

## La variation et les corpus : une entrée méthodologique, avec l'exemple du style

Françoise Gadet<sup>1</sup>

(1) MoDyCo & Université de Paris Ouest

Les corpus sont aujourd'hui incontournables en sciences du langage, avec pour effet la production d'une masse potentiellement illimitée de données. Mais tout est-il bon pour "faire corpus"? Habert avait plaisanté à ce sujet, avec l'heureuse formule "gros, c'est beau" (2000). Si une masse importante de données présente de l'intérêt pour certains objectifs, il en est d'autres qui exigent plutôt qualité et fiabilité (Cappeau & Gadet 2007), ce qui soulève la question des critères pour évaluer la qualité: qu'est ce que de "bonnes données"? Nous défendrons ici l'idée que c'est en amont, lors de la constitution d'un corpus, que se joue son traitement, et non *post hoc* sur des données recueillies à l'aveugle: c'est parce qu'on se pose certaines questions que l'on adopte telle ou telle démarche méthodologique, ce qui plaide pour une diversification des modalités de recueil.

### 1 Le style

Je ne parlerai ici que d'objectifs de corpus sur la *variation*, car si la documentation pour les faits de langue proches du standard est assez abondante, c'est loin d'être le cas pour les données non ou peu standard, qui soulèvent des questions sur la façon de traiter/situer les données.

#### 1.1 Des données pour étudier le style

Parmi les données les plus difficiles à recueillir, il y a ce qui relève du *style*, ou produit de l'instabilité des productions d'un même locuteur (variabilité stylistique, intra-locuteur, diaphasique). Cette variabilité donne lieu à un répertoire dont dispose tout locuteur, mis à contribution dans la vie sociale par son adaptabilité aux interactions diversifiées qu'il traverse (Goffman 1979 parle d'*ajustement*).

Le *style* n'est plus aujourd'hui la « dimension négligée » que pouvait dénoncer Bell 1984, il a au contraire donné lieu à des travaux, qui soulignent tous le caractère polysémique du terme (Coupland 2007, Auer 2007). Ce terme renvoie en effet, selon une opposition entre produits et processus, à des conceptions divergentes de la variabilité dans la langue :

- d'un côté, on localise le style en des points précis de la chaîne parlée, à partir de phénomènes linguistiques répertoriés, quantifiables, se manifestant dans la linéarité (c'est la notion de variable), autorisant l'hypothèse d'investissement social de phénomènes linguistiques;
- de façon radicalement opposée, une conception que l'on dira sémiotique regarde le style comme plus diffus, concernant tout ce qui contribue à la "signification sociale" des

énoncés.

On fait l'hypothèse que des données d'origines différentes ne documentent pas les mêmes faits linguistiques, qu'il y a donc des effets de la méthodologie de recueil des corpus, dont il faut dès lors diversifier les objectifs, les contextes, et les modalités de recueil.

Le style est souvent donné comme une évidence, maladroitement formulée dans la notion scolaire de "niveaux de langue" qui présente, parmi divers inconvénients, celui de supposer que c'est la même chose qui se dit en des styles différents. A partir de Labov, la sociolinguistique a renouvelé la problématique du style, en mettant en avant deux acquis décisifs: 1) il n'y a pas d'énoncé sans style, 2) il n'y a pas de locuteur monostyle, contrairement aux représentations de la doxa. La conception labovienne a deux effets concernant les données: encourager le recours à l'interview (voir *infra*), et favoriser la localisation du style dans des variables, lieux ponctuels dans le fil du discours où est censé se dire le style.

On se demandera dans quelle mesure les corpus peuvent contribuer à comprendre le style, difficile à documenter par son caractère fugace et instable. Il a surtout été étudié à partir d'interviews<sup>1</sup>, en particulier dans la tradition labovienne qui alterne différents moments afin de provoquer différents styles; avec pour limites une certaine monotonie et le formatage par le format de l'interview (qui n'est souvent qu'un faux dialogue, donnant au contraire lieu à un certain monologisme - voir Thibault & Vincent 1990).

## 1.2 Le style en français, et chez les jeunes

Le français est une langue très standardisée qui connaît une vaste amplitude et souplesse stylistique, en même temps qu'une certaine rigidité de ses genres. Les styles ont été bien étudiés à l'écrit, moins à l'oral (ce qui peut être un effet d'idéologie du standard). Ce sont surtout la phonologie et le lexique qui ont été étudiés, la prosodie un peu moins, le discursif, le pragmatique et le syntaxique encore moins. Quand le style est étudié à l'oral, c'est en général chez des locuteurs des classes moyennes censés disposer d'une palette plus large, ce qui répond à la conception traditionnelle en standard/non-standard et formel/informel. Il est cependant possible d'étudier le style partout, puisque tous les locuteurs sont *pluristyle*, disposant d'une amplitude de *répertoire* plus ou moins vaste, mais jamais monocorde. On peut en particulier l'étudier chez les jeunes, qui offrent un effet de loupe sur le style, par leur rapport direct au style et par leur croyance en la spécificité de leurs façons de parler.

L'ANR-09-FRBR-037-01 (*Multicultural London English/Multicultural Paris French*) a constitué un corpus en région parisienne (voir Gadet/Guerin 2012), étudiant les façons de parler des jeunes en train d'advenir en région parisienne, en rapport avec les contacts de langues. La population enquêtée concerne des urbains, jeunes (ou se voulant tels - voir Bulot 2004), de milieux sociaux modestes, issus de l'immigration. Nous élaborons une tentative de caractérisation "interne" non dominée par des critères socio-démographiques, qui impose un passage au crible des données recueillies. Nous disposons ainsi d'entretiens réalisés sur une base d'exploitation de réseaux, reposant pour partie sur une qualité de la relation antérieure entre l'enquêteur et l'enquêté. C'est pourquoi le corpus MPF est l'œuvre conjointe de 17 enquêteurs différents, qui à l'heure actuelle ont recueilli plus de 37 heures aujourd'hui transcrites/revues (437.000 mots).

---

<sup>1</sup> En préparant une banque de données sur les corpus de français hors de France pour la DGLFLF, on a pu constater sur 145 réponses répertoriées qu'environ 80% des corpus ont été constitués sur la base d'interviews.

Le focus porte donc sur le pôle "informel", celui du *vernaculaire*, défini par les effets sur les façons de parler de contextes relevant de l'immédiat communicatif (Koch & Oesterreicher 2001), et non sur l'échantillonnage socio-démographique, en cherchant à dépasser le mythe d'un degré zéro de surveillance sociale, auto- ou hétéro-, puisque c'est plutôt la qualité de l'interaction qui est en jeu. On suppose ainsi une conception du style non comme production solitaire, mais comme se jouant dans l'interaction (*recipient design* de Goffman, repris par Bell 1984 dans sa conception d'*audience design*).

Nous nous arrêterons maintenant à deux aspects des réflexions méthodologiques soulevées à l'occasion de la constitution de ce corpus, concernant un type de population (à tort) tenue pour ne disposer que d'un répertoire stylistique limité, peu susceptible d'être prise comme terrain pour illustrer la variabilité stylistique.

## **2 La "qualité" des interviews : se forger l'oreille par des écoutes réitérées**

Nous avons recueilli des données en gros caractérisables comme relevant de trois types<sup>2</sup>: entretiens "traditionnels" (relation préalable faible ou inexistante entre les protagonistes, script préparé à l'avance); entretiens "de proximité", avec des protagonistes sélectionnés par réseau; données "écologiques", événements discursifs qui prennent place indépendamment des sollicitations d'un enquêteur. Ces derniers sont les plus difficiles à recueillir, car ils ne peuvent l'être sans l'aide active d'enquêtés, qui n'ont pas toujours l'envie, la capacité ou le savoir pour répondre aux objectifs du chercheur<sup>3</sup> - et la plupart des gens ont du mal à comprendre pourquoi l'ordinaire peut être digne d'intérêt.

### **2.1 La diversité des données : comment**

L'hypothèse derrière la diversification recherchée ne recoupe ni la diversité en genres, ni la recherche de différents moments stylistiques qui était à l'œuvre dans les interviews laboviennes (moments du plus au moins formels, obtenus par la diversification des sujets traités).

Des entretiens peuvent-ils atteindre l'ordinaire ? Ils peuvent en tous cas y tendre, plus ou moins, car dès lors qu'on privilégie l'interaction, on est dans le qualitatif donc le continu. Faute de terme mieux adapté, on parlera ici d'*authenticité*, en étendant légèrement le raisonnement de Coupland 2003 ou de Palomarès & Tersigni 2001, car ces auteurs situent l'authenticité dans le fond, alors que le (socio)linguiste a aussi affaire à la forme.

La qualité d'un enregistrement n'est garantie ni par un quadrillage externe, ni par la reconduction de conditions d'interviews (script semi-directif, même lieu, même enquêteur)... Notre choix méthodologique est de ne pas compter sur des facteurs externes pour déterminer la validité d'un entretien, mais de chercher la perception de qualité (on a ainsi écarté environ 25% des enregistrements, pour des raisons qui vont de la qualité acoustique au sentiment d'inauthentique, de forçage, ou d'excès de "distance"). Les critères sur lesquels on fait reposer

---

<sup>2</sup> "En gros", car la frontière entre les trois types est souvent fragile, comme on peut s'y attendre. Voir Gadet/Guerin (2012) pour des précisions.

<sup>3</sup> Nous avons tout vu de ce point de vue, depuis les enquêtés qui ont d'abord accepté pour ensuite manquer tous les rendez-vous, jusqu'à l'enquêté qui se prend au sérieux et joue à l'enquêteur.

un tel jugement, qui fait l'objet d'évaluations et de discussions dans l'équipe, ne sont pas facilement généralisables (c'est pourquoi j'ai parlé de "se forger l'oreille")<sup>4</sup>.

Une certaine "qualité des enregistrements" peut être reliée (certainement pas de façon mécanique) à la familiarité avec un terrain/une personne. On peut ainsi s'appuyer sur des notions comme l'*histoire conversationnelle partagée* (Golopentia-Erescu 1988), la *communauté de pratique* (Eckert 2012) et leurs effets sur la co-construction du sens social, ou encore l'*immédiat communicatif* (Koch & Cesterreicher 2001). En voici deux exemples: Wajih qui a été 5 ans surveillant dans un collège de Mantes-la-Jolie, exploite ses relations avec des adolescents qu'il connaît depuis la 6<sup>e</sup>; Nacer, à l'association Zy'va de Nanterre, fait de l'aide aux devoirs et du soutien en arabe, ce qui facilite ses demandes d'entretien. Ces cas diffèrent d'enregistrements dans un même lieu, par un enquêteur qui n'a pas de connaissance préalable des enquêtés et qui suit un script<sup>5</sup>. Mais ces pratiques sont exigeantes en temps et en disponibilité personnelle, et elles limitent le rayon d'action.

L'enquêteur est donc confronté à la complexité et à l'intrication des facteurs en jeu, relevant de paramètres divers, peut-être non répertoriés et certainement non quantifiables. Ainsi, pour Nacer: il est Algérien, et n'appartient pas à la communauté des beurs qu'il interviewe. Mais son extériorité à voir avec une identité "imaginée" des jeunes enquêtés, ce qui d'ailleurs peut jouer dans des directions opposées, l'extériorité pouvant susciter l'empathie favorisant le naturel, ou être perçue comme une distance qui entraîne surjeu ou absence de collaboration (en l'occurrence, cela a joué favorablement).

Pour mesurer l'impact de l'enquêteur, l'histoire de la sociolinguistique a retenu l'expérience de Rickford & McNair-Knox 1994, qui fait interviewer une jeune fille afro-américaine successivement par une enquêtrice de la même communauté, puis par une enquêtrice hors communauté, avec des résultats probants sur le plan quantitatif (taux de variables). Il reste néanmoins à s'interroger sur le sens social que l'enquêté peut prêter au renouvellement d'une interview, même à long terme (avec probablement plus ou moins les mêmes questions). Car les relations humaines ne rejouent jamais la même pièce, et les interactions véritables ne sont pas répliquables.

## 2.2 La diversité des données : pourquoi

Nous nous interrogeons sur ces aspects des données qui, pour la syntaxe, ne sont que rarement, ou insuffisamment, ou jamais, documentés dans les corpus construits sur la seule base d'entretiens, même dits informels. On fait ainsi l'hypothèse que des données obtenues dans des conditions différentes ne montreront pas tout à fait les mêmes faits linguistiques, ni selon les mêmes fréquences, saillances ou contraintes (voir Gadet 2008), en supposant des effets de la différence des modes d'obtention, en particulier entre données sollicitées et données écologiques. Mais de quel ordre seraient-elles ? Les phénomènes vont-ils se présenter en dichotomie, ou selon une gradation fine ?

Se poser de telles questions constitue une invitation à dépasser la notion de variable,

---

<sup>4</sup> L'entretien étant une relation intersubjective, ce n'est pas l'enquêteur qui est plus ou moins bon, et un même enquêteur peut "réussir" un entretien et en "rater" d'autres, sans parler des maladroites, par exemple quand un enquêteur prend à son compte une expression comme *de la beuh*, sans attendre que ce soit l'enquêté qui introduise ses propres mots.

<sup>5</sup> Entendons-nous : ces pratiques ne sont pas à rejeter en bloc, car elles donnent des résultats d'un intérêt incontestable. Elles sont seulement à compléter par d'autres modalités de recueil.

davantage concernée par la place structurelle des phénomènes que par la réactivité sociale des enquêtés. Le quantitatif répond en effet à une conception du style qui, tout en imageant bien quelque chose, demeure d'un fort degré d'abstraction, et qui ne suit pas forcément (ou pas du tout) les cheminements des interactants pour produire/interpréter du style.

La linéarité est-elle en effet une bonne hypothèse? Elle est certes conforme au mode de production/réception de l'oral, mais les interactants évaluent-ils le style selon des processus linéaires? Tenter ainsi d'adopter le point de vue du locuteur conduit à se demander ce qu'il "fait" avec le style. Comment en joue-t-il pour son propre compte? A-t-il une maîtrise sur sa capacité à moduler le style dans sa langue? Comment perçoit-il le style de l'interlocuteur? Est-ce que tout contribue à l'évaluation? Ou bien est-ce qu'au contraire il y a saillance de certains phénomènes? Et y a-t-il toujours cohérence? Ou une sorte de pondération entre traits, valorisants, discriminants, neutres?

Il y a là plus de questions que de réponses, mais ne pas se les poser reviendrait à adopter implicitement les réponses apportées par les prédécesseurs.

### **3 La palette stylistique d'un même locuteur**

Nous avons voulu écarter la tentation d'essentialisme, prenant un aspect d'une personne ou d'un groupe comme explication d'un comportement linguistique, en posant un lien entre identité collective et langue (Mendoza-Denton 2002).

#### **3.1 Des enregistrements en doublets**

MPF offre pour le moment 13 enregistrements dédoublés, offrant pour un même locuteur un enregistrement en interview de proximité et un enregistrement en situation écologique choisie par l'enquêté lui-même. Le recueil de ces situations écologiques s'avère extrêmement difficile, que ce soit en pratique d'auto-enregistrement ou dans la tentative pour faire comprendre à des non-linguistes ce que l'on attend d'eux<sup>6</sup>.

Il s'agit donc de deux situations d'immédiat, l'une des deux étant sollicitée et l'autre pas. Ce n'est pas un schéma méthodologique classique, car la pratique habituelle est plutôt de rechercher un écart maximal (formel *vs* informel), ce qui renvoie en fin de compte à une certaine conception de la langue où le tranché l'emporte sur un continuum subtil et réversible.

On peut s'interroger sur la "comparabilité" de données ainsi constituées, sachant que le style doit avant tout être conçu comme distinctif (ou contrastif). La comparabilité est en effet en général considérée comme assurée par des facteurs externes, alors que nous cherchons à construire cette notion de façon interne, sur des considérations d'authenticité linguistique et langagière.

#### **3.2 Stéphane, en entretien et en écologique**

Stéphane (pseudo) n'a rien de jeune, il a 36 ans (et aurait donc échappé à un quadrillage socio-démographique). Pourtant, son mode de vie et ses options fondamentales font qu'il se

---

<sup>6</sup> Il faudrait être naïf pour se persuader que l'auto-enregistrement obtenu dans ces conditions sera exempt de toute conscience. Pourtant, les productions des interactants sont encore une fois en continuum, et les réflexions sociologiques sur les groupes ont montré que le poids de la situation en groupe pouvait l'emporter sur la situation *in absentia*, celle de l'engagement envers le chercheur. La seule alternative est le micro caché (voir à ce sujet les réflexions de Gasquet-Cyrus à paraître).

vit comme "jeune". Nous disposons de deux enregistrements transcrits, "Stéphane entretien" (enregistrement pratiqué par un ami d'une amie d'enfance) et "Stéphane écologique" (conversation ordinaire, chez lui, avec un intime). Les deux enregistrements ont été retenus pour la qualité interne et leur authenticité.

(1) *Stéphane entretien*

Les french pour euh pour être international en fait c'est un délire de sape moi je reviens dans la musique en solo avec un délire de sapes de fringues tu vois où c'est vintage de 50 à 80 tu vois et en fait euh je réinvente le style je vais essayer de réinventer un style hip hop parisien tu vois le délire parce qu'on en a marre d'être habillés comme des blaireaux d'Américains là non mais ça allait ça allait quand c'était pas populaire maintenant c'est trop populaire ici n'importe quel blaireau ici il porte un il porte tout un jordan il porte un baggy tu vois avant nous ça avait une signification aujourd'hui ça en a plus non c'est c'est devenu un peu la la l'habit des beaufs tu sais ce que c'est un beauf<sup>7</sup>

Un beauf c'est le français moyen c'est euh c'est le français lambda non un bobo c'est un bourgeois bohème c'est un fils de pute qui a plein d'argent et qui va faire genre il en a pas et il s'habille avec des guenilles tu vois moi je suis le contraire moi je suis euh j'appelle ça boto bourgeois du ghetto on a pas d'argent mais faut que tout brille

(2) *Stéphane écologique (récit ordinaire fait à son ami Farid)*

St je sais pas qui vous êtes tu vois ce que je veux dire je leur ai fait comme ça (.) genre je parfois il y a des jeunes ils ont la haine sur nous hein

F Bien sûr

St Donc moi je vois vous vous me faites des zephas j'accélère hein je me casse hein après euh il me dit vous savez qu'on peut vous enlever le permis ben je fais bien sûr vous avez tous les droits je je connais vos droits monsieur et après bon que ça se reproduise pas

F Tu lui dis ouais peut-être que les jeunes ils voulaient me faire un quetru hein tu sais pas

St Non mais c'était eux les nejeu c'était eux les nejeu en <vrai>

F <Ah> ouais

St Moi je croyais que c'était des des des

F Tu lui dis peut-être que c'est oit hein tu voulais me faire un faire un quetru tu vois

St Je voyais ils étaient <despee>

F <Attends> tu t'annonces pas tu es un keuf

St Ben ouais je dis attends euh vous êtes en vilci vous jouez les cailleras

F Ben ouais

St Non mais trop de fois man hein franchement je me serais fait autant arrêté mais comme ça en privé en civil je me serais chié dessus vingt cinq fois (.) mais quand tu sais quand tu as le truc Securitas tu sais ils savent tu vailletras tu vois

### 3.3 La comparabilité des énoncés

Les deux enregistrements diffèrent de plusieurs manières. Une caractérisation en termes de variables<sup>8</sup> n'est certes pas incongrue ici: une seule occasion de *ne*, absent en (1), davantage en (2) (mais c'est surtout l'absence d'énoncés négatifs en (1) que cela désigne - différence de genres discursifs?); trois absences de *que*, en (2) (*je voyais ils étaient despee; tu t'annonces pas tu es un keuf; ils savent tu vailletras*), aucune en (1) (mais même remarque)...

<sup>7</sup> L'enquêteur, Roberto, est italien, et il a adopté une attitude de type "aide-moi à mieux comprendre les banlieues françaises".

<sup>8</sup> Il est un peu malhonnête de recourir à la notion de variable sur des exemples brefs, parce qu'elle est justement destinée à traiter des corpus longs, avec l'idée que le quantitatif aplanit d'éventuelles idiosyncrasies.

Est-ce toutefois ce qui est décisif? Ces remarques n'épuisent pas les différences, devant des facteurs plus difficiles à repérer, comme la connivence et les allusions, opaques à un observateur extérieur, qui concernent peu des points précis mais plutôt des facteurs comme le débit et le rythme, la courbe intonative et plus particulièrement la présence de certains contours; le lexique, en particulier la présence de beaucoup de verlan (*nejeu, quetru, oit, keuf, cailleras, despee, vailletras, vilci, zephas*), plus spécifique que l'argot que l'on rencontre aussi en (1)<sup>9</sup>.

On peut donc se demander en quoi la syntaxe et le discours contribuent au jugement de différences entre (1) et (2). Peut-on l'interpréter en termes de saillance? (comme *genre* ou *même*), de cumul? (accumulation de traits non-standard)? Est-ce interprétable perceptivement par une oreille extérieure au contexte?

Il y a là de vraies questions.

## **Conclusion**

Qu'a-t-on appris sur le style? On a opposé la pratique de la variable au tissage de phénomènes, convergents ou divergents, cohérents ou non, qui contribuent à produire une évaluation pour les interactants (dont le linguiste ne fait que mimer l'attitude), et jouent un rôle dans la constitution du sens social. Le style, ou le stylage (*styling*) est une activité dynamique, un processus de co-construction, plus qu'un produit stabilisé.

Et qu'a-t-on appris sur les corpus? Que faudrait-il que soient les corpus pour mieux servir une réflexion sur le style? Il n'y a toujours pas assez de corpus de français (de tous les français), et ils ne sont pas suffisamment variés. Mais il n'y a pas non plus suffisamment d'exigences méthodologiques, pas suffisamment de sensibilité aux conditions d'enquêtes, auxquels pourtant les sociologues ou les ethnologues sont si sensibles (Bizeul 2007); et les linguistes se laissent trop souvent aller à la facilité de reconduire ce qu'ont fait les prédécesseurs. Plusieurs sociolinguistes ont d'ailleurs fait remarquer récemment qu'il était de moins en moins de bon ton d'afficher ses problèmes méthodologiques (contrairement à ce que font les ethnologues).

Les considérations sur le terrain ne doivent certes pas conduire à considérer qu'il faut empiler n'importe quoi comme données, mais au contraire, qu'il faut prêter la plus grande attention à chacun des gestes méthodologiques pratiqués par le linguiste, chaque choix pouvant avoir des conséquences dont l'impact est difficile à mesurer avant la phase d'exploitation, moment où il sera trop tard pour s'apercevoir qu'on a laissé passer quelque chose.

## **Références**

AUER P. (2007 Ed.). *Style and Social Identities. Alternative Approaches to Linguistic Heterogeneity*. Berlin: Mouton de Gruyter.

BELL A. (1984). Language Style as Audience Design, *Language in Society* 13, 145-204.

BIZEUL D. (2007). Que faire des expériences d'enquête? Apports et fragilité de l'observation directe. *Revue française de science politique*, 57, 69-89.

---

<sup>9</sup> Le verlan, omni-présent en (2), est-il évité en (1) parce que Stéphane pense que Roberto ne comprendra pas ? Impossible de répondre à cette question. Pour le lexique, voir le *Dictionnaire de la zone*.

BULOT T. (2004). Les parlers jeunes et la mémoire sociolinguistique. Questionnements sur l'urbanité langagière. *Cahiers de Sociolinguistique* 9, 133-47. Rennes : Presses Universitaires de Rennes

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

COUPLAND N. (2003). Sociolinguistic authenticities, *Journal of Sociolinguistics* 7, 416-31.

COUPLAND N. (2007). *Style. Language variation and identity*, Cambridge University Press.

*Dictionnaire de la zone*. [www.dictionnairedelazone.fr](http://www.dictionnairedelazone.fr)

ECKERT P. (2012). Three Waves of Variation Study: the Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology* 41, 87-100.

GADET F. (2010 [2008]). Les corpus oraux et la diversité des productions langagières, *Verbum*, Tome XXX n°4, 261-73.

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

GASQUET-CYRUS M. (à paraître). Cachez ce micro (caché) que je ne saurais voir.

GOFFMAN E. (1979). Footing, *Semiotica* 25, 1-29, tr. fr. 1981, La position, *Façons de parler*, Editions de Minuit.

GOLOPENTIA-ERESCU S. (1988). Interaction et histoire conversationnelle, in *Échanges sur la conversation*, J. Cosnier, N. Gelas et C. Kerbrat-Orecchioni (Dir.), Paris : Éditions du CNRS. 69-81.

HABERT B. (2000). Des corpus représentatifs : de quoi, pour quoi, comment ? In Bilger M. (éd.), *Linguistique sur corpus - Etudes et réflexions*, Perpignan, Presses Universitaires de Perpignan.

KOCH P. & CESTERREICHER W. (2001). Langage parlé et langage écrit, in *Lexikon der romanistischen Linguistik*, tome 1, 584-627, Tübingen, Max Niemeyer Verlag.

MENDOZA-DENTON N. (2002). Language and Identity, in Chambers *et al*, *The Handbook of Language Variation and Change*, Oxford, Blackwell Publishing, 475-99

PALOMARES E. & TERSIGNI S. (2001). Les rapports de place dans l'enquête: les ressources du malentendu. *Langage & Société* 97, 5-26.

RICKFORD J. & MCNAIR-KNOX F. (1994). Addressee- and topic-influenced style shift: a quantitative sociolinguistic study, in D. Biber and E. Finegan (Eds) *Sociolinguistic Perspectives on Register*, New York: Oxford University Press, 235-76.

THIBAUT P. & VINCENT D. (1990). *Un corpus de français parlé*, Québec, Recherches sociolinguistiques 1.