

Introduction

Le corpus à l'interface des humanités et des sciences sociales

Le corpus de textes constitue l'un des lieux les plus favorables à l'observation des réalisations de la langue pour la linguistique ; en témoigne une de ses principales, sinon sa principale branche méthodologique, la linguistique justement dite « de corpus ». Dans cette optique, deux conceptions du corpus coexistent et peuvent se nourrir l'une l'autre : il est tantôt considéré comme un recueil d'exemples ou d'attestations utilisé de façon déductive pour vérifier des hypothèses ou étayer des théories linguistiques (l'approche *corpus-based*), tantôt perçu comme un outil d'exploration à part entière, susceptible de faire émerger de manière inductive des savoirs linguistiques (l'approche *corpus-driven*) (Williams 2005, p. 13). En parallèle, à la suite du *discursive turn* qui s'affirme dans les années 1970, les chercheurs en sciences sociales ont conscience du fait que les objets qu'ils étudient émergent discursivement. Chacune des disciplines considère alors qu'elle a « constamment affaire au langage quels que soient ses méthodes et ses domaines de recherche » (Lahire 1994, p. 189). Ainsi, « c'est considérer que ce qui est dit et écrit médiatise une part de la réalité et y donne une prise » (Le Lay 2013, p. 3). Répondant à des démarches empiriques, le corpus de textes est ainsi devenu le terrain privilégié de la mise au jour et / ou de l'observation de phénomènes historiques, géographiques, sociologiques, à partir de leurs formulations et dénominations par différents acteurs.

La réflexion autour du corpus s'affirme comme un enjeu méthodologique central et, dans le champ linguistique, elle a été souvent menée, permettant de définir nettement le statut épistémologique du corpus ; le concept de corpus fait partie du quotidien du linguiste, référence banale et fréquente (Dalbera 2002). Cependant, le problème reste encore peu pris en charge en sciences sociales : si l'utilisation du terme devient de plus en plus fréquente, les développements sur cet objet, tant épistémologiques que méthodologiques, sont encore discrets. En premier lieu, le corpus peut être considéré à travers la notion de prisme. En effet, le corpus naît d'un prisme, celui du chercheur à la recherche d'un faisceau convergent de preuves ou de traces ; en linguistique, même dans le cas où le corpus préexiste au projet de recherche du chercheur ou dans le cadre de l'approche *corpus-driven*, les données initialement brutes et objectives finissent par être filtrées par la conscience du scientifique. Mais le corpus est souvent lui aussi un prisme puisqu'il retranscrit les objectifs et les questionnements du chercheur et qu'il constitue la source principale de résultats, voire d'interprétations. De fait, « tout corpus linguistique est d'abord représentatif du système d'hypothèses que le chercheur s'est fabriqué » (Tournier 1993, p. 31). Parmi les trois réductions – linguistique, heuristique et textométrique – que Maurice Tournier (1993) distinguait pour caractériser le chemin qui conduit au corpus, la deuxième désigne cet effort de focalisation : la réalité complexe est passée au prisme de la curiosité du chercheur qui définit une thématique, adopte un cadrage spatio-temporel et formule un petit nombre d'hypothèses susceptibles d'être testées. Même en retenant des invariants temporels (bornes chronologiques), spatiaux (une région), thématiques (objet d'étude) ou linguistiques (par exemple la langue française), il reste tant de variables – notamment les caractéristiques socioculturelles des énonciateurs – qu'il importe d'en minoriser quelques-unes pour évaluer l'influence des variables d'étude.

Questionner le corpus implique de prendre du recul sur la démarche de recherche sur les discours : révéler ses limites, ses interrogations et ses dits semble crucial dans une approche constructiviste. « Le caractère subjectif et singulier d'un corpus peut être relativisé, dans la mesure où il conduit à une exploitation et à des résultats s'inscrivant dans un cadre d'analyse commun et réutilisables » (Bommier-Pincemin 1999, p. 423). En effet, assumer la singularité et la subjectivité d'un corpus ne vise pas à remettre en cause sa pertinence, mais bien à le confronter à la littérature existante et à d'autres sources ou corpus pour parvenir à consolider les résultats. Ce positionnement inscrit le corpus entre matériel et immatériel : il est une représentation concrète (des formes textuelles et iconiques) mais aussi idéelle (à la suite de « Dis-moi quel est ton corpus, je te dirai quelle est ta problématique » de Patrick Charaudeau, en 2009). Dans cette représentation, le texte pourrait incarner la manifestation concrète consécutive à une sémiotisation, quand les discours seraient plutôt de

l'ordre du cognitif, voire de l'invisible ; c'est ainsi que l'on peut dire, à la suite des travaux de Jean-Michel Adam, que « le discours est défini comme le texte mis en rapport avec ses conditions historiques de production, d'où le schéma couramment répercuté : DISCOURS = TEXTE + CONDITIONS DE PRODUCTION, TEXTE = DISCOURS - CONDITIONS DE PRODUCTION » (Jaubert 2002). Tous les discours étudiés en linguistique et sciences sociales prennent la forme matérielle du texte, mais la dimension discursive d'un même texte fait qu'il est le terrain de différents niveaux d'expression et de représentation, chacun pouvant être étudié suivant l'orientation disciplinaire du chercheur.

La construction du corpus, son analyse, son interprétation et sa mise en perspective questionnent certes le contexte d'énonciation des discours, mais aussi la place ou la posture du chercheur. Au sein de chaque discipline, la place accordée aux corpus textuels, ainsi que la question de leur investigation, de leur taille, de leur représentativité et de leur interprétation doit être éclaircie et débattue. En outre, la situation centrale occupée par le texte interroge les cloisonnements disciplinaires et la possibilité de les dépasser. À quelles conditions peut-on faire communiquer les savoirs et les méthodes d'un champ à l'autre ? Les corpus de textes (considérés comme construits) représentent-ils un dénominateur commun, ou sont-ils le révélateur de visions segmentées qui reproduisent des clivages disciplinaires ?

De l'interdisciplinarité à la fécondité de la pensée par corpus

Damon Mayaffre (2002), chercheur à l'interface entre l'histoire et la linguistique, propose deux perspectives pour considérer les corpus : d'une part une approche sérielle et d'autre part une approche heuristique. La première approche fait du corpus un ensemble complexe de données, quand la seconde insiste sur son intérêt pour répondre à une problématique choisie (Duteil-Mougel 2005).

Ce point de vue peut être complété par celui exprimé par François Rastier qui affirme que deux conceptions du corpus coexistent :

L'une, *documentaire*, ne retient que des variables globales caractérisant les documents, sans tenir compte de leur caractère textuel, ni de leur structure. Dans cette conception *logico-grammaticale*, le corpus se résume à un échantillon de la langue, un réservoir d'exemples ou d'attestations. En revanche, la conception *philologique-herméneutique* tient compte des rapports de texte à texte, ce qui n'est possible qu'au sein d'un discours. (Rastier 2004)

Ces deux approches sont combinées : l'approche documentaire permet d'inventorier et de classer le corpus, mais elle n'est jugée éclairante que dans le cas de la deuxième théorisation du corpus. En effet, la présence ne prend sens qu'en lien avec le contexte d'énonciation ou la comparaison entre différents discours. Cette deuxième conception semble la plus exploitée par les sciences humaines

et sociales, quand les deux approches témoignent d'une réelle richesse pour les humanités ; certains linguistes, pour leur part, se contentent parfaitement de la première approche, suffisante pour faire émerger des faits de langue.

De la linguistique...

Les textes sont considérés à la fois comme l'objet empirique de la linguistique (Pincemin 2011), mais aussi comme son unité minimale à trois niveaux (discours, champ générique et genre) qui prend son sens dans un corpus (Rastier 2004).

« Tout ensemble de textes n'est pas un corpus » (Bommier-Pincemin 1999, p. 416). Un corpus est un recueil composite où est regroupée une collection de textes avec une volonté de cohérence (Mayaffre 2002). Ainsi, un corpus consiste en un regroupement de textes et donc à une addition ou à un cumul de données : ce rassemblement de textes semble fonctionner par agrégation. Néanmoins, constituer un corpus reconnu en tant que tel revient aussi à soustraire ou à ôter des textes. Un corpus est donc toujours un ensemble jugé représentatif d'un tout, sans pour autant correspondre *stricto sensu* à ce tout. Un corpus vérifie des conditions de signifiante, d'acceptabilité et d'exploitabilité (Bommier-Pincemin 1999). Il se présente donc comme un assemblage de différentes qualités poursuivant des objectifs particuliers.

Conditions	Qualités	Objectifs
Signifiante	Pertinence	Le corpus n'existe que pour certains objectifs et certaines hypothèses.
	Cohérence	Le corpus sous-entend une perspective donnée.
Acceptabilité	Représentativité	Le corpus est une représentation du réel échantillonnée et vraisemblable.
	Régularité	Le corpus ne doit pas être perturbé par des contraintes externes.
	Complétude	Le corpus doit permettre un niveau d'analyse suffisamment riche.
Exploitabilité	Homogénéité	Le corpus doit pouvoir fournir des unités comparables.
	Volume	Le corpus doit présenter une taille suffisante pour permettre les analyses.

Tableau. Quand un ensemble de textes devient corpus (d'après Bommier-Pincemin 1999).

Après avoir fixé des hypothèses (généralement à l'aune de la littérature ou d'une connaissance de l'objet ou du site d'étude), le chercheur construit son corpus (sauf, évidemment, dans les cas où il travaille sur un corpus déjà construit ou que, adepte de la linguistique *corpus-driven*, il fasse émerger sans *a priori* des données de son corpus). Ce corpus deviendra ensuite le lieu de l'analyse et joue ainsi un rôle central dans les résultats et dans les interprétations. Un retour sur

les hypothèses peut donc amener à reconsidérer son corpus : un corpus n'est ni fini, ni figé et peut être amené à évoluer (par enrichissement, mais aussi par simplification). Un corpus n'est donc pas un acquis, mais un construit dont l'élaboration semble parallèle à la constitution des problématiques de recherche.

Ce détour par la linguistique permet d'établir quelques critères pour l'appréhension du corpus. Avec l'émergence de l'analyse du discours dans les années 1970, la création du corpus est, en effet, devenue un questionnement central (Guilhaumou 2002). Son intérêt pour la linguistique est alors méthodologique, mais aussi épistémologique. Ainsi, l'intérêt des méthodes linguistiques

redouble, au contraire, si l'on considère les textes non seulement comme des témoignages institutionnels, mais encore comme des traces involontaires d'une activité qui les déborde ; si on leur pose, indépendamment de celle de leur sens, la question de ce qu'ils révèlent ou trahissent, malgré eux, des activités qui les ont produites. L'approche linguistique s'impose alors pour découvrir, en plus du sens explicite, au premier degré, un sens second, implicite et masqué. (Prost 1996, p. 270)

...aux sciences sociales

Pourquoi le corpus devient-il un terrain d'étude ? La documentation écrite et les transcriptions d'entretiens oraux sont au cœur des travaux des humanités et des sciences humaines et sociales : comment organiser, traiter et délimiter ces collections de textes ?

Antoine Prost (1996) identifie trois critères pour établir un corpus, chacun remplissant un objectif particulier : il est fondé sur le contraste pour permettre des comparaisons, il s'avère diachronique pour favoriser l'étude des continuités ou des bifurcations, et il est composé de textes vus comme significatifs pour présenter une pertinence. Damon Mayaffre (2002) propose lui aussi trois caractéristiques : si la contrastivité et la diachronicité sont maintenues, l'homogénéité est ajoutée. Ainsi, un corpus permet d'étudier des contrastes (spatiaux, temporels, dus aux acteurs...) à partir d'une base commune. De même, Francis Chateauraynaud (2003) propose trois règles méthodologiques : prendre au sérieux les discours dans leur explicite, sans oublier l'implicite ; considérer le rôle des rassemblements de textes pour éclairer ces mêmes textes selon des logiques de convergences ou d'oppositions ; et appréhender les sens des formes uniquement par le contexte où elles s'insèrent. Le propre du corpus est d'inscrire le discours dans la comparaison et de déterminer des invariants, des points communs, des différences, voire des contraires. Discours et corpus n'existent qu'en contexte :

Aucune étude qui entend appréhender les discours ou les textes dans leur dimension sociale ou politique n'est indifférente à la question des corpus pour la raison suffisante qu'une production linguistique revêt une dimension politique seulement si elle se trouve effectivement attestée dans la société. (Mayaffre 2005)

Damon Mayaffre décrit ces corpus comme « référentiels », c'est-à-dire comme des liens ou des allusions constants avec le monde parfois dit « réel ». En effet, le corpus intéresse les sciences humaines et sociales en tant que matériel situé spatialement, socialement et temporellement, et donc en contexte.

Enfin, réfléchir sur son corpus est aussi une manière d'appréhender son discours de chercheur dans une perspective de réflexivité : « texte et contexte se retrouvent dans un même espace construit, mais aussi sous la forme de corpus co-construits à l'intérieur desquels le point de vue du chercheur est l'une des données majeures de la construction de ces corpus » (Guilhaumou 2002). Il semble alors pertinent de questionner cette interaction entre texte et contexte (à la fois d'énonciation mais aussi de constitution du corpus).

Le corpus, une construction à géométrie variable

La clôture du corpus ne peut plus être aléatoire ni seulement d'ordre quantitatif ; des contraintes qualitatives viennent s'ajouter, le corpus est alors de l'ordre des *données pertinentes*. Par ailleurs la décision de garder le corpus ouvert a pour corollaire l'implication plus franche du linguiste dans le modelage de celui-ci ; le corpus est alors de l'ordre des *données construites*. (Dalbera 2002)

Le corpus est donc une base de données qui regroupe des matériaux de recherche. Même s'il se prête à des mesures, il n'est qu'une partie d'un tout et sous-entend la prise de conscience de ses biais et de ses limites. Admettre la présence d'un corpus, c'est donc reconnaître l'approche d'une représentation, qui n'est qu'une des représentations possibles de la réalité. Néanmoins, de telles réflexions préalables permettent aussi de raisonner sur les discours à travers le concept d'échelle.

La clôture du corpus ou les limites du corpus

D'après Damon Mayaffre (2002), la perspective sérielle sous-entend la clôture du corpus : où un corpus doit-il commencer et s'arrêter ? L'utilisation du concept de clôture en linguistique ne semble pas vraiment transférable aux sciences sociales où le terme de *limite* semble préférable pour insister sur le flou et la mouvance de son découpage.

Le choix de l'échantillonnage peut proposer une solution. Cette pratique consiste à ne couvrir que certains espaces, que certaines périodes, que certains auteurs... La majorité des corpus apparaissent comme des échantillons : il semble très difficile de réunir l'ensemble des textes sur une thématique donnée, excepté dans le cadre de monographies très spécifiques ou de biographies s'appuyant sur des pas de temps précis. « Le travail sur échantillon se révèle donc

être un impératif pratique. Mais pour que l'analyse prétende à quelque validité, on ne saurait se contenter d'un échantillon aléatoire » (Dalbera 2002). Ainsi, toute étude sous-entend une technique pour recueillir un ensemble cohérent de discours, mais ces situations complexes donnent naissance à une multiplicité de débats et donc de sources. L'échantillon pose alors la question de la représentativité :

[...]l'enjeu n'est plus celui de l'exhaustivité, mais celui de la représentativité. Il s'agit alors de constituer des échantillons représentatifs d'une réalité plus large – en statistique on dirait : d'une population. Les objectifs d'une telle entreprise peuvent être assez variés : appréhender et donner à voir cette réalité trop vaste pour être embrassée dans sa totalité. (Mellet 2002)

En effet, de nombreuses études en sciences sociales portent sur des réalités trop complexes pour parvenir à une exhaustivité des sources. Il s'agit alors de questionner son corpus et de faire apparaître ses biais pour en faire un terrain de recherches efficient, un lieu d'observation représentatif ou emblématique de certaines logiques.

Dans une approche constructiviste, le chercheur se doit d'interroger les modalités de constitution d'un corpus qui n'est pas un donné ou un préexistant. De plus, selon une logique systémique, le corpus met en interaction les formes matérielles et immatérielles du discours, via la figure concrète de textes.

Corpus, sous-corpus et hors-corpus : les échelles du corpus

« Parler d'échelle, c'est justement admettre qu'autre chose que la taille change quand change la taille » (Lévy 2013, p. 308-309). Cet extrait de la définition d'échelle, vue comme « un rapport de taille entre réalités », semble pouvoir s'appliquer au corpus. En effet, les discours, en tant que textes mis en contexte, présentent des visages différents selon les tailles considérées, et notamment l'ampleur du contexte pris en compte. Les discours sont tantôt découpés, tantôt agrégés, ce qui peut influencer les résultats et les interprétations des analyses.

Partant de l'ensemble ouvert des possibles, le chercheur construit son corpus en procédant à une réduction linguistique : pour être représentatif des hypothèses de recherche, le corpus d'étude résulte souvent d'un échantillonnage, effectué au sein de la population, des textes susceptibles d'être rassemblés (Tournier 1993). Bien plus, il importe de garder en tête que ce discours constatable ne constitue qu'une partie des discours avérés, une partie significative pouvant être perdue ou rester inaccessible.

Selon les objectifs poursuivis, le corpus peut donner naissance à une pluralité de corpus (figure). Bénédicte Bommier-Pincemin (1999, p. 420) évoque des « emboîtements ». Il n'y aurait pas un corpus unique mais bien différentes entités qualifiables de corpus. Le corpus existant correspond à l'ensemble des

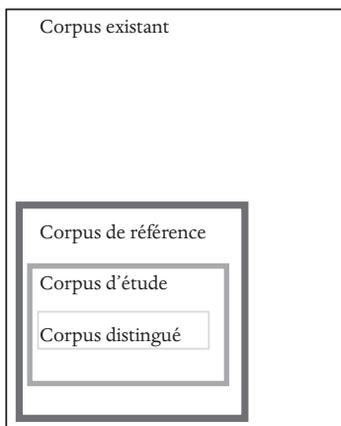


Figure. Quand les objectifs définissent l'échelle considérée des corpus (d'après Bommier-Pincemin 1999).

textes disponibles, ensemble flou et généralement non matérialisé pour des questions de faisabilité ou d'accessibilité. Le corpus de référence est celui qui inclut les textes analysés ainsi que des éléments de contexte. Le corpus d'étude a pour objectif de répondre aux questionnements de la recherche. Le terme de *corpus* utilisé sans qualificatif ou sans complément regroupe généralement les perspectives discursives du corpus de référence et d'étude. Enfin, le corpus distingué peut aussi être appelé sous-corpus.

Le corpus est donc constitué d'un ensemble de textes et d'images (le corpus d'étude), mais aussi du cadre de pensée qui favorise sa création (le corpus de référence). Le corpus existant peut aussi être qualifié d'archive (vue comme l'ensemble de documents accessibles), mais François Rastier (2004) considère alors qu'il ne faut pas parler de corpus : l'archive ne correspond ni aux qualités ni aux objectifs intrinsèques aux corpus. Des corpus plus restreints sont souvent nommés sous-corpus. Le postulat sous-jacent est qu'une réduction de la quantité de données est nécessaire pour des raisons de faisabilité, ou que le sous-corpus présente des particularités, sa propre cohérence et une homogénéité, et nécessite donc d'être analysé comme un objet en soi, comme un corpus. En effet, le sous-corpus est la portion d'un corpus principal, mais le corpus est lui-même une partie d'un tout généralement inaccessible (l'ensemble de la documentation qui constitue le corpus existant). Le nombre de sous-corpus semble quasiment illimité, même si un sous-corpus se doit de répondre aux attendus d'un corpus (cohérence, pertinence, diachronie, contraste...). À une grande échelle, le corpus peut donc être propice à des zooms : cette technique sous-entend le recours à des sous-corpus. Ainsi, le sous-corpus correspond à un ensemble cohérent de textes, une partition du corpus total ; en linguistique, il peut correspondre par exemple aux quatre facteurs de variation de la langue

généralement admis (variation diachronique, dialectale, sociale ou selon le type d'interaction engagé dans la production de l'énoncé), ce qui permet d'appréhender la langue de façon contrastive et de mettre au jour des variations, distinctes des invariants de la langue. Une forme particulière de sous-corpus est l'anti-corpus, soit un sous-corpus créé autour de l'exclusion d'une caractéristique (Chateauraynaud 2003). Ce sous-corpus reprend toutes les caractéristiques du corpus : cette partie de corpus n'est qu'une autre échelle d'analyse du tout, notamment pour développer des lectures plus fines. Ce découpage du corpus rejoint sa dimension spatiale (sa taille ou sa longueur), mais aussi sa dimension temporelle (en conservant ou non l'ampleur du pas de temps initialement étudié dans le corpus) ou thématique (en proposant des focales).

À une petite échelle,

on ne pouvait comprendre un mot sans la phrase et la phrase sans le discours, on ne pouvait comprendre le discours sans l'interdiscours, le texte sans le co-texte (sans même parler ici du hors-texte), c'est-à-dire aussi et de manière plus générale, le corpus sans le hors-corpus. (Mayaffre 2002)

Cette question du contexte et du hors-corpus semble importante : elle fonde le corpus et les résultats de son étude. Selon Maurice Tournier (1980, p. 200), « les variables présentes à l'entrée et reconnues valides constituent en effet les interprétants à venir [...]. L'interprétation n'est qu'une remontée épurée de l'énonciation ». Cette échelle plus globale que celle du corpus est cruciale dans la structuration du corpus mais aussi dans les investigations menées sur le corpus : « déterminer le contexte est bien sûr un acte herméneutique majeur, puisque c'est décider ce qui est accessible et même structurant, pour l'interprétation du texte, *en dernière instance* » (Bommier-Pincemin 1999, p. 422). Même si certains linguistes peuvent n'étudier que le plan de la séquence linguistique sans évoquer le niveau discursif, le contexte comme choix se positionne donc entre objectivité et subjectivité.

Cette réflexion sur les échelles du corpus se double d'un questionnement similaire à l'échelle du texte, qui concerne ses délimitations et la prise en compte de l'« ouvert du texte » qui correspond aux « éléments jusqu'ici désignés comme plus ou moins "extérieurs" à lui (péritexte, épitexte, intertexte, hypertexte, etc.) » (Viprey et Adam 2009). Cet « ouvert » peut être incorporé via des codages sur ou hors le texte. Ainsi, Yves-François Le Lay identifie deux échelles : « un texte gagne à être étudié d'après des échelles distinctes, tantôt infra-textuelle – celles des caractères, des mots, des phrases, des paragraphes – tantôt supra-textuelles ou contextuelles, à la faveur des notions d'intertextualité ou de situation d'énonciation par exemple » (Le Lay 2013, p. 3). Cette approche supra-textuelle correspond à une approche discursive, quand l'approche infra-textuelle se concentre sur les éléments constitutifs du texte.

★

Au total, un corpus sous-entend des données protéiformes rassemblées en une entité, le caractère protéiforme étant maîtrisé par les métadonnées, l'étiquetage et la hiérarchisation (par exemple en sous-corpus) :

Un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications. (Rastier 2004)

Ces deux points sont en relation : les applications envisagées à partir du corpus vont générer des enrichissements – comme l'étiquetage ou le codage de certaines variables – et vont donc engendrer des modalités de structuration du corpus. En effet, ce point n'a pas encore été abordé, mais semble fondamental : un corpus est un ensemble de textes structurés et dont les modalités d'organisation sont généralement lisibles et explicables. Francis Chateauraynaud (2003) considère que parmi les différents niveaux d'informations nécessaires pour traiter des textes, trois ont été conservés comme dignes d'intérêt en sciences sociales. Le premier niveau s'inscrit dans une perspective sérielle à travers la date et l'auteur. Le deuxième niveau conditionne des approches thématiques à travers les thèmes mais aussi les acteurs impliqués. Le troisième niveau questionne des réseaux de significations entre acteurs et thèmes, les points de vue ou les jugements adoptés (notamment à travers les agencements ou les formes stylistiques utilisées). Ainsi, construire un corpus sous-entend aussi l'élaboration et la réflexion sur les métadonnées, grilles de structuration d'une information. Penser par corpus permet aussi d'organiser les données.

Le corpus sera appréhendé dans un premier temps dans son utilisation linguistique : les contributions de Catherine Pinon et Soufiane Lanseur utilisent chacune un corpus construit comme lieu d'exploration de la langue afin de dégager des régularités d'emploi d'un verbe arabe pour la première, ou des nouveautés lexicales relatives au domaine de l'économie pour la seconde. Dans un deuxième temps, le corpus sera étudié dans sa dimension temporelle, que ce soit pour appréhender les évolutions historiques d'un discours de presse au fil des jours (Natacha Souillard *et al.* à propos des révolutions arabes, Émeline Comby sur les pollutions aux PCB) ou au contraire en tant qu'ensemble de documents permettant d'avoir accès à un passé révolu (Nadia Zidelmal et Azeddine Belakehal). Dans un troisième temps, le corpus sera exploité comme lieu d'observation de représentations sociales incarnées dans le discours : qu'il s'agisse de propos de professionnels sur leur activité (Yves-François Le Lay *et al.*), du discours de la presse sur des problèmes de santé adolescente (Lise Jacques et Audrey Arnoult) ou de textes relevant d'un ensemble d'auteurs ayant les mêmes conceptions politico-scientifiques (Konstantinos Delimitsos), les éléments réunis en corpus permettent de dégager des représentations qui portent sur, mais aussi modifient parfois, le contexte social.

Références

- ADAM Jean-Michel et VIPREY Jean-Marie, 2009, « Corpus de textes, textes en corpus. Problématique et présentation », *Corpus*, n° 8. En ligne : [<http://corpus.revues.org/1672>] (consulté le 13 mars 2015).
- BOMMIER-PINCEMIN Bénédicte, 1999, *Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle par la caractérisation des destinataires des documents*, thèse de doctorat de l'Université Paris IV-Sorbonne.
- CHARAUDEAU Patrick, 2009, « Dis-moi quel est ton corpus, je te dirai quelle est ta problématique », *Corpus*, n° 2. En ligne : [<http://corpus.revues.org/1674>] (consulté le 13 mars 2015).
- CHATEAURAYNAUD Francis, 2003, *Prospero. Une technologie littéraire pour sciences humaines*, Paris, CNRS éditions, Paris, 404 p.
- DALBERA Jean-Philippe, 2002, « Le corpus entre données, analyse et théorie », *Corpus*, n° 1. En ligne : [<http://corpus.revues.org/10>] (consulté le 13 mars 2015).
- DUTEIL-MOUGEL Carine, 2005, « Les mécanismes persuasifs des textes politiques », *Corpus*, n° 4. En ligne : [<http://corpus.revues.org/357>] (consulté le 13 mars 2015).
- GUILHAUMOU Jacques, 2002, « Le corpus en analyse de discours : perspective historique », *Corpus*, n° 1. En ligne : [<http://corpus.revues.org/8>] (consulté le 13 mars 2015).
- JACQUES Marie-Paul, 2005, « Pourquoi une linguistique de corpus? », *La linguistique de corpus*, G. Williams éd, Rennes, Presses universitaires de Rennes (Rivages linguistiques), p. 21–30.
- JAUBERT Anna, 2002, « Corpus et champs disciplinaires. Le rôle du point de vue », *Corpus*, n° 1. En ligne : [<http://corpus.revues.org/13>] (consulté le 15 mars 2015).
- LAHIRE Bernard, 1994, « Remarques sociologiques sur le *linguistic turn*. Suite au "Dialogue sur l'espace public" entre Keith M. Baker et Roger Chartier », *Polix*, n° 7, p. 189-192.
- LAKS Bernard, 2010, « La linguistique des usages: de l'*exemplum* au *datum* », *L'exemple et le corpus. Quel statut?*, P. Cappeau et al. éd., Rennes, Presses universitaires de Rennes, p. 13-28.
- MAYAFFRE Damon, 2002, « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus*, n° 1. En ligne : [<http://corpus.revues.org/11>] (consulté le 13 mars 2015).
- , 2005, « Rôle et place des corpus en linguistique : réflexions introductives », *Texto!*, vol. 10. En ligne : [http://www.revue-texto.net/Corpus/Publications/Mayaffre_Corpus.html] (consulté le 2 mars 2015).
- MELLET Sylvie, 2002, « Corpus et recherches linguistiques », *Corpus*, n° 1. En ligne : [<http://corpus.revues.org/7>] (consulté le 19 juillet 2016).
- PINCEMIN Bénédicte, 2011, « Sémantique interprétative et textométrie – Version abrégée », *Corpus*, n° 10. En ligne : [<http://corpus.revues.org/2121>] (consulté le 13 mars 2015).
- PLANCHON Philippe, 2010, « Paramètres de variation et modalités d'articulation entre l'exemple et le corpus », *L'exemple et le corpus. Quel statut?*, P. Cappeau et al. éd., Rennes, Presses universitaires de Rennes, p. 43-58.
- PROST Antoine, 1996, « Les mots », *Pour une histoire politique*, R. Rémond éd., Paris, Seuil, p. 255-285.
- RASTIER François, 2004, « Enjeux épistémologiques de la linguistique de corpus », *Texto!*, vol. 9. En ligne : [http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html] (consulté le 2 mars 2015).
- , 2011, *La mesure et le grain: sémantique de corpus*, Paris, Champion, Paris, 272 p.

- TOURNIER Maurice, 1980, «D'où viennent les fréquences de vocabulaire? La lexicométrie et ses modèles», *Mots*, n° 1, p. 189-209.
- , 1993, *Lexicometria*, Lisbonne, Universidade Aberta, 127 p.
- WILLIAMS Geoffrey, 2005, «Introduction», *La linguistique de corpus*, G. Williams éd., Rennes, Presses universitaires de Rennes (Rivages linguistiques), p. 13-18.

ENS ÉDITIONS