



**Lidil**

Revue de linguistique et de didactique des langues

**32 | 2005**

**Sémantique des noms et adjectifs d'émotion**

---

## « Autour du *Corpus de référence du français parlé* »

Recherches sur le français parlé n° 18, Équipe DELIC, 2004, Publications de l'université de Provence, 265 p.

**Mathieu Avanzi**

---



### Édition électronique

URL : <http://journals.openedition.org/lidil/120>

DOI : 10.4000/lidil.120

ISSN : 1960-6052

### Éditeur

UGA Éditions/Université Grenoble Alpes

### Édition imprimée

Date de publication : 1 décembre 2005

ISBN : 2-914176-13-9

ISSN : 1146-6480

### Référence électronique

Mathieu Avanzi, « « Autour du *Corpus de référence du français parlé* » », *Lidil* [En ligne], 32 | 2005, mis en ligne le 18 juillet 2007, consulté le 21 septembre 2020. URL : <http://journals.openedition.org/lidil/120> ; DOI : <https://doi.org/10.4000/lidil.120>

---

Ce document a été généré automatiquement le 21 septembre 2020.

© Lidil

---

## « Autour du Corpus de référence du français parlé »

Recherches sur le français parlé n° 18, Équipe DELIC, 2004, Publications de l'université de Provence, 265 p.

Mathieu Avanzi

---

- 1 Sous l'influence des autres pays européens qui possèdent « un corpus de référence de leur langue » bien fourni (Espagne, Portugal, Italie), et qui en ont extrait des grammaires ou des analyses (Biber pour les Anglo-saxons), il devenait urgent pour « la communauté des linguistes, chercheurs et enseignants, [de disposer d']un témoignage de la langue française parlée aujourd'hui dans l'hexagone » (p. 11). Cette exigence répondait de fait à une commande de la Délégation générale à la langue française (Ministère de la culture) qui avait confié en 1998 à C. Blanche-Benveniste et son équipe la réalisation d'un vaste corpus de référence pour la langue française. Le groupe *Description linguistique informatisée sur corpus (DELIC)* a hérité de ce vaste projet et expose dans ce numéro ses premiers résultats.
- 2 Si l'on peut regretter que le terme même de « référence » ne soit à aucun moment clairement défini<sup>1</sup>, il n'en reste pas moins que c'est avec intérêt que l'on (re)découvrira à la lecture de ce recueil certains des axes de recherche développés par le *Groupe aixois de recherche en syntaxe (GARS)* au cours de ces vingt-cinq dernières années : réflexion sur la collecte des données (*partie 1*) ; analyse syntaxique des productions langagières (*partie 2*) ; analyse des ratés, bribes et autres pauses liées à leur actualisation dans le discours (*partie 3*). Les co-auteurs, habitués à une longue collaboration possèdent de puissants instruments pour l'établissement d'analyses statistiques et ne se privent pas d'en faire usage. Ils nous livrent ainsi une série d'articles associant avec succès le quantitatif et le qualitatif dans le cadre d'une linguistique de corpus informatisés
- 3 Pour ce qui est des chiffres, le CRFP compte 134 enregistrements, soit environ 440 000 mots pour quelque 36 heures de parole ! Les situations d'enregistrements sont différenciées et équitablement réparties selon a) des critères géographiques (taille, localisation des villes d'enquête), b) la longueur de la transcription (durée retranscrite

de l'enregistrement), c) la situation de parole (privée, professionnelle, publique), d) le nombre des locuteurs et e) leur âge, sexe et niveau scolaire.

- 4 Cette description, qui concerne les variables d'échantillonnage, est suivie de celles concernant le recueil des données proprement dit : procédures d'enregistrement (matériel et conditions d'enquête), fiche signalétique (sorte de carte d'identité de l'enregistrement où toutes les variables énoncées ci-dessus sont circonscrites), conventions de transcription (librement inspirées de l'ouvrage de C. Blanche-Benveniste et C. Jeanjean, *Le français parlé - transcription, édition* (1987) : pas de ponctuation, orthographe normalisée notamment) et soucis d'anonymisation (formalités juridiques et déontologiques) des corpus.
- 5 Pour finir, les auteurs proposent plusieurs moyens d'exploitation pour les éventuels intéressés : on peut soit consulter de courtes concordances (écrites seulement) sur Internet<sup>2</sup>, soit disposer de la totalité des documents audio sur CD-ROM (avec logiciel d'exploitation *Contextes*<sup>3</sup>). Notons également qu'une troisième version offre la possibilité de lire le texte avec le son « aligné » (grâce au logiciel *Transcriber*).
- 6 Les contributions de la deuxième partie de l'ouvrage prennent en partie<sup>4</sup> appui sur ces données et les indices de variation afin de préciser les comportements syntaxiques de certaines catégories et unités lexico-grammaticales remarquables. Par exemple, l'article de F. Sabio traite du « pronom y » en tant que marque de valence locative. Il apparaît que l'usage de ce morphème est restreint par diverses contraintes : traits de référence lexico-sémantiques, canal de communication (écrit/oral) ou situation de parole (formelle/informelle).
- 7 Dans le texte suivant, M. Bilger adopte une démarche relativement similaire et relate d'abord la répartition des « adverbes en -ment » selon les trois situations de parole dans le CRFP, puis elle examine ensuite les caractéristiques syntactico-sémantiques des plus fréquents d'entre eux (*vraiment, justement, évidemment, effectivement*).
- 8 Ce qui intéresse C. Chanet, c'est la grande catégorie des « marqueurs discursifs » (*enfin, en fait, quoi...*). Problématiques quant à leur traitement dans les analyses contemporaines, ils le sont tout autant quand il s'agit de les identifier à l'aide d'outils informatiques. L'auteur montre par une série de remarques pertinentes comment contourner les pièges posés par le repérage automatique des marqueurs discursifs sur gros corpus.
- 9 Enfin, pour clôturer cette section, trois auteurs s'arrêtent sur le fonctionnement des « pronoms conjoints liés au sujet » : P. Cappeau étudie dans ce cadre « les formes disjointes du pronom sujet » ; alors que M. Blasco-Dulbecco et S. Caddéo ciblent respectivement leurs analyses sur les séquences « *moi je* », « *lui, le propriétaire, le propriétaire, lui* ». Les trois linguistes insistent sur la propre part syntaxique des tournures qu'ils décrivent, et qui sont trop souvent négligées au profit d'une approche plus spécifiquement pragmatique.
- 10 Les dernières analyses sont consacrées aux modes de production orale : M.-N. Roubaud fait le point sur les problèmes de transcription/interprétation des « amorces de mots » en se basant sur les brouillons de transcription du CRFP. Elle rappelle que l'importance de ces phénomènes d'achoppement réside dans le fait qu'ils renferment des informations à la fois sur les plans syntaxique, sémantique et énonciatif.

- 11 E. Campione s'interroge sur le rôle des « pauses » (silencieuses vs remplies – « euh ») et leurs similitudes avec les phénomènes d'allongements syllabiques dans le travail de formulation du discours et les faits d'hésitation.
  - 12 Pour finir, S. Henry et B. Pallaud explorent les mêmes phénomènes que ceux décrits précédemment (amorces et hésitations) mais vont plus loin. Ainsi, elles sont amenées à définir trois catégories pour « les amorces répétées » (complétées, modifiées ou inachevées) avec des statuts morphosyntaxiques différents.
  - 13 Au final, que dire de ce dix-huitième numéro de la série des *RSFP* ?
  - 14 – Certains éprouveront des sentiments partagés quant au contenu, qui vacille entre l'inédit (Sabio, Chanet...), le déjà-vu et la redondance (articles sur les pronoms disjoints du sujet, phénomènes d'amorce et d'hésitation entre autres).
  - 15 – D'autres seront ravis par ces descriptions tout à fait uniques et si particulières qui faisaient déjà le charme des avancées de la recherche sur l'oral des membres du GARS.
  - 16 – Quoi qu'il en soit, les professionnels de la langue (enseignants et chercheurs en linguistique) trouveront dans ce nouvel opus des informations plus précises sur une base de données solide susceptible de les intéresser.
- 

## NOTES

1. Pour une discussion du concept de « référence », voir les actes du colloque de Perpignan (mai 1998), édités par M. Bilger (2000) : *Corpus. Méthodologie et applications linguistiques*, Paris, Champion ; et les actes du colloque de Louvain-La-Neuve (novembre 1999), édités par M. Francard (2000) : *Le français de référence. Constructions et appropriations d'un concept*, Cahiers de l'Institut de linguistique de Louvain, 26, 1-4S.
2. <<http://www.up.univ-mrs.fr/delic/crfp>>
3. Logiciel développé par J. Véronis (université de Provence). Pour davantage d'informations et les mises à jour : <<http://www.up.univ-mrs.fr/veronis/logiciels/Contextes>>
4. Habités à recueillir des données, les auteurs possèdent d'autres corpus tout aussi sinon plus fournis que le CRFP. Ils usent de ces ressources (écrites et orales mais aussi variées en « genres ») pour enrichir leurs comptages, analyses et conclusions.

---

AUTEUR

MATHIEU AVANZI

Université de Neuchâtel