



**HAL**  
open science

**Traduction pragmatique, linguistique de corpus,  
traducteur : un ménage à trois explosif?. Session 3 - TA  
et Biotraduction**

Natalie Kübler

► **To cite this version:**

Natalie Kübler. Traduction pragmatique, linguistique de corpus, traducteur : un ménage à trois explosif?. Session 3 - TA et Biotraduction. Tralogy II. Trouver le sens : où sont nos manques et nos besoins respectifs?, Jan 2013, Paris, France. 15p. hal-02497335

**HAL Id: hal-02497335**

**<https://hal.science/hal-02497335v1>**

Submitted on 3 Mar 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# TRALOGY

## Traduction pragmatique, linguistique de corpus, traducteur : un ménage à trois explosif ?

**Natalie Kübler**

CIILAC-ARP, Université Paris Diderot, PRES Paris Sorbonne cité  
nkubler@eila.univ-paris-diderot.fr

TRALOGY II - Session 3  
Date d'intervention : 17/01/2013

**Depuis une quinzaine d'année, la linguistique de corpus s'intéresse à la traduction, d'une part dans l'observation des caractéristiques linguistiques de la traduction et des stratégies mises en oeuvre par les traducteurs, d'autre part, pour démontrer comment les corpus peuvent venir utilement compléter d'autres outils utilisés par le traducteur. Cet article cherche à démontrer les limites de l'emploi de corpus, mais aussi, comment, malgré celles-ci, ils peuvent répondre à des difficultés liées aux références culturelles, mariant ainsi la traduction aux corpus.**

lien video : [http://webcast.in2p3.fr/videos-traduction\\_pragmatique](http://webcast.in2p3.fr/videos-traduction_pragmatique)



# 1. Introduction

Cet article s'inspire à la fois de la problématique du traducteur pragmatique et de la question de la formation de ceux-ci à l'utilisation des corpus. En effet, plusieurs points ont nourri notre réflexion ; ils portent notamment sur l'usage des corpus numériques par le traducteur, sur le type de corpus à utiliser, le problème des données spécialisées et l'utilisation de corpus de référence, de corpus parallèles, quasi-comparables ou comparables.

## 1.1 Question de terminologie

Une précision terminologique s'impose quant à la définition des termes *corpus parallèle*, *corpus de traduction*, *corpus comparable*, *corpus bilingue*, *multilingue* etc... Nous suivrons la définition de Mc Ennery & Xiao (2007), mais aussi Maia (2002) en affirmant qu'un corpus parallèle est un corpus qui contient des textes source et leur traduction ; il peut être bilingue s'il a une langue source et une langue cible (voir le corpus *Hansard*), ou multilingue si, outre la langue source il a plusieurs langues cibles (voir le corpus MeLLANGE Castagnoli et al. 2011, ou le corpus du site *linguee*) ; un corpus parallèle peut être mono-directionnel, à savoir qu'il ne contient que des textes d'une langue A vers une langue B, mais il peut être bi-directionnel. Le corpus *COMPARA* est un corpus bi-directionnel car il contient des textes source en portugais traduits en anglais et des textes source en anglais traduits en portugais (Frankenberg-Garcia & Santos 2003). Il existe des corpus parallèle dont on n'est pas toujours sûrs ni de la source, ni de la cible, par exemple, le corpus des minutes du Parlement Européen *Europarl* ou le corpus du site *linguee*. Un corpus comparable se définit comme incluant des textes originaux dans différentes langues et répondant aux mêmes critères de genre, de temporalité, de registre, etc. Ici, aucun texte n'est la traduction de l'autre et on parle aussi de textes « natifs ». Un corpus comparable peut aussi être multilingue. Nous incluons aussi dans cette définition les corpus comprenant des textes rédigés dans une langue originale et des textes qui ont été traduits dans cette langue. Enfin, nous distinguons les corpus dits de référence, c'est-à-dire conçus pour être représentatifs d'une langue en général, tels que le *British National Corpus* ou le *Corpus of Contemporary American English* et les corpus spécialisés qui sont représentatifs d'un genre ou d'un domaine spécialisé. Tous ces corpus se définissent comme un ensemble de textes sur support électronique, qui ont été assemblés selon des critères spécifiques en fonction d'un objectif précis. Nous ne parlons donc pas ici de corpus d'exemples, ni de textes assemblés au hasard des trouvailles, sans cadre défini. Cette définition d'un corpus représente la définition canonique adoptée par la linguistique de corpus.

## 1.2 La linguistique de corpus

Ces dernières décennies, la linguistique de corpus a révolutionné les sciences du langage en s'attachant à l'étude des faits observables sur de grands ensembles de textes. Elle trouve des applications dans de nombreuses sous-disciplines des sciences du langage, comme par exemple l'enseignement des langues ou l'étude de la langue des apprenants d'une langue seconde.

La linguistique de corpus commence à être considérée comme une nouvelle discipline à la fin des années 80 et au début des années 90, notamment grâce à l'ouvrage définitoire publié par John Sinclair en 1991. L'on peut penser que, dans les années cinquante et soixante, la critique de Chomsky (1957 :159), qui rejette l'emploi du corpus en affirmant la chose suivante, se concevait sans trop de difficulté :

Any natural corpus will be skewed. Some sentences won't occur because they are obvious [...], false, [...] impolite. [...] the description [...] would be no more than a mere list.

Cette critique ne se justifie plus aujourd'hui pour plusieurs raisons. Chomsky avait besoin de se démarquer des structuralistes et des behavioristes et l'on peut donc envisager que cette critique relevait plus d'une posture visant à marquer sa différence par rapport à une linguistique qu'il considérait n'être qu'une collecte de papillons. Par ailleurs, la taille des corpus papier tout d'abord, et ensuite des premiers corpus électroniques ne leur permettaient pas d'être vraiment représentatifs d'une langue. On peut donc comprendre qu'à l'époque, le repli sur l'intuition du locuteur-auditeur natif ait paru être une solution raisonnable. Aujourd'hui cependant, et déjà en 2001, lorsque Chomsky jette dans la mare de la linguistique le pavé *corpus linguistics doesn't exist*, (repris par de nombreux linguistes dans le monde, mais aussi en France) les corpus rassemblent une quantité suffisamment importante de données pour que l'on puisse affirmer qu'ils sont représentatifs d'une langue ou d'une variété. Le BNC<sup>1</sup>, publié en 1992 et qui comprenait cent millions de mots, nous paraît pourtant relativement modeste aujourd'hui, face aux 450 millions du COCA<sup>2</sup> ou aux 7.3 milliards du corpus COSMAS<sup>3</sup>. Comme la linguistique de corpus s'attache à l'étude des faits observables, on peut la qualifier plutôt de linguistique du discours que de linguistique de la langue.

### 1.3 Des approches de la traduction diversifiées

Depuis bientôt une quinzaine d'année, les linguistes de corpus s'intéressent à la traduction de diverses manières. Au moment où se posent les questions de l'utilité des corpus en traduction, c'est toute une école qui se développe dès les années 1990, notamment avec les conférences CULT (Corpus Use and Learning to Translate) et TaLC (Teaching and Language Corpora). Dans son article fondateur, Aston (1999 : 289) démontre le rôle que peuvent jouer les corpus dans l'amélioration de la qualité et de la rapidité du processus de traduction. Il aborde immédiatement la question de la formation des traducteurs en montrant comment les corpus peuvent permettre aux futurs traducteurs de développer leurs compétences interprétatives et leurs stratégies de traduction, tout en améliorant leur sensibilité aux problèmes posés par la traduction. Toute la recherche dans ce domaine s'est ensuite développée dans cette direction. Nombreux sont les chercheurs qui ont démontré l'utilité des corpus en traduction, notamment pour résoudre des questions de terminologie de phraséologie, mais aussi des questions de style en fonction des genres et des domaines de spécialité ou encore pour évaluer la qualité des traductions (voir entre autres, Zanettin 1998, Varantola 2000, Bowker 2001, Maia 2002, Bowker & Pearson 2002, Zanettin et al. 2003, Beeby et al. 2009, Kübler, & Aston 2010 ; Kübler 2011).

Du côté des traductologues, on a cherché à décrire la langue des textes traduits comme étant un troisième code que l'on pouvait différencier de la langue originale à l'aide de la méthodologie de la linguistique de corpus (Baker 1999, Olohan 2004, Mauranen 2007). Cependant, ces hypothèses ne sont pas toujours validées et ne tiennent pas toujours compte de critères d'ordres très différents, tels que les genres de textes, les domaines de spécialité, mais aussi la formation des traducteurs et les outils d'aide à la traduction qu'ils utilisent.

Aujourd'hui, les linguistes issus de l'analyse contrastive s'intéressent de plus en plus à l'utilisation des corpus pour valider leurs observations sur les choix de traduction. Cependant, ces observations se font souvent sur des corpus parallèles littéraires, ce qui n'est pas complètement adapté à la problématique de la traduction pragmatique.

On peut donc dire que non seulement la linguistique s'interroge et tente d'apporter des réponses à ce qu'est la traduction, mais aussi, et surtout, que la linguistique de corpus représente un tournant épistémologique et méthodologique dans la description et la compréhension de ce qu'est la traduction.

(1) British National Corpus : représentatif de l'anglais britannique dans les années 90

(2) Corpus of Contemporary American English : représentatif de l'anglais américain aujourd'hui

(3) Corpus Search, Management and Analysis System : représentatif de l'allemand aujourd'hui.

A l'aide de cette approche, on peut donc étudier la traduction comme un troisième code et chercher à valider ou invalider l'hypothèse des universaux de traduction, indépendamment de la langue, tels que postulés par Baker. On peut aussi montrer à l'apprenti traducteur pragmatique comment utiliser les corpus pour s'approprier un domaine de spécialité, résoudre les problèmes terminologiques, retrouver la phraséologie spécifique à un domaine et à un genre, et comprendre les différences linguistiques entre langue source et langue cible.

Si cependant l'approche par corpus commence à être reconnue dans le monde de la traduction, elle reste limitée dans son utilisation par les acteurs. Il s'agit peut-être ici du reflet d'une certaine méfiance envers un type de linguistique plus formelle qui s'intéresse à la langue plutôt qu'au discours, domaine de la traduction. Cette approche est par ailleurs concurrencée par la facilité à trouver des informations qu'offre Internet, avec ses bases de données en ligne et simplement, la recherche sur un moteur bien connu. Une partie du monde de la traductologie peut penser que la traduction n'est pas une affaire de linguistique. Nous pensons que ces réticences prennent naissance dans les limites externes et internes que rencontre l'approche par corpus.

## 2. Corpus et traduction pragmatique

### 2.1 Traduction pragmatique et traduction littéraire

Nous interrogeons ici la relation entre corpus et traduction du point de vue de la traduction pragmatique, à savoir toute traduction non littéraire ou toute traduction se concentrant principalement sur la réception du texte par le destinataire. Il s'agit donc d'une traduction qui vise la clarté et la compréhensibilité en langue cible avant tout. On évoque généralement l'opposition entre traduction littéraire et traduction spécialisée, plutôt que pragmatique. Or, la traduction spécialisée se définit souvent négativement comme étant la traduction de tout sauf de ce qui est littéraire et pourtant, le nombre de traduction dans des domaines autres que littéraire est très élevé. La consultation du rapport 2009 de la SFT sur les tarifs des traducteurs (SFT 2010 : 12) nous informe à la fois sur les différents domaines, y compris la traduction littéraire, dans lesquels interviennent les traducteurs, ainsi que sur le pourcentage consacré à chaque domaine :

<b>Arts et littérature</b>	5 %
<b>Communication</b>	8 %
<b>Economie et gestion</b>	9 %
<b>Industries et techniques</b>	27 %
<b>Juridique et politique</b>	13 %
<b>Sciences humaines et sociales</b>	3 %
<b>Sciences pures et appliquées</b>	9 %
<b>Tourisme et loisirs</b>	4 %
<b>Autres</b>	5 %

On constate que la part de la traduction littéraire est donc très restreinte par rapport aux autres domaines. Or, la traduction de textes dans les domaines autres que littéraire relève plutôt de la traduction pragmatique que de la traduction littéraire. Parler de traduction spécialisée ne définit la traduction que par le domaine et laisse de côté la question cruciale du genre textuel qui a pour conséquence un style spécifique et donc des caractéristiques linguistiques spécifiques et différentes d'une langue à l'autre. Traduire dans un domaine spécialisé n'est certes pas qu'une question de terminologie. La traduction littéraire et peut-être, dans une certaine mesure, la traduction des sciences humaines et sociales, relève plus d'une approche sourcière que cibliste,

contrairement à la traduction pragmatique. Nous nous intéressons ici à l'apport des corpus dans le domaine de la traduction pragmatique.

## 2.2 Formation des traducteurs

La traduction pragmatique s'intéresse aux questions de traduction dans des domaines spécialisés<sup>4</sup>. Bien que l'on puisse considérer que former un traducteur en traduction littéraire lui permettra de traduire tout texte spécialisé, nous pensons que la formation des traducteurs spécialisés peut être abordée de deux manières. Soit l'on forme le traducteur aux domaines spécialisés en lui donnant par exemple, des cours de droit et d'économie, soit on le forme aux outils et méthodes qui l'aideront à s'approprier n'importe quel domaine de spécialité. Dans ce cas, ce sont les méthodes et techniques de la linguistique de corpus qui pourront l'aider à atteindre cet objectif.

Le traducteur doit donc être formé pour acquérir les connaissances et compétences suivantes :

- connaître et savoir interroger les gros corpus de référence, tels le BNC, le COCA, Frantext, etc.
- connaître et savoir utiliser les outils en ligne permettant d'interroger le Web comme corpus ou de constituer des corpus à partir du Web (WebCorp, The Web as Corpus, Sketchengine, BootCat etc.)
- savoir constituer et interroger un corpus parallèle dans le domaine spécialisé et le genre textuel nécessaires
- savoir constituer et interroger un corpus comparable dans le domaine spécialisé et le genre textuel nécessaires

L'acquisition de ces compétences demande cependant de nombreuses heures de pratique qu'il n'est pas toujours possible d'intégrer complètement dans les cursus de traduction. Le manque de pratique rend parfois le travail sur corpus lent et fastidieux, voire même incompréhensible pour les apprentis traducteurs. Cela a pour conséquence, qu'ils ne l'utiliseront pas dans leur vie professionnelle. C'est d'ailleurs l'objection émise le plus souvent par les étudiants : utiliser des corpus est dévoreur de temps, de même que la compilation du corpus. On préfère alors utiliser Internet, qui ne donne pas de contextes, ni de collocations, ou le site qui rencontre le plus de succès en ce moment, *linguee*. Nous reviendrons sur ce site qui offre aux utilisateurs un accès gratuit à un corpus parallèle. Par ailleurs, nous verrons en section 3, que le travail sur corpus rencontre des limites soit externes au corpus, soit internes à celui-ci. Dans la section suivante, sont rapidement abordés les différents types d'informations que le traducteur peut s'attendre à extraire des corpus.

## 3. Informations extraites des corpus

Les corpus spécialisés dans la langue source permettent tout d'abord au traducteur de s'approprier le domaine en navigant dans le corpus à partir de la lecture du texte à traduire. La recherche de contextes définitoires à l'aide de marqueurs linguistiques peut l'aider à mieux comprendre le domaine. En anglais, le marqueur *be defined as* par exemple (Pearson 1998) amène à trouver des contextes explicatifs. L'exemple suivant, extrait d'un corpus comparable en sciences de la terre, montre ce type de contextes pour le terme *cryptodome* est une variante du terme *debris flow discharge* :

(4) Nous n'aborderons pas ici la question des techniques de traduction qui sont enseignées dans les formations et qui peuvent être très utilement complétées par les résultats de la recherche en traductologie induite par le corpus et que nous avons brièvement mentionnée dans l'introduction.

(1) A cryptodome can **be defined as** a body of viscous magma which has risen to a high level in the crust. It actively deforms adjacent strata, but does not break through to the surface.

(2) The discharge of a debris flow can **be defined as** an amount of liquefied geomaterials passing through a fixed cross-section perpendicular to a debris flow path. Similar to water discharge in a river, debris flow discharge varies with both time and space. Consequently, a peak discharge can be defined and identified for each cross-section. The maximum discharge of a debris flow can be further defined as the largest peak discharge of all identified peak discharges along a flow path

On connaît aujourd'hui de nombreux marqueurs linguistiques permettant de relever des contextes riches en information. L'étude de ceux-ci reste cependant à compléter dans de nombreuses langues.

Ce type de contextes amène aussi à valider des candidats termes en recherchant leurs définitions. En effet, bien que les bases de données terminologiques comme *Termium* ou *Le grand dictionnaire terminologiques* soient riches en termes, elles restent encore incomplètes, notamment dans les domaines qui évoluent très rapidement.

Une fois les termes déterminés dans la langue source, la recherche d'équivalents en langue cible peut aussi parfois poser problème si ceux-ci ne se trouvent pas dans les bases de données ou les dictionnaires spécialisés. Ici encore, les corpus viennent pallier les manques des autres outils. Bien que l'idéal consiste à interroger un corpus parallèle qui donnera des équivalents en langue cible, cette situation est assez rare. On a donc recours à des corpus comparables dans lesquels le traducteur doit mettre en oeuvre de véritables talents de détective linguistique pour trouver des équivalents. Cette approche a déjà été abondamment décrite par les chercheurs. On peut donc se référer à Maia (1997, 2000), Bowker & Pearson (2002), Kübler (2003) entre autres. L'une des limites externes que l'on peut rencontrer dans ce cas est qu'il n'existe pas encore d'équivalent dans la langue cible si le terme est récent et a été créé dans la langue source. On ne trouvera donc aucune information dans le corpus à ce sujet. C'est alors au traducteur, en collaboration avec un expert du domaine, de proposer un équivalent dans la langue cible, en s'appuyant à la fois sur ses connaissances linguistiques de formation des termes, sur sa créativité et sur les connaissances de l'expert.

La recherche de collocations dans les corpus parallèles ou comparables renseigne aussi le traducteur sur l'emploi des termes en contexte. Dans ce cas, la comparaison avec l'emploi des collocations dans des corpus généraux peut parfois s'avérer très utile pour déterminer la spécialisation de la collocation. En effet, certaines collocations spécialisées peuvent être très différentes de celles de la langue générale, ou encore ne pas exister dans d'autres domaines pour le même terme. C'est par exemple le cas pour le terme *dépression* en médecine et en météorologie. Pour évoquer une dépression importante, on parlera de *grave dépression* dans le premier cas et de *forte dépression* dans l'autre. Enfin, les corpus généraux donnent des informations précieuses sur certaines structures collocationnelles qui se rencontrent à la fois en langue générale et en langue de spécialité. En effet, tout n'est pas spécifique à la langue de spécialité dans un texte spécialisé.

Un texte spécialisé s'inscrit toujours dans un genre textuel spécifique, caractérisé par des marques linguistiques et reflète ainsi un style particulier. La navigation dans un corpus parallèle ou comparable guide le traducteur dans les différences linguistiques qui existent d'une langue à l'autre pour le style. Ainsi, Bowker & Pearson (2002 : 205-206) montrent comment, dans un corpus de vulgarisation sur les virus informatiques en anglais, on peut détecter afin

de le reproduire dans la traduction, le style informel des articles en repérant ses marques linguistiques.

Ainsi, ce panorama très succinct du type d'informations que l'on peut extraire d'un corpus montre son intérêt dans des domaines dans lesquels les dictionnaires ou la recherche sur Google ne sont pas toujours suffisants. On peut objecter que certains éléments ne se trouvent pas dans les corpus, notamment ce qui relève des références culturelles

## 4. Où situer les limites ?

L'emploi des corpus, à la fois par le traducteur et par le chercheur, permet d'apporter des réponses à de nombreuses questions de recherche, mais aussi donne des réponses concrètes à des questions pratiques. Cependant, la linguistique de corpus et son application à la traduction spécialisée connaissent un certain nombre de limites, dues d'une part à des facteurs externes, et d'autre part, à des facteurs internes.

### 4.1 Limites externes

Comme nous l'avons mentionné plus haut, il existe un des corpus dit généraux, élaborés pour être représentatifs d'une langue de manière générale, et cela dans différentes langues. Ces corpus présentent cependant un certain nombre de limitations. Tout d'abord, ils n'existent pas dans toutes les langues et s'ils existent, ne sont pas constitués de la même manière selon les différentes langues. Des corpus comme le *COCA* pour l'anglais américain et le *Cosmas* pour l'allemand ne sont pas accessibles avec les mêmes outils. Les différences entre les interfaces ne permettent pas forcément d'interroger les corpus de la même manière et exigent donc de l'utilisateur d'apprendre à se servir d'autant d'interfaces que de langues. Par ailleurs, bien qu'il existe en français, plusieurs corpus dits généraux, ceux-ci ne sont pas équilibrés. Le corpus *FRANTEXT* par exemple n'est pas un corpus équilibré, ni entre les différents genres, ni entre l'écrit et l'oral. Il s'agit surtout d'un corpus littéraire, bien qu'il ait été enrichi d'un certain nombre d'écrits scientifiques (des essais surtout) et qu'il donne accès à une base de 1083 normes AFNOR. Le corpus français élaboré par l'Université de Leipzig n'est pas non plus un corpus équilibré puisqu'il contient surtout des sites de presse, *wikipedia* et divers sites Web. Par ailleurs, son interface d'interrogation a été récemment modifiée. Si auparavant, elle ne donnait pas accès à des concordances, elle permettait d'obtenir les collocats les plus fréquents d'un mot, illustré par un certain nombre d'exemples. Elle ne donne plus aujourd'hui qu'une représentation graphique du réseau des collocats, sans indiquer si ceux-ci se trouvent à gauche ou à droite.

Pour résoudre les problèmes d'équivalences terminologiques, l'utilisation d'un corpus spécialisé parallèle est idéal. Cependant, les corpus parallèles disponibles sont rares et encore plus rares sont les corpus parallèles spécialisés. Et bien souvent, même s'il existe un corpus parallèle spécialisé, le domaine traité ne sera pas tout à fait celui du texte que l'on a à traduire. Il est cependant possible au traducteur de créer lui-même un corpus parallèle spécialisé. Cependant, deux obstacles bien connus s'opposent à cela : d'une part, la constitution d'un corpus parallèle, aligné est longue et demande de la part de l'utilisateur des compétences techniques que souvent le traducteur n'a pas ; d'autre part, les ressources disponibles permettant de créer des corpus parallèles sont rares ou protégées par des droits d'auteurs.

Pour pallier le manque de corpus parallèle, on peut créer ce que Maia (1997) nomme *do-it-yourself corpora*. Ce sont des corpus comparables que l'on compile en fonction du domaine et du genre du texte à traduire, à partir de ressources trouvées sur Internet, ou de ressources



fournies par le demandeur d'ordre par exemple. Se posent alors souvent des questions de droits d'auteurs que Varantola (2000) se propose de résoudre en les appelant *disposable corpora*, des corpus jetables. Il est possible de compiler un petit corpus comparable à partir de documents disponibles, mais protégés et de jeter ensuite ce corpus, lorsque la traduction est faite. Elle affirme que cela permet d'éviter d'éventuelles poursuites liées aux droits d'auteur. Cependant, même si ces questions sont plus ou moins résolues, reste la disponibilité des ressources selon la langue. Il est bien connu aujourd'hui par exemple, qu'il est plus facile de trouver des ressources spécialisées en anglais qu'en français (Kübler 2011) ou dans d'autres langues. Et souvent, même si l'on trouve des ressources textuelles appartenant au même domaine de spécialité en anglais et en français, elles n'appartiennent pas forcément au même genre textuel. Dans le domaine des sciences dites dures par exemple, collationner des articles de recherche en anglais est une tâche aisément réalisable. En français, en revanche, et selon les disciplines, cette recherche peut se révéler plus ou moins réaliste. Les chercheurs en sciences du vivant ne publient quasiment plus qu'en anglais. Le *Bulletin de la Société Géologique de France* publie de plus en plus en anglais, alors qu'il représentait l'une des dernières revues à publier en français en sciences de la terre. La question se pose un peu différemment en sciences humaines, puisque l'on continue à éditer des revues dans d'autres langues que l'anglais. Malgré cela, les revues de sciences humaines considérées comme étant de portée internationale par la European Science Foundation sont très majoritairement en anglais.

Dans le domaine des techniques, du juridique et de l'économique, la question de la disponibilité des documents, quelle que soit la langue, se pose de manière aiguë. En effet, de nombreux documents sont protégés ou tout simplement non disponibles en ligne, ou encore payants.

Une autre limite rencontrée dans le travail sur corpus est celle du soin apporté à sa constitution et de sa taille. On sait que la taille du corpus dépend de l'objectif pour lequel on le compile. Pour un traducteur, les objectifs doivent être de multiples ordres : compréhension du domaine, terminologie, phraséologie spécialisée, style lié à un genre. Or, la compilation d'un corpus, même jetable, même spécialisé, requiert de l'utilisateur des compétences sur les outils permettant de télécharger automatiquement un grand nombre de textes potentiellement pertinents pour le genre et le domaine. Il s'agit d'outils comme *bootCat* ou *Sketchengine*. Ils permettent de gagner du temps, mais présentent l'imperfection inhérente à tout outil informatique de devoir trier le bruit. Par ailleurs, *Sketchengine* oblige à utiliser les outils d'interrogation en ligne intégrés au site. Or, un utilisateur peut avoir besoin de travailler hors ligne.

Par conséquent, les limitations rencontrées dans les corpus lorsque l'on travaille sur la terminologie bilingue ou la phraséologie spécialisées ne sont pas inhérentes aux corpus en tant que tels, mais dépendent de facteurs externes.

Nous n'envisageons pas par ailleurs, de n'utiliser que des corpus spécialisés, il est indispensable de faire appel aux grands corpus généraux pour des questions d'ordre plus général. C'est ainsi qu'en combinant différents types de corpus, mais aussi en possédant des compétences spécifiques lui permettant de compiler des corpus comparables et spécialisés rapidement et d'interroger tous les corpus à sa disposition de manière efficace, un traducteur devrait avoir tous les outils qui lui permettraient de résoudre ses problèmes de traduction. Tous ? Peut-être pas.

## 4.2 Un exemple de limite interne ?

La traduction de certains documents qui font référence à des connaissances d'ordre culturelles et qui sont sans rapport direct avec le domaine de spécialité peut poser problème. On peut faire l'hypothèse que les références culturelles ne se trouveront pas dans des corpus spécialisés. Il n'est par ailleurs pas certain que celles-ci fassent partie des corpus généraux.

Nous prenons ici l'exemple concret d'un article traduit par l'une de nos étudiantes de Master de traduction spécialisée. Il s'agit d'un article en anglais, portant sur les relations entre le droit et la détection de mensonges par imagerie cérébrale<sup>5</sup>. L'auteur du texte s'interroge sur l'utilisation des dispositifs de neuro-imagerie permettant de détecter si un témoin est sincère dans un tribunal ou repérer des terroristes en puissance dans les aéroports. Dans l'un des paragraphes, il est fait référence à l'ouvrage d'Aldous Huxley, intitulé *Brave New World* :

Because these two preexisting conditions have received relatively little attention from legal scholars anxious to be the first to map a **brave new world** of accurately neuroimaged cognition, they will be the focus of this article.

Ne connaissant pas cet ouvrage, l'étudiante ne pouvait pas traduire cette phrase en utilisant la traduction française du titre de Huxley ; le texte a donc été tout d'abord sous-traduit :

Ces deux conditions préalables seront les points centraux de cet article. En effet, elles ont été quelque peu délaissées par les professeurs de droit, tous désireux d'être les pionniers de ce **nouveau monde** de la neuro-imagerie cognitive.

Grâce une aide externe, l'étudiante a finalement correctement traduit le texte (l'ensemble de sa traduction étant par ailleurs excellente). On peut comprendre aujourd'hui la méconnaissance de l'ouvrage de Huxley qui appartient au genre de la science-fiction, surtout celle qui n'est pas très récente. Tout le monde ne s'intéresse pas forcément à ce genre. Par ailleurs, même si l'on a lu *Le meilleur des Mondes*, on ne connaît pas forcément son titre original en anglais : *Brave New World*. Dans les deux cas, ces titres font référence à des oeuvres littéraires appartenant à la culture anglaise pour l'un, puisqu'il a été emprunté à la *Tempête* de William Shakespeare, et à la culture française pour l'autre qui à l'origine vient du *Candide* de Voltaire.

Se pose alors la question de l'emploi du corpus : est-il possible de trouver la bonne traduction de *brave new world* dans un corpus ? Et dans lequel ? Il faut tout d'abord être capable de repérer que cette expression *brave new world* relève d'un certain figement en anglais. Une recherche simple dans le COCA donne 395 occurrences de *brave new world*. Cette expression est effectivement utilisée dans des contextes similaires à celui de notre article, comme par exemple dans :

*Nurses need to get ready to meet the challenge of the brave new world of the NSH by making sure they are part of the decision-making*

On comprend qu'il s'agit d'un monde en évolution et qui se doit, si l'allusion à Huxley reste inconnue, d'être courageux. Rester dans cette interprétation peut conduire à un contresens dans le processus de traduction. Heureusement, la cinquième, la septième et la quinzième occurrences relevées par le COCA permettent de clarifier l'usage de l'expression :

Images would be one such development, and what Aldous Huxley termed, in the **Brave New World**, « the feelies »

Them. In a way, when I think of « 1984 » and « **Brave New World** », two brilliant dystopic books

Technology becomes an instrument of super oppression. Aldous Huxley's **Brave New World** sets the literary standards here. And Isaac Asimov's I Robot

On a ici très vite l'information qu'il s'agit du titre d'un livre, rédigé par Aldous Huxley, et que celui-ci est utilisé pour faire référence au monde décrit dans le livre. Encore faudrait-il savoir de

(5) Moreno Joelle Anne. 2009. The Future of Neuroimaged lie-detection and the Law. *Akron Law Review*, Vol. 42. 717-734.

quel type de monde il s'agit. Deux solutions s'offrent alors : soit lire le livre, soit analyser l'emploi de *brave new world* en corpus.

Le BNC, corpus général de l'anglais britannique, donne 61 occurrences pour *brave new world*. L'examen de son emploi en corpus donne lieu aux observations suivantes. Il s'agit aujourd'hui d'un nom composé lexicalisé qui porte une connotation négative, ou du moins ironique :

ALN **289** Whether and how the **brave new world** of social care promulgated by government can be operationalized must in part be answered by reference to the experiences gained in field innovations of the kind we studied.

A1V **315** Hong Kong Special Report : A **brave new world** tumbles down: Hong Kong tried to be optimistic about Chinese rule.

A7T **280** After all, what would be more likely to gain from a **brave new** world of free capital flows than the greatest financial centre in Europe?

Par ailleurs, l'expression fait parfois directement référence au passage de la *Tempête* de Shakespeare:

ANY **2322** 'O **brave new world**', said Robyn, 'where only the managing directors have jobs.'

[1] P353 - \*COLLECTIF , *Arts et littérature dans la société contemporaine*, dir. Pierre Abraham : t. 2, 1936, p. 4012

\*Studies , \*I 927 / ; romancier, il construit ses oeuvres selon une philosophie amère qui émane logiquement du récit / \*Point \*Counter \*Point, \*I 928 ; \***Brave** \***New** \***World** , \*I 932 /. Son style, d'un rythme conscient, et sa langue élégamment châtiée, l'apparentent lui aussi à la tradition d'\*Anatole \*France. Cette

Haut dtu formulaire

Bas du formulaire

Toutes ces observations peuvent se faire très rapidement et mènent à la conclusion que ce composé doit avoir une traduction spécifique et figée elle aussi en français. Dans les langues de spécialité, il arrive qu'un terme français soit accompagné, entre parenthèses, du terme anglais d'origine. Ce phénomène est fréquent lorsque les membres d'une même communauté de discours utilisent à la fois le terme anglais et le terme français, ou, lorsque l'équivalent français n'est pas encore bien installé. Cette situation permet de retrouver les équivalents français, en recherchant le terme anglais dans un corpus français (Kübler 2004). La recherche de *brave new world* dans le corpus FRANTEXT, donne un seul résultat :

Malheureusement, la traduction du titre n'est pas donnée dans cet extrait. Il devient donc difficile de trouver le titre dans ce corpus français. Cependant, l'outil WebCorp qui permet d'effectuer des recherches sur le Web avec les outils de la linguistique de corpus (Renouf 2002, 2009), donne :

de pages 155 ISBN 978-2266131056 Retour au meilleur des mondes (Brave New World Revisited) est un essai

Cette indication peut aider à comprendre quel est l'équivalent de *brave new world*. Il reste aussi la possibilité d'utiliser le site *linguee* qui représente un corpus parallèle dont on ne sait cependant pas toujours quel est le texte source et quel est le texte cible. En outre, les traductions

ne sont pas toujours fiables. Bien que certaines traductions soient marquées comme étant peut-être erronées, il n'est pas toujours facile de savoir ce qu'il en est. Cependant, une recherche en anglais permet très vite de trouver une référence au livre et d'obtenir la confirmation de la traduction du titre en français :

<p>Whether we like it or not, we live in Huxley's <b>Brave New World</b>.</p> <p>europarl.europa.eu</p>	<p>Que nous le voulions ou pas, nous vivons dans <b>le Meilleur des mondes de Huxley</b>.</p>
---	---

On peut donc faire l'hypothèse que *Le Meilleur des Mondes* est une bonne traduction. Cependant, l'examen des données fournies par le site *linguee* soulève un certain nombre de questions. En effet, l'expression n'est pas toujours traduite par *le meilleur des mondes*. On trouve aussi *splendide nouveau monde, quelque chose de nouveau, le monde merveilleux, le nouveau monde fantastique, d'autres buts indirects, le nouveau monde, un homme nouveau* par exemple. Dans certains cas, on peut se demander si le traducteur a reconnu la référence à Huxley, dans d'autres, il semble que le choix d'une autre traduction soit délibéré, afin de rendre explicite la connotation négative, un peu ironique, de la citation en anglais. Ce choix s'explique peut-être par l'hypothèse faite par le traducteur que le lecteur final ne comprendra pas la teneur de la référence en anglais.

Cette variation dans la traduction, et même, le choix d'omettre complètement la référence voulue par l'auteur anglophone nous pousse à vérifier l'emploi de *meilleur des mondes* en français. En effet, ces différences de traduction laissent aussi à penser que l'emploi français n'a peut-être pas exactement la même connotation qu'en anglais. L'examen des occurrences de *meilleur des mondes* dans FRANTEXT montre qu'il est fait, la plupart du temps, référence à la citation du *Candide* de Voltaire : *tout va pour le mieux dans le meilleur des mondes possibles* et non à l'ouvrage d'Huxley. Il en va de même pour l'emploi donné par le concordancier WebCorp : sur le Web, la référence à Huxley en français se rencontre très souvent dans l'évocation même de son ouvrage ; sinon, c'est la plupart du temps la référence à Voltaire qui se retrouve dans l'expression *meilleur des mondes*. Par ailleurs, une recherche de cette expression en français dans *linguee* permet de souligner qu'en anglais, son équivalent n'est souvent pas la plupart du temps, mais une traduction littérale. Cela indique que l'emploi de le plus fréquent de l'expression en français est celui de Voltaire qui n'a pas la même signification que *Brave New World* de Huxley, ce qui explique pourquoi certains traducteurs font le choix de traduire l'expression *brave new world* autrement que par le titre français.

Partie de l'hypothèse qu'il était plus difficile de travailler sur les références culturelles en corpus, nous avons eu la surprise de découvrir que celui-ci permettait malgré cela de trouver les équivalents, mais aussi de mieux comprendre la teneur de la référence, et ses possibles connotations par l'observation linguistique de ses occurrences en corpus. Cependant, afin de compléter les connaissances sur les références que nous avons citées, et d'éviter de mal traduire parfois *brave new world* par *meilleurs des mondes*, il nous semble qu'il peut être utile de rechercher dans d'autres sources des informations permettant de mieux comprendre ces expressions, par exemple en consultant des encyclopédie sur les ouvrages de Huxley et Voltaire. Nous pensons aussi que seule une grande pratique de la langue source, que ce soit à l'aide de corpus ou par l'expérience, peut amener un traducteur à s'interroger sur la portée de l'expression *brave new world*. Cela relève, à notre avis, de connaissances qui vont au-delà des corpus et qui doivent faire partie de la culture générale d'un professionnel de la langue. Les traductions sur le site *linguee* sont d'ailleurs révélatrices de cet élément. C'est pourquoi nous pensons que ce site ne doit pas être le seul corpus utilisé par les traducteurs. Nous allons tenter de l'expliquer dans la sous-section suivante.

<p>Bénéficiant d'état des installations des arts, élégant, fonctionnel et flexible, Kuala Lumpur Convention Centre est <b>parmi le meilleur des mondes</b> quand il s'agit d'accueillir des expositions, des conférences, réunions, séminaires ou événements spéciaux.</p> <p>icmhesr5.my</p>	<p>Boasting <b>state of the arts</b> facilities, stylish, functional and flexible, Kuala Lumpur Convention Centre <b>is amongst the worlds best when it comes</b> to hosting exhibitions, conferences, meetings, seminars or special events.</p> <p>ic</p>
---	--

### 4.3 Pourquoi utiliser d'autres corpus que *linguee* ?

On pourrait penser que l'existence du site *linguee* permet de remplacer tout autre corpus. Cependant, les textes contenus dans ce site n'offrent pas la possibilité de trouver des équivalents terminologiques dans de très nombreux domaines de spécialité. En outre, on ne peut ni trouver de contexte définitoire, ni étudier un genre textuel à l'aide de ce site. Il n'est donc utile que pour une partie des utilisations que le traducteur peut faire d'un corpus. Outre le problème de la langue source qui n'est pas donnée, il présente tous les inconvénients inhérents aux corpus parallèles puisque toutes les traductions ne sont pas forcément fiables. Elles peuvent être issues de traduction automatique :

On rencontre des traductions erronées :

<p>Jose Luis Isidro ont eu la possibilité de s'entraîner <b>avec le meilleur des monde de l'Aikido</b>, et a été en mesure de vérifier la validité de l'efficacité de l'Aikido dans les situations de danger réel.</p> <p>artesguerreras.es</p>	<p>Jose Luis Isidro <b>had the opportunity</b> to train <b>with the best of the Aikido world, and has been able</b> to verify the validity of the effectiveness of Aikido in situations of real danger.</p>
---	---

Parfois, une traduction peut être signalée comme erronée, alors qu'elle est tout à fait correcte :

<p>Le Mudam présente jusqu'au 23 mai l'exposition <b>Le Meilleur des Mondes du point de vue</b> de la collection Mudam, qui invite le public à percevoir le monde à travers [...]</p> <p>centre-pompidou.fr</p>	<p><b>The Mudam</b> in Luxembourg is presenting <b>Brave New World from the perspective</b> of the Mudam's Collection until 23 May. The exhibition draws on the museum's collection</p>
---	---

Enfin, la méconnaissance des références culturelles peut mener un rédacteur à commettre des erreurs que la traducteur devrait corriger :

<p>À entendre cela, on pourrait croire, comme l'a si bien dit le naturaliste britannique, Aldous Huxley, <b>que tout est pour le mieux dans le meilleur des mondes</b>.</p> <p>www2.parl.gc.ca</p>	<p>That would lead us to think that, as Aldous Huxley the British naturalist <b>put it, it is all for the best in the best of worlds</b>.</p>
--	---

Il est inutile de nier l'utilité de ce corpus, mais il faut cependant l'utiliser avec précaution et parfois, vérifier dans un corpus monolingue si un équivalent donné est correct.

## 5. Conclusion

Nous avons tenté de montrer ici que le ménage à trois entre le traducteur, la traduction pragmatique et le corpus n'est pas si explosif que cela. Bien que l'approche par corpus connaisse des limites à la fois internes et externes, elle peut représenter une source d'information précieuse à la fois dans la formation des traductions et dans la pratique de la traduction. En traduction pragmatique, on ne peut parfois pas faire l'économie d'un expert pour comprendre la terminologie ou trouver des équivalents qui n'existent pas encore. La formation en traduction ne fait pas l'économie d'une grande pratique de celle-ci, ni de l'acquisition des processus bien connus qui sont utilisés. Ici, les résultats de la recherche sur corpus traduits ou sur des corpus d'apprenants de la traduction peuvent venir éclairer le formateur en révélant les caractéristiques à éviter dans les textes traduits ou en faisant la lumière sur les préférences de choix d'une langue à l'autre.

Utiliser les corpus n'écarte pas non plus l'emploi de bases de données ou de dictionnaires, mais vient les compléter là où ceux-ci sont insuffisants. Enfin, l'interrogation de corpus peut donner des pistes sur la manière de traduire une référence culturelle. Mais cette approche reste toujours limitée par les ressources textuelles dans les différentes langues. Enfin, elle ne fait pas l'économie, même pour les textes très spécialisés, d'une culture générale bien comprise, qui exigera, dans certains cas, de faire appel à la créativité du traducteur.

## Bibliographie

- Aston, G. (1999) 'Corpus use and learning to translate', *Textus*, 12, 289-313.
- Baker, M. (1999) 'The role of corpora in investigating the linguistic behaviour of translators', *International Journal of Corpus Linguistics*, 4 : 281-298.
- Baroni, Marco & Bernardini, Silvia. 2004. BootCaT : Bootstrapping corpora and terms from the web. Proceedings of LREC 2004.
- Beeby Allison, Rodriguez Inés Patricia and Sánchez-Gijón Pilar (eds). 2009 *Corpus Use and Translating*. Amsterdam : Benjamins
- Bowker, Lynne. 2001. 'Towards a Methodology for a Corpus-Based Approach to Translation Evaluation'. *Meta*, XLVI, 2, <http://id.erudit.org/iderudit/002135ar>
- Bowker, Lynne and Pearson, Jennifer. 2002. *Working with Specialised Language : a Guide to Using Corpora*, London : Routledge
- Brown, P., Lai J. & Mercer R. 1991 : « Aligning sentences in parallel corpora ». In 29th Annual Meeting of the Association for Computational Linguistics, Berkeley, CA. 169-176
- Castagnoli, Sara, Ciobanu, Dragos, Kübler, Natalie, Kunz, Kerstin, Volanschi, Alexandra "Designing a Learner Translator Corpus for Training Purposes". In N. Kübler. (ed) (2011) *Corpora, Language, Teaching, and Resources : From Theory to Practice*. Bern : Peter Lang. 221-248.

Corpas Pastor, Gloria and Seghiri, Myriam (2009) 'Virtual corpora as documentation resources : translating travel insurance documents (English-Spanish)', in A. Beeby, P. Rodriguez Inés and P. Sánchez-Gijón (eds) (2009) *Corpus Use and Translating*. Amsterdam : Benjamins.

Frankenberg-Garcia, A. & Santos, D. 2003. Introducing COMPARA, the Portuguese-English parallel translation corpus. In F. Zanettin, S. Bernardini & D. Stewart (eds.) *Corpora in Translation Education*, Manchester : St. Jerome Publishing. 71-87

Kübler, Natalie & Aston, Guy. 2010. "Using Corpora in Translation". in M. Mc Carthy & A. O'Keefe (eds) *The Routledge Handbook of Corpus Linguistics*, London : Routledge. 505-515

Kübler, Natalie 2011. 'Working with different corpora in translation teaching'. In Ana Frankenberg-Garcia, Lynne Flowerdew, and Guy Aston (eds) *New Trends in Corpora and Language Learning*. London : Continuum. 62-80

Kübler, Natalie. 2004. 'Using Webcorp for building specialized dictionaries'. In Aijmer K. (ed.) *Proceedings of the ICAME Conference, May 2002, Göteborg, Suède*. Amsterdam : Rodopi.

Kübler, Natalie. 2003. Corpora and LSP translation. In Zanettin, F., S. Bernardini & D. Stewart (eds), *Corpora in Translator Education*. Manchester : St Jerome.

Maia, Belinda (1997). 'Do-it-yourself corpora ... with a little bit of help from your friends !' in Barbara Lewandowska-Tomaszczyk and Patrick James Melia (eds) *PALC'97 Practical Applications in Language Corpora*. Lodz : Lodz University Press. 403-410.

Maia, Belinda (2002) 'Corpora for Terminology Extraction – the Differing Perspectives and Objectives of Researchers, Teachers and Language Services Providers'. In Elia Yuste-Rodrigo (ed) *Proceedings of the Workshop Language Resources for Translation Work and Research*. LREC 2002, Las Palmas de Gran Canaria, Spain.

Mauranen, Anna. 2007. 'Universal tendencies in translation', in M. Rogers and G. Anderman (eds) *Incorporating Corpora : The Linguist and the Translator*. Clevedon : Multilingual Matters.

McEnery, Anthony & Xiao, Richard. 2007. *Parallel and comparable corpora : What are they up to ?* In *Incorporating Corpora : Translation and the Linguist (Translating Europe)*. Multilingual Matters Ltd, Clevedon, UK. <http://eprints.lancs.ac.uk/59/> Consulté le 15 décembre 2012

Olohan, Maeve. 2004. *Introducing Corpora in Translation Studies*, London : Routledge.

Pearson, Jennifer. 1998. *Terms in Context*. Amsterdam : John Benjamins Publishing Company

Renouf, Antoinette. 2009. *Corpus Linguistics beyond Google : the WebCorp Linguist's Search Engine*. *Digital Studies*, vol. 1 (1). Consulté le 5 janvier 2013.

Renouf, Antoinette. 2002. 'WebCorp : providing a renewable data source for corpus linguists.' *Extending the scope of corpus-based research : new applications, new challenges*. Ed. S. Granger and S. Petch-Tyson. Atlanta, GA : Rodopi.

SFT Commission statistiques et étude du marché. 2010 : *SFT enquête tarifs 2009*. <http://www.sft.fr/tarif-de-traduction.html>. Consulté le 18/12/2012.

Varantola, K. (2000) "Translators, Dictionaries and Text Corpora", in Bernardini, S. & Zanettin, F (eds) (2000) pp 117-133.

Zanettin, Federico (1998). 'Bilingual Comparable Corpora and the Training of Translators'. *Meta*, XLIII, 4. <http://id.erudit.org/iderudit/004638ar> Zanettin

Zanettin, F., Bernardini, S. and Stewart, D. (eds) (2003) *Corpora in Translator Education*, Manchester : St. Jerome.

### **Corpus et outils :**

Corpus COMPARA : <http://www.linguateca.pt/COMPARA/>

Corpus HANSARD <http://www.isi.edu/natural-language/download/hansard/>

Corpus Europarl : <http://www.statmt.org/europarl/>

Corpus Cosmas : <https://cosmas2.ids-mannheim.de/cosmas2-web/>

Corpus MeLLANGE : <http://corpus.leeds.ac.uk/mellange/>

Corpus linguee : <http://www.linguee.fr>

WebCorp : <http://www.webcorp.org.uk>

Sketchengine : <http://www.sketchengine.co.uk/>

The Web as Corpus : <http://webascorpus.org/>

BootCat : <http://bootcat.sslmit.unibo.it/>