

Univerzita Karlova  
Pedagogická fakulta

DIPLOMOVÁ PRÁCE

2017

Bc. Mariana Pípalová

**Univerzita Karlova**  
**Pedagogická fakulta**

Katedra francouzského jazyka a literatury

DIPLOMOVÁ PRÁCE

**L'analyse des vrais amis anglo-français**  
**d'un corpus d'échantillons textuels authentiques**

Analýza anglo-francouzských pravých protějšků v korpusu autentických vzorků textů

Analysis of English and French true friends (vrais amis)  
in a corpus of authentic text samples

Bc. Mariana Pípalová

Vedoucí práce: Mgr. Jiří Jančík  
Studijní program: Učitelství pro střední školy  
Studijní obor: N AJ – FJ

2017

**Prohlášení:**

Prohlašuji, že jsem diplomovou práci na téma *L'analyse sur corpus des vrais amis en anglais et en français dans un ensemble authentique d'échantillons des textes* vypracovala pod vedením vedoucího práce samostatně za použití v práci uvedených pramenů a literatury. Dále prohlašuji, že tato práce nebyla využita k získání jiného nebo stejného titulu.

V Praze, dne 18. dubna 2017

.....

### **Poděkování:**

Ráda bych tímto poděkovala vedoucímu své diplomové práce, Mgr. J. Jančíkovi, za jeho trpělivost, vstřícnost a cenné připomínky. Dále bych chtěla poděkovat své konzultantce, PhDr. K. Matuchové, PhD., za její čas, cenné rady a podněty. Současně bych chtěla vyjádřit svou vděčnost MgA. D. Gigon za její ochotu provést jazykovou korekci.

## **ABSTRAKT**

Diplomová práce se zabývá parolovým výzkumem pravých protějšků (*vrais amis*) v současné francouzštině a angličtině. K tomuto účelu byl sestaven speciální korpus anglicko-francouzských překladů složený ze třech dílčích subkorpusů stejného rozsahu (biblický diskurz, politický diskurz a diskurz umělecké prózy), v rozsahu cca 60 000 slov. Pomocí nástroje AntConc byly z korpusu vygenerovány užité pravé protějšky, které se v práci pojímají vždy jakožto jev konkrétního registru. Na základě frekvenčních kritérií byl vymezen centrální okruh 64 protějšků, (takřka) shodných frekvencí a identických kontextů a registrů, který byl následně podroben mnohaaspektové analýze. Zkoumala se jejich výslovnost i psaná podoba, jejich slovnědruhové zařazení, zastoupení slovotvorného postupu derivace; dále se sledovala absolutní četnost jak typů, tak i tokenů v korpusu, a nepřímou (na základě pásem frekventovanosti) se zjišťovala i integrovanost těchto výpůjček do slovní zásoby angličtiny jakožto jazyka přijímajícího. Dle většiny zkoumaných aspektů lze pravé protějšky přiřadit do tří pásem od identity přes blízkost až po relativní nepodobnost. Podle teorie centra a periferie (Daneš 1966) lze tak pravé protějšky uspořádat na pomyslné škále od identity až po vzdálenou podobnost ve většině aspektů analýzy. K práci je připojen abecední seznam všech centrálních pravých protějšků zastoupených v jednotlivých subkorpusech. Zjištěné tendence by mohly přispět k objasnění podobností mezi zkoumanými jazyky a zejména tak usnadnit proces jejich osvojování.

## **KLÍČOVÁ SLOVA**

pravé protějšky, angličtina, francouzština, korpus, překlad, registr, mnohaaspektová analýza

## **ABSTRACT**

This final project provides a parole analysis of vrais amis (true counterparts) in current French and English. To this end a specialized English-French translation corpus was assembled, composed of three subcorpora equal in length, namely Religious, Political and Fiction discourse, amounting altogether to approximately 60,000 words. With the help of the AntConc instrument, true friends employed in the corpus were generated, here conceived of as a register-specific phenomenon exclusively. Using the frequency criteria, a central set of 64 most frequent counterparts was delineated. These central counterparts, marked by (almost) identical frequencies, identical contexts and the same registers, were subjected to a multiaspectual analysis, scrutinizing the pronunciation, spelling, word classes, share of derivation, and frequency of types and tokens. Since English proved to be the borrowing language in all instances, the research also indirectly addressed the degree of their integration in the English words stock by reference to frequency bands. For most of the researched aspects, three zones of counterparts were identified, namely those exhibiting identity, close similarity and relative difference. As a result, employing the Theory of Centre and Periphery (Daneš 1966), we may arrange true counterparts along an imaginary scale ranging from identity to relative difference in nearly all the criteria. The final project includes an alphabetically arranged list of all true counterparts pertinent to the individual subcorpora. The tendencies detected may enhance the current understanding of similarities between the two languages and particularly facilitate the process of their acquisition.

## **KEYWORDS**

true counterparts (vrais amis), English, French, corpus, translation, register, multiaspectual analysis

## Sommaire:

|         |  |    |
|---------|--|----|
| 1       | Introduction .....                                       | 1  |
| 2       | Partie théorique .....                                   | 3  |
| 2.1     | L'Arrière-plan.....                                      | 3  |
| 2.1.1   | L'Arrière-plan historique.....                           | 3  |
| 2.1.2   | Les langues en contact .....                             | 5  |
| 2.2     | Définition des termes lexicologiques.....                | 6  |
| 2.2.1   | Les vrais amis.....                                      | 8  |
| 2.2.2   | Les faux amis .....                                      | 9  |
| 2.2.2.1 | Les faux amis dans le cadre sémantique .....             | 11 |
| 2.2.3   | Les faux amis partiels .....                             | 11 |
| 2.2.4   | Le lexique.....  | 13 |
| 2.3     | La conception saussurienne .....                         | 15 |
| 2.3.1   | La sémiologie et le caractère social de la langue .....  | 15 |
| 2.3.2   | La théorie du signe linguistique .....                   | 16 |
| 2.3.3   | Diachronie vs. synchronie.....                           | 18 |
| 2.3.4   | Langue vs. parole .....                                  | 18 |
| 2.4     | Les fonctions du langage .....                           | 19 |
| 2.5     | Approches envers la classification des textes .....      | 20 |
| 2.5.1   | Typologie des typologies pour classer les textes .....   | 20 |
| 2.5.2   | Du type de texte à la séquence .....                     | 21 |
| 2.5.3   | Approches textuellement extérieures et intérieures ..... | 22 |
| 2.5.4   | Les types de textes .....                                | 22 |
| 2.5.5   | Le style.....  | 23 |

|       |  |    |
|-------|--|----|
| 2.5.6 | Les styles fonctionnels .....                          | 23 |
| 2.5.7 | Le registre .....                                      | 24 |
| 2.6   | La traduction .....                                    | 25 |
| 2.7   | Études de corpus .....                                 | 27 |
| 2.7.1 | La linguistique de corpus .....                        | 27 |
| 2.7.2 | Le corpus .....  | 28 |
| 3     | Partie pratique .....                                  | 34 |
| 3.1   | Corpus .....   | 34 |
| 3.2   | Méthode .....  | 37 |
| 3.2.1 | Critères .....   | 41 |
| 3.2.2 | Analyse d'échantillons (extraits des sous-corpus)..... | 45 |
| 4     | Résultats .....  | 51 |
| 4.1   | La bible .....   | 51 |
| 4.1.1 | La prononciation .....                                 | 52 |
| 4.1.2 | L'orthographe .....                                    | 52 |
| 4.1.3 | La morphologie .....                                   | 52 |
| 4.1.4 | La fréquence.....                                      | 53 |
| 4.2   | Le parlement .....                                     | 54 |
| 4.2.1 | La prononciation .....                                 | 55 |
| 4.2.2 | L'orthographe .....                                    | 55 |
| 4.2.3 | La morphologie .....                                   | 55 |
| 4.2.4 | La fréquence.....                                      | 56 |
| 4.3   | Le roman .....   | 57 |
| 4.3.1 | La prononciation .....                                 | 57 |



|       |                             |    |
|-------|-----------------------------|----|
| 4.3.2 | L'orthographe .....         | 58 |
| 4.3.3 | La morphologie .....        | 58 |
| 4.3.4 | La fréquence.....           | 58 |
| 4.4   | Noms propres .....          | 59 |
| 5     | Conclusion.....             | 60 |
| 5.1   | Prononciation.....          | 62 |
| 5.2   | Orthographe .....           | 62 |
| 5.3   | Morphologie.....            | 63 |
| 5.4   | Fréquence.....              | 65 |
| 5.5   | Sens.....                   | 67 |
| 6     | Bibliographie.....          | 72 |
|       | Résumé diplomové práce..... | 76 |

# 1 Introduction

Au premier regard, l'anglais et le français semblent bien éloignés l'un de l'autre – surtout du point de vue des caractéristiques de la langue parlée, donc de la prononciation, dont les sons, l'accentuation et l'intonation sont différents.

Mais si l'on regarde de plus près, on s'aperçoit qu'il y a de nombreuses ressemblances assez fines, non seulement dans la grammaire mais aussi dans le domaine du vocabulaire, qui fait (avec ses aspects morphologiques) le sujet de ce mémoire de master.

Notamment dans le registre des termes abstraits et dans la terminologie scientifique et technologique, beaucoup de termes sont quasi identiques en anglais et en français. Cela s'explique par leurs nombreux contacts à travers des siècles.

Le présent travail continue et approfondit le thème de notre mémoire de licence intitulé *Les vrais et les faux amis en anglais et en français et leurs aspects lexicaux et morphologiques*, en ce qu'il porte toujours sur l'analyse des notions comme celle des vrais amis, des faux amis et des faux amis partiels. Au centre de l'attention sont cette fois les vrais amis.

Dans notre mémoire de licence, qui était conçu du point de vue de la *langue*, nous avons constaté que les vrais amis étaient plutôt rares. Traditionnellement, ce sont les mots longs et abstraits, qui forment le lexique intellectuel (et précisément, ils se trouvent dans les sphères scientifiques et techniques de la terminologie). Ils sont avantageux pour un Français apprenant l'anglais et vice versa, un Anglais apprenant le français, puisqu'ils ne présentent aucun risque par rapport au transfert négatif. Par contre, ils offrent la possibilité de profiter au maximum du transfert positif. Le champ lexical au travers de la catégorie des vrais amis comprend traditionnellement la religion, l'administration, l'art, le droit, l'architecture, etc.

Le but des pages suivantes est de présenter, comme le titre de ce mémoire de master l'indique déjà, les vrais amis en anglais et en français dans un ensemble authentique d'échantillons des textes. Particulièrement, il s'agit d'une étude sur corpus rassemblé et

sélectionné spécialement pour les buts de ce travail. Dans notre recherche, nous nous concentrons sur la représentation réelle des vrais amis dans les textes choisis. Il nous a intéressé, si les vrais amis font partie de l'usage d'aujourd'hui, s'ils sont employés aussi fréquemment dans les différents types de textes et s'ils sont facilement reconnaissables au premier regard. Au-delà, nous avons étudié lesquels sont les plus fréquents.

Puisque nous tenons à l'objectivité des résultats, il nous faut dans un certain sens enjamber les deux langues, et donc considérer les deux sens de la traduction avec autant d'attention: de l'anglais au français et vice versa, du français à l'anglais. Par rapport aux types des textes choisis, nous avons décidé de fonder notre choix sur des critères extérieurs aux textes.

Ce qu'il est donc important d'indiquer avant la rédaction-même, c'est que nous avons adopté une perspective singulière pour cette analyse. Plus précisément, nous allons nous concentrer sur la langue anglaise et la façon dont (autant que la mesure où) le français l'a influencé et enrichi – cela pour des raisons avant tout historiques: le sens de l'emprunt lexical original était presque exclusivement du français à l'anglais. (Ce n'est que très récemment que le processus s'est inversé avec l'anglais prenant de l'ampleur au niveau mondial.)

Cette fois, il ne s'agit donc plus de l'étude de la norme abstraite dans les deux langues, mais de l'accrochage de sa réalisation pratique dans les textes des traductions authentiques concrets.

Il nous paraît aussi important d'indiquer que ce mémoire de master porte sur le français et l'anglais *d'aujourd'hui*.

## 2 Partie théorique

### 2.1 L'Arrière-plan

#### 2.1.1 L'Arrière-plan historique

Si l'anglais et le français se distinguent par des différences considérables en termes de prononciation et d'accentuation, ils disposent d'une grande proximité au niveau grammatical, lexical et de type de langue (analytique). C'est surtout grâce à l'histoire qui a étroitement interconnecté leurs destins.

Au-delà de l'ancien fond (proto-)indo-européen (commun aux Celtes et aux Saxons), c'était a priori l'influence du latin qui a largement contribué à la création d'une base commune partagée par les deux langues.

La romanisation, durant jusqu'à la fin du V<sup>e</sup> siècle, a laissé peu de traces linguistiques. L'évangélisation, par contre, commençant au IV<sup>e</sup> siècle après J.-C., est considérée comme déterminante dans l'introduction et ensuite dans la propagation du latin – c'est à partir de cette période-là que les clercs écrivaient surtout en latin et cela a duré plusieurs siècles.

L'apport de ce latin de l'Église avait un rôle important dans l'évolution de la langue anglaise (germanique), encore qu'il soit souvent difficile d'établir si tel mot ou telle structure provient directement du latin ou de l'ancien français imposé par les Normands, suite à la victoire de Guillaume le Conquérant à Hastings en 1066.

L'invasion normande a donc eu pour conséquence que le français est devenu la langue « officielle » de la cour et du Royaume d'Angleterre pour plusieurs siècles pendant lesquels l'histoire de la France et celle de sa voisine d'outre-Manche sont restées intimement liées.

Au commencement du XIII<sup>e</sup> siècle, après que Jean de sans Terre avait perdu ses territoires continentaux, le français normand a décliné et l'influence du français central – parisien a été introduit en Angleterre. Ce qui, aujourd'hui, aboutit aux emprunts parallèles, puisque dans ce temps-là, il arrivait qu'un mot ait été repris en même temps du français normand et du français central, par ex. *catch* (du français normand, A) – *chase* (du français central, A)).

Au XVI<sup>e</sup> siècle, l'époque de la Renaissance, des auteurs grecs et latins furent redécouverts, ce qui a provoqué de nombreuses traductions d'ouvrages italiens et français vers l'anglais, ce qui a continué à enrichir l'anglais d'apports similaires (termes et tournures) à ceux qui orientaient l'évolution du français.

Ce n'est que pendant une époque beaucoup plus récente, que la notion du *snobisme* fut introduite – comprenant la francophilie autant que l'anglomanie – que la tradition *d'emprunts réciproques* s'est perpétuée. « Les mots anglais viennent (en français) en partie parce que les Français trouvent un certain plaisir snob à parler en anglais et en partie parce que certains mots ont été inventé par des anglophones pour nommer des choses que les Français auraient pu nommer eux-mêmes. » (Edwards, 2017)

Si l'influence du français sur l'anglais s'est exercée fortement dans ses domaines de compétences reconnues et culturelles, comme la gastronomie, la diplomatie, la politique, ou encore celui des considérations abstraites, l'anglais a laissé son empreinte dans les domaines sportifs et technologiques, les Français adoptant le vocable en même temps qu'ils adoptaient la technique.

Cette tendance à intégrer les termes anglais en français s'est renforcée principalement dans la seconde moitié du XX<sup>e</sup> siècle, notamment sous l'influence de la suprématie économique et technologique des États-Unis; l'anglais était alors perçu comme une *lingua franca* du monde actuel.

Dans la période récente, c'est en particulier l'Union européenne qui a remarquablement contribué à des emprunts réciproques ou à des créations de termes à partir des racines anglaises ou françaises.

Il n'est donc pas surprenant qu'après tant de siècles de contacts et d'échanges linguistiques que de nombreux mots aient une forme identique ou voisine dans ces deux langues. Mais cela ne veut pas dire qu'ils aient adopté exactement le même sens du mot. D'une part, à partir d'une origine commune, souvent latine ou grecque, ils ont pu évoluer différemment au cours des siècles. D'autre part, l'emprunteur a pu déformer, directement ou progressivement, le vocable initial, en le marquant soit de sa propre interprétation,

soit de son système linguistique (par ex. la resegmentation<sup>1</sup>). Ce pouvait même être le cas, où le terme déjà existait dans les deux langues, mais dans une des langues, le mot présentait une nuance (acception) de plus. Et précisément, cette acception a pu être adoptée par la deuxième langue, élargissant ainsi le sens du mot original. « Au XVIII<sup>e</sup> siècle, ils (les Français) ont pris de l'anglais une nuance du mot *responsability* qui n'était pas en France, la responsabilité d'un ministre devant le parlement, par exemple. » (Edwards, 2017)

Le thème de ce projet peut être perçu par deux approches, synchronique et diachronique. La littérature suggère que la majorité des emprunts venait du français en anglais, et précisément, pour les raisons historiques. Dans les pages suivants, nous proposons une perspective uniquement synchronique sur la problématique. Et particulièrement, parce qu'elle pourrait être pertinente pour l'usager d'aujourd'hui – soit de l'anglais ou du français, qui vise à acquérir la deuxième langue (celle, qui lui est étrangère), soit d'une toute autre langue, qui a l'intention d'acquérir les deux. Le présent travail pourrait non seulement rendre le processus d'acquisition plus facile, mais aussi le doter de l'aspect motivant.

### **2.1.2 Les langues en contact**

Du point de vue sociolinguistique, le contact entre deux (ou plusieurs) langues peut aboutir à une variété de cas. Parfois, seulement quelques mots sont empruntés; parfois, les nouvelles langues se forment (en tant que le résultat d'un transfert linguistique étendu ou d'un transfert linguistique entre de différents types des langues). Les résultats d'un tel contact diffèrent suivant plusieurs facteurs, y compris la longueur et l'intensité du contact parmi les groupes d'interlocuteurs; les types des relations sociales, économiques et politiques; des fonctions que la communication entre eux doit remplir; et le degré de la ressemblance entre les deux langues en contact.

La plupart des langues ont été influencée par le contact – à un moment ou l'autre – au cours de leur histoire, ce qui a aboutit aux degrés de transfert (des traits d'un à l'autre) différents.

---

<sup>1</sup> Par ex. le mot *ammunition* (A), introduit en anglais en résultat de l'erreur des soldats français qui ont pris le mot « la munition » (du latin *munitio*) comme si c'était « l'ammunition »

Dans le cas des langues géographiquement voisines, nous parlons de *l'adstrat*, l'influence réciproque de deux langues utilisées dans des régions voisines.

« L'**adstrat** est, par conséquent, le résultat d'une situation géographique ou sociale particulière. » (Martinet, 1969)

L'anglais a emprunté un grand nombre de mots vocabulaire du français, latin et grecque au cours de son développement. Un tel transfert ne revendique même pas que les interlocuteurs des langues différentes se parlent véritablement, puisqu'il peut être accompli par l'apprentissage de livres. Les enseignants transmettent le nouveau vocabulaire aux autres interlocuteurs via la littérature, les textes religieux, les dictionnaires, à présent l'internet, etc.

Il semble que les situations de contact langagier sont généralement soumises à deux tendances, qui sont (en plus) souvent en conflit. D'un côté, il s'agit d'une nécessité d'achever une effectivité de la communication adéquate pour le but d'interaction (le dynamisme d'accommodation). De l'autre côté, c'est la volonté de maintenir la conscience distincte d'identité de groupe (la loyauté envers le groupe). La première tendance stimule la convergence et le compromis entre les langues, la deuxième encourage la divergence ou la conservation des frontières langagières.

## **2.2 Définition des termes lexicologiques**

Entre les mots semblables du français et de l'anglais existent des relations sémantiques différentes; de *vrais amis* (mots de même sens et de même forme), de *faux amis* (mots d'apparence semblable mais de sens différent) et de *faux amis partiels*, qui, selon les contextes, peuvent avoir soit le même sens, soit un sens tout autre que leur homologue dans la deuxième langue.

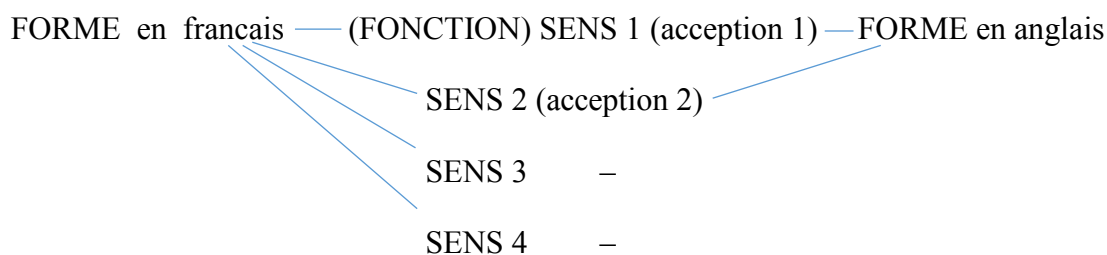
Il faut souligner le fait que les vrais amis, les faux amis et les faux amis partiels n'existent pas a priori dans la langue (comme par exemple les synonymes ou les paronymes), il s'agit d'une relation entre les deux éléments des langues différentes : c'est la connaissance linguistique inégale du locuteur bilingue qui les produit.

Les vrais amis anglo-français peuvent (légèrement) différer l'un de l'autre par rapport à *l'extension* et *l'intension*. L'extension est un ensemble des objets que le mot désigne. L'intension, par contre, est une classe déterminée par ses caractéristiques.

De manière schématique, il est possible d'impliquer que si une langue a adopté la forme d'un mot (laquelle elle a pu modifier dans des différents degrés), ce ne signifie pas qu'elle a obligatoirement repris tous ses sens aussi. Elle a pu choisir un seul ou au contraire plusieurs de sens originaux du mot emprunté. Elle a même pu changer le(s) sens plus ou moins légèrement. En plus, les langues développent naturellement, et donc le sens original du mot dans la langue donneuse ne doit même plus exister. Au-delà, les nouveaux sens peuvent émerger aussi.

Complétons que l'anglais et le français ont de différentes variétés régionales et sociales, par ex. le français canadien, l'anglais australien, etc., qui peuvent, eux aussi, différer l'un de l'autre en ce qui concerne le nombre de sens d'un mot. Le présent travail peut ainsi montrer uniquement les correspondances dans l'usage du français et l'anglais moderne, et ceci sur la base des variétés, registres et des contextes concrets étudiés.

Voici un exemple d'un mot polysème, qui en français, dispose de 4 sens. Toutefois, l'anglais l'a repris avec seuls deux ces sens.





### 2.2.1 Les vrais amis

Les vrais amis ou autrement, selon Walter (2001) bons amis, sont des termes figurant dans les deux langues avec une graphie identique (ou très voisine), donc homographes, de même sens et de prononciation similaire. D'une manière générale, il s'agit du groupe le plus compréhensible et transparent; des synonymes internationaux/interlangagiers (par ex. possible (A) – possible (F)). Exceptionnellement, on peut trouver deux emprunts provenant d'une seule langue (par ex. *piano* (A, F) de l'italien).

Il est possible de diviser les termes dans cette catégorie en trois types: les vrais amis « sûrs », les vrais amis « proches » et les vrais amis à l'écrit qui deviennent des « traîtres à l'oral ». (Marcheteau, Dahan, 2009: 14-15)

Les vrais amis « sûrs » sont ceux, qui sont, comme l'indique leur dénomination même, presque identiques sous tous les aspects. Voici quelques exemples: *accident, agent, cigarette, distance, elephant, hotel, moustache, noble, patience, restaurant, transport, village, etc.*

Les vrais amis « proches » présentent un même sens, une prononciation quasiment identique, mais une orthographe légèrement modifiée dans les deux langues, par exemple: *abbreviation* (A) – *abréviation* (F), *city* (A) – *cit  * (F), *development* (A) – *d  veloppement* (F), *hospital* (A) – *h  pital* (F), *language* (A) – *langage* (F), *number* (A) – *nombre* (F), *to observe* (A) – *observer* (F), *victory* (A) – *victoire* (F).

Remarque: il peut s'agir aussi des mots anglais emprunt  s par le fran  ais, par exemple *bifteck* (F) – *beefsteak* (A), m  me si ces cas sont peu fr  quents.

Les vrais amis   crit qui deviennent « des traîtres    l'oral » sont des termes dont la signification m  connaissable est soumise    une diff  rence de l'accentuation et/ou prononciation. Ainsi: *archive* [a:kaiν] (A) – *archive* [aʁʃiv] (F), *lieutenant* [lef'ten  nt] (A) – *lieutenant* [lj  tn  ] (F), *rhythm* [rið  m] (A) – *rythme* [ritm] peuvent   tre consid  r  s comme une sous-cat  gorie de ce groupe des mots identiques (ou quasi identiques) en fran  ais, qui ne posent pas de probl  mes de compr  hension. Par contre, ils sont souvent mal prononc  s par les locuteurs fran  ais. Par exemple: *guardian*

[ˈgɑːdiən] (A) – *gardien* [gardjɛ̃] (F), *promis* [ˈprɒmɪs] (A) – *promesse* [promes] (F), *to examin* [ɪgˈzæmɪn] – *examiner* [ɛgzamine] (F).

Walter (2001) à ce propos, présente une classification tripartite: les bons (vrais) amis étrangers, les bons amis apparentés et les « aller et retour ».

Les bons amis étrangers ont leur origine dans des langues étrangères pour l'anglais aussi bien que pour le français. Par exemple: *anaconda*, *alligator*, *banjo*, *vodka*, *malaria*, *polka*, *calypso*, *kimono*, *ravioli*, *zen* etc.

Les bons amis apparentés (formant le groupe majoritaire) sont les mots « ayant le même étymon, la forme identique ayant été transmise soit par le français à l'anglais, soit à l'inverse. » (Walter, 2001: 101-112.) Par exemple: *adroit*, *discipline*, *sabotage* (F → A); *dealer*, *leader*, *jockey* (A → F).

Le groupe des « aller et retour » sont « des mots qui, venus de la France, ont été adoptés en anglais, et sont ultérieurement revenus en français, sous une forme altérée et avec un nouveau sens ». (Walter, 2001: 101-112.) Par exemple: *cash* (1. « allé » en anglais de l'ancien français *ca(i)sse*, 2. « retourné » de l'anglais en français comme *argent liquide*), *budget* (1. « allé » en anglais de l'ancien français *bougette*, 2. « retourné » de l'anglais en français comme *finances de l'état*), *interview* (1. « allé » en anglais de l'ancien français *enterveue*, 2. « retourné » de l'anglais en français comme *entretien public*).

### 2.2.2 Les faux amis

Le terme a été introduit en 1928 pour la première fois, et précisément par les linguistes Maxime Kœssler et Jules Derocquigny, dans le livre *Les Faux Amis ou les trahisons du vocabulaire anglais*. Les faux amis désignent en effet un groupe des faux amis absolus, complets ou totaux – ceux, dont la traduction littérale produit un contre-sens. Fondièrément, il est possible de les « appeler les vrais faux amis, dans la mesure où leur apparence formelle est proche, alors que leur signification en est franchement différente. Certains de ces termes ne doivent leurs apparentes ressemblances qu'à des rencontres fortuites, purs produits du hasard. » (Marcheteau, Dahan, 2009: 22)

Igot, le traducteur professionnel entre l'anglais et le français, propose sur son site une définition des faux amis encore plus élargie et précise: « les faux amis sont des

mots, des expressions, des structures, etc. du français qui ressemblent à l'anglais et qui semblent corrects, mais qui ne le sont pas, parce qu'un aspect particulier — le sens, l'emploi, la place dans la phrase, la construction ou autre chose encore — n'est pas le même en français qu'en anglais. » (Igot, [www.fauxamis.fr](http://www.fauxamis.fr)<sup>2</sup>)

Il s'agit donc de paires de mots « traîtres » dont la ressemblance lexicale (formelle) peut mystifier leur interprète. Voici deux exemples: *bride* (A, épouse) ≠ *bride* (F), *casket* (A, coffret) ≠ *casquette* (F), *to attend* (A, fréquenter) – *attendre* (F). Il s'agit en fait des homonymes inter-langagiers.

Cependant la plupart des faux amis sont des mots qui, à partir d'une origine commune – grecque, latine, française ou anglaise – ont évolué différemment dans les deux langues. Il est même bien possible qu'à un moment donné de leurs histoires, les deux mots aient eu un sens identique, mais aujourd'hui, ils se trouvent sémantiquement bien éloignés l'un de l'autre. Par exemple: *cave* (A, grotte) – *cave* (F).

Dans un certain nombre de cas la divergence sémantique s'explique par le fait que le vocabulaire anglais est resté (contrairement au vocabulaire français qui évoluait davantage) fidèle au sens du terme latin dont il est souvent dérivé (par ex. du latin « *persona* » : *person* (individu, A) – *personne* (F)).

Une sous-catégorie spéciale sur l'échelle entre les faux amis et les faux amis partiels est constituée par les faux anglicismes, autrement dit franglicismes ou aussi des pseudo-anglicismes: « les mots formés en totalité ou en partie d'éléments venant de la langue anglaise et non généralement employés tels quels dans celle-ci – c'est à dire, avec les différences d'emploi au niveau de la classe, de la forme ou du sens. » (Trescases, 1983 : 86-101) ou « les termes pris à l'anglais avec un autre sens. » (Humbley, 1974 : 55 ) Les mots empruntés donc reçoivent un nouveau référent dans la langue emprunteuse. Par exemple: *campsite* (A) – *camping* (F), *crisps* (A) – *chips* (F), *parking lot* (A) – *parking* (F), *make-over* (A) – *relooking* (F), *channel hopping* (A) – *zapping* (F), *suit* (A) – *smoking* (F).

---

<sup>2</sup> <http://fauxamis.fr> Le site consulté le 2 février 2017.

### 2.2.2.1 Les faux amis dans le cadre sémantique

Certains cas de faux amis sont en fait des paronymes ou des homonymes.

#### **Polysémie**

La notion de la polysémie marque des significations ou fonctions multiples. C'est un spectre des sens d'un seul lexème, qui présentent une connexion l'un avec l'autre. Il s'agit d'un phénomène assez fréquent pour les langues analytiques. (voir Čermák, 2001 : 190-192)

#### **Homonymie**

L'homonymie est en fait un accord formel et fortuit entre deux ou plusieurs lexèmes – des opposés dans le cadre d'une langue ou provenant de deux langues différentes<sup>3</sup> (dont l'origine n'est pas commune), tandis que leur signification est différente. L'homonymie, elle aussi, est assez typique pour les langues analytiques.

#### **Paronymie**

La paronymie est définie comme une relation entre des mots de forme non pas identique, mais similaire ou voisine, qui peuvent être aisément mêlés ou échangés, et donc présentent une source des difficultés. Ce phénomène peut apparaître soit dans le cadre d'une seule langue (par ex. *affect* – *effect* (A), *amener* – *emmener* (F), *attention* – *intention* (F)), soit à travers des frontières entre les deux langues différentes (par ex. *bride* (A, épouse) ≠ *bride* (F)).

### 2.2.3 Les faux amis partiels

Le terme des faux amis partiels ou, selon Walter (2001), les bons amis partiels désigne les paires de mots dans les deux langues dont la forme se ressemble, mais dont le recouvrement sémantique<sup>4</sup> n'est que partiel (sur l'échelle sémantique, ils constituent

---

<sup>3</sup> Le homonymes donc peuvent être trouvés soit dans l'intérieur d'une langue ou au travers deux langues différentes.

<sup>4</sup> Le recouvrement sémantique peut être défini comme (au moins) un sème partagé entre deux mots.

un degré intermédiaire entre les vrais et les faux amis). Il s'agit donc d'homographes internationaux/inter-langagiers. Plus particulièrement, les faux amis partiels se caractérisent par au moins un sens partagé avec son homologue et au moins un sens différent, que le terme de la deuxième langue ne connaît pas. Il s'agit donc soit de mots anglais qui ont le même sens que leur correspondant français et au-delà, ils ont encore des sens supplémentaires en anglais. Soit il s'agit des mots dont la définition est plus restreinte que celle de leurs opposés français. Ce qui constitue une source d'erreurs particulièrement fréquentes pour les interlocuteurs non natifs, qui, par erreur, en employant ou interprétant les faux amis partiels en accordance avec le(s) sens dont ils disposent dans leur langue maternelle, le peuvent faire facilement de manière déplacée dans la deuxième langue. Par exemple: *figure* (silhouette, chiffre, A) – *figure* (silhouette, ligne, F), *distraction* (divertissement, confusion, trouble de l'esprit, folie, A) – *distraction* (divertissement, F). (Ayasch, Hillion, 2006)

Tous les termes (vrais amis, faux amis et faux amis partiels) au-dessus sont de caractère essentiel pour ce travail.

Il paraît important de mentionner que, pour les buts de cette étude lexicale, on a adopté le point de vue de *parole* (Saussure, 1997), la réalisation concrète de la *langue*.

C'est pour cette raison que ce chapitre sera divisé en trois parties. Premièrement, nous éclaircirons le lexique et la théorie du signe linguistique; ensuite, nous présenterons les fonctions du langage et les typologies des textes et finalement, nous nous concentrerons sur la linguistique de corpus moderne.

#### 2.2.4 Le lexique

Dans toutes les langues, le lexique est un ensemble des mots dont une langue donnée dispose, autrement dit le système de mots qui existent dans la langue concrète. Le présent projet se concentre uniquement sur les mots lexicaux, en tant que des porteurs du sens par excellence. En évoluant sans arrêt, le lexique appartient aux éléments de la langue les moins stables: il s'approprie constamment et entièrement aux besoins actuels du locuteur (il faut un *signifiant* pour représenter un *signifié* (Saussure, voire au-dessous) ; si le signifiant pour un signifié nouvellement apparu manque dans le lexique d'une langue, une de possibilités est de l'emprunter d'une autre langue, qui l'a déjà inventé).

Plus particulièrement, le lexique est un ensemble des phénomènes interconnectés qui s'influencent. Le lexique montre un certain manque de l'équilibre. Sous forme d'un modèle, il est constitué d'un centre et d'une périphérie. (Daneš, 1966) Le centre représente le gros du vocabulaire; ce qui est déjà bien intégré, stable, typique, plus fréquent et neutre (par rapport au style), cependant que la transition et la périphérie comprennent les phénomènes exactement opposés. Somme toute, elles comprennent des couches lexicales comme l'élevé (par ex. les poétismes), le neuf (les néologismes) ou l'obsolète (les historicismes).

Un échange constant prend place entre les deux couches ou zones de l'usage, selon les revendications de l'usage et de la réalité actuels. Par exemple, au moment où un mot devient archaïque, il est déplacé du centre à la périphérie, et à l'opposé, si un mot archaïque acquiert un sens nouveau, relevant à l'usage actuel, il revient au centre du lexique.

De plus, travaillant dans le système, il y a deux forces opposées: une menant vers l'intégration et l'uniformité, l'autre s'orientant vers la diversification et la distinction fine du sens.

Très important pour les besoins de ce mémoire de master est le processus de l'enrichissement du lexique qui se passe à travers des procédés diverses : la dérivation (affixale et non affixale/ impropre ; préfixale et suffixale), la composition (populaire et savante), l'abréviation (la troncation, les mots valises ou la siglaison), les emprunts, les

néologismes, la formation récursive et la dérivation inverse/ régressive. Pour la perspective des homologues entre l'anglais et le français sont pertinents les procédés suivants:

### **La dérivation**

La dérivation affixale consiste à former de nouveaux mots à partir de mots déjà existant par l'adjonction d'un affixe: un préfixe, un suffixe ou plus rarement un infixe, à une base lexicale donnée (un radical). Un préfixe se situe avant le radical du mot tandis qu'un suffixe se situe après le radical du mot.

### **Les emprunts**

Considéré comme un des phénomènes indispensables pour l'enrichissement lexical d'une langue donnée, l'emprunt linguistique est défini comme un transfert d'éléments linguistiques d'une langue à une autre. « Acte par lequel une langue accueille un élément d'une autre langue ; élément (mot, tour) ainsi incorporé. » (Petit Robert, édition 2008). Un emprunt est donc un mot ou une expression qu'une communauté emprunte à une autre langue, sans le traduire, mais en l'adaptant généralement aux règles morphosyntaxiques, phonétiques et prosodiques de sa propre langue.

Il faut souligner qu'au centre de l'attention du présent travail se trouvent aussi les emprunts, qui ont été traditionnellement repris tels quels, comme des unités incluant soit la dérivation, soit des éléments composés.

### **Les calques**

Les calques sont puis des cas spéciaux des emprunts, où la langue donnée emprunte un mot ou une expression d'une autre, mais ensuite traduit tous les éléments mot à mot.

### **Les néologismes**

Un néologisme est résultat de la création d'un mot nouveau, ou de l'acte de doter un mot existant d'un sens nouveau.

## **2.3 La conception saussurienne**

### **2.3.1 La sémiologie et le caractère social de la langue**

Dans l'effort d'éclaircir la position de la linguistique et sa relation envers les autres branches scientifiques, Saussure a introduit sa théorie de *la sémiologie*, une discipline qui étudie la vie des signes travaillants à l'intérieur de la société humaine. Les systèmes dont la société dispose sont selon lui nombreux: l'écriture, les salutations, le code de la route, etc. Mais le système des signes essentiel pour la société est justement la langue. La linguistique fut désormais classifiée comme une discipline sociale, l'élevant ainsi au niveau des sciences naturelles.

La théorie saussurienne de la sémiologie – et de la linguistique en tant que sa partie – est en rapport proche avec ses théories sur le caractère systémique de la langue et du signe linguistique. Ce qui est évident par ex. du fait qu'il allègue que dans la langue (comme dans tout autre système sémiotique), les signes individuels sont caractérisés a priori par ce qui les rend différents des autres signes. Ce n'est que cette différence qui produit la valeur de chaque signe.

A présent, on développe et, dans une certaine mesure, on change la conception de la sémiologie et son rapport envers les autres disciplines, autant que le nom de cette théorie des signes, peu à peu plus fréquemment désignée comme *la sémiotique*.



### 2.3.2 La théorie du signe linguistique<sup>5</sup>

La théorie du signe linguistique, qui est devenue indispensable pour la linguistique moderne, occupe une place éminente dans le *Cours de la linguistique générale*, 1916 . Bien qu'elle exige une prestation de la pensée abstraite considérable, elle a apporté un grand nombre d'idées nouvelles et, avant tout, elle a établi des principes solides de la sémantique moderne.

Saussure comprend le signe linguistique comme un fusionnement de deux composantes, qu'il nomme « le concept » et « l'image acoustique ». Contrairement à la conception idéaliste, Saussure ne croit pas qu'il s'agit d'une unité d'une chose et d'un nom, mais d'une union des deux composantes de la nature psychique, connectées dans le cerveau par l'association.

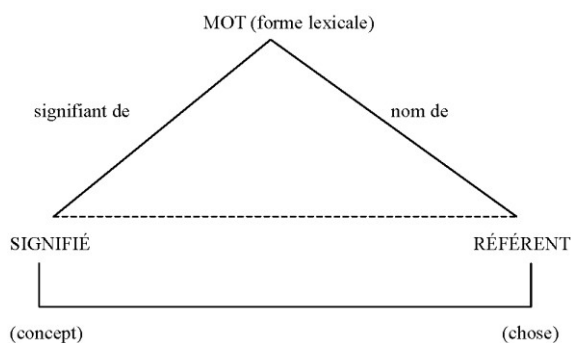
Par ex., « l'image acoustique » n'est pas interprétée comme un son purement physique, mais comme son empreinte ou sa piste dans notre cerveau. Précisément, ceci peut être affirmé par le fait qu'on est capable de « parler en pensée », sans que notre appareil phonatoire bouge et sans qu'on entende le moindre bruit.

« Le concept » n'existait pas au-paravant, cependant, les deux éléments du signe se formaient en même temps et sont indissolublement accrochés.

Vu que le terme « signe » pouvait être aisément mal interprété comme l'équivalent de « l'image acoustique », qui, en fait, désigne des terme « concept »,

---

<sup>5</sup> Il faut mentionner qu'à présent, le linguistes préfèrent le modèle du triangle sémiotique par Ogden-Richards, qui intègre aussi la réalité extra-linguistique dans la conception du signe linguistique. (voir par ex. Čermák 2001: 24)



Saussure a introduit les termes nouveaux, qui éliminent les risques de mauvaise compréhension. Et particulièrement, pour « l'image acoustique » il a créé l'équivalent « signifiant », et à la place du « concept », il a introduit « le signifié ». Ces deux composantes ensemble puis produisent le signe linguistique.

Il faut souligner que Saussure comprend cette relation en tant qu'une unité de deux opposés. Même s'il s'agit de deux éléments opposés, il est impossible d'abstraire l'une de l'autre.

Au-delà, Saussure a défini les trois attributs de base du signe linguistique : son caractère arbitraire, linéaire et sa discontinuité.

Le caractère arbitraire touche à la relation entre les deux composants. Il n'existe aucune relation intérieure entre *le signifié* et *le signifiant*, ce qui est affirmé par le fait que des différentes langues ont des différents signifiants (par ex. horse (A), cheval (F)) pour le même signifié.

La linéarité du signe linguistique indique que la communication se déroule dans le temps. Par conséquent, il est nécessaire de ranger tous les signes inclus successivement, l'un après l'autre.

Finalement, la discontinuité (ou la différence) exprime le fait que le matériel sonore de la langue, autant que la somme du contenu extralinguistique, sont amorphes. Ce n'est qu'au moment où un nombre des sons concrets d'une langue concrète commencent à s'enchaîner, que le signe linguistique se produit.

Après avoir introduit les notions de signifié et de signifiant, Saussure présente *la valeur* comme un élément de la signification du mot. Plus précisément, la signification est la relation entre le signifié et le signifiant d'un mot tandis que la valeur est la mise en rapport des mots entre eux. La valeur est donc un fait de langue (ou le système abstrait), tandis que la signification est un fait de parole (l'utilisation concrète ou la réalisation de la langue).

### 2.3.3 Diachronie vs. synchronie

Saussure dit qu'on peut adopter deux points de vue différentes et complémentaires en linguistique : synchronique et diachronique. L'approche synchronique étudie la langue d'une période donnée de son histoire. L'approche diachronique par contre étudie la langue dans son évolution au cours de son histoire.

«...nous préférons parler de linguistique *synchronique* et de linguistique *diachronique*. Est *synchronique* tout ce qui se rapporte à l'aspect statique de notre science, *diachronique* tout ce qui a trait aux évolutions. De même synchronie et diachronie désigneront respectivement un état de langue et une phase d'évolution. » (Saussure, 1916: 117) Ajoutons que le mot synchronique souvent désigne l'état de la langue actuel ou récent.

### 2.3.4 Langue vs. parole

Saussure fait aussi une distinction entre *la langue* et *la parole*.

*La langue* est un système de communication conventionnel. Par « système », on comprend que ce n'est pas seulement une collection d'éléments, mais bien un ensemble structuré, composé d'éléments et de règles permettant de décrire un comportement régulier. La langue est *acquise*.

*La parole*, quant à elle, puis consiste en l'utilisation concrète de la langue. La parole est en fait le résultat de l'utilisation de la langue.

Selon Saussure, la langue est le résultat d'une convention sociale transmise par la société à l'individu et sur laquelle ce dernier n'a qu'un rôle accessoire. Par opposition, la parole est l'utilisation personnelle ou la réalisation concrète de la langue (y compris toutes les variantes personnelles possibles: style, rythme, syntaxe, prononciation, etc.).

Il paraît inutile de constater que chaque mot est en fait un signe, ce qui le rend essentiel pour la présente étude. Vu qu'on vise à une étude concernant l'emploi des mots (signes), il faut se rendre compte du fait que la langue remplit des nombreuses fonctions.

## 2.4 Les fonctions du langage

Le linguiste R. Jakobson a introduit une classification de six fonctions du langage: la fonction référentielle (ou dénotative), expressive (ou émotive), conative, métalinguistique, poétique et phatique.

La fonction référentielle constitue la fonction primordiale du langage, comme elle a pour but de décrire le monde (elle est donc centrée sur le référent).

La fonction expressive communique les sentiments de l'émetteur.

La fonction conative du langage est centrée sur le destinataire, et vise à provoquer un certain comportement chez lui.

La fonction métalinguistique sert à parler (en utilisant le langage) du langage-même.

La fonction poétique est centrée sur le côté esthétique du langage.

La fonction phatique cherche soit à établir, soit à maintenir le contact. (Jakobson, 1963 : 209-248)

Avant que nous choissions des échantillons des textes convenables pour ce travail, le choix devait être appuyé sur une classification des textes adéquate. Il existe des nombreuses approches et conceptions proposées par des différentes écoles linguistiques.

Tout d'abord, il est possible de constater que chaque texte écrit reflète un projet. Au-delà, on identifie de six à huit types de texte dans la littérature linguistique: le texte narratif, descriptif, argumentatif, explicatif (ou didactique), injonctif (ou prescriptif), expressif, rhétorique et dialogal.

Le texte narratif a pour but de raconter des événements et des histoires (par ex. le roman).

Le texte descriptif vise à décrire des objets, des lieux et des personnages (par ex. le texte documentaire).

Le texte argumentatif est conçu pour convaincre, persuader et critiquer (par ex. le pamphlet).

Le texte explicatif explique et instruit (par ex. le manuel scolaire).

Le texte injonctif veut conseiller ou ordonner (par ex. la recette).

Le texte expressif a pour but d'exprimer les émotions et les sentiments (par ex. la lettre d'amour).

Le texte rhétorique crée un effet esthétique (par ex. le poème).

Le texte dialogal rapporte les propos.

## **2.5 Approches envers la classification des textes**

### **2.5.1 Typologie des typologies pour classer les textes**

A. Petitjean (1989) propose une typologie des typologies pour classer les textes, qui comprend quatre grandes familles :

1) Les typologies à base *énonciative*, inspirées des travaux de Benveniste (1966, 1970), qui se fondent sur l'analyse des rapports entre le locuteur, l'interlocuteur et l'ancrage spatio-temporel. Au centre de l'attention se trouvent les opérations linguistiques et les marqueurs par lesquels la situation énonciative s'inscrit dans l'énoncé : pronoms personnels, déictiques, système des temps, etc.

2) Les typologies à base *communicative*, ou *fonctionnelle*. Dans ce cas, on se réfère à Jakobson (1963) qui distingue six fonctions principales : la fonction référentielle, la fonction conative, la fonction phatique, la fonction métalinguistique, la fonction expressive et enfin la fonction poétique (voir section 2.4).

3) Les typologies *situationnelles* ou *institutionnelles* sont celles dont l'inspiration est essentiellement sociologique, et qui distinguent le discours publicitaire, le discours religieux, etc.

4) Les typologies fondées sur les *structures cognitives* mises en jeu. Celle de Werlich (1975) distingue cinq grands types de textes : le descriptif, le narratif, l'expositif, l'argumentatif, l'instructif (voir section 2.5.4).

### **2.5.2 Du type de texte à la séquence**

Suivant la typologie des typologies, à la fin des années 1980 le linguiste Adam établit sa conception sur le fait que tout texte est fondamentalement hétérogène. Il a défini le texte comme un lieu où « apparaissent immédiatement une hétérogénéité et une complexité qui semblent constitutives des discours en langue naturelle » (Adam, 1987 : 51). Il introduit davantage la notion de *séquence* pour « ainsi désigner une segmentation qui délimite des espaces textuels hétérogènes faisant partie d'un plus grand ensemble textuel, hétérogène, mais se donnant à lire comme homogène » (Benali, 2012). Il distingue entre la notion de la séquence *dominante* et de la séquence *dominée*, son hyponyme. Cela lui permet de ne pas concevoir la structure du texte comme une suite linéaire de séquences, et de mettre en évidence les phénomènes d'hiérarchisation globale du texte. Selon Adam (1992), les types textuels traduisent des procédures cognitives, des modes d'interpréter la réalité et de donner à cette vision une structure typifiée qui agit en profondeur sur le texte. Ces types de textes se reconnaissent par une dominante dans leur organisation séquentielle.

Selon ce même auteur, il existe cinq prototypes de séquences textuelles : narrative ou récit ; descriptive ou description ; argumentative ou argumentation ; explicative ou explication ; dialogique ou dialogue. « Cette classification est basée sur le raisonnement suivant : la séquence se définit comme l'un des plans d'organisation de la textualité. Une séquence, et plus encore un texte complet, n'offre jamais qu'une actualisation plus ou moins approchée du modèle prototypique de référence. Cela n'empêche nullement sa caractérisation mais au contraire, elle la rend plus exacte et plus fine. La séquence est une unité textuelle qui peut être définie comme une structure, c'est-à-dire comme un réseau relationnel hiérarchique, autrement dit une entité décomposable en parties reliées à la fois entre elles mais aussi à tout dont elles font

partie ; ou une entité relativement autonome, dotée de sa propre organisation interne et reliée également à l'ensemble plus vaste dont elle fait partie. » (Benali, 2012)

### **2.5.3 Approches textuellement extérieures et intérieures**

Selon Esser (2009), la distinction à faire est celle entre les approches textuellement extérieures et intérieures.

Les approches textuellement extérieures sont fondées sur les distinctions qui, par leur nature, ne sont pas intrinsèquement linguistiques, mais qui peuvent être constatées par ex. par les laïcs ou les bibliothécaires.

Les approches textuellement intérieures peuvent être soit indépendantes de leur médium, soit dépendantes de lui. Quant à la leur structure lexico-grammaticale, des textes peuvent être étudiés sur la base des approches sémantiques et formelles. On distingue deux tendances de recherche dans le cadre de ceux dernières. D'un côté, il y a l'étude des éléments démonstratifs et cohésifs. De l'autre côté, il y a l'étude dans le domaine de co-présence des éléments linguistiques qui sont étudiés avec des méthodes de la linguistique de corpus.

Finalement, il y a des approches textuellement intérieures, dépendantes de leur médium, où la présentation graphique ou orale constitue un facteur décisif pour l'attribution d'un type de texte (par. ex. faire-part de décès ou une chanson pop).

### **2.5.4 Les types de textes**

« Tout texte a plusieurs dimensions : il exprime des relations ou situations sociales, il a un contexte, des contenus, des buts, des éléments formels. Il a aussi une organisation et une cohérence. » (Benali, 2012)

Werlich (cité dans Esser, 2009) a basé sa typologie textuelle sur des *phrases types*. Voici sa définition du type de texte :

« Une phrase type est une norme idéalisée de la structure du texte concrète, qui sert comme une grille des règles et des éléments disponibles à l'encodeur (l'émetteur) pour répondre linguistiquement aux aspects spécifiques de son expérience. » (Werlich, 1976 : 39)

Werlich a davantage précisé que la structure profonde d'un texte, le « type », devrait être mise en relation avec les facteurs contextuels de la situation de communication et les procédés cognitifs: la description est liée à la perception de l'espace, la narration est associée à la perception dans le temps, l'exposition entre en correspondance avec la compréhension, l'argumentation est organisée autour du jugement et finalement, l'instruction concerne la planification.

### **2.5.5 Le style**

Le style linguistique est un mode de sélection consciente et d'organisation des moyens linguistiques, qui se met en valeur pendant la genèse du texte; en résultant de la communication, il se manifeste comme un principe d'organisation des éléments linguistiques, qui, des éléments individuels, constitue une unité convenante à l'intention de l'auteur. Dans le texte, il existe en correspondance avec le contenu, mais en analysant le style, il est possible de ne prendre en considération que l'expression.

« Le style de langue est un système des moyens sémantiquement clos, expressivement limité et rationnellement organisé, qui correspond à un genre 'littéraire', à une sphère de l'activité sociale (par exemple le style officiel et des affaires publiques, style administratif, style télégraphique, etc.) à une situation sociale (par exemple, style solennel, style d'extrême politesse), à un type de rapports langagiers entre les membres différents ou entre les couches de la société différentes. [Les tâches de la stylistique consistent à] dégager les principaux styles de la langue 'littéraire' et à déterminer leur hiérarchie, leur poids sémantique et fonctionnel, leurs relations, leur interaction et leurs sphères d'emploi. » (Bally, 1921)

La théorie du texte ensuite comprend le texte comme une unité définie non seulement au niveau de l'expression, mais aussi par son intention, thématique, sa situation d'origine et sa perception.

### **2.5.6 Les styles fonctionnels**

La notion de style fonctionnel a été introduite dans les *Thèses du Cercle linguistique de Prague* en 1929.

Les origines des styles fonctionnels et langues ont été établis au début des années trente dans le travail de B. Havránek *Úkoly spisovného jazyka a jeho kultura*.



Sur la base des fonctions primordiales de la langue littéraire, Havránek a introduit les langages fonctionnels fondamentaux: familier, de travail (matériel), scientifique et poétique. Et plus particulièrement, il l'a fait selon le but et le mode d'expression.

### **2.5.7 Le registre**

M.A.K. Halliday, dans la publication *Language, context, and text: aspects of a language in a social-semiotic perspective*, 1989, définit « le contexte de situation » en utilisant le cadre des termes suivants: le champ du discours (« field »), la nature du discours (« tenor ») et le mode du discours (« mode »).

Il propose davantage la notion de « la fonction du langage », les composants fonctionnels du système sémantique du langage : la fonction d' idées (divisée en fonction logique et fonction d'expérience), interpersonnelle et textuelle.

Finalement, il soutient que les deux notions sont interconnectées dans le sens qu'il y a une corrélation entre les catégories de situation et celles du système sémantique. De manière générale, le champ est reflété dans les significations d'expérience, la nature dans les relations interpersonnelles et le mode dans les significations textuelles. Ou vice-versa, les significations d'expérience sont activées par les traits du champ, les significations interpersonnelles par ceux de la nature et les significations textuelles par les traits du mode.

Le registre selon Halliday est un concept sémantique. Il est possible de le définir comme une configuration des significations typiquement associées avec une configuration de situation particulière du champ, de la nature et du mode. Cependant, parce qu'il s'agit d'une configuration des significations, un registre doit aussi inclure des expressions, des traits lexico-grammaticaux et phonologiques, qui d'habitude accompagnent ou réalisent ces significations. Le registre est donc une variante suivant son application.

On distingue les registres fermés et (plus) ouverts. Le registre fermé dispose d'un nombre totale et fixe des messages disponibles, comme par ex. la Langage International de l'Air. Dans ce type de registre, il n'y a pas de place pour l'individualité ni pour la créativité.

Le registre (plus) ouvert n'est pas prescrit par un livre, comme par ex. une conversation spontanée, mais il faut aussi constater que même les registres les plus ouverts ne sont jamais entièrement ouverts; il n'existe pas une situation qui n'est pas, dans une certaine mesure, prescrite pour nous ou associée avec une application du langage typique.

Selon Esser (2009), les registres peuvent être réduits en un seul critère, par ex. le français du sport, l'anglais parlé etc.

## **2.6 La traduction**

Les traductions surtout de la belle littérature étaient traditionnellement soumises à l'intérêt plutôt littérairement esthétique. Ce n'est qu'à la deuxième moitié du XX<sup>e</sup> siècle que l'approche linguistique a été introduite. Au début, les deux chemins existaient à parallèlement, et à l'occasion, ils se rivalisaient. Finalement, leurs statuts ont atteint l'équivalence sous forme d'une méthode intégrée qui prend en considération a priori l'aspect pragmatique.

Depuis les années soixante du XX<sup>e</sup> siècle, de nombreux travaux ont été écrits par rapport à l'approche linguistique envers la théorie de la traduction. La question d'équivalence était considérée en tant que le problème cardinal de la traduction. On soulignait la possibilité de transmettre toutes les informations du texte de la langue de départ dans la langue cible malgré la différence entre les deux systèmes grammaticaux des deux langues.

Aujourd'hui, l'approche fonctionnelle est considérée comme le principe fondamental de la traduction. Ce qui signifie qu'il n'est pas important si l'on emploie des moyens du langage identiques ou différents. Ce qui compte, c'est qu'ils réalisent la même fonction. Plus particulièrement, ceci dans tous les domaines – pas seulement dans le domaine dénotatif et référentiel, mais aussi dans le domaine expressif, associatif et pragmatique.

Le composant sémantique (du contenu) constitue l'essentiel du texte. En étant exprimé par les éléments lexicaux, le système grammatical instaure les relations entre eux. A côté de l'information dénotative (centrée sur la situation réelle et matérielle),

c'est l'information connotative (réalisée à travers la coloration stylistique et expressive de l'expression langagière) qui fait une partie indissociable du texte.

La traduction, pour être adéquate et réussie, doit remplir un certain nombre des critères. Il faut tenir compte du fait que le récepteur du texte résultant, en langue cible, perçoit le produit dans toute sa complexité, comme un ensemble (un résultat qui se réalise à travers des décisions individuelles). Au cas où la traduction est bonne, il n'est pas conscient de toutes les étapes, décisions et dilemmes qui font partie du processus de la traduction. Une traduction réussie ne devrait pas être perçue comme une traduction, mais comme une oeuvre originale, conçue en la langue cible. Une traduction de qualité devrait remplir les trois critères suivants:

- a) l'expression langagière produit un effet tout à fait naturel et organique
- b) le texte final dans la langue cible communique le même sens et produit un effet identique chez le lecteur à l'original
- c) l'expression langagière dans la langue cible garde le dynamisme du modèle.

Selon Knittlová (2010), les critères ci-dessus mènent à la transposition réussie dans la langue cible, qui, en même temps, ne la déforme pas selon la langue de départ. D'autant plus, il est nécessaire de garder l'équivalence sémantique ou d'information, stylistique et hypersyntactique (la transition réussie de l'acte illocutoire).

Jakobson (1987) distingue trois types de traduction:

- la traduction intralinguistique « ou la reformulation est une interprétation des signes verbaux au moyen d'autres signes de la même langue » (Jakobson, 1987) – il peut s'agir soit de la réiteration, soit de la citation, ou aussi de la paraphrase
- la traduction inter-sémiotique « ou transmutation est une interprétation des signes verbaux au moyen de systèmes de signes non-verbaux » (ibid) – il peut s'agir par ex. de l'interprétation linguistique des données sur les écrans des machines diverses, ou tout simplement des montres
- la traduction inter-linguistique « ou traduction propre est une interprétation des signes verbaux au moyen d'une quelconque autre langue » (ibid) (ce qui sera le plus important dans notre travail)

Knittlová (2010), introduit quatre sous-types de la traduction inter-linguistique:

- la traduction interlinéaire, considérée comme un exemple extrême de la traduction littérale, puisqu'elle garde exclusivement l'information linguistique
- la traduction littérale, aussi dénommée servile, qui traduit les éléments lexicaux sans respect de leur appartenance/placement contextuel
- la traduction libre, qui se trouve à l'opposition parfaite de la traduction interlinéaire, prouve un respect secondaire envers le texte original, comme elle ignore davantage le registre et le style ainsi que la couche connotative, et donc a une tendance à être fautive ou imprécise, éventuellement appauvrie
- la traduction communicative est liée étroitement avec l'aspect pragmatique de la discipline et il faut la respecter avant tout dans le domaine des formules conventionnelles (par ex. les salutations)

## 2.7 Études de corpus

### 2.7.1 La linguistique de corpus

La linguistique de corpus attire beaucoup d'attention dans les années récentes. Après avoir été considérée comme une approche marginale, utile avant tout dans la sphère de la linguistique anglaise (et particulièrement, dans les études de grammaire), la linguistique de corpus a graduellement commencé à élargir son emprise. Aujourd'hui, il s'agit d'un sujet de reconnaissance vivante, qui peut servir à répondre aux questions différentes de la recherche linguistique.

La linguistique de corpus devient aussi progressivement multilingue, avec beaucoup de langues et beaucoup de variantes de celle-ci sont étudiés à l'aide des données du corpus.

Avant d'élaborer la notion du corpus, nous aimerions l'encadrer pour les raisons d'orientation. Čermák (1995) distingue quatre types des *collections des textes* :

#### 1) Les archives

Une archive est un dépôt des textes libre sous la forme électronique, où les textes ne sont ni unifiés, ni élaborés.

## 2) Les textothèques

Une textothèque est définie comme une bibliothèque électronique des textes dans un format standard avec certaines revendications sur leur contenu etc., mais sans des restrictions rigoureuses de sélection.

## 3) Les corpus

Un corpus désigne une partie de la textothèque créée sur la base des principes de compilation de corpus pour des buts spéciaux.

## 4) Les sous-corpus

Un sous-corpus représente un sous-ensemble de corpus; soit le composant statique du corpus, soit une sélection dynamique du corpus. (Čermák, 1995 :18)

### 2.7.2 Le corpus

#### Définition

Étymologiquement, le terme « corpus » provient du latin, portant le sens de « corps, ensemble ou tout ». Dans la terminologie du droit par exemple, il désigne davantage « le recueil de pièces ou de documents concernant une même discipline. » (Le Nouveau Petit Robert 2008 : 548)

Plus précisément, le corpus linguistique peut être défini comme « un ensemble de textes établi selon un principe de documentation exhaustive, un critère thématique ou exemplaire en vue de leur étude linguistique » ([www.cnrtl.fr](http://www.cnrtl.fr)<sup>6</sup>), « une collection des textes considérés représentatifs d'une certaine langue, un dialecte, ou un autre sous-ensemble langagier, créé pour les besoins de l'analyse linguistique » (Francis, 1982: 7, cité in Tognini-Bonelli, 1996) ou aussi comme « un ensemble fini d'énoncés réels réuni en vue de l'étude d'un phénomène linguistique ». (Le Nouveau Petit Robert 2008 : 548)

L'étude de corpus vise donc à produire des descriptions et des analyses du langage basées sur les observations du langage *appliqué*, écrit autant que parlé, ce qui

---

<sup>6</sup> <http://www.cnrtl.fr/definition/corpus> Le site consulté le 14 février 2017.

classifie la linguistique de corpus parmi les disciplines de parole selon la division de Ferdinand de Saussure.

### **Corpus électroniques**

Stimulée par la venue des ordinateurs dans les années soixante, la linguistique de corpus incarne une nouvelle approche de l'investigation du langage. L'analyse du corpus linguistique sous la forme électronique avec les logiciels exécutifs de révélation et d'analyse a rendu possible de pénétrer dans les données du langage produit naturellement ou spontanément, ce qui n'était pas accessible auparavant.

« Un corpus électronique est un corpus qui est encodé de manière standardisée et homogène pour permettre des extractions non limitées à l'avance » (Sinclair, 1996 : 5).

La linguistique de corpus est ainsi étroitement liée avec l'analyse empirique des données linguistiques qui, dans ce cas, sont accessibles sous la forme électronique.

« En effet, la simple existence sur support électronique ne fait pas d'un ensemble de textes un corpus électronique. Il faut encore que ce document obéisse à des conventions de représentation et de codage répandues, voire faisant consensus, qui permettent la transmission et la réutilisation des données textuelles en cause. » (Guilhamou, Habert, Nazarenko, Salem : 1998)

La technologie informatique a donc rendu possible la découverte d'une nouvelle perspective sur la langue, et elle représente l'aspect central à la linguistique de corpus :

« Les ordinateurs rendent possible l'identification et analyse des formules complexes de du langage utilisé, permettant à la fois la déposition et l'analyse d'une base de données plus grande du langage naturel qu'il serait possible de le traiter manuellement. » (Biber, Conrad, Reppen, 1998 : 4)

Le traitement manuel ne peut pas concurrencer à l'efficacité, la véracité et l'exactitude projetées par l'ordinateur. Ce développement a ajouté une nouvelle dimension à la recherche empirique dans les études linguistiques.

## **Histoire**

Originellement, la linguistique de corpus a été introduite par les pays anglophones. Et même si aujourd'hui, il s'agit d'un phénomène assez répandu, les pays anglophones figurent toujours au centre de la recherche.

## **Types des corpus linguistiques**

Le corpus peut servir aux buts et recherches linguistiques différentes. C'est pourquoi il y a un caléidoscope des types des corpus – constitués en vue de répondre aux besoins particuliers de chaque entreprise dans le domaine de la recherche linguistique. Les corpus diffèrent par leur étendue, leur intervalle de temps, leur sélection des textes (respectivement, leur orientation) ou par les langues employées.

Nous proposons d'ouvrir ce chapitre par deux sous-divisions majeures : premièrement, on distingue les corpus diachroniques et synchroniques, et deuxièmement, on diffère les corpus d'une langue (traditionnellement de la langue nationale) et les corpus parallèles, de deux ou plusieurs langues.

### Les corpus diachroniques et synchroniques

Dans ce cas, la distinction est fondée sur l'aspect de temps, respectivement sur l'intervalle du temps à la base de laquelle le corpus était constitué.

#### ***A1 Le corpus diachronique***

Les corpus diachroniques, contrairement au corpus synchroniques, portent sur des longues périodes de temps. Et particulièrement, ils étudient ou rassemblent des textes de différentes époques (par exemple siècles) pour dévoiler l'évolution de la langue et le changement linguistique concret (le changement d'orthographe, le changement de sens, etc.).

#### ***A2 Le corpus synchronique***

Par contre le corpus synchronique se concentre sur une période de temps relativement courte, et ceci jusqu'à 50 ans. Dans la majorité des cas, il s'agit de corpus de la langue actuelle (lors de la naissance du corpus concret).

Ils existent des corpus synchroniques non-tranchés, dits généraux; ceux, qui manquent un but spécifique. Ce type ne prouve aucune préférence parmi des domaines

linguistiques, les styles ni les niveaux de langue. En plus, le trait principal de ce type-là incarne un rare équilibre entre les genres littéraires, les registres de textes choisis et les médias de transmission représentés.

Cependant la majorité des corpus est conçue pour les buts de recherche concrets; les corpus spécialisés.

La publication de Bowker et Pearson (2002 : 11-13) introduit plusieurs genres de corpus :

**B1 *Corpus de référence*** (« general reference corpus »)

« Un corpus de référence est conçu pour fournir une information en profondeur sur une langue. Il vise à être suffisamment étendu pour représenter toutes les variétés pertinentes du langage et son vocabulaire caractéristique, de manière à pouvoir servir de base à des grammaires, des dictionnaires et d'autres usuels fiables. » (Sinclair, 1996 : 10)

**B2 *Corpus spécialisé*** (« special purpose corpus »)

« Les corpus spécialisés sont limités à une situation de communication, ou à un domaine.» (Guilhamou, Habert, Nazarenko, Salem, 1998) Grâce à sa nature spécialisée (ou délimitée), ce type de corpus ne peut pas être utilisé pour créer des observations générales sur une langue.

Utilisés conjointement, un corpus de référence et un corpus spécialisé peuvent permettre d'identifier les différences entre une langue spécialisée et la langue générale.

**C1 *Corpus oral***

Un corpus oral est constitué de transcriptions de discours oraux (conversations, émissions, cours, etc.).

**C2 *Corpus écrit***

Un corpus écrit contient des textes destinés à être lus.

(Certains corpus, comme le *British National Corpus*, sont un mélange des deux.)

**D1 *Corpus monolingue ou national***

Le corpus monolingue ne contient que des textes de la langue choisie.



## **D2 *Corpus bilingue ou multilingue***

Contrairement à un corpus monolingue, un corpus multilingue rassemble des textes venant au moins de deux langues.

Les corpus multilingues peuvent être davantage divisés en corpus parallèles et en corpus comparables:

### a) Corpus parallèle

Un corpus parallèle est constitué des textes traduits dans une ou plusieurs langues.

### b) Corpus comparable

Un corpus comparable ne contient pas des traductions mais des textes écrits dans la langue de source qui ont tous le même sujet, la même fonction de communication, qui sont de même nature (manuel d'instruction, rapport technique, etc.) et qui ont des dates de publication proches.

## **E1 *Corpus ouvert ou de suivi* (Sinclair, 1996 : 4)**

Un corpus ouvert (« monitor corpus ») s'étend constamment. C'est le type généralement utilisé en lexicographie. « Il devient alors possible d'étudier l'évolution de certains phénomènes langagiers : néologismes, emplois privilégiés à un moment donné de certains suffixes ou préfixes, etc., un peu comme les éditions papier de certains dictionnaires d'usage (Le Petit Larousse, Le Petit Robert) servent de « sonde » sur le lexique et ses changements. » (Guilhamou, Habert, Nazarenko, Salem, 1998)

## **E2 *Corpus clos***

Un corpus clos reste tel quel l'est. Une fois qu'il est compilé aucun autre texte n'y est ajouté.

### **Les textes en tant que composants du corpus**

Un corpus qui est conçu en vue de constituer un échantillon représentatif d'un type délimité du langage, se concentre sur les spécimens des textes. Dans ce cas, le créateur du corpus doit prendre en considération plusieurs critères.

La première décision à prendre touche la quantité des textes : soit, les textes peuvent être considérés dans leur étendue complète, originale et peuvent donc former un « corpus des textes (complets) », soit, il peut s'agir seulement d'une sélection contingente, un extrait et donc un « corpus d'échantillons ». (Sinclair, 1996 : 4)

Le deuxième choix concerne l'unité textuelle, respectivement l'étendue des textes individuels, comme il est nécessaire d'établir une unité de base pour pouvoir procéder à l'analyse.

Le troisième critère porte ensuite sur l'échelle des variétés textuelles.

### 3 Partie pratique

#### 3.1 Corpus

Le corpus proposé dans ce mémoire de master, dont l'étendue est environ de 60 000 mots, consiste en trois sous-corpus. Ces sous-corpus recouvrent trois différents registres selon Halliday, et particulièrement, ils forment une échelle imaginaire du registre fermé au registre ouvert. Ce sont le registre politique (le plus fermé), le registre religieux (entre fermé et ouvert) et le registre de la prose littéraire<sup>7</sup> (le plus ouvert).

Nous avons choisi le concept du registre étant donné qu'il paraît le plus apte à capturer les fines différences sur l'échelle imaginaire de l'ouverture à la clôture par rapport à la variabilité. Le registre comprend trois variables: le champ du discours ou « field », la nature du discours ou « tenor » et le mode du discours ou « mode ». Par rapport à nos échantillons des textes, le champ du discours présente les différences les plus remarquables, étant donné qu'il s'agit des textes politique, religieux et artistique. Le mode de discours présente par contre les plus grandes similarités: dans tous les cas, les textes ont les traits de l'écriture et de l'oralité, mais leur chaîne est graphique (le roman présente la plus grande proportion de l'oralité dans cet égard, comme il contient le discours direct fréquent). En ce qui concerne la nature de discours, les textes proposés sont plutôt ou entièrement formels (le roman peut être classifié comme « normal » à cet égard et la bible comme rigide, voire très formelle). Ils incluent en premier lieu l'interaction entre les personnages ou les participants à la situation de communication et en deuxième lieu la relation du lecteur et l'auteur (voir éditeur).

Le registre est une classification basée sur les critères de caractère extérieur au texte selon Esser (2009). Nous avons opté pour le registre étant donné qu'il convient au domaine de contacts culturels entre les deux langues.

---

<sup>7</sup> Ici, on pourrait employer le terme du Cercle linguistique de Prague, le style fonctionnel poétique qui incarne la fonction littéraire ou poétique.

Au-delà, les sous-corpus ont été étudiés en tant que les traductions « en miroir », en anglais<sup>8</sup> et en français. Dans tous les cas il donc s'agit de la traduction interlinguistique. Ces textes ont été choisis aussi parce que nous avons supposé que le texte biblique serait le plus fidèle à son origine (présentant ainsi plutôt la traduction littérale), la traduction de roman serait la plus libre et la traduction politique aurait les traits de communicatifs et pragmatiques. Nous aspirions à la représentation équilibrée de ces approches vers la traduction.

Vue qu'il s'agit de la langue emprunteuse, une étendue « fondamentale » a été établie pour la version anglaise, et précisément, de 10 000 mots. (Ici il faut noter que, en raison de la traduction, l'étendue de la version française puis était légèrement différente).

Tous les textes dont le présent corpus est compilé n'ont plus que trente ans et sont disponibles en ligne<sup>9</sup>:

**la Bible** : version Darby (l'évangile de Matthieu)

<http://unbound.biola.edu/><sup>10</sup>

<http://unbound.biola.edu/index.cfm?method=searchResults.doSearch> – version anglaise

<http://unbound.biola.edu/index.cfm?method=searchResults.doSearch> – version française

**un ensemble des négociations au Parlement européen (1996 – 2011)**

<http://www.statmt.org/europarl/><sup>11</sup>

**le roman *L'insoutenable légèreté de l'être* par Kundera**

KUNDERA, Milan, KÉRÉL, François (traduction), *L'insoutenable légèreté de l'être*, France: Gallimard, 1990, 476 pages.

---

<sup>8</sup> Complétons que la variété de l'anglais proposée dans ce travail est *britannique* dans le cas des corpus parlementaire et biblique et *américaine* dans le cas du roman. Ce choix a été fait en vue de la représentation de deux dialectes équilibrés.

<sup>9</sup> Les traductions du roman *L'insoutenable légèreté de l'être* (Kundera) sont disponibles en format pdf.

<sup>10</sup> <http://unbound.biola.edu/> Le site consulté le 14 février 2017.

<sup>11</sup> <http://www.statmt.org/europarl/> Le site consulté le 10 février 2017.

KUNDERA, M., *The unbearable lightness of being*, New York City: Harper Perennial, 1999, 166 pages.

Il s'agit des textes compliqués, complexes, dans lesquels s'alternent des différentes séquences (Adam, 1987), par ex. dans le texte littéraire, la fonction principale est expressive mais on peut y trouver aussi les séquences narratives, contemplatives ou descriptives. Pareillement, la bible a a priori la fonction conative, mais il est possible d'y rencontrer des passages narratifs, descriptifs ou argumentatifs. En dernier lieu, le registre politique présente surtout les séquences argumentatives et référentielles.

Tous les textes présentent les traits de l'écriture et de l'oralité (tous comprennent le discours direct), c'est pourquoi on les désigne comme complexes. Ne présentant pas de lapsus, les textes sont apparemment édités ou modifiés; la norme finale est écrite et littéraire.

Le corpus proposé dans ce travail peut être décrit comme synchronique et récent (A2), spécialisé – rassemblé pour les buts de cette analyse (B2), écrit (C2), bilingue (D2) et clos (E2).

En plus, il s'agit d'un corpus composé, c'est-à-dire compilé des échantillons de textes cohérents, disponibles en ligne. Leur étendue correspond au critère d'une longueur uniforme établie pour les buts de ce travail, et précisément, aux 10 000 mots.

Les présents échantillons des textes sont choisis à la base des domaines sémantiques qui ont été décrits comme « influencés par les contacts interlangagiers » dans la littérature. Ces domaines comprennent entre autre justement la religion, l'administrative et l'art (et d'autres).

En dernier lieu, mentionnons que les textes ont été analysés à l'aide du programme AntConc 3.4 2014 qui est apte à travailler avec les diverses langues.

## 3.2 Méthode

La présente étude ne pose pas un regard diachronique, mais strictement synchronique, qui peut offrir une contribution à la didactique des langues.

Le phénomène est étudié du côté de *parole*, c'est à dire qu'il observe la présence réelle des éléments lexicaux dans les registres différents (la typologie textuelle étant basée sur des critères extérieurs – voir Esser, 2009).

Étant donné que c'est surtout la prononciation qui risque de torcher la compréhension, le présent corpus ne contient que des textes écrits (publiés, imprimés), où le trait traditionnellement prépondérant est la forme standard et la langue soutenue. Ce sont des textes qui se trouvent à la frontière de l'oralité et de l'écriture (ils contiennent les traits des deux, par ex. le discours direct), le médium est néanmoins toujours graphique. Ce fait nous mène à supposer qu'ils sont tous adaptés aux normes de l'écriture ; il s'agit des traductions bien rédigées et officielles. Tous les échantillons sont relativement récents, aucun parmi eux n'a pas été publié avant plus de trente ans ; ils sont tous disponibles en ligne.

Pour les buts de ce travail, nous avons choisi trois échantillons des textes recouvrant différents registres : le registre religieux (la Bible : Darby, l'évangile de Matthieu), le registre politique (un ensemble des négociations au Parlement européen entre les années 1996 – 2011) et le registre de prose artistique (le roman *L'insoutenable légèreté de l'être* par Kundera, 1990 et 1999). Tous les trois textes ont été élaborés en tant que des traductions « en miroir », anglaises et françaises, et ceci à l'aide du moyen pour les études sur corpus, le programme AntConc 3.4 2014, qui nous a paru facile à utiliser, et qui maîtrise les deux langues considérées.

Sur la base de la littérature, notre hypothèse était que l'anglais serait a priori dans le rôle de la langue emprunteuse. Au commencement, on a donc établi l'étendue des échantillons individuels, qui compte 10 000 mots pour la version anglaise du texte. Ici, il faut souligner qu'il s'agit d'une étendue dit « du départ », comme la longueur du texte français ne doit pas correspondre précisément à ce-nombre-là :

**La Bible :**

la traduction anglaise : 10 000 mots,

la traduction française : 9857 mots.

**Le Parlement :**

la traduction anglaise : 10 000 mots,

la traduction française : 10 547 mots.

**Le roman :**

la traduction anglaise : 10 000 mots,

la traduction française : 10 729 mots.<sup>12</sup>

A l'aide du programme AntConc, et précisément, sa fonction « la liste de mots », nous avons généré le répertoire de tous les mots présents/utilisés dans le texte. Puis, nous avons éliminé tous les mots grammaticaux. Ensuite, – cette fois uniquement des mots lexicaux – nous avons choisi les 100 premiers cas les plus fréquents de vrais amis dans les deux listes, ceci en gardant l'ordre proposé par le programme AntConc, et particulièrement, de la fréquence la plus haute jusqu'à la fréquence la plus basse. Et finalement, nous avons comparé les deux ensembles de résultats, et ceci premièrement par rapport à leur fréquence dans les deux langues (présentant la perspective statistique) et puis par rapport à leur traduction dans les passages des textes concrets (moyennant ainsi la perspective contextuelle). Au-delà, nous n'avons pris en considération que des cas qui présentaient la même classe de mot dans les deux traductions.

Dans l'étape suivante, nous nous sommes concentrés sur les mots, dont fréquence ou présence dans la liste anglaise est égale ou dépasse le chiffre 5 (donc nous n'avons pris en considération que des cas de l'abondance de valeur 5 ou plus). Cette fréquence a été choisie arbitrairement, pour mieux comprendre (et présenter) les tendances et pour qu'il s'agisse au minimum de 0,05% de l'échantillon (anglais) entier.

---

<sup>12</sup> Vu que les échantillons anglais comptent exactement 10 000 mots, nous avons opté pour les versions anglaises en tant que des versions « de départ ». Au-delà, il paraît approprié de partir de la langue emprunteuse pour mieux comprendre des tendances concernant les vrais amis.

Il est nécessaire de préciser que notre analyse porte uniquement sur les cas, dont il était possible de trouver l'homologue dans la liste « partenaire » (nous avons éliminé les cas où il n'était pas possible de les joindre pour les buts de ce travail), c'est-à-dire ceux, qui ont été réellement employés par le traducteur en tant que termes parfaitement opposés dans les deux langues. Et particulièrement, deux catégories de paires des vrais amis ont ainsi été générées comme pertinentes pour ce travail: la catégorie de fréquence *identique* (comprenant les vrais amis employés aussi fréquemment dans les deux traductions) et la catégorie de fréquence *proche* ou presque identique (comprenant les vrais amis employés avec une légère différence – de plus ou moins – 1 ou 2 fois).

Ensuite, nous avons séparé deux catégories, qui ont été négligées dans ce travail, et précisément, en vu qu'elles n'ont pas rempli soit le critère de la fréquence (égalant ou dépassant le chiffre 5 dans la liste anglaise), soit de la différence de la présence du mot dans le texte différent (par la valeur 2 au maximum). Pour ces catégories, il serait difficile voire impossible de juger les tendances valides, vu que les fréquences ont des valeurs si basses. Il s'en suit que ce domaine mériterait une étude particulière et spécialisée, qui dépasserait les buts de ce travail.

Les pages suivantes sont donc consacrés uniquement aux cas, dont fréquence/présence dans le texte égale ou dépasse le chiffre 5 dans la liste anglaise et en même temps, dont valeur de la présence dans les deux traductions varie jusqu'au chiffre 2. (Il a été constaté dans le contexte concret que parfois, les auteurs ont remplacé un mot lexical par un pronom, une ellipse grammaticale, etc., parfois, ils ont employé une construction différente, par ex. la voix passive, l'expression synonyme, etc. Leurs opposés pouvaient être soit plus fréquents, soit plus rares – ce qui était exceptionnel, introduisant seulement 4 cas dans le registre littéraire. Dans la plupart des cas, même les homologues français montraient des valeurs de 5 ou plus.

Pour des raisons d'orientation, les cas correspondant aux critères cités ci-dessus sont désormais désignés de l'étiquette « la sélection centrale ».

Tous les registres ont été soumis à la même analyse.

On a aussi trouvé des ensembles des éléments pairs (c'est-à-dire des vrais amis, qui avaient leur homologue dans la liste « partenaire »), dont la fréquence oscillait entre



les valeurs 1 et 4 dans la liste anglaise dans tous les trois registres. Nous attachons les listes de ces cas, même si leur analyse dépasserait les buts de notre travail, puisque sur la base des valeurs aussi basses, il serait difficile voire impossible de vérifier des tendances. Nous devrions analyser ces cas sur un échantillon considérablement plus grand, pour qu'ils soient valides.

Selon notre étude, il est possible de ranger les vrais amis sur une échelle de l'identité absolue (la minorité des cas) jusqu'à la parenté plus haute ou plus basse, et ceci sur la base de nombreux critères, qui peuvent entrer en différentes combinaisons : la prononciation, l'orthographe, la morphologie et la fréquence. Ces critères sont pour la plupart formelles, ils concernent surtout *le signifiant*, parce que le sens, et donc *le signifié*, a été considéré comme identique. Et précisément, cette identité de sens a été vérifiée par une fréquence presque pareille dans les deux traductions du même registre et dans le même contexte. Ici, il faut tenir compte du fait que tout échantillon du texte (un sous-corpus) est toujours délimité. Il lui est impossible de contenir tous les sens dont un mot (polysémique) dispose.

Tous les critères proposés ci-dessous sont toujours subdivisés entre les zones suivantes : *identique*, *proche* et *différente* (c'est-à-dire éloignée parmi les trois catégories, mais toujours reconnaissable : il s'agit des vrais amis).

La catégorie *identique* ne comprend que des divergences mineures, négligeables (par ex. l'accent dans l'orthographe ou la prononciation de « r » différente dans les deux langues) : c'est vraiment l'aspect de l'identité qui prévaut.

La *proximité* est définie comme une similarité comprenant une légère différence dans un seul aspect, par ex. une lettre différente, ou même l'inversion des lettres à l'écrit.

La *différence* est puis référée à ces vrais amis, qui diffèrent, ou sont éloignés l'un de l'autre, par deux ou plusieurs aspects (par ex. par 3 lettres différentes à l'écrit).

### 3.2.1 Critères

#### 3.2.1.1 La prononciation

On classe la prononciation dans cette étude en prononciation (presque) *identique* (par ex. Madam [mæd.əm] (A) – Madame [ma.dam] (F)), *proche* (par ex. place [pleis] (A) – place [plas] (F)) et *différente* (par ex. cohesion [kəʊ'hi:.ʒən] (A) – cohésion [kœzjɔ̃] (F)).

Vu que la prononciation de l'anglais et du français sont assez différentes, la prononciation identique devrait présenter un phénomène assez rare – en effet, nous avons déjà indiqué que les vrais amis sont beaucoup plus faciles à reconnaître dans la forme écrite.

#### 3.2.1.2 L'orthographe

Aussi l'orthographe peut également être divisée entre *identique*, *proche* (ou légèrement différent) et *différent*.

L'orthographe *identique* présente un accord parfait dans la forme écrite (par ex. compassion (A, F), long (A, F), question (A, F)), tandis que l'orthographe *proche* montre des détails langagiers spécifiques et/ou typiques, comme par ex. un déplacement des lettres (member (A) – membre (F)), une substitution des lettres (fund (A) – fond (F)) ou soit une addition, soit une réduction des lettres (second (A) – seconde (F), homage (A) – hommage (F), sabbath (A) – sabbat (F)) – dans ce cas, il dépend de la direction de la comparaison. Dans notre travail, nous avons utilisé le critère d'une lettre différente ou déplacée pour qualifier l'orthographe de la paire comme *proche*.

L'orthographe *différente* comprend des changements plus complexes, par ex. une réduction, une substitution et une addition des lettres en même temps (mountain (A) – montagne (F)). Pour qualifier l'orthographe en tant que différente, nous avons établi une règle d'au moins 2 lettres qui sont différentes dans les deux mots.

Remarque : Vu que les accents en français ne présentent qu'une légère modification à la forme écrite du mot, inférieure au changement, addition ou réduction des lettres – ceci de la perspective d'un locuteur non natif (sans ses accents, le mot reste dans la majorité des cas toujours compréhensible pour lui), pour les buts de ce travail,

nous avons considéré la forme graphique du mot sans y compter les accents français. Les accents français ont ainsi été toujours interprétés comme un signe de l'orthographe *identique*.

### 3.2.1.3 La morphologie

(Remarque : Il faut distinguer entre *la présence*, type d'un terme comme tel et *sa variante concrète* dans le texte (token)<sup>13</sup>. Voilà pourquoi dans cette analyse, un nom par ex. recouvre les deux, singulier et pluriel du mot et un verbe en infinitif comprend toute sa conjugaison, etc. Grâce à la richesse des terminaisons grammaticales en français (surtout chez les verbes), l'échelle imaginaire des formes de mots des types concrets est d'habitude plus riche qu'en anglais. C'est pourquoi les listes et tableaux de ce projet proposent uniquement les mots dans leur forme de base (*lemma*) – comme dans les dictionnaires (infinitif du verbe, singulier du nom). Le nombre auprès du type donc recouvre toutes les figures du terme (tokens) dans le texte et il peut comprendre soit les formes identiques à celle citée dans la liste, soit les formes différentes, par ex : *adopter* comprenant *nous adoptons, adoptez !, en adoptant*, etc. ; *question* impliquant *question, questions*.)

Les suffixes en tant qu'aspects morphologiques peuvent, eux aussi, être subdivisés en trois catégories : les suffixes *identiques*, *proches* (ou presque identiques) et *différents*.<sup>14</sup>

Le suffixe *identique* peut exister même si les vrais amis ont une orthographe différente, par ex. **judgement** (A) – **jugement** (F). Dans la majorité des cas, il s'agit d'emprunts tels quels, incluant le suffixe.

Le suffixe *proche* (ou presque identique) peut être accompagné par une prononciation ou un accent différents ou par une orthographe modifiée, par ex. **economic** (A) – **économique** (F), **development** (A) – **développement** (F).

On met aussi des suffixes « aborigènes », propres à la langue emprunteuse, par ex. **divinely** (A) – **divinement** (F).

---

<sup>13</sup> type vs. token (Soutet, 2005 : 66) (Čermák, 2001: 26)

<sup>14</sup> Inutile de dire que tous les vrais amis sont le résultat du processus de l'enrichissement du vocabulaire par l'emprunt. Dans la section de morphologie, cette étude s'intéresse à déterminer la proportion des emprunts comportant la dérivation.

Un groupe spécial de divergences morphologiques (concernant les suffixes) est conçu par le suffixe « zéro », fréquent surtout en anglais, opposé par le suffixe « standard » en français (par ex. concern (A) – concerner (F)). Et précisément, il s'agit des verbes, puisque les verbes doivent fonctionner dans le système de la langue particulière. En anglais, la conversion est facilement possible sans même des changements formels (par ex. du nom au verbe), ce qui n'est pas toujours le cas en français.

Bien que les vrais amis soient perçus comme des mots entiers (et non pas seulement les radicaux de mots), y compris leur appartenance aux classes de mots, il est souvent arrivé que leur classe de mot (originale) n'a pas du été conservée dans la traduction du texte concret. Par conséquent, il y a des différences statistiques (par ex. divinement (F) – in a divine way (A)). Les raisons pour cela peuvent être diverses, ressortant par ex. de la construction de la phrase, en but de renforcer la nature idiomatique de la traduction de la prose, de l'idiolecte de traducteur, etc. (généralement désignés par la mesure de la quantité constante des classes de mots en traduction).

#### **3.2.1.4 La fréquence**

La fréquence des vrais amis est différente dans les deux langues (et traductions) au cas où le mot dans une des langues dispose d'une restriction de collocation ou d'autres attributs stylistiques (par ex. il peut être plus ou moins formel en le comparant avec son « ami » dans la deuxième langue, il peut être de caractère marginal, spécialisé, archaïque, etc.), éventuellement, il est employé uniquement dans certains registres, ce qui conduit à son remplacement par une autre variante.

L'idiolecte de traducteur peut sans doute aussi jouer un rôle important par rapport à la fréquence des mots (par ex. les mots favoris, etc.).

La statistique (la fréquence) est influencée même par les paramètres grammaticaux et textuels (ou plus précisément, de formation du texte) : certains mots lexicaux peuvent être remplacés par des mots grammaticaux dans le processus de la traduction (par ex. la pronominalisation : Jésus – il), une autre fois, il est possible de les omettre complètement (l'ellipse, au cas où les mots sont en proximité ou dans le cas la

deuxième mention) ou à cause d'une autre construction grammaticale (par ex. la voix passive ou les synonymes).

### 3.2.1.5 La signification

Au premier regard, la signification des vrais amis devrait être la même dans les deux langues, ce qui peut être problématique et déroutant. Si dans un sens, le mot a un correspondant fidèle, mais dans un autre sens, les mots divergent, il faut les percevoir a priori comme les faux amis partiels. Le cas extrême peut présenter le sens identique dans un registre (particulièrement fermé) et le sens différent dans un autre, cette fois un registre ouvert. Pour cette raison, nous avons analysé les vrais amis uniquement dans le cadre du registre concret et davantage, nous les avons vérifiés dans le contexte donné. Ce qui nous mène à supposer et à croire, qu'il s'agit d'une garantie suffisante de l'identité du sens.

(Les mots apparentés peuvent aussi échouer à saisir toutes les nuances de leur homologue dans la deuxième langue (par ex. *demand* (A, revendiquer, réclamer) – *demander* (F)). Pour cette raison, nous nous prononçons exclusivement envers le sens dans un registre concret.)

On sait que les langues analytiques (y compris l'anglais et le français) se caractérisent par une polysémie et homonymie des mots remarquable (voir par ex. Skalička, 2004). Il est donc plus que vraisemblable que certains mots ont un nombre de nuances de sens, qui s'emploient avec une probabilité et fréquence différente dans différents registres (par ex. un ange hors du contexte religieux, quelqu'un de très gentil ou innocent). Pour ce mémoire de master, nous avons analysé uniquement les mots (et les sens) qui ont été employés dans les échantillons de textes déposés et élaborés.

Une autre difficulté potentielle dans les statistiques est constituée par la conversion (la dérivation non-affixale), en tant qu'un phénomène très fréquent en anglais (par ex. *face* en anglais peut signifier les deux, le nom et le verbe : *face* – nom (visage, A), *face* verbe (faire front à quelque chose, F)). C'est pour des raisons de clarté et simplicité qu'on a opté chaque fois pour *une* classe de mots prédominante, indiquée auprès du mot.

### 3.2.2 Analyse d'échantillons (extraits des sous-corpus)<sup>15</sup>

#### LÉGENDE:

abc = les noms propres/ les antroponymes    abc =les noms géographique/ les toponymes    abc = les vrais amis

## 1. LA BIBLE

### 1.1. Version anglaise

#### Matthew 2

1. Now **Jesus** having been born in **Bethlehem** of **Judaea**, in the days of **Herod** the king, behold **magi** from the east **arrived** at **Jerusalem**, saying,
2. Where is the king of the **Jews** that has been born? for we have seen his star in the east, and have come to do him **homage**.
3. But **Herod** the king having heard [of it], was **troubled**, and all **Jerusalem** with him;
4. and, **assembling** all the chief priests and **scribes** of the **people**, he inquired of them where the **Christ** should be born.
5. And they said to him, In **Bethlehem** of **Judaea**; for thus it is written through the **prophet**:
6. And \*thou\* **Bethlehem**, land of **Juda**, art in no wise the least among the **governors** of **Juda**; for out of thee shall go forth a leader who shall shepherd my **people Israel**.
7. Then **Herod**, having **secretly** called the **magi**, inquired of them accurately the time of the star that was **appearing**;
8. and having sent them to **Bethlehem**, said, Go, search out accurately concerning the child, and when ye shall have found [him] bring me back word, so that \*I\* also may come and do him **homage**.
9. And they having heard the king went their way; and lo, the star, which they had seen in the east, went before them until it came and stood over the place where the little child was.
10. And when they saw the star they **rejoiced** with exceeding great **joy**.
11. And having come into the house they saw the little child with **Mary** his mother, and falling down did

---

<sup>15</sup> Les listes des vrais amis dans le présent travail contiennent uniquement ceux cas, dont fréquence/présence dans le texte égale ou dépasse le numéro 5 dans la liste anglaise et en même temps, dont valeur de l'abondance dans les deux traductions varie jusqu'au nombre 2. Au-delà, nous ne nous prononçons pas a priori sur les noms propres dans notre analyse, mais nous les commentons brièvement dans la section 4.4.

him **homage**. And having opened their **treasures**, they **offered** to him gifts, gold, and frankincense, and **myrrh**.

12. And being **divinely** instructed in a dream not to **return** to **Herod**, they departed into their own country another way. (...)

<http://unbound.biola.edu/index.cfm?method=searchResults.doSearch>

## 1.2. Version française (extrait)

### Matthieu 2

1. Or, après que **Jésus** fut né à **Bethléhem** de **Judée**, aux jours du roi **Hérode**, voici, des **mages** de l'orient arrivèrent à **Jérusalem**, disant:
2. Où est le roi des **Juifs** qui a été mis au monde? Car nous avons vu son étoile dans l'orient, et nous sommes venus lui rendre **hommage**.
3. Mais le roi **Hérode**, l'ayant ouï dire, en fut **troublé**, et tout **Jérusalem** avec lui;
4. et ayant **assemblé** tous les principaux sacrificateurs et **scribes** du **peuple**, il s'enquit d'eux où le **Christ** devait naître.
5. Et ils lui dirent: A **Bethléhem** de **Judée**; car il est ainsi écrit par le **prophète**:
6. "Et toi, **Bethléhem**, terre de **Juda**, tu n'es nullement la plus petite parmi les **gouverneurs** de **Juda**, car de toi sortira un conducteur qui paîtra mon **peuple Israël**".
7. Alors **Hérode**, ayant appelé **secrètement** les **mages**, s'informa exactement auprès d'eux du temps de l'étoile qui **apparaissait**;
8. et les ayant envoyés à **Bethléhem**, il dit: Allez et enquérez-vous exactement touchant le petit enfant; et quand vous l'aurez trouvé, faites-le-moi savoir, en sorte que moi aussi j'aie lui rendre **hommage**.
9. Et eux, ayant ouï le roi, s'en allèrent; et voici, l'étoile qu'ils avaient vue dans l'orient allait devant eux, jusqu'à ce qu'elle vint et se tint au-dessus du lieu où était le petit enfant.
10. Et quand ils virent l'étoile, ils se **réjouirent** d'une fort grande **joie**.
11. Et étant entrés dans la maison, ils virent le petit enfant avec **Marie** sa mère; et, se prosternant, ils lui rendirent **hommage**; et ayant ouvert leurs **trésors**, ils lui **offrirent** des dons, de l'or, et de l'encens, et de la **myrrhe**.
12. Et étant avertis **divinement**, en songe, de ne pas **retourner** vers **Hérode**, ils se retirèrent dans leur pays par un autre chemin. (...)

<http://unbound.biola.edu/index.cfm?method=searchResults.doSearch>

## LES NÉGOCIATIONS DU PARLEMENT EUROPÉEN

### 1.3. Version anglaise (extrait)

I declare resumed the session of the European Parliament adjourned on Friday 17 December 1999, and I would like once again to wish you a happy new year in the hope that you enjoyed a pleasant festive period.

Although, as you will have seen, the dreaded 'millennium bug' failed to materialise, still the people in a number of countries suffered a series of natural disasters that truly were dreadful.

You have requested a debate on this subject in the course of the next few days, during this part-session.

In the meantime, I should like to observe a minute's silence, as a number of Members have requested, on behalf of all the victims concerned, particularly those of the terrible storms, in the various countries of the European Union.

Please rise, then, for this minute's silence.

(The House rose and observed a minute's silence)

Madam President, on a point of order.

You will be aware from the press and television that there have been a number of bomb explosions and killings in Sri Lanka.

One of the people assassinated very recently in Sri Lanka was Mr Kumar Ponnambalam, who had visited the European Parliament just a few months ago.

Would it be appropriate for you, Madam President, to write a letter to the Sri Lankan President expressing Parliament's regret at his and the other violent deaths in Sri Lanka and urging her to do everything she possibly can to seek a peaceful reconciliation to a very difficult situation?

Yes, Mr Evans, I feel an initiative of the type you have just suggested would be entirely appropriate.

If the House agrees, I shall do as Mr Evans has suggested

Madam President, on a point of order. (...)

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



## 1.4. Version française (extrait)

Je déclare reprise la session du Parlement européen qui avait été interrompue le vendredi 17 décembre dernier et je vous renouvelle tous mes vœux en espérant que vous avez passé de bonnes vacances.

Comme vous avez pu le constater, le grand "bogue de l'an 2000" ne s'est pas produit. En revanche, les citoyens d'un certain nombre de nos pays ont été victimes de catastrophes naturelles qui ont vraiment été terribles.

Vous avez souhaité un débat à ce sujet dans les prochains jours, au cours de cette période de session.

En attendant, je souhaiterais, comme un certain nombre de collègues me l'ont demandé, que nous observions une minute de silence pour toutes les victimes, des tempêtes notamment, dans les différents pays de l'Union européenne qui ont été touchés.

Je vous invite à vous lever pour cette minute de silence.

(Le Parlement, debout, observe une minute de silence)

Madame la Présidente, c'est une motion de procédure.

Vous avez probablement appris par la presse et par la télévision que plusieurs attentats à la bombe et crimes ont été perpétrés au Sri Lanka.

L'une des personnes qui vient d'être assassinée au Sri Lanka est M. Kumar Ponnambalam, qui avait rendu visite au Parlement européen il y a quelques mois à peine.

Ne pensez-vous pas, Madame la Présidente, qu'il conviendrait d'écrire une lettre au président du Sri Lanka pour lui communiquer que le Parlement déplore les morts violentes, dont celle de M. Ponnambalam, et pour l'inviter instamment à faire tout ce qui est en son pouvoir pour chercher une réconciliation pacifique et mettre un terme à cette situation particulièrement difficile.

Oui, Monsieur Evans, je pense qu'une initiative dans le sens que vous venez de suggérer serait tout à fait appropriée.

Si l'Assemblée en est d'accord, je ferai comme M. Evans l'a suggéré.

Madame la Présidente, c'est une motion de procédure. (...)

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

## 2. LE ROMAN

### 2.1. Version anglaise (extrait)

Soul and Body

It would be senseless for the author to try to convince the reader that his characters once actually lived. They were not born of a mother's womb; they were born of a stimulating phrase or two or from a basic situation. Tomas was born of the saying Einmal ist keinmal. Tereza was born of the rumbling of a stomach.

The first time she went to Tomas's flat, her insides began to rumble. And no wonder: she had had nothing to eat since breakfast but a quick sandwich on the platform before boarding the train. She had concentrated on the daring journey ahead of her and forgotten about food. But when we ignore the body, we are more easily victimized by it. She felt terrible standing there in front of Tomas listening to her belly speak out. She felt like crying. Fortunately, after the first ten seconds Tomas put his arms around her and made her forget her ventral voices.

Tereza was therefore born of a situation which brutally reveals the irreconcilable duality of body and soul, that fundamental human experience.

A long time ago, man would listen in amazement to the sound of regular beats in his chest, never suspecting what they were. He was unable to identify himself with so alien and unfamiliar an object as the body. The body was a cage, and inside that cage was something which looked, listened, feared, thought, and marveled; that something, that remainder left over after the body had been accounted for, was the soul.

Today, of course, the body is no longer unfamiliar: we know that the beating in our chest is the heart and that the nose is the nozzle of a hose sticking out of the body to take oxygen to the lungs. The face is nothing but an instrument panel registering all the body mechanisms: digestion, sight, hearing, respiration, thought. (...)

KUNDERA, M., *The unbearable lightness of being*, New York City: Harper Perennial, 1999, 166 pages.

## 2.2. Version française (extrait)

L'ÂME ET LE CORPS

Il serait vain, de la part de l'auteur, de prétendre que ses personnages ont réellement existé. Ils ne sont pas nés

d'un corps maternel, mais de quelques mots évocateurs

ou d'une situation clé. Tomas est né du dicton (einmal ist

keinmal). Tereza est née de ses borborygmes.

La première fois qu'elle franchit le seuil de

l'appartement de Tomas, ses entrailles furent prises de

gargouillements. Il ne faut pas s'en étonner, elle n'avait ni

déjeuné ni dîné, s'étant contentée d'un sandwich sur le

quai en fin de matinée, avant de monter dans le train.

Toute à l'idée de son audacieux voyage elle en oublia de

manger. Mais à ne point se soucier de son corps, on en

devient plus facilement la victime. Ce supplice d'entendre

ses tripes prendre la parole au moment où elle se

retrouvait face à face avec Tomas ! Elle était au bord des

larmes. Au bout de dix secondes, heureusement, Tomas

l'enlaçait, et elle put oublier les voix de son ventre.

Tereza est donc née d'une situation qui révèle

brutalement l'inconciliable dualité du corps et de l'âme,

cette expérience humaine fondamentale.

Jadis, il y a très longtemps, l'homme écoutait avec

stupéfaction le martèlement régulier qui lui parvenait du fond

de sa poitrine et se demandait ce que ça voulait dire.

Il ne

pouvait s'identifier à un corps, cette chose tellement

étrangère et inconnue. Le corps était une cage et à

l'intérieur se dissimulait quelque chose qui regardait,

écoutait, s'effrayait, pensait et s'étonnait; ce quelque

chose, ce reliquat qui subsistait, déduction faite du corps,

c'était l'âme.

Bien sûr, aujourd'hui, le corps a cessé d'être un

mystère, ce qui cogne dans la poitrine, c'est le cœur, on le

sait, et le nez n'est que l'extrémité d'un tuyau qui dépasse

pour amener l'oxygène aux poumons. Le visage n'est que

le tableau de bord auquel aboutissent les mécanismes

physiques : la digestion, la vue, l'ouïe, la respiration, la

réflexion.

(...)

KUNDERA, Milan, KÉRÉL, François (traduction), *L'insoutenable légèreté de l'être*, France: Gallimard, 1990, 476 pages.

## 4 Résultats

### 4.1 La bible

Cités ci-dessous, nous proposons les résultats pour le registre religieux, présentés selon les critères choisis.

#### VRAIS AMIS CENTRAUX – BIBLE

| la liste de mots anglaise: |   |            |   |    |   | la liste de mots française: |    |  |
|----------------------------|---|------------|---|----|---|-----------------------------|----|--|
| 1*                         | P | prophet    | ~ | 22 | N | prophète                    | 22 |  |
| 1*                         | P | disciple   | ° | 21 | N | disciple                    | 21 |  |
| 4*                         | P | fruit      | * | 15 | N | fruit                       | 16 |  |
| 3*                         | P | judgement  | * | 11 | N | jugement                    | 12 |  |
| 1*                         | P | parable    | * | 11 | N | parabole                    | 11 |  |
| 4*                         | P | generation | * | 10 | N | génération                  | 9  |  |
| 2*                         | D | angel      | ~ | 9  | N | ange                        | 10 |  |
| rare                       | D | pharisee   | * | 9  | N | pharisien                   | 9  |  |
| 1*                         | P | homage     | * | 7  | N | hommage                     | 6  |  |
| 5*                         | D | nation     | * | 7  | N | nation                      | 8  |  |
| rare                       | P | sabbath    | * | 7  | N | sabbat                      | 8  |  |
| 4*                         | P | secret     | * | 7  | N | secret                      | 7  |  |
| rare                       | P | baptise    | ~ | 6  | V | baptiser                    | 4  |  |
| 3*                         | I | enemy      | ~ | 6  | N | ennemi                      | 6  |  |
| 5*                         | P | people     | * | 6  | N | peuple                      | 6  |  |
| rare                       | P | scribe     | * | 6  | N | scribe                      | 6  |  |
| 1*                         | P | synagogue  | * | 6  | N | synagogue                   | 8  |  |
| 4*                         | D | commit     | ° | 5  | V | commettre                   | 5  |  |

#### LÉGENDE:

N = nom, ADJ = adjectif, V = verbe

abc = orthographe différente, abc = orthographe proche, abc = orthographe identique

la fréquence/présence proche la fréquence /présence identique

I = prononciation identique, P = pronon. proche, D = pronon. différente

\* = suffixe identique ~ = suffixe proche ° = suffixe différent

#### 4.1.1 La prononciation

Dans la liste des vrais amis basée sur le registre religieux, nous avons trouvé tous les cas de la prononciation : (presque) *identique*, *proche* et *différente*. La prononciation (presque) *identique* compte un seul cas (enemy [enəmi ] (A) – ennemi [ɛnmi] (F)), la prononciation *proche* est représentée par 12 cas (par ex. demon [di:mən] (A) – démon [demɔ̃] (F), fruit [fru:t], [fɪʁi] (A, F), judgement [dʒʌdʒmənt ] (A) – jugement [ʒyʒmɑ̃] (F)) et la prononciation *différente* trouve sa représentation dans 5 cas (par ex. angel [eɪn.dʒəl] (A) – ange [ɑ̃:ʒ] (F), enter [entə(r)] (A) – entrer [ɑ̃ tre] (F), pharisee [fær.i.si:] (A) – pharisien [farizjɛ̃] (F)).

#### 4.1.2 L'orthographe

En ce qui concerne l'orthographe, l'orthographe *identique* est représentée par 7 cas (par ex. disciple (A, F), generation (A) – génération (F), nation (A, F), secret (A, F)). L'orthographe *proche* compte 9 cas (par ex. angel (A) – ange (F), homage (A) – hommage (F), prophet (A) – prophète (F)). Et finalement, il y a seulement 2 cas de l'orthographe *différente* (commit (A) – commettre (F), pharisee (A) – pharisien (F)).

#### 4.1.3 La morphologie

Pour ouvrir la section de morphologie, nous aimerions tout d'abord énumérer les classes de mots trouvées dans la liste de « la sélection centrale », en les accompagnant par leurs nombres exacts : 17 noms et un verbe (aucun adjectif ni adverbe ne répondait aux critères de la « sélection centrale »).

Au-delà, la liste présente les différents suffixes : les suffixes *identiques*, les suffixes *proches* et les suffixes *différents*. Les suffixes *identiques* comptent 4 cas au total (par ex. judgement (A) – jugement (F), generation (A) – génération (F), homage (A) – hommage (F), nation (A, F)), les suffixes *proches* ont été relevés 2 fois (par ex. prophet (A) – prophète (F)), et le suffixe *différent* une seule fois (pharisee (A) – pharisien (F)).

#### 4.1.4 La fréquence

Par rapport à la fréquence, nous nous permettrons d'introduire cette section par le nombre de termes constituant « la sélection centrale », dont fréquence dépasse le chiffre 5 dans la liste, et qui ont trouvé leur opposé. Particulièrement, il s'agit de 18 cas qui ont été employés aussi fréquemment dans les deux traductions (dont la différence de fréquence/présence compte jusqu'au plus ou moins 2 fois).

La somme de toutes leurs fréquences (y compris toutes les variantes concrètes employées dans le texte) donne le nombre 171, donc 1,71% du nombre totale de 10 000 mots de l'échantillon entier.

Il paraît aussi important de mentionner le nombre total de mots qui disposent d'une fréquence (ou présence) identique dans les deux traductions : 33 mots, c'est-à-dire 33% en tout (par ex. secret (A, F)), dont 9 mots provenant de « la sélection centrale », délimitée par la fréquence 5 fois ou plus. Il y a aussi des mots qui présentent une fréquence similaire (plus ou moins deux cas) : 26 mots (26%) en tout (par ex. generation (A) – génération (F)), dont de nouveau 9 mots de la liste délimitée par la fréquence 5 fois ou plus. En totalité, il s'agit de 59% des cas qui se reflètent presque parfaitement dans les deux versions du même texte ; en prenant en considération uniquement la liste des fréquences de valeur 5 ou plus, autrement dit « la sélection centrale », il ne s'agit que de 18%.

Finalement, il paraît intéressant de présenter les nombres de la fréquence des termes en anglais (toujours dans la liste délimitée par la fréquence 5 fois ou plus) : selon le dictionnaire Collins COBUILD (2001), on a compté 2 cas de mots très fréquents (5\*), 4 cas de mots classifiés comme assez fréquents (4\*), 2 cas de mots fréquents (3\*), un cas de mots moins fréquents (2\*) et 5 cas de mots peu fréquents (1\*) et 4 mots rares. Ce qui indique qu'une plus grande moitié des mots sont plutôt spécialisé (marqués), religieux, dans le cadre du lexique anglais.

## 4.2 Le parlement

Cités ci-dessous, nous proposons les résultats pour le registre parlementaire, présentés selon les critères choisis.

### VRAIS AMIS CENTRAUX – PARLEMENT

|      |   |                   |   |    |     |               |    |
|------|---|-------------------|---|----|-----|---------------|----|
| 4*   | P | <u>parliament</u> | * | 49 | N   | parlement     | 51 |
| 1*   | I | Madam             | ~ | 35 | N   | Madame        | 34 |
| rare | P | <u>report</u>     | * | 29 | N   | rapport       | 29 |
| 5*   | I | programme         | * | 27 | N   | programme     | 29 |
| 5*   | P | fund              | * | 25 | N   | fond          | 26 |
| 5*   | D | point             | * | 15 | N   | point         | 14 |
| 5*   | D | <u>council</u>    | * | 10 | N   | conseil       | 11 |
| 2*   | D | structural        | * | 10 | ADJ | structurel    | 11 |
| 5*   | D | case              | ~ | 10 | N   | cas           | 11 |
| 5*   | D | <u>concern</u>    | ° | 9  | V   | concerner     | 10 |
| 5*   | I | possible          | * | 9  | ADJ | possible      | 8  |
| 4*   | D | principle         | ~ | 9  | N   | principe      | 7  |
| 1*   | D | harmonisation     | * | 8  | N   | harmonisation | 7  |
| 4*   | D | <u>particular</u> | ° | 8  | ADJ | particulier   | 8  |
| 4*   | P | period            | ~ | 8  | N   | période       | 8  |
| 1*   | D | cohesion          | * | 7  | N   | cohésion      | 7  |
| 5*   | I | conference        | * | 7  | N   | conférence    | 6  |
| 5*   | D | course            | ~ | 7  | N   | cours         | 6  |
| 5*   | P | development       | * | 7  | N   | développement | 7  |
| 5*   | I | party             | * | 7  | N   | parti         | 5  |
| 4*   | P | session           | * | 7  | N   | session       | 6  |
| 1*   | P | coordination      | * | 6  | N   | coordination  | 6  |
| 5*   | I | <u>economic</u>   | ° | 6  | ADJ | économique    | 6  |
| 3*   | P | importance        | * | 6  | N   | importance    | 5  |
| 4*   | D | <u>necessary</u>  | ° | 6  | ADJ | nécessaire    | 8  |
| 5*   | P | <u>process</u>    | ° | 6  | N   | procéssus     | 6  |
| 4*   | D | propose           | ~ | 9  | V   | proposer      | 8  |
| 2*   | I | tourism           | ~ | 6  | N   | tourisme      | 7  |
| 4*   | P | activity          | ~ | 5  | N   | activité      | 7  |
| 4*   | P | annual            | ~ | 5  | ADJ | annuel        | 5  |
| 5*   | P | place             | * | 5  | N   | place         | 5  |
| 3*   | D | procedure         | * | 5  | N   | procédure     | 6  |
| 4*   | D | <u>sector</u>     | ° | 5  | N   | secteur       | 6  |

#### LÉGENDE:

N = nom, ADJ = adjectif, V = verbe

abc = orthographe différente, abc = orthographe proche, abc = orthographe identique

la fréquence/présence proche la fréquence /présence identique

I = prononciation identique, P = pronon. proche, D = pronon. différente

\* = suffixe identique ~ = suffixe proche ° = suffixe différent

### 4.2.1 La prononciation

Dans la liste des vrais amis basée sur le registre politique (et particulièrement, dans la section délimitée par la fréquence 5 ou plus, dit « la sélection centrale »), la prononciation *identique* compte 7 cas (par ex. conference [kɒnfərəns] (A) – conférence [kɔ̃ fɛrɑ̃ ˈs] (F), party [pa:ti] (A) – parti [partɪ] (F), programme [prəʊgræm], [prɔ̃grɑm] (A, F)). La prononciation *proche* est représentée par 12 cas (par ex. development [dɪ'veləpmənt] (A) – développement [devlɔpmɑ̃] (F), fund [fʌnd] (A) – fond [fɔ̃] (F), parliament [pa:ləmənt] (A) – parlement [parləmɑ̃] (F)), et la prononciation *différente* trouve sa représentation dans 14 cas, (par ex. case [keɪs] (A) – cas [kɑ] (F), necessary [nesəsəri](A) – nécessaire [nesesɛ:r] (F), particular [pə'tɪkjələ(r)] (A) – particulier [partikylje] (F)).

### 4.2.2 L'orthographe

En ce qui concerne l'orthographe, l'orthographe *identique* est représentée par 10 cas (par ex. programme (A, F), point (A, F), possible (A, F)). L'orthographe *proche* compte 13 cas (par ex. Madam (A) – Madame (F), case (A) – cas (F), period (A) – période (F)). Et finalement, il y a 9 cas d'orthographe *différent* (par ex. council (A) – conseil (F), parliament (A) – parlement (F), necessary (A) – nécessaire (F)).

### 4.2.3 La morphologie

Pour ouvrir la section de morphologie, énumérons d'abord les classes de mots trouvées dans la liste, en les accompagnants par leurs nombres exacts : 25 noms, 6 adjectifs et 2 verbes (aucun adverbe ne réponde aux critères de « la sélection centrale »).

Au-delà, la liste de « la sélection centrale » présente des différents affixes. Les suffixes *identiques* comptent 9 cas au totale (par ex. coordination (A,F), importance (A,F), parliament (A) – parlement (F), possible (A, F),), les suffixes *proches* ont été révélés 5 fois (par ex. structural (A) – structurel (F), economic (A) – économique (F)),



et les suffixes *différents* seulement 2 fois (par ex. **necessary** (A) – **nécessaire** (F), **sector** (A) – **secteur** (F)).

#### 4.2.4 La fréquence

Quant à la fréquence, nous proposons d'ouvrir cette section par le nombre de termes choisis, dont fréquence a dépassé le chiffre 5 dans la liste, et particulièrement, ce groupe compte 33 cas appartenant à « la sélection centrale ».

La somme de leurs fréquences (y compris toutes les variantes concrètes employées dans le texte) donne le nombre 344, donc 3,44% de l'échantillon anglais comptant 10 000 mots. Il paraît aussi approprié de mentionner le nombre de mots qui disposent d'une fréquence *identique* dans les deux traductions : 10 (par ex. **economic** (A) – **économique** (F), **place** (A, F), **report** (A) – **rapport** (F)), et aussi de mots qui présentent une fréquence *similaire* (plus ou moins deux cas) : 23 (par ex. **importance** (A, F), **possible** (A, F), **programme** (A, F)).

Finalement, voici les nombres de la fréquence des termes en anglais : on a compté 14 cas de mots très fréquents (5\*), 10 cas de mots assez fréquents (4\*), 2 cas de mots fréquents (3\*), 2 cas de mots moins fréquents (2\*), 4 cas de mots peu fréquents (1\*) et un cas de mot rare. Ce qui indique que les mots sont déjà bien intégrés dans le lexique anglais.

### 4.3 Le roman

Cités ci-dessous, nous proposons les résultats pour le registre religieux, présentés selon les critères choisis.

#### VRAIS AMIS CENTRAUX – ROMAN

| la liste de mots anglaise: |   |            |   |    |     | la liste française: |    |
|----------------------------|---|------------|---|----|-----|---------------------|----|
| 2*                         | P | compassion | * | 19 | N   | compassion          | 19 |
| 2*                         | D | mistresse  | * | 11 | N   | maîtresse           | 9  |
| 4*                         | P | table      | * | 8  | N   | table               | 8  |
| 2*                         | P | eternal    | ~ | 7  | ADJ | éternel             | 8  |
| 5*                         | P | place      | * | 7  | N   | place               | 8  |
| 4*                         | I | positive   | ~ | 6  | ADJ | positif             | 5  |
| 5*                         | P | second     | ~ | 6  | N   | seconde             | 4  |
| 5*                         | D | change     | ~ | 5  | V   | changer             | 5  |
| 5*                         | D | decision   | * | 5  | N   | décision            | 6  |
| 2*                         | I | erotic     | ° | 5  | ADJ | érotique            | 4  |
| 5*                         | I | letter     | ~ | 5  | N   | lettre              | 5  |
| 5*                         | D | offer      | ° | 5  | V   | offrir              | 3  |
| 5*                         | P | press      | ° | 5  | V   | presser             | 4  |

#### LÉGENDE:

N = nom, ADJ = adjectif, V = verbe

abc = orthographe différente, abc = orthographe proche, abc = orthographe identique

la fréquence/présence proche la fréquence /présence identique

I = prononciation identique, P = pronon. proche, D = pronon. différente

\* = suffixe identique ~ = suffixe proche ° = suffixe différent

#### 4.3.1 La prononciation

Même dans la liste des vrais amis basée sur le roman, nous avons trouvé des cas de la prononciation *identique*, *proche* et *différente*. La prononciation *identique* compte cette fois 3 cas (erotic [I'rɒt.ɪk] (A) – érotique [ɛʁ ɔtik] (F), letter [letə(r)] (A) – lettre [lɛtʁ] (F), positive [pɒzətɪv] (A) – positif [pozitif] (F)). La prononciation *proche* est représentée par 6 cas (par ex. compassion [kəm'pæʃ.ən], [kɔ̃'pæsʃɔ̃] (A, F), eternal [I'tɜː.nəl] (A) – éternel [etɛʁnɛl] (F), place [pleɪs], [plas] (A, F)), et la prononciation *différente* trouve sa représentation dans 4 cas (par ex. change [tʃeɪndʒ] (A) –

changer [ʃ ɑ̃ʒe] (F), decision [di'siʒn] (A) – décision [desizjɔ̃] (F), mistress [mis.trəs] (A) – maîtresse [mɛtrɛs] (F)).

#### 4.3.2 L'orthographe

En ce qui concerne l'orthographe, l'orthographe *identique* est représentée par 4 cas (par ex. compassion (A, F), decision (A) – décision (F), place (A, F)). L'orthographe *proche* compte aussi 4 cas (par ex. eternal (A) – éternel (F), letter (A) – lettre (F), second (A) – seconde (F)), et finalement, il y a 5 cas d'orthographe *différente* (par ex. erotic (A) – érotique (F), mistress (A) – maîtresse (F), positive (A) – positif (F)).

#### 4.3.3 La morphologie

Pour ouvrir la section de morphologie, nous aimerions tout d'abord présenter les classes de mots trouvées dans la liste de « la sélection centrale », en les accompagnants par leurs nombres exacts : 7 noms, 3 adjectifs et 3 verbes (aucun adverbe ne répondait aux critères de « la sélection centrale »).

Sinon, la liste présente les différents suffixes : les suffixes *identiques* comptent 3 cas au totale (par ex. compassion (A,F), decision (A) – décision (F), mistress (A) – maîtresse (F)), les suffixes *proches* ont été révélés 3 fois (par ex. eternal (A) – éternel (F), erotic (A) – érotique (F), positive (A) – positif (F)), et aucun suffixes *différent*.

#### 4.3.4 La fréquence

En ce qui concerne la fréquence, nous nous permettrons d'introduire cette section par le nombre de termes choisis, dont fréquence dépasse le chiffre 5 dans la liste, et particulièrement, il s'agit de 19 cas. Ce groupe compte 13 cas de mots qui appartiennent à « la sélection centrale ».

La somme de leurs fréquences (y compris toutes les variantes concrètes employées dans le texte) donne le nombre 94, donc 0,94% de de l'échantillon anglais comptant 10 000 mots. Il paraît aussi important de mentionner le nombre de mots qui disposent d'une fréquence *identique* dans les deux traductions : 4 (par ex. compassion (A, F), letter (A) – lettre (F), table (A, F)), et aussi de mots qui présentent une fréquence *similaire* (donc plus ou moins deux cas) : 9 (par ex. decision (A) – décision (F), place (A, F), second (A) – seconde (F)).

Finalement, il paraît nécessaire de présenter les nombres de la fréquence des termes en anglais : on a compté 7 cas de mots très fréquents (5\*), 2 cas de mots assez fréquents (4\*), 0 cas de mots fréquents (3\*), 4 cas de mots moins fréquents (2\*), 0 cas de mots peu fréquents (1\*) et 0 cas de mots rares. Ce qui indique que les mots sont déjà bien intégrés dans le lexique anglais, qui n'appartiennent pas au lexique spécialisé, puisque ce sont les termes de la langue parlée.

#### 4.4 Noms propres

Complétons le chapitre des résultats par la constatation que les échantillons des textes contiennent aussi un type marginal de vrais amis, les noms propres, et précisément, les anthroponymes et les toponymes.<sup>16</sup> Par rapport aux classes de mots, il s'agit soit de noms/substantifs (par ex. Tomas (A, F), Prague (A, F)), soit des adjectifs (par ex. European (A) – européen (F)). Dans certains cas, les noms sont très fréquents (par ex. dans le registre religieux, Jesus (A) 47 – Jésus (F) 49). Même dans cette catégorie, on rencontre des modifications linguistiques en ce qui concerne la prononciation et l'orthographe.

Voici des exemples de la prononciation *identique* : Babylon (A) – Babylone (F), de la prononciation *proche* : David (A) – David (F) et finalement, de la prononciation *différente* : Jordan (A) – Jourdain (F).

Au-delà, l'orthographe *identique* peut être illustré par Anna (A, F), *proche* par Egypt (A) – Égypte (F) et *différent* par Czech (A) – tchèque (F).

On peut mentionner que certains noms sont propres au registres concrets, comme l'est par ex. Jesus (A) – Jésus (F) et Herod (A) – Hérode (F) dans le registre religieux.

---

<sup>16</sup> Par la notion de noms propres nous comprenons uniquement ceux qui ne peuvent pas être employés comme des noms communs.

## 5 Conclusion

La présente analyse porte sur les vrais amis entre l'anglais et le français, des langues voisines, interconnectées culturellement et historiquement.

Pour récapituler, un corpus spécial a été rassemblé pour les buts de ce travail, contenant environ 60 000 mots. Particulièrement, trois échantillons de textes traduits « en miroir » ont été choisis, dont les versions anglaises comptent chacune 10 000 mots (leurs traductions françaises contiennent les nombre de mots légèrement différents).

Vu que les échantillons anglais comptent exactement 10 000 mots, nous avons opté pour les versions anglaises en tant que des versions « de départ ». Au-delà, il paraît approprié de prendre comme point de départ la langue a priori emprunteuse pour mieux comprendre les tendances concernant les vrais amis.

Les trois échantillons recouvrent des différents registres : le registre religieux (la Bible), le registre politique (un ensemble des négociations au Parlement européen) et le registre de prose artistique (le roman). Ces registres ont été choisis en considération du contact intensif langagier et culturel entre les deux langues. Le corpus proposé dans ce travail peut être décrit comme synchronique et récent, spécialisé – rassemblé pour les buts de cette analyse, écrit, bilingue et clos.

Tous les trois textes ont été élaborés en tant que des traductions « en miroir », anglaises et françaises. Ces textes ont été choisis aussi parce que nous avons supposé que le texte biblique serait le plus fidèle à son origine (s'approchant ainsi vers la traduction littérale), la traduction du roman serait la plus libre et la traduction politique aurait les traits des caractéristiques communicatives et pragmatiques.<sup>17</sup> Nous aspirions à la représentation équilibrée de ces approches vers la traduction.

La présente analyse a été réalisée à l'aide du moyen pour les études sur corpus, le programme AntConc 3.4 2014.

Sur la base des statistiques de fréquence (vérifiées dans le contexte), une liste des vrais amis a été générée pour chaque registre, comptant en totalité 135 mots (attaché

---

<sup>17</sup> Il est possible que des données de fréquence dans les sous-corpus concrets ont été influencées par le style et le type de traduction.

ci-dessous). Il est nécessaire d'en tenir compte surtout par rapport aux résultats proposés par le présent travail : ces résultats ne touchent que les registres concrets.

Les deux listes des résultats (en anglais et en français) ont été comparées, et ceci ensuite par rapport à leur fréquence dans les deux langues (ainsi donnant la perspective statistique) et puis par rapport à leur traduction dans les passages des textes concrets (moyennant ainsi la perspective contextuelle). Au-delà, nous n'avons pris en considération que des cas qui présentaient la même classe de mot dans les deux traductions.

En particulier deux catégories de paires des vrais amis ont été considérées comme pertinentes pour le présent mémoire de master: la catégorie de fréquence *identique* (comprenant les vrais amis employés aussi fréquemment dans les deux traductions), la catégorie de fréquence *proche* ou presque identique (comprenant les vrais amis employés avec une légère différence – de plus ou moins – 1 ou 2 fois).

Particulièrement, nous nous sommes surtout concentrés sur les mots dont fréquence ou présence dans la liste anglaise égalait ou dépassait le chiffre 5 (donc nous n'avons pris en considération que des cas de l'abondance de valeur 5 ou plus).

Ces vrais amis ont été établies pour ce travail comme « centrales », vérifiées du côté statistique et contextuel, et elles ont été étudiées en détail. L'analyse-même touchait à plusieurs critères : la prononciation, l'orthographe, la morphologie et la fréquence. Le critère stylistique a été négligé parce que les résultats sont de caractère valide uniquement pour le registre concret et contexte ; notre hypothèse de départ était que le statut stylistique sera le même dans les deux traductions.

L'identité du sens (et de l'usage) suit des statistiques identiques (ou presque identiques) et du fait qu'elle était vérifiée dans les passages concrets.

Trois sphères ont été déterminées dans le cadre de la majorité des critères : la sphère de *l'identité*, la sphère de *proximité* et la sphère de *différence*, et précisément en but de dévoiler les tendances valides. (Toutes les sphères ont été définies de manière explicite.)

Tous les registres ont été soumis à la même analyse.

A ce mémoire de master est joint l'inventaire complet du « centre » des vrais amis générés. Il comporte un ordre classement selon les registres individuels et l'autre citant tous les cas de vrais amis « centraux » relevés dans les 30 000 mots analysés, ordonnés par l'ordre alphabétique.

Les résultats concernent le corpus entier autant que les échantillons individuels (les sous-corpus).

## 5.1 PRONONCIATION

Par rapport à la prononciation, le cas le moins fréquent de manière générale, c'est la prononciation *identique*. En revanche, la prononciation *proche* constitue le cas le plus fréquent dans tous les trois registres. – En totalité, de 64 vrais amis centraux, 11 cas sont de prononciation *identique*, 30 cas de prononciation *proche* et 23 cas représentent la prononciation *différente*. On peut donc constater que la prononciation oscille entre *proche* et *différente*.

Le nombre le plus élevé de la prononciation *identique* se trouve dans le registre du roman. La bible, par contre, offre le nombre le plus bas des cas de la prononciation *identique*.

Nous avons enregistré des contrastes moindres dans le registre parlementaire et par contre, les contrastes considérables dans le registre de la bible.

Il est clair que même s'il s'agit de vrais amis, la prononciation en est légèrement différente : enfin, ils proviennent des deux langues différentes (aussi, grâce aux traits typiques de prononciation concrète, on peut distinguer, de quelle langue il s'agit).

## 5.2 ORTHOGRAPHE

Des 64 vrais amis « centraux », les sphères de vrais amis *proches* (26 cas) et *identiques* (22 cas) par rapport à l'orthographe paraissent être très équilibrées, les vrais amis *proches* constituant le groupe majoritaire. Par contre, les vrais amis de l'orthographe *différente* sont partout en minorité (16 cas). Ce qui indique que l'identification des vrais amis à l'écrit est plutôt facile.

Le registre du parlement est un équilibre entre les trois zones (dans les deux perspectives, des types et des variantes concrètes) : 11 cas d'orthographe *identique*, 13 cas d'orthographe *proche* et 9 cas d'orthographe *différent*.

Dans le registre de la bible, il est possible de trouver la majorité de l'orthographe *proche* (12 cas), puis vient le groupe beaucoup moins nombreux de l'orthographe *différente* (5 cas) et en minorité les cas de l'orthographe *identique* (1).

Pareillement, le registre du roman dispose des 3 cas de l'orthographe *identique*, 6 cas *proches*, et 4 cas de l'orthographe *différente*. Et donc, les phénomènes prédominants sont ici *l'identité* et *la proximité*.

Si on projette les résultats sur l'abondance de l'emploi dans les textes, les tendances sont un peu différentes. Par rapport à l'orthographe, dans les registres parlementaire et biblique la *proximité* orthographique (l'orthographe *proche*) prédomine toujours, tandis que dans le registre du roman, cette fois c'est *l'identité*, ou l'orthographe *identique* qui prédomine.

De manière générale, on peut constater que le cas le plus fréquent est l'orthographe *proche*, en deuxième place par rapport à la fréquence se trouve l'orthographe *identique* et finalement, l'orthographe *différente* est plutôt rare.

### **5.3 MORPHOLOGIE**

En ce qui concerne les classes de mots, on peut constater que des 64 vrais amis « centraux », les noms ont la représentation majoritaire (49 cas), le reste comptant les adjectifs (9 cas) et les verbes (6 cas). Au-delà, par rapport à la distribution des classes de mots en registres individuels, les noms prédominent dans toutes les trois listes.

La bible ne dispose d'aucun adjectif. Par contre, dans le registre du parlement, les adjectifs prédominent aux verbes.

Par rapport aux suffixes, parmi les 64 vrais amis présentés dans ce travail, 29 était produits de la dérivation suffixale (45,32%), dont seulement 2 cas présentaient des préfixes. La suffixation donc prédominait à la préfixation. Au-delà, la suffixation est



relativement fréquentée, formant une plus grande moitié par rapport aux tokens (52,83%).

Les plus fréquents sont les suffixes formant les noms : -ation, -ion, -ment, etc.

La représentation la plus grande des éléments dérivés figurait dans le registre parlementaire, où c'était 16 mots sur 33 (48,48%). Le registre littéraire, ensuite, compte 46,14% des dérivations et la bible 38,88%. Il s'en suit que ce sont a priori des mots longs, dont le statut stylistique est plutôt élevé en anglais.

En ce qui concerne la formation des mots des vrais amis mentionnés dans ce travail, il s'agit des emprunts à la langue romaine (le latin ou l'ancien français) pour l'anglais. Dans la section de morphologie, cette étude s'intéresse à déterminer la proportion des emprunts comportant la dérivation. Soit, les emprunts ne disposaient pas des affixes, soit les emprunts étaient reçus tels quels, en tant qu'une unité, y compris les suffixes originales.

Dans les registres parlementaire et artistique prédomine la dérivation (contrairement au registre religieux). Traditionnellement, on parle des mots longs.

Si on prend en considération tous les trois échantillons et l'aspect de la dérivation par rapport à la quantité, le trait dominant c'est l'*identité* du suffixe, la deuxième place est occupée par les suffixes *proches* et finalement, les suffixes *différents* sont en minorité. Pour la langue emprunteuse, il paraît plus facile de garder l'orthographe que la prononciation du mot.

Le groupe majoritaire est constitué par les noms, en seconde position par rapport aux nombres viennent les adjectifs et finalement, il y a les verbes. Le plus grand nombre de noms parmi les trois échantillons peut être constaté dans le registre religieux, pendant que le plus grand nombre d'adjectifs et verbes se trouve dans le registre littéraire, où les valeurs éprouvent le meilleur équilibre. Il est intéressant que dans le cas du roman, on a enregistré une diminution remarquable des noms, ce qui s'explique probablement par le dynamisme de discours croissant.

## 5.4 FRÉQUENCE

Notre étude a prouvé que la plus haute probabilité de trouver les vrais amis « centraux » parmi ces échantillons, est – en ce qui concerne la fréquence – dans le cas du registre du parlement (contenant 3,44% de tout l'échantillon de 10 000, incluant les mots grammaticaux). Le registre de la bible se trouve en deuxième place par rapport à la fréquence ou présence de vrais amis « centraux » dans le texte (1,71%), et le registre du roman occupe la dernière place à cet égard (0,94%). Il s'en suit que plus que le registre est fermé, plus la présence des vrais amis augmente (il faudrait vérifier cette tendance sur un corpus plus large).

Le nombre le plus grand de mots éprouvant la fréquence la plus haute (5 fois et plus) peut être trouvé dans la liste de vrais amis « politiques » (33 cas). La liste religieuse et littéraire incluent des mots beaucoup moins fréquents, comme la bible ne présente que 18 et le roman 13 cas de mots égalant ou surpassant la présence ou fréquence de valeur 5 dans les échantillons.

Le nombre le plus grand de mots disposant d'un recouvrement sémantique parfait se trouve de nouveau dans la liste de registre politique (10 cas), tout aussi comme la plus grande quantité des mots disposant d'un recouvrement sémantique approximatif (plus ou moins deux cas), 23 cas.

Le registre de la bible présente une divergence intéressante : presque une moitié des cas de vrais amis est peu fréquenté ou rare, tandis que la deuxième moitié est classifiée comme fréquentée ou même très fréquentée dans la langue anglaise selon Collins COBUILD (2001).

Les registres du roman et du parlement, à leur tour, disposent du pourcentage le plus haut des éléments d'une très haute fréquence dans le système entier de la langue anglaise.

La fréquence absolument la plus haute a été constatée chez le mot parlement (A, 49 cas) – parlement (F, 51 cas), la suivante chez le mot Madam (A, 35) – Madame (F, 34 répétitions) et la fréquence la plus haute en troisième place a été attribuée au mot report (A, 27) – rapport (F, 29). Il faut noter que tous les mots proviennent d'un seul registre : celui du parlement.

La fréquence égale ou dépassante le numéro 10 a été enregistrée chez 9 mots appartenant au registre du parlement (parliament (A) – parlement (F), Madam (A) – Madame (F), programme (A, F), fund (A) – fond (F), report (A) – rapport (F), point (A, F), council (A) – conseil (F), structural (A) – structurel (F), case (A) – cas (F)), 6 mots de la liste biblique et seulement 2 (disciple (A,F), generation (A) – génération (F), fruit (A, F), judgement (A) – jugement (F), parable (A) – parabole (F), prophet (A) – prophète (F)) mots des vrais amis du registre littéraire (compassion (A, F), mistress (A) – maîtresse (F)). Ce qui symbolise le caractère plutôt fermé de l'échantillon parlementaire et une variété du roman.

Le registre de la bible possède le plus haut nombre de vrais amis (59) en totalité parmi les trois échantillons choisis, quand seulement un tiers incomplet a atteint la fréquence 5 ou plus – pour la plupart, ce sont des mots qui ne constituent pas le centre du lexique anglais. Et précisément, ce sont les mots associés au discours religieux, qui en anglais disposent des valeurs de la fréquence assez basses dans le système langagier anglais selon Collins COBUILD (2001).

L'échantillon parlementaire compte 38 cas des vrais amis (de fréquence différente de valeur 2 au maximum dans le cadre de chaque pair), dont la majorité (98,84 %) dispose de présence dans le texte surpassant 5 occurrences. Ce qui indique qu'il s'agit de mots (et des formules) stéréotypés et régularisés.

En ce qui concerne l'échantillon littéraire, celui-ci comprend le même nombre de vrais amis comme son opposé parlementaire, 38 cas, mais cette fois-ci, seul un tiers est de fréquence 5 ou plus. Dans ce cas, ce sont traditionnellement les mots constituant le centre du lexique, de fréquence très haute (5\*) ou assez haute (4\*) dans le lexique anglais selon Collins COBUILD (2001). Quant au reste, les deux autres tiers sont de caractère rare, ce qui implique que l'auteur aspire à employer le langage riche et varié.

Si on projette ces constatations de fréquence sur les nombres de 10 000 mots en perspective des registres concrets, la représentation la plus basse a été découverte dans l'échantillon littéraire (0,94%), moyenne dans le registre religieux (1,71%) et la plus haute dans le registre de parlement (3,44%). Ceci indique la plus haute mesure des

tokens chez les registres fermés, qui se distinguent par un appareil délimité et réglé des termes.

## 5.5 SENS

Le sens des vrais amis figurants dans cette étude a été vérifié dans le contexte et les termes ont été analysés dans le cadre du registre concret. Les résultats proposés sont donc pertinent uniquement pour un registre concret. S'il s'agit d'un mot polysémique, et il ne présente qu'un sens concret dans les échantillons étudiés dans cette analyse, le présent travail ne le contient pas. Nous ne généralisons pas, nous ne concluons pas sur d'*autres* registres qui peuvent contenir les vrais amis présents dans cette étude.

De manière générale, on peut conclure que l'échantillon de la bible est le plus contrasté, par contre le registre littéraire est le plus équilibré en ce qui concerne la fréquence des mots dans le lexique anglais.

Il a été vérifié à l'aide du dictionnaire étymologique que 64 cas de vrais amis « centraux », dont la fréquence égalait ou surpassait le chiffre 5, étaient tous d'origine romaine, emprunté par l'anglais. Même si les termes individuels peuvent différer par rapport à l'origine, qui était traditionnellement latine, ils ont été introduits dans l'anglais par le français ancien. La question de la direction de l'emprunt n'est pas pertinente. En effet, l'anglais, dans cette étude, figurait toujours en tant que langue emprunteuse, cependant que le français était la langue donneuse.

Nous nous permettons de résumer les résultats concrets en pourcentage dans les tables :

LÉGENDE :

le pourcentage dans le cadre du registre de **la bible**, **du roman**, **du parlement**

**la valeur/proportion la plus remarquable** parmi les trois registres

**le pourcentage du corpus entier**

**TYPES<sup>18</sup> de mots (lexèmes) :**

|      |                      |       | BIBLE<br>(corpus entier) |        | ROMAN<br>(corpus entier) |        | PARLEMENT<br>(corpus entier) |        | (corpus<br>entier) |
|------|----------------------|-------|--------------------------|--------|--------------------------|--------|------------------------------|--------|--------------------|
|      |                      |       |                          |        |                          |        |                              |        |                    |
| 1    | prononciation        | I     | 5,56%                    | 1,56%  | 23,07%                   | 4,69%  | 21,21%                       | 10,94% | 17,19%             |
|      |                      | P     | 66,67%                   | 18,76% | 46,15%                   | 9,38%  | 36,36%                       | 18,75% | 46,88%             |
|      |                      | D     | 27,77%                   | 7,81%  | 30,77%                   | 6,25%  | 43,42%                       | 21,88% | 35,93%             |
| 2    | orthographe          | I     | 38,88%                   | 10,93% | 30,77%                   | 6,25%  | 34,33%                       | 17,19% | 34,37%             |
|      |                      | P     | 50%                      | 14,06% | 30,77%                   | 6,25%  | 39,39%                       | 20,31% | 40,62%             |
|      |                      | D     | 11,12%                   | 3,13%  | 38,46%                   | 7,81%  | 27,27%                       | 14,07% | 25,01%             |
| 3    | (suffixes)           | I     | 22,22%                   | 6,25%  | 23,07%                   | 4,69%  | 27,27%                       | 14,06% | 25%                |
|      |                      | P     | 11,11%                   | 3,13%  | 23,07%                   | 4,69%  | 15,15%                       | 7,81%  | 15,63%             |
|      |                      | D     | 5,55%                    | 1,56%  | 0%                       | 0%     | 6,06%                        | 3,13%  | 4,69%              |
|      | (classes de<br>mots) | _     | 61,11%                   | 17,19% | 53,86%                   | 10,94% | 51,52%                       | 26,56% | 54,69%             |
|      |                      | N     | 94,44%                   | 26,55% | 53,86%                   | 10,94% | 75,76%                       | 39,06% | 76,55%             |
|      |                      | ADJ   | 0%                       | 0%     | 23,07%                   | 4,69%  | 18,18%                       | 9,38%  | 14,07%             |
|      |                      | V     | 5,56%                    | 1,56%  | 23,07%                   | 4,69%  | 6,06%                        | 3,13%  | 9,38%              |
| 4    | fréquence            | I     | 50%                      | 14,06% | 30,77%                   | 6,25%  | 30,3%                        | 15,63% | 35,94%             |
|      |                      | P     | 50%                      | 14,06% | 69,23%                   | 14,06% | 69,7%                        | 35,94% | 64,06%             |
|      |                      | 5*    | 11,11%                   | 3,13%  | 53,85%                   | 10,94% | 42,42%                       | 21,86% | 35,93%             |
|      |                      | 4*    | 22,22%                   | 6,25%  | 15,38%                   | 3,13%  | 30,3%                        | 15,63% | 25,01%             |
|      |                      | 3*    | 11,11%                   | 3,13%  | 0%                       | 0%     | 6,06%                        | 3,13%  | 6,26%              |
|      |                      | 2*    | 5,55%                    | 1,56%  | 12,12%                   | 6,25%  | 6,06%                        | 3,13%  | 10,94%             |
|      |                      | 1*    | 27,77%                   | 7,81%  | 0%                       | 0%     | 12,12%                       | 6,25%  | 14,06%             |
| rare | 22,22%               | 6,25% | 0%                       | 0%     | 3,03%                    | 1,56%  | 7,8%                         |        |                    |

<sup>18</sup> type vs. token (Soutet, 2005 : 66) (Čermák, 2001: 26)

### VARIANTES CONCRÈTES de types de mots (tokens<sup>19</sup>):

|   |                   |      | BIBLE<br>(corpus entier) |               | ROMAN<br>(corpus entier) |              | PARLEMENT<br>(corpus entier) |               | (corpus<br>entier) |
|---|-------------------|------|--------------------------|---------------|--------------------------|--------------|------------------------------|---------------|--------------------|
|   |                   |      |                          |               |                          |              |                              |               |                    |
| 1 | prononciation     | I    | 3,51%                    | 0,94%         | 17,02%                   | 2,5%         | 26%                          | 15,2%         | 18,64%             |
|   |                   | P    | <b>75,44%</b>            | 21,16%        | <b>55,32%</b>            | 8,15%        | <b>42,36%</b>                | <b>24,76%</b> | <b>54,07%</b>      |
|   |                   | D    | 21,05%                   | 4,7%          | 27,66%                   | 4,07%        | 31,64%                       | <b>18,52%</b> | 27,29%             |
| 2 | orthographe       | I    | 42,11%                   | 11,29%        | <b>41,49%</b>            | 6,11%        | 27,35%                       | <b>15,99%</b> | 33,39%             |
|   |                   | P    | <b>49,70%</b>            | 13,32%        | 24,47%                   | 3,61%        | <b>38,34%</b>                | <b>22,41%</b> | <b>39,34%</b>      |
|   |                   | D    | 8,19%                    | 2,19%         | 34,04%                   | 5,02%        | 34,31%                       | <b>20,06%</b> | 27,27%             |
| 3 | (suffixes)        | I    | 29,91%                   | 5,49%         | 37,23%                   | 5,49%        | <b>45,31%</b>                | <b>26,49%</b> | 37,47%             |
|   |                   | P    | 16,37%                   | 4,39%         | 19,15%                   | 2,82%        | 8,58%                        | <b>5,02%</b>  | 12,23%             |
|   |                   | D    | 5,26%                    | 1,41%         | 0%                       | 0%           | 2,95%                        | <b>1,72%</b>  | 3,13%              |
|   | (classes de mots) | _    | <b>57,89%</b>            | 15,52%        | <b>43,62%</b>            | 6,43%        | 43,16%                       | <b>25,24%</b> | <b>47,17%</b>      |
|   |                   | N    | <b>97,08%</b>            | 26%           | <b>64,89%</b>            | 9,56%        | <b>83,38%</b>                | <b>48,75%</b> | <b>84,33%</b>      |
|   |                   | ADJ  | 0%                       | 0%            | 19,15%                   | 2,82%        | 11,8%                        | <b>6,9%</b>   | 9,72%              |
|   |                   | V    | 2,92%                    | 0,78%         | 15,97%                   | 2,35%        | 4,82%                        | <b>2,82%</b>  | 9,95%              |
| 4 | fréquence         | I    | <b>54,38%</b>            | <b>14,58%</b> | <b>39,36%</b>            | 5,8%         | 23,32%                       | 13,63%        | 34,02%             |
|   |                   | P    | 45,62%                   | 12,24%        | 60,64%                   | 8,93%        | <b>76,68%</b>                | <b>44,82%</b> | <b>65,98%</b>      |
|   |                   | 5*   | 7,6%                     | 2,04%         | 40,42%                   | 5,96%        | <b>40,21%</b>                | <b>23,51%</b> | <b>31,51%</b>      |
|   |                   | 4*   | 21,64%                   | 5,8%          | 14,89%                   | 2,19%        | 29,76%                       | <b>17,4%</b>  | 25,39%             |
|   |                   | 3*   | 9,94%                    | <b>2,66%</b>  | 0%                       | 0%           | 2,94%                        | 1,72%         | 4,38%              |
|   |                   | 2*   | 5,26%                    | 1,41%         | <b>44,69%</b>            | <b>6,58%</b> | 4,29%                        | 2,5%          | 10,49%             |
|   |                   | 1*   | <b>39,18%</b>            | <b>10,5%</b>  | 0%                       | 0%           | 15,03%                       | 8,78%         | 19,28%             |
|   |                   | rare | 16,38%                   | 4,39%         | 0%                       | 0%           | 7,77%                        | <b>4,56%</b>  | 8,95%              |

<sup>19</sup> type vs. token (Soutet, 2005 : 66) (Čermák, 2001: 26)

Pour la majorité des critères, nous avons établi des zones d'identité, de proximité et de différence. Inutile de dire qu'à la base de l'harmonie de ces similitudes, il était possible de ranger les vrais amis sur une échelle fictive. Et particulièrement, le spectre comprend des cas qui se ressemblent le plus (donc ceux, qui ont atteint l'identité dans la plupart des critères) à travers la proximité dans la majorité des critères jusqu'à ceux qui sont le plus éloignés (donc ceux, qui n'ont montrés que la différence dans plusieurs critères). Comme un exemple de vrais amis identiques dans presque tous les sens, citons *synagogue* (A, F) ou *long* (A,F), d'un exemple des vrais amis généralement proches, citons *period* (A) – *période* ou *development* – *développement* (F), et finalement, citons *mistresse* (A) – *maîtresse* et *council* (A) – *conseil* (F) comme des exemples de faux amis différents.<sup>20</sup> Il est évident que même les vrais amis peuvent être classés selon le modèle du centre, de la transition et de la périphérie (voir Daneš, 1966). Cette conception-ci concerne l'usager d'un des langues en question, mais, si on considère le côté statistique, la prononciation et l'orthographe prédominantes sont *proches*. Ce qui indique que, dans une certaine mesure, les homologues s'éloignent.

Les résultats de notre analyse montrent que les vrais amis centraux (ou « la sélection centrale » des vrais amis) comprennent 64 types et 638 tokens, et donc un type correspond en moyenne à 9,97 tokens. Par contre, les vrais amis de la périphérie contient 71 types et seulement 214 tokens, ce qui, cette fois, correspond à 3,01 tokens par un type en moyenne. Au totale, nous avons trouvé 135 types auxquels correspondent 852 tokens, et donc un type présente en moyenne 6,31 tokens.

Il s'induit de notre étude qu'il n'y a qu'un cas de vrais amis centraux<sup>21</sup> (de « la sélection centrale ») partagés par deux registres (ils sont présents et en français, et en anglais). Et particulièrement, ces vrais amis sont partagés par les registres parlementaire et littéraire : *place* (A, F). Cette constatation affirme qu'il s'agit d'un phénomène propre au registre concret il n'est pas universellement valide et applicable.

---

<sup>20</sup> Il est intéressant que notre classification des vrais amis ressemble à la classification proposée par Marcheteau, et Dahan: les vrais amis « sûrs », les vrais amis « proches » et les vrais amis à l'écrit qui deviennent des « traîtres à l'oral ». (Marcheteau, Dahan, 2009: 14, 15)

<sup>21</sup> En dehors de « la sélection centrale », il y avait 5 tels cas de plus: *compassion* (A, F), *condem* (A) – *condamner* (F), *letter* (A) – *lettre* (F), *point* (A, F), *table* (A, F).

Pour conclure, en montrant combien les deux langues ont en fait *en commun* par rapport aux homologues (les vrais amis), cette étude est en opposition complète avec la majorité des publications touchant surtout aux différences de sens entre les opposés dans deux langues (les faux amis), ce qui pourrait avoir un effet motivant aux étudiants de l'anglais et du français. Au-delà, bien que cette analyse *parole* ne considère qu'un corpus fermé, de 60 000 mots, certains vrais amis introduits dans cette étude excèdent les listes présentées dans les publications lexicographiques sur ce sujet.



## 6 Bibliographie

ADAM, J. M. Textualité et séquentialité. L'exemple de la description, *Langue française* 74, 1987, p. 51-72.

ADAM, J. M., *Les textes : types et prototypes*, Paris: Nathan, 1992, 296 p.

AYASCH, I. et HILLION, L. *Les faux amis réconciliés*, Paris: Ellipses, 2006, 263 p.

BALLARD, M. et WECKSTEEN, C., *Les faux amis en anglais*, Paris: Ellipses, 2005, 207 p.

BALLY, C., *Traité de stylistique française*, 2nd ed., Heidelberg: 1921, 349 p.

BENALI, A., Les problèmes de la catégorisation textuelle: entre fondements théoriques et fondements structurels, *Synergies Algérie* 17, 2012, p. 35-49.

BIBER, D., CONRAD, S., et REPPEN, R., *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge: Cambridge University Press, 1998, 300 p.

BLAIN, R. Discours, genres, types de textes, textes...De quoi me parlez-vous?, *Québec français* 98, 1995, p.22-25.

BOWKER, L., PEARSON, J., *Working with Specialized Language. A Practical Guide to Using Corpora*, New York: Routledge, 2002, 265 p.

CHERIGUEN, F., Typologie des procédés de formation du lexique, *Cahiers de lexicologie* 55, Paris: Didier, 1989.

ČECHOVÁ, M. et al. *Současná česká stylistika*. Praha: ISV nakladatelství, 2003, 342 p.

ČERMÁK, F., Jazykový korpus: prostředek a zdroj poznání, *Slovo a slovesnost* 56, 1995/(2000), p. 119-14.

ČERMÁK, F. et al., *Souborné dílo Vladimíra Skaličky, Díl 2. (1951-1963)*, Praha: Karolinum, 2004, 454 p.

ČERMÁK, F. *Jazyk a jazykověda*. Praha: Pražská imaginace, 1994, 251 p.

ČERMÁK, F. *Jazyk a jazykověda*. Univerzita Karlova v Praze : Karolinum, 2001, 341 p.

- CHISS, J.-L. , BALLY, C. (1865 - 1947) *Historicité des débats linguistiques et didactiques*, Stylistique, Énonciation, Crise du français Peeters, 2006, 257 p.
- DANEŠ, F., The relation of centre and periphery as a language universal. *Travaux linguistiques de Prague, 2: Les problèmes du centre et de la périphérie du système de la langue*, Prague: Académie tchécoslovaque des sciences, 1966, p. 9-21.
- ESSER, J., *Introduction to English Text-Linguistics*, Frankfurt am Mein: Peter Lang, 2009, 209 p.
- GUILHAMOU, J., HABERT, B., NAZARENKO, A. et SALEM, A., *Les linguistiques de corpus*. In: *Histoire & Mesure*, 1998 volume 13 - n°3-4. Varia. p. 414-416.
- HABOURY, F. (ed.), *Le Dictionnaire érudit de la langue française*, Paris: Larousse, 2009, 2109 p.
- HALLIDAY, M.A.K et HASAN, R., *Language, Context and Text: Aspects of Language in a Social-semantic Perspective*, Oxford: Oxford University Press, 3 ed, 1991, 125 p.
- HAVRÁNEK, B., Úkoly spisovného jazyka a jeho kultura. In *Spisovná čeština a jazyková kultura*. Praha: Melantrich, p. 32-84.
- HUMBLEY J., *Vers une typologie de l'emprunt linguistique*, Cahiers de lexicologie 25, 1974, 334 p.
- JAKOBSON, R. On Linguistic Aspects of Translation. In *On Translation*. R. A. Brower (ed.), Harvard Studies in Comparative Literature. Cambridge, MA: Harvard University Press, 1959, p. 232-239.
- JAKOBSON, R. Linguistique et poétique, *Essais de linguistique générale*. Paris: Minuit, 1963, 209-248.
- KNITTLOVÁ, D. a kol., *Překlad a překládání*, Olomouc: Univerzita Palackého v Olomouci, FF, 2010, 291 p.
- Kolektiv autorů, *Francouzsko-český, Česko-francouzský slovník*, Olomouc: Fin Publishing, 1998, 1488 p.
- KUNDERA, M., *L'insoutenable légèreté de l'être*, KÉRÉL, F. (traduction), Paris: Gallimard, 1990, 476 p.

- KUNDERA, M., *The unbearable lightness of being*, New York City: Harper Perennial, 1999, 166 p.
- MARCHETEAU, M. et DAHAN, L., *Vrais et faux amis en anglais*, Paris: CPI, 2009, 95 p.
- MARTINET, A., *La Linguistique, guide alphabétique*, Paris: Denoel, 1969.
- ONIONS, C.T. (ed.), *The Oxford dictionary of English Etymology*, Oxford: Oxford University Press, 1966, 1024 pages.
- PEPRNÍK, J. *English Lexicology*, Olomouc: Univerzita Palackého v Olomouci, 2001, 186 p.
- PETITJEAN, A., Les typologies textuelles, *Pratiques* 62, 1989, p. 86-125.
- PICOCHÉ, J. *Précis de lexicologie française*. Paris: Nathan, 1992, 190 p.
- RADINA, O., *Zrádná slova ve francouzštině*, Praha: Státní pedagogické nakladatelství, 1988, 184 p.
- REY, A. et REY-DEBOVE, J. (eds.) *Le Nouveau Petit Robert 2008. Dictionnaire alphabétique et analogique de la langue française*. Paris: Le Robert, 2008, 2837p.
- SAUSSURE, F. de, *Cours de linguistique générale. Edition critique*, ed. & comm. Tullio de Mauro, Paris: Payot, 1972/1997, 509 p.
- SCRIVEN, R. (ed.), *French Dictionary*, New York: HarperCollins Publishers, 2013, 618 p.
- SINCLAIR, J. (ed.) *Collins Cobuild English Dictionary for Advanced Learners*, Glasgow: HarperCollins Publishers, 2001, 1824 p. (COLLINS COBUILD)
- SINCLAIR, J. *EAGLES Preliminary Recommendations on Corpus Typology*. EAG-TCWG-CTYP/P, Pisa: ILC-CNR, 1996.
- TOGNINI-BONELLI, E., *Corpus Linguistics at Work*, Amsterdam: John Benjamins, 1996, 219.
- TRESCASES, P., Aspects du mouvement d'emprunt à l'anglais reflétés par trois dictionnaires de néologismes, *Cahiers de lexicologie* 42, 1983, p. 86-101.

VACHEK, J. *English Past and Present. An Introductory Course*. Praha : Státní pedagogické nakladatelství, 71p.

WALTER, H., Les “faux amis” anglais et l'autre côté du miroir, *La linguistique* 37/2, 2001, p. 101-112.

WERLICH, E., *A Text Grammar of English*, Heidelberg: Quelle & Meyer, 1976.

### **RÉSORCES ELÉCTRONIQUES:**

EDWARDS, Michael, *Un membre de l'Académie française peut-il employer des anglicismes?*, interview vidéo par RFI, 2017/3/26.  
[https://www.facebook.com/RadioFranceInternationale/videos/10154793246549442/?autoplay\\_reason=all\\_page\\_organic\\_allowed&video\\_container\\_type=0&video\\_creator\\_product\\_type=2&app\\_id=2392950137&live\\_video\\_guests=0](https://www.facebook.com/RadioFranceInternationale/videos/10154793246549442/?autoplay_reason=all_page_organic_allowed&video_container_type=0&video_creator_product_type=2&app_id=2392950137&live_video_guests=0)

IGOT, Pierre, <http://fauxamis.fr> - site consacré aux faux amis entre l'anglais et le français

<http://www.cnrtl.fr/definition/corpus>

<http://unbound.biola.edu/> (la source de notre corpus basé sur les textes bibliques)

<http://unbound.biola.edu/index.cfm?method=searchResults.doSearch> – version anglaise de la Bible : Darby

<http://unbound.biola.edu/index.cfm?method=searchResults.doSearch> – version française de la Bible : Darby (Version DARBY Copyright (c) 1991 Bible et Publications Chrétiennes, Valence FRANCE.)

<http://www.statmt.org/europarl/> (la source de notre corpus basé sur les négociations politiques)

<http://www.laurenceanthony.net/software.html> (AntConc)

## **Résumé diplomové práce:**

### *Analýza anglo-francouzských pravých protějšků v korpusu autentických vzorků textů*

L'analyse des vrais amis anglo-français d'un corpus d'échantillons textuels authentiques

Tato diplomová práce se zabývá parolovým výzkumem lexikálního tématu, a sice tzv. pravých protějšků (vrais amis) v současné francouzštině a angličtině. Jakkoli jde o jev pramenící z historických kontaktů mezi oběma sousedními jazyky i kulturami, zaměření této práce je výlučně synchronní.

Diplomová práce se člení do několika kapitol. Po stručném Úvodu, v němž se téma nastiňuje, přichází Teoretický oddíl, v němž se nejprve pozornost věnuje historickému pozadí a jazykům v kontaktu, dále se představují a vymezují lexikologické koncepty, které jsou pro práci klíčové (zejména různé vztahy mezi podobnými slovy dvou jazyků a charakteristika slovní zásoby). Značná pozornost se dále věnuje přínosu F. de Saussura relevantnímu této práci, zvláště jeho vkladu do poznání znaku (koncepty signifiant a signifié), odlišení pohledu synchronního a diachronního a rovněž jeho rozlišení langue a parole. Protože je práce zaměřena parolově, po nástinu různých funkcí jazyka se představuje výběr výrazných přístupů ke klasifikaci textů. V závěru teoretického oddílu se pojednává o překladu a také se podává nástin tradice korpusové lingvistiky, včetně přehledu různých typů korpusů.

Následující Praktická část diplomové práce vychází z kapitoly o metodice výzkumu, v níž se zvolený korpus i dílčí subkorpusy popisují z hlediska různých teoretických východisek uvedených v prvním oddílu a nastiňuje se postup mnohaaspektové analýzy pravých protějšků. Zkoumala se jejich výslovnost i psaná podoba v obou jazycích, jejich slovnědruhové zařazení, (a vedle samozřejmého postupu vypůjčování) i míra zastoupení slovtvorného postupu derivace; dále se sledovala

celková četnost jak typů, tak i tokenů v korpusu, a nepřímo (na základě pásem frekventovanosti) se zjišťovala i integrovanost těchto výpůjček do slovní zásoby angličtiny jakožto jazyka přijímajícího. Využívá se teorie centra a periferie F. Daneše (1966), neboť u většiny zkoumaných aspektů lze pravé protějšky přiřadit do tří pásem od identity přes blízkost až po relativní nepodobnost. Tyto zóny jasně vymezeny a byly ilustrovány odpovídajícími vrais amis. Oddíl o metodice výzkumu přináší také ukázkou provedené analýzy na vzorcích všech tří subkorpusů.

Další kapitola pak předkládá vlastní výsledky výzkumu, a sice pro jednotlivé registry zvlášť. Zjištění se pro každý registr člení dle jednotlivých aspektů mnohaaspektové analýzy. Závěr práce pak přináší souhrn relevantních zjištění jak z hlediska zkoumaných aspektů, tak z hlediska jednotlivých registrů. Pro přehlednost jsou výsledky předloženy též ve dvou tabulkách, z nichž jedna se týká příslušných jednotlivých typů vrais amis a druhá jejich tokenů.

K práci je připojen také obsah, český i anglický abstrakt, soupis klíčových slov v češtině a angličtině, dále české resumé, soupis použité bibliografie, a dále celá řada příloh, včetně tabulek zjištěných pravých protějšků v jednotlivých subkorpusech a abecedního seznamu všech centrálních pravých protějšků zastoupených v celém zkoumaném korpusu.

Pro účely této práce byl sestaven speciální korpus anglicko-francouzských překladů složený ze třech dílčích subkorpusů stejného rozsahu, demonstrujících biblický diskurz, politický diskurz a diskurz umělecké prózy, v celkovém rozsahu přibližně 60 000 slov. Pokud jde o délku jednotlivých subkorpusů, jak známo, překlady se přirozeně mohou délkou nepatrně lišit. Poněvadž z historické situace, jakož i relevantní odborné literatury vyplývá, že přejímajícím jazykem bývá většinou angličtina, rozsah 10.000 slov byl stanoven jako kritérium jen pro anglickou verzi, přičemž byly francouzské překlady rozsahem pouze přibližně stejné. Celkově byly tedy

analyzovány celkem tři typy registrů v současné angličtině v celkovém rozsahu 30 000 slov, a jejich protějšky, tj. francouzské překlady, přibližně stejné délky.

Byl zvolen korpusový nástroj AntConc, který lze dobře využít pro výzkum různých jazyků. Jeho pomocí byly z korpusu vygenerovány užité pravé anglicko-francouzské protějšky. Je třeba podtrhnout, že se anglo-francouzské *vrais amis* v této práci pojímají vždy výlučně jakožto jev konkrétního registru. Poněvadž se četnost výskytu těchto protějšků navzájem lišila, aby bylo vůbec možno vysledovat jisté tendence, pro zvolený multiaspektový výzkum byl stanoven jako minimální požadavek frekvence pět tokenů v angličtině a současně bylo požadováno, aby se frekvence tokenů mezi oběma jazyky lišila maximálně o dva. Dodejme, že někdy byla frekvence zcela identická, někdy byla maximálně o dva výskyty vyšší v angličtině, jindy byla naopak až o dva výskyty vyšší ve francouzštině. Následně byl každý případ ověřen v příslušných konkrétních kontextech překladů, aby byla vyloučena náhodná frekvenční shoda a současně bylo požadováno, aby se překladem neměnil slovní druh. Na základě uvedených kritérií, tj. minimálních frekvenčních požadavků, (takřka) shodných absolutních frekvencí tokenů (lišící se v překladech maximálně o dva tokeny), identických registrů a konkrétních kontextů, byl vydělen centrální okruh 64 protějšků, které se považovaly za zcela shodné z hlediska signifié (i jazykového úzu). Proto byly tyto prvky následně podrobeny mnohaaspektové analýze zejména s důrazem na aspekty signifiant.

Nejčastější skupinu tvořila podstatná jména, méně frekventovaná byla přídavná jména a nejmenší zastoupení měla slovesa, což je překvapivé, neboť dva ze zkoumaných subkorpusů měly prvek narativnosti. V subkorpusu biblického diskursu nebyla zjištěna adjektivní *vrais amis*, slovesa měla okrajové zastoupení. Zdá se, že v tomto registru substantiva mezi *vrais amis* dominují, což může plynout z přejímání především klíčových religiozních konceptů.

Svébytnou skupinu *vrais amis* tvoří vlastní jména, z hlediska slovnědruhov $\acute{e}$  příslušnosti hlavn $\acute{e}$  substantiva, okrajov $\acute{e}$  adjektiva. Tak $\acute{e}$  ta podl $\acute{e}$ hala modifikaci formy, neboť lze vysledovat případy výslovnostní i pravopisné identity, blízkosti, ale i vzdálen $\acute{e}$  podobnosti.

Délka slov je v angličtin $\acute{e}$  podstatná, protože nezřídka bývá ukazatelem stylov $\acute{e}$  příslušnosti. I když se jednalo ve sto procentech o výpůjčky, navíc výhradn $\acute{e}$  o výpůjčky do anglick $\acute{e}$ ho jazyka (prov $\acute{e}$ řené dle Etymologick $\acute{e}$ ho slovníku), bylo zajímav $\acute{e}$  zjistit, jaký podíl v t $\acute{e}$ chto výpůjčkách m $\acute{e}$ l proces derivace. Výzkum ukázal, že slovovorný postup derivace, v drtiv $\acute{e}$  většin $\acute{e}$  případů vlastn $\acute{e}$  sufixace, se vyskytuje takřka u poloviny prot $\acute{e}$ jšků. Jednalo se však o prvky značně frekventované, neboť z hlediska počtu tokenů derivované prvky tvoří mírnou většinu. Vzhledem k čast $\acute{e}$  blízkosti a někdy dokonce identit $\acute{e}$  sufixů je zřejm $\acute{e}$ , že se přejímala celá slova včetn $\acute{e}$  sufixů, velmi odlišn $\acute{y}$ ch sufixů je minimum. Lze doplnit, že zejm $\acute{e}$ na angličtina pak m $\acute{u}$ že přejatá slova dále snadno konvertovat a přepisovat jim jiné slovn $\acute{e}$ druhov $\acute{e}$  charakteristiky mnohdy i bez formáln $\acute{y}$ ch zm $\acute{e}$ n.

Okruh centráln $\acute{y}$ ch prav $\acute{y}$ ch prot $\acute{e}$ jšků zahrnoval celkem 64 typů, jímž v subkorpusech odpovídalo 638 tokenů. To znamená, že prům $\acute{e}$ rný počet tokenů na jeden typ dosahoval takřka deseti výskytů (9,97). Naproti tomu prvky periferní z hlediska využití v analyzovaných subkorpusech (kter $\acute{e}$  nebyly podrobeny multiaspektuáln $\acute{y}$  analýze, nicmén $\acute{e}$  jsou pro úplnost uvedeny v celkov $\acute{e}$ m záv $\acute{e}$ rečn $\acute{e}$ m přehledu v tabulce), čítaly celkem 71 typů. Jim odpovídalo v subkorpusech 214 tokenů. V tomto periferním okruhu tak prům $\acute{e}$ rná četnost tokenů na jeden typ dosahuje pouze zhruba t $\acute{r}$ í (3,01). Tudíž z tohoto okruhu nebylo možno dovodit žádn $\acute{e}$  podložen $\acute{e}$  tendence. Pokud pro úplnost výsledky obou frekvenčních zón sloučíme, v cel $\acute{e}$ m korpusu bylo zjišt $\acute{e}$ no 135 prav $\acute{y}$ ch anglo-francouzsk $\acute{y}$ ch prot $\acute{e}$ jšků-typů, jímž odpovídalo 852 tokenů. Prům $\acute{e}$ rn $\acute{e}$  tak byl každ $\acute{y}$  prav $\acute{y}$  prot $\acute{e}$ jšek zaznamenán více než šestkrát (četnost 6,31). Pokud jde o porovnání tendencí, které se týkají typů a tokenů, nejpřesn $\acute{e}$ ji je shrnují ob $\acute{e}$  připojen $\acute{e}$  tabulky v samotn $\acute{e}$ m záv $\acute{e}$ ru studie.



Nejvíce *vrais amis* bylo celkově zastoupeno v subkorpusu politického diskursu. To patrně svědčí o stereotypnosti a regulovanosti jazyka parlamentních diskusí, dále o jeho stylistické povaze, zejména o jeho formálnosti. V tomto subkorpusu byly objeveny též tři protějšky s absolutně nejvyšším výskytem tokenů v celém zkoumaném korpusu. Počet prvků, jejichž četnost překročila 10 tokenů, byl v tomto subkorpusu též vůbec nejvyšší, šlo celkově o 18 případů.

Subkorpus biblického diskursu se vyznačoval průměrným zastoupením *vrais amis*. Šlo přitom jednak o prvky takřka výlučně biblické, a jednak o prvky relativně velmi běžné. Význačný je také počet prvků s výskytem překračujícím 10 tokenů, kterých bylo celkem šest.

Subkorpus umělecké prózy se vyznačoval celkově nejnižším počtem *vrais amis*, která splňovala frekvenční kritéria. Rovněž počet těch párů, které dosáhly nejméně 10 tokenů, byl vůbec nejnižší: šlo pouze o dva případy z jádra slovní zásoby. Důvodem je patrně poměrně velmi pestrý jazyk umělecký, který se zpravidla vyhýbá přílišnému opakování prvků, stereotypnosti a ztělesňuje zcela otevřený registr.

Studie potvrdila, že pravopisná stránka pravých protějšků obvykle osciluje mezi identitou a blízkostí, zatímco výslovnostní aspekt charakterizuje spíše blízkost až relativní nepodobnost. Z toho lze dovodit, že psaný projev je poměrně přístupnější mluvcím druhého jazyka, a tudíž *vrais amis* v něm lze identifikovat snáze.

Podle teorie centra a periferie (Daneš 1966) lze tak na základě souhry kritérií pravé protějšky uspořádat na pomyslné škále od centra, s převažujícími rysy identity, přes přechod, s převažujícími rysy blízké podobnosti, až po periferii, s převažujícími charakteristikami vzdálené podobnosti ve většině aspektů / kritérií. Tento pohled z hlediska uživatele obou jazyků však doplňuje pohled statistický, podle kterého je dominujícím rysem (centráním) pouze formální blízkost pravých protějšků.

Výsledky naznačují, že zjištěné pravé protějšky byly takřka výlučně specifické pro ten který registr, neboť průniky mezi registry jsou spíše okrajové. Zdá se též, že *vrais amis* celkově nepředstavují příliš častý jev; v politickém diskursu, ve kterém byly relativně nejčetnější, pokryly 3,44% užití lexikální zásoby. Z toho lze dovodit, že se jazyky vydaly spíše samostatnou cestou a spřízněné prvky se jak po formální, ale zejména po stránce významové a stylistické od sebe v různé míře vzdalovaly.

Dále se zdá, že u uzavřených registrů tento jev zřejmě nabývá na významu, zatímco u otevřených registrů jeho zastoupení patrně klesá. Z toho lze usoudit, že by uzavřené registry pro mluvčí druhého jazyka mohly být přístupnější. Tyto tentativní závěry by však bylo třeba prověřit a potvrdit výzkumem rozsáhlejšího korpusu, pokud možno složeného z většího souboru subkorpusů.

Díky svému parolovému zaměření však i přes omezený rozsah analyzovaného korpusu tato práce přináší některá zjištění, která obohacují seznamy *vrais amis* z některých publikací na toto téma (např. Marcheteau, Dahan: 2009). Zjištěné tendence by mohly přispět k objasnění podobností mezi slovními zásobami obou zkoumaných jazyků a mohly by najít také své didaktické uplatnění. Práce by tak mohla usnadnit proces osvojování angličtiny a francouzštiny, ať již pro rodilé mluvčí některého ze studovaných jazyků, tak pro mluvčí jiných jazyků, kteří si chtějí osvojit oba zkoumané jazyky.