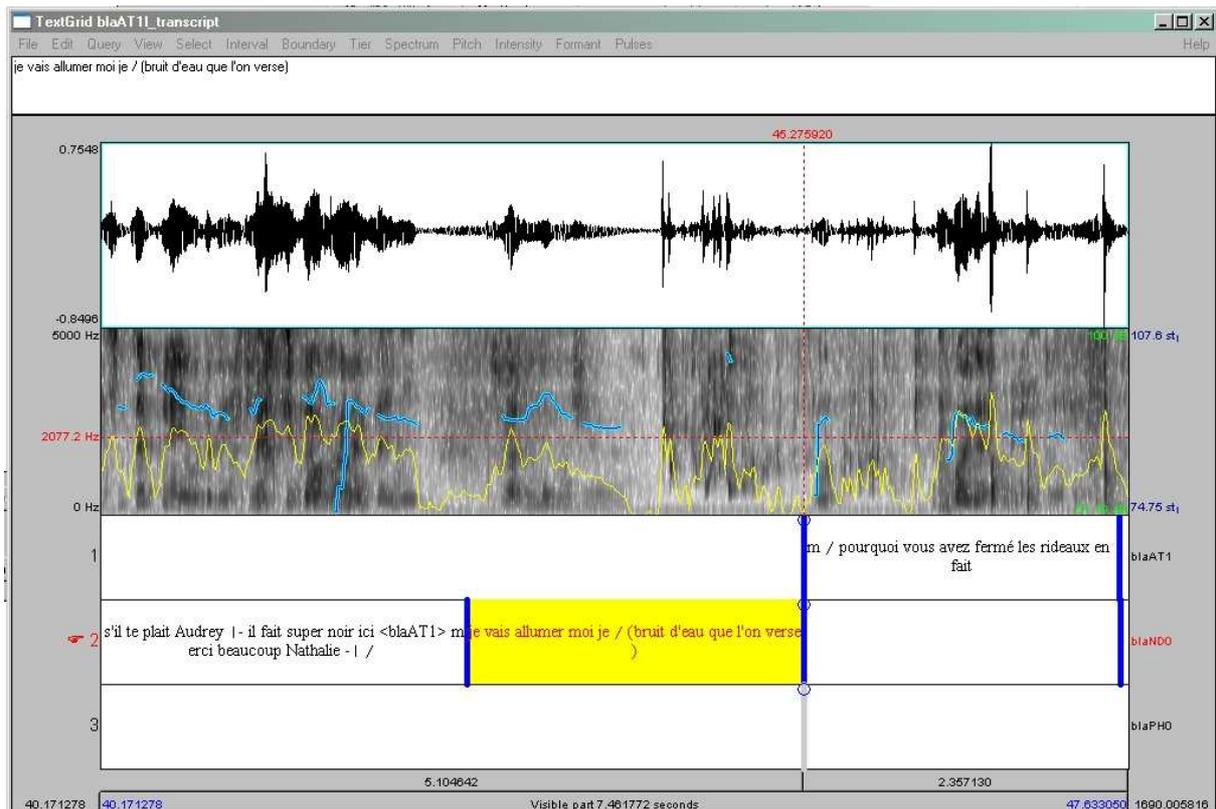


Figure 1 : fenêtre d'édition de Praat utilisée pour la transcription synchronisée

- la nécessité d'un script ou d'un logiciel pour transformer le textgrid dans un format plus lisible (pour la relecture, la publication ou la consultation des transcriptions) impose une opération supplémentaire et distingue le format des données telles qu'elles sont stockées et échangées du format de publication.

En conclusion, il n'y a plus aucune raison technique, aujourd'hui, pour transcrire un fichier son de manière non synchronisée. La transcription alignée pose par contre les questions suivantes :

- L'accès aisé au son modifie-t-il la manière d'envisager la transcription (degré de précision, notation des formes non standard, etc.) (question à laquelle nous avons répondu à la section 3.1) ?
- Quelle unité d'alignement choisir (cf. 4.2) ?
- Comment concevoir et organiser les transcriptions ou les annotations multiples d'un même enregistrement (cf. 4.3) ?

4.2. Quelle(s) unité(s) d'alignement choisir ?

L'utilisation de Praat permet de créer un fichier contenant des intervalles temporels qui segmentent le fichier son en « portions ». On insère dans chaque intervalle la transcription de

la portion de son correspondante. La structure du fichier d'alignement force le transcripteur à adopter une règle pour segmenter en intervalles temporels. Plusieurs types d'intervalles sont envisageables :

- des intervalles de taille identique (5 ou 10 secondes par exemple) avec pour conséquence que les paroles transcrites sont interrompues de manière aléatoire d'un intervalle au suivant ;
- des intervalles détectés automatiquement, par exemple en insérant une frontière à chaque pause silencieuse (choix opéré par Bertrand 1999) et en transcrivant des « unités séparées par des pauses », qui ne correspondent pas nécessairement à des unités linguistiques comme la proposition syntaxique ou l'énoncé ;
- des « tours de parole » (avec les problèmes de définition posés par cette unité émergente), lesquels peuvent s'avérer très longs dans des discours à structure monologique et / ou de style formel (conférence, émission radiophonique) ;
- des unités définies sur la base de critères prosodiques (dont la détection sera automatisée ou auditive) comme la période (Lacheret-Dujour et Victorri 2002) ou l'unité bornée par une frontière majeure (Mertens 1997, Simon 2004).

La liste pourrait encore être longue. Les deux premiers types n'ont a priori aucun statut linguistique ; les deux seconds ont un certain rôle à jouer dans l'organisation du discours ou de la conversation, sans pour autant qu'on puisse, lors de la transcription, pousser l'analyse suffisamment loin pour appliquer des critères de définition homogènes : on sait que le tour de parole (tel qu'il est défini dans la linguistique interactionnelle, cf. note 11) n'est pas simplement une unité bornée par des changements de locuteurs (la preuve, le statut de « pseudo-tour » de certaines productions verbales qui viennent « interrompre » un tour en cours de construction) ; on sait aussi que les unités intonatives, qu'on aurait parfois tendance à assimiler à des « actes discursifs » ou à des « énoncés », se définissent par une série de critères (l'intonation, mais aussi la présence ou l'absence de pause, les accélérations de débit, etc.) qui se combinent de manière subtile. Au mieux, on peut dire qu'elle définissent des « pseudo-énoncés ».

Pour résoudre la question de l'unité d'alignement, il faut tenir compte d'une part du fait que ce choix aura des implications sur des étapes ultérieures du traitement des données (ceci constitue un inconvénient), et d'autre part que les outils informatiques ne limitent pas nécessairement le chercheur à un seul alignement par fichier son : il est possible de superposer ou de combiner plusieurs types d'alignement en fonction des objectifs de la recherche (ceci constitue un avantage, cf. 4.3).

Dans notre banque de données, gérée par l'interface [moca]²⁷, l'unité d'alignement est l'unité intonative majeure (Bachy *et al.* 2006). Elle est à la fois une unité de consultation (lecture / écoute) du corpus, et une unité d'annotation.

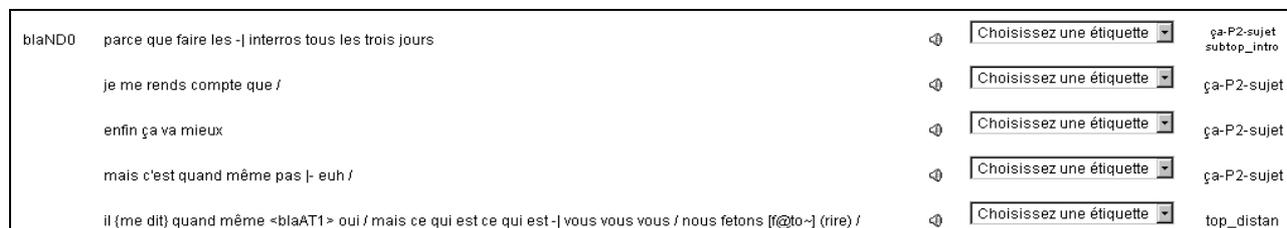


Figure 2: visualisation d'un enregistrement annoté dans l'interface [moca]

D'une part, l'utilisateur qui consulte un corpus aligné et déposé sur l'interface web verra ce corpus disposé ligne par ligne, chaque ligne correspondant à un intervalle dans le fichier de transcription alignée généré sous Praat (voir figures 1 et 2). Le tour de parole d'un locuteur est donc divisé en autant de lignes que d'unités intonatives majeures, et l'utilisateur peut écouter chaque ligne individuellement, ou cliquer sur une ligne pour l'écouter ainsi que les lignes suivantes, de manière continue.

D'autre part, l'utilisateur peut attribuer des étiquettes à des lignes de transcription²⁸ dans lesquelles il observe un phénomène linguistique intéressant qu'il voudra retrouver par la suite. Chaque étiquette (cf. figure 2) est attribuée à une ligne de transcription. L'unité d'alignement crée donc un artefact pour l'étiquetage qui sera fait ultérieurement sur les données²⁹. Si l'unité intonative majeure convient bien comme domaine temporel pour identifier des phénomènes

²⁷ [moca] est une interface en ligne qui permet tout à la fois d'archiver les données (primaires, secondaires, analyses), de les consulter et de les annoter. L'interface est spécifiquement conçue pour une exploitation sociolinguistique des données, via la formulation de requêtes exploitant la description des locuteurs et des situations d'enregistrement (cf. Kevers *et al.* 2006). L'interface [moca] permet une gestion des données cohérente avec la politique de VALIBEL de constituer « une banque de données ouverte, et non un méga-corpus clos : les corpus s'ajoutent au fil des collaborations, et ce indépendamment des thématiques et des exploitations. [...] Cette option, qui avait été préférée initialement à des techniques d'échantillonnage "représentatif" [...], s'est avérée pertinente dès lors qu'une masse critique de données a été réunie : non seulement elle permet des recherches multiples sur l'ensemble du matériau disponible, mais elle atteint l'essentiel des objectifs qui pourraient être assignés à un méga-corpus clos » (Francard *et al.* 2002 : 71-72).

²⁸ Sur le modèle de l'attribution des *keywords* qu'on utilise pour identifier et retrouver des extraits de vidéos dans les bases de données générées par les utilisateurs de Transana (cf. <http://www.transana.org/>).

²⁹ Noter cependant qu'*étiquetage* n'équivaut pas à *analyse*. Dans notre procédure, l'étiquetage permet d'identifier des phénomènes qu'on suppose intéressants, de les repérer dans de nombreux enregistrements, et de rassembler les occurrences étiquetées afin de les comparer et de les analyser. Il s'agit du premier balisage d'un corpus – lequel peut aussi se faire sur un corpus non transcrit (mais ce n'est pas l'objet de cet article). Lors du processus de récupération des lignes étiquetées, l'utilisateur peut toujours afficher les lignes autour de la ligne qui a reçu une étiquette afin de valider son étiquetage, ou de le modifier, le cas échéant.

micro-syntaxiques ou prosodiques, elle se révèle trop réduite pour mener une analyse de contenu (en étiquetant par exemple les types d'arguments utilisés par les locuteurs lors d'interviews sociolinguistiques³⁰), voire trop étendue pour repérer des phénomènes de phonétique articulatoire³¹.

En conclusion, l'unité intonative majeure est une unité d'alignement qui s'établit à un certain niveau de segmentation du discours. Elle facilite en général la consultation d'une transcription, car elle correspond grossièrement à des « actes de discours », à des « pseudo-énoncés » ou à des « unités de construction de tour », mais elle ne peut en aucun cas servir de base à l'analyse. En effet, identifiée sur une base auditive, elle n'est actuellement pas une unité linguistique objectivable qui garantit une segmentation en unités homogènes³².

Dans notre système de consultation des corpus en ligne, l'unité d'alignement sert aussi d'unité d'annotation, avec les inconvénients décrits. Ces inconvénients sont tempérés par le fait que, pour les corpus intégrés dans la base de données VALIBEL, chaque chercheur a la liberté d'opter pour une autre unité d'alignement, si elle convient mieux à l'objectif de sa recherche.

4.3. Combiner plusieurs transcriptions et annotations d'un enregistrement

L'inconvénient du choix d'une unité d'alignement est aussi partiellement compensé par la possibilité offerte, dans le format textgrid de Praat, de multiplier les couches d'annotation, et donc les unités d'alignement. En général, le mouvement va plutôt vers un alignement de plus en plus fin, dont la borne ultime est l'alignement sur une transcription phonétique, qui peut passer par un alignement en groupes accentuables, en mots ou en syllabes. Ces différents alignements, qu'ils soient strictement hiérarchisés (encapsulés) ou qu'ils se chevauchent partiellement, peuvent se combiner dans un même textgrid et être utilisés par le chercheur pour divers types d'analyses qui concernent le plus souvent la structure prosodique ou intono-syntaxique.

Le mouvement inverse peut être opéré aussi, en regroupant des unités de type « pseudo-énoncé » en unités plus grandes, comme le tour de parole.

³⁰ En effet, un argument se développe souvent sur plusieurs unités intonatives majeures, et donc sur plusieurs lignes de transcription qui doivent recevoir la même étiquette.

³¹ Une étiquette comme « assourdissement consonantique » attribuée à une ligne de transcription de quelques secondes de parole est relativement imprécise (plusieurs contextes sont potentiellement concernés).

³² Les résultats de travaux en cours, qui visent à établir une segmentation semi-automatisée sur la base de critères prosodiques et à partir d'un alignement syllabique, réduiront ces inconvénients.

Chaque niveau d'annotation peut ensuite être dupliqué si, à la transcription orthographique ou phonétique, le chercheur ajoute un codage spécifique (par exemple un codage alpha-numérique ou un étiquetage morphosyntaxique) utilisant un ensemble de symboles qui lui sont propres. De cette manière, on peut garder une transcription en orthographe standard qui sert de base à d'autres codages. Le protocole de codage des données mis au point par le projet *Phonologie du français contemporain*³³ (cf. Durand *et al.* 2002) adopte cette possibilité technique pour coder la réalisation des schwas et des liaisons de manière *parallèle* (non intégrée) à une transcription en orthographe standard. Chaque couche (de transcription, d'annotation) peut ensuite être interrogée et exploitée par des outils spécifiques (essentiellement des scripts).

5. Conclusion

La transcription de données orales est une démarche cruciale, de laquelle dépendent étroitement les résultats de la recherche. Transcrire, ce n'est pas simplement écouter et mettre sur le papier ce qu'on entend (Blanche-Benveniste et Jeanjean 1987, Blanche-Benveniste 2002). Depuis que la transcription est vue comme une démarche digne d'être théorisée (cf. l'article fondateur de Ochs 1979), le mythe du copiste est mort.

Le centre de recherche VALIBEL a réfléchi depuis sa création à la pratique de transcription, se dotant de conventions explicites, qui ne gomment pas le travail d'élaboration du discours non planifié. Nous avons montré ici comment ces conventions peuvent évoluer, sur des points mineurs, notamment en fonction des outils informatiques dont on dispose (concordanciers, logiciels d'aide à la transcription). Ces outils permettent des études sur les données secondaires et primaires que l'on n'envisageait pas auparavant, non seulement en termes de qualité mais également de taille des données. Nous avons enfin insisté sur les répercussions que peut avoir l'alignement du texte et du signal sonore sur les informations présentes dans les transcriptions et sur la nécessité de définir une unité d'alignement.

Références bibliographiques

- BACHY Sylviane, HAMBYE Philippe, SIMON Anne Catherine (2006). *Tutoriel Praat, Cahier 2, Transcription et alignement du texte sur le son*. Document non publié.
- BALLY Charles (1935). *Le Langage et la Vie*, Zurich, Max Niehans (2^e éd.).

³³ Les différentes étapes du protocole, depuis le recueil des données jusqu'à leur transcription et leur codage, sont expliquées sur le site du projet : <http://www.projet-pfc.net/>

BÉGUELIN Marie-José (dir.) (2000). *De la phrase aux énoncés : grammaire scolaire et descriptions linguistiques*, Bruxelles, De Boeck & Larcier.

BÉGUELIN Marie-José (2002). « Clause, période ou autre ? La phrase graphique et la question des niveaux d'analyse », *Verbum XXIV 1-2 (Y a-t-il une syntaxe au-delà de la phrase ?)*, M. Charolles, P. Le Goffic et M.-A. Morel Ed.), pp. 85-107.

BERTHOUD Anne-Claude, MONDADA Lorenza (Eds) (2000). *Modèles du discours en confrontation*, Berne, Lang

BERTRAND Roxane (1999). *De l'hétérogénéité de la parole. Analyse énonciative de phénomènes prosodiques et kinésiques dans l'interaction interindividuelle*. Thèse de doctorat, Université de Provence.

BLANCHE-BENVENISTE Claire, JEANJEAN Colette (1987). *Le français parlé. Transcription et édition*. Paris : Didier Érudition.

BLANCHE-BENVENISTE Claire (1997). *Approches de la langue parlée en français*. Paris, Ophrys.

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

BLANCHE-BENVENISTE Claire, ROUGET Christine, SABIO Frédéric (Éds.) (2002). *Choix de textes de français parlé. 36 extraits*. Paris, Honoré Champion.

BOERSMA Paul, WEENINK David (2007). Praat : doing phonetics by computer (Version 4.5.16) [Computer program]. Retrieved February 18, 2007, from <http://www.praat.org>.

CANDEA Maria (2000). *Contribution à l'étude des pauses silencieuses et des phénomènes dits « d'hésitation » en français oral spontané*, Thèse non publiée.

CHEEPEN Christine (1995). « Discourse considerations in transcription and analysis », G. Leech, G. Myers, J. Thomas (Éds.), *Spoken English on Computer. Transcription, Mark-up and Application*, New York, Longman, pp. 135-143.

DISTER Anne (2007). *De la transcription à l'étiquetage morphosyntaxique. Le cas de la banque de données textuelle orale VALIBEL*, Thèse de doctorat non publiée, Université de Louvain.

DISTER Anne, FRANCARD Michel, GERON Geneviève, GIROUL Vincent, HAMBYE Philippe, SIMON Anne Catherine, WILMET Régine (2006). *Conventions de transcription régissant les corpus de la banque de données VALIBEL* (<http://valibel.fltr.ucl.ac.be>, corpus oraux, conventions de transcription).

DUEZ (Danielle). 1997. « La signification des pauses dans la production et la perception de la parole », *Revue Parole*, 3-4, pp. 275-299.

DURAND Jacques, LAKS Bernard, LYCHE Chantal (2002). *Protocole, conventions et directions d'analyse*. Toulouse, CNRS ERSS-UMR5610 et Université de Toulouse-Le Mirail, *Bulletin PFC* n° 1, 74 p.

EDWARDS Jane A. (1995). « Principles and alternative systems in the transcription, coding an mark-up of spoken discourse », *Spoken English on computer. Transcription, mark-up and application* (G. Leech, G. Myers, T. Jenny Eds), New York, Longman, pp. 19-34.

EHMER Oliver (2006). « The Tranformer – a tool for working with aligned language data », communication présentée à la 3^e rencontre fribourgeoise de la linguistique sur corpus appliquée aux langues romanes, Freiburg-im-Breisgau, 14-17 septembre 2006.

ESPESSER Robert (2003). « L'enregistrement et la prise de son », *Bulletin de Phonologie du français contemporain* 2, 23-24 (http://www.projet-pfc.net/?u_s=0&u_a=34&)

FRANCARD Michel, PERONNET Louise (1989). « La transcription de corpus oraux dans une perspective comparative. la démarche du projet PLURAL », *Recherche en linguistique appliquée à l'informatique (RELAI)*, CIRB, Québec, pp. 295-307.

FRANCARD Michel, GERON Geneviève, WILMET, Régine (2002). « La banque de données VALIBEL : des ressources textuelles orales pour l'étude du français en Wallonie et à Bruxelles », *Romanistische Korpuslinguistik – Korpora und gesprochene Sprache / Romance Corpus Linguistics – Corpora and Spoken Language* (C. Pusch, W. Raible, Wolfgang Ed.) (= ScriptOralia; 126), Tübingen, Gunter Narr, pp. 71-80.

GADET Françoise (1992). *Le Français populaire*, Paris, Presses universitaires de France (coll. « Que sais-je ? »).

GADET Françoise (2003). *La variation sociale en français*, Paris, Ophrys.

GIOVANNONI Dominique-Catherine, SAVELLI Marie-Josée (1990). « Transcrire, traduire, orthographier le français parlé. De l'impossible copie à la falsification des données orales », *Recherches sur le français parlé* 10, pp. 19-37.

GUMPERZ John J., BERENZ Norine (1993). « Transcribing Conversational Exchanges », *Talking Data. Transcription and in Coding Discourse Research* (J.A. Edwards adnt M.D. Lampert Eds), Hillsdale, Lawrence Erlbaum Associates, pp. 91-120.

KEVERS Laurent, GILLES Peter, SIMON Anne Catherine (2006). « [moca], un système de gestion et d'annotation de données orales, communication », communication présentée à la 3^e rencontre fribourgeoise de la linguistique sur corpus appliquée aux langues romanes, Freiburg-im-Breisgau, 14-17 septembre 2006.

LACHERET-DUJOUR Anne, VICTORRI Bernard (2002). « La période intonative comme unité d'analyse pour l'étude du français parlé : modélisation prosodique et enjeux linguistiques », *Verbum* XXIV/1-2, pp. 55-72.

LAFOREST Marty (1992). *Le Back-channel en situation d'entrevue*, Québec, CIRAL (Recherches sociolinguistiques).

LOSIER Line, KASPARIAN Sylvia, CHEVALIER Gisèle, GAUVIN Karine (2002). *Guide de présentation de mémoires et de thèses en linguistique et conventions pour la transcription de conversations*, Université de Moncton.

MERTENS Piet (1997). « De la chaîne linéaire à la séquence de tons », *TAL (Traitement automatique des Langues)* 38/1, 27-51, Paris, Klincksieck.

MONDADA Lorenza (2000). « Les effets théoriques des pratiques de transcription », *LINX*, 42, revue de l'Université de Paris X-Nanterre, pp. 131-150.

OCHS Elinor (1979). « Transcription as theory », *Developmental pragmatics* (E. Ochs et B. B. Schieffelin Éds), New York, San Francisco, London, Academic Press, pp. 43-72.

PAUMIER Sébastien (2006). *Unitex 1.2. Manuel d'utilisation*, <http://www-igm.univ-mlv.fr/~unitex/manuel.html>

RAINGEARD Martine, LORSCHIEDER Ute (1977). « Édition d'un corpus de français parlé », *Recherches sur le français parlé* 1, pp. 14-29.

SIMON Anne Catherine. (2001). « Le rôle de la prosodie dans le repérage des unités textuelles minimales », *Cahiers de linguistique française* 23, pp. .

SIMON Anne Catherine (2004). *La structuration prosodique du discours en français. Une approche multidimensionnelle et expérientielle*, Berne, Peter Lang.

VERSTRAETE Estelle (2000). « Digressions dans les interactions verbales entre élèves travaillant en groupe », *Recherches* 33, pp. 151-179.