

Les incidences de quelques aspects de la transcription outillée

Paul Cappeau, Françoise Gadet, Emmanuelle Guerin, Roberto Paternostro

▶ **To cite this version:**

Paul Cappeau, Françoise Gadet, Emmanuelle Guerin, Roberto Paternostro. Les incidences de quelques aspects de la transcription outillée. Linx, 2011, Les genres de discours vus par la grammaire, n° 64-65, p. 85-100. 10.4000/linx.1403 . hal-01635064

HAL Id: hal-01635064
https://hal.science/hal-01635064

Submitted on 13 Dec 2017

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

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

Linx
Revue des linguistes de l'université Paris X Nanterre

64-65 | 2011
Les genres de discours vus par la grammaire

Les incidences de quelques aspects de la transcription outillée

Paul Cappeau, Françoise Gadet, Emmanuelle Guerin et Roberto Paternostro

revues.org

Éditeur
Presses universitaires de Paris Nanterre

Édition électronique
URL : http://linx.revues.org/1403
DOI : 10.4000/linx.1403
ISSN : 2118-9692

Édition imprimée
Date de publication : 1 juillet 2011
Pagination : 85-100
ISSN : 0246-8743

Référence électronique
Paul Cappeau, Françoise Gadet, Emmanuelle Guerin et Roberto Paternostro, « Les incidences de quelques aspects de la transcription outillée », *Linx* [En ligne], 64-65 | 2011, mis en ligne le 01 juillet 2014, consulté le 23 février 2017. URL : http://linx.revues.org/1403 ; DOI : 10.4000/linx.1403

Département de Sciences du langage, Université Paris Ouest

Les incidences de quelques aspects de la transcription outillée

Paul Cappeau, Françoise Gadet,
Emmanuelle Guerin, Roberto Paternostro[1]

> La problématisation du rapport aux informateurs, aux procédures de fabrication des données par les technologies de l'enquête, aux contextes [...] de recueil des données, qui ont marqué le tournant critique et réflexif de l'anthropologie, est remarquablement absente de l'historiographie et des réflexions contemporaines en linguistique.
>
> Mondada 1998, p. 41

De façon régulière, la linguistique connaît des mouvements d'engouement et de rejet (Moreau, 1985). L'actuelle omniprésence des corpus, si elle peut être regardée avec sympathie et intérêt, ouvre une phase d'unanimisme qui mérite probablement que soient soulevées des questions sur les pratiques retenues et les objectifs poursuivis. Cet article part de l'observation d'une surprenante discrétion dans les réflexions méthodologiques quant aux premiers moments de la collecte de données et à certains gestes effectués lors des phases préliminaires du traitement d'une masse importante de données (Cappeau et Gadet, 2007, 2010).

---

[1] Selon l'ordre alphabétique, les auteurs relèvent des institutions suivantes : Université de Poitiers et Forell ; Université Paris Ouest & MoDyCo ; Université d'Orléans et LLL ; Université de Brescia et de Paris Ouest-MoDyCo. Ce sont les enjeux des choix de transcription soulevés dans le traitement des données françaises du projet ANR-09-FRBR-037-01 (*Multicultural London English/Multicultural Paris French*), et les discussions auxquelles ils ont donné lieu qui ont conduit les auteurs à mettre à plat des discussions récurrentes entre eux.

On aurait pu espérer que la généralisation du travail sur la base de grands corpus conduise les linguistes à intensifier les discussions méthodologiques, à reconnaître l'incidence des choix lors de ces premières étapes, et du même coup à mieux peser l'impact des actes alors réalisés[2]. Mais ce n'est pas ce qui s'est passé, chacun défendant ses propres options comme les meilleures qui soient, faisant silence sur ces moments, ou reconduisant sans état d'âme ce qui a été fait par les prédécesseurs. En fin de compte, on débouche sur très peu de confrontations critiques. La forte susceptibilité de ceux qui ont le sentiment d'avoir effectué un gros travail (ce qui est incontestable) pour recueillir des corpus rend d'ailleurs risqué de soulever ce type d'interrogations, des questions légitimes passant facilement pour des attaques.

Nous ne cherchons ici qu'à participer à une prise en compte des paramètres sur lesquels le linguiste peut intervenir, sans traiter des phases ultérieures de l'exploitation. Ce que l'on cherche à montrer, c'est à quel point les premiers moments méthodologiques ne sont jamais que des enjeux méthodologiques, mais engagent en dernière instance des options théoriques sur la conception de la langue. Ce dont il n'y avait d'ailleurs pas à douter depuis au moins Ochs (1979) et son beau titre, provocateur voire apparemment contre-intuitif, « Transcription as theory ». Il s'agit seulement de ce premier moment on l'on « graphie » l'oral, dans lequel il n'y a pas de place pour le métalinguistique : prendre en compte l'oral en orientant minimalement sa lecture, même s'il est hors de doute que les étapes qui suivront orientent dans une certaine mesure les décisions.

1. Pas d'unanimité mais pas non plus de grands débats sur la transcription

La communauté des linguistes francisants réitère régulièrement le souhait de disposer d'un corpus de français parlé qui pourrait être dit « de référence[3] », par exemple réunissant/unifiant divers corpus. Mais pour les corpus existants, différents facteurs rendent complexe, coûteux voire inaccessible un tel objectif : la diversité des situations enregistrées, le disparate des objectifs affichés, les différences de

[2] Ces préoccupations sur les premiers moments du terrain s'avèrent difficilement audibles pour la plupart des linguistes, alors qu'elles font le quotidien des ethnologues (voir la citation en exergue). Il nous est ainsi apparu bien naïf de faire comme si pratiquer une activité (ici, transcrire) pouvait aplanir les réflexions méthodologiques qu'elle soulève et les choix qu'elle implique.

[3] L'expression mêle différentes valeurs, notamment sur la qualité (que serait censé posséder un tel corpus), sur la représentativité (que lui attribueraient les utilisateurs), sur son existence même (le fait qu'on puisse s'y référer). Certains voient l'idée de référence du côté de la langue (maximum de faits langagiers documentés), d'autres dans un échantillonnage social (proche de l'idée de représentativité, en général selon une sociolinguistique rudimentaire). Voir par exemple un exposé de Brigitte Garcia lors d'une séance de travail de l'IRCOM, le 22 octobre 2012 : « Corpus de référence/évolutif, données primaires/secondaires ». L'Institut de Linguistique Française a aussi organisé en mars 2013 deux journées de réflexion sur le thème « Initiative Corpus de référence du français », qui illustre à quel point cette ambition perdure.

conventions de transcription, et depuis peu l'hétérogène des formats d'encodage. Il en va autrement des corpus à venir, pour lesquels une certaine unification peut être envisagée[4].

1.1. Une dommageable division des tâches

L'interrogation d'un ouvrage récent (*Mais que font les linguistes ?*) présente, pour les aspects qui touchent aux interfaces linguistique/technique, une résonance forte. Ainsi, dans les premières étapes de la réflexion sur les étiquetages (avant 1990), a-t-on vu les informaticiens se saisir de la question. Le rouleau compresseur de l'efficacité a rendu peu audibles les questions que la reconduction d'étiquettes grammaticales – héritées de la tradition et maintes fois critiquées – auraient pu légitimement soulever.

La tentation est grande d'instaurer une semblable division du travail dans les étapes qui vont du recueil du corpus oral à son exploitation. Une coupure (que l'institution entérine tacitement par la faible valorisation de la rubrique « édition de corpus » dans les CV de chercheurs – malgré quelques progrès récents qui laissent entrevoir un futur différent) accorderait aux uns le recueil de données et/ou la transcription, et à d'autres l'annotation[5], le traitement, les analyses, et la publication des résultats. Soit une division où de petites mains font les tâches ingrates, ce qui pourrait couper l'analyste d'une certaine qualité des données constituées. Des chercheurs avec une longue expérience de la transcription, comme Blanche-Benveniste qui revient encore, dans l'une de ses toutes dernières publications, sur un point qui pourrait apparaître comme un détail (2010 : il s'agit du traitement de *il* dans *il y a*) ont insisté sur le fait que la phase de transcription ne peut sans dommage être déléguée à des auxiliaires peu formés (et non partie prenante du projet théorique) et que des corrections réitérées sont indispensables, même si l'on sait bien qu'il s'agit d'un processus sans fin.

Parmi les gestes préalables à l'exploitation de corpus oraux (choix du terrain, sélection des informateurs, mise en place du recueil, numérisation, anonymisation…), la transcription occupe une place de choix, aux enjeux sans doute encore souvent sous-estimés. On en veut pour preuve le fait qu'il n'en est à peu près jamais parlé dans les articles, qui tendent à aller tout de suite vers « l'essentiel », l'étiquetage, l'analyse, les résultats[6]. L'exception qu'est l'ouvrage de Blanche-Benveniste et Jeanjean, 1986

[4] Celle-ci passe en particulier par la connaissance éclairée et la diffusion de choix communs aux diverses étapes (conventions, transcription), ce à quoi œuvre l'IRCOM Corpus Oraux, qui met notamment un glossaire de notions à disposition de la communauté.

[5] Force est de constater que « transcription » peut être ambigu par rapport à d'autres termes désignant des actions proches : « annotation », « codage », « étiquetage ». Si l'étiquetage réfère plutôt à l'attribution de catégories morpho-syntaxiques et le codage à celle d'un code alpha-numérique à un phénomène faisant l'objet d'analyses automatiques (comme le e muet ou la liaison), la limite entre transcription et annotation s'avère souvent difficile à poser. Voir Delais-Roussarie & Yoon (2011), qui parlent de transcription prosodique pour des phénomènes où d'autres parleraient d'annotation.

[6] La question de la transcription n'est le plus souvent évoquée que sous l'angle, bien compréhensible dans la situation actuelle, de son coût (une bonne part des budgets de projets de corpus est engloutie

constitue un véritable OVNI – d'ailleurs rarement cité. Pas davantage cité (et sans doute pas lu, car rédigé en allemand), l'ouvrage de Dittmar, 2002, dont les 240 pages ne s'intéressent qu'aux techniques pour la « Verdauerung flüchtiger mündlicher Rede » (p. 15, « faire perdurer la parole éphémère »).

Même sur les sites des grands projets, les indications sur ce que des chercheurs en sciences sociales nomment plaisamment « arrière-cuisine de la recherche » ne sont pas toujours clairement exposées, ou du moins elles passent sous silence les échecs ou les impasses ; contrairement aux conventions de transcription, qui, elles, donnent toujours lieu à un très précis tableau des conventions, en bonne et due forme.

1.2. Les grands projets de corpus et les logiciels d'aide à la transcription[7]

L'intérêt de recourir à un logiciel d'aide à la transcription fait à peu près l'unanimité des grands projets, qui font tous connaître leur choix – voir d'ailleurs, sur le site d'ICAR[8], un tableau comparatif (mais plus technique que méthodologique). Ainsi, en ne parlant que de corpus du français[9], *PFC*, *VaLiBel*, *CIEL_F*, *MPF* ou *OFROM* ont adopté *Praat* ; *TCOF*, *CFPP2000*, le *CRFP* ou *ESLO-2*, *Transcriber* ; *CLAPI*, à la fois *Praat* et *CLAN* ; et *FRAN*, *CLAN*. Quant au *CFPQ*, il a recours à *Transana* (qui, comme *CLAN*, « permet un alignement du son, de l'image et du texte », site du *CFPQ*)[10].

Les discussions qui ont pu prendre place pour retenir tel logiciel et les comparatifs entre solutions en restent semble-t-il à l'état de débats internes aux équipes, et nous n'en connaissons guère d'expression publique (sauf de points de vue techniques). Pourtant, cette étape engage des enjeux, en amont sur la représentation de l'oral, et en aval quant au produit écrit (la façon dont l'oral sera donné à lire) qui constituera le support le plus fréquent de l'analyse. Ce choix relève précisément de l'étape préliminaire qui nous concerne ici, et ne peut rester sans incidence sur les étapes de traitement des données à suivre.

Un point décisif concerne la future exploitation du corpus, selon qu'il y a ou non un objectif disciplinaire orientant d'emblée vers certaines options : la phonologie pour PFC ou l'analyse de conversation pour CLAPI ou le CFPQ, dont les visées

---

dans la transcription). Mais dès 1979, un article de Houdebine soulevait des questions autour de ce qu'elle appelle la « notation » de l'oral, en insistant sur l'incontournable nécessité d'écoutes réitérées.

[7] *Praat* n'a pas été conçu à l'origine pour la transcription, mais comme un logiciel d'analyse acoustique. Nous nous limitons ici à discuter de ses caractéristiques en tant qu'outil d'aide à la transcription.

[8] http://icar.univ-lyon2.fr/projets/corinte/confection/comparatif_logiciels.htm

[9] Le choix d'un logiciel est soumis à de nombreux facteurs, dont certains purement pratiques, comme la facilité de prise en main par des non-experts. Malgré la coutume, nous ne donnons pas, faute de place, les références des logiciels (trouvables sur le web).

[10] Le choix de *CLAN* est plus fréquent chez les psycholinguistes ou les acquisitionnistes (en particulier spécialistes de la langue de l'enfant). Il est enfin à noter la rareté du choix de *Winpitch*, à notre connaissance retenu seulement par *C-Oral-Rom* et le *Corpus di italiano parlato*. Comme on s'intéresse ici à la mise à disposition sur le web, il apparaît difficile de faire état d'autres corpus canadiens que le *CFPQ* et le *FRAN* (en cours de constitution), les grands corpus historiques étant trop anciens pour avoir fait appel à des logiciels d'aide à la transcription. Voir Vincent, 2009.

conversationnalistes imposent de pouvoir mesurer les pauses, ce que par exemple *Transcriber* ne pourrait pas faire aussi aisément que d'autres logiciels. En revanche, *Transcriber* semble parfaitement convenir lorsque le recueil n'est pas, en amont, orienté vers des objectifs disciplinaires précis. Pour ESLO2 par exemple, le choix de *Transcriber* est motivé par le format de sortie : la simplicité du logiciel limite les pré-analyses au moment de la transcription, ce qui a conduit les chercheurs à parler de transcriptions de « degré zéro » (Baude et Abouda, 2007), ou transcriptions minimalement contraintes par les visées de traitement. Mais un tel choix impose que ces données ne soient pas considérées comme « prêtes à l'emploi », une analyse fine nécessitant le recours à d'autres logiciels. L'intérêt de *Transcriber* réside notamment dans la facilité de conversion, le format de sortie des fichiers (XML) étant une garantie d'interopérabilité. Cette conception oriente vers l'idée d'état transitoire. ESLO 2 défend ainsi, comme OFROM, une conception contributive des transcriptions : devant la masse de données recueillies, on comptera sur les utilisateurs pour améliorer des transcriptions insuffisamment vérifiées. Les résultats permettront de valider (ou non) cette démarche.

1.3. Avantages/inconvénients de logiciels d'aide à la transcription

Deux arguments principaux sont avancés en faveur de l'usage de logiciels d'aide à la transcription. L'un concerne la facilité de manipulation du signal sonore : on peut découper le continuum sonore en séquences courtes qui peuvent être réécoutées à loisir. Cet aspect pratique a des effets positifs sur la qualité de la transcription, en outre facilitée par la visualisation du signal. Un deuxième argument, plus technique, concerne l'alignement son/texte, qui pourrait si on le souhaite alléger les conventions de transcription : par exemple, pourquoi noter graphiquement une proéminence[11] ou un contour intonatif que l'on peut entendre et voir en déroulant les deux objets parallèlement ? Une qualité souvent reconnue à *Praat* concerne la création de tires, qui indiquent/codent un même type d'informations (différencier des locuteurs, noter des évènements extralinguistiques, introduire des annotations phonétiques…). Davantage de reconnaissance semble alors accordée à la complexité de la production orale, que la mise en mots (ou en sons transcrits) réduite à la seule prise en compte de la dimension verbale. Ainsi, *Praat* permet, par les tires, une visualisation simultanée de codages de plans différents.

[11] Nous visons ici le recours à des majuscules pour transcrire une proéminence syllabique non finale (« accent secondaire » ou « d'insistance »), résultant d'une augmentation d'énergie articulatoire. Exemple : c'est SUper ! (voir Léon, 2007 : ch. 9). Certaines conventions de transcription recommandent l'annotation de contours intonatifs à l'aide de barres: montants /, descendants \ ou montant-descendants /\. Cette façon d'annoter des phénomènes supra-segmentaux ne se fonde pas sur des critères acoustiques, mais sur la perception du/des transcripteurs(s). Outre engendrer des ambiguïtés entre transcription et annotation (voir note 5), cette pratique risque de sur-interpréter ce qui exigerait une analyse plus fine.

Un balisage précis des pauses est aussi facilité, notamment par *Praat*, et c'est pourquoi il est souvent privilégié dans les projets à orientation écologique. *Transcriber* présente sur ce point plusieurs limites (attribution arbitraire de la pause à l'un des protagonistes, difficulté pour gérer les chevauchements[12]…).

2. Ce que donnent à voir les logiciels d'aide à la transcription

Les logiciels d'aide à la transcription fournissent une représentation de l'oral qui entre dans un format prédéfini. Ils ne peuvent pas se présenter comme des outils neutres, sans incidence.

2.1. Découpage et représentations

Les logiciels fournissent une visualisation graphique du signal, manipulable par le linguiste. Leur interface reflète la priorité que chacun accorde aux composantes visibles à l'écran : *Praat* place le signal en haut de la fenêtre, soit une place primordiale ; dans *Transcriber*, c'est la transcription qui est visuellement mise en avant, le signal sonore, relégué en bas de l'écran, semble plus secondaire.

D'autres différences se manifestent concernant les variables du signal. *Praat*, comme *Winpitch*, présente un spectrogramme ainsi que des détails acoustiques (courbe intonative, intensité…) qui peuvent s'avérer utiles pour certains types d'analyses, notamment le codage des contours intonatifs ou des syllabes proéminentes. *Transcriber* se limite à fournir une visualisation du signal en conformité avec l'objectif d'être, du fait de l'absence de sophistication, un outil de transcription simple et rapide à prendre en mains.

Le découpage en tires dans *Praat* apporte une grande malléabilité, puisque leur nombre est laissé à l'utilisateur. Cette liberté entraine une diversité (disparité ?) des transcriptions sous *Praat*, chacun pouvant ajouter une couche selon ses intérêts. Mais plus que sur ce point (qui touche un débat récurrent en linguistique - unification *vs* diversification), il semble opportun de s'arrêter sur les changements que ce dispositif technique provoque quant à la visualisation de la transcription.

Dans les transcriptions sous *Transcriber*, le texte est généralement découpé en tenant compte des pauses, et chaque paragraphe correspond à un découpage du signal sonore. La transcription se présente sous une forme conforme à nos habitudes de lecture des textes écrits :

[12] Pour noter les chevauchements, on peut recourir à un codage, mais on ne peut pas visualiser où commencent et où finissent les superpositions. Pourtant, le repérage des éléments qui se chevauchent est souvent essentiel pour comprendre la co-construction du discours. Là encore, il en va de la fidélité de la représentation graphique d'un événement phonique.

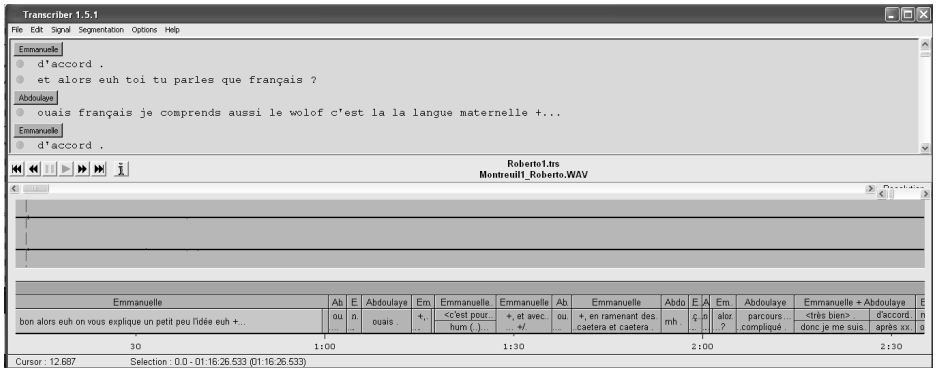


Fig. 1. Extrait de transcription sous *Transcriber*

Dans une transcription sous *Praat*, une tire attribuée à chaque intervenant conduit à une disposition qui fait ressortir la structure de l'échange. En revanche, dans la fenêtre, seuls de courts fragments du texte sont lisibles. L'intégralité du texte transcrit correspondant à chaque séquence alignée sur le signal est accessible, une fois ladite séquence sélectionnée, dans un espace à part, indépendant de l'alignement sur le signal. Cette organisation est souvent perçue comme une difficulté lors de la prise en main du logiciel, car elle oblige à jongler entre deux modes de lecture, l'un, inhabituel, qui permet de considérer dans un même temps l'oral et l'écrit selon une disposition pluridimensionnelle dans l'espace, l'autre, traditionnel, qui présente le texte linéairement. Traiter des données sous *Praat* impose de circuler dans cette double lecture :

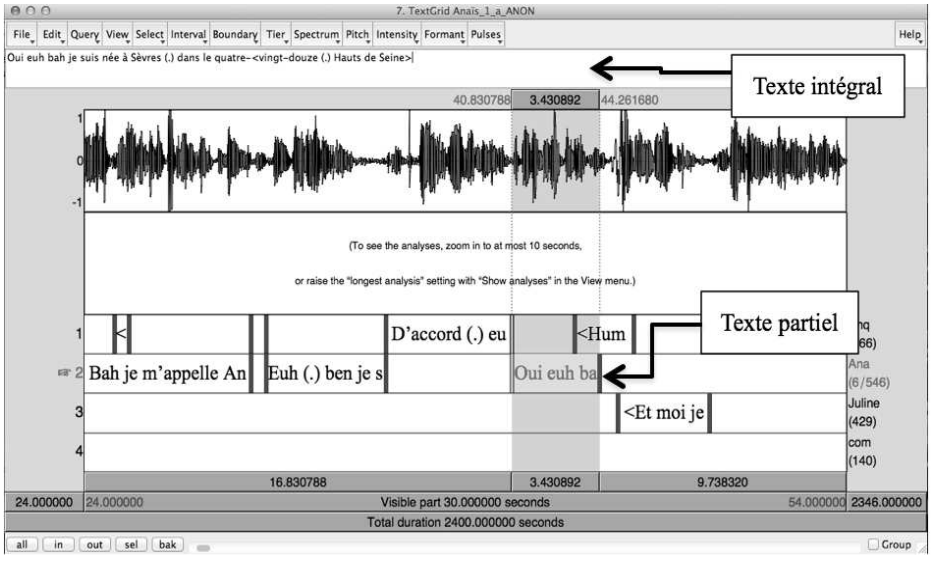


Fig. 2. Extrait de transcription sous *Praat*

Le caractère enchevêtré et pluridimensionnel de la conversation est à la fois montré visuellement et organisé, par l'empilement des tires où chaque intervenant se voit attribué son texte[13]. Le mode de transcription ainsi engagé est donc très éloigné de la représentation classique. *Praat* rompt avec la représentation habituelle de l'écrit linéaire, de deux manières : par la dimension horizontal/vertical en intégrant la superposition des tires, qui apportent un fractionnement des phénomènes et tentent de gérer la simultanéité (mais est-ce un objectif du logiciel, ou une conséquence de la représentation en tires ?). D'autre part, par le focus large/étroit que permet la fonction 'zoom'. A ce mouvement correspond une prise en compte plus ou moins détaillée des phénomènes : un focus large correspond à une transcription strictement verbale, un focus étroit construit un mille-feuille qui montre la simultanéité des différentes composantes.

Les enjeux conceptuels débordent du même coup le phonique, avec des questions concernant la prosodie, dont le syntacticien a forcément à connaître, en particulier l'intonation, qui n'est pas toujours repérable à l'oreille. Ainsi, *Praat* permet, à certaines conditions (sélectionner un segment inférieur ou égal à 10ms et activer la visualisation du pitch), de visualiser la courbe intonative : on « voit » si un contour est montant ou descendant, ce qui contraste avec le seul codage à l'oreille, qui laisse passer beaucoup d'erreurs perceptives voire fait courir le risque de sur-interprétation. Il en va de même pour tous les codages imposés par un logiciel, comme dans *CLAN* où des codages spécifiques alourdissent la lecture du texte et risquent de générer une sur-interprétation.

Au contraire, des choix de transcription les plus « légers »[14] possibles, et l'absence de contraintes sur les codages, permettent de garder ouverte toute perspective d'analyse de phénomènes émergents. Nous poursuivrons cette ligne de réflexion dans la troisième partie.

2.2. Les enjeux et les incidences

Outre les aspects techniques, non négligeables en ce qu'ils imposent des prises de décision aux utilisateurs, il y a lieu de s'interroger sur les représentations de la langue que véhiculent ces outils, et sur leur incidence dans le travail qu'ils ont pour but de faciliter. En effet, ces logiciels imposent des méthodes de travail et une conception de l'oral qui méritent d'être analysées de plus près.

Le son et la transcription sont considérés comme deux entités distinctes, qui le resteront puisqu'elles sont dans deux fichiers physiquement distincts, même si lisibles de conserve. Ces outils reposent sur une propriété, qui peut d'ailleurs être vue comme un avantage ou un inconvénient : mettre en jeu des exigences déclaratives initiales, comme préciser les paramètres qui doivent être dissociés et se verront chacun attribuer une tire. Il est donc imposé de projeter avant toute démarche d'analyse un « format » de découpage des données. Parmi les effets induits : cette représentation matérielle

[13] Cela laisse peu de place à l'incertitude et laisserait entendre que tout serait identifiable et attribuable. Quiconque a fait de la transcription d'oral non monologal sait qu'il n'en est rien.

[14] Chacun croit voir la neutralité et l'absence de lourdeur dans sa propre pratique.

impose de concevoir, *volens nolens*, la chaine discursive comme un empilement de couches d'informations, dissociant les plans (verbal, suprasegmental, segmental…) – ce qui de fait correspond à une certaine doxa sur le langage.

Donner à voir, c'est aussi forcer à voir, donc imposer une vision du monde linguistique et langagier. On ne peut qu'être frappé par ce type de représentation, fréquente dans les descriptions linguistiques : ainsi la variation a en des temps anciens (révolus ?), été conçue comme une « couche » qui s'ajoutait au standard et venait le déranger. De même, le style reste parfois envisagé comme « en plus » d'un niveau neutre, apportant une coloration à une production.

De fait, rien de plus récurrent qu'une conception en couches empilées : ce modèle est accepté comme une représentation satisfaisante des faits langagiers, bien que très vite soient apparues les limites de cette représentation. En ira-t-il de même quand une « garantie » technologique vient en renfort d'une telle conception de la langue ? Ne peut-on craindre que tant de corpus décrits selon un tel schéma, ne finissent par figer les conceptions ? Des représentations alternatives auraient offert une tout autre image : on peut penser à la forme d'hypertexte (popularisée par les pages web), ou la représentation sous forme de soleil de Rivenc (1971) pour la structure du lexique effectif dans la langue parlée. Loin de superposer des strates étanches, sa représentation (empruntée aux astronomes) montre les imbrications ou les intersections entre les zones identifiées. Bref, d'autres représentations d'une structure complexe sont envisageables.

Mais qu'ont donc fait les linguistes ? On pourra (re)poser cette question dans quelques années… et il apparaîtra que beaucoup ont été des utilisateurs (plus ou moins éclairés) d'outils, qu'ils se sont pliés à des contraintes externes, mais qu'ils ont peut-être trop peu interrogé les représentations que ces modèles induisaient. Nous y reviendrons en conclusion.

3. Quelques points d'enjeux de transcription

Le corpus *MPF* nous servira à illustrer nos remarques. C'est un corpus d'énoncés éloignés du standard, peu formatés par l'alternance question-réponse de l'interview (encore moins pour la partie écologique[15]). Ses principaux enjeux sont en lien avec l'absence de prévisibilité de phénomènes discursifs ou syntaxiques qui seront dits « émergents ». Au delà de la diversité des préoccupations des membres de l'équipe (à quoi s'ajoute l'horizon de mise à la disposition de la communauté scientifique), nous nous efforçons de rendre visible la réflexion autour des choix méthodologiques de la transcription des données, qui demeurent souvent internes aux équipes alors que, dans une perspective de mutualisation des corpus, il faudrait sinon s'accorder du moins dialoguer aussi à ce stade de prises de décisions.

---

[15] *MPF* désigne ici la partie française du projet ANR évoqué en note 1. Les présents commentaires concernent cette partie française, dont le corpus comporte en janvier 2013 environ 35 heures transcrites et révisées (420.000 mots). Pour le design de *MPF*, voir Gadet et Guerin, 2012, et Gadet, 2013. Des objectifs écologiques sont aussi à l'œuvre dans le grand projet *CIEL_F*, pour l'ensemble de la francophonie (voir Gadet *et al.*, 2012). Il va de soi que ce choix ne constitue en rien un modèle de ce qu'il faudrait faire, mais souligne quelques-uns des problèmes rencontrés.

Les questions soulevées par la transcription du corpus seront ici reliées aux outils techniques, afin de faire émerger des problématiques nouvelles.

3.1. Coder le discours rapporté ?

Plusieurs projets de grands corpus codent le discours rapporté dès la phase de transcription (par exemple, FRAN, CFPP2000, MLE), ce qui ne les empêche pas de reconnaître les difficultés, notamment sur les indices de fin de segment rapporté. Ce qui signifie qu'il y a aussi des arguments pour ne pas le coder. Quelle incidence la transcription outillée apporte-t-elle dans l'identification des paroles rapportées ? Permet-elle un changement de perspective ?

Les problèmes vont au-delà du repérage des frontières du discours rapporté. Certains exemples semblent soit échapper au codage proposé, soit se voient assigner une valeur catégorique (DR ou non) qui ne rend pas justice à la complexité des faits :

(1) mais après pour dire que ouais j'ai le seum pour dire ouais je suis éner- je suis énervé là dernièrement euh.

(2) par exemple moi c'est une expression que moi j'utilise pas que j- que je connais hein lorsque je les je les entends euh lorsque je les écoute le dire euh ouais j'ai le seum je comprends ce que ça veut dire (MPF : Paris).

En (1) et en (2), l'enquêteur a demandé à Farid le sens de *seum*, à quoi Farid répond que c'est « poison » en arabe, mais que les jeunes l'emploient au sens de « énervé ». En présentant le discours rapporté à l'oral, il est habituel de le dire introduit par un verbe à la première ou troisième personne du singulier du présent ou de l'imparfait, et suivi d'une particule introductive (*ouais*, *mais*, *non*, *ben*…) prononcée sur une plage sonore plus haute ou plus basse que le discours porteur et le segment rapporté (parmi d'autres, Danon-Boileau et Morel, 1998 : 131-133). Or, Farid prononce effectivement le *ouais* sur un ton plus élevé, mais le verbe introducteur en (2) est à l'infinitif. L'énoncé est ainsi ambigu, entre un simple exemple et un cas de discours rapporté non recensé par les descriptions traditionnelles. Un codage trop rapide aurait ici eu pour effet de refermer l'interprétation, et éventuellement d'empêcher de relever une forme inédite.

Par ailleurs, un codage d'emblée pose la particule introductive comme constitutive de l'appareillage signalant un discours rapporté. Or, une étude plus fine interroge le statut de ces particules, entre discours cité et discours citant. La thèse en cours d'Anaïs Moreno montre l'intérêt d'une position intermédiaire entre les deux cadres énonciatifs. Alors, comment systématiser l'annotation ? On voit l'intérêt d'une lecture de la transcription sans orientation préalable, qui repose sur la disposition simultanée d'informations relevant de différents niveaux d'analyse : en l'occurrence ici, phonétique et syntaxique.

3.2. L'analyse d'une particule

La présence de termes ou expressions en langue étrangère dans des énoncés en français est fréquente dans MPF. Dans la plupart des cas, il ne s'agit pas de code-switching, mais d'unités intégrées dans le discours sans référence au concept évoqué

dans la langue d'origine. Aussi ne suffit-il pas de référer à l'autre langue pour interpréter un énoncé. De plus, souvent, les unités prennent une valeur syntaxique et/ou discursive autonome. Dès lors se pose la question du codage, de la mise à disposition d'une transcription analysable. Le cas de *wesh* est intéressant car si on est en mesure d'en fournir une traduction (l'interrogatif *quoi* en arabe populaire), les exemples suivant montrent que ses emplois imposent de poursuivre l'étude :

(3) non c'est Kosona il a mon âge lui wesh (MPF : Mantes la Jolie)

(4) la plupart du temps tu dis ouais tu viens ou quoi wesh (MPF : Mantes la Jolie)

En (3), la valeur interrogative semble avoir disparu. En (4), une traduction stricte serait *a priori* redondante. Pour fournir au syntacticien les moyens d'analyser ces énoncés, il faut plus que la seule traduction, qui en l'occurrence n'est pas pertinente. Une transcription linéaire impose que ces informations figurent dans le texte même (et non dans une tire spécifique), alourdissant de fait la lecture. En revanche, une représentation permettant une lecture à la fois horizontale et verticale offre un tout autre rendu[16]. Créer une tire de *Praat* dédiée à ce type de renseignements est apparu comme une solution. On peut aligner la transcription sur le signal et noter en dessous des commentaires sur le segment concerné. La distinction entre transcription et commentaire facilite l'exploitation des données. En contrepartie, cela impose d'appréhender autrement la lecture des données, en acceptant la rupture de la linéarité pour entrer dans une lecture (au moins) bidimensionnelle.

3.3. Le cas des chevauchements

Certains phénomènes d'interaction peuvent être représentés, donc pris en compte de manière plus efficace dans le mode de représentation retenu : c'est le cas des chevauchements. Le discours spontané en interaction étant une co-construction, les fréquents chevauchements ne peuvent pas être traités comme de simples tours de parole, puisqu'il s'agit souvent d'une collaboration dans l'élaboration d'unités plus petites que le discours (l'énoncé, voire le syntagme). Ainsi, certains énoncés ne sont ni analysables ni interprétables sans tenir compte des produits d'au moins deux locuteurs (ce que Jeanneret, 1999 a nommé « co-énonciation », et que Blanche-Benveniste a théorisé sous la forme du « locuteur collectif » – voir entre autres 1990). La division en tires offre une visualisation, puisque les éléments de l'énoncé sont présentés à la suite bien qu'ils ne soient pas sur le même axe horizontal :

[16] Passons sur le caractère paradoxalement intenable de cette position, qui devrait conduire à ne fournir d'exemples que sous la forme de copie d'écrans du logiciel, puisque la représentation écrite oblige à respecter la linéarité de l'écrit.

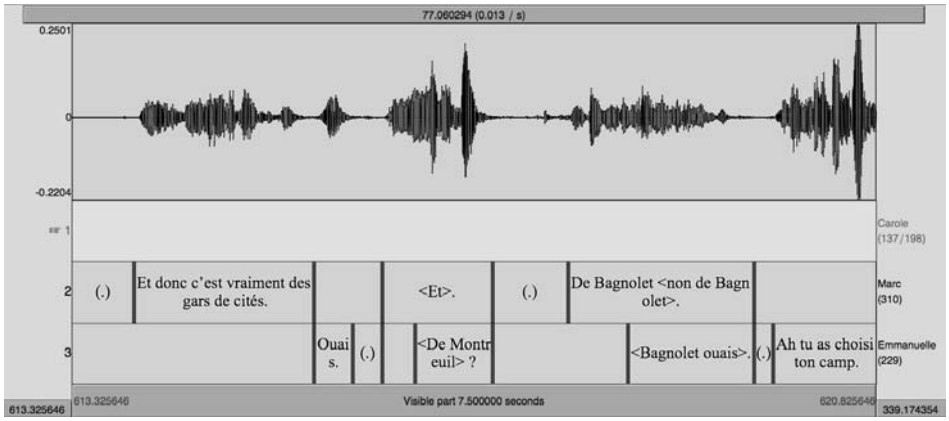


Fig. 3. Exemple de co-énonciation (extrait MPF : Montreuil)

L'accès simultané aux propos des deux intervenants rend visible le mode de production, qui donne lieu à un énoncé cohérent : *C'est vraiment des gars de cités (pas) de Montreuil (mais) de Bagnolet ouais*. La lecture à la fois horizontale et verticale aide à comprendre pourquoi la séquence amorcée par *et* n'aboutit pas. De même, dans l'exemple de la Fig. 4, on accède facilement à l'interprétation de *J'ai (.) voilà*[17]. Ces visualisations, en éloignant l'idée d'énoncés mal formés ou mal organisés, offrent une représentation bien plus structurée de l'oral.

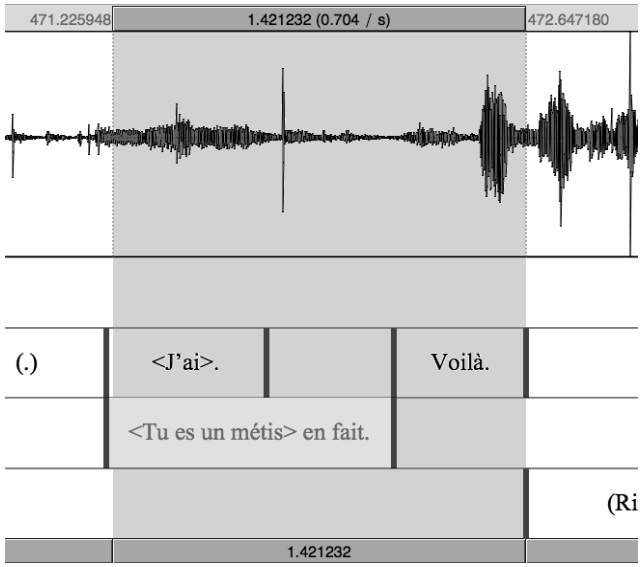


Fig. 4. Exemple de chevauchement (extrait MPF : Noisy-le-Sec)

[17] On peut être surpris, dans la figure 4, que le chevauchement mette en cause d'un côté un segment très court et de l'autre un segment nettement plus long. Pourtant, tel est bien le cas, et il faut tenir compte des effets de nos conventions de transcriptions, minimales : le *j'ai* est très allongé, quand au contraire *tu es* est prononcé [te].

Certes, il est toujours possible d'étudier un énoncé co-construit sans logiciel de transcription, mais la visualisation facilite une lecture à la fois horizontale et verticale.

3.4. La transcription comme entre-deux

Le choix de *Praat* pour transcrire les données de MPF a donc été largement motivé par la volonté de limiter les préconstruits en amont, renvoyant l'analyse au stade ultérieur d'exploitation des données. Néanmoins, il serait illusoire de penser que les transcriptions ainsi présentées soient vierges d'interprétation du transcripteur, dont le rôle se réduirait au geste mécanique de poser par écrit : il est besoin d'une certaine expertise. Dans MPF, la plupart des transcriptions sont assurées par les enquêteurs, à qui la connaissance de la situation permet de contextualiser certains phénomènes. Les enquêteurs/transcripteurs ont tous une culture (socio)linguistique, limitant ainsi les interprétations ou codages trop naïfs. Ces précautions sont prises parce que les choix méthodologiques, l'usage de *Praat*, les décisions sur les conventions, font de cette étape du traitement des données un incontournable entre-deux, entre livraison brute et analyse proprement dite.

Si les chevauchements peuvent facilement être représentés sans que le transcripteur ait à interpréter, tout n'est pas également représentable, et d'autres phénomènes ne peuvent être ni représentés ni interprétés dès cette étape. Tel est par exemple le cas des reprises – voir (5), phénomène fréquent qui peut avoir autant d'impact dans la co-construction du discours que le chevauchement : la présentation en couches de *Praat* rate, en partie, la mise en évidence du phénomène qui est décalé dans le temps, et le signalement du chevauchement ne peut se faire sans une décision du transcripteur :

(5) L1. Donc c'est le mercredi après-midi principalement ou vous avez d'autres jours aussi ?

L3. C'est le mercredi après-midi principalement. (MPF : Marly-le-Roy)

Contrairement au cas du chevauchement, *Praat* (pas plus qu'un autre logiciel) ne permet pas une visualisation instantanée de la reprise (est-ce même possible ?). Se trouve ainsi posée la question de la limite de cet entre-deux, qui est illustrée par la différence entre le chevauchement, phénomène trans-couches qui concerne les deux locuteurs, et les reprises, pour lesquelles on perd le processus.

On peut certes considérer que c'est peu important, et on peut aussi ne pas voir d'inconvénient à renvoyer cela au stade de l'analyse.

Conclusion

On ne s'est pas arrêtés ici à ce que les transcriptions mettent en jeu quant à la relation entre oral et écrit (Cappeau et Gadet, 2013), car notre objectif était de montrer à quel point les logiciels, et plus largement l'outillage technologique, font courir au linguiste le risque de perdre le contrôle de l'objet « conception de la langue », qui tend ainsi à se trouver formaté par des exigences techniques, loin que ce soient ces

dernières qui s'adaptent à une vision de la langue et de la linguistique[18]. Cela relève sans doute du même type d'illusion que celle selon laquelle pourrait exister une transcription basique sur laquelle tout le monde s'accorderait, chacun pouvant ensuite ajouter une strate pour ses besoins d'analyse (voir à ce sujet les pistes de convergence avancées par différentes équipes dans Bilger (2008)).

Les transcriptions outillées offrent une façon différente de voir les choses, qui a des effets dans nos pratiques, mais il serait naïf de croire qu'il s'agit d'une pure et simple avancée technologique, neutre, sans incidence sur nos représentations et notre vision de l'objet langue. C'est cette vision de la langue qui mérite d'être questionnée. Il reste alors à se demander si le fait de consolider le rôle des métadonnées, impliquant que le tout de l'interprétation n'est pas dans les seules séquences langagières, permet de rééquilibrer cette conception imposée de la langue.

Certains linguistes imaginent pouvoir demeurer maîtres d'œuvre dans leur domaine et même renforcer leur position, en déléguant certaines tâches aux informaticiens[19]. Tout l'enjeu de cette collaboration, à la fois souhaitable et inévitable, est de déterminer précisément l'effet-retour que la vision technologique aura : l'un des risques réside dans un formatage appuyé sur un souci d'efficacité et de rapidité. C'est déjà au nom de ces principes qu'il a parfois été proposé de nettoyer la production orale de tout ce qui l'alourdit (bribes, hésitations,…). Finalement, c'est rien moins que la persistance d'un certain type de réflexion linguistique qui est en jeu, accordant une place à ce que l'on dira, pour faire vite, la linguistique de terrain. Même si on lui attribue en l'occurrence pour objectif, qui n'est certainement pour elle ni prioritaire ni exclusif, une intervention dans la collecte de grands corpus, il ne nous apparaît pas que la linguistique de corpus ait à gagner à faire l'impasse sur les préoccupations liées au terrain. En somme, il s'agit de considérer chaque étape de la constitution de corpus pour ce qu'elle est, avec ses objectifs propres, l'expertise et les moyens nécessaires à chacune, sans négliger un dialogue équitable entre elles, ce qui implique que les intérêts des uns n'orientent pas le travail des autres.

[18] Est-ce vraiment nouveau ? A des périodes antérieures, il y a déjà eu de semblables débordements par la technique. L'article de Mondada, 1998 laisse même entendre qu'il ne saurait en aller autrement, ce qui n'est sûrement pas une raison pour ne pas réfléchir sur la portée des gestes.

[19] La double compétence de linguiste et d'informaticien de nombreux jeunes chercheurs n'interfère pas fondamentalement sur les questions posées ici, qui relèvent de l'indispensable réflexion que les gestes des linguistes doivent provoquer.

BIBLIOGRAPHIE

ABOUDA, L., BAUDE, O., 2007, « Constituer et exploiter un grand corpus oral : choix et enjeux théoriques. Le cas des ESLO », *in Actes du colloque Corpus en lettres et sciences sociales, des documents numériques à l'interprétation*, Colloque d'Albi Langages et Signification. Presses universitaires de Toulouse, p. 161-168.

BILGER, M. (Coord.), 2008, *Données orales. Les enjeux de la transcription*, Cahiers de l'Université de Perpignan, Presses universitaires de Perpignan.

BLANCHE-BENVENISTE, C. *et al.*, 1990, *Le français parlé. Études grammaticales*, Paris, Éditions du CNRS.

BLANCHE-BENVENISTE, C., 2010, « Où est le *il* de *il y a* ? », *Travaux de linguistique*, 61, p. 137-153.

BLANCHE-BENVENISTE, C., JEANJEAN, C., 1986, *Le français parlé, transcription et édition*, Paris, Didier-érudition.

CAPPEAU P., GADET F., 2007, « L'exploitation sociolinguistique des grands corpus. Maître-mot et pierre philosophale », *Revue Française de Linguistique Appliquée* XII-1, p. 99-110.

C-Oral-Rom, Integrated Reference Corpora for spoken Romance languages, *lablita.dit.unifi.it/coralrom/*

CAPPEAU, P., GADET, F., 2010, « Transcrire, ponctuer, découper l'oral : bien plus que de simples choix techniques », *Cahiers de linguistique*, 35/1, p. 187-202.

CAPPEAU P., GADET F., 2013, « Quand *l'œil écoute* : que donnent à lire les transcriptions d'oral ? », Communication orale au CILPR de Nancy.

CFPP2000, Corpus de Français Parlé Parisien des années 2000, http://ed268.univ-paris3.fr/syled/ressources/Corpus-Parole-Paris-PIII/index.html

CFPQ, Corpus de Français Parlé Québécois, http://pages.usherbrooke.ca/cfpq/index.php

CIEL_F, Corpus International Écologique de la Langue Française, www.ciel-f.net

CLAPI, Corpus de LAngue Parlée en Interaction, http://clapi.univ-lyon2.fr/

Corpus di italiano parlato, http://www.ciao.it/Corpus_di_italiano_parlato_Con_CD_Emanuela_Cresti__2015181.

DELAIS-ROUSSARIE, E. et YOON, H.-Y., 2011, « Transcrire la prosodie : un préalable à l'échange et à l'analyse des données ». *Journal of French Language Studies* 21, p. 13-37.

DITTMAR, N., 2002, *Transkriptionsanalyse. Ein Leitfaden mit Aufgaben für Studenten, Forscher und Laien*. Oplade. Leske + Budrich.

ESLO-2, Etude SocioLinguistique sur Orléans, www.univ-orleans.fr/eslo/

FRAN, Corpus des Français d'Amérique du Nord, http://continent.uottawa.ca/corpus-et-ressources-electroniques/

GADET, F., 2013, « Un nouveau corpus recueilli dans la région parisienne », *Langage & Société* n° 143, p. 109-17.

GADET, F., GUERIN, E., 2012, « Des données pour étudier la variation : petits gestes méthodologiques, gros effets », *Cahiers de linguistique* 38-1, p. 41-65.

GADET, F., LUDWIG, R., MONDADA, L., PFÄNDER, S., SIMON, A-C., 2012, « *CIEL_F* : choix épistémologiques et réalisations empiriques d'un grand corpus de français parlé », *Revue Française de Linguistique Appliquée* XVII-1, p. 19-54.

HOUDEBINE, A.-M., 1979, « Pour qui, pourquoi et comment transcrire ? », *Le français dans le monde* n° 145, p. 28-35.

JACQUET-PFAU, C., SABLAYROLLES, J-F., PRUVOST, J., 2005, *Mais que font les linguistes ? : Les sciences du langage, vingt ans après*, Paris, L'Harmattan.

JEANNERET, T., 1999, *La coénoncation en français*, Berne, Peter Lang.

LÉON, P., 2007, *Phonétisme et prononciation du français*, Paris, A. Colin (5e édition).

MONDADA, L., 1998, « Technologies et interactions dans la fabrication du terrain du linguiste », *In* M. Mahmoudian et L. Mondada (eds.), *Le travail du chercheur sur le terrain. Questionner les pratiques, les méthodes, les techniques de l'enquête*, *Cahiers de l'ILSL*, n° 10, p. 39-68.

MOREAU, M.-L., 1985, « La boutonique et son objet d'étude. Conte linguistique », *Revue de Phonétique Appliquée*, 76, p. 385-395.

MOREL, M.-A., DANON-BOILEAU, L., 1998, *Grammaire de l'intonation. L'exemple du français*, Paris, Ophrys.

MORENO, A. (en cours), *Le discours rapporté dans la langue des jeunes : délimitations et modes d'interaction, de points de vue sociolinguistique, prosodique et pragmatique*, Thèse de l'Université de Paris Ouest.

OCHS, E., 1979, « Transcription as theory », *In* E. Ochs et B. Schieffelin (*Eds*), *Developmental pragmatics*, New York, Academic Press.

OFROM, Corpus Oral de Français parlé en suisse ROMande, http://www11.unine.ch/uploads/Documents/AM-MJB_OFROM-[1.0]x.pdf

PFC, Phonologie du Français Contemporain, http://www.projet-pfc.net/?accueil:intro

RIVENC, P., 1971, « Lexique et langue parlée », *in* A. Rigault (ed.), *La grammaire du français parlé*, Paris, Hachette, p. 51-69.

TCOF, Traitement de Corpus Oraux en Français, www.cnrtl.fr/corpus/**tcof**/

VaLiBel, VAriétés LInguistiques de BELgique, http://valibel.fltr.ucl.ac.be

VINCENT, D., 2009, « Corpus, banques de données, collections d'exemples. Réflexions et expériences », *Cahiers de linguistique*, 33 (2), p. 81-96.