

## Réexplorer la langue de la traduction : une approche par corpus

Mona Baker

Volume 43, Number 4, décembre 1998

L'approche basée sur le corpus  
The Corpus-based Approach

URI: <https://id.erudit.org/iderudit/001951ar>  
DOI: <https://doi.org/10.7202/001951ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)  
1492-1421 (digital)

[Explore this journal](#)

Cite this article

Baker, M. (1998). Réexplorer la langue de la traduction : une approche par corpus. *Meta*, 43(4), 480–485. <https://doi.org/10.7202/001951ar>

Article abstract

This paper discusses the need to develop a coherent corpus-based methodology for identifying the distinctive features of the language of translation. The aim of this endeavour is not merely to unveil the nature of the 'third code' per se, but most importantly, to understand the specific constraints, pressures, and motivations that influence the act of translating and underlie its unique language.

# RÉEXPLORER LA LANGUE DE LA TRADUCTION : UNE APPROCHE PAR CORPUS

MONA BAKER

UMIST, Manchester, Royaume-Uni

## **Résumé**

*L'approche basée sur le corpus requiert le développement d'une méthodologie cohérente pour l'identification des caractéristiques propres à la langue de la traduction. En plus de corroborer l'hypothèse du «troisième code», cette méthodologie servira à comprendre les contraintes, les pressions et les motivations qui influencent spécifiquement l'acte traductionnel et sous-tendent cette forme unique de communication.*

## **Abstract**

*This paper discusses the need to develop a coherent corpus-based methodology for identifying the distinctive features of the language of translation. The aim of this endeavour is not merely to unveil the nature of the 'third code' per se, but most importantly, to understand the specific constraints, pressures, and motivations that influence the act of translating and underlie its unique language.*

La plupart des linguistes ont tendance à considérer la traduction — quand ils se donnent la peine d'y réfléchir — comme un sous-domaine de la linguistique appliquée. C'est dire que la plupart d'entre eux n'ont pas encore admis que la «théorie de la traduction» est une théorie autonome. Pour eux, traduire n'est rien d'autre qu'une aptitude linguistique qui peut être «affinée» par une exposition aux structures naturelles de la langue cible et par une meilleure appréciation des subtilités de la langue source. Paradoxalement, la linguistique offre certains des outils les plus performants qui soient pour prouver qu'il n'en est rien. En effet, la traduction est une forme unique de communication linguistique et culturelle parce qu'elle implique bien plus que la simple compréhension des subtilités d'une langue et de la structuration de la langue source et de la langue cible. Les caractéristiques propres à la traduction la rendent donc apte à devenir l'objet d'une discipline indépendante que nous appelons traductologie.

Deux traits spécifiques suggèrent que la traduction constitue vraisemblablement un comportement linguistique et culturel unique, de sorte que sa structuration propre doit apparaître au niveau de surface<sup>1</sup>. Le premier trait, c'est qu'un texte/énoncé traduit est normalement contraint par un texte (ou énoncé) entièrement articulé dans une autre langue. Cela ne se produit pas dans d'autres types de productions linguistiques. Le second trait est que les traducteurs ont tendance, consciemment ou non, à tenir compte de la perception du statut social du texte qu'ils produisent. Ils savent que les traductions ne sont pas perçues comme des textes originaux dans la plupart des sociétés contemporaines, même si, bien entendu, cette perception diffère d'un contexte social et historique à un autre<sup>2</sup>. La prise de conscience de ce contexte spécial de réception permet de mieux comprendre pourquoi les textes traduits ont tendance à se conformer aux caractéristiques typiques de la langue cible et même à les exagérer.

La traduction doit donc être reconnue comme un «communicative event which is shaped by its own goals, pressures and context of production» (Baker 1996 : 175). Si nous acceptons que toute structuration linguistique est influencée par les contraintes liées à l'environnement et à une époque donnée de l'utilisation de la langue, et ce à n'importe quel moment, alors nous devons admettre que la structuration d'un texte traduit diffère de celle d'un original. En d'autres mots, la nature et les pressions du processus traductionnel doivent laisser des traces dans la langue utilisée par les traducteurs (Baker 1996 : 176). La question est de savoir comment étudier ces traces concrètement. La méthodologie employée ici ne consiste pas à comparer des textes sources à leurs traductions, mais plutôt à comparer des textes originaux et des traductions *dans une même langue* et dans des domaines apparentés. Les deux ensembles de textes, sous forme électronique et donc utilisables par traitement automatique ou semi-automatique, sont appelés «corpus comparables» (*comparable corpus*, cf. Baker 1995). Plusieurs chercheurs en traduction sont en train de rassembler un «corpus comparable» de l'anglais de traduction. Leur but est d'identifier les traits qui distingueraient la langue anglaise de traduction de celle des usagers ordinaires (Baker 1996; Laviosa 1997, 1998; Kenny 1997).

De nombreuses affirmations touchant la langue de traduction ont été avancées ces dernières années. Elles doivent être corroborées par l'analyse d'un grand nombre de textes traduits de toutes sortes. Le «Corpus traductionnel anglais» (*Translational English Corpus*, Laviosa 1997), de l'UMIST, actuellement subventionné par l'Académie britannique, contient les textes traduits de romans, de biographies, d'articles de journaux et de périodiques de lignes aériennes. Par exemple, Anthony Pym devait affirmer, dans l'*Oxford Guide to Literature in English Translation* (à paraître) : «translations into English have tended to be in a language that is less specific, more international, than most works originally written in English». Les chercheurs engagés dans l'analyse basée sur le corpus de textes traduits souhaitent étudier de près ce genre d'affirmation : ils veulent, entre autres, identifier les éléments tangibles qui rendent la langue de l'anglais de traduction moins spécifique.

#### LA NOTION DE «TROISIÈME CODE»

L'idée que la langue de traduction diffère de la langue courante a été, et est toujours, débattue dans les écrits en dehors des études sur le corpus. Cependant, abstraction faite des travaux sur les études de corpus, aucune étude à grande échelle n'a examiné ce phénomène. Le premier à avoir parlé de «la langue de traduction» comme ayant une existence autonome est Frawley (1984). Ce dernier suggère que la confrontation du texte source et de la langue cible pendant le processus de traduction crée ce qu'il appelle un «troisième code» (*third code*). En d'autres mots, le code (ou la langue) qui évolue pendant la traduction, et dans lequel le texte cible est rédigé, serait unique. Il s'agit d'un compromis entre les normes ou structures de la langue source et ceux de la langue cible. Bernard Dulsey en fournit un exemple concret dans sa traduction vers l'anglais du roman équatorien de Jorge Icaza, *Huasipungo* (Mafla Bustamante 1992). Le mot *pes*, un dérivé du *pues* espagnol, y est directement emprunté. *Pes* peut être traduit en anglais par *well* ou *then* et peut servir simplement à ponctuer le discours. On en observe deux types dans la traduction anglaise. Le premier est un emprunt direct comme dans «Dónde estáis, *pes*?» / «Where are you, *pes*?». Ici, *pes* apparaît en fin de phrase, là où il apparaîtrait normalement dans le dialecte équatorien et où *pues* serait placé en espagnol. Le deuxième type, dans la même traduction, suit la norme du *well* anglais comme dans «Dos o tres veces he sido capataz, *pes*» / «*Pes*, I have been a fore-

man a few times». En fait, cette seconde façon de traduire est une tentative de compromis entre le lexique du texte source et la syntaxe de la langue cible. Dans les deux cas, nous sommes en présence du «troisième code» dont parlait Frawley : on a affaire à une structure où le terme emprunté n'apparaît pas dans son environnement original, ou bien l'équivalent du terme emprunté ne saurait être présent dans l'environnement cible.

La notion de troisième code se révèle utile pour rendre compte de certains des problèmes auxquels les traductologues se heurtent en essayant d'appliquer les techniques de la linguistique du corpus à l'étude de la langue de traduction. En effet, elle permet à Frawley d'aborder sous un angle descriptif et théoriquement complexe les caractères propres à la langue de traduction. Il n'est pas question ici de parler de «jargon de traduction» (*translationese*) avec toutes ses connotations péjoratives. La traduction crée un troisième code parce qu'elle est une forme de communication unique, et non parce qu'elle est une forme de communication fautive, déviante ou non conforme à la norme.

La notion de troisième code doit être examinée plus à fond si nous voulons rendre compte des particularités de la langue de traduction d'une façon plus nuancée en tant que phénomène distinct. Par exemple, il est indéniable que la coprésence de codes n'est pas la seule contrainte à jouer en traduction : d'autres contraintes s'exercent et contribuent à la structuration distinctive du texte traduit. De plus, la nature même de cette coprésence de codes demande à être clarifiée. Cette exigence découle de ce que la simple présence de deux codes dans un même événement ne suffit pas à distinguer la traduction d'autres processus, tels un énoncé ou le texte écrit par un apprenant de langue seconde, où on reconnaît également que l'apprenant ou l'interlocuteur emploie consciemment deux codes.

Plus important encore, les chercheurs travaillant dans ce domaine reconnaissent la nécessité de développer une méthodologie pour l'étude des particularités de la langue de traduction, et cette méthodologie n'est pas encore disponible dans les références actuelles sur la traduction, y compris celles qui traitent des études basées sur le corpus. L'accès, sous forme électronique, à un grand nombre de textes traduits et l'utilisation d'un ensemble d'outils informatiques pour les traiter n'est qu'un point de départ. Il faut encore créer une méthode pour traiter les données, donc savoir ce que l'on cherche et comment le chercher. Finalement, une description détaillée de la langue de traduction ne sera utile que si nous pouvons émettre quelques hypothèses sur la motivation qui sous-tend certains types de structures.

Par exemple, admettons, comme on l'affirme souvent, qu'il existe une tendance à la normalisation dans l'anglais de traduction. Ainsi, les traducteurs vers l'anglais auraient tendance à être plus conservateurs dans leur usage de la langue, en utilisant très souvent les structures essentielles de l'anglais et en évitant le plus possible les structures plus marquées. En fait, il existe des preuves de cette tendance dans le *Translational English Corpus*, mais ce dernier ne comporte pas encore assez de données pour vraiment corroborer des hypothèses fortes : le corpus ne compte qu'environ trois millions de mots, ce qui est très peu pour la linguistique du corpus. Mais il en ressort un trait spécifique : il y a une homogénéité frappante entre les textes traduits vers l'anglais lorsqu'on les compare aux textes rédigés d'abord en anglais et portant sur le même sujet. En effet, peu importe la partie de la structure linguistique étudiée, les textes traduits semblent se comporter de la même façon alors que les textes originaux montrent plus de diversité.

Une des explications possibles de ce phénomène, s'il était largement corroboré par les textes traduits, est que les traducteurs réagissent inconsciemment à ce qu'ils perçoivent comme le statut textuel et social de la traduction. Ils savent que les traductions

ne sont pas reçues de la même façon que les originaux. En littérature, les lecteurs et les critiques ont des critères différents selon qu'il s'agit de juger de la qualité d'une traduction ou de celle d'une œuvre originale. La réception de l'*Ulysse* de Joyce aurait certainement été différente si l'œuvre avait été présentée à la communauté littéraire comme la traduction d'une autre langue. Les traducteurs vont donc inconsciemment surveiller leur rendement linguistique afin qu'il soit conforme aux attentes des lecteurs et des critiques. C'est ainsi qu'ils en viennent à utiliser largement les structures les plus «sûres», les plus typiques de la langue.

Pym (à paraître) propose une autre source de motivation pour expliquer le type de structuration que nous commençons à identifier dans les romans anglais. Les traductions vers l'anglais ont eu tendance à utiliser une forme d'anglais moins spécifique, plus internationale, à cause du statut de la langue sur le marché international et aussi parce que «translators and publishers set about to address many regions at once». Ceci pourrait expliquer certains types de structures que l'on observe dans le *Translational English Corpus*. Par exemple, les mots les plus fréquents en anglais sont considérablement plus fréquents en anglais de traduction. Cela vaut tant pour les textes journalistiques que les romans et les nouvelles. De plus, les deux rapports types/occurrences et densité lexicale sont considérablement plus faibles en anglais de traduction (Laviosa-Braithwaite 1995, 1996). Le rapport types/occurrence mesure la variété du vocabulaire utilisé dans un texte ou dans un corpus : plus le rapport est élevé plus le vocabulaire utilisé est varié, plus il est faible, moins un texte utilise de mots différents. La densité lexicale mesure la redondance contenue dans un texte : elle considère la proportion d'éléments lexicaux ou mots pleins par rapport aux éléments grammaticaux. Plus la proportion d'éléments lexicaux tels que *love* ou *peace* est élevée, plus la proportion d'éléments grammaticaux tels que *in* ou *the* est faible et, donc, plus le texte est dense. Inversement, plus la proportion d'éléments grammaticaux est élevée, plus le texte est redondant et plus il se prête facilement au traitement automatique.

Ces observations suggèrent que les traducteurs pourraient simplifier inconsciemment la langue qu'ils utilisent pour faire passer un message étranger à un nouveau public : par la redondance dans le premier cas et par la limitation de la variété du vocabulaire dans le second. De même, les traducteurs ont tendance à normaliser la ponctuation, même lorsqu'elle est de nature expérimentale dans le texte original (May 1997; Malmkjær 1997), et les interprètes ont tendance à terminer les phrases laissées en suspens, à grammaticaliser les énoncés agrammaticaux et à omettre les hésitations et les faux départs même si l'orateur les produit volontairement (Shlesinger 1991 : 150; Berk-Seligson 1990). Cette tendance à «normaliser» la langue de traduction et à suivre de près les conventions de la langue cible laisse croire, encore une fois, que les traducteurs répondent inconsciemment à la perception qu'ils ont du statut du texte ou de l'énoncé qu'ils produisent.

#### LES RÉGULARITÉS DANS LES ÉTUDES BASÉES SUR LE CORPUS

Certains traducteurs et traductologues pourront trouver que cette approche qui semble mettre l'accent sur des traits tels que la normalisation ou la structure récurrente en général est inquiétante, car, à première vue, elle semble ne pas tenir compte du comportement individuel ou créatif qui caractérisent pourtant les textes traduits, particulièrement les textes littéraires. Il convient de noter que bien que les études basées sur le corpus s'intéressent d'abord aux régularités, elles ne s'intéressent pas moins à la créativité. Bien au contraire. John Sinclair, un des principaux chercheurs dans la linguistique du corpus, explique que l'intérêt premier pour les phénomènes répétitifs «does not

mean that unique, one-off events are necessarily ignored, but rather that they cannot be evaluated in the absence of an interpretative framework provided by the repeated events» (1996 : 81). Mais la créativité, associée en particulier aux textes littéraires, ne peut s'expliquer qu'en regard de la structure standard de la langue.

En voici un exemple : Louw (1993) décrit un phénomène de structure linguistique connue maintenant sous l'appellation de «prosodie sémantique». Par prosodie sémantique, on entend le sens constant qu'un élément lexical acquiert par son association répétée avec d'autres éléments de la langue; elle exprime souvent un sens comportemental. Louw se sert de cette notion pour expliquer un trait du poème *Days* de Philip Larkin dont les premiers vers sont : «What are days for? / Days are where we live». «Days are where we live» devrait exprimer des associations mélioratives pourtant, il s'agit de «a line which leaves the reader with inexplicable feelings of melancholia» (Louw 1993 : 162) et qui annonce le thème de la mort, lequel émerge dans la seconde moitié du poème. L'étude des occurrences de *days* dans un corpus de 37 millions de mots permet à Louw d'affirmer que ce mot est intimement associé à des éléments tels que *gone, over, good and over*. Notre expérience de la langue, notre exposition régulière à ce type de structure nous font inconsciemment associer *days* au lieu où nous étions, celui où nous avons vécu et où, peut-être, nous ne pouvons retourner, et non pas au lieu où nous sommes.

De même, l'intérêt actuel des études traductologiques basées sur le corpus pour des phénomènes tels que la normalisation et la simplification ne doivent pas laisser croire à un manque d'intérêt pour le comportement créatif présent/absent du traducteur. Il existe des textes dans le *Translational English Corpus* qui se distinguent nettement des autres par la longueur moyenne des phrases, le rapport types/occurrences, etc. C'est précisément la disponibilité d'une toile de fond d'événements linguistiques répétés dans le corpus qui permet de décrire ces textes comme «différents» et donc «intéressants», de sorte qu'ils méritent d'être étudiés plus en profondeur.

Les analyses de traductions basées sur le corpus doivent donc faire la part entre le général et le spécifique, entre la norme et l'exception. Certains chercheurs étudiant les structurations des textes traduits finissent par les traiter comme des règles absolues. Cette tendance est inquiétante et nous espérons qu'elle ne deviendra pas une pratique représentative de l'analyse des traductions basée sur le corpus. Évidemment, ce problème découle du fait qu'une fois qu'on dispose d'un très grand nombre de textes en format électronique et qu'on peut créer du bout du doigt toutes sortes de statistiques et d'études des fréquences, la tentation est forte de mettre l'accent sur la norme, sur ce qui est spécifique, aux dépens du comportement présent/absent, de l'utilisation de la langue la plus créative, au point, parfois, de décrire cette utilisation comme «fautive» parce qu'elle dévie de la norme. Mais, comme nous l'avons dit plus haut, une des raisons qui nous poussent à étudier la structuration de toutes sortes de productions textuelles, y compris la traduction, est que les structures constituent la toile de fond qui donnera forme à la créativité : les normes autorisent l'usage créatif de la langue et leur identification permet non seulement de cerner les traits récurrents de la traduction, et par conséquent de comprendre la traduction comme un phénomène en soi, mais elle nous autorise également à comprendre l'exemple original.

## CONCLUSION

Le texte traduit a toujours été traité en parent pauvre de la linguistique de corpus. Il a été exclu des corpus monolingues où il est généralement considéré comme non représentatif de la langue étudiée, quel que soit le sens de la traduction. Même un texte

traduit dans sa propre langue ne trouvera normalement pas de place dans un corpus monolingue. Si on a étudié le texte traduit, c'était pour montrer que le jargon traductionnel est courant ou que certains faits de langue qu'étudie le linguiste du corpus ont été influencés par une autre langue. Les traductologues, qui travaillent sur une description de la langue de traduction, tentent de dissiper cette impression de «déviance» en créant une méthodologie cohérente qui permette d'identifier les traits distinctifs et qui lie la description à un jeu de contraintes et de motivations influençant le comportement traductionnel.

La disponibilité de techniques développées par les chercheurs en linguistique de corpus fournit un point de départ raisonnable pour ce genre de recherche, mais la méthodologie pour l'interrogation d'un corpus en vue d'identifier des caractéristiques abstraites, par exemple la normalisation, reste à faire. De plus, les chercheurs doivent prendre garde de ne pas confondre les objectifs des études traductologiques avec ceux de la linguistique de corpus (ou de les y réduire). Pour les études traductologiques, une description détaillée des caractéristiques linguistiques n'est pas une fin en soi : c'est plutôt un moyen en vue d'une fin, soit la compréhension des pressions et contraintes qui font partie intégrante de la vie des traducteurs et qui laissent inévitablement des traces dans leur travail.

#### Notes

1. Cette discussion présume que la structuration linguistique de surface est conditionnée par un jeu de contraintes de nature essentiellement sociale, cognitive et culturelle dont elle est le reflet.
2. Il paraît important de souligner que cette difficulté est un problème de perception et non pas d'essence. On pourrait affirmer que les traductions n'ont pas à être différentes de ce que l'on nomme l'«original», mais le fait est qu'elles sont souvent perçues comme telles.

#### REFERENCES

- BAKER, Mona (1993) : «Corpus Linguistics and Translation Studies: Implications and Applications», Mona Baker, Gill Francis & Elena Tognini-Bonelli (Eds), *Text and Technology: In Honour of John Sinclair*, Amsterdam & Philadelphia, John Benjamins, pp. 233-250.
- BAKER, Mona (1995) : «Corpora in Translation Studies: An Overview and Some Suggestions for Future Research», *Target*, 7 (2), pp. 223-243.
- BAKER, Mona (1996) : «Corpus-based Translation Studies: The Challenges that Lie Ahead», Harold Somers (Ed.), *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sager*, Amsterdam & Philadelphia, John Benjamins, pp. 175-186.
- BERK-SELIGSON, Susan (1990) : *The Bilingual Courtroom: Court Interpreters in the Judicial Process*, Chicago and London, The University of Chicago Press.
- KENNY, Dorothy (1997) : «(Ab)normal Translations: a German-English Parallel Corpus for Investigating Normalization in Translation», Barbara Lewandowsk-Tomaszczyk and Patrick Janes Melia (Eds), *Practical Applications in Language Corpora. PALC'97 Proceedings*, pp. 387-392.
- FRAWLEY, William (1984) : «Prolegomenon to a Theory of Translation», William Frawley (Ed.), *Translation: Literary, Linguistic, and Philosophical Perspectives*, London & Toronto, Associated University Presses, pp. 159-175.
- LAVIOSA, Sara (1998) : «The English Comparable Corpus: A Resource and a Methodology», Lynne Bowker, Michael Cronin, Dorothy Kenny & Jennifer Pearson (Eds), *Unity in Diversity: Current Trends in Translation Studies*, Manchester, St. Jerome Publishing.
- LAVIOSA, Sara (1997) : «How Comparable Can Comparable Corpora Be?», *Target*, 9 (2), pp. 289-319.
- LAVIOSA-BRAITHWAITE, Sara (1995) : «Comparable Corpora: Towards a Corpus Linguistic Methodology for the Empirical Study of Translation», M. Thelen and B. Lewandoska-Tomaszczyk (Eds), *Translation and Meaning (Part 3)*, Maastricht, Hogeschool Maastricht.
- LAVIOSA-BRAITHWAITE, Sara (1996) : *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation*, Ph.D. Thesis, Manchester, Department of Language Engineering, UMIST.

- LOUW, Bill (1993) : «Irony in the Text or Insincerity in the Writer. The Diagnostic Potential of Semantic Prosodies», Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds), *Text and Technology: In Honour of John Sinclair*, Amsterdam, John Benjamins, pp. 157-176.
- MAFLA BUSTAMANTE, Cecilia (1992) : *Dialects in Translation. A Study of the English Translations of Icaza's "Huasipungo"*, Unpublished MA Dissertation, School of English, University of Birmingham, UK.
- MALMKJÆR, Kirsten (1997) : «Punctuation in Hans Christian Andersen's stories and in their translations into English», Fernando Payotas (Ed.), *Nonverbal Communication and Translation. New Perspectives and Challenges in Literature, Interpretation and the Media*, Amsterdam, John Benjamins.
- MAY, Rachel (1997) : «Sensible Elocution: How Translation Works in and upon Punctuation», *The Translator*, 3 (1), pp. 1-20.
- PYM, Anthony (à paraître) : «Titre à déterminer», Peter France (Ed.), *The Oxford Guide to Literature in English Translation*, Oxford, Oxford University Press.
- SHLESINGER, Miriam (1991) : «Interpreter Latitude vs. Due Process. Simultaneous and Consecutive Interpretation in Multilingual Trials», Sonja Tirkkonen-Condit (Ed.), *Empirical Research in Translation and Intercultural Studies*, Tübingen, Gunter Narr, pp. 147-155.
- SINCLAIR, John (1996) : «The Search for Units of Meaning», *Textus*, IX, pp. 75-106.