

Deux méthodes d'annotation de la conceptualisation métaphorique dans un corpus parallèle : *On the Origin of Species* de Charles Darwin

Eve-Marie Gendron-Pontbriand¹, Yan Bodain², Sylvie Vandaele¹

¹Département de linguistique et de traduction, Université de Montréal, Montréal, Canada

²Département de systèmes d'information et méthodes quantitatives de gestion, Université de Sherbrooke, Sherbrooke, Canada

Abstract

Metaphorical conceptualization is ubiquitous in scientific texts and its translation raises several issues. Here, we propose two XML annotation methods, adapted to the study of parallel corpora and derived from a previous method designed for the characterization of comparable corpora. These two methods are tested against a corpus comprised of Charles Darwin's *On the Origin of Species* and its French translation. This seminal work is known for the richness of its metaphors. In the first strategy, the XLink language is used to create a unidirectional link between the files containing the source and target texts. In the second strategy, the source and target texts are aligned, and the resulting alignment file is annotated. Both methods are evaluated for feasibility, user-friendliness, and downstream applications of the annotated files. They generate quantitative data that can be treated statistically with the help of non-parametric tests such as the X^2 .

Résumé

La conceptualisation métaphorique est omniprésente dans les textes scientifiques et soulève des problèmes quant à sa traduction. Nous proposons deux méthodes d'annotation faisant appel au langage XML, adaptées à l'étude de corpus traduits, et qui découlent d'une méthode antérieurement décrite pour la caractérisation de corpus comparables. La problématique a été soulevée dans le cadre d'une analyse des traductions françaises de *On the Origin of Species* de Charles Darwin, une œuvre scientifique phare reconnue pour sa richesse métaphorique. La première stratégie d'annotation fait appel au langage XLink pour établir un lien unidirectionnel entre fichiers contenant les textes source et cible. Dans la deuxième stratégie, l'annotation est subséquente à un alignement des textes source et cible. Les deux méthodes sont évaluées en fonction de leur faisabilité, de l'ergonomie de l'annotation et de l'usage en aval des fichiers annotés. Elles permettent d'obtenir des résultats quantitatifs exploitables statistiquement à l'aide de tests non paramétriques tels que le X^2 .

Keywords: modes of metaphorical conceptualization, fictive representation, annotation, X^2 test

Mots-clés : mode de conceptualisation métaphorique, représentation fictive, annotation, test du X^2

1. Introduction

Le présent travail s'inscrit dans un programme de recherche qui porte sur les modes de conceptualisation en science (cf. Vandaele 2002 ; Vandaele 2009 pour une revue), en synchronie (époque moderne) et en diachronie (histoire des sciences; voir Vandaele 2012). Fondée essentiellement sur l'approche de sémantique cognitive des métaphores conceptuelles proposée par Lakoff et Johnson (1980/2003) ainsi que sur le modèle général de la fictivité de Talmy (2000), notre analyse fait appel au repérage, en corpus, des indices de conceptualisation métaphorique (ICM ; Vandaele et Lubin 2005) et des représentations

fictives associées. Nous renvoyons le lecteur à Vandaele (2009) pour une description détaillée de ces concepts et de leur application aux textes dans les sciences de la vie.

Notre étude a pour objectif le repérage des modes de conceptualisation métaphoriques présents dans l'œuvre princeps de Charles Darwin, *On the Origin of Species* (1859; dorénavant *OS*) et la manière dont ils ont été rendus dans les différentes traductions en français (Gendron-Pontbriand 2013; Vandaele et Gendron-Pontbriand 2014). L'*OS*, qui donne le coup d'envoi à la théorie de l'évolution, constitue une œuvre phare qui pave la voie des sciences de la vie dès l'orée du XX^e siècle (Ruse 1999). Tout comme du temps de Darwin, elle fait toujours l'objet de contestations pour des raisons idéologiques ou religieuses (créationnisme). Une particularité bien connue de la théorie de l'évolution est qu'elle est construite de façon hautement métaphorique (voir par ex. Young 1985 ; Ruse 2000). De fait, appliquant l'approche décrite plus haut, nous y trouvons des ICM prédictifs véhiculant différents types de représentations fictives. Par exemple, le tableau 1 présente les verbes *to select* et *to tend*, qui attestent de la conceptualisation de la nature comme une personne : la représentation fictive est une action fictive. On remarquera que *Nature*, qui exprime le premier actant de ces ICM, prend une majuscule et que Darwin a eu recours au pronom *she*, ce qui, en plus du réseau lexical mobilisé, renforce la personnification de la nature. Le tableau 1 montre également que Clémence Royer (1862) et Aurélien Berra (2009) ont repris la même conceptualisation métaphorique dans leurs traductions françaises respectives.

Action fictive, personnification : NATURE IS A PERSON	
Darwin (1861 : 87)	Man selects only for his own good; <i>Nature</i> [<i>selects</i>] only for that of the being which <i>she</i> <i>tends</i> .
Action fictive, personnification : LA NATURE EST UNE PERSONNE	
Royer (1862 : 119)	L'homme ne choisit qu'en vue de son propre avantage, et la <i>nature</i> [<i>choisit</i>] seulement en vue du bien de l'être dont elle <i>prend soin</i> .
Berra (2009 : 365)	L'homme sélectionne seulement pour son propre bien : la <i>nature</i> [<i>sélectionne</i>] seulement pour celui de l'être dont elle <i>s'occupe</i> .

Tableau 1 : Exemples d'action fictive avec personnification

Le tableau 2 fait état des ICM *Tree* et *branches*, qui sous-tendent la conceptualisation du signifié de *Life*, c'est-à-dire l'ensemble des êtres vivants (le vivant)¹. Combinant une métaphore et une métonymie, il s'agit ici d'un cas, assez rare, de métaphtonymie (Goossens, 1990). Ici aussi, les deux traducteurs ont repris les mêmes conceptualisations.

État fictif : LIFE [→ LIVING ORGANISMS] IS A TREE	
Darwin (1861 : 147)	[...] I believe it has been with the great Tree of Life , which <i>fills</i> with its dead and broken branches the crust of the earth [...].
État fictif : LA VIE [→ LE VIVANT] EST UN ARBRE	
Royer (1862 : 190)	[...] il en a été, je crois, du grand arbre de la <i>vie</i> qui <i>remplit</i> l'écorce de la terre des débris de ses branches mortes [...].
Berra (2009 : 422)	[...] il en a été de même, je crois, par la génération, pour le grand Arbre de la Vie , qui <i>emplit</i> de ses branches mortes et brisées l'écorce de la terre [...].

Tableau 2 : Exemples d'état fictif

¹ Cette métaphtonymie est ailleurs explicitée sous forme d'analogie par Darwin : « The affinities of all the beings of the same class have sometimes been represented by a great tree » (Darwin 1861 : 146).

L'*OS*, dont on compte six éditions² originales successives, a été abondamment traduite et retraduite dans de nombreuses langues. Plusieurs traductions sont parues en français, les premières étant contemporaines de Darwin. Parmi celles-ci, celle de Clémence Royer (1861) reniée par Darwin lui-même. À l'époque moderne, outre Aurélien Berra (2009), déjà cité, qui a retraduit la 6^e édition de 1876, Daniel Becquemont (1992/2008) s'est attaqué à la traduction de la 1^{re} édition, et Thierry Hocquet (2013), à sa retraduction, tous deux avec l'objectif explicite de retrouver « l'authenticité » de la pensée darwinienne. Nous avons ainsi affaire à un corpus parallèle multiple, diachronique et historique (voir Vandaele et Gendron-Pontbriand 2014, pour le détail des éditions et des traductions).

Le caractère essentiel des modes de conceptualisation métaphoriques dans l'*OS* et dans la théorie darwinienne, conjugué à l'existence de multiples versions et retraductions, soulève ainsi la question de savoir comment ces conceptualisations, initialement exprimées en anglais par un auteur particulièrement soucieux de son écriture et de la portée de ses idées, ont été réexprimées dans une autre langue, en l'occurrence le français.

Dans le présent travail, nous nous concentrerons sur les aspects méthodologiques et statistiques. À cette fin, nous nous limiterons à la comparaison de deux traductions, celle de Royer (1962) – dont Darwin était insatisfait –, et celle de Berra (2009) – présentée par le traducteur lui-même comme plus fidèle à l'original. Après avoir présenté la méthodologie (section 2), nous exposerons et discuterons les résultats (section 3) avant de conclure.

2. Matériel et méthodes

2.1. Constitution du corpus

L'ampleur de l'*OS* (plus de 150 000 mots pour la 1^{re} édition, près de 200 000 pour la 6^e) a imposé de cerner stratégiquement le corpus à l'étude. Nous avons d'abord sélectionné le chapitre IV, qui traite du concept fondamental de la théorie de l'évolution, soit la sélection naturelle³. Nous avons ensuite annoté l'intégralité du chapitre IV en anglais selon la méthode décrite à la section 3, ce qui nous a ainsi donné une vue d'ensemble de la conceptualisation métaphorique dans le chapitre⁴. Cependant, les traductions sont issues d'éditions originales différentes. Ainsi, en prévision de l'étude de toutes les traductions par la méthode ici décrite, nous avons identifié les phrases inchangées dans les six éditions originales, afin de pouvoir comparer les traductions qui en sont issues (sous-corpus EN311 : 311 phrases). Enfin, pour des raisons de faisabilité liée au caractère chronophage de l'annotation, nous avons retenu les phrases dont la densité en ICM était la plus élevée (sous-corpus EN55 : 55 phrases totalisant 2645 mots). L'annotation complète du chapitre IV en anglais nous a permis d'évaluer la représentativité de EN55. Enfin, les phrases correspondantes à celles de EN55 dans les traductions respectives de Royer (sous-corpus FR1862 : soit 65 phrases et 2 964 mots⁵) et de Berra (sous-corpus FR2009 : 55 phrases et 2 946 mots) ont été repérées et annotées. La préparation des fichiers constituant le corpus suit le protocole détaillé dans Vandaele, Bodain et Raffo (2012) (numérisation, nettoyage des textes, conversion en format TXT puis en format

² Éditions originales : 1859, 1860, 1861, 1866, 1869, 1872 ; plus une édition presque identique à celle de 1872, dite édition 6b, en 1876.

³ Le même choix a été fait, de manière indépendante, par Vandepitte *et al.* (2011) pour une étude des traductions néerlandaises de l'*OS*.

⁴ L'annotation a été effectuée sur le chapitre IV de la 3^e édition anglaise (1861).

⁵ Les traductions suivent, globalement, l'enchaînement phrastique du texte darwinien mais, à l'occasion, une phrase peut être rendue par deux phrases dans le texte d'arrivée, ou l'inverse.

XML). L'annotation a été réalisée avec l'éditeur XML⁶ *Oxygen 14.0* pour Mac OS. L'extraction des données s'est faite à l'aide du langage de requête XQuery.

2.2. Annotation : balises et schéma

Nous avons repris et modifié la méthode d'annotation des ICM au sein de corpus en format XML décrite dans Vandaele et Boudreau (2006). Initialement conçue pour l'annotation de corpus comparables (non traduits) ou exploités comme tels (sans tenir compte de l'opération de traduction), elle a été aménagée afin d'autoriser, sur des corpus parallèles, une mise en correspondance point par point de segments originaux et de leurs traductions, avec pour objectif la mise en œuvre de tests statistiques permettant de tester une hypothèse nulle de fidélité totale entre l'original et les traductions.

L'annotation fait appel à plusieurs balises. Deux balises structurantes, `<phr>` et `<par>`, permettent de segmenter le texte en phrases et en paragraphes, respectivement. Toutes deux sont numérotées automatiquement lors de conversion des fichiers TXT en fichiers XML (attribut `@position`). Les balises `<concIndPred>` et `<concInd>` servent au repérage des ICM prédicatifs (tableau 3-A) et non prédicatifs (tableau 3-B). La balise `<lingEl>` repère les éléments linguistiques réalisant les actants des ICM prédicatifs. Ces trois balises contiennent des attributs permettant de consigner un numéro d'identification (`@id`), la partie du discours (`@lex`) et la forme lemmatisée (`@lem`) de l'élément annoté. La valeur de l'attribut `@id` est construite à partir de celle de l'attribut `@position` de la phrase correspondante, suivi du numéro de l'élément annoté au sein de la phrase (par ex., `id="_51-1"` : élément 1 dans la phrase 51). L'essentiel pour la suite est que les valeurs de `@id` sont uniques. Auparavant défini dans le schéma de validation par un type `xs:token`, `@id` s'est vu attribuer un type `xs:ID`, ce qui permet de valider l'unicité des valeurs de l'attribut au sein du fichier.

Dans la balise `<concIndPred>`, autant d'attributs `@actn` que nécessaire permettent de consigner l'identifiant unique des actants de l'élément annoté (`act1="_51-2"`). Cet attribut, lui aussi antérieurement de type `xs:token`, est dorénavant défini comme un type `xs:IDREFS`, lequel fonctionne en tandem avec le type `xs:ID`. En effet, un attribut de type `xs:IDREFS` doit pointer vers un identifiant unique de type `xs:ID` dans le même fichier. Le type `xs:IDREFS` permet également de consigner un ou plusieurs identifiants uniques⁷, ce qui est utile, par exemple, lorsque l'actant sémantique d'un ICM prédicatif est réalisé en discours par au moins deux expressions linguistiques coordonnées par une conjonction de coordination. Enfin, la balise `<concIndPred>` contient deux attributs permettant de caractériser la conceptualisation métaphorique : `@metn`, pour indiquer la catégorie conceptuelle la plus saillante se projetant sur l'actant *n* (par ex. : `metn="personne"`) et `@repFict`, pour indiquer la représentation fictive (par ex. : `repFict="personnification"`). L'attribut `@repFict` permet aussi de caractériser les ICM non prédicatifs, lesquels comportent, en plus, l'attribut `@proj`. Cet attribut accepte une chaîne de caractères, permettant ainsi de consigner le ou les traits sémantiques projetés (par ex. : `proj="forme"`). La balise `<concInd>` contient aussi l'attribut `@synt`, qui permet de colliger les syntagmes auxquels participent des ICM quasi-prédicatifs.

⁶ XML 1.0 (Second Edition), World Wide Web Consortium. *Extensible Markup Language (XML) 1.0, Second Edition*. Disponible à : <http://www.w3.org/TR/2000/WD-xml-2e-20000814>

⁷ Contrairement au type `xs:IDREF`, qui n'admet qu'une seule référence.

A	<p><i>Man selects only for his own good; Nature only for that of the being which she tends.</i> (Darwin 1861 : 87)</p> <pre><phr position="_51"> Man selects only for his own good; <lingEl id="_51-1" lex="nom" lem="nature">Nature</lingEl> only <concIndPred id="_51-2" lex="verbe" lem="select, to" act1="_51-1" met1="agent" repFict="personnification"/> for that of the being which<lingEl id="_51-3" lex="pronom" lem="nature">she</lingEl> <concIndPred id="_51-4" lex="verbe" lem="tend, to" act1="_51-2" met1="agent" repFict="personnification"> tends</concIndPred>.</phr></pre>
B	<p><i>[...] I believe it has been with the great Tree of Life [...]</i> (Darwin 1861 : 147)</p> <pre><phr position="_664"> [...] I believe it has been with the great <concInd id="_664-9" lex="nom" synt="Tree of Life" proj="forme" repFict="état fictif">Tree </concInd> of Life [...].</phr></pre>

Tableau 3 : Exemple d'annotation d'ICM. A) ICM prédicatif *selects* et *tends* : le verbe étant elliptique dans la seconde proposition, une balise vide (id="_51-2") permet de lui associer son premier actant (id="_51-1"). Pour permettre l'analyse, la valeur de @lem dans l'élément balisant *she* est celle de l'antécédent (*Nature*). B) ICM quasi-prédicatif *Tree* : l'unité *Life* n'étant pas vraiment un actant, elle n'est pas annotée comme tel, mais l'attribut @synt rend compte du syntagme *Tree of Life*.

2.3. Mise en relation des corpus parallèles

Plusieurs scénarios de traduction des ICM sont envisageables (figure 1) : 1) un ICM du texte source (TS) est traduit par un ICM dans le texte cible (TC) ; l'équivalence peut être complète (si la représentation fictive est identique) ou partielle (si la représentation fictive est différente) ; 2) un ICM du TS est traduit par une expression non métaphorique dans le TC (perte d'une occurrence d'ICM) ; 3) un élément non métaphorique dans le TS est traduit par un ICM dans le TC (gain d'une occurrence d'ICM).

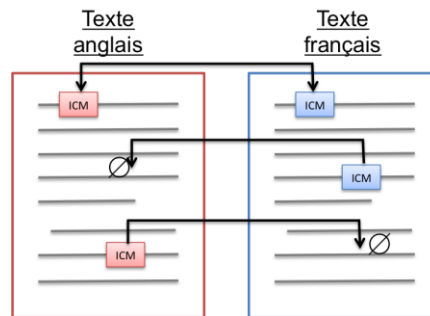


Figure 1 : Scénarios de traduction

L'étude des ICM exprimés dans un corpus parallèle impose donc la mise en correspondance point par point du texte source avec les traductions. Pour ce faire, nous avons testé la pertinence de deux méthodes.

2.3.1. Méthode 1 : mise en relation de fichiers XML distincts

La première méthode s'applique au texte source et aux textes cibles annotés séparément dans des fichiers distincts. Le langage XLink (*XML Linking Language*) permet de créer un lien unidirectionnel entre les fichiers (tableau 4). Deux attributs supplémentaires, pointant vers des balises du TS, sont ajoutés dans les balises <concIndPred>, <concInd> et <lingEl> des TC : @xlink:type (spécifie le type de lien : simple, dans notre cas) et @xlink:href (spécifier l'identifiant unique de la balise ciblée dans le TS).

<p><i>[...] les espèces natives ont été si complètement vaincues par des espèces naturalisées qu'elles ont laissé ces étrangères prendre définitivement possession du sol.</i> (Royer 1862 : 119)</p>
<pre><phr position="_82"> [...] les <lingEl id="_82-11" [...]>espèces</lingEl> natives ont été si complètement <concIndPred id="_82-13" act1="_82-15" [...] xlink:type="simple" xlink:href="Darwin1861OriginCh4.xml#_46-6">vaincues</concIndPred> [...].</phr></pre>

Tableau 4 : Pointeur Xlink dans la traduction de Royer (1862). Les détails de l'annotation ont été ôtés pour faciliter la lecture : les crochets [...] ne font pas partie du langage XML.

2.3.2. Méthode 2 : annotation de fichiers alignés

La deuxième méthode consiste à annoter un bitexte généré par un outil tel que *LogiTerm Pro* (Terminotix), un logiciel commercial destiné aux langagiers professionnels qui comprend un aligneur et un éditeur d'alignement. Une fois l'alignement vérifié, celui-ci peut être exporté en format XML. Le fichier XML obtenu ne respectant pas le schéma utilisé pour l'étude, il faut le nettoyer pour le rendre le plus conforme possible (numéroté les phrases, renommer les balises, etc.), idéalement à l'aide d'une transformation XSLT, surtout pour des grands corpus. Le schéma doit néanmoins être adapté, avec l'ajout d'une balise permettant de repérer chaque texte aligné. Un attribut @src (pour « source »), de type xs:IDREFS, est ajouté dans les balises <concIndPred>, <concInd> et <lingEl>, afin de consigner l'identifiant unique de la balise correspondante dans l'original anglais. Cette méthode est particulièrement intéressante pour annoter des alignements multiples (tableau 5).

[...] <i>there will have been much extinction of the less improved forms [...].</i> (Darwin 1861 : 114)
[...] <i>et beaucoup des formes inférieures ont dû s'éteindre.</i> (Royer 1862 : 150)
[...] <i>il y aura une forte extinction des formes les moins améliorées [...].</i> (Berra 2009 : 392)
<pre> <phr id="_313"> <en1861 id=" _313-1">[...] there will have been much <concIndPred id="_313-1-1" [...]>extinction </concIndPred> of the <lingEl id="_313-1-2" [...]>less improved forms</lingEl> [...].</en1861> <fr1862 id=" _313-2">[...] et beaucoup des <lingEl id=" _313-2-1" [...] src=" _313-1-2">formes inférieures</lingEl> out dû <concIndPred id=" _313-2-2" [...] src=" _313-1-1">s'éteindre</concIndPred> [...].</fr1862> <fr2009 id=" _313-5">[...] il y aura eu une forte <concIndPred id=" _313-5-1" src=" _313-1-1">extinction </concIndPred> des <lingEl id=" _313-5-2" [...] src=" _313-1-2">formes les moins améliorées</lingEl> [...].</fr2009> </phr> </pre>

Tableau 5 : Exemple d'alignement multiple. <en1861> : texte source de Darwin ; <fr1862> : traduction de Royer (1862) ; <fr2009> : traduction de Berra (2009).

3. Résultats et discussion

3.1. Comparaison des deux méthodes d'annotation

Chaque méthode comporte des avantages et des inconvénients. Pour la première, la préparation des fichiers avant l'annotation est plus brève et plus facilement automatisable. Cependant, elle impose une annotation séquentielle des fichiers, d'abord le texte source, puis le texte cible. L'utilisation du pointeur XLink alourdit considérablement l'annotation et rend plus difficile l'extraction des données par le langage XQuery. L'utilisation de XLink se justifie par les applications en aval, notamment la visualisation de bitextes en format HTML, et la facilité à ajouter des textes à des corpus déjà annotés.

Pour la deuxième méthode, la validation de l'alignement effectué est chronophage en raison du nettoyage à effectuer, et la transformation XSLT permettant de générer des alignements multiples est complexe. L'annotation est plus simple, l'attribut @src étant plus léger à manipuler que le lien XLink, et plus flexible, car elle peut être, selon les besoins, séquentielle (un texte à la fois) ou parallèle (annotation simultanée). L'extraction des données est plus facile, car les requêtes, menées sur un seul fichier, peuvent facilement être prises en charge par XQuery. Malgré tout, une contrainte de fond est que les textes à aligner présentent une structure aussi proche que possible (pas de réécriture libre, pas de déplacements majeurs, etc.), ce qui peut représenter un défi en cas d'alignement multiple. Le choix de la méthode dépend donc du corpus à l'étude et de l'usage que l'on veut faire du fichier annoté

(visualisation Web, exploitation statistique, etc.). Les deux méthodes fournissent les mêmes données et n'ont pas d'influence sur le choix de l'analyse statistique.

3.2. Représentations fictives repérées dans le texte source et représentativité de l'échantillon

L'annotation du chapitre IV complet, qui comprend l'échantillon de 55 phrases utilisées pour l'analyse des traductions, a mis en évidence six catégories de représentations fictives (action, état, intention, processus, qualité, résultat), dans lesquelles la personnification des êtres vivants (représentation anthropomorphique), d'une part, et de la nature, de la sélection naturelle et de la sélection sexuelle, d'autre part ont été distinguées, soit onze catégories au final (tableau 6). Les autres sous-catégories de représentation fictives ont été regroupées dans chacun des cas sous « autres ».

Représentations fictives		Exemples
action fictive	anthropomorphisme ¹	One large group will slowly conquer another large group [...]. (Darwin 1861 : 131)
	personnification ²	Natural Selection acts solely through the preservation of variations [...]. (Darwin 1861 : 115)
	indéterminée ³	[...] most of the many now existing low forms have not in the least advanced [...]. (Darwin 1861 : 135)
	autres	[...] the great Tree of Life, which fills with its dead and broken branches [...]. (Darwin 1861 : 147)
état fictif	anthropomorphisme ¹	[...] if any one species does not become modified and improved in a corresponding degree with its competitors [...]. (Darwin 1861 : 107)
	autres	She can act on [...] the whole machinery of life. (Darwin 1861 : 87)
intention fictive ⁴		[...] the accumulation of variations tending towards specialisation is within the scope of Natural Selection. (Darwin 1861 : 134)
processus fictif	anthropomorphisme ¹	[...] for, had the area been open to immigration , these same places would have been seized on by intruders. (Darwin 1861 : 86)
	autres	[...] there will have been much extinction of the less improved forms [...]. (Darwin 1861 : 114)
qualité fictive	personnification ²	[...] I can see no limit to the amount of change [...] which may be effected [...] by nature's power of selection. (Darwin 1861 : 115)
résultat fictif	anthropomorphisme ¹	[...] such differences would effectually settle which variety [...] should succeed . (Darwin 1861 : 90)

Tableau 6 : Exemple d'ICM pour chaque catégorie de représentation fictive. 1. Anthropomorphisme : personnification des êtres vivants. 2. Personnification de concepts : nature, sélection naturelle et sélection sexuelle. 3. Indétermination : la conceptualisation évoquée peut être interprétée de plusieurs façons. 4. La catégorie Intention n'a pas été sous-catégorisée (petit nombre d'ICM).

La première question porte sur la représentativité de l'échantillon de 55 phrases, comparativement à l'ensemble du chapitre IV⁸. Le tableau 7 présente, en fonction des

⁸ Le chapitre IV de l'OS a été annoté à partir de la 3^e édition (1861), la première qui a été traduite, par Royer (1862). Des différences pourraient être observées si l'échantillon était comparé aux autres éditions, mais elles ne paraissent pas suffisantes pour remettre en question notre démarche. Le *variorum* de l'OS (darwin-online.org.uk/Variorum/, édité par Barbara Bordalejo) montre en effet que plus de la moitié des phrases du chapitre IV demeurent identiques dans les 3^e, 5^e et 6^e éditions (c'est-à-dire les éditions qui ont été traduites).

catégories fictives, la distribution du nombre d'occurrences d'ICM dans l'échantillon (EN55, 2645 mots) et dans l'ensemble du chapitre (ENTot, 19 996 mots), ainsi que le nombre d'ICM distincts par catégorie. La première série de données permet d'accéder à la densité des ICM dans le corpus, tandis que la deuxième série informe sur la richesse lexicale mobilisée. Par exemple, un ICM donné peut apparaître 3 fois dans le corpus (3 occurrences). L'échantillon contient donc 178 occurrences de 84 ICM différents, et le chapitre IV, 538 occurrences de 134 ICM distincts.

Représentations fictives	N occurrences d'ICM		N ICM distincts	
	EN55	EnTot	EN55	EnTot
action fictive -- anthropomorphisme	16	40	11	18
action fictive -- personnification	79	267	27	51
action fictive -- indéterminée	25	63	15	21
action fictive -- autres	11	25	6	7
état fictif -- anthropomorphisme	27	85	11	18
état fictif -- autres	5	6	4	5
intention fictive	1	1	1	1
processus fictif -- anthropomorphisme	6	33	4	7
processus fictif -- autres	2	6	1	1
qualité fictive -- personnification	1	1	1	1
résultat fictif -- anthropomorphisme	5	11	3	4
Total	178	538	84	134
X²	30,56		3,68	
ddl	10			
X² théorique	23,21		[3,94-4,86]	
p	p < 0,01		0,90 < p < 0,95	

Tableau 7 : Distribution des occurrences d'ICM et des ICM dans le texte source : échantillon (EN55) et chapitre IV complet (EnTot)

Le sous-corpus En55 représente 13,2 % du chapitre, mais contient 33 % des occurrences d'ICM et 47 % des ICM. Il est « enrichi » (volontairement) en ICM. Mais cette population « enrichie » se répartit-elle dans les différentes catégories de représentation fictive de la même manière dans EN55 et dans le chapitre entier ? On ne peut en aucun cas postuler que les données suivent une loi normale, ce que confirme un test de Kolmogorov-Smirnov appliqué aux distributions (réalisé avec le logiciel *GraphPad Prism v.6* pour MacOS). Le test requis pour comparer les distributions doit donc être non paramétrique, soit le test de Pearson (X^2), que nous avons appliqué selon la technique proposée par Muller (1968 : 95-110), qui stipule les conditions de son application à des corpus et la manière de déterminer la distribution théorique : en l'occurrence, elle est calculée à partir de l'échantillon et ramenée à l'effectif observé pour le chapitre complet. Le test montre que la répartition des occurrences d'ICM est différente de celle du chapitre ($p < 0,01$). En effet, l'échantillon est particulièrement enrichi en ICM d'action fictive témoignant d'une personnification des concepts fondamentaux de l'OS (conceptualisation de la nature, de la sélection naturelle et de la sélection sexuelle) et de processus fictifs témoignant d'une tendance à l'anthropomorphisme (personnification des espèces animales et végétales).

Par contre, on ne peut pas conclure que la distribution des ICM soit significativement différente entre EN55 et ENTot ($0,90 < p < 0,95$). Compte tenu du fait que EN55 contient 47 % des ICM du chapitre, bien que ne comptant que 13,2 % des mots du chapitre, on pourrait envisager une stratégie d'extension du corpus pour capter des occurrences de tous les ICM du chapitre et en étudier la traduction, sans avoir à analyser l'ensemble du chapitre pour toutes les traductions considérées. L'ensemble de ces données nous indique que l'examen des

traductions sur cet échantillon devrait pouvoir nous renseigner sur les stratégies utilisées par les traducteurs dans les textes cibles avec une certaine fiabilité, tout au moins pour les représentations fictives les plus riches.

3.3. Analyse statistique des corpus traduits

Nous cherchons à déterminer dans quelle mesure les traducteurs ont suivi les modes de conceptualisation exprimés dans le texte source. Nous présentons les données relatives à l'échantillon anglais (EN55) et aux traductions de Royer (FR1862) et de Berra (FR2009). L'attribution d'une catégorie de représentation fictive aux ICM repérés dans le texte source et les textes cibles autorise la comparaison des effectifs de chacune de ces catégories à l'aide d'un test de Pearson (X^2).

3.3.1. Nombre d'occurrences des ICM

L'hypothèse nulle est que les textes cibles sont identiques au texte sources sur le plan de l'expression de la conceptualisation métaphorique (c'est-à-dire sur le plan du sens) : en raison de la nature unidirectionnelle du processus de traduction, nous considérons donc que la distribution théorique équivaut à celle de l'échantillon anglais (EN55). En effet, l'hypothèse nulle impliquerait l'existence d'une traduction (imaginaire...) parfaitement équivalente, dans laquelle le nombre d'occurrences d'ICM serait le même que celui de l'anglais, avec une répartition identique dans les mêmes catégories de représentation fictive, aucun transfert d'ICM d'une catégorie à l'autre, aucune perte d'ICM et aucun ajout. Les critères de faisabilité du χ^2 (effectif identique pour les distributions à comparer) nous obligent cependant à prendre en compte une catégorie additionnelle, que nous avons nommée *représentation factive*, qui est celle dans laquelle tombent les expressions non métaphoriques : cette catégorie rend compte de la figure 1, qui illustre le fait que, dans la réalité, certaines expressions non métaphoriques sont traduites par un ICM, et inversement.

Le tableau 8 présente le nombre d'occurrences d'ICM relevées dans EN55, FR1862 et FR2009, en fonction de la catégorie de représentation.

N occurrences d'ICM		Catégories de représentation	N occurrences d'ICM	
EN55	FR1862		EN55	FR2009
16	20	action fictive -- anthropomorphisme	16	17
79	73	action fictive -- personnification	79	87
25	20	action fictive -- indéterminée	25	20
11	9	action fictive -- autres	11	11
27	18	état fictif -- anthropomorphisme	27	25
5	10	état fictif -- autres	5	5
1	0	intention fictive	1	0
6	5	processus fictif -- anthropomorphisme	6	5
2	2	processus fictif -- autres	2	2
1	2	qualité fictive -- personnification	1	2
5	2	résultat fictif -- anthropomorphisme	5	5
15	32	<i>représentation factive (non-ICM)</i>	9	8
193	193	Effectifs totaux	187	187
34,05		X^2	4,299	
11		ddl	11	
24,72		X^2 théorique	[4,57 - 5,57]	
p < 0,01		p	(0,95 < p < 0,99)	

Tableau 8 : Occurrences d'ICM dans EN55, FR1862 et FR2009 selon la catégorie de représentation

On peut constater que les distributions des occurrences d'ICM sont statistiquement différentes entre EN55 et FR1862, pour une valeur de $p < 0,01$. Par contre, on ne peut pas mettre en évidence de différence statistiquement significative entre EN55 et FR2009 ($p > 0,95$). Une analyse plus fine, qui découle directement de la méthode d'annotation mettant en relation les ICM sources et leurs traductions et les ICM ajoutés dans le texte cible, montre en fait qu'il y a un mouvement de certains ICM d'une catégorie à l'autre⁹. Royer est celle qui joue le plus librement avec les solutions de traduction, ce qui corrobore, de manière quantitative, les observations qualitatives (par ex., Conry 1974), ce qui expliquerait le rejet de sa traduction par Darwin. Par contre, la traduction de Berra suit de très près l'original anglais, ce qui corrobore également les affirmations du traducteur lui-même quant à son projet.

3.3.1. Nombre d'ICM distincts

Le tableau 9 présente le nombre des ICM distincts auxquels les traducteurs ont eu recours, comparativement aux ICM présents dans le texte source. Les effectifs sont ici trop faibles pour effectuer un test du X^2 fiable (Muller 1968 : 101), même avec des regroupements de catégories, lesquels enlèveraient toute signification à l'analyse. Une étude sur un corpus plus vaste serait nécessaire pour obtenir des données statistiquement significatives. Qualitativement, on observe toutefois que Royer emploie une variété d'ICM plus grande que celle du texte source et que celle du texte traduit par Berra, celui-ci se distinguant à peine de l'original anglais. Ainsi, la tendance observée plaide pour une plus grande liberté de Royer quant au texte source, ce qui est cohérent avec les données décrites ci-dessus.

Catégories de représentation fictive	ICM		
	EN55	FR1862	FR2009
action fictive -- anthropomorphisme	11	15	13
action fictive -- personnification	27	34	31
action fictive -- indéterminée	15	13	10
action fictive -- autres	6	5	6
état fictif -- anthropomorphisme	11	12	11
état fictif -- autres	4	9	4
intention fictive	1	0	0
processus fictif -- anthropomorphisme	4	1	3
processus fictif -- autres	1	2	1
qualité fictive -- personnification	1	2	2
résultat fictif -- anthropomorphisme	3	2	3
Total	85	94	84

Tableau 9 : Nombre d'ICM distincts dans l'échantillon anglais et les traductions de Royer et de Berra

4. Conclusion

De manière générale, les approches quantitatives et statistiques sont encore très peu utilisées en traductologie (de Sutter *et al.* 2012), et ce, malgré l'avènement des linguistiques de corpus (Habert *et al.* 1997). Avec Baker (1993), nous pensons qu'il est important de les développer, afin de pouvoir confirmer les intuitions obtenues qualitativement par des données probantes. Le présent article décrit une approche originale d'annotation de corpus multilingue constitué d'un texte source et de plusieurs traductions. Nous l'appliquons à une analyse de la conceptualisation métaphorique dans les sciences, mais la méthode est potentiellement applicable à de nombreux contextes. Les données obtenues, sous certaines conditions

⁹ L'explicitation détaillée de ces données dépasse le cadre du présent article.

d'applicabilité, sont analysables à l'aide de tests statistiques. Ils permettent de juger de la validité des résultats obtenus avec les corpus d'analyse pour, le cas échéant, envisager leur extension.

Nous avons déterminé qu'il fallait avoir recours à des statistiques non paramétriques, plus précisément le test de Pearson, appliqué selon la méthode décrite par Muller (1968). Le point clé est la détermination de la distribution théorique : dans un cas, l'analyse porte sur la représentativité de l'échantillon, dans l'autre, il s'agit de comparer des traductions à un texte source, l'hypothèse nulle étant celle d'une équivalence de sens. Il est crucial de souligner que les deux traductions ne peuvent être comparées entre elles : additionner les valeurs observées dans les textes cibles pour obtenir une distribution théorique reviendrait à postuler que les deux textes proviennent d'un même corpus homogène. Or, c'est faux, car cela reviendrait à créer artificiellement un texte de référence qui représenterait une moyenne des changements effectués par les traducteurs. La comparaison doit donc être faite avec le texte source.

Enfin, les valeurs obtenues avec le test de Pearson nous fournit une indication sur la proximité des textes cibles avec le texte source pour les paramètres choisis, ce qui nous permet d'envisager un classement des traductions : un faisceau d'indices quantitatifs convergent pour établir que Royer (1862) s'éloigne du texte original de façon marquée, tandis que Berra (2009) le suit au plus près. Des travaux futurs nous permettront d'explorer les autres traductions de cette œuvre princeps qui a révolutionné la science.

Remerciements

Le présent travail a reçu l'appui financier du Conseil de recherches en sciences humaines (CRSH) du Canada, sous forme de bourse doctorale (attribuée à Eve-Marie Gendron-Pontbriand ; Programme de bourses d'études supérieures Joseph-Armand-Bombardier) et d'une subvention de recherche Savoir (attribuée à Sylvie Vandaele). Les auteurs remercient Nathan Ménard pour les discussions fructueuses sur l'applicabilité du X^2 .

Références

- Baker, M. (1993). Corpus Linguistics and Translation Studies. Implications and Applications. In M. Baker, G. Francis et E. Tognini-Bonelli (dirs.), *Text and Technology : In honour of John Sinclair* (pp. 223-250). Amsterdam/Philadelphie : John Benjamins Publishing Company.
- Conry, Y. (1974). *L'introduction du darwinisme en France au XIX^e siècle*. Paris : Vrin.
- Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life* (1^{re} éd.). Londres : John Murray.
- Darwin, C. (1861). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life* (3rd ed.). Londres : John Murray.
- Darwin, C. (1862). *De l'origine des espèces ou des lois du progrès chez les êtres organisés* (Trad. de la 3^e éd. anglaise par C.-A. Royer). Paris : Guillaumin et Cie/Victor Masson et Fils.
- Darwin, C. (2009). *L'origine des espèces par le moyen de la sélection naturelle, ou la préservation des races favorisées dans la lutte pour la vie* (Trad. de la 6^e éd. anglaise par A. Berra). Paris : Honoré Champion.
- De Sutter, G., Goethals, P., Leuschner, T. et Vandepitte, S. (2012). Toward Methodologically More Rigorous Corpus-Based Translation Studies. *Across Language and Cultures*, 13(2), 137-143.
- Goossens, L. (1995). Metaphonymy: The interaction of metaphor and metonymy in figurative

- expression for linguistic action. In L. Goosses, P. Pauwels, B. Rudzka-Ostyn, A.-M. Simon-Vandenbergen et J. Vanparys (dirs), *By Word of Mouth. Metaphor, Metonymy and Linguistic Action in a Cognitive Perspective* (pp. 159-174). Amsterdam/Philadelphie : John Benjamins.
- Habert, B., Nazarenko, A. et Salem, A. (1997). *Les linguistiques de corpus*. Paris : Armand Colin.
- Lakoff, G. et Johnson, M. (1980/2003). *Metaphors We Live by*. Chicago : The University of Chicago Press.
- Muller, C. (1968). *Initiation à la statistique linguistique*. Paris : Librairie Larousse.
- Oakes, M. P. et Ji, M. (2012). *Quantitative Methods in Corpus-Based Translation Studies. A practical guide to descriptive translation research*. Amsterdam/Philadelphie : John Benjamins.
- Polguère, A. (2012). Propriétés sémantiques et combinatoires des quasi-prédicats sémantiques. *Scolia*, 26, 131-152.
- Ruse, M. (1999). *The Darwinian Revolution: Science Red in Tooth and Claw*. 2e éd. Chicago/Londres : The University of Chicago Press.
- Ruse, M. (2000). Metaphor in evolutionary biology. *Revue Internationale De Philosophie*, 54(214), 593-619.
- Talmy, L. (2000). Fictive Motion in Language and "Ception". In L. Talmy (dir.), *Towards a cognitive semantics. Volume I : Concept structuring systems* (pp. 99-175). Cambridge : The MIT Press.
- Vandaele, S. (2002). Métaphores conceptuelles en traduction médicale et cohérence. *TTR*, 15(1), 223-239.
- Vandaele, S. (2009). Les modes de conceptualisation du vivant : une approche linguistique. In S. David, J. Przychodzen & F.-E. Boucher (dirs.), *Que peut la métaphore? Histoire, savoir et poétique* (pp. 187-207). Paris : L'Harmattan.
- Vandaele, S. et Béland, M.-C. (2012). Les modes de conceptualisation des unités d'hérédité au XIX^e siècle : Spencer, Haeckel et Elsberg. *Linguistica Antverpiensia*, 11, 227-246.
- Vandaele, S., Bodain, Y. et Raffo, M. (2012). Protocole de préparation et d'annotation des corpus en vue de l'étude de la conceptualisation métaphorique. *Rapport interne*. Montréal, Université de Montréal.
- Vandaele, S. et Boudreau, S. (2006). Annotation XML et interrogation de corpus pour l'étude de la conceptualisation métaphorique. *Lexicometrica*, Actes JADT (2006), 943-951.
- Vandaele, S. et Gendron-Pontbriand, E.-M. (2014). Des « vilaines infidèles » à la postérité : traduction et retraduction de l'œuvre de Charles Darwin. In J. Pinilla et B. Lépinette (dirs), *Traducción y difusión de ciencia y de la técnica en España en los siglos XVIII y XIX* (pp. 249-275). Valencia : Universitat de València.
- Vandaele, S. et Lubin, L. (2005). Approche cognitive de la traduction dans les langues de spécialité : vers une systématisation de la description de la conceptualisation métaphorique. *Meta*, 50(2), 415-431.
- Vandepitte, S., Vandenbussche, L. et Algoet, B. (2011). Travelling Certainties : Darwin's Doubts and Their Dutch Translations. *The Translator*, 17(2), 275-299.
- Young, R. M. (1985). *Darwin's Metaphor : Nature's Place in Victorian Culture*. Cambridge : Cambridge University Press.