

L'apport des corpus échantillonnés aux descriptions grammaticales. L'exemple des formes "contre" et "entre

Mireille Bilger, Paul Cappeau

► **To cite this version:**

Mireille Bilger, Paul Cappeau. L'apport des corpus échantillonnés aux descriptions grammaticales. L'exemple des formes "contre" et "entre. *Linguistik Online*, Bern Open Publishing, 2016, 78 (4), <10.13092/lo.78.2948>. <hal-01394073>

HAL Id: hal-01394073

<https://hal-univ-perp.archives-ouvertes.fr/hal-01394073>

Submitted on 14 Nov 2016

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.

L'apport des corpus échantillonnés aux descriptions grammaticales. L'exemple des formes *contre* et *entre*

Mireille Bilger (Perpignan) et Paul Cappeau (Poitiers)

Abstract

The following article sets out to show how different kinds of sampled corpus data can be used to provide a critical view of certain descriptions given in reference works (grammars and dictionaries) which, for French, tend to concentrate on literary usage. We show how quantitative and qualitative methods can be proposed to nuance the descriptions, using the example of the forms *contre* and *entre* in four different corpora: private speech, public speech, literature, press.

1 Introduction

Depuis plus de 15 ans, de nombreux travaux linguistiques prennent appui sur des corpus afin de parfaire les analyses développées. Reste que la question de la représentativité des données constituées pourrait parfois être interrogée. On est loin de l'orientation adoptée dans Biber et al. (1999) où les distributions sont différenciées selon divers « registres » disponibles. Dans le domaine français, il faut bien reconnaître que le caractère modélisant de la langue littéraire ne s'affaiblit que modérément depuis le 17^{ème} siècle. Cela est si vrai que de nombreux outils disponibles (comme le *Trésor de la Langue Française* ou des manuels de grammaire) ne représentent et décrivent principalement que l'usage littéraire¹. Or, en ce qui concerne la description syntaxique, le recours à des corpus de divers types (oral, écrit, presse, littérature, etc.) et de genres différents (explications techniques, récits de vie, etc.) se révèle primordial si l'on souhaite proposer une description moins parcellaire et plus sensible à la diversité des usages. L'objet de cet article est de montrer comment les données de corpus échantillonnés conduisent à adopter un regard critique sur certaines descriptions proposées dans des ouvrages de référence (grammaires, dictionnaires ou manuels), et à indiquer quels aménagements pourraient être apportés (Hunston 2002). Pour ce faire, nous nous appuyons sur l'étude des formes *contre* et *entre*.

¹ D'autres usages (presse, langue technique...) peuvent parfois être signalés, souvent sous forme de remarques.

2 Présentation des données exploitées

2.1 Les corpus utilisés

Nos commentaires des présentations disponibles dans les ouvrages de référence s'appuieront sur un certain nombre d'observations tirées de corpus. Certes à l'heure actuelle, cette approche n'a rien de bien original, si ce n'est que nos données seront diversifiées et feront intervenir pour chaque médium (oral et écrit) deux types de productions assez nettement différenciées. Nous aurons ainsi à notre disposition plus d'un million de mots répartis équitablement dans les quatre sous-ensembles suivants :

- pour l'oral, nous aurons à la fois de la parole privée : conversations, entretiens, etc. (O-PRI) et de la parole publique : enregistrement dans les médias (libre antenne, débats, ...), dans des réunions, etc. (O-PUB) ; ces corpus sont extraits, pour partie, du Corpus de Référence du Français Parlé² (CRFP, DELIC 2004).
- pour l'écrit, nous disposerons de textes de presse : journaux ou magazines d'actualité comme *Le Monde*, *le Nouvel Observateur*, *Libération* (E-PRESSE) et d'extraits de romans du 19^{ème} siècle, Balzac, Flaubert, Zola (E-LITT).

Nous disposons ainsi de trois corpus de langue contemporaine et d'un corpus de littérature dont la langue est certes datée, mais qui présente un indéniable intérêt dans la mesure où celle-ci a longtemps constitué le terreau des exemples et des descriptions développées dans les grammaires et les dictionnaires. Le concordancier Contextes (développé par Jean Véronis) a permis d'extraire les exemples requis.

Ce choix nous inscrit dans la problématique des genres en linguistique de corpus (Biber 1988 ou Rastier 2005). On montrera d'ailleurs rapidement l'incidence de ce choix de données variées sur les résultats.

2.2 Les formes retenues

Nous appuierons nos observations sur l'analyse de deux formes : *contre* et *entre*. Il ne s'agit pas, dans ces pages, d'en fournir une description détaillée³. Ces deux formes nous ont paru se prêter à notre démonstration car elles sont relativement bien représentées dans nos données (même si comme on le verra par la suite, cette première indication demande à être nuancée ou affinée) :

	O-PRI	O-PUB	E-PRESSE	E-LITT
<i>contre</i>	81	39	208	167
<i>entre</i>	101	90	302	219

Tableau 1 : Nombre d'occurrences de *contre* et *entre*

² Cf. *Recherche sur le français parlé*, n° 18, 2004 et Cappeau (2007) pour une présentation plus détaillée de ce corpus et de la distinction parole privée/parole publique.

³ Pour une telle description, voir Bilger et Cappeau (2003), Bilger et Cappeau (2013).

Globalement, la forme *entre* est plus attestée que la forme *contre* et c'est dans les productions écrites (presse ou littérature) qu'on les retrouve le plus⁴.

3 Des déséquilibres dans la présentation des emplois

La thèse que nous défendons dans cet article consiste à pointer l'homogénéité ou la similitude des données utilisées dans les ouvrages de référence. On peut reprendre les critiques de Habert (2000) qui soulignait le danger de ne travailler que sur un seul type de données. Si par exemple, on décide de développer la description d'une forme grammaticale en n'utilisant que des articles d'un seul journal, on obtiendra une description certes intéressante mais représentative de cette seule source. Peu importe alors que l'on décide d'augmenter la taille du corpus : on risque de retrouver sans cesse les mêmes emplois. Les ouvrages de référence s'appuient de façon souvent implicite sur les textes littéraires dans lesquels ils puisent leurs exemples mais aussi – et c'est plus gênant car cette dimension n'est pas revendiquée – ils organisent la présentation des formes syntaxiques selon ces mêmes données.

Pour ce qui touche aux emplois, ce biais descriptif peut avoir deux incidences : certains usages seront soit mis en avant et occuperont une place usurpée, soit occultés ou minorés.

3.1 Emplois mis en avant

Dans la *Nouvelle Grammaire du Français* (Delatour et al. 2004), quelques pages sont consacrées à l'examen des prépositions. Celles-ci sont présentées sous forme de couples, probablement dans le but d'attirer l'attention des apprenants sur leur proximité de forme ou d'emploi (*dans* et *à*, *avant* et *devant*, etc.). La préposition *entre* est contrastée avec *parmi*. Les données quantitatives tirées des corpus laissent penser que ce parallélisme devrait être commenté de façon plus détaillée.

	O-PRI	O-PUB	E-PRESSE	E-LITT
<i>entre</i>	101	90	302	219
<i>parmi</i>	9	12	46	82

Tableau 2 : Nombre d'occurrences de *entre* et *parmi*

Il existe dans nos corpus un écart important entre ces deux prépositions : *entre* domine largement et *parmi* est bien plus présent dans la langue écrite (surtout dans la langue littéraire). Voilà deux informations (le décalage quantitatif dans l'emploi de ces deux formes et la spécialisation de *parmi* pour l'écrit littéraire) qu'il aurait été judicieux d'apporter aux apprenants de français. La proximité entre les deux prépositions, induite par la présentation, résiste mal à l'observation des usages. Une autre distorsion apparaît dans le choix des exemples, *parmi* y figurant chaque fois en position initiale :

- 1) Parmi toutes les fleurs sauvages, ce sont les coquelicots que je préfère. (Delatour et al. 2004: 165)

⁴ Ces résultats sont confortés par les calculs de fréquence portant sur des corpus plus volumineux comme ceux effectués par Brunet (1999) pour l'écrit littéraire et Véronis (2004) pour l'oral. Concernant la catégorie des prépositions dans les textes littéraires, ces analyses signalent qu'*entre* arrive à la 12^{ème} place, *contre* à la 16^{ème}. Dans les productions orales, *entre* est une fois et demi plus fréquent que *contre*. Nos données révèlent que ces deux formes sont nettement moins utilisées à l'oral qu'à l'écrit, quasi trois fois moins.

- 2) Parmi les spectateurs, on remarquait le Premier ministre. (ibid.).

Or cet emploi initial est marginal dans l'écrit de presse (4 exemples, soit moins de 10 %) et faible dans l'écrit littéraire (14 exemples, soit 17 %). La présentation introduit des biais qui ne manquent pas d'intriguer.

Un autre emploi (*tout contre*) est mis en avant, ce qui peut sembler surprenant. Le TLFi⁵ signale très rapidement cette locution (dans le paragraphe I.A). Le *Grand Robert* (Rey 2013) effectue le même choix. Dans nos données, cette tournure est absente de l'oral (privé comme public) et de l'écrit de presse. Dans l'écrit littéraire (3) elle reste peu utilisée (3 occurrences). Le tropisme en faveur de l'écrit littéraire semble avoir favorisé le choix de cette locution :

- 3) A force de rompre, le Parisien se sentit acculé tout contre le rempart (E-LITT).

A l'inverse, on ne trouve pas de trace de *rien contre* (4) qui est pourtant autant attesté. On peut subodorer que sa présence dans l'oral privé n'est pas indifférente à cet « oubli » :

- 4) le technique j'avais rien contre + et je voulais euh partir euh partir en pension (O-PRI)

3.2 Emplois minorés

L'observation de l'article *contre* dans les dictionnaires illustre pleinement le déséquilibre dans la présentation que la langue littéraire provoque. On sait que la forme *par contre* est fortement critiquée par la norme. Le TLFi relègue cette locution en toute fin d'article (rubrique IV. B) avec une description relativement neutre (*Loc. adv. marquant l'opposition à un énoncé antérieur*) et la position des puristes n'est signalée et critiquée que dans la remarque 2⁶. Dans le *Grand Robert*, c'est dans la partie I.B.1 que figure cette locution et un encadré fait état des réserves que suscite cette forme auprès de « certains pédagogues puristes ». Bref, ces deux ouvrages ne semblent pas endosser les réserves normatives. Se pose toutefois la question de la place relativement discrète qu'y occupe cette locution. Les données de corpus (Tableau 3) rendent cette question légitime :

O-PRI	O-PUB	E-PRESSE	E-LITT
61 (75 %)	16 (41 %)	3 (1,4 %)	1 (0,4 %)

Tableau 3 : Nombre d'occurrences et part des emplois de *par contre*

Ce tableau constitue une clé pour comprendre l'organisation des articles de dictionnaire. La locution *par contre* est résiduelle à l'écrit – quasiment absente de la littérature (5), très peu attestée dans la presse (6)⁷ –, c'est probablement cette rareté qui conduit les dictionnaires à traiter à la marge cet emploi. On peut facilement imaginer qu'une description du français qui s'appuierait plus fortement sur la langue parlée chamboulerait l'organisation de cet article. Les emplois de *par contre* sont très largement majoritaires dans l'oral privé (7), ce qui inciterait à

⁵ Tous les renvois se feront à la version en ligne (informatisée) du *Trésor de la Langue Française* disponible à l'adresse suivante : <http://atilf.atilf.fr/>.

⁶ « 2. Sur l'offensive des puristes à propos de la loc. *par contre*, on verra A. DOPPAGNE, *Trois aspects du fr. contemp.* Paris, Larousse, 1966, pp. 193–197 : « le succès qu'ont réservé à *par contre* la plupart des écrivains du XXe siècle, le fait qu'il ne soit pas toujours remplaçable par les locutions par lesquelles on propose de le remplacer, légitiment tout à fait l'utilisation de cette locution. » (TLFi)

⁷ Il faut d'ailleurs noter dans ce cas un emploi en incise, alors que l'emploi le plus répandu semble être celui en position initiale d'une construction verbale.

leur accorder une place de choix dans la présentation. L'oral public (8) en fait un usage important, même si l'on peut imaginer que la norme (pourtant si peu défendue dans les ouvrages consultés) exerce une influence non négligeable.

- 5) Bien des noms y étaient effacés, sans doute les noms de ceux qu'il avait pu joindre. Par contre, il y avait deux noms nouveaux qui remplissaient les blancs. (E-LITT)
- 6) Le Venezuela, par contre, risque une aggravation de ses problèmes (E-PRESSE)
- 7) je ne peux pas dire que j'ai peur de la mort non par contre j'ai peur de souffrir (O-PRI)
- 8) il y a des choses intéressantes par contre rapport à l'environnement + je pense que + on on on peut ici se poser beaucoup de questions (O-PUB)

4 Des liens avec le lexique sous-exploités

Le travail sur corpus a conduit à revoir la frontière entre syntaxe et lexique (Sinclair 1991). Les faits de collocation (Legallois 2012) prennent d'ailleurs une importance grandissante dans les analyses contemporaines. Des associations privilégiées, différentes selon les productions, peuvent être relevées.

Pour *contre*, deux séquences se dégagent nettement et apparaissent comme très spécifiques des productions :

1. Dans l'écrit de presse, on trouve *lutte* ou *lutter contre* (9), soit la valeur d'opposition ;
2. Dans le corpus de textes littéraires, c'est *appuyer contre* qui ressort (10), et dans ce cas c'est donc la valeur locative qui domine :

- 9) Ils servent maintenant à lutter contre les contrebandiers. (E-PRESSE.)
- 10) La princesse les prit à pleines mains et les appuya contre ses lèvres en fermant les yeux. (E-LITT.)

Sans surprise, le TLF signale bien le verbe *s'appuyer* mais omet le verbe *lutter*. En cela il ne reflète que l'écrit littéraire. Le *Grand Robert* signale bien les deux séquences mais sans préciser dans quels types de productions elles sont concentrées.

La préposition *entre* apparaît majoritairement avec des verbes tels que *discuter* ou avec des noms tels que *différence(s)*, *discussion(s)*, *relation(s)* dans l'oral privé (11) et dans la presse (12). Aucune de ces séquences n'est signalée dans le TLFi. Le *Grand Robert* ne recense que *différence entre*. Une nouvelle fois, ce qui ne se rencontre pas dans la littérature semble ne pas (vraiment) exister.

- 11) avant les voisins ils pouvaient discuter entre eux (O-PRI)
- 12) Depuis 1989 et la loi du 6 juillet régissant les relations entre propriétaires et locataires (E-PRESSE)

5 Des distinctions syntaxiques négligées

S'il est établi que les dictionnaires ne sont pas des grammaires, on peut s'étonner tout de même que certaines informations grammaticales éclairantes ne soient pas (mieux) utilisées quand elles pourraient aider à l'emploi d'une forme. Ainsi *entre* possède une syntaxe particulière qui impose que le complément qui suit soit un pluriel (13) ou un syntagme de type *A et B* (14). C'est pourquoi on ne trouvera pas un nom ou un pronom renvoyant à un singulier (15), ni même un nom collectif (16). Le *Grand Robert* n'apporte cette précision que de façon fragmentaire en indiquant (paragraphe II. 4) que *entre* peut être suivi d'« un pronom personnel ou d'un nom de

personne au pluriel, sans article ». Dans le TLF, c'est de façon implicite (avec des formules telles que : *à l'intérieur de deux limites, délimité par deux éléments, une série d'éléments, ...*) qu'elle est évoquée.

- 13) les élèves ont l'habitude de se retrouver entre amis (O-PUB)
- 14) il y a un lien comme dit Bourdieu + entre + la position + et la prise de position (O-PUB)
- 15) * se méfier des interactions qu'ils peuvent avoir entre moi
- 16) * quelles sont les différences entre la police

Ces deux remarques grammaticales sont plus intéressantes qu'il n'y paraît dans la mesure où elles vont pouvoir caractériser certains types de productions. Nos données révèlent par exemple que les syntagmes en *A et B* sont nettement plus fréquents dans la presse ; cela correspond à 66% des emplois de *entre*. De même, cette forme introduit un nom pluriel sans prédéterminant (17) ou une séquence en *A et B* (18) également sans prédéterminant, mais essentiellement dans le corpus presse (9%) et dans le corpus de parole publique (7%).

- 17) les conversations entre diplomates (E-PRESSE)
- 18) le dialogue entre psychiatrie et univers pénitentiaire est à réinventer (E-PRESSE)

Concernant les pronoms qui figurent après la préposition, le *Grand Robert* signale *soi* auquel est attribuée une valeur littéraire. Dans nos données, aucune occurrence de *entre soi* n'est apparue. Les formes liées à l'interaction (*nous/vous*) sont quasiment absentes de la presse, ce qui est assez prévisible, dans la mesure où elles sont réservées aux dialogues. Ces séquences *entre nous* et *entre vous* sont en revanche nombreuses en littérature. Mais c'est en fait la séquence *entre eux* (19) qui est la mieux répartie et la plus utilisée ; en presse, c'est plutôt *d'entre eux* (20) que l'on retrouve. Comme l'indiquent les dictionnaires, cette suite se réalise en prolongement d'un quantifieur (*plusieurs, certains, ...*) :

- 19) les gens euh entre eux ils parlaient le français ou le patois (O-PRI)
- 20) Trois d'entre eux sont aujourd'hui réfugiés à l'étranger (E-Presses)

S'interroger sur l'élément qui construit la préposition peut être intéressant. Ainsi, pour *contre*, sous la rubrique SYNT. (pour syntaxe), le TLFi fournit une liste des verbes et des substantifs les plus fréquents. Pour « l'idée d'hostilité » ou l'argumentation, sont signalés : *s'acharner, se battre, combattre, se révolter, attaque, combat, révolte*, etc. Ces listes sont intéressantes en ce qu'elles pointent des régularités. Trois précisions auraient peut-être pu être apportées : d'une part, comme on l'a signalé en [3], certains lexèmes sont plus étroitement associés à un type de production (*lutter* se rencontre plutôt dans la presse) ; d'autre part, la répartition entre tête verbale et tête nominale est en partie sensible à la variation de médium (les verbes sont plus présents à l'oral, les noms à l'écrit) ce qui n'est pas indiqué ; enfin peut-être serait-il plus économique d'indiquer que dans de nombreux cas, on pourrait se contenter de poser qu'une racine construit la préposition en question (et que cette racine peut se réaliser sous forme de verbe ou de nom). Par exemple, dans l'écrit littéraire la racine [défen-] se rencontre comme nom *défense* (21) ou verbe *défendre* (22). Il pourrait alors être intéressant de s'interroger sur la présentation la plus adéquate :

- 21) Madame Grandet fut sans défense contre les harmonies de la voix de sa fille. (E-LITT)
- 22) et si je ne sais pas défendre la fille de Nevers contre ce danger, comme je l'ai défendue contre tous les autres (E-LITT)

6 Une hiérarchisation à interroger

Si la présentation des articles du dictionnaire reproduit les emplois les plus fréquents dans l'écrit littéraire, il s'agit d'un choix qui n'est pas assumé comme tel mais qui est présenté comme une sorte d'évidence. Les dictionnaires se veulent d'ailleurs des dictionnaires de langue (et non de la langue littéraire).

Par exemple, pour *contre*, on trouve en première position la valeur locative (*se serrer contre quelqu'un*) et dans une deuxième partie la valeur d'opposition (*agir contre*). La présentation aurait été inversée si l'on avait pris appui sur un corpus de presse.

La même observation vaut pour *entre*. La valeur locative qui est la première décrite dans le TLFi ou dans le *Grand Robert* est celle que le corpus littéraire développe le plus. C'est cette valeur qui est également signalée dans les ouvrages concernant l'enseignement du FLE (cf. *La grammaire des premiers temps*, vol. 1, 2004). Il est dommage cependant que, dans ces ouvrages, rien ne soit dit sur un emploi somme toute bien représenté dans les textes littéraires, celui dans lequel *entre* introduit des éléments en relation avec le corps (*entre les deux sourcils, les deux yeux, les mains, les bras, dents, doigts, lèvres, jambes*) puisque cela correspond à plus de 20% des emplois de notre corpus littéraire (23)–(24) :

- 23) Touché en plein front, entre les deux yeux, et trois fois de suite... (E-LITT)
- 24) Au lieu de me répondre, il me planta sa rapière entre les deux sourcils (E-LITT)

En revanche, les valeurs sémantiques de cette forme sont plus variées dans le corpus oral et le corpus de presse. L'emploi locatif y est nettement moins représenté que la valeur temporelle, (comme dans *entre le mois de juillet et le mois de septembre*) notamment dans la presse, où il correspond à près de 10% des occurrences.

7 Des équivalents discutables

Le dictionnaire, et c'est là une de ses caractéristiques, propose régulièrement des synonymes et des antonymes qui aident le lecteur à cerner le sens des unités. Là encore, cette information précieuse pourrait être améliorée par l'apport de données diversifiées. On prendra un exemple simple concernant la locution *par contre*. On a vu que celle-ci, même si le discours normatif était atténué, restait peu mise en valeur par les dictionnaires. Il n'est pas étonnant alors que le *Grand Robert* fournisse une liste étoffée de substituts : *compensation (en), contraire (au), mais, revanche (en)*. Cette suite est présentée dans l'ordre alphabétique, principe organisateur du dictionnaire, encore qu'en l'occurrence on puisse douter de sa pertinence. La recherche dans nos corpus (Tableau 4) des formes indiquées fournit des suggestions pour retoucher cette présentation.

	O-PRI	O-PUB	E-PRESSE	E-LITT
<i>par contre</i>	61	16	3	1
<i>en compensation</i>	0	0	1	0
<i>au contraire</i>	6	12	31	13
<i>mais</i>	2 135	444	927	1 193
<i>en revanche</i>	1	1	33	2

Tableau 4 : Nombre d'occurrences de *par contre* et de ses équivalents

Du seul point de vue quantitatif (mais il est certain que des analyses linguistiques – en particulier syntaxiques – plus poussées seraient bienvenues), les résultats obtenus conduisent à chambouler la présentation des équivalents. On a choisi de laisser de côté *mais* qui est la forme la plus utilisée dans toutes les productions (même si d'importantes fluctuations sont perceptibles). Il faudrait se livrer à un examen plus attentif des emplois et des valeurs de cette conjonction. Les autres formes permettent des commentaires en lien avec les productions. Ainsi, il n'est pas très judicieux de commencer par *en compensation* dont l'emploi est extrêmement limité et qui possède parfois un sens bien différent de *par contre* (25). Mettre cette locution en première position induit le lecteur en erreur. Pour *en revanche*, il faudrait accompagner cette forme d'une spécification : elle se rencontre abondamment dans le corpus de presse (26) mais quasiment pas ailleurs (27). Enfin *au contraire* apparaît, après *mais* probablement, comme l'équivalent le plus équilibré : présent dans toutes les productions, plus utilisé à l'écrit (28), bien représenté dans la presse (29)⁸ :

25) Celles-ci, en compensation, purent décrocher d'autres marchés dans le département (E-PRESSE)

26) Rien, en revanche, n'empêche de mettre en place une procédure informelle de préparation des révisions (E-PRESSE)

27) c'est très facile d'arriver et de semer la pagaille + alors qu'en en revanche si on les cadre et qu'on les implique pour le nettoyage et tout (O-PRI)

28) Maître Louis sortait au contraire fort souvent, et rentrait même assez tard dans la nuit. (E-LITT)

29) Lorsque Bouillon de Culture a lancé ses invitations, avant les fêtes de Noël, aucune polémique n'a encore éclaté autour de son livre Une aveuglante absence de lumière. Au contraire. Dans l'imagerie médiatique, cet auteur marocain, installé à Paris depuis 1971, est la figure même du « gentil contre les méchants » (E-PRESSE)

En l'absence de telles indications, la liste proposée dans le dictionnaire apparaît comme factice, artificielle et d'un intérêt très limité. Les termes censés être des équivalents à *par contre* sont, pour plusieurs d'entre eux, de très mauvais candidats pour remplir cette fonction⁹.

8 Des variations liées aux productions délaissées

Comme déjà signalé, chacune de ces deux formes est beaucoup plus utilisée à l'écrit (notamment dans la presse) qu'à l'oral (cf. le tableau 1). Le recours à des corpus diversifiés (médium et type de production) et échantillonnés a permis de détecter des emplois partagés quasi équitablement dans les différents corpus et d'autres emplois plus spécifiques, reliés à un type de production.

Ainsi pour *contre* (Tableau 5), si l'on s'en tient aux contextes les plus fréquents, on voit se dégager de nombreux contextes spécifiques et peu de contextes communs à l'ensemble des productions :

⁸ Comme on le constate dans cet exemple, une étude plus précise des fonctionnements des unités listées permettrait de mieux tracer les différences entre elles. Ainsi, *au contraire* dans (27) possède un emploi macrosyntaxique de noyau que l'on n'a pas trouvé pour *en revanche* ou *par contre*. Sur l'analyse de type macrosyntaxique on renvoie à Blanche-Benveniste (2010) et Debaisieux (2013).

⁹ On a voulu mettre ici en avant l'argument de la fréquence mais il va de soi que d'autres différences d'ordre sémantique ou syntaxique pourraient aussi être invoquées.

	O-PRI	O-PUB	E-PRESSE	E-LITT
contexte Gauche	par contre		lutte/lutter contre ci-contre	porté contre appuyé contre attiré contre
	rien contre	deux contre un		
contexte Droit	contre + article défini			contre moi contre sa / son

Tableau 5 : Les contextes récurrents avec *contre*

On ne reviendra pas sur les structures déjà commentées auparavant. Certains contextes sont peu exploitables en l'état (comme *contre* + article défini). Notons simplement que quelques tournures (en gras dans le tableau) ne sont pas signalées dans le TLF. Ainsi pour *contre* entre deux nombres, utilisé notamment dans les commentaires sportifs (30), le TLF recense une structure légèrement différente dans laquelle un nom est présent (31). Quant à *contre moi* (32), le TLF n'en fait aucune mention. Remarquons aussi un emploi en quelque sorte technique (*ci-contre*) que l'on trouve uniquement dans la presse pour renvoyer à des informations présentes dans la page (33) :

- 30) parce qu'au départ de l'action + il y avait pratiquement deux contre un (O-PUB)
- 31) Nous opposons à l'ennemi un homme contre trois. (TLF)
- 32) Faites ce que vous voudrez contre moi, je n'ai plus peur de vous. (E-LITT)
- 33) Alors que si les Etats-Unis caracolent en tête du tableau des médailles, c'est uniquement grâce à leurs athlètes et nageurs (voir tableau ci-contre). (E-PRESSE)

Pour *entre*, on peut proposer la synthèse suivante (Tableau 6). Les séquences en gras n'ont pas été rencontrées dans l'article :

	O-PRI	O-PUB	E-PRESSE	E-LITT
contexte Gauche	<i>différence entre</i> <i>lien entre</i>	d'entre... <i>différence entre</i> <i>communication</i> <i>entre</i>	d'entre... <i>rappports entre</i> <i>relations entre</i>	d'entre... <i>tête entre</i>
contexte Droit	entre <i>eux</i>			
	<i>entre guillemets</i> entre <i>autres</i> entre-temps (entretemps)		entre-temps (entretemps)	entre <i>ses</i> entre (<i>les</i>) <i>deux</i>

Tableau 6 : Les contextes récurrents avec *entre*

La situation présente des différences avec celle observée pour *contre*. Ainsi, le médium peut être déterminant comme pour la séquence *entre guillemets* qui, sans surprise, ne se rencontre qu'à l'oral (34)–(35). Il en est de même pour *entre autres* (36) nettement plus présent à l'oral et quasi absent du corpus littéraire (37). Pour *entre-temps* (*entretemps*), il se réalise dans l'oral privé (38) et dans l'écrit de presse (39). Quant à *entre nous* que le TLFi étiquette comme familier quand il peut être glosé par *franchement*, on n'en relève que deux occurrences dans le corpus d'oral privé (40). Le corpus ici apporte un contrepoint intéressant aux indications du dictionnaire.

- 34) et donc là j'ai commencé à secouer donc mon oncle entre guillemets quoi à lui parler (O-PRI)
- 35) tous ces problèmes-là sont + entre guillemets normaux (O-PUB)
- 36) plusieurs critères qui rentrent en compte + et entre autres deux critères essentiels (O-PRI)

- 37) mais les escarmouches entre partisans continuaient aux frontières, et la vallée de Louron, entre autres, avait bon nombre de ces hôtes incommodes (E-LITT)
 38) parce que je + entre-temps nous avons rencontré à Mopti euh les gens du coin (O-PRI)
 39) Entre-temps, la crise économique est passée par là, marquée par une dette extérieure gigantesque (E-PRESSE)
 40) entre nous c'est vrai que bon ils ont tendance à plus me respecter (O-PRI)

Le tableau confirme aussi que les associations de *entre* avec du lexique (question déjà évoquée en [3]) sont assez spécifiques à chaque type de production.

9 Conclusion

L'objectif principal de cet article était d'illustrer, à partir de ces deux formes (qui mériteraient chacune une étude plus approfondie), l'intérêt de travailler sur des productions diversifiées. En prenant appui sur des corpus échantillonnés, il est possible de fournir des indications à la fois quantitatives et qualitatives qui permettent de donner aux phénomènes étudiés une représentation plus juste et plus complète. De fait, la linguistique sur corpus a permis de remettre en jeu la conception même de la grammaire : il n'y aurait pas d'un côté un « système abstrait » et de l'autre des « réalisations individuelles », mais bien comme le dit Halliday (1991), un système linguistique composé du cumul de l'ensemble des usages. S'il est indéniable que la prise en compte de la diversité des usages permet de peaufiner, voire de modifier, la description de certains faits de langue, il n'en demeure pas moins vrai que les outils pédagogiques sont encore lents à tirer profit des recherches et des résultats obtenus (Debaisieux/Boulton 2008; Boulton/Tyne 2014). Soit les manuels (dictionnaires et grammaires) favorisent encore trop souvent les seules données littéraires et donnent à cet usage particulier un statut quasi « systémique », soit ils envisagent une simple opposition écrit/oral qui s'avère bien trop réductrice pour représenter la variété des usages. L'appui sur corpus diversifiés permettrait, nous semble-t-il, de remédier judicieusement à cette situation.

Bibliographie

- Abry, Dominique/Chalargon, Marie-Laure (2004): *La grammaire des premiers temps*. Vol. 1. Grenoble: Ellug.
- Biber, Douglas (1988): *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas et al. (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.
- Bilger, Mireille/Cappeau, Paul (2013): « Comment les données de corpus pourraient renouveler les manuels de grammaire ? Illustrations à partir de quelques pronoms et de la forme même ». *LINX* 68/69: 177–200.
- Blanche-Benveniste, Claire/Martin, Philippe (2010): *Le Français. Usages de la langue parlée*. Leuven/Paris: Peeters.
- Boulton, Alex/Tyne, Henry (2014): *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*. Paris: Didier.
- Brunet, Etienne (1999): *Liste de fréquence lexicale*. <http://eduscol.education.fr/cid50486/liste-de-frequence-lexicale.html> [21.07.2016].

- Cappeau, Paul (2007): « Constituer et exploiter un gros corpus oral : l'exemple du CRFP ». In : Ballard, Michel/Pineira-Tresmontant, Carmen (eds.): *Les corpus en linguistique et en traductologie*. Arras, Artois Presses Université: 145–158.
- Debaisieux, Jeanne-Marie (ed.) (2013): *Analyses linguistiques sur corpus. Subordination et insubordination en français*. Paris : Lavoisier. (= *Traité Cognition et traitement de l'information*).
- Debaisieux, Jeanne-Marie/Boulton, Alex (2007): « Alors la question c'est...? Questions pragmatiques et annotations pédagogiques des corpus ». *Cahiers de l'AFLS* 13/2: 31–59.
- Delatour, Yvonne/Jennepin, Dominique/Léon-Dufour, Maylis/Teyssier, Brigitte (2004): *Nouvelle grammaire du français. Cours de Civilisation Française de la Sorbonne*. Paris: Hachette.
- Habert, Benoît (2000): « Des corpus représentatifs : de quoi, pour quoi, comment ? ». In: Bilger, Mireille (ed.): *Linguistique sur corpus – Etudes et réflexions* 31. Perpignan, Presses Universitaires de Perpignan: 11–58.
- Halliday, Michael Alexander Kirkwood (1991): “Corpus studies and probabilistic grammar”. In: Aijmer, Karin/Altemberg, Bengt (eds.): *English Corpus Linguistics*. London/New York, Longman: 30–43.
- Hunston, Susan (2002): *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Legallois, Dominique (2012): « La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ? ». *Corpus* [En ligne] 11 mis en ligne le 21 juin 2013, consulté le 29 juin 2013. URL : <http://corpus.revues.org/2202>.
- Rastier, François (2005): « Enjeux épistémologiques de la linguistique de corpus ». In: Williams, Geoffrey (ed.): *La linguistique de corpus*. Rennes, Presses Universitaires de Rennes: 31–46.
- Rey, Alain (ed.) (2013): *Le Grand Robert de la langue française*. Paris: Dictionnaires Le Robert.
- Sinclair, John (1991): *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Veronis, Jean (2004): <http://sites.univ-provence.fr/veronis/donnees/index.html>