



HAL
open science

Ce que les corpus pourraient apporter aux grammaires et/ou aux dictionnaires : l'exemple de contre et même

Mireille Bilger, Paul Cappeau

► To cite this version:

Mireille Bilger, Paul Cappeau. Ce que les corpus pourraient apporter aux grammaires et/ou aux dictionnaires : l'exemple de contre et même. Dossiers d'HEL, 2017, Analyse et exploitation des données de corpus linguistiques, 11, pp.41-51. hal-01511249

HAL Id: hal-01511249

<https://hal.science/hal-01511249v1>

Submitted on 20 Apr 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.

CE QUE LES CORPUS POURRAIENT APPORTER AUX GRAMMAIRES ET/OU AUX DICTIONNAIRES : L'EXEMPLE DE *CONTRE* ET *MÊME*

Mireille BILGER (U. de Perpignan-Via-Domitia, CRESEM-ELADD)

Paul CAPPEAU (U. de Poitiers, FoReLL)

Résumé

Si les corpus occupent une place majeure dans la linguistique contemporaine, il n'en demeure pas moins vrai que la question des données ainsi rassemblées est majeure et que l'on interroge trop peu, parfois, la question de leur représentativité. De fait, de nombreux outils disponibles (comme le *Trésor de la Langue Française* (TLF) et bon nombre de grammaires) continuent à ne représenter et décrire, pour l'essentiel, que l'usage écrit littéraire. Or, en ce qui concerne la description syntaxique, l'appui sur des corpus de nature et de genres différents (oral, écrit, littérature, presse, explications techniques, récits de vie, etc.) se révèle primordial si l'on souhaite proposer une représentation plus fournie et plus sensible à certains faits de variation. On retrouve là des éléments du débat devenu classique entre *corpus-driven* et *corpus-based* (Tognini-Bonelli, 2001).

Nous prendrons l'exemple de deux formes grammaticales (*contre* et *même*) afin d'illustrer le décalage qui se manifeste entre les descriptions existantes (dans les grammaires ou les dictionnaires) et les pistes que fournissent les emplois attestés dans les corpus.

Mots clés

syntaxe, corpus, écrit, oral

Summary

While corpora play a major role in contemporary linguistic studies, the underlying implications of using such collections of data are very important, and the issue of representativeness remains largely unquestioned. Thus, many available tools and databases (such as the *Trésor de la Langue Française* (TLF) and grammars) continue to confine their descriptions to what are basically written, literary usages. However, concerning syntactic description, recourse to different corpora representing different genres (spoken, written, literary, newspaper, technical, life stories, etc.) appears to be all important in providing a fuller description, more akin to dealing with certain variable elements (cf. the distinction between *corpus-driven* and *corpus-based* – Tognini-Bonelli, 2001).

As a basis for our discussion, we look at the forms *contre* and *même*, showing how existing descriptions (in grammars and dictionaries) fall short of what can be found from exploring authentic corpus data.

Key words

syntaxis, corpora, written French, spoken French

INTRODUCTION

S'il est acquis que les corpus occupent une place majeure dans la linguistique contemporaine, il n'en demeure pas moins vrai que la question des données ainsi rassemblées est majeure et que l'on interroge trop peu, parfois, la question de leur représentativité. De fait, de nombreux outils disponibles (comme le *Trésor de la Langue Française* (TLF¹) et bon nombre de grammaires) continuent à ne représenter et décrire, pour l'essentiel, que l'usage écrit littéraire. Or, en ce qui concerne la description syntaxique, l'appui sur des corpus de nature et de genres différents (oral, écrit, littérature, presse, explications techniques, récits de vie, etc.) se révèle primordial si l'on souhaite proposer une représentation plus fournie et plus sensible à certains faits de variation. On retrouve là des éléments du débat devenu classique entre *corpus-driven* et *corpus-based* (Tognini-Bonelli, 2001).

Nous prendrons l'exemple de deux formes grammaticales (*contre* et *même*) afin d'illustrer le décalage qui se manifeste entre les descriptions existantes (dans les grammaires ou les dictionnaires) et les pistes que fournissent les emplois attestés dans les corpus.

1. INCIDENCE DES DONNÉES SUR LA DESCRIPTION

Travailler sur corpus est susceptible aujourd'hui encore de soulever des oppositions et doit donc susciter une réflexion (Cori et David, 2008) sur le travail engagé pour en préciser le cadre et la portée (Arbach, 2015). La mise à disposition de grands corpus dont la composition est plus ou moins bien appréhendée par l'utilisateur peut, de fait, provoquer des biais dans les analyses conduites. Par exemple, la description des environnements de l'adverbe *délicatement* est fort différente selon la ressource utilisée :

- avec *Frantext*, l'adverbe *délicatement* s'associe à des parties du corps (*joues, lèvres, doigts*) et à des verbes comme *poser* ou *prendre* ;
- dans *Wortschatz*², les principaux co-occurents de l'adverbe sont *blancs, incorporer* et *spatule*.

Cette illustration montre l'impact décisif que les réalisations majoritaires dans le corpus produisent sur les descriptions. On décrit autant la forme elle-même que le corpus lui-même : *délicatement* ne se réalise pas, de toute évidence, dans les mêmes environnements dans les romans et les recettes de cuisine.

La question de la légitimité et de l'exhaustivité des descriptions faites à partir de corpus a alimenté de nombreux débats entre linguistes et débouche sur le constat que la variété des données est un paramètre déterminant, même s'il est fréquemment négligé ou trop faiblement effectué. Cela peut tenir notamment à la minoration des productions orales souvent ignorées ou marginalisées dans les descriptions syntaxiques, mais aussi à une diversité des genres trop étriquée (Blanche-Benveniste, 2010). L'ouvrage de Biber et *alii* (1999) pour l'anglais donne un bon exemple de ce que pourrait être une grammaire élaborée à partir de corpus diversifiés et échantillonnés, dans laquelle les dimensions quantitatives et qualitatives partagent le même poids descriptif.

¹ Nous avons utilisé la version en ligne <http://atilf.atilf.fr/>

² Il s'agit de ressources mises à disposition par des chercheurs de l'université de Leipzig : http://corpora.informatik.uni-leipzig.de/?dict=fra_mixed_2012S

2. PRÉSENTATION DES CORPUS UTILISÉS

Dans de nombreux corpus d'études, on observe souvent un fort déséquilibre en faveur de l'écrit qui représente entre 80 ou 90% des données exploitées. Cela s'explique souvent par des raisons techniques (il est plus facile de constituer un corpus volumineux de textes écrits que de productions orales) et par un prestige traditionnellement accordé à l'écrit³. Tel n'a pas été notre choix. Nos analyses distributionnelles, menées à l'aide du concordancier CONTEXTES⁴, s'appuient sur des données recueillies dans des corpus de petite taille (280 000 mots chacun), mais fortement contrastés du point de vue externe. L'ensemble constitue donc un recueil de plus de 1 million de mots⁵. L'écrit et l'oral sont équilibrés du point de vue quantitatif. La diversité reste cependant modeste dans ce travail puisque nous ne distinguons que 4 types de situations (2 pour l'oral et 2 pour l'écrit) :

- l'oral de parole privée (O-PRI) rassemble des prises de parole sur des sujets variés⁶. Il s'agit souvent d'interviews en face-à-face. Ce type de données donne accès à un oral que l'on peut qualifier de tout-venant ;
- l'oral de parole publique (O-PUB) collecte des interventions dans les médias, des prises de parole dans des réunions (assemblée, cours, etc.). Il s'agit là d'un oral plus surveillé de la part des locuteurs ;
- l'écrit de presse (E-Press) est constitué d'extraits de journaux (*Le Monde*, *Libération*) et d'un magazine (*Le Nouvel Observateur*). La langue utilisée illustre ce que l'on pourrait qualifier comme un écrit de style journalistique ;
- l'écrit littéraire du XIX^e siècle (E-Litt) regroupe quelques auteurs (Zola, Balzac, Maupassant, etc.). Il s'agit de la grande période romanesque française avec une langue légèrement datée mais intéressante à observer dans la mesure où c'est elle qui a longtemps constitué le terreau des exemples (et des descriptions) utilisés dans les grammaires.

Comme on le constate, les corpus présentent de fortes différences et il n'est pas question de rechercher une correspondance entre oral et écrit mais plus modestement de disposer de variations sensibles tant pour la langue parlée que pour la langue écrite. Bien évidemment, cette variété délaisse de nombreuses autres possibilités (la langue orale professionnelle, le parler des jeunes, la littérature contemporaine, etc.), mais elle reste suffisante pour notre démonstration. L'un de nos objectifs est d'illustrer comment les descriptions sont dépendantes des données utilisées et, en ce sens, peuvent difficilement être proposées sans tenir compte de ce critère variationnel. On se situe ici dans le fil des travaux sur les genres tels qu'ils ont été conduits dans le cadre de la linguistique sur corpus (Biber, 1988 ; Rastier, 2005)

³ On ne s'attardera pas sur les difficultés qui persistent pour l'accès à des données orales suffisamment importantes (en quantité) et diversifiées. De nombreux projets (ESLO, PFC, MPF...) ont modifié le paysage, d'autres (ORFEO) devraient encore faciliter l'accès à des données orales. Mais la lourdeur du travail de transcription et d'édition de corpus oraux rend cette tâche particulièrement lourde.

⁴ Ce concordancier a été développé par Jean Veronis.

⁵ Chaque corpus est identifié par une lettre initiale O ou E qui indique d'emblée s'il s'agit de données orales ou écrites.

⁶ Ces corpus sont extraits du CRFP (*Corpus de Référence du Français Parlé*), Cf. *Recherche sur le français parlé*, n° 18, 2004 et Cappeau 2007.

3. LE TRAITEMENT DE LA FORME *CONTRE*

Nous allons présenter ici quelques faits importants que les données de corpus font ressortir, ce qui nous permettra de vérifier, d'une part, comment ces aspects sont reflétés dans les dictionnaires et, d'autre part, si les descriptions présentées dans ces ouvrages correspondent aux tendances qui se dégagent des corpus.

3.1. *Les données quantitatives brutes*

L'emploi de *contre* est très contrasté selon que l'on s'appuie sur des productions orales ou écrites (Bilger & Cappeau, 2003). Cet écart se manifeste notamment à travers les résultats quantitatifs (cf. le tableau 1) :

ORAL		ECRIT	
O-PRI	O-PUB	E-Presse	E-Litt
81	39	208	167

Tableau 1. Les occurrences de *contre*

La forme *contre* est nettement plus présente dans les corpus écrits. Le corpus E-Presse comporte 5 fois plus d'occurrences de *contre* que le corpus d'oral public (O-PUB). Toutefois, il existe aussi une disparité à l'intérieur de chacun des deux ensembles : à l'oral, par exemple, la parole privée en contient deux fois plus que la parole publique.

Les faits ne se réduisent donc pas à une simple différence oral vs écrit, d'autant plus qu'une autre subdivision peut être tracée entre données orales et données écrites. Dans ces dernières, on ne rencontre quasiment pas d'attestations de *par contre* ; à l'inverse, dans les corpus oraux cette locution adverbiale représente au minimum 40 % des cas (cette proportion atteint même $\frac{3}{4}$ des emplois dans la parole privée).

3.2. *La locution par contre*

On peut penser que l'influence de la norme qui continue à pourchasser l'utilisation de cette expression reste forte dans les textes écrits publiés, qu'ils soient littéraires ou tirés de la presse (Tableau 2). C'est peut-être aussi ce qui explique pourquoi les dictionnaires accordent une place réduite à cette locution et/ou la signalent tardivement dans l'entrée *contre* (Cf., par exemple, le TLF).

ORAL		ECRIT	
O-PRI	O-PUB	E-Presse	E-Litt
61 (75%)	16 (41%)	3 (1,4%)	1 (0,4%)

Tableau 2. Les occurrences de *par contre*

On voit l'importance dans un tel cas de bien identifier la source des données. En effet, un raisonnement sur la totalité du corpus donnerait une proportion de 16% d'emplois de *par contre*, valeur qui serait peu interprétable puisqu'elle ne refléterait aucun type d'usage : ce nombre serait très excessif pour l'écrit et fortement sous-évalué pour l'oral.

La façon dont les dictionnaires présentent les équivalents à la locution *par contre* mérite aussi d'être confrontée aux données de corpus. *Le Grand Robert* propose la liste suivante :

Par contre → **Compensation** (en), **contraire** (au), **mais**, **revanche** (en).

Cette série est présentée dans l'ordre alphabétique qui semble le principe de classement suprême retenu par le dictionnaire. Pour autant, pour le lecteur, l'ordre de succession peut être aussi compris – à tort – comme un ordre d'importance. Le recours au corpus apporte ici des informations fort utiles pour mieux connaître l'emploi de ces équivalents :

- *en compensation* est quasi inexistant dans nos données et apparaît donc comme un très mauvais équivalent. Sa mention en première position est en ce sens peu habile ;
- *au contraire* est la locution la mieux répartie. Elle se retrouve dans tous les sous-corpus (de façon plus importante toutefois dans l'E-Pressé) ;
- *mais* est de très loin la forme la plus utilisée dans tous les corpus (on en trouve ainsi plus de 2 000 occurrences dans l'O-PRI, soit 35 fois plus que *par contre*...). Cet emploi massif est aussi l'indice que *mais* possède des valeurs et des emplois bien plus larges que *par contre* et que cette conjonction peut difficilement être traitée comme un équivalent ;
- *en revanche* : cette locution est souvent présentée dans les grammaires comme un substitut normatif de *par contre*. Cette proposition ne vaut, dans nos données, que pour l'E-Pressé. Dans tous les autres sous-corpus, l'emploi est résiduel.

Cette rapide recherche dans les corpus inciterait à présenter les équivalents d'une façon foncièrement différente :

Par contre → *au contraire* (équivalent le plus neutre), *en revanche* (dans l'écrit de presse), *mais* (dans certains emplois à préciser). *En compensation* est une locution peu utilisée et n'est pas un équivalent bien adapté.

3.3. L'étude du contexte antérieur

La présentation des dictionnaires présente souvent des différences importantes d'un article à l'autre. Ainsi, pour décrire certaines prépositions, par exemple la préposition *en*, le TLF va s'appuyer fortement sur le fonctionnement syntaxique et les catégories (pour le terme complété, seront ainsi successivement signalées comme catégories : un verbe, un substantif, un adjectif). Cette sensibilité grammaticale n'est toutefois pas générale et ne se retrouve pas dans la présentation d'autres prépositions (pour *sur*, l'orientation sémantique est plus prononcée).

Pour ce qui touche à *contre*, *Le Grand Robert* recense des expressions sans s'attarder sur la catégorie (verbale ou nominale) de l'élément constructeur (*pousser le lit contre le mur*, *front contre front*). Le TLF opère, quant à lui, la distinction et présente deux listes :

Verbe + *contre* : *s'acharner*, *aller*, *(s')armer* [...]

Subst. + *contre* : *attaque*, *attentat*, *campagne* [...]

Sur ce point, les corpus apportent deux indications. Le repérage de la tête n'est pas toujours facile, comme l'exemple (1) permet de le constater :

(1) ces tentatives ont suscité une réaction contre les idées universalistes (E-Pressé)

Il est en effet difficile ici de déterminer si *contre* est construit par le verbe *susciter* ou par le nom *réaction*. Cette remarque conduit à s'interroger sur la présentation sous forme de deux listes. Elle est pertinente dans un certain nombre de cas (par exemple pour un nom tel que *colère* qui n'a pas de correspondant verbal). Dans d'autres cas (*s'acharner, attaquer, conspirer...*), on peut se demander s'il ne serait pas plus judicieux de considérer un lexème plus abstrait régissant la préposition *contre* et pouvant se réaliser sous forme de nom ou de verbe. C'est très nettement le cas de *lutte / lutter* dans les corpus.

La distinction nom / verbe serait en revanche être maintenue à un autre niveau (Halliday, 1987) puisque l'on pourrait dégager, à partir des données de nos corpus, la cartographie suivante :

O-Privé et E-Littéraire : plutôt verbe + *contre*

O-Public et E-Presse : plutôt nom + *contre*

Une nouvelle fois, les corpus fournissent des indications complémentaires sur la répartition des formes (Halliday *et al.*, 2004), éclairage dont on peut regretter qu'il ne soit pas mieux pris en compte.

3.4. L'organisation de l'article *contre*

Comme on l'a déjà indiqué en introduction, les dictionnaires apparaissent comme très étroitement liés à la langue écrite littéraire. Cette consanguinité conduit les dictionnaires à ne refléter qu'une partie de la langue, alors que leur titre « de la langue française » laisse attendre une toute autre ambition. La présentation de l'article *contre* confirme notre critique.

L'article du TLF rappelle l'histoire de cette forme qui possède d'abord une valeur d'opposition (les *Serments de Strasbourg* dès 842 contiennent la forme *contra* dérivée du latin), la valeur de contact est plus tardive (le TLF signale *cuntre* en 1100 dans la *Chanson de Roland*). L'entrée *contre* adopte pourtant un ordre inverse puisque la première partie regroupe les emplois locatifs et la deuxième partie la valeur d'opposition.

La motivation du choix de cet ordre n'est pas explicitée. Nos données nous permettent cependant d'avancer une hypothèse : c'est dans le corpus littéraire – et lui seul – que la valeur locative de *contre* domine (avec des verbes comme *appuyer, attirer...*). Dans les autres corpus, la valeur d'opposition est majoritaire. Par exemple dans l'écrit de presse, on trouve : *lutter / lutte contre, crime contre, etc.*

Le dictionnaire apparaît bel et bien dépendant de l'écrit littéraire qui alimente en profondeur la construction de l'article (rubriques et sous-rubriques, hiérarchie des faits présentés, etc.). Ce choix ne serait pas forcément problématique s'il était plus explicite et s'il était enrichi d'informations complémentaires relevant d'autres usages de la langue.

4. LE TRAITEMENT DE LA FORME *MÊME*

Il ne s'agira pas ici de revenir sur la description de cette forme, mais de montrer là encore comment la prise en compte de données diversifiées permet d'interroger la présentation et les commentaires fournis dans les manuels (dictionnaires et/ou grammaires).

4.1. Quelques données quantitatives brutes

Dans les corpus, la forme *même* est apparemment deux fois plus employée à l'oral qu'à l'écrit (cf. le tableau 3).

	ORAL		ECRIT	
	Privé	Public / Pro	Littéraire	Presse
Nombre occurrences	802	885	440	393

Tableau 3. Les occurrences de *même(s)*

Ceci dit, cette première impression est grandement faussée par le poids de la locution *quand même* qui représente entre 30% et 45 % des emplois à l'oral contre moins de 2 % à l'écrit (Tableau 4).

Oral Privé	Oral Public/Pro	Ecrit Littéraire	Ecrit Presse
361 (45%)	272 (30%)	9 (2%)	8 (1,6%)

Tableau 4. Les occurrences de *quand même*

Si l'on écarte cette locution de l'inventaire, les données sont bien plus équilibrées entre corpus oraux et écrits.

4.2. Même adjectif, marqueur de l'écrit ?

La forme *même* peut être variable (adjectif) ou invariable (adverbe) et ce critère morphologique peut déjà révéler des distributions différentes selon les corpus (cf. tableau 5) :

	ORAL		ECRIT	
	Privé	Public / Pro	Littéraire	Presse
<i>même</i> variable	22%	36%	57%	48%
<i>même</i> invariable	78%	64%	43%	52%

Tableau 5. Répartition des 2 emplois de *même*

Ce simple relevé esquisse déjà une certaine singularité du corpus littéraire, puisque c'est le seul qui présente la forme variable comme majoritaire. En ce sens, il n'est pas surprenant que les dictionnaires, et en particulier le TLF qui s'appuie sur *Frantext* (constitué à 80% de textes littéraires), commencent régulièrement leur article sur *même* par la description de l'adjectif, et notamment quand ce dernier est antéposé. Cependant, *même* peut aussi bien être antéposé que postposé, et la prise en compte de ce nouveau critère permet d'affiner les descriptions et de caractériser les types de productions. De fait, nos données révèlent des répartitions (et des collocations) différentes selon les corpus :

– Comme déjà vu, le corpus littéraire se caractérise par un emploi majoritaire de la forme *même* variable mais aussi par une répartition assez équilibrée entre 3 emplois : postposé au pronom (notamment *lui-même*) et au nom (*le jour même*) et antéposé au nom (*le même jour*).

– Les corpus oraux se caractérisent par un emploi minoritaire de la forme variable mais aussi par le fait que les 2/3 des emplois adjectivaux sont antéposés au nom (*la même chose*).

– Le corpus Presse occupe une position intermédiaire : proche du littéraire car la forme variable est bien représentée, mais aussi proche de l’oral par l’emploi massif de *même* antéposé au nom (*du même coup, au même moment, au même titre*).

En ce sens, si une forte présence de *même* variable est sans aucun doute un marqueur de l’écrit, la postposition de cette forme serait un marqueur d’un écrit particulier, l’écrit littéraire.

4.3. La locution quand même, marqueur de l’oral ?

Concernant la forme *même* invariable, et notamment la séquence *quand même*, celle-ci est toujours étiquetée comme relevant du « familier » (cf. le TLF) :

2. **Quand même**, tout de même, loc. adv. concessives. Synon. malgré tout.

a) **Fam.** Quand même. En dépit de tel fait, de telle situation.

Les données montrent cependant que, si cette séquence est bien spécifique de l’oral, elle est utilisée par tous les locuteurs, quelle que soit la situation de parole (privée et publique), et semble stable, produite tout autant en 1968 qu’aujourd’hui, comme le signale Beeching (2005 et 2007). Quoiqu’il en soit, dans les manuels, des équivalents moins marqués comme relevant de l’oral, et notamment de l’oral dit « familier », sont régulièrement proposés ; c’est le cas pour *tout de même*, et dans un registre plus littéraire, pour *néanmoins* et *cependant* (cf. *La grammaire des premiers temps* : 1999).

	ORAL		ECRIT	
	Privé	Public / Pro	Littéraire	Presse
<i>cependant</i>	3	2	47	22
<i>néanmoins</i>	2	5	10	18
<i>tout de même</i>	4	8	5	9
<i>quand même</i>	361	272	8	9

Tableau 6. Des alternatives à *quand même* d’après les manuels

Ces alternatives proposées mériteraient d’être justifiées ou vérifiées. Les données de nos corpus révèlent que *tout de même* est un piètre équivalent, qui est rarement utilisé aussi bien à l’écrit qu’à l’oral et que *cependant* est nettement plus littéraire que *néanmoins*.

4.4. La locution quand (bien) même

La séquence *quand même* versus *quand bien même* est régulièrement signalée dans les ouvrages (*Le Grand Robert, La grammaire des premiers temps* : 1999, Aslanides :

2001) comme moyen d'exprimer la concession adversative, quand le verbe est au conditionnel, comme dans l'exemple suivant produit à l'oral, par un locuteur professionnel (il s'agit en l'occurrence d'un avocat) :

(2) c'est la stricte vérité elle est très simple extrêmement délicate **quand bien même** je serais avocat **quand bien même** je parlerais bien **quand bien même** j'aurais les moyens de me défendre parce que l'on va me dire + peut-être vrai peut-être faux (PUB, GRE07)

C'est cependant un emploi très peu usité, on n'en relève d'ailleurs aucun autre exemple dans l'ensemble de nos corpus. Cette locution occupe pourtant dans les grammaires une place non négligeable, sur laquelle on ne peut être que dubitatif ou réservé, et elle ne s'accompagne dans les dictionnaires d'aucun commentaire particulier.

Dans ces ouvrages, seul sera signalé le fait que l'expression de la concession/hypothèse en français peut s'exprimer par d'autres structures, entre autres, par *même si*. C'est d'ailleurs ce que suggère *La grammaire des premiers temps* qui présente comme équivalent de *quand bien même* la séquence en *même si*, comme dans l'exercice suivant (p. 247):

– Lisez et transformez en utilisant **même si**

Exemple :

Quand bien même je le voudrais, je ne le pourrais pas

Même si je le voulais, je ne le pourrais pas

Il est d'ailleurs intéressant de relever que, dans nos corpus, cette séquence en *même si* est bien représentée, sauf dans l'écrit littéraire où elle est quasi inexistante. En revanche, cela représente 24% des occurrences de *même* forme adverbiale dans la presse (Tableau 7).

	ORAL		ECRIT	
	Privé	Public / Pro	Littéraire	Presse
<i>même si</i>	33	39	2	48

Tableau 7. Nombre d'occurrences de *même si*

Le décalage observé entre la place occupée par cette locution *quand bien même* dans les manuels et celle que révèlent nos données mériterait un certain nombre de développements, tout comme d'ailleurs l'équivalence proposée avec *même si*.

Ces quelques exemples et remarques signalent qu'en définitive les commentaires faits dans les manuels et ouvrages de référence prêtent plus à confusion qu'à description, tout comme d'ailleurs leur absence. Par exemple, *quand même* est bien un marqueur de l'oral, mais pas forcément « familier » ; *quand bien même* pourrait être signalé comme « peu fréquent » et pourquoi pas comme « vieilli »... Par ailleurs, rien n'est explicitement présenté comme emploi spécifiquement littéraire dans ces sources, pourtant il semble bien qu'il en existe. On a vu par exemple que c'est la forme variable qui est la plus représentée dans ce type de productions, alors que dans les autres types, c'est plutôt la forme adverbiale.

De même, les données du corpus révèlent un emploi de *même* qui semble être spécifique de l'écrit littéraire, c'est quand ce dernier est précédé de *pas*, comme dans :

(3) Personne, pas même Mme Grandet, n'avait la permission de venir. (E-LITT)

(4) – et rien ne pouvait l'en empêcher – pas même elle. (E-LITT)

dans lesquels la forme *pas* est très proche de *ni*. Il est alors intéressant de noter que cet emploi, bien représenté dans l'écrit littéraire – cela représente près de 11% des occurrences de la forme *même* adverbiale – est totalement absent des productions orales et très peu usité dans la presse.

L'oral semble plutôt privilégier l'ordre inverse *même pas*, qui produit une séquence à valeur restrictive, plutôt que l'ordre *pas même* :

(5) je me rappelle de rien même pas où j'étais (O-PRI)

De fait, la prise en compte de l'ensemble des données semble révéler que la littérature classique pourrait se caractériser par la postposition de la forme *même* (variable ou pas) : *lui-même*, *le jour même*, *pas même*, alors que la langue contemporaine (oral et Presse) aurait tendance à privilégier l'antéposition : *même lui*, *le même jour*, *même pas*, *même si*...

CONCLUSION

L'exploitation de corpus diversifiés et échantillonnés permet de jeter un éclairage neuf sur de nombreux faits de langue et d'en renouveler à la fois l'analyse et la présentation. Cependant, aujourd'hui encore, les travaux de ce type sont peu nombreux, car fortement dépendants des données disponibles.

En l'absence d'un grand corpus de référence du français (dans le sens anglo-saxon du terme), sur lequel pourraient s'appuyer de nouvelles études à la fois qualitatives et quantitatives, il va de soi que les indications, les commentaires (ou leur absence) relevés dans les manuels continueront à prêter souvent le flanc à des critiques, tant ils méritent d'être affinés ou actualisés. Comme nous avons essayé de le démontrer, la prise en compte des différents types de productions présente plusieurs avantages, entre autres, ceux d'éviter de présenter la langue littéraire comme l'usage général, de mieux cerner le lien qui existe entre certains emplois et les sources, de mieux gérer certains traits attribués aux formes (*familier*, *rare*, etc.), de hiérarchiser au mieux les équivalents.

ÉLÉMENTS BIBLIOGRAPHIQUES

Sources primaires

- ARBACH, Najib, 2015. *Constitution d'un corpus oral de FLE. Enjeux théoriques et méthodologiques*, Doctorat, Université Rennes 2.
- BEECHING, Kate, 2005. "Politeness-induced semantic change: the case of *quand même*", *Language Variation and Change*, 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 ? », *Langue Française*, 154, 78-93.
- BIBER, Douglas, 1988. *Variation accross speech and writing*, Cambridge, Cambridge University Press.
- BIBER, Douglas, JOHANSSON, Stig, LEECH, Geoffrey, CONRAD, Susan & FINEGAN, Edward, 1999. *Longman Grammar of Spoken and Written English*, London & New-York, Longman.
- BILGER, Mireille & CAPPEAU, Paul, 2003. « Les emplois de "contre" dans les corpus de français parlé et de presse écrite », *Recherches linguistiques*, 26, 91-111.

- BLANCHE-BENVENISTE, Claire, 2010. *Le Français. Usages de la langue parlée*, Leuven & Paris, Peeters.
- CAPPEAU, Paul, 2007. « Constituer et exploiter un gros corpus oral : l'exemple du CRFP », in Ballard, Michel & Pineira-Tresmontant, Carmen (éd.), *Les corpus en linguistique et en traductologie*, Arras, Artois Presses Université, 145-158.
- CORI, Marcel et DAVID, Sophie, 2008. « Les corpus fondent-ils une nouvelle linguistique ? », *Langages*, 171, 111-129.
- HALLIDAY, Michael A.K., 1987. "Spoken and Written Modes of Meaning", in Horowitz, Rosalind & Samuels, S. Jay (ed.), *Comprehending oral and written language*, New York, Academic Press.
- HALLIDAY, Michael A.K., TEUBERT, Wolfgang, YALLOP, Colin & CERMÁKOVÁ, Anna, 2004. *Lexicology and Corpus Linguistics. An introduction*, London, Continuum.
- RASTIER, François, 2005. « Enjeux épistémologiques de la linguistique de corpus », in Rey, Alain (dir.) (2013), *Le Grand Robert de la langue française*, Paris, Dictionnaires *Le Robert*.
- TOGNINI-BONELLI, Elena, 2001. *Corpus linguistics at work*, Amsterdam, Benjamins.

Bibliographie secondaire

- BRUNOT, Ferdinand & BRUNEAU, Charles, 1956 [1949]. *Précis de grammaire historique*, Paris, Masson.
- BULTON, Alex & TYNE, Henry, 2014. *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*, Paris, Didier.
- GREVISSE, Maurice & GOOSSE, André, 2008. *Le bon usage*, 14^e édition, Bruxelles, De Boeck.
- LABOV, William, 2001. « Qu'est-ce qu'un fait linguistique ? », *Marges linguistiques*, 1, 1-44.
- LAKS, Bernard. 2010. « La linguistique des usages : de l'*exemplum* au *datum* », *Travaux linguistiques du CerLiCO*, 23, 11-26.
- RIEGEL, Martin, PELLAT, Jean-Christophe & RIOUL, René, 1995. *Grammaire méthodique du français*, Paris, PUF (éd. 2008).
- SINCLAIR, John, 1991. *Corpus Concordance Collocation*, Oxford, Oxford University Press.
- WILLIAMS, Geoffrey (éd.), 2005. *La linguistique de corpus*, Rennes, PUR, 31-46.

Dictionnaires et manuels

- ABRY, Dominique & CHALARON, Marie-Laure, 2004. *La grammaire des premiers temps*, vol. 1. et 2, Grenoble, Ellug.
- ASLANIDES, Sophie, 2001. *Grammaire du français. Du mot au texte*, Paris, Champion.
- REY, Alain (dir.), 2009. *Le Grand Robert*, Paris, Le Robert.
- Trésor de la Langue Française*. Accès en ligne : <http://atilf.atilf.fr/>