

BILGER MIREILLE
Université de Perpignan-via-Domitia
CRESEM, axe Langages et Identités
bilger@univ-perp.fr

Réflexions sur les apports/les limites des corpus pour la description en syntaxe

Reflections on the Contributions/Limitations of Corpora for Syntactic Description

Abstract

The present real consensus is that corpora are excellent tools for observing, describing, teaching and learning language (Boulton & Tyne 2014). After a brief reminder of some of the discrepancies that this consensus may conceal and of the somewhat special situation of French in the field, compared, among others, to Anglo-Saxon work (Biber 2006), we will show how the exploitation, both quantitative and qualitative, of corpora of different media and genres, proves to be essential if we wish to propose a more accurate representation of the facts of language (Bilger & Cappeau 2015). The notion of frequency, which is regularly exploited from a variationist perspective, will also enable us to question certain oppositions that are still too often presented as fundamental, in particular that between oral and written language.

Keywords: corpus linguistics, oral/written French language, variation, register

Mots-clés : linguistique sur corpus, français parlé et écrit, variation, genres

Introduction

Aujourd'hui, il existe un réel consensus pour considérer les corpus comme d'excellents instruments d'observation, de description et d'enseignement-apprentissage de la langue (entre autres, Boulton & Tyne 2014). Après un bref rappel de certaines divergences que peut cacher ce consensus et de la situation quelque peu particulière du français dans le domaine, par rapport, entre autres, aux travaux anglo-

saxons (Biber 2006, Cheng 2012), nous montrerons comment l'exploitation, à la fois quantitative et qualitative, de corpus de médium et de genres différents, se révèle primordiale, si l'on souhaite proposer une représentation plus juste des faits de langue (Bilger & Cappeau 2015). La notion de fréquence, qui est régulièrement exploitée dans une perspective variationniste, nous permettra, par ailleurs, d'interroger certaines oppositions, encore trop souvent présentées comme fondamentales, en particulier celle que l'on pose entre oral et écrit.

1. Rapide état des lieux

Le débat sur l'exploitation des corpus pour mener à bien la description d'une langue semble clos. L'ensemble de la communauté des linguistes est d'accord pour dire que la linguistique sur corpus a été et est utile pour « repenser » la grammaire et l'enseignement de celle-ci (via la production de ressources pédagogiques à partir de données authentiques), qu'elle a permis et permet encore de donner un éclairage neuf sur de nombreux faits de langue, qu'elle en renouvelle à la fois l'analyse et la présentation.

Ce constat est indéniable, et relativement ancien, notamment pour l'anglais, via l'ouvrage de Biber *et al.* (1999) pour le domaine grammatical, via le Projet Touchstone, dirigé par MacCarthy (2004) pour le domaine didactique et pour le domaine de la lexicologie via le Projet Cobuild, initié par Sinclair en 1991.

Ce consensus à la fois théorique et méthodologique peut cependant cacher des divergences :

- a) Le terme « corpus » peut renvoyer à des réalités bien différentes issues de pratiques et de traditions qui se distinguent entre elles par la façon dont les textes sont regroupés ; on oppose ainsi souvent « corpus textuel » et « corpus de référence ». Le corpus dit « textuel » est constitué de textes entiers non organisés ; on parle aujourd'hui plutôt « d'archive » que de corpus pour désigner ce type de collection et de regroupement. C'est ce type de corpus qui a prévalu en France. En revanche, dans les pays anglo-saxons, c'est la constitution de corpus dit « de référence » qui a été favorisée. Ce type de corpus se caractérise cette fois par le souci de présenter de manière équilibrée et sous la forme d'échantillons le plus grand nombre d'usages possibles à l'écrit et à l'oral (cf. le British National Corpus).
- b) Les fonctions que l'on va accorder au corpus peuvent également être différentes. Les données seront-elles utilisées comme simple réservoir d'exemples ou comme matériau source pour guider la description ? On retrouve ici les deux approches désignées par les expressions « corpus based » ou « corpus driven ».

Ces divergences ne sont pas neutres, elles peuvent au contraire avoir des incidences sur la description syntaxique envisagée. Aussi, sans vouloir remettre en cause l'intérêt des corpus, il convient de s'interroger sur leur utilisation, faire le tri entre les services qu'ils rendent et les mirages qu'ils peuvent faire naître. De fait, selon la façon dont le corpus a été conçu, selon les raisons pour lesquelles ce dernier l'a été et selon la manière dont il a été exploité, la portée - ou encore la validité - de la description basée sur ce type de données risque d'être fort différente, voire discutable et certaines questions méritent d'être posées¹ :

¹ Cf. entre autres, Charaudeau (2009).

- Quelle est la portée réelle d'une description basée sur un corpus homogène construit en fonction d'un objectif spécifique ?
- Quelle est la représentativité d'un usage relevé dans un corpus dont les locuteurs ont été sélectionnés selon des catégories prédéfinies ?
- À quel moment peut-on extrapoler et parler d'un usage général ? particulier ?
- Comment éviter d'avoir une démarche circulaire ? Concevoir un corpus pour vérifier une hypothèse et forcément l'y retrouver.
- Enfin, en l'absence d'un grand corpus de référence, comment contraster les résultats de l'analyse ?

1. 1. Pour le français : quelles ressources ?

Nous n'avons toujours pas de grand corpus de référence du français. La France occupe en ce sens une situation particulière et atypique par rapport à d'autres pays européens (Espagne, Portugal, Grande-Bretagne, etc².)

Pour l'écrit, on a bien entendu la base FRANTEXT³ et pour l'oral, depuis « le corpus Aixois » initié en 1985⁴, les corpus se sont multipliés. Ils sont de plus en plus variés tant par les caractéristiques des locuteurs que par les situations de paroles ou les lieux de recueils ; et à partir des années 2000, l'accès à leurs données s'est beaucoup simplifié. Nous pouvons citer ici les principaux :

- Phonologie du Français Contemporain (PFC) – Corpus de Langue Parlée en Interaction (CLAPI) – Enquête Socio-Linguistique à Orléans (ESLO-1et ESLO-2) – Corpus Oral de Français de Suisse Romande (OFRON) – Corpus Multicultural Paris French (MPF) - Corpus d'interactions écologiques de la langue française (CIEL-F) – Corpus du Français Parlé de nos Régions (CFPR).
- Enfin, depuis 2020, la communauté a à sa disposition un corpus constitué de données orales (4 millions) et écrites (6 millions) en accès libre et téléchargeables ; c'est le Corpus d'Etude pour le Français Contemporain (CEFC) disponible via la plate-forme ORFEO (Outils et Ressources sur le Français Ecrit et Oral sous la direction de J.-M. Debaisieux [2013–2020]).

Ce corpus qui est constitué d'échantillons de corpus déjà existants a été conçu pour refléter les diverses variétés du français, aussi bien à l'oral qu'à l'écrit. En ce sens, il est assez proche d'un corpus dit « de référence », susceptible de servir d'étalon à l'ensemble des recherches ; il est cependant trop tôt pour savoir comment ce dernier sera exploité et s'il le sera dans ce sens-là.

2 Par rapport aussi à la Pologne qui s'est dotée de ce type de corpus dans les années 2006–2010, et qui est constitué de 250 millions de mots dont 10% d'oral.

3 La base Frantext initiée dans les années 1970 est constituée à 80% de sources littéraires. C'est en tant que telle une archive, même s'il est aujourd'hui possible de « créer » des sous-corpus échantillonnés selon des besoins spécifiques.

4 Premier grand corpus de productions orales constitué et analysé par le Groupe Aixois de Recherches en Syntaxe (GARS) dirigé par Cl. Blanche-Benveniste.

2. Quelles incidences pour la description ?

24

En dépit de ces nombreuses ressources, l'absence d'un grand corpus de référence (oral/écrit) sur lequel appuyer les recherches a eu des incidences pour la description du français, on peut en rappeler cinq :

1° – La parution tardive d'une « grammaire encyclopédique du français » censée prendre en compte aussi bien les données de l'oral que celles de l'écrit, le standard que le non standard, cf. *La Grande Grammaire du Français* de Abeillé & Godard, parue seulement en 2021 (plus de 20 ans après celles d'autres langues comme l'italien, le castillan, l'anglais, le portugais, etc.).

A noter cependant que dès la parution de cet ouvrage, il y a eu un certain nombre de réserves quant à la place de l'oral et le traitement des données. Par exemple, Cappeau (2021 : 12–13) signale que « l'oral est envisagé comme une dimension variationnelle au même titre que les français régionaux ou non hexagonaux ». Il ajoute « ne sont traités que des faits singuliers, atypiques dans le sens absents de l'écrit standard et généralement dépréciés par la norme ».

Coltier et Masseron (2022 : 203) font le même constat : « C'est une grammaire finalement centrée sur la norme de l'écrit en dépit des dénégations, sur le modèle phrastique et les catégories traditionnelles ». Elles poursuivent en précisant « À ces réserves, s'ajoute celle du traitement des exemples : les exemples construits occupent centralement le contenu des chapitres et relèguent dans une sous-partie finale les exemples de variations attestées ».

Sans remettre en question l'intérêt indéniable de cette grande grammaire, on peut regretter ce traitement somme toute assez peu original des données attestées de l'oral, qui rappelle celui que l'on relève dans d'autres ouvrages de référence (cf., entre autres, *La grammaire méthodique du français* de Riegel, Pellat & Rioul 2009).

2° – Contrairement au domaine anglo-saxon riche de deux descriptions grammaticales (issues des approches « corpus based » et « corpus driven »), le domaine français apparaît encore largement déficitaire (cf. Deulofeu & Debaisieux 2012). Les travaux sur le français à partir de l'exploitation des corpus sont certes de plus en plus nombreux, mais encore bien trop parcellaires, pour fournir une description de la langue prise dans son ensemble.

3° – Les descriptions des formes grammaticales dans les manuels (dictionnaires/grammaires LM/FLE) sont encore celles qui correspondent plutôt à l'usage écrit littéraire (issu notamment des données de FRANTEXT), ou encore à un usage dit « naturel, non marqué, conforme à la norme ». Encore aujourd'hui, certains travaux se fondent sur des données quelque peu discutables ou qui interrogent, cf. ce que dit Le Goffic (2019 : 14) au sujet du choix de ses exemples :

Les exemples donnés sont tirés de sources très diverses (écrites ou orales) mais sont dans leur grande majorité des exemples fabriqués, visant à représenter un usage naturel, non marqué, tout en étant conforme à la norme (français standard). Les emplois s'écartant de cette norme (déviants, recherchés, archaisants) sont mentionnés en tant que tels (non-standards).

Cette pratique pose évidemment question, tant il est difficile de savoir ce que l'on doit entendre par « usage naturel » ? À quoi cela fait-il référence ? Cet usage, est-il largement partagé ? Et où le trouve-t-on ?

4° – Le fait que ces ouvrages/manuels fondent leurs descriptions sur une langue proche de l'usage littéraire, ou sur un usage « idéalisé », a pour conséquence annexe de diffuser (de manière plus ou moins implicite) des considérations normatives sur les faits de langue étudiés.

De fait, l'idée de norme et du « bon usage » a fortement brouillé la relation entre *français écrit* et *français parlé*, en installant des associations étroites, telles que : l'écrit avec *le bon usage* et l'oral avec l'usage *familier* voire *fautif*.

5°- Enfin, l'absence d'une réflexion suffisamment générale sur la sélection des données, fait que les phénomènes traités sont souvent présentés, sur un même niveau, qu'ils soient rares ou ordinaires, cf. le traitement de la locution « *quand bien même* » (Bilger & Cappeau 2017) que l'on a dans :

(1) Quand bien même je le voudrais, je ne le pourrais pas

Cette locution est régulièrement signalée dans les ouvrages de référence – *Le Grand Robert*, *La grammaire des premiers temps* (Abry & Chalaron 1999), la *Grammaire du français. Du mot au texte* (Aslanides 2001) – comme un des moyens pour exprimer la concession adversative, quand le verbe est au conditionnel. Cette locution est cependant très peu usitée, elle est rare dans l'ensemble des corpus, qu'ils soient littéraires, issus de la presse ou des productions orales. Elle occupe pourtant dans les grammaires (LM et FLE) une place non négligeable, sur laquelle on ne peut être que réservé, et elle ne s'accompagne dans les dictionnaires d'aucun commentaire particulier, du type « peu fréquent », voire « vieilli ».

Pour conclure, sur les incidences que l'absence d'un « grand corpus de référence » a eu pour la description du français, un constat encore d'actualité : comme le signalaient déjà Debaisieux et Boulton (2007), les outils pédagogiques sont encore lents à tirer profit des recherches basées sur corpus (si ce n'est peut-être dans le domaine du FLE). Dans les manuels destinés au grand public, la prise en compte de la variation, du point de vue des productions, a du mal à s'imposer, pour des raisons beaucoup plus « idéologiques » (prégnance encore vivace de certaines représentations de la langue) et « économie éditoriale » que par la carence de ressources ou de méthodologie descriptive.

3. Le traitement des données

Si la dimension variationnelle est un aspect important pour la description et la représentativité des résultats, la façon dont sont et seront traitées les données l'est tout autant. Sur ce point, l'informatisation des corpus a grandement modifié la relation du chercheur avec son objet d'étude, et a orienté les recherches en syntaxe. Ainsi, dans les années 1980/1990, la description s'appuyait sur une lecture suivie des corpus (notamment des transcriptions des corpus oraux) pour y rechercher des exemples ; cette tâche fastidieuse et chronophage entraînait cependant une connaissance fine et complète des textes, ce qui a permis pour l'oral de montrer l'existence d'une rhétorique fondamentale basée sur des phénomènes de répétitions et de ruptures de construction, cf. Blanche-Benveniste *et al.* (1990). Ce type d'approche a permis également de mieux cerner la variation intra-personnelle des locuteurs (présence ou absence du *ne* de négation, du *passé simple*, du *nous* en sujet, etc.) qui est souvent liée au « genre » ou à la « posture » du locuteur. Les travaux ont montré que, dans une même production orale, le locuteur change de « registre » lorsqu'il devient porte-parole d'un groupe ou quand il mentionne son activité professionnelle.

Par la suite, la numérisation des corpus a facilité l'exploitation de corpus de grande taille, et l'utilisation du concordancier (qui permet à la fois une étude qualitative et quantitative des données) a eu pour effets la multiplication de travaux en syntaxe portant sur des formes, dont il était/est facile de formaliser les requêtes (ce qui n'est pas toujours le cas). Cette linguistique « outillée » a d'ailleurs obligé de repenser les relations entre *lexique* et *grammaire* (cf. Halliday 1991) via la prise en compte des collocations,

des phénomènes phraséologiques, dans lesquels le poids du lexique est plus important que ce que l'on pouvait/pourrait penser. De fait, le concordancier permet de mieux percevoir la « résonance » d'un terme, d'en étudier mieux son usage en contexte. Il permet aussi de montrer que les emplois, les usages ne sont pas uniformes mais se distribuent de manière différenciée selon les types et les « genres » de productions. Cette notion de fréquence et sa prise en compte dans les descriptions a permis de réinterroger certaines oppositions, (encore trop souvent présentées comme fondamentales), en particulier, celle que l'on pose entre oral et écrit.

3.1. L'opposition « oral/écrit »

Cela fait plus de 30 ans que nombre de travaux ont permis de contester cette vision naïve d'une syntaxe différente entre ces deux modes de productions. Plus de 30 ans donc que l'on sait que du point de vue *qualitatif*, il y a en fait peu de différences (cf. entre autres, Blanche-Benveniste *et al.* 1990 et Blanche-Benveniste 2010), exceptés les phénomènes liés aux modes de production, tels que les bribes, les amorces de syntagme, les recherches de dénominations, etc.

En revanche, ce qui peut être très différent entre l'écrit et l'oral, ce sont les faits de distribution et l'importance quantitative de certains phénomènes. Mais ces différences vont pouvoir être modulées selon les types de productions orales ou écrites sur lesquelles on travaille. L'opposition « binaire » (oral/écrit) s'avère donc rapidement insuffisante, et d'autres découpages risquent d'influer tout autant sur les résultats des analyses.

Cette problématique en relation avec la typologie des productions est d'ailleurs, aujourd'hui, un domaine qui suscite de nombreuses réflexions, d'autant plus que la prise en compte de ces différents « genres » a permis et permet encore de montrer comment les phénomènes linguistiques se distribuent différemment en fonction de la « nature » des productions. Les travaux qui s'appuient sur différents types de productions, dont l'exemple le plus abouti reste encore l'ouvrage de Biber *et al.* (1999), donnent une idée de ce que pourrait être une grammaire fondée sur une méthode empirique, à partir des données de corpus diversifiés et échantillonnés.

3.2. L'aspect quantitatif

Si travailler sur des données quantifiées issues de corpus diversifiés fournit un angle d'observation intéressant, il n'en demeure pas moins vrai que le recours à la quantification et à la notion de fréquence mérite d'être interrogé. Par exemple, comment déterminer un nombre d'occurrences significatif ? Et significatif de quoi ? Comment interpréter l'absence d'une forme ou les faits équilibrés ? Ceux qui ne le sont pas ?

En ce qui concerne le nombre d'occurrences, Blanche-Benveniste et Biber préconisaient le fait d'avoir au minimum 50 exemples pour pouvoir décrire une forme. Cette précaution méthodologique semble raisonnable mais demande à être explicitée : faut-il 50 exemples d'une seule forme ou 50 exemples de chacune des constructions que peut avoir cette forme ? Il est évident que selon la réponse, la taille du corpus nécessaire à l'analyse risque d'être fort différente.

L'absence d'attestation d'une tournure soulève le problème du choix du corpus étudié ; il se peut que ce dernier ne soit pas adapté ou qu'il y ait eu des défaillances dans sa constitution liées à sa taille, ou qu'il soit trop homogène, etc.

L'étude de la séquence « sur ce » (Bilger 2020 : 101–114) que l'on trouve dans les exemples suivants :

(2) Il en donna une également à Rouletabille et continua son chemin vers la tour carrée. **Sur ce**, Rouletabille lui demanda ce qu'il allait faire à la tour carrée (Frantext)

(3) **Sur ce**, je te laisse (Frantext)

a posé ce type de problème. La description de cette forme à partir de divers corpus⁵ montre qu'elle est rare, même si elle est régulièrement citée dans les manuels (grammaires LM et FLE). On en relève seulement 110 occurrences sur les 150 millions de mots des corpus utilisés et essentiellement à l'écrit dans les données issues de Frantext. Cette séquence semble nécessiter cependant un contexte discursif interactionnel (dialogue) dans lequel la première personne est souvent présente. Les données issues de Sketch Engine confortent cette prédominance, mais dans le contexte, cette fois, d'une « correspondance immédiate » via les blogs. Comment interpréter alors l'absence de cette séquence dans les corpus oraux ? Sans aucun doute par le fait que les corpus n'étaient pas adaptés, en n'étant pas assez narratifs ou pas suffisamment interactifs, avec une distance trop grande entre les locuteurs⁶.

Quand on est amené à comparer différents types de corpus, il est intéressant de s'interroger sur les informations données par les faits équilibrés ou pas, par ceux qui sont massifs ou au contraire rares. Par exemple, un fait équilibré dans divers types de corpus peut s'interpréter comme un usage de la langue en général ; c'est le cas de la conjonction « lorsque » qui exprime la simultanéité et qui se répartit de manière homogène dans les différents corpus (cf. Bilger 2013).

En revanche, une distribution fortement différenciée selon les corpus donnera à la forme en question le statut de marqueur variationnel, sensible aux « genres ». C'est par exemple le cas de la locution « quand même », comme dans :

(4) il est venu **quand même** pour nous saluer (Oral Privé)

(5) ils étaient **quand même** contents de de s'être rencontrés (Oral Public)

que l'on relève essentiellement à l'oral, et ce quel que soit le type de production « oral privé ou public, formel ou informel », comme l'avait déjà signalé Beeching (2005 et 2007).

C'est aussi le cas de la conjonction « tandis que » qui a deux emplois, soit elle s'interprète comme un marqueur temporel :

(6) ils remontèrent les pentes du Sahel, **tandis que** la lune tardive émergeait des bois d'eucalyptus (Litt.)

soit comme un marqueur d'opposition:

(7) les figures de femmes avaient été découpées au couteau, **tandis que** les visages d'hommes avaient les yeux crevés (Presse)

(8) le terrain de pétanque c'est six à onze mètres **tandis que** nous ça commence déjà à douze mètres cinq (Oral)

L'étude de cette conjonction à partir de trois corpus de types différents (Litt., Presse et Oral) de 440.000 mots chacun montre que ces deux emplois ne sont pas équilibrés mais se répartissent de manière différente selon les corpus. Dans le corpus littéraire, la valeur temporelle est largement majoritaire (80%

5 L'étude a été menée à partir des corpus suivants : Frantext (période littéraire 1900–2013, 140 millions de mots) + Corpus CERF (Presse, Forum, 9 millions de mots) + corpus oraux (2 millions de mots).

6 Cette absence dans les corpus rappelle celle qui concerne la forme « comme » dans son emploi d'exclamatif (« *comme c'est joli, s'émerveilla Béliard* ») que l'on ne retrouve là encore que dans les corpus littéraires et très rarement.

des emplois). Dans le corpus Presse, la conjonction est plutôt utilisée comme un connecteur marquant l'opposition (56%) et dans le corpus Oral, seul ce dernier emploi est relevé. Une étude limitée à un seul type de corpus (excepté peut-être celui de la Presse) aurait biaisé l'analyse⁷.

Travailler sur des données quantifiées issues de corpus diversifiés fournit donc un angle d'observation intéressant, mais comme déjà dit, cette approche est assez peu exploitée dans les ouvrages/manuels LM. Prendre en compte ces considérations quantitatives modifierait pourtant les présentations habituelles. Nous n'en fournirons ici qu'un seul exemple concernant le traitement des prépositions « contre » et « entre » :

- dans les manuels et dictionnaires, c'est la valeur locative de ces deux prépositions qui est présentée en premier, ce qui correspond effectivement à l'usage littéraire. Dans ce type de corpus, ces prépositions dépendent de verbes, tels que : *s'appuyer, attirer contre, planter/se blottir entre...* Et l'on peut relever aussi dans le corpus littéraire que 20% des occurrences de la préposition *entre* construisent des syntagmes prépositionnels particuliers du type : *entre les sourcils/doigts/lèvres*.

En revanche, en ce qui concerne le corpus Presse, ces prépositions ne sont plus majoritairement locatives ; elles expriment d'autres valeurs, comme l'opposition pour *contre*⁸ (*lutter contre, le crime contre*), et la réciprocité pour « entre » (*les relations entre*) ou des limites temporelles (*entre 1914 et 1918*).

Dans le corpus de productions orales, on retrouve les mêmes valeurs que dans le corpus Presse mais ces deux prépositions sont deux à trois fois moins utilisées.

Ces informations d'ordre quantitatif semblent également peu présentes dans les ouvrages de FLE, contrairement à certaines préoccupations apparues lors de l'élaboration du français fondamental (Gougenheim *et al.* 1956). Elles joueraient pourtant un rôle important pour l'enseignement de la langue en permettant de hiérarchiser les emplois et leur apprentissage.

4. Pour conclure

Le rappel de ces quelques exemples signale comment les descriptions sont forcément tributaires des données sur lesquelles on travaille. Ils révèlent pourquoi les usages signalés en premier dans les ouvrages de référence (qui s'appuient encore essentiellement sur les données de Frantext) sont ceux qui correspondent aux usages littéraires, et non pour d'autres raisons, liées à une fréquence plus grande ou à « la norme » - ceci dit, en favorisant encore trop souvent ces seules sources littéraires, cela donne aisément à cet usage particulier le statut de « norme » et de « bon usage ». Les études sur corpus ont permis de tourner la page avec un certain « impressionnisme linguistique », mais on est encore loin, de pouvoir « contraster » les faits de langue susceptibles d'appartenir à un usage général de ceux qui sont spécifiques à un type de productions (littéraires, oral de conversation, etc.) ou à un type de locuteurs. Certes, les travaux allant dans ce sens sont de plus en plus nombreux, mais comme déjà dit, ils sont encore

7 Cf. Le Goffic (2019) qui ne mentionne que la valeur temporelle. La valeur d'opposition est seulement suggérée sous les termes : « une nuance possible d'opposition » (p. 147). Ce qui laisse supposer que son étude a été menée essentiellement sur des données littéraires.

8 Les données du corpus Presse permettent aussi de relever un emploi qui semble spécifique de ce « genre », c'est l'emploi de « contre » qui signale un rapport de deux grandeurs, comme dans : *c'est en province que la hausse a été la plus élevée : 2,1% contre 1, 8% pour Paris*. Cela correspond à 24% des occurrences du corpus journalistique.

bien trop fragmentés pour donner une vision descriptive d'ensemble. La prise en compte de la dimension variationnelle et de la dimension quantitative des données oblige de fait à revenir sur des oppositions souvent présentées comme fondamentales qui dépassent largement l'opposition entre les deux médiums oral et écrit.

Références

- Abeillé, Anne, Danielle Godard (éds.) (2021) *La grande grammaire du français*. Paris : Actes Sud Imprimerie nationale Éditions.
- Abry, Dominique, Marie-Laure Chalaron (1999) *La grammaire des premiers temps*. Grenoble : Presses universitaires de Grenoble.
- Aslanides, Sophie (2001) *Grammaire du français. Du mot au texte*. Paris : Champion.
- Beeching, Kate (2005) « Politeness-induced semantic change : the case of *quand même*. » [In :] *Language Variation and Change*. Vol. 17 ; 155–186.
- Beeching, Kate (2007) « La co-variation des marqueurs discursifs *bon, c'est-à-dire, enfin, hein, quand même, quoi et si vous voulez* : une question d'identité ? » [In:] *Langue Française*. Vol. 154 ; 78–93.
- Biber, Douglas (2006) *A Corpus-based Study of Spoken and Written Registers*. Amsterdam : John Benjamins Publishing Company.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan (1999) *The Longman Grammar of Spoken and Written English*. London : Longman.
- Bilger, Mireille (2013) « Etude de *alors que* et *tandis que* à partir de corpus oraux et écrits. » [In :] Jeanne-Marie Debaisieux (éd.) *Analyses linguistiques sur corpus, subordination et insubordination en français*. Paris : Hermès-Lavoisier ; 293–314.
- Bilger, Mireille (2020) « Les séquences *et ce* et *sur ce* à l'aune de la macrosyntaxe. » [In :] Anne Theissen, Annie Kuyumcuyan (éds.) *La polyvalence de «ce» en français : de la syntaxe à la cohésion textuelle*. *Studii de Linguistica*. Vol. 10 ; 101–114.
- Bilger, Mireille, Paul Cappeau (2015) « Comment les données de corpus pourraient renouveler les manuels de grammaire ? » [In :] *LINX*. Vol. 68–69 ; 177–199.
- Bilger, Mireille, Paul Cappeau (2017) « Ce que les corpus pourraient apporter aux grammaires et/ou aux dictionnaires : l'exemple de *contre* et *même*. » [In :] *Dossier HEL*. Vol. 11. « *Analyses et exploitation des données de corpus linguistiques* » ; 41–51.
- Blanche-Benveniste, Claire, Mireille Bilger, Christine Rouget, Karel van den Eynde (1990) *Français parlé. Analyses grammaticales*. Paris : édition du CNRS.
- Blanche-Benveniste, Claire (2010) *Le français : Usages de la langue parlée*. Leuven : Peeters.
- Boulton, Alex, Henry Tyne (2014) *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*. Paris : Didier.
- Cappeau, Paul (éd.) (2021) *Une grammaire à l'aune de l'oral*. Rennes : Presses Universitaires de Rennes.
- Charaudeau, Patrick (2009) « Dis-moi quel est ton corpus, je te dirai quelle est ta problématique. » [In :] *Corpus*. Vol. 8 ; 37–66.
- Cheng, Winnie (2012) *Exploring Corpus Linguistics: Language in Action*. London : Routledge.
- Coltier, Danielle, Caroline Masseron (2022) Compte rendu « Anne Abeillé, Danièle Godard (dir.), Grande Grammaire du Français. » [In :] *Scolia*. Vol. 36 ; 193–206.

- Debaisieux, Jeanne-Marie, Alex Boulton (2007) « Alors la question c'est... ? Questions pragmatiques et annotation pédagogique des corpus. » [In :] *Cahiers de l'Association for French Language Studies*. Vol. 13.2 ; 31–60.
- Deulofeu, José, Jeanne-Marie Debaisieux (2012) « Une tâche à accomplir pour la linguistique française du xx^e siècle : élaborer une grammaire des usages du français. » [In :] *Langue française*. Vol. 176 ; 27–46.
- Gougenheim, Georges, René Michéa, Paul Rivenc, Aurélien Sauvageot (1956) *L'élaboration du français élémentaire : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris : Didier.
- Halliday, Michael A.K (1991) « Corpus Studies and Probabilistic Grammar. » [In :] Karin Aijmer, Bengt Altemberg (éds.) *English Corpus Linguistics*. Londres-New-York : Longman ; 30–43.
- Le Goffic, Pierre (2019) *Grammaire de la subordination en français*. Paris : Ophrys.
- MacCarthy, Mike (2004) *Touchstone. From Corpus to Course Book*. Cambridge : Cambridge University Press.
- Riegel, Martin, Jean-Christophe Pellat, René Rioul (2009) *Grammaire méthodique du français*. Paris : P.U.F.
- Sinclair, John (1991) *Corpus, Concordance, Collocation*. Oxford : Oxford University Press.