

Introduction

Introduction

Charlotte Danino

**Electronic version**

URL: <http://journals.openedition.org/corpus/3099>

ISSN: 1765-3126

Publisher

Bases ; corpus et langage - UMR 6039

Electronic reference

Charlotte Danino, « Introduction », *Corpus* [Online], 18 | 2018, Online since 03 July 2018, connection on 08 September 2020. URL : <http://journals.openedition.org/corpus/3099>

This text was automatically generated on 8 September 2020.

© Tous droits réservés

Introduction

Introduction

Charlotte Danino

« *If there was one development we did not predict it was that, as well as getting larger, corpora would also get smaller !* » McEnery et Wilson (2001 : 191)

Doit-on qualifier de « petit corpus » tout corpus qui n'est pas grand ?

- 1 Et d'ailleurs, pourquoi faudrait-il se poser la question ? Celle-ci peut paraître triviale mais elle signale un moment particulier des sciences du langage. L'informatique a permis l'établissement de grands corpus et le traitement d'impressionnantes quantités de données. L'évolution des outils technologiques et statistiques notamment a transformé les pratiques en sciences du langage. La question du corpus s'est reposée, avec en son cœur, le problème de son statut. Le corpus est-il outil ou objet ? Point de départ ou d'arrivée ? Les différents domaines de recherche et les différentes approches ont répondu diversement à ces deux questions auxquelles toutes études continuent de devoir répondre. Le corpus est en effet une idée très malléable : recueil de productions linguistiques authentiques, voilà le cœur de sa définition – tout le reste semble adaptable.
- 2 Si bien que le principal enjeu apparaît alors comme la tâche de dresser une typologie des corpus, laquelle ferait ensuite apparaître la structuration du champ des sciences du langage. Considérons seulement la longue liste de qualificatifs à même de modifier le mot corpus : oral, écrit, diachronique, synchronique, web, phonologique, sémantique, pragmatique, audio, vidéo ou audiovisuel ; test, exploratoire, *ad hoc*, au vol, homogène ou hétérogène, de référence, annoté, enrichi, outillé, aligné, parallèle, bi- ou multilingue, de travail, fini ou clos, représentatif ; les adjectifs anglais empruntés tel quel se rajoutent à ces termes : *sampling*, *monitoring*, *DIY* (pour *do-it-yourself*)

notamment. Un rapide survol des contributions à ce volume achèvera de convaincre le lecteur.

- 3 Et évidemment *grand* et *petit corpus*. *Petit* est d'ailleurs rarement donné en tant que tel : cet adjectif est utilisé pour dire « pas grand », souvent affublé d'un qualificatif comme *spécialisé*, *test* ou *oral* – et parfois *tout petit* –, qui excuse la petite taille par le type. Les (tout) petits corpus semblent souvent susciter la méfiance ou la suspicion si bien qu'il faut s'en justifier prestement.
- 4 Le présent volume vise à établir les petits corpus en eux-mêmes, à en donner une définition, théorique, méthodologique et fonctionnelle positive, qui cherche entre autres choses à « dédramatiser la notion de petitesse », pour emprunter les mots de Viollain et Chatellier (ce volume). Il s'agit aussi de répondre à la question tutélaire : tout corpus qui n'est pas grand est-il petit ? Cela impliquerait une bipartition mais on pourrait après tout envisager très sérieusement une tripartition qui intégrerait des *moyens* corpus. Les contributions au présent volume ont été rassemblées dans le but de répondre à la question de la définition des petits corpus et à celle de la partition des corpus linguistiques (bipartition, tripartition, catégories multiples, partition épistémologique, qualitative, etc.) selon l'idée familière : dis-moi quel corpus tu utilises et je te dirai quel linguiste tu es/quelle linguistique tu fais.
- 5 La linguistique de corpus, celle des grands corpus, semble se stabiliser autour d'un consensus, sans avoir mis un point final à toutes les discussions. En témoigne le nombre de manuels à destination des étudiant.e.s ou des chercheur.e.s qui veulent se former à ce champ de recherche, signe d'une stabilisation des connaissances à leur égard dans un ensemble de bonnes pratiques qui forment un faisceau cohérent d'approches linguistiques (Tognini-Bonelli 2001, McEnery et Hardie 2011, Baker 2012, Biber et Reppen 2015 en anglais, Williams 2005, Landragin et Poudat 2017 pour des ouvrages en français).
- 6 Le moment serait donc venu de se pencher sur ces corpus qui ne sont pas grands. En effet, de très nombreuses études de linguistiques reposent sur des corpus sans pour autant être représentatives de la linguistique dite de corpus. Aussi faudrait-il distinguer linguistique *de* et *sur* corpus. Si la première est un champ précis, structuré autour de conceptions théoriques explicites et de méthodologies dédiées aux grands corpus principalement, la linguistique *sur* corpus est-elle alors par défaut la linguistique des petits corpus ?

Le possible et le nécessaire

- 7 Travailler à la définition de l'objet et de l'outil *petit corpus* est devenu nécessaire ne serait-ce qu'en vertu d'un constat très simple : ils sont très nombreux ; et très utilisés, alors même que l'on peut techniquement recourir aux grands. « *My analysis of the continuing creation of small corpora when it is technologically possible to create larger ones is that here necessity is playing a larger role* », écrivait déjà Antoinette Renouf (2007 : 29) il y a une dizaine d'années. Plusieurs auteurs de ce volume se font l'écho de ce constat.
- 8 Viollain et Chatellier par exemple parlent d'une survivance. Doualan rappelle les conditions institutionnelles de la thèse de doctorat qui empêche bien souvent la constitution et surtout le traitement de corpus trop grands. C'est d'ailleurs dans la recherche doctorale de l'auteur de ces lignes que se trouve l'origine de ce projet

(Danino 2014). Debras aussi prend l'exemple, entre autres, de son propre corpus de thèse et pose la question des corpus multimodaux, extrêmement chronophages.

- 9 Ross, Hennecke ou encore Carruthers et Vergez-Couret sont confrontés à la quantité de sources ou de données disponibles ou accessibles : linguistique diachronique, langues peu dotées ou éteintes peuvent et doivent assumer la petite quantité de données sans renoncer à des objectifs de recherche partagés avec les langues (ou productions) mieux dotées.
- 10 La pragmatique et la didactique des langues, enfin, sont deux champs qui se sont emparés très explicitement de cette problématique (citons ainsi Vaughan et Clancy 2013, Ghadessy *et al.* 2001). Non que les données manquent : mais leurs objets d'étude empêchent le recours aux grands corpus. Prenons un exemple : les grands corpus reposent, pour être maniables, sur le traitement automatique (au moins partiel). Mais comment entraîner un ordinateur à repérer toutes les fautes possibles produites par des apprenants ? Comment lui apprendre à défaire la subtilité d'inférences contextuelles reposant sur l'implicite, c'est-à-dire largement sur l'absence d'observables¹ ? Landragin dans ce volume rend compte de ce problème dans une étude sur la coréférence.
- 11 Pour tous ces chercheurs, le petit corpus est un ou le seul objet d'étude possible. Résumer les petits corpus à une trop faible quantité de données reviendrait à disqualifier les acquis de ces champs de recherche, ou du moins à les nuancer fortement, en les cantonnant aux rôles d'études préliminaires ou génératrices d'hypothèses. Car le corpus continue de poser des questions d'épistémologie et de méthodologie aux sciences du langage. Entre théorie et pratique, entre objet et outil, le corpus continue de différencier les approches selon les réponses que l'on apporte aux questions qu'il pose. La contribution de Raccah se fait l'écho de cette tension, qui peut mener à certains contresens.
- 12 Dans cette logique-là le petit corpus ne saurait se définir qu'en tant que contrepoint, nécessité ou phase préliminaire. La difficulté est donc de reposer la question du corpus (ou des corpus plutôt) sans recourir aux notions de représentativité, de taille, ou d'outillage qui ont présidé à la définition des grands corpus. Le critère de la taille n'en est d'ailleurs pas un, ni pour les grands ni pour les petits vu le peu de consensus qui existe sur l'établissement de seuils. Ainsi, le seuil de 1 million de mots en dessous duquel le petit corpus doit se situer est fréquemment repris (*e.g.* Fachinetti [éd.] 2007, McEnery et Wilson 2001) mais ce chiffre ne fait pas l'unanimité et ne vaut, de plus, qu'en comparaison des grands corpus. Certains auteurs comme Vaughan et Clancy (2013 : 55) remarquent la fluidité de cette frontière :

While it seems to be accepted that the upper limit of a small corpus is approximately 200,000-250,000 words (see Aston 1997 ; Flowerdew 2004), one- to five-million-word samples have also been described as 'small' (McCarthy 1998 ; Sinclair 2001). Aston (1997) notes that small corpora exist in the 20,000-200,000 word range, and are more specialized in terms of topic and/or genre than large corpora.

Size doesn't matter : so what does ?

- 13 Dans ce volume, les auteurs assument tous l'étiquette de petits corpus sans forcément travailler sur les mêmes nombres de mots. Tous préfèrent d'autres unités de mesure que la quantité d'unités (mots, phonèmes, constructions, selon les études). Ainsi, les

petits corpus oraux de quelques heures permettent d'atteindre largement la **densité** nécessaire pour étudier un phénomène phonologique ou gestuel (Viollain et Chatellier, Debras). Mais ces corpus sont-ils encore petits s'ils permettent de franchir un certain seuil quantitatif critique ? La réponse tient en fait largement à la manière de leur constitution.

- 14 D'autres critères sont mis en avant par les auteurs : fréquence relative par rapport à des corpus dits représentatifs (*monitor, sampling corpora*), **comparabilité** avec d'autres corpus (Biros, Rossi et Sahakyan sur ce point notamment, Hennecke pour la question des corpus parallèles), **homogénéité** des productions, type de situations, sont autant d'idées qui tâchent de reposer le débat dans d'autres termes que ceux dans lesquels il s'est déjà tenu. Citons ainsi la réflexion de Verdier et de Fornel (ce volume), qui, au sein de l'analyse conversationnelle, envisagent les situations sociales au prisme des interactions verbales auxquelles elles donnent lieu. En confrontant la notion de corpus à celles de classe et de collection, la notion de représentativité fait face à celle d'exemple et de ce qui détermine sa valeur, et donc sa portée. Newton aurait-il eu besoin d'étudier trois pommes pour penser la gravité ? Raccah pose cette exacte question en se plaçant au cœur de la sémantique. **L'exemplarité** dispense l'analyse du nombre. Encore faut-il l'établir.
- 15 Ainsi, la notion de représentativité n'est pas gommée mais déplacée. Un grand corpus vise la représentativité d'une langue. Le petit corpus vise lui la représentativité d'un usage de la langue. Doualan reprend cette idée de spécialisation, en s'alignant sur les critères de Flowerdew (2004). Elle en explore ainsi la validité pour l'étude de la synonymie dans un contexte très précis. Se profile l'idée de **typicalité** d'une production linguistique dont les paramètres ont été soigneusement étudiés. Métadonnées, contexte de production, contraintes : la constitution des petits corpus est en fait la première donnée de leur analyse. O'Keefe, McCarthy et Carter (2007 : 4) soulevaient déjà ce point. La notion de *corpus design* est cruciale pour tout corpus bien évidemment, mais elle est consubstantielle des petits construits autour d'un biais, d'une entrée dans la langue. La **documentation** des petits corpus, comme le rappelle clairement Landragin (ce volume) devient alors décisive. Tout se passe comme si les défauts des grands corpus deviennent les leviers qui président à la constitution des petits.
- 16 La littérature habituée à cet objet linguistique se fait largement l'écho de ces déplacements conceptuels. Bianchi et Passaglia peuvent ainsi écrire (in Facchinetti [éd.] 2007 : 266) : « *in terms of size, several authors agree that the most specific the type of language to investigate, the smaller the corpus can be* ». Dans le même ouvrage, S. Granath explicite le titre de son chapitre « Size matters » par la glose « *matters having to do with size* ». L'auteur en vient d'ailleurs (*op. cit.* 169) à défendre l'idée d'une complémentarité des grands et petits corpus, idée largement reprise par la communauté scientifique (Cameron et Deignan 2003, Mair in [Renouf et Kehoe 2016], Debras, ce volume notamment). Citons enfin Laurence Anthony, qui prolonge l'analogie proposée par Stubbs (1996, cité par l'auteur) pour qui les corpus sont comme les télescopes des astronomes. Anthony reprend, et nous le citons au long :

[...] *corpora are more like regions of space such as stars, planets and nebulae that we observe with telescopes. Seeing corpora in this way allows us to resolve some of the common issues in the field. For example, there has been a great debate about the value of small corpora with some people seeing them as simply an unfortunate limitation (Sinclair 2004), and others seeing their great potential (McEnery and Wilson 2011). If we see corpora as regions of space, the value of both large and small corpora is immediately apparent.*

Investigating large corpora can be compared with investigations of galaxies from which we can learn about the building block of the universe and create models about how it all works. A study of a small corpus can be compared with looking at a small group of stars, from which we can gain more detailed information about their unique properties. [...] However limited the corpus is in terms of size and representativeness, nobody can doubt the value of astronomical observations of the nearest star to us, the Sun. The usefulness of a corpus is dependent not on its size but on what kind of information we hope to extract from it with our tools.

(Anthony, in [Baker 2012 : 91-92])

- 17 Le corpus est ici un objet linguistique à traiter avec certains outils qui permettent de dégager des données. **Données** et corpus ne peuvent être confondus et les outils se pensent et se forment selon la finalité de la recherche. Tognini-Bonelli (2001) dans son introduction se fait l'écho de cette distinction (voir notamment ses notions d'application et de pré-application, 2001 : 1-2). Si l'on adopte la définition minimale du corpus comme ensemble raisonné de productions linguistiques attestées, les données sont en deçà du corpus, ce qui émerge lorsque le corpus est traité méthodologiquement. On comprend alors que l'auteur propose un chapitre par **type d'approche**, inductive et déductive (*corpus-based* ou *corpus-driven* – que l'on différencie maintenant de plus en plus d'approches dites *data-driven*). Williams (2005) rappelle que :

La linguistique de corpus [cherche] [...] à révéler les choix linguistiques opérés par les locuteurs dans des contextes réels. [...] Deux approches fondamentales se côtoient, l'une déductive, *corpus-based* en anglais, qui utilise le corpus pour confirmer ou infirmer une hypothèse, et une linguistique de corpus inductive, *corpus-driven*, qui cherche à explorer les données sans *a priori*. Les deux sont complémentaires [...].

(Williams [dir.] 2005 : 13)

- 18 Ainsi, au sein même de la linguistique de (et sur) corpus, plusieurs approches coexistent que les chercheurs utilisent déjà de manière complémentaire. Les petits corpus semblent reposer ces questions avec force. Ces quelques mots d'introduction mettent au jour la variété des problématiques soulevées par les petits corpus aux disciplines des sciences du langage, de même que la diversité des approches et des utilisations de cet objet – et outil. Elle laisse entrevoir également la défense d'une complémentarité des outils à la disposition des chercheurs. Cela implique alors de consacrer aux petits corpus la même attention dont ont bénéficié les grands.

Structure du volume

- 19 Les auteurs de ce volume partagent cet objectif et permettent de proposer au lecteur un parcours qui tente d'organiser ce paysage théorique et méthodologique, en quatre temps.
- 20 Les quatre premières contributions permettent de (re)situer le débat dans ses ENJEUX THÉORIQUES, ÉPISTÉMOLOGIQUES ET PRATIQUES.
- 21 **Pierre-Yves Raccah** ouvre paradoxalement ce recueil en arguant de « l'inutilité des *corpora d'occurrences* en sémantique des langues ». Derrière cette critique, l'auteur réexamine le statut de l'empirie en linguistique, dans son opposition traditionnelle avec l'introspection : le (petit) corpus se tiendrait-il au milieu du champ de bataille ?

- 22 En contrepoint, **Maud Verdier** et **Michel de Fornel** abordent le problème tel qu'il se pose à l'analyse conversationnelle où le petit corpus est quasiment une conséquence nécessaire de la théorie ethnométhodologique qui sous-tend cette méthodologie d'analyse. L'analyse conversationnelle « sera toujours aussi une linguistique de petit corpus » même si l'analyse en classes peut permettre le rapprochement avec des approches quantitatives².
- 23 **Cécile Viollain** et **Hugo Chatellier** proposent une réflexion et une étude de cas dans le domaine de la phonologie. Leur contribution définit la notion de corpus phonologique qui leur permet de présenter le programme PAC (*Phonologie de l'Anglais Contemporain*) : plusieurs petits corpus rassemblés en une grande base de données font-ils « grand corpus ? ».
- 24 Pour clore ce cadrage autant théorique qu'épistémologique, **Camille Debras** propose une réflexion sur les petits corpus dans le domaine relativement émergent, et donc en construction, de la gestualité coverbale. À mi-chemin entre recension et état des lieux, le texte invite à se concentrer sur l'interprétation des données et le travail de théorisation qui le précède.
- 25 La seconde étape de notre parcours envisage le PROBLÈME DE L'HÉTÉROGÉNÉITÉ, assumée ou de fait, au sein des corpus de taille variée. Chaque corpus peut être divisé en sources, sous-corpus selon des critères différents d'une étude à l'autre, de même qu'ils peuvent être eux-mêmes conçus comme test ou première version d'une entreprise plus large.
- 26 **Gaëlle Doualan** revient sur les notions de représentativité et de spécialisation pour leur préférer l'idée de typicalité liée aux types de discours, mais aussi aux types de situations qui ont mené à leur production. Dans une étude, par ailleurs doctorale, sur la synonymie, l'auteure teste la validité d'un protocole qui aura mené à la constitution d'un petit corpus dans la droite ligne des idées de Koester (2007) cité par Doualan :
- The patterns can first of all be linked to a particular context, because the corpus analysis shows that they are concentrated within that context. We can see that these patterns are localised, as they are traceable to local situational conditions, such as gender, power or discourse goal. As a result, the patterns can be linked to pragmatically specialised uses within that particular context of situation.*
- (Koester 2010 : 74)
- 27 **Camille Biros**, **Caroline Rossi** et **Inesa Sahakyan** envisagent ensuite des types variés de corpus qui posent tous la question du sous-corpus : est-il un petit corpus ? Comment le traiter indépendamment des autres ? Quel statut de la collection finale ? En prenant l'exemple du discours sur l'environnement selon les communautés de locuteurs, problématique qui oblige à sous-diviser les données recueillies, les auteurs tâchent de penser les lignes de sous-catégorisation et les découpages que la volonté d'équilibrage et de comparabilité rend nécessaire.
- 28 À l'inverse de cette diffraction, **Frédéric Landragin** présente un corpus unique, un « tout petit corpus », pensé comme exploratoire et préliminaire. Dans le cadre d'un projet d'étude de la coréférence avec les outils de la linguistique computationnelle, l'auteur fait le récit d'une série de difficultés qui aura forcé l'équipe de chercheurs à reprendre leur approche, permettant ce faisant de repenser les pratiques de recherche associées aux petits corpus tests ou exploratoires dans l'élaboration de projet plus important. Sa contribution se situe entre l'épistémologie et l'institution.
- 29 Enfin, **Inga Hennecke** propose une étude dans le champ du bilinguisme et du multilinguisme. Là encore se pose la question de ce qui fait (petit) corpus. Dans les

corpus parallèles, où est le corpus ? En tentant de disperser « l'insécurité méthodologique » qui règne autour de ces corpus, l'auteure affine la typologie des corpus plurilingues qu'elle met à l'épreuve d'une recherche sur un corpus bilingue qui confronte l'anglais, langue démesurément dotée au français manitobain, nettement moins dotée.

- 30 La troisième étape de notre parcours propose d'ancrer la réflexion dans LA NOTION DE CONTRAINTE.
- 31 **Janice Carruthers** et **Marianne Vergez-Couret** font également face aux exigences posées par l'étude d'une « langue minorisée ». Leur recherche a mené à la constitution d'un petit corpus comparatif en occitan dont les différentes parties font varier le degré d'oralité. La contrainte géographique se double d'une réflexion obligatoire sur la forme orale ou écrite des sources intégrées, sur l'époque du collectage et sur les facteurs sociolinguistiques afin de maîtriser la comparabilité au sein du corpus.
- 32 Ce n'est plus le lieu mais le moment qui agit comme principale contrainte de la contribution de **Sophie Moirand** qui s'intéresse aux « instants discursifs ». Ses objets discursifs sont de courte durée dans la sphère publique, si bien que le corpus doit être constitué « au vol » selon des critères que l'on souhaite le plus stable possible mais qui restent malgré tout coordonnés au développement d'un discours en devenir. Et le corpus s'arrête quand sa source discursive s'éteint.
- 33 La dernière contribution du recueil pose autant de questions qu'elle reprend de problèmes représentés ailleurs dans le volume. En effet, **Daniel Ross**, dans une étude approfondie que l'on propose au lecteur en tant que somme (ou clôture temporaire) étudie la construction anglaise *try and* + verbe. En tâchant d'être exhaustif sur cette construction l'auteur doit prendre en charge une hétérogénéité multipliée par différentes contraintes : diachronie, variation et acquisition sont trois des approches que Ross doit combiner dans cette étude qui accumulent les sources restreintes.
- 34 Il est évident que ce parcours ne résume pas les possibilités de lecture de ce recueil : approche par domaine, opposition de l'oral et de l'écrit, parti pris méthodologique sont autant de manières de tracer son propre passage dans la problématique des petits corpus. Chaque article travaille par ailleurs chacun à sa manière les questionnements épistémologique, théorique et méthodologique.
- 35 On souhaite enfin signaler deux pistes de recherche auxquelles nos choix éditoriaux n'ont pas pu rendre justice. La première concerne l'acquisition et l'apprentissage. Il semble qu'il y a là matière à réflexion tant en prenant le point de départ des petits corpus que celui du corpus en général. Une communication récente de Christophe Parisse dans le séminaire Corpus et recherches linguistiques STIH (28 mars 2018) semble aller dans le même sens. Citons aussi l'ouvrage de Timmis (2015) dans lequel l'auteur tente de présenter de manière synthétique ce que les corpus peuvent apporter à l'enseignement de l'anglais³.
- 36 La seconde remarque est en fait un appel à la communauté scientifique : les petits corpus, plus encore que les grands, nous obligent à un effort de recension, pour la mise en valeur des travaux menés mais aussi pour rendre ces corpus visibles, disponibles, et finalement encore plus utiles. Les corpus de thèse notamment gagneraient à être mieux recensés, archivés, diffusés. Ce paragraphe est une invitation aux chercheurs concernés à faire connaître leurs corpus « personnels » auprès d'initiatives telles que celles de l'IRCOM ou ORTOLANG⁴.

- 37 Ces deux perspectives nous incitent à poursuivre la réflexion autour des petits corpus, et nous serions heureux de le faire avec les lecteurs de ce volume.
- 38 Cette introduction ne saurait être close sans l'expression de notre sincère reconnaissance à ceux dont l'expertise et les remarques ont permis l'élaboration de ce numéro thématique. Nous sommes heureux de pouvoir les remercier ici, par ordre alphabétique :
- 39 Sophie Alby, Adrien Barbaresi, Laure Calabrese, Shirley Carter-Thomas, Gilles Col, Catherine Collin, Alice Deignan, Richard Faure, Gaëlle Ferré, Peter Follette, Cédric Gendrot, Vincent Hugou, Véronique Magri, Pascale Manoïlov, Michèle Olivieri, Paola Pietrandrea, Alain Rabatel, Geneviève Salvan, Tobias Scheer et Geoffrey Williams.
- 40 Nos remerciements vont tout spécialement à Sylvie Mellet et Damon Mayaffre pour leur patience, leur aide et leur attention durant toutes les phases de réalisation de ce volume.
- 41 Toutes les erreurs et imperfections de ce volume nous reviennent évidemment, et nous espérons que le lecteur nous en excusera.

BIBLIOGRAPHY

- Anthony, Laurence (2012). « Issues in the design and development of software tools for corpus studies – The case for collaboration », in Baker Baker, Paul (éd.) *Contemporary Trends in Corpus Linguistics*. London/NewYork : Continuum.
- Baker, Paul (éd.) (2012). *Contemporary Trends in Corpus Linguistics*. London/NewYork : Continuum.
- Bianchi, Francesca, et Pazzaglia, Roberto (2007). « Student writing of research articles in a foreign language : metacognition and corpora », in *Corpus Linguistics 25 Years on*. Boston, USA : Brill / Rodopi.
Doi : https://doi.org/10.1163/9789401204347_015.
- Biber, Douglas et Randi, Reppen (éd.) (2015). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge : Cambridge University Press.
- Cameron, Lynne et Deignan, Alice (2003). « Combining Large and Small Corpora to Investigate Tuning Devices Around Metaphor in Spoken Discourse », *Metaphor and Symbol* 18(3) : 149-160.
- Danino Charlotte (2014). *Les mécanismes de construction du sens dans le discours sur l'événement en cours : étude de sémantique sociocognitive sur le direct de CNN le 11 septembre 2001*. Thèse de l'Université de Poitiers.
- Fachinetti, Roberta (éd.) (2007). *Corpus Linguistics 25 years on*. Amsterdam : Rodopi.
- Flowerdew Lynne (2004), « The Argument for Using English Specialized Corpora to Understand Academic and Professional Settings », in U. Connor et T. Upton (éd.) *Discourse in the Professions : Perspectives from Corpus Linguistics*. Amsterdam : John Benjamins : 11-33.
- Ghadessy, Mohsen, Alex, Henry et Robert L. Roseberry (éd.) (2001). *Small corpus studies and ELT : Theory and Practice. Studies in Corpus Linguistics*. Amsterdam : John Benjamins.

- Koester A. (2010). « Building small specialized corpora », in A. O’Keeffe et M. J. McCarthy (éd.) *The Routledge Handbook of Corpus Linguistics*. London : Routledge Publ., 66-79.
- Landragin, Frédéric et Poudat, Céline (2017). *Explorer un corpus textuel. Méthodes – Pratiques – Outils*. Louvain-la-Neuve, Belgique : De Boeck Supérieur.
- Mair, Christian (2016). « Tracking ongoing grammatical change and recent diversification in present-day standard English : the complementary role of small and large corpora », in Renouf et Kehoe (éd.) *The Changing Face of Corpus Linguistics*. Boston : USA Brill.
- McEnery, Tony et Andrew, Wilson (2001). *Corpus Linguistics. An Introduction*. 2^e éd. Édinburgh : Edinburgh University Press.
- McEnery, Tony, et Hardie, Andrew (2011). *Corpus Linguistics : Methods, Theory and Practice*. Cambridge : Cambridge University Press.
- McIlvenny, Paul, et Raudaskoski, Pirkko Liisa (1996). « The mutual relevance of CA and Linguistics : a discussion in reference to interactive discourse », in L. Heltoft (éd.) *Proceedings of the Thirteenth Scandinavian Conference of Linguistics*, Roskilde, Denmark, 263-277 [En ligne] <http://ntpaul.sprog.auc.dk/Paul/research/cv/pubs/ca-linguistics96.htm>.
- O’Keeffe A. (2007). « The Pragmatics of Corpus Linguistics », communication présentée à la 4^e Conférence de linguistique de corpus, Université de Birmingham, Birmingham, juillet 2007.
- O’Keeffe Anne, McCarthy Michael et Carter Ronald (2007). *From Corpus to Classroom*. Cambridge : Cambridge University Press.
- Parisse, Christophe (2018). « Corpus et recherche sur l’acquisition de langage », Séance du séminaire Recherches Linguistiques et Corpus, STIH Paris Sorbonne. 28 mars 2018.
- Partington, Alan Scott (2014). « Mind the gaps. The role of corpus linguistics in researching absences », *International Journal of Corpus Linguistics* 19 : 1, p. 118-146.
- Renouf, Antoinette et Kehoe, Andrew (éd.) (2016). *The Changing Face of Corpus Linguistics*. Boston : USA Brill.
- Renouf, Antoinette. 2007. « Corpus Development 25 years on : from super-corpus to cyber-corpus ? », in Facchinetti R. (éd.) *Corpus Linguistics 25 Years on*. Amsterdam : Rodopi.
- Timmis, Ivor (2015). *Corpus linguistics for ELT, Research and Practice, Routledge Corpus Linguistic Guides*. London & New York : Routledge.
- Tognini Bonelli, Elena (2001). *Corpus Linguistics at Work*. Amsterdam : John Benjamins.
- Vaughan, Elaine et Brian, Clancy (2013). « Small corpora and pragmatics », *The Yearbook of Corpus Linguistics and Pragmatics* 2013 : 53-73.
- Williams, G. (dir.) (2005). *La linguistique de corpus*. Presses Universitaires de Rennes.

NOTES

1. C’est le problème des données négatives face auquel l’ordinateur est particulièrement mal armé. Tognini-Bonelli (2001 : 17) rappelle que les approches sur corpus reposent sur la présence de preuves postulées (*corpus-based approach*) ou avérées (*corpus-driven approach*). Partington (2014) tente cependant de nuancer l’exclusion de l’absence du champ de réflexion.
2. Rappelons que linguistes et analystes de conversation sont depuis plusieurs décennies engagés dans un débat parfois tendu sur les mérites comparés des deux approches du langage. McIlvenny

et Raudaskoski (1996), par exemple, rappellent précisément les termes du débat tout en prônant une conciliation.

3. Nous remercions Claire Tardieu de nous avoir présenté cet ouvrage.

4. IRCOM : ircom.huma-num.fr/

ORTOLGANG : <https://www.ortolang.fr/>

AUTHOR

CHARLOTTE DANINO

Université Sorbonne Nouvelle Paris 3