

Les traducteurs sont-ils des linguistes comme les autres ? L'intégration des outils de corpus dans la formation des futurs traducteurs

Rudy LOOCK¹²

rudy.loock@univ-lille.fr

La question posée dans le titre de notre article – *Les traducteurs sont-ils des linguistes comme les autres ?* – est volontairement provocatrice. Dans le cadre de colloques et de conférences, nous avons en effet pu régulièrement assister à des échanges et à des débats parfois passionnés sur cette question. Si le traducteur professionnel est sans aucun doute un expert linguistique de ses langues de travail, il est néanmoins possible de se demander dans quelle mesure il partage l'approche de la langue du chercheur en linguistique. Cette question est en fait loin d'être anodine et se pose de façon cruciale lorsqu'il s'agit de s'interroger sur la façon de former les traducteurs de demain : quelle place les méthodes d'analyse linguistique doivent-elles occuper au sein des formations ?

En particulier, nous souhaitons nous pencher ici sur la façon dont la compilation et l'exploitation de corpus électroniques, autrement dit de bases de données linguistiques interrogeables de façon automatique via une interface en ligne ou hors ligne, peuvent intégrer la liste des compétences pertinentes pour les futurs traducteurs professionnels. Un certain nombre de formations en traduction ont en effet intégré un enseignement spécifique à l'utilisation des outils de corpus : Frérot et Karagouch (2016), s'appuyant sur différentes enquêtes aux niveaux national français et européen, avancent en effet qu'environ deux tiers des formations proposent de tels enseignements, même si les pratiques sont parfois très différentes. De même, les programmes de formation continue, écoles d'été, ou ateliers à destination des professionnels en activité sur cette question semblent se multiplier, jusqu'au sein de la Direction générale de la traduction de la Commission européenne. Or, notre expérience de formateur à l'Université de Lille au sein d'un programme de master en traduction, en l'occurrence le master « Traduction Spécialisée Multilingue », mais aussi dans le cadre d'écoles d'été ou d'ateliers spécifiques pour traducteurs en exercice, nous a montré que l'appropriation de ces outils, et surtout la démarche linguistique qui l'accompagne, ne vont pas de soi. En particulier, le fait qu'un corpus électronique ne s'utilise pas comme un dictionnaire ou un glossaire, mais exige une démarche autre qui fait la part belle à la formulation d'hypothèses que l'on teste auprès de données attestées et donc sujettes à variation, n'est pas sans poser problème. Il nous semble que l'origine de la « méfiance », même si les choses changent rapidement avec notamment l'intégration désormais possible, au moins partiellement, des outils de corpus aux logiciels de TAO (Tagne & Loock, 2017), se situe justement au niveau de l'approche vis-à-vis de la langue : le chercheur en linguistique adopte une approche que l'on qualifie de descriptive ; à l'inverse le traducteur est à la recherche d'outils d'aide à la prise de décision à différents niveaux linguistiques (terminologie, phraséologie, syntaxe, organisation discursive) et en cela, il adopte une

¹ Université de Lille & UMR « Savoirs, Textes, Langage » (STL) du CNRS.

² Rudy Loock est Professeur des Universités au sein de l'UFR des Langues Etrangères Appliquées de l'Université de Lille, où il est responsable du programme de master « Traduction Spécialisée Multilingue ». Il est par ailleurs affilié à l'UMR « Savoirs, Textes, Langage » du CNRS. Ses recherches portent sur la traductologie de corpus, notamment sur l'utilisation des corpus comme outils d'aide à la traduction et comme outils de recherche en traductologie, l'enseignement de la traduction ainsi que sur la qualité des textes traduits. Il a publié en 2016 *La Traductologie de corpus* aux Presses Universitaires du Septentrion.

approche qui est davantage prescriptive. En consultant des données réunies en corpus, le traducteur et le linguiste n'auront donc pas la même façon d'aborder les choses, et il nous semble que pour que le traducteur s'empare des outils de corpus comme outils d'aide à la traduction, une approche spécifique est nécessaire, pour la compilation comme l'exploitation, qui prenne en compte les besoins et les attentes de la profession : fiabilité des informations collectées, rapidité d'obtention de la réponse à une question donnée, le tout pour une rentabilité satisfaisante, car n'oublions pas que le traducteur reste majoritairement payé au nombre de mots qu'il traduit.

Notre article est organisé de la façon suivante. Dans un premier temps, nous passons en revue les différents types de corpus électroniques présents dans la boîte à outils du traducteur, de façon « cachée » (c'est-à-dire au sein d'autres outils) ou utilisés indépendamment comme outils d'aide à la traduction, afin de montrer leur omniprésence, puis expliquons rapidement le type d'informations qu'ils peuvent fournir. Nous présentons ensuite les différents enjeux en lien avec leur enseignement (décloisonnement, complémentarité, maîtrise technique vs. utilisation en contexte, techniques de compilation, ergonomie...) sans perdre de vue la question posée en début d'article : quelle(s) spécificité(s) pour le traducteur par rapport au chercheur en linguistique ? Enfin, nous abordons la question de l'approche linguistique qu'implique une telle exploitation de bases de données linguistiques, avec une réflexion sur les enjeux liés à une approche empirique de la langue, qui explique selon nous pourquoi l'appropriation de ces outils par les traducteurs n'est pas si aisée.

Cet article est avant tout un retour sur expérience vis-à-vis de nos enseignements consacrés à l'utilisation des outils de corpus auprès de futurs traducteurs ou de traducteurs en exercice.

1. Omniprésence des outils de corpus dans le cadre des outils d'aide à la traduction

L'exploitation de bases de données linguistiques pour la traduction n'est pas chose nouvelle, mais son histoire est un peu complexe et certainement pas linéaire. Avec les développements de l'informatique à la fin du 20^{ème} siècle, deux technologies se sont développées sans toutefois converger (la convergence est seulement en train de se faire) : la traduction automatique (TA), d'abord à base de règles mais assez rapidement statistique, et les mémoires de traduction au sein des logiciels d'aide à la traduction (également appelés outils de traduction assistée par ordinateur, ou TAO). Dans les deux cas, l'exploitation de corpus électroniques est indispensable : dans le premier il s'agit d'associer des données, bilingues ou monolingues, à un algorithme permettant de déterminer la traduction la plus probable ; dans le second d'exploiter d'anciennes traductions alignées au niveau de la phrase afin qu'un logiciel puisse rechercher des correspondances, totales ou partielles, à proposer au traducteur lors de la traduction d'un nouveau texte. Dans les deux cas également, et de façon un peu paradoxale, l'appellation « corpus » n'est jamais, ou de façon très marginale, utilisée : les mémoires de traduction ne sont par exemple rien d'autre que des corpus parallèles bilingues alignés au niveau de la phrase ; la TA statistique, et depuis peu neuronale, exploite également de (très) grandes quantités de données linguistiques, bilingues et parfois monolingues, mais il reste rare que le terme « corpus » soit mentionné alors que c'est bien de cela qu'il s'agit. Le non recours à l'appellation « corpus » ne poserait pas de problème en soi si cela ne débouchait pas sur un manque de réflexion sur la façon de compiler et d'exploiter des bases de données linguistiques dans le cadre d'un projet de traduction, et si pour revenir à notre problématique, elle ne sous-entendait pas une dichotomie entre l'approche du chercheur en linguistique et celle du traducteur. Or, dans ce cas comme ailleurs, la pertinence et la qualité des données sont des critères essentiels, tout comme une réflexion sur la façon la plus pertinente de les

interroger. Par ailleurs, il est important de faire remarquer que le développement du recours à des corpus électroniques au sein des logiciels de TAO et des moteurs de traduction automatique n'est pas en lien direct avec le développement des corpus électroniques en linguistique dans la seconde moitié du siècle dernier (*cf.* Somers, 2003 ou Garcia, 2005 pour un historique des mémoires de traduction). Il y a donc là une forme de paradoxe : les corpus électroniques sont bien présents dans la boîte à outils du traducteur professionnel, jusque dans les dictionnaires en ligne proposant des traductions en contexte (www.linguee.com ou <http://context.reverso.net/traduction/>), mais pas toujours visibles, et nombreux sont les traducteurs, lorsqu'on leur demande s'ils ont recours à des corpus dans leur pratique quotidienne, qui répondent par la négative alors qu'ils utilisent bien ces différents outils.

Au-delà de leur présence au sein des logiciels de TAO, des outils de TA ou encore de dictionnaires en ligne, les corpus électroniques peuvent être exploités de façon indépendante par les traducteurs, par le biais d'une interface en ligne (*cf.* les corpus proposés par la Brigham Young University, <https://corpus.byu.edu/>, ou le portail Sketch Engine, <https://www.sketchengine.co.uk/>) ou par le biais de programmes informatiques hors ligne (concordanciers). Leur utilisation est aujourd'hui très répandue dans les milieux universitaires, où compilation et exploitation de corpus électroniques font partie des compétences enseignées (*cf. inter alia* Beeby, Rodríguez et Sánchez-Gijón, 2009 ; Bowker, 1998, 1999 ; Bowker & Pearson, 2002 ; Pearson, 2003 ; Frankenberg-Garcia, 2015 ; Kübler, 2003, 2011a, 2011b ; Looock, 2016a, 2016b ; Varantola, 2003 ; Zanettin, 2012 ; Zanettin, Bernardini et Stewart, 2003). Ces corpus peuvent être des corpus monolingues de langue source ou de langue cible, ou encore des corpus dits « comparables » composés de différents sous-corpus de langues originales afin d'effectuer des comparaisons inter-langagières, de corpus dits « parallèles » contenant des textes originaux et leur(s) traduction(s), voire dans le cadre de réflexions traductologiques de corpus comparables composés cette fois de sous-corpus permettant la comparaison entre textes originaux et traduits pour la même langue (comparaison intra-langagière).³ Nous n'entrons pas dans le détail ici car nous l'avons fait ailleurs (voir Looock, 2016a/b ; voir également sur ce sujet Kübler & Aston, 2010), mais les objectifs sont multiples : amélioration de la compréhension du texte source en langue étrangère dans ses moindres détails, recherche d'invisibilité en langue cible dans le cas de textes spécialisés afin de mettre au jour la terminologie mais aussi la phraséologie et l'organisation du discours par des experts du domaine concerné, mise au jour de différences d'usage entre langue source et langue cible, création de glossaires multilingues, vérification de l'homogénéité linguistique entre textes traduits et textes originaux. Les expériences ont été et sont encore nombreuses en la matière, avec des résultats tout à fait probants (*cf.* p. ex. Kübler, Mestivier, Pecman et Zimina, 2016 ; Frérot, 2016). Le recours à des corpus électroniques dans le cadre des formations a d'une façon générale tendance à se démocratiser dans le cadre de l'apprentissage de langues étrangères (Boulton & Tyne, 2014) : on parle alors de *data-driven learning* (Johns, 1990).

Les corpus électroniques, prêts à l'emploi ou à compiler soi-même pour reprendre la distinction *manufactured* vs. *DIY* (pour *do-it-yourself*) de Bernardini et Ferraresi (2013), ont donc toute leur place dans la vie quotidienne du traducteur, utilisés au sein d'outils ou de façon indépendante. En ce sens, nous les considérons comme des outils d'aide à la traduction et donc comme relevant de la traduction assistée par ordinateur, que nous définissons de façon

³ La distinction que nous opérons ici entre corpus comparables et corpus parallèles est assez classique, même si la littérature entretient parfois un certain flou sur le sujet. Les premiers sont des corpus de données qui appartiennent à deux langues distinctes ou à deux variétés de la même langue, tandis que les seconds, parfois également appelés corpus de traduction, sont des corpus bilingues/multilingues qui contiennent textes originaux et leur(s) traduction(s) dans une ou plusieurs langue(s). Nous renvoyons le lecteur à Looock (2016a : 85-86) pour plus d'informations sur cette distinction.

large, à l'instar de Frérot et Karagouch (2016 : 1), pour qui est outil de TAO « tout outil informatique mis à la disposition des professionnels de la traduction et facilitant leur travail ». Cette façon de présenter les choses nous paraît importante, que ce soit auprès des traducteurs ou des futurs traducteurs, afin que ceux-ci puissent s'appropriier ces outils. En revanche, l'approche de la langue par le biais de ces outils nous semble spécifique, d'où la question soulevée ici.

2. Comment intégrer l'enseignement des outils de corpus au sein des formations ?

Si la quasi-totalité des formations au niveau européen forme les futurs traducteurs aux logiciels de TAO et à l'utilisation des mémoires de traduction, ce n'est pas tout à fait le cas pour les outils de corpus. Frérot et Karagouch (2016) dressent à ce sujet un état des lieux intéressant de l'enseignement des outils de corpus au sein des formations, françaises (enquête réalisée auprès des membres de l'AFFUMT, Association française des formations universitaires aux métiers de la traduction) et européennes (enquêtes Optimale et réseau EMT, *European Masters' in Translation*, de la Commission européenne) : 60 % des répondants intègrent ce type d'enseignement, proportion qui monte à 75% pour les formations françaises membres de l'AFFUMT. Néanmoins, les pratiques semblent très diverses. Il y a également le risque, comme pour tout outil informatique, de focaliser l'enseignement sur le développement de la compétence technique : il convient en effet de distinguer, à l'instar de Beeby et al. (2009), « *corpus use for learning to translate* » et « *learning corpus use to translate* ». Il n'en reste pas moins que la capacité à gérer des bases de données linguistiques, autrement dit des corpus, fait partie des compétences actuelles à acquérir pour le futur traducteur : la nouvelle roue de compétences du réseau EMT, publiée fin 2017, mentionne explicitement cette compétence au sein des compétences technologiques (compétence numéro 16) : « *Make effective use of search engines, corpus-based tools, text analysis tools and CAT tools* » (Toudic & Krause, 2017 : 9).

Dans cette partie nous souhaitons, à travers des exemples précis expérimentés avec nos étudiants au sein du programme de master « Traduction Spécialisée Multilingue » de l'Université de Lille,⁴ exposer les enjeux de l'utilisation de corpus électroniques avec de futurs traducteurs : (i) la nécessité de « décloisonner » cet enseignement, (ii) l'importance de présenter cet outil comme un outil complémentaire permettant de résoudre des problèmes de traduction non résolus par ailleurs, (iii) les enjeux liés à la pertinence des données, (iv) les enjeux liés à l'exploitation, et enfin (v) la nécessité de proposer une méthode de compilation rapide et une approche ergonomique de ces outils, question souvent négligée. Afin d'illustrer notre propos, nous exploiterons différents corpus « maison », encore appelés DIY (*do-it-yourself*), compilés soit manuellement soit de façon semi-automatisée à l'aide du logiciel BootCaT (*Bootstrap Corpora And Terms from the Web*), un logiciel *front-end* permettant de récupérer des données du web à partir de mots-clefs (Baroni & Bernardini, 2004), afin de résoudre des problèmes dans le cadre de la traduction de textes spécialisés (économie, informatique) rédigés en anglais vers le français.

2.1. Le décloisonnement

Pour que les étudiants puissent s'approprier les outils de corpus comme outils d'aide à la traduction, il importe selon nous de ne pas isoler un tel enseignement, ce qui est pourtant le risque au sein des structures d'enseignement telles qu'elles sont proposées aujourd'hui au sein

⁴ L'enseignement des outils de corpus pour la traduction est un enseignement continu sur les deux années du master (niveau 1 en première année, niveau 2 en seconde année). La première année est consacrée aux questions d'ordre technique et en lien avec l'exploitation comme outils d'aide à la traduction ; la seconde année est consacrée à leur exploitation dans le cadre de la recherche en traductologie. Les expériences décrites ici relèvent du niveau 1.

des universités (chaque enseignement doit en effet faire l'objet d'une évaluation spécifique, donnant droit à l'attribution de crédits européens [ECTS]). Si une présentation théorique reste importante et ne peut être évitée (définitions des concepts clés comme représentativité, échantillonnage, annotation, bruit, silence...), le lien avec d'autres enseignements nous semble primordial. Ainsi, l'application concrète dans un cours de traduction pratique est une bonne façon de sensibiliser les étudiants à l'utilité des outils de corpus. D'après notre expérience, il convient de choisir un texte suffisamment spécialisé pour que la thématique abordée ne soit que vaguement connue des étudiants, ce qui demandera une recherche documentaire importante. Nous prenons ici l'exemple d'un document sur la politique de la Banque centrale européenne en termes de rachats d'actifs, pour lequel nous avons proposé un déclouisonnement entre trois différents enseignements : (i) le cours de terminologie, (ii) le cours sur les corpus, et (iii) le cours de traduction pratique anglais-français, le premier étant assuré par une collègue traductrice professionnelle⁵ et les deux autres par nous-mêmes. Dans un premier temps, les étudiants ont effectué des recherches terminologiques : après avoir repéré les termes anglais qui leur semblaient relever de la terminologie économique, ils ont recherché les équivalents en langue cible (français) au moyen de glossaires en ligne spécialisés (p. ex. IATE ; Termium Plus) ou de dictionnaires en ligne monolingues et bilingues, ce qui leur a permis de compiler un glossaire bilingue détaillé (terme en langue source, domaine, équivalent en langue cible, catégorie grammaticale, définition, source, date de la recherche). L'objectif était de préparer le terrain pour la traduction à venir, mais d'un point de vue pédagogique, un autre objectif était de leur montrer l'insuffisance de la compilation d'un simple glossaire bilingue et de l'importance de la prise en compte du contexte pour déterminer l'équivalent en langue cible mais aussi de la phraséologie (avoir trouvé le bon équivalent ne signifie pas que l'on sache l'utiliser au sein d'une phrase). C'est là qu'intervient la sensibilisation à l'utilité d'une base de données linguistiques (de qualité), autrement dit de corpus, permettant de vérifier la pertinence des équivalents trouvés et leur utilisation en contexte.

L'insuffisance du glossaire se démontre assez aisément lors de la traduction du texte (cours de traduction pratique) : le glossaire fournit des équivalents mais ne permet pas d'utiliser ces équivalents en contexte. Il est alors possible de faire appel aux compétences développées dans le cadre d'un cours sur les outils de corpus, en compilant et en exploitant par exemple un ensemble de textes portant sur le même sujet (politique de la BCE en termes de rachat d'actifs). Ceci peut se faire en rassemblant de façon manuelle un ensemble de textes publiés sur l'internet de bonne qualité sur le même sujet (il peut s'agir de rapports économiques publiés sur le site du Fonds monétaire international par exemple, ou encore des textes issus de la presse spécialisée), ou bien en compilant un corpus de façon automatisée grâce au logiciel BootCaT à l'aide de mots clés. Dans les deux cas, le corpus sera à exploiter à l'aide d'un concordancier, programme informatique permettant d'interroger hors ligne un ensemble de documents réunis en corpus. Celui que nous avons retenu ici et pour nos enseignements est AntConc (Anthony, 2014a).

Une fois le corpus compilé, compétence technique acquise dans le cadre du cours sur les corpus, il devient aisé de montrer lors du cours de traduction pratique comment accéder aux phénomènes phraséologiques et en particulier aux phénomènes de collocations : par exemple, le terme *rendement* comme équivalent d'*earnings* dans le contexte qui nous intéressait s'associe avec des verbes spécifiques. Une exploitation du corpus compilé à l'aide du concordancier permet d'observer le co-texte linguistique du terme : la saisie de la requête

⁵ Nous souhaitons remercier ici Nathalie Moulard, traductrice professionnelle et intervenante au sein du Master TSM, avec qui nous avons travaillé en collaboration.

rendement* (l'astérisque permettant d'inclure la forme plurielle) permet de mettre au jour l'utilisation de certains verbes grâce à l'observation des lignes de concordance (Figure 1) : un investisseur peut *délivrer* un rendement, des rendements peuvent être *orientés à la baisse*, peuvent à l'inverse *bondir* ; on parle également de *remontée* des rendements suite à une décision politique, etc. Les étudiants peuvent également y trouver des termes techniques comme *rendements obligataires*, plus spécialisé que *rendements des obligations* et dont l'utilisation permet une traduction spécialisée adaptée là où les étudiants font souvent preuve d'une tendance à la détechnicisation du texte source.

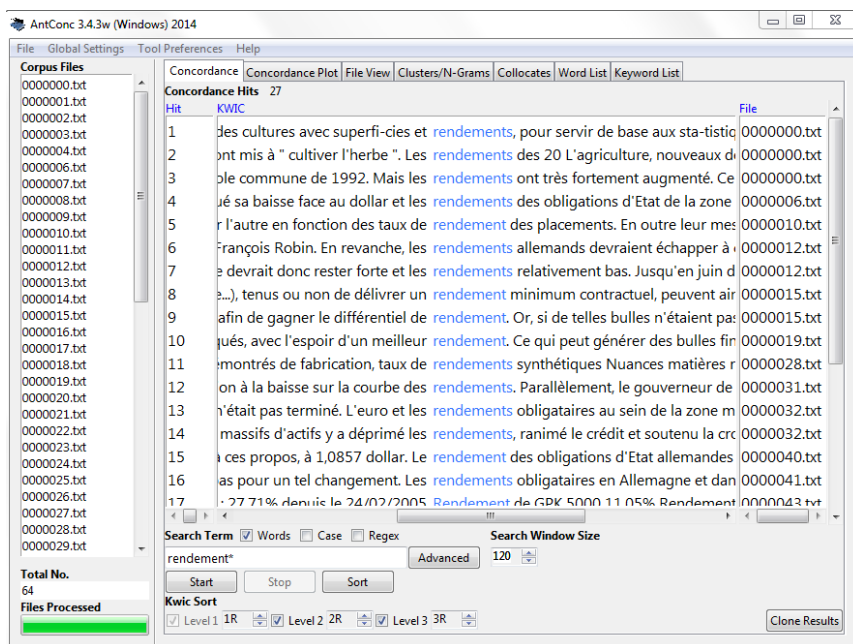


Figure 1. Résultats de la requête *rendement**

L'utilisation des corpus permet donc de montrer que la terminologie ne suffit pas, et le décloisonnement entre les trois enseignements mentionnés ici permet d'y sensibiliser les étudiants. L'enseignement consacré aux outils de corpus ne doit donc pas selon nous être organisé de façon isolée, le décloisonnement maximal pouvant se faire dans le cadre d'un *skills lab* ou agence virtuelle de traduction (Vandepitte, Kerremans, Van Egdome et Looock, 2017), où les étudiants doivent mobiliser l'ensemble de leurs connaissances (linguistiques, techniques et technologiques, interpersonnelles) afin de gérer des projets de traduction.

2.2. Les corpus comme outils complémentaires

La boîte à outils du traducteur professionnel n'a jamais été aussi remplie qu'en ce début de 21^{ème} siècle. Il est même envisageable de dire qu'il y a aujourd'hui pléthore d'outils informatiques entre les outils « traditionnels » (dictionnaires et glossaires en ligne), les logiciels de TAO avec leurs mémoires de traduction, sans oublier les outils de gestion de projets de traduction en ligne, ou encore la traduction automatique de plus en plus présente sur le marché.⁶ L'introduction d'outils supplémentaires doit donc se justifier, et si l'on souhaite convaincre de leur utilité, il importe de montrer comment les outils de corpus jouent un rôle non rempli par les autres outils. Au-delà des phénomènes mentionnés dans la section précédente, il est possible de montrer comment l'utilisation du même corpus permet de

⁶ D'après l'enquête *European Language Industry Survey*, la proportion de prestataires de services linguistiques n'utilisant pas la traduction automatique est passée de 80% à 36% entre 2016 et 2018 ; du côté des traducteurs indépendants, la proportion est passée de 69% à 48% (Tirry, 2018).

résoudre des questions d'ordre terminologique non résolues par ailleurs. Ceci est particulièrement vrai des termes qui possèdent plusieurs équivalents et pour lesquels le choix doit se faire en fonction du contexte. Ainsi, les étudiants ont rencontré des difficultés pour traduire le terme *budget blueprint*. Si dans le cadre de leur recherche terminologique, le terme a bien été identifié comme un terme technique, les équivalents trouvés dans les différents glossaires à leur disposition, ont été divers : *planification budgétaire*, *plan de financement*, *budget prévisionnel*. Le contexte de l'apparition du terme anglais (*the House of Representatives passed a budget blueprint on Thursday*) fournit suffisamment d'informations pour qu'une recherche au sein d'un corpus consacré à des questions économiques puisse être fructueuse : il s'agit d'un vote à propos de décisions relatives au budget du pays. Cette recherche permet de mettre au jour des expressions du type *adopter un budget*, *approuver un budget prévisionnel* dans le cadre d'une *réforme fiscale*, terme qui pourra être utilisé pour la traduction de *tax-overhaul goals*, qui a également posé problème à certains étudiants puisque certains glossaires en ligne proposent le terme *révision fiscale*. Dans le cadre de la traduction d'un autre texte économique, les étudiants avaient éprouvé de la difficulté pour traduire le terme *annual rate* dans la phrase *Real gross domestic product (GDP) increased 4.2 percent at an annual rate in the second quarter*. Tandis que les glossaires proposent des solutions différentes (*rythme annuel*, *taux annuel* notamment), la question de l'utilisation de son équivalent en français se pose également : doit-on écrire *sur*, *selon*, *d'après*... un rythme annuel ? Une recherche sur corpus avait très vite montré (Loock 2016a : 131-132) que l'équivalent d'*annual rate* dans ce contexte était *rythme annualisé* au sein d'un schéma figé : *progresser/reculer de x% en rythme annualisé de [date1] à [date2]*. Aucun autre outil n'avait été en mesure de fournir des informations aussi précises ; de plus, le nombre d'occurrences (7 dans notre cas), faisant toutes état d'un schéma linguistique identique, permet de prendre une décision de façon plus sécurisée qu'un exemple unique fourni au sein d'un glossaire, aussi sophistiqué soit-il.

2.3. La pertinence des données

L'une des grandes difficultés que semblent rencontrer les étudiants est le choix du corpus en fonction du projet de traduction. Or, puisqu'il s'agit de régler des problèmes de traduction, il importe que les informations fournies par les corpus soient fiables. Pour cela, il faut que les données collectées soient elles-mêmes fiables, mais aussi pertinentes. La phase de sélection des données est donc une étape cruciale lors de la compilation d'un corpus, bien plus que les considérations d'ordre technique qui ne posent plus problème avec les outils d'aujourd'hui, et il importe d'inviter les utilisateurs à bien prendre le temps de la réflexion, et c'est en ce sens que la démarche du traducteur doit être proche de celle du chercheur en linguistique. Le choix d'un corpus ou la sélection des données dans le cadre de la compilation d'un corpus ne sont jamais anodins mais impliquent des choix : choix d'inclusion (données conformes), d'exclusion (données non conformes), ou encore de clôture (le corpus étant par définition fini, un terme doit nécessairement être mis à la compilation des données qui se doivent idéalement de constituer un échantillon représentatif). Par conséquent, cette étape est soumise à la subjectivité du chercheur, plus ou moins consciente. Comme l'indique très justement le titre de l'article de Charaudeau (2009), « Dis-moi quel est ton corpus, je te dirai quel est ta problématique ».

Il importe donc de sensibiliser les étudiants aux questions de variation, liée au registre mais aussi géographique. Un corpus de textes médicaux ne peut servir pour traduire un texte économique, et vice-versa. Un corpus de textes en français de France ne peut servir pour traduire un texte vers le français du Québec, et vice-versa. Un corpus en langue dite

« générale »⁷ ne peut être utilisé pour traduire un texte en langue spécialisée, même si les documents au sein du corpus abordent le même sujet (risque, une fois encore, de détechnicisation du texte source au moment de la traduction). Il semble s’agir là d’évidences, mais notre expérience montre que les étudiants utilisent des corpus en ligne ou compilent leurs propres corpus sans prendre en compte ces phénomènes de variation, et obtiennent ainsi des informations non pertinentes. Ainsi nos étudiants ont pu traduire *heart-shaped* dans un texte médical spécialisé par une expression de la langue générale, *en forme de cœur*, alors que dans un texte scientifique et technique on préférera l’adjectif à racine savante *cordiforme*. De la même manière, lors de la compilation de données grâce au logiciel BootCaT, nous avons constaté un manque de recul vis-à-vis des pages web récupérées (les pages en question sont-elles extraites de blogs, de la presse, de pages encyclopédiques ? de quel type de variante géographique s’agit-il ?). Pour remédier à ce manque de prise de conscience, des exercices simples peuvent être mis en place, à partir par exemple du *Corpus of Contemporary American English* (Davies, 2008-) multi-registre, en amenant les étudiants à comparer des mots ou des termes synonymes mais à la distribution différenciée en fonction du registre. Ainsi les requêtes *heart attack* et *myocardial infarction* montrent que le second terme se trouve majoritairement au sein des textes « académiques » au détriment du premier terme (Figure 2).⁸

SPOKEN	1,609	109.4	14.71	
FICTION	1,288	104.9	12.28	
MAGAZINE	2,367	110.1	21.50	
NEWSPAPER	1,564	106.0	14.76	
ACADEMIC	284	103.4	2.75	
SPOKEN	4	109.4	0.04	
FICTION	12	104.9	0.11	
MAGAZINE	37	110.1	0.34	
NEWSPAPER	13	106.0	0.12	
ACADEMIC	367	103.4	3.55	

Figure 2. Distribution des termes *heart attack* (haut) et *myocardic infarction* (bas) au sein du COCA

De même, la distribution en fonction du registre des paires de mots comme *answer/response* et *whole/entire*, qui relèvent du schéma bien connu en anglais des paires de mots synonymes, l’un d’origine germanique l’autre d’origine romane, permet de sensibiliser les étudiants à l’illusion de la synonymie. On constate en effet que les termes d’origine romane (*response*, *entire*) sont plus fréquents dans le sous-corpus « académique » du COCA, alors qu’en anglais parlé, ce sont les termes d’origine germanique qui sont les plus fréquents (Tableau 1). Il s’agit là d’une autre façon de sensibiliser les étudiants à l’importance du registre sur les choix de traduction, en particulier dans le cas de traductions spécialisées, la présence d’un terme d’origine romane dans le texte source indiquant un registre plus élevé. S’agissant de l’anglais, et donc de la langue source pour nos étudiants qui traduisent vers le français qui est leur langue maternelle, ces résultats permettent de déterminer le caractère plus ou moins formel du texte à traduire, et donc la stratégie de traduction à adopter. Un texte formel se traduit en effet, sauf consigne contraire de la part du donneur d’ordre, en respectant le niveau de langue.

⁷ L’utilisation du terme « langue générale » est un raccourci. En tant que telle, celle-ci n’existe pas : il s’agit d’une langue abstraite qui correspond en fait à l’addition des différents registres, tout acte de communication appartenant nécessairement à un registre spécifique de la langue.

⁸ La première colonne indique la fréquence brute, à savoir le nombre d’occurrences dans les différents sous-corpus ; la deuxième indique le nombre de mots dans chaque sous-corpus (en millions) ; la troisième indique la fréquence dite « normalisée » par million de mots, soit le chiffre le plus intéressant pour des comparaisons pertinentes.

	<i>answer</i> (N)	<i>response</i> (N)	<i>whole</i> (Adj)	<i>entire</i> (Adj)
COCA academic subcorpus	113.51	407.06	97.75	119.02
COCA spoken subcorpus	177.43	88.08	331.18	103.14

Tableau 1. Fréquences normalisées par million de mots des tokens *answer*, *response*, *whole*, et *entire*

Il importe donc de sensibiliser les étudiants à l'adéquation entre les données au sein du corpus et le projet de traduction en question. Nous aimons à ce sujet mentionner l'adage « *garbage in, garbage out* » (GIGO pour les intimes) utilisé en informatique pour signaler l'inadéquation ou la corruption de données, et qui est également utilisée dans le cadre de la linguistique de corpus (p. ex. Viana, Zyngier et Barnbrook, 2011 : 183) afin de sensibiliser les étudiants à cet enjeu crucial : de mauvaises données, en termes de qualité ou d'adéquation avec le texte à traduire, fourniront de mauvaises informations.

2.4. La formulation d'hypothèses

Un corpus ne se consulte pas comme un dictionnaire et requiert une approche qui repose sur la formulation intuitive d'hypothèses et l'observation de données devant conduire à une prise de décision, soit une démarche qui n'est pas sans rappeler celle du chercheur en linguistique. Il s'agit en effet de tester des hypothèses afin d'effectuer des vérifications d'ordre linguistique sur le contexte d'apparition de tel(le) ou tel(le) mot ou séquence de mots, ou encore son/leur existence elle-même. Or, il importe non seulement de savoir quelles hypothèses tester (on ne trouve pas ce qu'on ne cherche pas...) mais aussi de savoir prendre du recul vis-à-vis des résultats affichés afin de prendre une décision éclairée (présence inévitable de bruit et de silence). Nous prendrons pour cela deux exemples : la traduction en français du sigle *OS* pour *operating system* dans le domaine de l'informatique et la traduction du terme *interest rates* dans le contexte de la politique de rachats d'actifs de la Banque centrale européenne.

Dans le premier cas, le traducteur pourrait s'attendre à ce que l'équivalent du sigle *OS* soit *SE* en français puisque l'équivalent d'*operating system* est *système d'exploitation*. Il s'agit d'ailleurs d'une proposition du portail linguistique de Microsoft (<https://www.microsoft.com/fr-fr/language>), utilisable comme glossaire multilingue de référence. Or, une recherche effectuée au sein d'un corpus composé de textes spécialisés relatifs au lancement d'un nouveau système d'exploitation en Chine afin de concurrencer les autres systèmes d'exploitation, ne donne presque aucune occurrence du sigle *SE*, ce qui pose véritablement question quant à l'existence de celui-ci même si un corpus reste un ensemble fini de données qui par définition ne peut être exhaustif. Il convient alors d'être imaginatif et de tester des hypothèses, en saisissant par exemple *système d'exploitation*, ou directement le sigle *OS* (Figure 3), afin d'obtenir confirmation qu'il s'agit là de l'équivalent du sigle anglais : les spécialistes du domaine informatique recourent au terme traduit *système d'exploitation* mais au sigle non traduit *OS*, les quelques occurrences de *SE* reflétant une utilisation marginale.

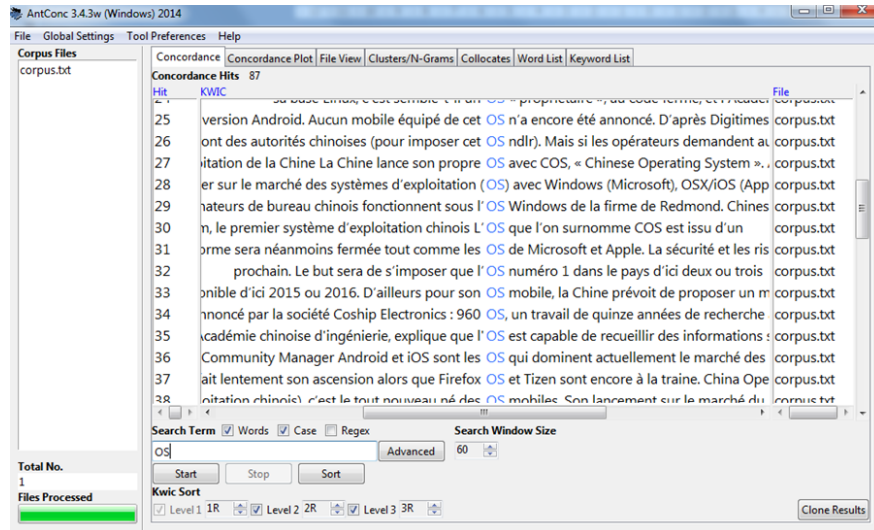


Figure 3. Résultats de la requête OS

Dans le second cas, il s'agit de traduire le terme *interest rates* dans l'exemple suivant (le contexte étant la politique de rachats d'actifs de la BCE) :

- (1) The ECB, as expected, left **interest rates** unchanged and reiterated that it anticipated them to remain at present levels for an “extended period” and beyond the eventual end of its asset buying program.

Si à première vue l'équivalent *taux d'intérêt* proposé par différents glossaires en ligne peut sembler convenir, il importe néanmoins d'effectuer une vérification au sein d'un corpus contenant des textes économiques spécialisés, afin de vérifier l'hypothèse selon laquelle le terme, utilisé dans de nombreux sous-domaines économiques et financiers, a un équivalent unique en français. L'interrogation du corpus déjà utilisé *supra* donne 744 occurrences du mot *taux*. L'observation du co-texte linguistique s'avère donc fastidieuse, même en opérant un tri en fonction du co-texte linguistique droit, place privilégiée des modificateurs en français. En revanche, il est possible de n'exploiter que les lignes de concordance faisant intervenir *taux* et *BCE* : à l'aide de la fonctionnalité avancée « *Use context words* » d'AntConc, il est en effet possible de ne faire apparaître que les occurrences de *taux* avec le sigle *BCE* apparaissant parmi les 5 mots (nombre de mots personnalisable) précédant ou suivant le terme recherché (pivot). Le nombre de lignes de concordance tombe alors à 26 (Figure 4), soit un nombre plus facile à observer, qui fait apparaître notamment l'utilisation de *taux d'intérêt*, *taux directeurs*, ou encore *taux des réserves obligatoires*. L'observation du co-texte immédiat, devant évoquer une décision de la BCE, permet de déterminer que si le terme *taux directeurs*, en tant qu'hypéronyme, est acceptable, c'est le terme hyponyme qui est le plus usité en français, à savoir *taux directeurs*, comme l'indiquent les lignes de concordance 14, 15 et 16.

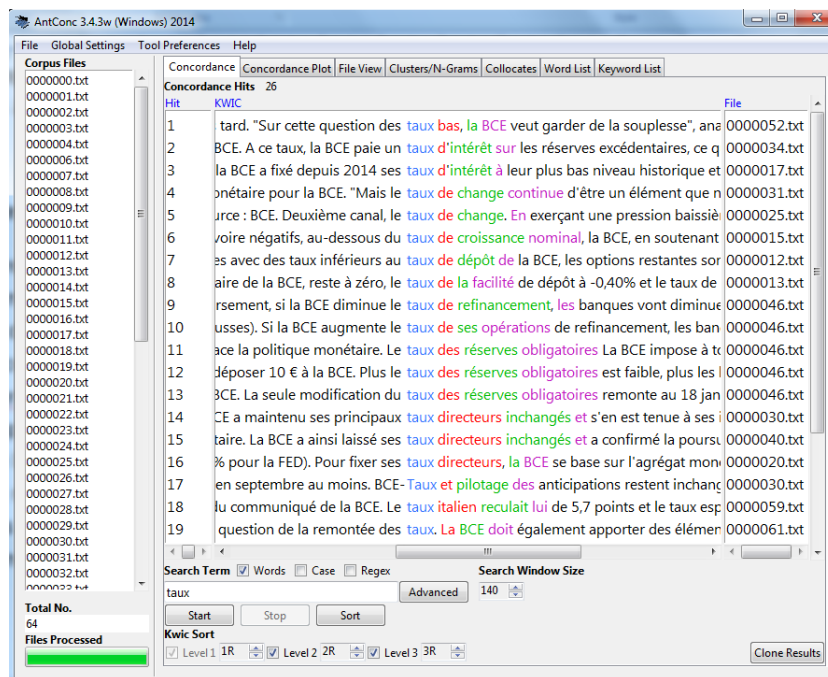


Figure 4. Résultats de la requête `taux` avec filtre BCE 5 mots avant/après le pivot

Il importe donc d'être « malin » lorsque l'on consulte un corpus ; la démarche est très différente de l'utilisation d'un dictionnaire, et celle-ci doit s'apprendre. Notre expérience nous montre qu'il s'agit là d'un enjeu important et que l'étudiant en formation de traduction n'est pas toujours à l'aise avec cette approche intuitive, proche de celle du chercheur en linguistique.

2.5. Rapidité de compilation et ergonomie : des enjeux à ne pas négliger

Parce qu'ils travaillent sur des textes souvent techniques relevant de la langue spécialisée, les traducteurs ne trouvent pas souvent en ligne le corpus adapté à leurs besoins et se voient donc contraints de compiler les leurs, une activité souvent perçue comme chronophage et technique qui serait réservée aux spécialistes de la linguistique de corpus. Le traducteur étant encore généralement payé au mot traduit, tout investissement de temps dans l'utilisation d'un outil doit donc être rentable (augmentation de la productivité et/ou de la qualité du texte traduit). Dans la mesure où c'est le marché qui dicte l'utilisation de tel ou tel outil comme nous avons pu le voir avec les logiciels de TAO, cette question est cruciale. Car force est de constater que les outils de corpus n'ont pas encore trouvé leur place en dépit de tous les avantages qu'ils présentent, y compris chez les jeunes professionnels qui y ont pourtant été formés (Frérot & Karagouch, 2016), l'une des raisons avancées étant l'argument économique. Afin de prendre en compte cet enjeu, la phase de compilation doit donc être optimisée. Un choix doit être fait entre une compilation manuelle, plus longue, et une compilation semi-automatisée à l'aide d'un logiciel comme BootCaT : dans le premier cas, la qualité sera optimale ; dans le second cas, il conviendra d'être prudent au moment de la consultation des résultats. Dans les deux cas, en revanche, nous prôtons la compilation de corpus « non nettoyés », ce qui, nous en avons bien conscience, fera frémir les linguistes de corpus : la sélection des données peut en effet être « grossière », c'est-à-dire sans éliminer titres, intertitres, légendes de photographies, textes publicitaires, liens vers d'autres documents, citations mises en exergue... En effet, ces éléments n'apparaîtront pas lors des recherches effectuées au sein du concordancier, ou seront très repérables, et il n'est donc pas nécessaire de passer un temps non négligeable à nettoyer les fichiers, d'autant que le

traducteur, contrairement au chercheur en linguistique, ne compilera pas de statistiques à partir de ses observations. Il importe en revanche de rester prudent et de ne pas incorporer, lors de la récupération de pages web, les commentaires des internautes situés en bas de page, dont la qualité linguistique ne peut pas être garantie.

Par ailleurs, les questions d'ergonomie des outils de corpus ne sont pas à négliger. Nous avons déjà eu l'occasion ailleurs (Loock, 2016b) d'aborder la question de cette importance, et serons donc assez brefs ici. Il s'agit de s'intéresser aux « interactions entre les humains et les autres composantes d'un système, tant technologiques que sociales », à différents niveaux : physique, cognitif, et organisationnel (Lavault Olléon, 2011). Pour que le traducteur s'approprie un outil d'aide à la traduction, il importe en effet que celui-ci s'intègre à son espace de travail et vienne l'accompagner dans ses prises de décision, sans venir interrompre de façon brutale ou contre-productive le cheminement de sa pensée. Les outils de corpus, de ce point de vue, ne sont pas sans poser problème : il existe beaucoup de types de corpus différents (monolingues, multilingues, comparables, parallèles) exploitables de façons différentes, en ligne au moyen d'interfaces non homogènes et parfois peu faciles d'utilisation, ou hors ligne au moyen de concordanciers plus ou moins faciles d'utilisation. De nombreux efforts ont été faits et un certain nombre d'outils sont désormais très faciles à prendre en main, comme le concordancier AntConc que nous utilisons avec nos étudiants. Un autre aspect de la question est l'intégration à l'espace de travail : si jusqu'à présent les outils de corpus étaient interrogeables en dehors des logiciels d'aide à la traduction avec leurs mémoires de traduction, on constate à l'heure actuelle une convergence avec de nouvelles fonctionnalités intégrées à ces logiciels permettant d'effectuer des recherches au sein de bases de données linguistiques : des recherches de concordance sont désormais possibles au sein de SDL Studio et de memoQ, tout comme la possibilité d'interroger des corpus disponibles en ligne directement au sein du logiciel (fonctionnalités memoQ WebSearch et SDL Web Lookup!), et l'on voit même apparaître des plug-ins spécifiques comme le Sketch Engine plug-in de SDL Studio, qui permet d'interroger le corpus EUR-lex (corpus multilingue contenant les textes législatifs européens en 24 langues, cf. Vít, Michelfeit, Medved' et Jakubíček, 2016) et d'obtenir des informations sur un terme sélectionné au sein du texte source ou du texte cible (occurrences en contexte, collocations, synonymes). D'autres possibilités d'intégration, nécessitant parfois quelques manipulations assez techniques, sont possibles (Tagne & Loock, 2017). Nous assistons donc à l'heure actuelle à une convergence de ces outils, plus précisément à une véritable intégration des outils de corpus au sein des logiciels de TAO, preuve s'il en fallait une de leur appropriation progressive par les traducteurs professionnels.

3. Les raisons d'une appropriation pas si aisée : quelques réflexions

Au-delà des questions relatives à l'ergonomie, à la nécessité de développer son intuition afin de formuler des requêtes pertinentes et de se montrer prudent lors de la phase de compilation des données si l'on souhaite disposer d'un corpus pertinent pour le projet de traduction en cours, nous avons pu remarquer chez nos étudiants, mais aussi chez les traducteurs en exercice, qu'une certaine forme de méfiance subsistait vis-à-vis de la consultation de données attestées par opposition à d'autres outils comme les dictionnaires ou les glossaires. En cela, il nous semble que deux approches de la langue s'opposent, celle des traducteurs et celle des chercheurs en linguistique.

Comme le soulignent Penke et Rosenbach (2004) en introduction du numéro spécial de *Studies in Language* consacré à la question des données en linguistique, « *nowadays most linguists will probably agree that linguistics is indeed an empirical science* » (Penke & Rosenbach, 2004 : 480, c'est nous qui soulignons). Le débat sur cette question, qui date d'il y

a trente ans environ, semble avoir été tranché. Ce qui en revanche subsiste, c'est la question du type d'empirisme et au-delà, du type d'observables à partir desquels la connaissance théorique sur le langage et les langues peut être extraite : doit-on travailler à partir de données attestées réunies en corpus, électronique ou non, ou à partir de données fabriquées pour la circonstance permettant une démonstration ciblée et efficace ? En ce début de 21^{ème} siècle, et grâce à l'accessibilité des corpus électroniques depuis le développement de l'internet à haut débit, nombreux sont les linguistes qui penchent pour la première solution. Nous n'entrerons pas ici dans le débat du type d'observables à exploiter en sciences du langage, et ne condamnerons certainement pas les chercheurs ayant recours à des données fabriquées. En dehors d'études très théoriques, dans les deux cas il s'agit de s'appuyer sur l'observation de données, que le chercheur soumet à des expérimentations afin de tester des hypothèses et d'en tirer des conclusions théoriques. Grâce aux observables réunis en corpus d'étude, des hypothèses peuvent être testées, et ainsi confirmées ou falsifiées. L'empirisme se veut donc être une méthode, une méthode d'observation à partir desquelles la connaissance, la théorie pourra émerger. Tout ceci fait aujourd'hui plutôt consensus chez les chercheurs en linguistique.

A partir du moment où l'on travaille sur des données attestées, écrites ou orales, produites dans le cadre de situations de communication réelles, on sort nécessairement des exemples « normés », et les phénomènes de variation ne sont pas rares puisque l'on s'intéresse à l'usage de la langue. Ainsi des exemples « non-standard » surgiront inévitablement, à savoir des exemples qui s'écartent des normes attendues telles que décrites par les ouvrages de grammaire. Et c'est là selon nous qu'une réaction négative vis-à-vis de l'utilisation des corpus électroniques peut apparaître chez les traducteurs. Ces phénomènes ne doivent pas être passés sous silence, et il importe de sensibiliser à leur existence puisque ces données « atypiques » ne manqueront pas d'apparaître lors de recherches. Là où le chercheur en linguistique est habitué à ce type de données, et souvent même les recherche expressément, le traducteur, dont le but n'est pas de décrire la langue, peut s'en inquiéter, voire s'en effrayer. Nous voyons en effet une opposition entre l'attitude descriptive du chercheur en linguistique et celle, plus prescriptive, du traducteur, dans la mesure où ce dernier exploite des données réunies en corpus afin de résoudre des problèmes de traduction, et a donc besoin de données desquelles il puisse s'inspirer en toute confiance. Par conséquent, découvrir des exemples qui dévient des normes attendues pose problème, et peut l'amener à ne pas faire confiance aux outils de corpus. Comment en effet s'inspirer de données réunies en corpus lorsque l'on constate, que ce soit dans le *Corpus of Contemporary American English* ou le *British National Corpus* (Davies, 2004-), que la négation de *used to* se fait majoritairement par l'emploi de *didn't used to*, condamné dans les ouvrages de grammaire du fait de la présence d'un double marquage morphologique pour le prétérit (2) ?

(2) You need estradiol to reproduce. But humans didn't used to live as long as they do today, so now you're exposed to it longer, and it's causing cancers in your body. (COCA, magazine)

De même, que penser de la présence significative de *malgré que* au sein du corpus Frantext, qui regroupe près de 300 millions de mots de français écrit, principalement littéraire, du 12^{ème} au 21^{ème} siècle (<http://www.frantext.fr/>) ? Tout ceci pose problème au traducteur, qui doit apprendre à repérer ces phénomènes de variation, et comprendre que s'inspirer de l'usage demande une prise de recul critique, car il ne faut pas oublier non plus que tout corpus contient inévitablement des erreurs (erreurs de saisie, d'étiquetage...) et que toute requête mettra en jeu une part de bruit et de silence. Il nous semble que ceci est un véritable problème pour les étudiants en formation de traduction, qui doivent donc adopter l'attitude d'un chercheur en linguistique lorsqu'ils consultent un corpus électronique.

L'expérience nous montre que deux types de réponse semblent les rassurer. Premièrement, il convient d'insister sur la phase de compilation du corpus : pour traduire un texte spécialisé, il convient de rassembler en corpus un ensemble de textes spécialisés, qui par définition, sont moins sujets à ce type de variation que d'autres registres (le premier étant naturellement le registre parlé, et il importe de dire aux étudiants d'exclure ce registre lorsqu'ils consultent des corpus de référence en ligne comme le COCA ou le BNC). Deuxièmement, il est important de transformer cette faiblesse en force, et la traduction dans les domaines du marketing et de la publicité est parfaite pour cela, puisqu'il s'agit justement de domaines où le jeu avec la langue est omniprésent, ce qui peut poser problème au traducteur lorsqu'il s'agit de la langue source étrangère. Que faire en effet d'exemples du type *When life gives you lemons, keep them because free lemons!* Là où un schéma *because* + proposition complète (*because you've been given free lemons*) ou un schéma *because of* + SN (*because of free lemons*) correspondraient aux normes attendues, cet exemple fait intervenir un schéma non décrit par les ouvrages de grammaire ou les dictionnaires : *because* + SN, soit l'emploi de *because* comme préposition. Quelle différence de sens faire avec les schémas traditionnels ? Y en a-t-il une ? C'est là que les données réunies en corpus qui font la part belle à l'usage deviennent intéressantes et permettent au traducteur d'observer d'autres exemples du même type (3). Ceci lui permet non seulement de comprendre qu'il ne s'agit pas d'une erreur de performance mais bien d'un phénomène significatif de la langue, et également d'affiner sa compréhension du texte source par le biais de l'observation de données en contexte (pour d'autres exemples, voir Loock 2016a : 111):

- (3) a. People die of heart attacks and strokes because diabetes. (BNC magazine, 2007)
- b. TOM: I definitely kind of viewed him as a suspect. KOTB: Why? TOM: Well, because motive. (BNC spoken, NBC, 2005)
- c. And "Primary Colors" I think has hit the country like a dud, because behavior. (BNC spoken, Fox News 1998)
- d. I think I prefer just a curtain and I don't think I could put both there because the roller. (BNC, spoken, meeting)

De ce point de vue, le traducteur adopte alors une approche critique des données, similaire à celle du chercheur en linguistique.

4. Conclusion

Dans cet article, nous avons souhaité revenir sur notre expérience d'enseignement des outils de corpus et discuter des différents enjeux et obstacles à leur appropriation que nous avons pu constater chez les traducteurs en formation. Nous avons argumenté que la plupart de ces difficultés étaient liées à une approche de la langue spécifique, qui n'est pas tout à fait la même que celle du chercheur en linguistique. Nous avons proposé des solutions afin de les pallier et de sensibiliser les étudiants aux enjeux d'une approche de la langue qui fait la part belle aux données attestées, et donc à l'usage. En cela, et pour revenir à notre question initiale, il nous semble que le traducteur, qui n'est certes pas chercheur en linguistique, doit néanmoins adopter une démarche assez similaire s'il souhaite pouvoir bien exploiter les outils de corpus dans le cadre de son activité. Ceci n'empêche pas une démarche spécifique, où la question de la rentabilité économique (augmentation de la productivité et/ou de la qualité des textes traduits) ne doit pas être négligée.

Références

- ANTHONY, Laurence (2014), AntConc (Version 3.4.3) [Programme informatique], Tokyo, Waseda University, disponible sur <http://www.laurenceanthony.net/> [consulté le 22/02/2018]
- BARONI Marc et BERNARDINI Silvia (2004), « BootCaT: Bootstrapping corpora and terms from the web », *Proceedings of LREC 2004*, disponible sur <http://www.lrec-conf.org/proceedings/lrec2004/> [consulté le 22/02/2018]
- BEEBY, Allison, RODRÍGUEZ Inés, Patricia, et SÁNCHEZ-GIJÓN, Pilar (éds) (2009), *Corpus Use and Translating*, Amsterdam/Philadelphie, John Benjamins.
- BERNARDINI, Silvia et FERRARESI, Adriano (2013), « Old Needs, New Solutions: Comparable Corpora for Language Professionals », in Sharoff, S., Rapp, R., Zweigenbaum, P. et Fung, P. (éds), *Building and Using Comparable Corpora*, Dordrecht, Springer, p. 303-319.
- BOULTON, Alex et TYNE, Henry (2014), *Des Documents authentiques aux corpus : Démarches pour l'apprentissage des langues*, Paris, Didier.
- BOWKER, Lynne (1998), « Using Specialized Monolingual Native-Language Corpora as a Translation Resource: a Pilot Study », *Meta : journal des traducteurs / Meta: Translators' Journal* 43(4), pp. 631-651.
- BOWKER, Lynne (1999), « Exploring the Potential of Corpora for Raising Language Awareness in Student Translators », *Language Awareness* 8(3-4), pp. 160-173.
- BOWKER, Lynne et Pearson, Jennifer (2002), *Working with Specialized Language: A Practical Guide to Using Corpora*, Londres, Routledge.
- CHARAUDEAU, Patrick (2009), « Dis-moi quel est ton corpus, je te dirai quelle est ta problématique », *Corpus* 8, pp. 37-66.
- DAVIES, Mark (2004-), *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Disponible sur <http://corpus.byu.edu/bnc/> [consulté le 22/02/2018]
- DAVIES, Mark (2008-), *The Corpus Of Contemporary American English: 450 million words, 1990-present*. Disponible sur <http://corpus.byu.edu/coca/> [consulté le 22/02/2018]
- FRANKENBERG-GARCIA, Anna (2015), « Training Translators to Use Corpora Hands-on: Challenges and Reactions by a Group of 13 Students at a UK University », *Corpora* 210 (3), pp. 351-380.
- FRÉROT, Cécile (2016), « Corpora and corpus technology for translation purposes in professional and academic environments. Major achievements and new perspectives », *Cadernos de Tradução*, 36(1), <https://doi.org/10.5007/2175-7968.2016v36nesp1p36> [consulté le 21/03/2018]
- FRÉROT, Cécile et KARAGOUCHE, Lionel (2016), « Outils d'aide à la traduction et formation de traducteurs : vers une adéquation des contenus pédagogiques avec la réalité technologique des traducteurs », *ILCEA* 27. Disponible sur <http://ilcea.revues.org/3849> [consulté le 22/02/2018]
- GARCIA, Ignacio (2005), « Long Term Memories: Trados and TM Turn 20 », *The Journal of Specialized Translation* 4, pp. 18-31.
- JOHNS, Tim (1990), « From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-driven Learning », *CALL Austria* 10, pp. 14-34.
- KÜBLER, Natalie (2003), « Corpora and LSP Translation », in Bernardini, S., Stewart, D. et Zanettin, F. (éds), *Corpora in Translator Education*, Manchester/Northampton, St Jerome, p. 25-42.
- KÜBLER, Natalie (éd.) (2011a), *Language Corpora, Teaching, and Resources: From Theory to Practice*, Bern, Peter Lang.
- KÜBLER, Natalie (2011b), « Working with Corpora for Translation Teaching in a French-speaking Setting », in Frankenberg-Garcia, A., Flowerdew, L., et Aston, G. (éds), *New Trends in Corpora and Language Learning*, Londres, Continuum, p. 62-80.
- KÜBLER, Natalie et ASTON, Guy (2010), « Using Corpora in Translation », in M. Mc Carthy et A. O'Keefe (éds), *The Routledge Handbook of Corpus Linguistics*, Londres : Routledge, p. 505-515
- Natalie Kübler, Alexandra Mestivier, Mojca Pecman, Maria Zimina (2016), « Exploitation quantitative de corpus de traductions annotés selon la typologie d'erreurs pour améliorer les méthodes d'enseignement de la traduction

spécialisée », *JADT2016 Journées internationales d'Analyse statistique des Données Textuelles*, Nice, p.731-741.

LAVAUULT-OLLEON, Élisabeth (2011), « Une introduction à la problématique Traduction et Ergonomie », *ILCEA* 14, Disponible sur <http://journals.openedition.org/ilcea/1118> [consulté le 22/02/2018].

LOOCK, Rudy (2016a), *La Traductologie de corpus*, Villeneuve d'Ascq, Presses Universitaires du Septentrion.

LOOCK, Rudy (2016b), « L'Utilisation des corpus électroniques chez le traducteur professionnel : quand ? comment ? pour quoi faire ? », *ILCEA* 27, Disponible sur <http://journals.openedition.org/ilcea/3835> [consulté le 22/02/2018].

PENKE, Martina et ROSENBAACH, Anette (2004), « What counts as evidence in linguistics? – An introduction », *Studies in Linguistics* 28 (3), pp. 480-526.

TAGNE, John Borel et LOOCK, Rudy (2017), « Ça “corpus” pas mal du côté des logiciels de TAO : l'heure de la convergence aurait-elle sonné ? », Communication lors de la journée d'études « La gestion des contraintes génériques/textuelles par les traducteurs : annotation, modélisation, extraction de l'information dans les corpus électroniques », 24 novembre 2017, Dijon, France.

TOUDIC, Daniel et KRAUSE, Alexandra (2017), *European Master's in Translation - Référentiel de compétences EMT*, rapport de la Direction générale de la traduction, Commission européenne, Disponible sur https://ec.europa.eu/info/sites/info/files/emt_competence_fw_k_2017_fr_web.pdf [consulté le 22/02/2018].

SOMERS, Harold (2003), « Translation Memory Systems », in Somers, H. L. (éd.), *Computers and Translation: a Translator's Guide*, Amsterdam/Philadelphie : John Benjamins, p. 31-48.

VANDEPITTE, Sonia, KERREMANS, Koen, VAN EGDOM, Gys-Walt, et LOOCK, Rudy (2017), « Teaming up in translator education: simulated translation bureaus as a means of increasing professionalism and employability », Colloque « BAAHE 2017 - Let's Inter-Act! Innovative Teaching Practices in English Studies », Louvain-la-Neuve, 15 décembre 2017.

TIRRY, Rudy (2018), « L'impact de la Traduction Automatique d'ici 2020 », Communication dans le cadre le journée d'études « Traduction & Qualité 2018 : Biotraduction et traduction automatique », Université de Lille, 2 février 2018.

VANDER VIANA, ZYNGIER, Sonia et BARNBROOK, Geoff (éds.) (2011), *Perspectives on Corpus Linguistics*, Amsterdam/Philadelphie : John Benjamins.

VARANTOLA, Krista (2003), « Translators and Disposable corpora », in Bernardini, S., Stewart, D. et Zanettin, F. (éds), *Corpora in Translator Education*, Manchester, St. Jerome, p. 55-70.

VÍT, Baisa, MICHELFEIT, Jan, MEDVEĎ, Marek, et JAKUBÍČEK, Miloš (2016), « European Union Language Resources in Sketch Engine », in *The Proceedings of tenth International Conference on Language Resources and Evaluation (LREC'16)*, European Language Resources Association (ELRA). Disponible sur https://www.sketchengine.co.uk/wp-content/uploads/eur_lex_2016.pdf [consulté le 22/02/2018].

ZANETTIN, Federico (2012), *Translation-driven Corpora*, Manchester, St Jerome Publishing.

ZANETTIN, Federico, BERNARDINI, Silvia, et STEWART, Dominic (2003), *Corpora in Translator Education*, Manchester, St Jerome Publishing.