La main de Thôt
ISSN : 2272-2653
Éditeur : Carole Filière

9 | 2021
La traduction littéraire et SHS à la rencontre des nouvelles technologies de la traduction : enjeux, perspectives et défis

Traduire ou post-éditer la traductologie : une expérience comparative

Caroline Rossi et Alice Carré

http://interfas.univ-tlse2.fr/lamaindethot/987

Référence électronique
Caroline Rossi et Alice Carré, « Traduire ou post-éditer la traductologie : une expérience comparative », *La main de Thôt* [En ligne], 9 | 2021, mis en ligne le 11 décembre 2021, consulté le 14 mai 2023. URL : http://interfas.univ-tlse2.fr/lamaindethot/987

Traduire ou post-éditer la traductologie : une expérience comparative

Caroline Rossi et Alice Carré

PLAN

1. Post-éditer des articles de sciences humaines et sociales (SHS) : le jeu en vaut-il la chandelle ?
2. Une expérience comparative
   2.1 Méthodologie
   2.2 Résultats
3. Conclusion
La traductologie de corpus : une évolution quantitative ou qualitative ?
   1. Introduction
   2. Quelle place pour les corpus en traductologie ?
   3. Les corpus nous ont-ils aidés à définir notre objet d'étude ?

TEXTE

Nous remercions Dorothy Kenny pour son aide, ainsi que la direction éditoriale de la revue Journal of Translation Studies qui nous a donné la permission de traduire et de publier la traduction française de l'article original de Dorothy Kenny.

1 Si l'anglais est actuellement la langue de publication la plus courante (Bowker & Buitrago Ciro 2019), il peut sembler paradoxal que la traductologie ne fasse pas exception. Il s'agit en tout cas certainement d'un défaut qui nuit au développement d'une traductologie internationale (Susam-Sarajeva 2003, Froeliger, 2015), et l'on peut se poser la question suivante : les progrès récents de la traduction automatique (TA) permettront-ils d'établir une traductologie multilingue ?

2 Sans prétendre apporter une réponse définitive à cette question, nous proposons ici d'évaluer les progrès des systèmes à partir d'une expérience qui concerne la traduction et la post-édition vers le français de travaux de recherche écrits en anglais. Notre expérience se fonde sur un article publié il y a 15 ans, dans lequel Dorothy Kenny fait un tour d'horizon des apports de la traductologie de corpus (Kenny, 2006).

3 Pour ce texte, nous présentons d'abord le choix raisonné d'un outil de TA, en nous appuyant sur une activité proposée aux étudiants du Master de Traduction spécialisée multilingue de l'Université Grenoble Alpes dans le cadre de leur cours sur la traduction automatique. Nous décrivons ensuite la démarche, inspirée par celle d'O'Brien et al. (2018), qui a consisté à alterner traduction et post-édition. Le moteur de TA choisi est celui de la Commission européenne, dans sa version neuronale (eTranslation, « cutting edge »). La discussion des résultats porte sur les apports et limites d'outils de TA pour la production d'articles multilingues en traductologie : nous examinons quelques mesures chronométriques avant de comparer plusieurs passages post-édités et traduits.

1. Post-éditer des articles de sciences humaines et sociales (SHS) : le jeu en vaut-il la chandelle ?

4 La question de savoir si la post-édition permet vraiment de gagner du temps, tout en conservant les mêmes attentes qualitatives, n'a pas été tranchée pour les articles s'inscrivant dans le champ des SHS, pour autant que nous puissions en juger. Un grand projet européen s'est intéressé à la traduction de cours en ligne (TraMOOC) en utilisant des systèmes de TA statistique et neuronale (CASTILHO et al., 2017). Un groupe de travail « Traductions et science ouverte » s'est aussi penché récemment sur la pertinence de solutions de traduction assistée par ordinateur et automatique pour renforcer le multilinguisme dans les publications scientifiques : les SHS ne font pas partie des premières expérimentations prévues, mais le rapport suggère « d'inclure à moyen terme d'autres disciplines, comme la philosophie, la psychologie ou la linguistique » (FIORINI, 2020 : 14).

5 Les publications qui montrent un bénéfice de la post-édition évaluent souvent celui-ci à partir de textes courts (BOWKER & BUITRAGO CIRO, 2015). Dans certains travaux, les évaluations sont faites avec des moteurs entraînés pour le domaine en question, et la TA est appliquée après pré-édition du texte source, puis post-éditée par des

traducteurs professionnels (GARCIA, 2011 : 218). D'autre part, beaucoup de travaux concernent la publication en anglais langue étrangère. Ainsi, une étude récente d'O'Brien et al. (2018) montre que la TA peut aider à produire un premier jet qui ne pourra néanmoins pas se passer d'une relecture approfondie.

6 Nous nous situons dans un cadre différent, et bien en amont d'expérimentations à grande échelle, avec des solutions de traduction automatique neuronale (TAN) qui auraient pu être entraînées sur de grands corpus d'articles traduits en SHS (à condition de pouvoir identifier et rassembler ces corpus), et nous souhaitons poser ici une question simple à laquelle nous tenterons de répondre : le jeu en vaut-il la chandelle ? L'expression imagée a été choisie à dessein. Le jeu reflète en effet la dimension ludique de toute traduction. Quant à la chandelle, elle est le signe de l'énergie que nous consommons, et représente, outre l'effort de traduction ou de post-édition, la « dévoration » de données produites par des traducteurs humains qui constitue la face cachée de la TA basée sur corpus, de ses formes statistiques jusqu'aux plus récents moteurs dits neuronaux (KENNY, 2011, voir aussi KENNY et al., 2020).

2. Une expérience comparative

7 Pour évaluer la TA, on procède souvent à la comparaison de plusieurs systèmes en tenant assez peu compte du contexte de traduction ou de l'environnement textuel (TORAL et al., 2018,115). L'évaluation est certes différente selon si elle est pratiquée dans un cadre universitaire ou en entreprise (DRUGAN ,2013 in CASTILHO et al., 2018, 1), mais ces pratiques ont un dénominateur commun : on évalue au niveau du segment (qui correspond presque toujours à une phrase), avec parfois des moyennes obtenues pour un texte entier, sans qu'il existe pas à notre connaissance d'évaluation effectuée. Dans ce cadre, il n'est pas rare que les segments soient présentés dans le désordre : c'est par exemple une possibilité qui est offerte dans l'interface d'évaluation de KantanMT (LQR) et obligatoire dans l'outil PET (AZIZ et al., 2012). Si la visée, qui est d'atteindre des mesures plus fiables en évitant notamment l'effet de la présentation de segments dans le même ordre à tous les évaluateurs, est louable, les résultats risquent d'être faussés de diverses manières, et nous avons montré que nos étu-

diants de Master pouvaient éprouver des difficultés lors de tâches effectuées dans ces conditions (ESPERANÇA-RODIER et ROSSI, 2019 où des mesures ont été prises en utilisant tour à tour ces deux interfaces sont présentées).

8 S'il existe au moins un environnement de traduction assistée par ordinateur (TraduXio) qui privilégie des unités plus larges, pour répondre aux besoins de la traduction culturelle (GONCHAROVA et LACOUR, 2011), ces unités sémantiques, et les questions de cohésion textuelle qui les accompagnent, ne sont à notre connaissance pas encore intégrées à l'évaluation de sorties de traduction automatique.

9 La démarche que nous présentons ci-dessous est donc différente : elle s'inspire des « explorations » originales de Sharon O'Brien, Michel Simard et Marie-Josée Goulet (O'BRIEN et al., 2018) qui s'attachent à l'usage de la TA comme aide à la rédaction en anglais, mais qui nous ont semblées transposables à une tâche de traduction/post-édition comme la nôtre.

2.1 Méthodologie

10 Notre expérience concerne la traduction vers le français d'un article rédigé en anglais : il s'agit d'un article de Dorothy Kenny, publié en 2006 dans le *Journal of Translation Studies* et intitulé « Corpus-based Translation Studies: A Quantitative or Qualitative Development? ». Nous avons choisi un texte déjà ancien, mais cité dans plusieurs publications récentes, pour trois raisons principales : d'abord, la permission qui nous a été donnée par la revue où est publié le texte source, de traduire et de publier le texte traduit ; ensuite, la possibilité de contacter l'auteure, avec qui nous collaborons régulièrement ; et enfin, la présence d'unités terminologiques et phraséologiques intéressantes et qui nous ont semblé représentatives des travaux de traductologie contemporains publiés en anglais. Il s'agit en effet d'un article qui effectue un bilan sur les apports des corpus, au terme de 10 ans d'utilisation de ceux-ci en traductologie, et qui fait référence tour à tour à des notions de linguistique souvent utilisées dans le champ de la traductologie (*repeated patterns*, *co-text*, *cooccurrence tendencies*, *extended units of meaning*, *sampling* etc.) et à des concepts ou notions clés de la traductologie (*translation universals*, *taget text information*, *textual interventions made by translators*, etc.)

11 Nous n'avons traité que les trois premières parties, les deux suivantes étant davantage dédiées à l'analyse qualitative d'exemples issus de traductions de l'allemand vers l'anglais, plus susceptible de poser problème dans le cadre de la tâche proposée. Au total, ce sont donc environ 2000 mots de l'article original qui sont pris en compte ici. L'expérience que nous avons faite comporte plusieurs étapes : une première étape d'évaluation comparative réalisée avec les étudiants du Master de Traduction spécialisée multilingue de l'Université Grenoble Alpes, une seconde étape de traduction/post-édition alternées et chronométrées, et une dernière étape de révision par une traductrice professionnelle, co-auteure du présent article. Nous ferons le même usage de ces révisions qu'O'Brien et al. 2018, c'est-à-dire un décompte des marques de révision à des fins de comparaison.

12 Pour la première étape, voici les consignes que nous avons données aux étudiants : « Il vous est demandé d'effectuer une comparaison raisonnée d'outils de TA, située dans le contexte d'une commande. Un éditeur de revues en sciences humaines et sociales souhaite publier une traduction de l'article de D. Kenny (2006) que vous trouverez ci-dessous. Il vous consulte pour évaluer la pertinence de solutions de traduction automatique pour cette commande, en vous proposant de confronter deux outils : eTranslation (dont la sortie vous est fournie par l'éditeur) et un autre moteur de votre choix. »

13 Les étudiants avaient au préalable participé à plusieurs séances de cours sur le choix raisonné d'un outil de TA, choix qui implique aussi que l'on puisse refuser d'utiliser la TA si l'évaluation ne lui est pas favorable. La commande proposée et le niveau de qualité attendu ne permettaient pas d'envisager la pré-édition du texte source : en effet, et comme O'Brien l'a très justement fait remarquer (2010), si la préparation du texte est de nature à l'optimiser, encore faut-il veiller à ne pas mettre sur le même plan l'optimisation pour la machine (MT *friendly*) et celle qui concerne les utilisateurs finaux (*user friendly*). D'autre part, les mesures qui avaient été présentées aux étudiants et qu'ils avaient appris à utiliser au cours de ces séances sont les suivantes : durée de la post-édition (mesure chronométrique qui peut être prise par segment, ou par paragraphe ou extrait choisi), adéquation et fluidité (sur une échelle de 4 ou de 5, la seconde étant moins discriminante), jugement de préférence entre deux sorties de TA (souvent appelé « A-B *Test* » en anglais), relevé des erreurs les plus

importantes, et usage de métriques d'évaluation automatique comme BLEU (pour une présentation complète de ses mesures et de leurs usages possibles dans le cadre d'un cours sur la traduction automatique, voir Rossi & Carré, sous presse).

14 Les étudiants ont tous comparé eTranslation et DeepL, et leurs conclusions ne font pas ressortir de différence nette entre les deux sorties : elles soulignent plutôt des erreurs récurrentes de terminologie (par exemple pour *translation studies*, et *corpus/corpora* qui sont répétés tout au long de l'article. Deux choix sont majoritaires : le premier est une préférence pour DeepL, qui n'est pas justifiée par des mesures chronométriques fiables[1] mais davantage par des considérations ergonomiques (possibilité d'utiliser une base de données terminologique et d'avoir accès à d'autres propositions de traduction dans DeepL), et le second est une préférence pour la traduction humaine pour ce projet.

15 Nous avons tenu compte de ces jugements et intégré la traduction humaine à notre évaluation. En nous appuyant sur la démarche d'O'Brien et al. (2018), *mutatis mutandis*, nous avons découpé le texte en unités d'une longueur comparable et qui constituaient des unités de sens (correspondant le plus souvent aux paragraphes de l'article), et alterné traduction et post-édition en prenant des mesures chronométriques. Pour choisir une solution de traduction automatique, nous avons préféré eTranslation qui nous a semblé être une solution plus adaptée à des tâches entreprises par des chercheurs dans le service public européen (même si nous nous situons ici assez loin des domaines d'entraînement du moteur), mais aussi plus susceptible de garantir la confidentialité des données si nécessaire[2].

2.2 Résultats

16 Le tableau ci-dessous fait apparaître le nombre de mots du texte source (TS) traités dans chacun des passages (en traduction humaine, TH ou en post-édition, PE), la durée correspondante, et les modifications effectuées par notre réviseuse (marques de révision), en considérant celles-ci à l'échelle du mot (O'Brien et al. 2018 : 249), les marques de ponctuation étant elles-aussi comptées comme un mot, tandis que les espaces ajoutées ou supprimées ne le sont pas. Les déplacements de segments entiers ne sont comptés qu'une fois, et non

mot à mot, pour aboutir à une mesure qui reflète au mieux la complexité des opérations de révision. Nous avons ajouté une mesure subjective de l'effort, dont nous avons évalué l'intensité sur une échelle de 1 (très bas) à 9 (effort très élevé) : même si cette mesure ne nous semble pas extrêmement fiable, il apparaît qu'elle varie le plus souvent avec la durée (Paas 1992, p. 429-430, cité par Vieira 2016).

Passage concerné	**Nombre de mots TS**	**TH / PE**	**Durée (min)**	**Mots par minute**	**Mesure d'effort (1-9)**	**Nombre de marques de révision**	**Nombre de marques / nombre de mots en %**
1. (résumé)	193	TH	23	8,39	6	18	9,33
2	224	PE	17	13,18	3	16	7,14
3	167	TH	19	8,79	5	14	8,38
4	178	PE	18	9,89	5	27	15,17
5	163	TH	16	10,19	4	30	18,40
6	123	PE	11	11,18	3	29	23,58
7	137	TH	12	11,42	4	35	25,55
8	113	PE	12	9,42	6	2	1,77
9	199	TH	22	9,05	5	7	3,52
10	167	PE	17	9,82	4	32	19,16
11	304	TH	37	8,22	4	30	9,87
Nombre total de mots et moyenne HT (n=6)	1968	Nb. de TH : 6	21,5	9,34	4,5	22,33333333	12,51
Moyenne PE (n=5)		Nb. de PE : 5	15	10,70	4,2	21,2	13,36

17 Que nous apprennent ces mesures ? Il apparaît tout d'abord difficile de constater un bénéfice important de la post-édition, lorsque nous la comparons à la traduction humaine : les moyennes ne sont pas très différentes, et si un écart entre les mesures temporelles semble parfois observable dans notre tableau, les résultats d'une comparaison statistique de nos moyennes (test-T) montrent que ces différences ne sont pas significatives (t=0,25791 et p=0,401493, soit $p > 0,05$). Ce constat reflète bien notre perception subjective de l'effort requis, qui

ne varie pas du tout en fonction de l'alternance entre traduction humaine et post-édition.

18 Les ressemblances ainsi observées entre le travail de post-édition et celui de traduction sont peut-être liées au choix de ne pas utiliser d'outil de traduction assistée par ordinateur (TAO), et donc de travailler dans un éditeur de texte, sur un document préparé au préalable pour faire apparaître tour à tour une sortie de TA ou un passage du TS. La post-édition nous obligeait, dans cette configuration, à consulter le texte source sur un second écran, alors que pour la traduction humaine nous y avions un accès plus immédiat, sur le même document. Il faut cependant remarquer aussi que lors des tâches de traduction nous devions effacer le texte source alors que seuls les parties post-éditées étaient supprimés dans les autres passages. Nous ne répondons pas ici à la question de savoir si nous aurions travaillé mieux dans un outil de TAO, même si le découpage en segments phrastiques semble peu adapté à ce type de texte : elle aurait nécessité une autre analyse comparative.

19 En regardant de plus près les marques de révision, nous avons repéré quelques différences saillantes : les passages 8 et 9 comportent très peu révisions, à la différence du passage 7, ou encore du passage 10, qui en comportent beaucoup. Nos mesures chronométriques font apparaître un traitement plus rapide de 7 (avec environ deux mots de plus par minute), si bien que l'on peut se demander s'il comporte par exemple des phrases moins longues et/ou moins complexes, s'il s'agit d'un passage globalement plus simple du point de vue de son contenu. Le passage 7 contient effectivement 4 phrases, dont trois relativement courtes (20, 23 et 11 mots dans le TS), alors que dans les trois passages suivant, et en particulier dans les passages 9 et 10, on relève des phrases qui comportent jusqu'à 67 mots, et des nombres moyens de mots par phrase bien plus élevés, comme indiqué dans le tableau ci-dessous.

Passage	Nombre de mots par phrase (et moyenne) dans le TS
7	20, 44, 23, 11 (24,5)
8	35, 41, 28, 10 (28,5)
9	67, 31, 32, 11, 48 (37,8)
10	66, 32, 56 (51,33)

20 Il apparaît aussi que les marques de révision relativement nombreuses pour le passage 7 sont liées à la présence de nombreuses références bibliographiques dont certaines avaient été omises dans la traduction initiale, et d'autres ont été revues et adaptées. Ce sont donc bien les caractéristiques textuelles des passages en question qui permettent de rendre compte de l'effort mesuré, indépendamment du fait qu'il s'agisse d'une tâche de traduction ou de post-édition.

3. Conclusion

21 Au terme de cette expérience, nous n'avons pas observé de gain important, tant du point de vue du temps passé à traduire que de l'effort perçu, liée à l'usage d'un outil de TA (eTranslation) pour la production d'articles multilingues en traductologie. Notre expérience présente de nombreuses limites. D'abord, elle ne concerne qu'un seul article, dont les caractéristiques sont évidemment liées au style de l'auteure, aux thèmes abordés etc. Ensuite, elle ne concerne qu'une seule paire de langues et une seule direction de traduction : il faudra aller beaucoup plus loin pour développer une traductologie internationale, et nous découvrirons probablement en chemin d'autres usages possibles et fructueux de la traduction automatique. La création d'un environnement intégré pour la traduction d'articles de traductologie, qui permette la gestion de mémoires de traduction et de bases de données terminologiques, ainsi que l'accès à une solution de TAN respectueuse des données et personnalisée, serait certainement une étape importante. En l'état actuel de nos moyens, nous ne pouvons pas recommander l'usage de la traduction automatique pour d'autres raisons que celle de l'exercice pratiqué ici, et dont un bénéfice intéressant a cependant été de rompre la monotonie de la tâche. Dans le prolongement du projet de N. Froeliger (2015), nous appelons de nos vœux la mise en place d'autres expériences et d'autres dispositifs susceptibles de faciliter la publication d'articles multilingues et de faire vivre le multilinguisme au sein d'une traductologie véritablement internationale.

Annexe : traduction révisée

La traductologie de corpus : une évolution quantitative ou qualitative ?

22 Dorothy Kenny

[Passage 1. Résumé. TH]

23 Dans son article fondateur sur les apports possibles d'une traductologie de corpus, Baker (1993) proposait que les corpus constituent le socle empirique d'une nouvelle traductologie descriptive. Depuis, les corpus ont surtout été appliqués à l'étude des « universaux » (ou caractéristiques pour ainsi dire « générales ») de traduction, et à celle du style du traducteur. Nous disposons désormais d'études de cas nombreuses et approfondies, et de propositions consolidées de méthodologies basées sur corpus, mais l'apport des corpus à notre compréhension des concepts de base de la traductologie reste difficile à définir. Le tournant quantitatif s'est-il accompagné d'un tournant qualitatif, comme en linguistique (Tognini-Bonelli 1996) ? Dans cet article, je me demande si notre utilisation des corpus en traductologie depuis 10 ans a infléchi notre compréhension de certains aspects de notre objet d'étude. À partir de concepts comme les unités de traduction, l'équivalence et la voix du traducteur, j'interroge l'impact de données plus nombreuses et de nouvelles façons de les analyser, en cherchant à cerner d'éventuels ajustements ou refontes de ces concepts.

[Passage 2.PE]

1. Introduction

24 Dans son article programmatique de 1993 sur la linguistique de corpus et la traductologie, Mona Baker a suggéré que le recours aux corpus, ces grands ensembles de textes numériques, allait bouleverser la traductologie. Le champ de la traductologie avait, selon Baker, suffisamment évolué pour accueillir les corpus : le texte source n'était plus la préoccupation centrale, et l'on s'était détaché de la notion

prescriptive d'équivalence pour s'intéresser aux caractéristiques des textes traduits et à leurs contextes, notamment sous l'influence de la théorie des polysystèmes d'Even-Zohar (1979) et de la traductologie descriptive de Toury (1980, 1995). En disposant de grands corpus de textes traduits, accompagnés de leurs originaux, les théoriciens de la traduction seraient à même de mieux définir et expliquer leur objet de recherche, puisqu'ils auraient « l'occasion inédite d'observer leur objet d'étude et de découvrir ce qui le distingue des autres, notamment de la langue générale et de tout autre type d'interaction culturelle » et « d'explorer, à une échelle jamais encore atteinte, les principes qui régissent le comportement traductionnel et les contraintes à l'œuvre »3 (Baker 1993 : 235). Les corpus, espérait-on, allaient aider la traductologie à passer « des énoncés normatifs aux énoncés descriptifs, de considérations méthodologiques à une véritable théorisation, et de recherches isolées et fragmentaires à des généralisations efficaces »4 (ibid. : 248).

[Passage 3. TH]

25 On comprend aisément l'enthousiasme de Baker si l'on considère les bouleversements opérés par les corpus dans d'autres disciplines comme la lexicographie et la linguistique. En 1987 était paru le premier dictionnaire de l'anglais basé sur corpus (Sinclair 1987), et les corpus avaient très rapidement commencé à occuper une place de choix. Ils sont désormais au cœur de la compilation de presque tous les dictionnaires (Čermák 2003 : 18). En linguistique, même si l'utilité des corpus avait d'abord été mise en doute, on a commencé dès le début des années 1990 à s'appuyer sur des productions langagières authentiques, et les linguistes de corpus se réjouissaient de voir que les corpus étaient de plus en plus plébiscités (Svartvik 1996). Très vite, on a commencé à considérer que les généralisations linguistiques qui n'étaient pas confirmées par des données de corpus n'étaient pas satisfaisantes (voir par ex. Aitchison 2005 : 185). Les corpus en sont également venus à occuper une place prépondérante en linguistique computationnelle (cf. les travaux non publiés de Julia Hirschberg, cités par Sampson 2005), et l'induction automatique à partir de corpus est désormais perçue, entre autres, comme la seule manière fiable de créer des grammaires computationnelles et de dépasser les limites de l'acquisition de connaissances en traduction automatique.

[Passage 4. PE]

26 En lexicographie et en lexico-grammaire, cette branche de la linguistique qui considère l'interaction entre les mots et les structures grammaticales dans lesquelles ils se trouvent habituellement (et cette interaction n'intéresse pas seulement la linguistique de corpus, mais aussi d'autres domaines de la linguistique théorique et computationnelle – voir par ex. Pustejovsky 1995 : 5), la révolution des corpus implique d'abord que les linguistes se tournent désormais vers de grandes bases de données, faciles à consulter et constituées avec une visée généraliste, pour valider leurs hypothèses sur le langage. Elle a aussi et surtout opéré un changement qualitatif, lié à la quantité de données disponibles : les données de corpus ont en effet conduit les linguistes à remettre en question l'hypothèse selon laquelle le lexique pouvait être séparé de la grammaire ; la forme, du sens ; ou même l'occurrence, du système. Les lignes de concordance, ces ensembles sur lesquels se fonde la pratique du linguiste de corpus, ont été utilisées pour montrer que les rapports verticaux du langage devenaient visibles lorsque les rapports horizontaux étaient suffisamment déployés, si bien que les relations paradigmatiques et syntagmatiques de Saussure (1916) étaient en fait deux façons de voir les mêmes données. En d'autres termes, « l'effet cumulatif des occurrences répétées est considéré comme le reflet du système sémiotique »6 (Tognini-Bonelli 1996 : 71).

[Passage 5. TH]

27 C'est en analysant les relations syntagmatiques que Sinclair (1996) est parvenu à la proposition suivante : « les tendances à la cooccurrence des mots, des phénomènes grammaticaux, des mots ou syntagmes appartenant aux mêmes champs lexicaux et des mots ou syntagmes véhiculant la même prosodie sémantique sont si fortes qu'il nous faut adopter une vision plus large des unités de sens et admettre qu'elles dépassent en extension et en variété ce que l'on observe au niveau du mot »7 (Sinclair 1996 : 94). Il revient en particulier à Michael Stubbs (2001, 2002) d'avoir ensuite utilisé des données de corpus pour tenter de quantifier la participation de lexèmes fréquents de l'anglais à de telles unités étendues de signification. Les données de corpus ont ainsi apporté des preuves à l'appui de propositions ambi-

tieuses qui concernaient l'objet d'étude de la linguistique et les unités d'analyse applicables à des productions langagières authentiques. L'importance des corpus pour la lexicographie contemporaine, le traitement automatique des langues et la linguistique européenne ne fait aucun doute (Teubert 2001 : 125).

[Passage 6. PE]

28 Au terme d'une décennie d'utilisation des corpus en traductologie, il semble donc opportun de s'interroger sur leurs apports à la discipline : les corpus sont-ils devenus indispensables ? Le « tournant de la linguistique de corpus en traductologie » 8 (Laviosa 2004 : 9) s'applique-t-il à la seule traductologie de corpus, ou ses ramifications se sont-elles propagées à l'ensemble de la discipline ? Les corpus nous ont-ils aidés à définir notre objet d'étude et à en expliquer certaines spécificités (si toutefois les traductions présentent des caractéristiques spécifiques), comme Baker (1993) l'avait prédit ? Nous ont-ils obligés à repenser l'une ou l'autre des catégories analytiques que nous appliquons aux traductions, ou nous ont-ils apporté quoi que ce soit de nouveau sur la façon d'étudier le style du traducteur ? Dans cet article, j'essaie de répondre à certaines de ces questions, en commençant par délimiter la place qu'ont pris les corpus dans la traductologie en général.

[Passage 7. TH]

2. Quelle place pour les corpus en traductologie ?

29 À l'évidence, les travaux qui s'inscrivent dans le champ de la traductologie de corpus constituent une somme remarquable. Depuis le milieu des années 1990, plusieurs monographies (voir par ex. Kenny 2001 ; Laviosa 2002 ; Olohan 2004), ouvrages collectifs (par ex. Zanettin et al. 2003 ; Mauranen et Kujamäki 2004) et numéros spéciaux de revues (Laviosa 1998 ; Tagnin 2002 ; Kruger 2004a) ont été publiés et, en août 2006, la *Bibliography of Translation Studies* recensait 370 publications en traductologie de corpus (à noter que ce chiffre comprenait probablement des travaux qui se fondaient sur des

corpus non numériques). Il faut y ajouter le nombre croissant de thèses de doctorat inscrites dans ce champ, ainsi qu'un ensemble important de conférences et sessions thématiques (cf. Kruger 2004b : 2). La traductologie de corpus a bel et bien acquis ses lettres de noblesse.

[Passage 8. PE]

30 La contribution plus générale des corpus à la traductologie a également été reconnue par des théoriciens de la traduction dont les travaux se situent en dehors du champ de la traductologie de corpus, notamment lorsqu'ils présentent une vue d'ensemble de la discipline. Les corpus ne sont cités que brièvement dans l'introduction à la traductologie de Munday (2001), et l'auteur suggère que leur utilisation en traductologie était encore « limitée »9 au début des années 2000, alors que dans son anthologie de la traductologie, Venuti (2000 : 335) évoque l'apport des « puissants outils d'analyse »10 issus de la linguistique de corpus. Prudent et audacieux à la fois, Mason (2004) décrit la traductologie de corpus comme « ce qui constitue, pour de nombreux chercheurs, l'avenir de la traductologie empirique »11.

[Passage 9. TH]

31 Ce mélange d'optimisme et de prudence est probablement de rigueur. Certes, il est désormais évident que les requêtes (semi) automatiques sur des corpus numériques sont parfaitement adaptées aux besoins des chercheurs qui étudient des faits de langue (1) aux caractéristiques formelles fortement prévisibles et (2) si fréquents que même l'analyse des données extraites d'un corpus de taille modeste doit être outillée. En revanche, pour de nombreuses caractéristiques textuelles dont l'étude est tout aussi utile, les formes linguistiques sont beaucoup moins prévisibles et donc bien moins faciles à détecter à l'aide d'outils de corpus. De plus, même si les outils de corpus peuvent être utilisés pour repérer des exemples, ils n'en permettent pas l'analyse approfondie, et ils n'autorisent pas non plus la contextualisation de textes entiers. Ces deux tâches restent l'apanage de l'analyste humain. C'est ainsi que les chercheurs qui ne pratiquent pas la traductologie de corpus en surveillent les progrès, d'un œil à la fois sceptique et intéressé, et que nous ne pouvons pas affirmer que la

traductologie empirique ait entièrement fait siennes les techniques de la linguistique de corpus, pas plus que nous ne recommanderions leur adoption pour l'ensemble des travaux empiriques du domaine.

[Passage 10. PE]

3. Les corpus nous ont-ils aidés à définir notre objet d'étude ?

32 La question de la représentativité a ouvert la voie à une première tentative de définition des objets d'étude de la traductologie dans la perspective de la création de corpus : Halverson (1998) a proposé de constituer des corpus de traductions dont les éléments centraux seraient des traductions « prototypiques », c'est-à-dire effectuées par des traducteurs professionnels (du moins pour les cultures occidentales) et qui comprendraient aussi d'autres traductions (de statut inférieur) à la périphérie. Tymoczko (1998) a en revanche rejeté l'idée que les corpus deviennent un moyen d'institutionnaliser certaines pratiques de traduction ou de valoriser un type de traduction par rapport à un autre, arguant qu'ils pourraient favoriser une traductologie du décentrage, inclusive et dynamique. Si définir notre objet d'étude signifiait fixer les limites de ce que l'on appellerait « traduction », alors ce projet faisait fausse route, d'après Tymoczko, et il ne semble d'ailleurs pas avoir fait l'objet de travaux nombreux en traductologie de corpus. Cependant, l'échantillonnage demeure une vraie question pour tous les créateurs de corpus.

[Passage 11. TH]

33 L'étude des caractéristiques les plus fréquentes de la traduction constitue une seconde piste susceptible de distinguer la traduction d'autres objets d'étude. Dans son article de 1993, Baker proposait d'utiliser les corpus pour explorer certaines régularités qui semblaient caractéristiques de la traduction au point de mériter le nom d'« universaux ». Alors que l'idée même d'« universaux » fait débat, l'ensemble des recherches menées pour les révéler a permis de rassembler quantité de données à interpréter (Mauranen et Kujamäki 2004 : 2). De nombreuses études ont montré, par exemple,

que la tendance à l'explicitation était très répandue en traduction. On retrouve cette caractéristique dans des traductions où s'illustre l'explicitation grammaticale, entre autres, et dans des traductions qui rendent explicites des informations implicites dans le texte source. Un exemple désormais célèbre d'explicitation grammaticale est celui d'Olohan et Baker (2000), qui ont montré que le connecteur (ou « complémenteur ») *that* était bien plus fréquent dans les textes traduits vers l'anglais que dans ceux qui avaient été rédigés d'emblée en anglais. Dans le prolongement de cette première étude, un article de Kenny (2005) nous fournit un exemple d'explicitation d'informations restées implicites dans le texte source : alors que l'omission de *that* après le verbe *say* dans des textes traduits en anglais était souvent le reflet, dans les textes sources allemands étudiés, de l'omission du connecteur *daß* (ou *dass*, selon la nouvelle orthographe), qui n'est pas non plus obligatoire, les patrons d'occurrence de *that* ne reflétaient d'aucune manière ceux de *daß* : il s'agissait bien plutôt d'une tendance des traducteurs à l'introduction de *that* dans leurs textes indépendamment des caractéristiques des textes sources. Outre ces études, d'autres travaux ont suggéré qu'il faille accorder une place particulière à l'explicitation dans le répertoire des interventions textuelles du traducteur, même s'il ne faudrait pas y voir une caractéristique exclusive des textes traduits.

NOTE DE FIN

1 Cet exercice nous a en effet permis de constater que nos étudiants n'avaient pas compris qu'ils n'obtiendraient pas de mesures fiables en retraduisant le même passage, puisque le premier travail de PE rendait presqu'inévitablement le second plus rapide.

2 En l'occurrence, si nous disposons de l'autorisation de traduire le texte et d'en publier la traduction, cela ne signifie pas pour autant que l'éditeur nous autorise à mettre dans DeepL le texte source, sur lequel nous n'avons en réalité aucun droit.

3 « [by providing them] with a unique opportunity to observe the object of their study and to explore what it is that makes it different from other objects of study, such as language in general or indeed any other kind of cultural interaction » et « to explore, on a larger scale than was ever possible before, the principles that govern translational behaviour and the constraints under which it operates ».

4 « prescriptive to descriptive statements, from methodologising to proper theorising, and from individual and fragmented pieces of research to powerful generalisations ».

5 Il reste cependant des champs de la linguistique dans lesquels les chercheurs ne se fondent pas sur des données de corpus.

6 « the cumulative effect of repeated instances is taken to reflect the semiotic system ».

7 « so strong are the cooccurrence tendencies of words, word classes, meanings and attitudes that we must widen our horizons and expect the units of meaning to be much more extensive and varied than is seen in a single word ».

8 « corpus linguistic turn in translation studies ».

9 « limited ».

10 « powerful analytic tools ».

11 « what many would see as the future of empirical studies of translating ».

AUTEURS

Caroline Rossi
Univ. Grenoble Alpes, ILCEA4caroline.Rossi@univ-grenoble-alpes.fr

Alice Carré
Univ. Grenoble Alpes, ILCEA4alice.Carre@univ-grenoble-alpes.fr