

Univerzita Karlova
Pedagogická fakulta

RIGORÓZNÍ PRÁCE

2017

Mgr. Mariana Pípalová

Univerzita Karlova
Pedagogická fakulta

Katedra anglického jazyka a literatury

RIGORÓZNÍ PRÁCE

Analysis of English and French True Friends (Vrais Amis)
in a Corpus of Authentic Text Samples

Analýza anglo-francouzských pravých protějšků v korpusu autentických vzorků textů

Mgr. Mariana Pípalová

Vedoucí práce: PhDr. Klára Matuchová, Ph.D.

Studijní program: Učitelství pro střední školy

Studijní obor: Anglický jazyk

2017

Prohlášení:

Prohlašuji, že jsem rigorózní práci na téma *Analysis of English and French true friends (vrais amis) in a corpus of authentic text samples* vypracovala pod vedením vedoucí 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 27.11. 2017

.....
Mariana Pípalová

Poděkování:

Ráda bych tímto poděkovala zejména vedoucí své rigorózní práce, PhDr. Kláře Matuchové, Ph.D., za její podporu, trpělivost, vstřícnost, cenné rady a podněty. Současně bych chtěla vyjádřit svou vděčnost Mgr. Jiřímu Jančíkovi za odborné konzultace z hlediska francouzského jazyka. V neposlední řadě bych ráda poděkovala Matthew Wagnerovi za jazykovou korekturu své rigorózní práce.

ABSTRAKT

Rigorózní 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 korpusu i 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 PhDr. dissertation provides a parole analysis of vrais amis (true counterparts) in current French and current 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 computer tool, 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 both, the types and the 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 lexicon 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 friends (counterparts) along an imaginary scale ranging from identity to relative difference in nearly all the criteria. The dissertation includes an alphabetically arranged list of all true counterparts generated and the lists of the findings pertinent to the individual subcorpora. The tendencies discovered 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

Contents:

1	Introduction	1
2	Theoretical part	3
2.1	Background	3
2.1.1	Historical background	3
2.1.2	Languages in contact	5
2.2	Definition of selected lexicological terms	7
2.2.1	True friends	8
2.2.2	False friends	10
2.2.2.1	The false friends in a semantic context	12
2.2.3	Partially false friends	13
2.2.4	Lexicon	14
2.3	Saussure's conception	17
2.3.1	Semiology and the social character of a language	17
2.3.2	The theory of a linguistic sign	18
2.3.3	Syntagmatic vs. paradigmatic relationships	19
2.3.4	Diachrony vs. synchrony	20
2.3.5	Lingue vs. parole	20
2.4	Functions of language	21
2.5	Text typology	21
2.5.1	Text Types	21
2.5.2	From text type to a sequence	22
2.6	Approaches towards the classification of texts	23
2.6.1	Typology of typologies for text classification	23

2.6.2	Text-external and text-internal approaches.....	23
2.6.3	Functional styles	24
2.6.4	Register	25
2.7	Translation	26
2.8	Studying the corpus	29
2.8.1	Corpus linguistics.....	29
2.8.2	The corpus.....	30
3	Practical part.....	35
3.1	Corpus.....	35
3.2	Method.....	38
3.2.1	Criteria	43
3.2.2	Illustrative Analysis of Text Samples (extracts from sub-corpora)	48
4	Results.....	54
4.1	The Bible.....	54
4.1.1	Pronunciation	55
4.1.2	Spelling	55
4.1.3	Morphology.....	55
4.1.4	Frequency.....	55
4.2	Parliament.....	57
4.2.1	Pronunciation	58
4.2.2	Spelling	58
4.2.3	Morphology.....	58
4.2.4	Frequency.....	59
4.3	Novel.....	60

4.3.1	Pronunciation	60
4.3.2	Spelling	61
4.3.3	Morphology.....	61
4.3.4	Frequency.....	61
4.4	Proper nouns	62
5	Conclusion.....	63
5.1	PRONUNCIATION.....	66
5.2	SPELLING.....	67
5.3	MORPHOLOGY	68
5.4	FREQUENCY	69
5.5	MEANING.....	72
6	LITERATURE	78
6.1	SECONDARY SOURCES.....	78
	Résumé rigorózní práce:	84
7	APPENDICES.....	90

1 Introduction

At first sight, English and French seem rather dissimilar, disparate, estranged one from another – especially from the point of view of the spoken language characteristics, for their pronunciations feature different sounds, stress and intonation patterns.

However, when looked upon more closely, numerous – by nature quite fine and delicate – similarities may be discovered, not only in grammar, but most importantly in the domain of vocabulary. In fact, lexical similarities together with some relevant morphological characteristics will be the subject of this dissertation.

It is especially in the layer of abstract terms and in the scientific terminology that some identical or nearly identical items in English and French may be detected. This similarity follows from frequent contacts between the two languages across/throughout centuries.

This project in fact continues and elaborates on the topic of the bachelor's thesis entitled *Les vrais et les faux amis en anglais et en français et leurs aspects lexicaux et morphologiques*, and the diploma thesis entitled *L'analyse des vrais amis anglo-français d'un corpus d'échantillons textuels authentiques*, in the sense that it is still concerned with the broad area of the true friends, the false friends and the partially false friends. This dissertation, however, focusses exclusively on the true friends.

The above mentioned bachelor thesis, which was conceived of from the langue perspective, suggested that the true friends were rather rare. Traditionally, they include long and abstract words forming the intellectual layers of the vocabulary. More precisely, they can be found especially in the scientific and technical spheres of the terminology. Usually, they are of use to a French speaker studying English and vice versa, an English speaker studying French, because they pose no risk of a negative transfer. Quite the opposite, they actually make it possible to benefit significantly from a positive transfer. The lexical fields where the true friends seem to be abundant include religion, administration, art, law, architecture, etc.

As the title of the present study suggests, the purpose of the following pages is to explore the true friends in English and French in authentic samples of text types. More specifically, the analysis is based on a corpus gathered exclusively for the purposes of this project. This research will concentrate on the actual representation of the true friends in the selected text samples. It will strive to find out whether the true friends take part in the modern usage of both the languages, whether they are employed equally frequently in different types of texts and whether they are easily recognisable at first sight. Moreover, the dissertation will aim to find out which of them are the most frequent.

Given the fact that objectivity of the results is of primary importance, it was decided to somehow straddle between the two languages, and thus give the two directions of translation the same attention: from English to French and vice versa, from French to English. As far as the chosen text types are concerned, it was deemed best to base the choice initially on text-external criteria.

Before the analysis itself, it seems appropriate to indicate that a rather unique perspective was adopted. This dissertation does not study the abstract system, but scrutinizes practical realizations of true friends in the texts of concrete, authentic translations. Since particularly for historical reasons the dominant tendency was for the borrowings to proceed from French to English (and only very recently this trend started to reverse, with English becoming a global language), special attention will be given to the English language and to the French influence on its lexicon.

It is deemed important to point out that this dissertation concerns itself solely with the *modern* varieties of English and French.

2 Theoretical part

2.1 Background

2.1.1 Historical background

Even though English and French differ in many respects, particularly in terms of their pronunciation features, the two languages display a great deal of similarities in grammar, vocabulary, and they belong to the same dominant language type (both being largely analytical/isolating). This may follow from the fact that historically, their destinies were closely interconnected.

Moreover, apart from the ancient (Proto-)Indo-European substrate, common to the Celts and to the Saxons, it was primarily the influence of Latin, which contributed significantly to the creation of a shared base between the two languages.

The Roman period, which lasted till the Vth century, left only a few linguistic traces behind. The Christianization, however, which started in the VIth century, is considered to be essential for the introduction and the subsequent spread of Latin. As a result, the clergy employed Latin in their writing, and this tradition was pursued for centuries.

This religious Latin had an important impact on the evolution of the (Germanic) English language, even though it is often difficult to establish whether a certain word was borrowed directly from Latin, or whether it was later mediated by old French, imposed on the Anglo-Saxons by the Normans after William the Conqueror's victory at Hastings in 1066.

Following the Norman invasion French was introduced as the official language of the English court. In fact, French retained its dominant position for several centuries during which the history of France and its neighbour across the Channel were closely interconnected.

At the beginning of the XIIIth century, after John the Lackland had lost his continental territories, the influence of Norman French declined in England, while the impact of the central (Parisian) French increased. The impact may still be traced in the English lexicon, for occasionally a word was borrowed from both, the Norman French

and the central French, for ex. *catch* (from Norman French) – *chase* (from central French).

In the XVIth century, during the New Learning and Renaissance period, Greek and Latin authors were rediscovered, and numerous Italian and French works were translated into English, which again helped to enrich the English wordstock.

It was not until a much more recent period that the notion of snobbism became associated with borrowing – including francophily and anglomanie – and that the trend of mutual borrowings was established.

The French influence on English lexis is particularly noticeable in the elevated, noble layers and domains, such as culture, gastronomy, diplomacy, politics, or education. English, on the other hand, has affected the French domains of sport and technology, allowing the French to adopt the term together with the pertinent skill or equipment.

This tendency to integrate English terms into French has increased especially in the second half of the XXth century, primarily due to the influence of the United States' economic and technological supremacy. English acquired the status of *lingua franca* of the modern world.

At present, it is particularly the European Union which has contributed significantly to the reciprocal borrowings and/or to the creation of new terms, originating from English and/or French roots.

Consequently, it does not come as a surprise that after centuries of mutual contacts and linguistic exchanges, in the two languages there are numerous words which are identical or closely related in form. However, this does not mean that they have always exactly the same meaning. In fact, they could have evolved differently over the centuries from one, usually Greek or Latin origin, or once borrowed, the receiving language could have modified (at once or gradually) the original word, by ascribing it its own meaning or by integrating it within its linguistic system (for ex. rebracketing,

i.e., resegmentation¹). A different case arises when the term had existed in the two languages already, but in one of them, it displayed an additional shade of meaning. It would then be precisely this extra sense which was adopted, thus enriching the existing range of meaning(s) of the original word.

According to Edwards (2017), “in the XVIIIth century, the French took one shade of the English word responsibility which was originally missing in French, the responsibility of a minister before the Parliament, for instance”. (translation mine)²

The topic of this dissertation can be addressed from two distinct perspectives, namely from a synchronic or a diachronic one. Secondary literature suggests that for historical reasons the majority of the borrowings proceeded from French to English. Therefore, this dissertation adopts the same direction.

It should be noted that the following pages take a synchronic approach to the subject matter exclusively. It is held that this perspective may prove useful for a current English or French speaker/user who intends to learn the other language (as a foreign language) or even for a speaker of a completely different language who strives to study both, English and French. This dissertation might facilitate the process of such learning, and it might provide some motivation as well.

2.1.2 Languages in contact

From the point of view of sociolinguistics, the contact between two (or more) languages can result in a number of cases. Sometimes, only a few words are borrowed; sometimes the rate of borrowing is more noticeable, and in extreme cases, whole new languages are formed. As Croft (2000) points out, there is a borrowing continuum which depends on the degrees of social contact between the two communities. In extreme cases, an immense borrowing rate may even cause language death.

¹ For instance the word ammunition (E), introduced into English as a result of the French soldiers having mistaken the word “la munition” (from Latin *munitio*) for “l'ammunition”.

² “Au XVIII^e 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)

The results of such language contacts differ according to several factors, including the length and intensity of the contact between the particular communities of speakers, the types of social, economic and political relationships between them, the functions that the communication between them aims to fulfill and the measure of similarity that the two languages in contact exhibit.

Hence borrowing has its prominent sociolinguistic dimension – it contributes to language change. According to Thomason and Kaufman (1988: 37, cited in Croft, 2000: 201), borrowing is the incorporation of foreign features into a group’s native language by speakers of that language: the native language is maintained but is changed by the addition of the incorporated features.”

In fact, during their histories, most languages were influenced at least to some degree by contact with others, which results in various degrees of transfer (of features from one to another).

If this is the case and the languages happen to be geographic neighbours, we speak about an *adstrat*, suggesting a mutual/reciprocal influence of two languages used in neighbouring regions.

According to Martinet (1969), “adstrat is thus the result of a particular geographical or social situation.” (translation mine)³

During its history, English has borrowed numerous words from especially from Scandinavian (i.e., Old Danish and Norse), French, Latin and Greek. Such a transfer need not be established on direct interaction between the speakers of the particular languages, since it can be mediated by writing and stem for example from book studying. The students may then adopt new vocabulary from literature, religious texts, dictionaries, and in modern times, the internet, etc.

It seems that the situations embracing language contact may generally produce two tendencies, which may be frequently in conflict. On the one hand, it is the need to

³ “L'**adstrat** est, par conséquent, le résultat d'une situation géographique ou sociale particulière. ” (Martinet, 1969)

achieve efficiency in communication, adequate for the purpose of the interaction (the dynamism of accommodation). On the other hand, it is the endeavour to maintain the group identity (the loyalty towards the group). The first tendency stimulates the convergence and the compromises between the languages, whereas the other encourages the divergence or the maintenance of language boundaries. (i.e., Bell, 2014)

2.2 Definition of selected lexicological terms

Theoretically speaking, there are different semantic relationships between the similar words in English and French; the *true friends* (les vrais amis, i.e., words of the same meaning and form), the *false friends* (les faux amis, i.e., words of similar form but different meaning) and the *partially false friends* (les faux amis partiels, i.e., words which, according to the particular contexts, can have either the same meaning as the counterpart in the other language, or a completely different one.

It is necessary to stress the fact that the true friends, the false friends and the partially false friends do not exist primarily in the language (like for example the synonyms or paronyms) – they represent a relationship between the two elements in two different languages; it is the uneven and inadequate linguistic knowledge of a speaker that produces them.

The true friends between English and French can differ one from another with respect to their extension and/or intension. Extension is understood to include “a range of individuals, etc. to which a term correctly applies” (Matthews, 123), i.e., a set of objects that the word can designate. Intension, on the other hand, involves “the properties that define the word or a concept” (Matthews, 182).

Schematically, it is possible to suggest that if a language has adopted a word form (which it could have deformed and modified to different extents), this does not necessarily mean that it has taken over all of its meanings (senses) as well. The language could have chosen to adopt only one or several of the original meanings of that borrowed word. It could have even – more or less significantly – modified or changed the original meaning(s). Moreover, languages develop constantly, and the

original meaning of the word (in the providing language) does not have to exist any longer. Over some period, new meanings can also emerge.

It might be added that both English and French have different regional and social varieties, for ex. Canadian French, Australian English, etc. Such varieties or dialects of a single language also frequently differ from one another in the number of meanings (senses) the particular word displays. For all the above reasons, this dissertation can only show correspondences in the use of modern standard English and French, based on the studied samples in the corpus, embodying the particular varieties, registers and specific co(n)texts.

Figure 1 below provides a schematic representation of a polysemic word, which has four senses in French. However, English has borrowed it only with two of the original meanings.

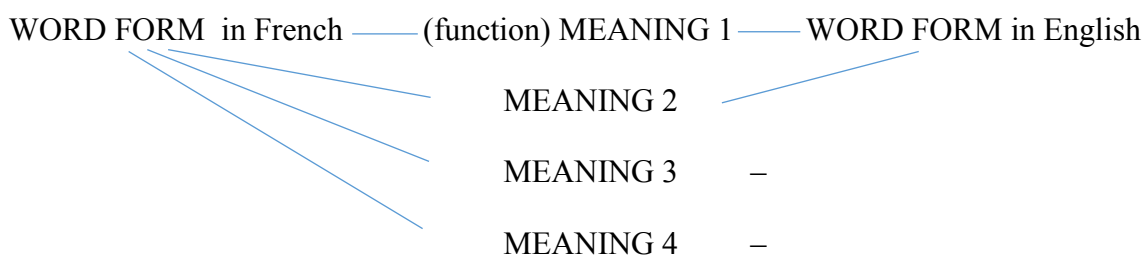


Fig.1

2.2.1 True friends

The so called true friends (in French *les vrais amis* or, according to Walter (2001), *les bons amis*) are such words which occur in two different languages, which have an (almost) identical word form (homographs), the same meaning and a similar pronunciation. Generally speaking, this group of words represents the most mutually intelligible and transparent set, for they may be considered as international or inter-language synonyms (for ex. possible (E) – possible (F)). Exceptionally, even examples originating in a third language can be found (for ex. *piano* (E, F) from Italian).

According to Marcheteau and Dahan (2009), the units in this category can be divided into three types: the “clear or evident” true friends, the “similar” true friends and the true friends only in the written language, with “misleading pronunciation”. (translation mine)⁴

As follows from the label itself, the clear true friends are those which are identical in almost every respect, for ex. : *accident, agent, cigarette, distance, elephant, hotel, moustache, noble, patience, restaurant, transport, village, etc.*

The “approximate, close” true friends bear the same meaning, exhibit an almost identical pronunciation, but a slightly different spelling in the two languages, for ex. *abbreviation (E) – abréviation (F), city (E) – cité (F), development (E) – développement (F), hospital (E) – hôpital (F), language (E) – langage (F), number (E) – nombre (F), to observe (E) – observer (F), victory (E) – victoire (F).*

Note: This can also be the case with English words borrowed by French, for ex. *bifeck (F) – beefsteak (E)*, even though they are not very common.

The true friends with “misleading pronunciation” are words whose meaning relatedness may be obscured by a difference in stress and/or pronunciation. Thus, *archive [ɑ:kɑiv] (E) – archive [arʃiv] (F), lieutenant [lef'tenənt] (E) – lieutenant [ljøtnã] (F), rhythm [rɪðəm] (E) – rythme [ritm]* can be considered as a sub-category of this group of (almost) identical words in French, which presumably do not pose a comprehension problem. Nevertheless, they are often mispronounced by the French/English speakers, for ex. *guardian ['gɑ:diən] (E) – gardien [gardjẽ] (F), promis ['prɒmɪs] (E) – promesse [promes] (F), to examine [ɪg'zæmɪn] – examiner [ɛgzamine] (F).*

Walter (2001) proposes a three-member classification as follows: *les bons (vrais) amis étrangers* (“the foreign true friends”, translation mine), *les bons amis apparentés* (“the evident true friends”, translation mine) and *les « aller et retour »* (the “go-and-return” type of true friends, translation mine).

⁴ “*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)

The origin of the foreign true friends can be traced to a third language, in this case to languages other than English or French, for ex. *anaconda, alligator, banjo, vodka, malaria, polka, calypso, kimono, ravioli, zen*, etc.

The related true friends (forming the majority) are the words “of the same etymology, where an identical form was transmitted either from French to English, or vice versa” (translation mine)⁵, for ex. *adroit, discipline, sabotage* (F → A); *dealer, leader, jockey* (A → F).

The group of the “go-and-return” type are “the words that were borrowed from France and adjusted to English, and eventually returned back to French with a modified form and with a new meaning” (translation mine)⁶, for ex. *cash* (1. “went” to (was borrowed by) English from an old French word *ca(i)sse*, 2. “returned” (was borrowed) from English back to French as *liquid money*); *budget* (1. “went” to English from an old French word *bougette*, 2. “returned” from English back to French as *state finances*); *interview* (1. “went” to English from an old French word *enterveue*, 2. “returned” from English back to French as “*public questioning*”).

2.2.2 False friends

The term was first introduced in 1928 by the linguists Maxime Kœssler and Jules Derocquigny in their book *Les Faux Amis ou les Trahisons du Vocabulaire Anglais*. The false friends, in fact, designate a group of *entirely* false friends – where using the counterpart in literal translation would be non-sensical. Basically, it is possible to call them “true false friends, in the sense that their formal appearance is similar, while their meaning is completely different. Some of these words do not owe their apparent similarities but to accidental correspondences, and are thus products of pure coincidence.” (translation mine)⁷

⁵ “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.)

⁶ “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.)

⁷ “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, a professional translator between English and French, offers a more precise and elaborated definition on his website: “the false friends are words, expressions and structures, etc. from French which resemble the English ones and which seem correct (as perfect counterparts), but in fact they are not, because a particular aspect – the meaning, the use, the position within a sentence, the construction or something else – is not the same in French and in English.” (translation mine)⁸

Thus, these are pairs of words the formal resemblance of which can mislead, confuse and mystify the translator/interpreter. Here are a few examples: *bride* (E) ≠ *bride* (F, rein), *casket* (E) ≠ *casquette* (F, cap), *to attend* (E) – *attendre* (wait) (F). These words represent in fact inter-language homonyms.

However, considering English and French, the majority of their inter-language false friends are words of Greek, Latin, English or French origin which have evolved differently in the two languages. It is even perfectly possible that at one moment of their histories, the two words carried (an) identical sense(s), but nowadays, they are significantly semantically apart, for ex. *cave* (E) – *cave* (F, cellar).

In a certain number of instances the semantic divergence can be explained by the fact that the English vocabulary has retained the original meaning, derived from Latin or another language, for ex. “persona” : *person* (E) – *personne* (F, nobody), whereas the French word has evolved further.

False anglicisms, pseudo-anglicisms or franglicisms then constitute a special sub-category on the scale of false friends: the words formed entirely or partly using the elements from the English language but generally not employed in this way in English – that is to say, displaying differences in use, (word-)class, form or meaning (translation mine)⁹ or the terms adopted/borrowed from English with a different meaning. (translation mine)¹⁰ The borrowed words thus obtain a new denotate in the borrowing

⁸ “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 – site accessed on 2nd February 2017)

⁹ “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)

¹⁰ “ou « les termes pris à l'anglais avec un autre sens.” (Humbley, 1974 : 55)

language, for ex. *campsite* (E) – *camping* (F), *crisps* (E) – *chips* (F), *parking lot* (E) – *parking* (F), *make-over* (E) – *relooking* (F), *channel hopping* (E) – *zapping* (F), *suit* (E) – *smoking* (F).

2.2.2.1 The false friends in a semantic context

Certain cases of false friends represent in fact paronyms or homonyms.

Paronymy

Paronymy is defined as a relation between words of similar form, which can be easily mistaken, and thus embody a source of potential errors. This phenomenon can be present within one language (for ex. *affect* – *effect* (E), *amener* – *emmener* (F)), or it may arise across the boundaries of two different languages (for ex. *bride* (E) ≠ *bride* (F, rein)).

Homonymy

Homonymy is in fact a formal (and random) correspondance between two or more lexemes, which obtains either within *one* language or across *two* different languages. In meaning terms, however, the two forms are completely unrelated and disparate. Moreover, their origin is never the same. Homonymy could be described as a feature typical of analytical languages. In English, it can be illustrated by *race* (competition) and *race* (group of people sharing skin colour and other physical features).

Polysemy

The notion of homonymy should be distinguished from that of polysemy. A polysemic word comprises multiple meanings and functions. Polysemy arises when a single lexeme has two or more meanings which are mutually related. Again, polysemy represents a phenomenon typical of analytical languages. (see Čermák, 2001 : 190-192)

In English, polysemy can be illustrated by *bed*, which can denote at least the following: 1. piece of furniture, 2. ground at the bottom of a sea or river and 3. layer of rock inside a larger area of rock.

2.2.3 Partially false friends

The term partially false friends or, according to Walter (2001), the partially true friends (*les bons amis partiels*), designates those pairs of words between the two languages whose word form is similar but the semantic overlap is only partial. Hence, on the semantic scale, they represent an intermediate element between the true and false friends. In fact they may be conceived of as international/inter-language homographs. More particularly, the partially false friends are characterised by at least one shared meaning (sense) and by at least one meaning (sense) that the words do not share. In practice, then, these may be either English words with the same meaning(s) as their French counterparts, but in addition to that, the English words have an extra meaning/extra meanings, or vice versa, the definition of the word is more restricted in English than it is in French, for some senses are not featured. Hence such pairs represent a source of particularly frequent errors for non-native speakers. By mistake they often employ or interpret the partially false friends in accordance with the meaning(s) the words have in their mother tongue, which can be incorrect in the other language, for example: *figure* (E) – *figure* (body, face or image, F), *distraction* (E) – *distraction* (entertainment, inadvertence, F). (Ayasch, Hillion, 2006)

All of the terms introduced above (the true, the false and the partially false friends) will be of primary importance to this project.

It seems appropriate to mention that for the purposes of this dissertation, a decision was made to adopt the *parole* (Saussure, 1997) perspective on the lexicological subject matter, thus focusing on the concrete usage of the languages. To set the research entitled *Analysis of English and French True Friends (Vrais Amis) in a Corpus of Authentic Text Samples* in appropriate context, the theoretical chapter is going to be divided into three parts. Firstly, we are going to discuss the vocabulary and the theories of linguistic sign; then we are going to deal with the functions of language and the text typologies; and finally, we are going to concentrate on the main approaches to translation and on the modern corpus linguistics.

2.2.4 Lexicon

In all languages, the lexicon consists of the sum total of words that the language comprises and their mutual relationships. In other words, it represents the system of words which exist within the particular language. This project concentrates solely on the lexical (content) words as the carriers of meaning par excellence. Evolving all the time, the lexicon belongs to the least stable language components: it adapts constantly and entirely to the current communicative needs (a *signifiant* may be needed in order to represent a *signifié* (Saussure, see also below)); if the signifiant for a currently needed signifié is missing in a language, one of the options available is to borrow it from another language where it already exists.

More specifically, in Saussure's understanding, the lexicon is a system of subsystems, a set of mutually related items which influence each other. The lexicon also displays a certain lack of balance. Schematically, it consists of a conspicuous centre, a transition, and a diffuse periphery. (Daneš, 1966) The centre represents the core of the lexicon, and is composed of expressions which are very frequent, unmarked, relatively stable, typical and stylistically neutral, while the transition and even more pronouncedly so, the periphery of the wordstock, embody the perfect opposite. The periphery is characterized by relatively infrequent and stylistically rather marked expressions, those which may be leaving the system, or may have been recently coined and therefore are not (yet) fully integrated in it. Generally speaking, the periphery may be illustrated by such lexical layers which connote literariness (for ex. the poetisms), novelty (neologisms) or which bear archaic traits (historicisms).

A constant interchange and flow is taking place between the two marked zones of usage, depending on the requirements of the current usage, communicative needs and the reality. For example, when a word becomes archaic, it is removed from the centre and is gradually attracted to, and finally integrated in, the periphery. Conversely, if a word acquires a new meaning, relevant to the modern usage, it may re-enter the central zone of the lexicon and become part of it again.

Furthermore, there are two opposing forces operating in the system: one aiming at integration and uniformity, whereas the other working towards diversification and finer distinctions/discriminations in meaning.

2.2.4.1 Word-formation processes

There are numerous word-formation processes through which the wordstock may be enriched. In English, they are divided into major and minor ones. The major word-formation processes include derivation (affixation, i.e., prefixation and suffixation, exceptionally also infixation), conversion, and compounding. Minor word-formation processes, on the other hand, involve back-formation, blending, reduplication, clipping and acronymy. (Quirk et al., 1985). In French, most of the processes are the same, even though the frameworks do not correspond perfectly. For example, in French lexicology, neologism is considered a separate word-formation process, or compounding (composition) is further distinguished into popular and scientific types. Nevertheless, for the purposes of this dissertation, the processes may be assumed to be similar, which seems to follow, among other things, from the fact that the two languages in question rank among Indo-European ones, neighbouring ones, historically and typologically related ones (their essential type being analytic (isolating)).

The word-formation processes relevant to this dissertation are especially the following:

Derivation

New words may be coined from the already existing ones by means of attaching an affix (i.e., a prefix, a suffix or, rarely an infix) to the root or stem of a given word. (Unlike the stem, a root itself is not derivable from any smaller or simple form. (Matthews)) A prefix goes before the root or stem, while the suffix follows the root or stem. A word may feature more affixes simultaneously.

Conversion

According to Quirk et al. (1985: 1558), conversion is the derivational process whereby an item is adapted or converted to a new word class without the addition of an

affix. The authors maintain that in this respect it is analogous to suffixation (unlike prefixation).

Compounding

Quirk et al. (1985: 1567) define a compound as a lexical unit consisting of more than one base and functioning both grammatically and semantically as a single word. They remark that typically the compound consists of two bases, even though in principle there may be more.

Calques

The calques represent a special case of borrowings. According to Matthews (1997), they are formed “by translation of a corresponding word or expression in another language”. For instance, the French gratet-ciel (lit. scratch-sky) is a calque on the English skyscraper.

Neologisms

Neologisms are the result of a new word creation using diverse processes, or of the act of ascribing a new meaning to an already existing word.

2.2.4.2 Borrowing

Although no word is actually formed or produced in a particular language using borrowing, borrowing is considered to be one of the indispensable sources of the vocabulary enrichment. It should be noted that the term is in fact a misnomer since the borrowed word is never returned.

Linguistic borrowing is defined as a transfer (or transference) of linguistic elements from one language to another. “An act by which a language accepts an element from another language; the element (word or expression) thus becomes incorporated.” (translation mine)¹¹ Similarly, Matthews (1997) delimits it as a

¹¹ “Acte par lequel une langue accueille un élément d'une autre langue ; élément (mot, tour) ainsi incorporé.” (Petit Robert, édition 2008)

“conventional term for the introduction into language *a* of specific words, constructions, or morphological elements of language *b*.”

According to Crystal (1995: 126), English is an insatiable borrower. The author estimates numbers of loanwords historically borrowed by English from diverse languages, stating that there are only a few Celtic words in English, around 2, 000 words from Scandinavian languages, but around 100, 000 words from French. Moreover, by the end of the Renaissance period, the borrowings from classically-derived languages, including Latin, have doubled the size of the English lexicon. Interestingly, as Crystal maintains, the process of borrowing has been active even since 1950s, when English achieved the status of a lingua franca.

As follows from the topic, borrowings will be in the centre of attention of this dissertation.

2.3 Saussure's conception

2.3.1 Semiology and the social character of a language

At the turn of the 19/20th centuries, Ferdinand de Saussure, the father of structural linguistics, introduced his theory of semiology, a discipline studying the life of signs in a human society. In his view there are many sign systems available for a society, nevertheless language is essential. Therefore, he treats linguistics as part of semiology. Saussure considers language as a system of signs for expressing ideas, and stresses its social character. It should be noted that semiology has recently undergone some development and has come closer to other disciplines. Also, the name of the discipline is gradually being replaced by *semiotics*.

Saussure has introduced a number of indispensable terms into modern linguistics, including the concept of sign, the paradigmatic vs. syntagmatic relationships, synchrony vs. diachrony, and langue vs. parole.

2.3.2 The theory of a linguistic sign¹²

The theory of linguistic sign, which has become indispensable in modern linguistics, occupies a prominent place in the *Cours de la Linguistique Générale*, 1916 (*Course of General Linguistics*). This theory has provided the foundations of modern semantics.

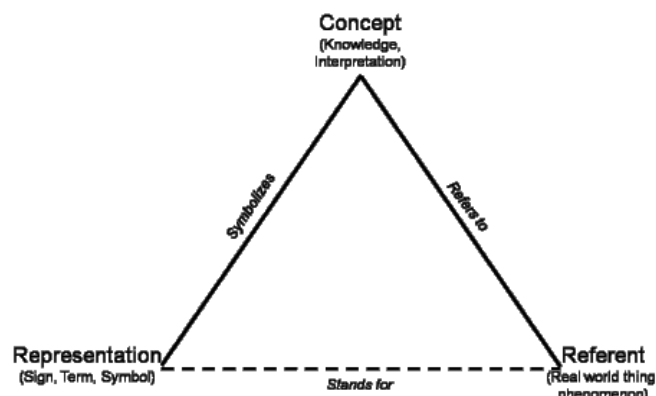
Saussure understands the linguistic sign as an entity uniting the “concept” and the “acoustic image”. Contrary to the idealist conception, he does not believe that a sign relates a thing and its name. He argues that the two components are of psychological nature, connected by the association in the brain.

For instance, the acoustic image is not interpreted as a purely physical sound, but as a psychological imprint in our brain. This can be affirmed by the fact that we are capable of talking to ourselves without making a sound.

“The concept” did not exist independently before, for the two aspects of a sign have formed simultaneously and are interconnected.

Given the fact that the term “sign” can be easily mistaken for an equivalent of the “acoustic image”, Saussure introduces new terminology to avoid incorrect interpretation. Instead of the “acoustic image”, he employs “signifiant” and for the

¹² It is necessary to mention that recently, linguists seem to prefer the model of a semiotic triangle developed by Ogden-Richards, which integrates also the extra-linguistic reality into the conception of a linguistic sign. (see for ex. Čermák 2001: 24)



“concept”, he introduces “le signifié”. These two components together then produce the linguistic sign.

It needs to be stressed that Saussure understands sign as a two-sided psychological entity uniting the concept with the sound-image.

Moreover, Saussure defines three basic features of a linguistic sign: arbitrariness, linearity and difference.

The arbitrariness concerns the relation between the two components. In fact, there is no natural and internal relationship between the *signifié* and the *signifiant*, which can be confirmed by the fact that different languages have different signifiants for the same signifié (for ex. horse (E), cheval (F)).

The linearity of linguistic signs indicates that the communication takes place in time. Consequently, it is necessary to arrange the signs successively, one after another.

Finally, the discontinuity (or the difference) suggests that there are discrete sign units, although the sound aspect of the language is amorphous. It is not until a number of particular sounds are chained together that the signifiant is established.

Having introduced the notions of signifié and signifiant, Saussure presents *the value (la valeur)*. In his view, the value follows from the particular link between the signifiant and the signifié, which distinguishes the sign from all the other signs in the system. Hence the individual signs are characterised primarily by what differentiates them from the other signs. It is precisely this difference (contrast) that produces their unique value.

2.3.3 Syntagmatic vs. paradigmatic relationships

Saussure maintains that signs enter into two types of relationships, namely syntagmatic and paradigmatic ones. The former are the relationships which are produced by combining signs successively, i.e., the concatenative ones, whereas the latter are established between signs of the same category, based on oppositions, associations or substitutions.

2.3.4. Diachrony vs. synchrony

Saussure gives two different (and complementary) points of view which can be adopted in linguistics: synchronic and diachronic. The synchronic approach is static, and studies the language of a given period of its existence. In contrast, the diachronic approach focusses on the evolution of the language throughout its history. It may be added that the term synchronic is often used to denote the present state of a language.

2.3.5. Langue vs. parole

Furthermore, Saussure differentiates between *langue* and *parole*.

Langue is understood as the abstract system of rules and conventions which is internalized by a particular language community. The “system” is interpreted not only as a collection of elements, but as a structured set, composed of elements, their relationships and the rules governing them.

Parole, on the other hand, is the concrete manifestation or realization of langue means. Parole is in fact the result of langue in actual use, and corresponds to both the written and spoken language portions.

According to Saussure, langue is common to a language community and represents the result of a social convention, imposed by the society on an individual. Conversely, parole is the personal use and concrete realisation of langue, and displays various personal characteristics: style, rhythm, syntax, pronunciation, etc.

It is needless to add that each word represents in fact a sign. Therefore, the concept of sign will be essential for this dissertation. Given the fact that this analysis is concerned with the actual use of words (signs), some attention will now be given to various functions the language fulfills.

2.4 Functions of language

In his well-known framework, Roman Jakobson (1963) distinguishes six language functions, i.e., referential, expressive (emotive or affective), conative, metalinguistic (metalingual or reflexive), poetic and phatic. Nevertheless, he maintains that messages seldom fulfill only one function.

Jakobson's referential function represents the primary language function, as its task is to describe the extralinguistic world (referents). The emotive (expressive) function conveys the speaker's attitude towards what s/he communicates. The conative function is oriented towards the addressee as it aims to induce a certain behavior on their part.

The metalinguistic function is instrumental in using the language to depict itself. The poetic function concerns the aesthetic aspects of the language. And lastly, the phatic function aims at establishing or maintaining the contact between speakers. (Jakobson, 1963: 209-248)

2.5 Text typology

Prior to choosing the most convenient text types for this research, it seems important to introduce an adequate theoretical background for their classification. However, this is by no means a simple task, for numerous (and frequently mutually conflicting) conceptions dealing with text typology have been proposed by different linguistic schools. Therefore, the following brief overview cannot but be very selective.

2.5.1 Text Types

Most linguists maintain that each text reflects a particular purpose or intention. In this respect linguists identify between five and eight essential text types.

For example, in his well-known study, Werlich (1976, see Esser: 2009) distinguishes between five basic text-types which he associates with the typical cognitive processes they involve. In his approach, description is associated with the cognitive process of perception in space, narration with perception in time, exposition is established on the process of comprehension, argumentation on judging, and instruction

is linked to the process of planning. Other linguists have expanded Werlich's list by additional categories, such as didactic, rhetorical or dialogical text-types.

2.5.2 From text type to a sequence

Apart from linguists who strived to classify texts using consistent criteria, there were also others who realized that texts were by no means always monolithic in nature.

For example, Adam (1987) bases his conception on the fact that every text is in fact heterogenous. He defines the text as a place where “there immediately appears heterogeneity and complexity which seem to be constitutive of discourses in a natural language.” (translation mine)¹³ Furthermore, Adam introduces the notion of a *sequence* “to designate a segmentation which delimits the heterogenous text spaces, forming a part of a greater textual collection – a heterogenous one but allowing itself to be read as homogenous.” (translation mine)¹⁴ A sequence, and even more pronouncedly a complete text, suggests only one actualization more or less approximating the prototypical model of reference. Adam distinguishes between a dominant (*dominant*) sequence and a dominated-on (*dominée*) sequence, its subordinate category, which allows him not to perceive the structure of a text as a linear succession and to demonstrate the globally hierarchizing phenomena of the text. According to Adam (1992), individual text types translate cognitive processes, the modes of interpreting the reality, giving a typifying structure to this view. These text types are characterized by a dominant in their sequential organization. According to Adam, there are five prototypes of textual sequences: narrative (story); descriptive (description); argumentative (argumentation); explanative (explanation); dialogic (dialogue).

¹³ “apparaissent immédiatement une hétérogénéité et une complexité qui semblent constitutives des discours en langue naturelle ” (Adam, 1987 : 51)

¹⁴ “ 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)

2.6 Approaches towards the classification of texts

2.6.1 Typology of typologies for text classification

Furthermore, some authors noticed that the term “text type” was given diverse meanings and could be established on very different criteria. For the sake of transparency, for example, A. Petitjean (1989) offers the following typology of text typologies, comprising four major type families:

1) *Propositional* typologies, inspired by Benveniste (1966, 1970), which are based on the analysis of the relationships between the two interlocutors and the temporal and spatial setting. In this case, special attention is paid to the linguistic operations and the markers by which the communicative situation imprints itself onto the proposition: personal and deictic pronouns, the system of tenses, etc.

2) Typologies resting on *communicative* or *functional* criteria adhere to Jakobson’s (1963) six main language functions: referential, conative, phatic, metalinguistic, expressive and finally, the poetic function (see section 2.4).

3) *Situational* or *institutional* typologies are those whose inspiration is essentially sociological, distinguishing between the journalistic, religious and other types of discourse.

4) Typologies based on the pertinent *cognitive structures*. For example, Werlich (1975) differentiates between five major text types established on the related cognitive processes (see section 2.5.1 above).

From the aforementioned, in this dissertation, more attention will be given to the second and third families.

2.6.2 Text-external and text-internal approaches

Another general typology of approaches to text classification has been provided by Esser (2009), who distinguishes crucially between text-external and text-internal approaches.

In his treatment, text-external approaches are based on the distinctions which, by their nature, are not intrinsically linguistic, and therefore even a lay person or a librarian can assign texts to diverse genres, such as to an editorial, face-to-face conversation or a professional letter.

Text-internal approaches, on the other hand, are further divided into those which are dependent on, and those which are independent of, the medium. Among those which are independent of the medium, Esser includes the concept of register and that of the text type in Finegan and Biber's (1986, cited in Esser 2009) understanding. Nevertheless Esser maintains that there is a progress of sophistication from genre to register and to text-type (in the sense adopted by Finegan and Biber). In fact, register is a semantic concept (text-internal configuration of meanings) which relates both, to text-external situational parameters, and to the text-internal cluster of language features (realizations). On the other hand, the so called text-type (according to Finegan and Biber) assigns samples to text categories on purely linguistic grounds, ie., solely on the basis of co-occurrences of linguistic features, chiefly morphological and syntactic ones, and this task is accomplished by a specialized computer programme.

Finally, Esser recognizes the text-internal approaches dependent on the medium, where the graphic or the oral presentation constitutes a decisive factor for the attribution of a concrete text type for example to a funeral notice or a pop song.

Since this dissertation will adopt a functional approach, more attention will be now given to two relevant concepts, namely to those of functional style and register.

2.6.3 **Functional styles**

The notion of functional style was introduced in the *Prague linguistic circle*. The origins of functional styles and languages may be traced to the beginning of the thirties, in B. Havránek's *Úkoly Spisovného Jazyka a jeho Kultura*. Based on the primary functions of standard language, Havránek identifies the fundamental functional dialects as layers of langue means (conversational, workaday/matter-of-fact, scientific and poetic) and relates them to the parole concept of functional styles. Havránek maintains that functional style is determined by the specific purpose of the given verbal response

and identifies a number of functional styles. His framework is established on two criteria, namely the specific purpose (e.g., matter-of-fact communication, exhortation, general explanation, technical explanation and codifying formulation) and on the manner of presentation (oral vs. written; private vs. public; monologue vs. dialogue). This functional approach to style became the foundation for further elaboration by other members of the Prague Linguistic Circle (e.g., Jedlička, Trávníček, and others). At present, the number of functional styles distinguished by authors varies but may include the following categories: technical/scientific, administrative, journalistic, rhetorical, colloquial/conversational, literary and essayistic. Among the authors elaborating on Havránek, Vachek (1991) based his framework of functional styles on three dimensions, namely the approach to subject matter, the approach to recipient and the medium. The first criterion establishes whether the content and form are regarded equally or whether one concern prevails over the other. The second criterion looks at the measure of bias, distinguishing texts which are plain from those that are held to be propagandist (not in the political sense, though). Lastly, the third criterion distinguishes between spoken and written samples.

2.6.4 Register

In his work *Language, Context, and Text: aspects of a language in a social-semiotic perspective*, 1989, M.A.K. Halliday defines the context of situation using three features, i.e., field, tenor and mode.

He introduces the notion of “language functions”, i.e., the functional components of the semantic system of a language and discriminates between the ideational function (divided further into logical and experiential functions), the interpersonal function, and the textual function.

Finally, he states that the two notions are closely interconnected in the sense that there is a correlation between the categories of situation and those of the semantic system. Generally speaking, the field is reflected in the experiential meanings, the tenor in the interpersonal meanings and the mode in the textual meanings. Conversely, the

experiential meanings are activated by the features of the field, the interpersonal meanings by those of the tenor and the textual meanings by the features of mode.

For Halliday, register is a semantic concept. It is possible to define it as a configuration of meanings typically associated with a particular situational configuration of field, tenor and mode. However, due to the fact that it is a configuration of meanings, a register has to include expressions, i.e., lexicogrammatical and phonological features, which usually accompany or realize these meanings. A register is therefore a variety according to use.

Halliday (1989) maintains that registers range from the more closed and restricted ones to those which are more open and open-ended. The closed registers are marked by a fixed and limited number of possible meaning choices, for ex. the International Air Language. In this type of register, there is virtually no place for individuality or creativity.

As Halliday argues, a more open register is much more creative, much less constrained in the range of meanings it can convey. Such registers may be illustrated by face-to-face interaction. Nevertheless, even very open registers are never totally open and therefore there is never a complete freedom of choice of language resources.

According to Esser (2009), registers are sometimes simplified and interpreted as one-dimensional varieties, established solely on a single criterion, for ex. formal French, spoken English, etc.

2.7 Translation

The translations of fiction traditionally conformed to and served primarily literarily aesthetic interests. It was only in the second half of the XXth century that the linguistic approach to the subject matter was introduced. Initially the two paths co-existed side by side and, from time to time, they competed. Eventually, their status has reached equality in an integrated method which takes into consideration mainly the pragmatic aspect.

Since the 1960s, numerous works have been written on the linguistic approach to the theory of translation. The question of adequate representation was considered a serious problem of translation. It was deemed possible to transfer all the information from the original text (in the source language) into the target language despite the differences in the grammatical systems of the two languages.

Today, the functional approach is considered as the guiding principle in translation. This means that it is not important whether exactly the same language means are employed. What counts is whether they realize the same function and more importantly, whether they do so in all the aspects (modes) – not only in view of the denotative and referential meaning, but in the expressive, associative and pragmatic ones as well.

Naturally, the semantic component (the content) of the text is crucial. Whereas the content is expressed by lexical elements, the grammatical system provides the relationships between them. Apart from the denotative information, there is also the connotative information (which follows chiefly from the stylistic and expressive colouring of the expressions employed in the text).

In order to be adequate, a quality translation has to meet certain criteria. It is necessary to ensure that the recipient of the translated text perceives the product in its entire complexity, as a unit producing the same effect in the recipient as the original. In fact, the recipient should not even be aware of the individual phases, decisions or dilemmas during the process of translation.

A quality translation has to adhere to the following criteria :

- a) the linguistic expression is natural and entirely organic in its effect
- b) the final product in the target language conveys the same meaning and has the same effect on its audience as the original
- c) the linguistic expression in the target language retains the dynamism of the original text.

According to Knittlová (2010), the criteria listed above lead to successful translation into the target language, which, at the same time, does not misrepresent the

original. Furthermore, it is necessary to ensure semantic (informational), stylistic and hypersyntactic equivalence (i.e., a successful flow of illocutionary acts).

Jakobson (1971: 232-9) distinguishes three types of translation as follows:

- intralingual translation or *rewording* is an interpretation of verbal signs by means of other signs of the same language – this may be the case of repeating, reformulating, quoting or paraphrasing;
- intersemiotic translation or *transmutation* is an interpretation of verbal signs via the means of the signs of non-verbal sign systems which can mean for ex. the linguistic interpretation of on-screen data;
- interlingual translation or *translation proper* is an interpretation of verbal signs by means of some other language (which, in our case, is the most important type of translation);

Knittlová (2010) introduces four sub-types of the interlingual translation, the first two focusing primarily on the forms, whereas the other two stress mainly the content:

- the interlineal translation, considered to be the extreme case of literal translation; this type is difficult to follow, for it preserves the linguistic information exclusively
- the literal (slavish) translation, which is grammatically correct, but translates individual lexical elements without giving adequate attention to their contextual use, to their collocability
- the (unduly) free translation, which is in perfect opposition to the interlineal translation, for it respects the source text only approximately/vaguely; it may ignore the register, stylistic, aesthetic features or connotative meanings, and thus has the tendency to convey a false, imperfect or even impoverished representation of the original
- the communicative or idiomatic translation is closely related to the pragmatic aspect of translation; it is necessary to employ it especially in the domain of various conventional formulae (for ex. the greetings)

2.8 Studying the corpus

2.8.1 Corpus linguistics

Corpus linguistics originated in the Anglophone countries. And although today this approach has become rather widespread, the Anglophone countries still remain at the centre of this type of empirical research.

Corpus linguistics has recently received a lot of attention. Although originally it was considered only a marginal approach, used primarily in English linguistics (and especially, in the domain of its grammar), it has gradually acquired a more prestigious and influential status. At present, it represents a widely accepted approach, which has the potential of giving answers to diverse questions of linguistic research. Moreover, corpus linguistics has become gradually multilingual, with a lot of languages and language varieties being studied through the corpus data.

Stimulated by the spread of computers, corpus linguistics currently represents a new approach to the investigation of language. It is the information technology which has facilitated the storage and the analysis of linguistic data available in electronic form. It has provided a new perspective on language, representing a central aspect to corpus linguistics.

The corpus studies aim at performing language descriptions and analyses established on the observations of language usage, spoken as well as written, which ranks corpus linguistics among the parole approaches according to Ferdinand de Saussure.

2.8.2 The corpus

Definition

Etymologically, the term “corpus” originated in Latin, bearing the meaning of “body”. In linguistics, corpus can be defined as “an electronic collection of texts assumed to be representative of a given language, dialect or other subset of a language, to be used for linguistic analysis”. (Francis, 1982: 7, cited in Kohnen Mukberjee, 2011)

According to Sinclair (1996: 5), a computer corpus is “a corpus which is encoded in a standardised and homogenous way for open-ended retrieval tasks.”

Electronic corpora

Elaborating on the the notion of corpus, Atkins et al. (1992: 1) distinguish four types of *text collections*, namely archive, electronic text library (ETL), corpus and subcorpus. In their treatment, archive is understood as “a repository of readable electronic texts not linked in any coordinated way.” As for the electronic text library, this is defined as “a collection of electronic texts in standardised format with certain conventions relating to content, etc., but without rigorous selectional constraints.” In the same study, a corpus is delimited as “a subset of an ETL, built according to explicit design criteria for a specific purpose.” And finally the subcorpus is defined as “a subset of a corpus, either a static component of a complex corpus or a dynamic selection from a corpus during on-line analysis.” (1991: 1)

Corpus linguistics is closely connected with the empirical analysis of linguistic data, which are accessible in electronic form. In fact the relevant softwares have made it possible to process huge linguistic data produced naturally and spontaneously, which was not possible before. As Biber et al. (1998: 4) state, “computers make it possible to identify and analyse complex patterns of language use, allowing for the storage and analysis of a larger database of natural language than could be dealt with by hand.” Naturally, the manual treatment cannot compete in efficiency, reliability, and precision with processing performed by a computer.

When discussing the corpus-based analysis, Biber et al. (1998: 4) give the following key features:

- “It is empirical, analysing the actual patterns of use in natural texts;
- it utilizes a large and principled collection of natural texts known as a “corpus” as the basis for analysis;
- it makes extensive use of computers for analysis, using both automatic and interactive techniques;
- it depends on both quantitative and qualitative analytical techniques.”

The new development has undoubtedly introduced a new dimension to the empirical research in linguistics, offering a valuable insight into the real usage of languages.

Types of linguistic corpora

A linguistic corpus can serve numerous purposes and research objectives. That is why there is a whole range of corpus types. Even though there are almost “as many types of corpora as there are types of investigation” (Bowker and Pearson, 2002: 11), a number of attempts have been made to classify them. The corpora may differ in many respects, such as in their size or in the number of the involved languages.

The present chapter strives to give attention especially such types of corpora which are relevant to the research. Therefore, it will be selective, deliberately paying no attention to certain categories, such as learner corpora, etc.

A systematic account of corpora types is provided by Bowker and Pearson (2002) who distinguish the following corpora:

A1 Diachronic corpus

Diachronic corpora, unlike the synchronic ones, cover long periods of time. More specifically, they assemble texts from different periods (for instance spanning centuries) to discover the language evolution and to pursue the particular language change(s) (such as changes in spelling, changes in meaning, etc.).

A2 *Synchronic corpus*

Conversely, the synchronic corpora focus on a particular, relatively short period of time, typically up to 50 years. Typically, a synchronic corpus contains samples of the current language usage (at the time of the corpus creation).

B1 *General reference corpus*

This type of corpus lacks a specific purpose and therefore may be employed to a variety of observations and research objectives. It does not show any preference as regards linguistic domains, styles or layers of language. Moreover, the main feature of this corpus type is its considerable size and an exceptional balance in the representation of various genres, registers, text types, etc.

B2 *Special purpose corpus* (or else specialized corpus)

However, the majority of corpora are assembled for specific and concrete purposes of research and are referred to as special purpose (specialized) corpora. Such corpora focus on the particular area the researcher wants to explore, such as on a specific domain, a particular text type or language variety. Thanks to its specialized nature, this type of corpus cannot be used for making general observations about the language use.

When correlated, a general reference corpus and a special purpose corpus may help identify the differences between a specialized and general language usage.

C1 *Spoken corpus*

A spoken corpus consists of transcriptions of spoken material (face-to-face conversations, telephone calls, etc.)

C2 *Written corpus*

This type of corpus comprises texts written to be read.

It should be noted, however, that certain corpora, such as the *British National Corpus*, represent a combination of the two, ie., spoken and written data.

D1 *Monolingual (national) corpus*

Monolingual corpus contains texts produced exclusively in the chosen language.

D2 Bilingual or multilingual corpus

Contrary to a monolingual corpus, a multilingual one assembles the texts produced in at least two different languages.

Multilingual corpora can be further divided into parallel and comparable corpora. A parallel corpus is composed of texts translated into one or more other languages. A comparable corpus does not contain any translations, but consists of texts written in two or more distinct languages, meeting particular criteria. In most cases, they discuss the same subject matter, share communicative functions, text type characteristics (instruction: manual, technical report, etc.) and period of origin.

E1 Open corpus

An open corpus (sometimes referred to as monitor corpus) is constantly being expanded. Since this type of corpus is continually upgraded, it is especially useful for the research in lexicology and lexicography areas, for example to study neologisms.

E2 Closed corpus

A closed corpus remains fixed, static, and the data it includes are finite. That means that it is not subjected to any expansion, modification or upgrading.

Moreover, other authors introduced additional types of corpora. For example, Laviosa (2002: 35) discriminates between Full Text corpora and Sample corpora. Whereas the former include unabridged texts, the latter are composed of text passages selected according to clear principles. For example, the size, the location and the method of selection may be provided. Furthermore, she identifies general and terminological corpora. The former “are made of texts assumed to be representative of everyday, non-specialized language”, whereas the latter include texts “originating within specialized subject fields, e.g., chemistry or geology, which are characterized by heavy use of recurring terms.”

Texts as components of a corpus

Corpora are compiled in order to provide representative samples of a particular language usage. Hence their collection has to be principled and has to be established on criteria. Rizzo (2010) gives several criteria, namely representativeness, size, language variety, type of text, genre and chronology.

To meet the first factor, the corpus should ensure representation of the characteristic and typical features in the particular area or language community. As the author admits, however, this can always be only approximate. Considering the size, the texts may be employed in their entire, original length or the corpus may be composed of text samples. Since noticeable differences have been found across texts of different text types, registers or genres, the corpus should embrace a range of such textual varieties. As for chronology, this criterion gives the span of time during which the texts were produced.

Similarly, Bowker and Pearson (2002: 9) hold that corpora should be composed of authentic texts which should be available electronically, their size should be adequate to the purpose of research and finally, the texts should be representative of the language or variety in question (i.e., not selected randomly).

3 Practical part

3.1 Corpus

The corpus upon which this dissertation is established, comprises approximately 60,000 words, and consists of three sub-corpora. These three sub-corpora illustrate three different registers according to M.A.K.Halliday. In fact, they can be arranged along an imaginary scale ranging from relatively closed to relatively open registers: the political discourse (which represents the most closed of all), the religious discourse (embodying a transition between closed and open registers) and the literary discourse (prose)¹⁵, being the most open of all.

According to Esser (2009), the register classification ranks among frameworks established on text-internal criteria. Nevertheless, it systematically relates features of the context of situation (text-external features) with typical text-internal meaning configurations, the latter being in turn realized by language means. This is also why, unlike purely linguistically oriented approaches (such as Finegan, Biber, cited in Esser 2009), the concept of register was prioritized as a more balanced approach. Furthermore, this framework was also selected for it was able to take into account the crucial domains of cultural contacts between the two languages and relate them to configurations of meanings in the two languages. Moreover, the concept of register was employed for it seemed capable of capturing various fine differences on the imaginary scale from closedness to openness, which corresponded also to gradual increase in the share of individuality and creativity.

The notion of a register is established on three features of the context of situation, namely field, tenor and mode. As far as the text samples scrutinized in this project are concerned, the most remarkable differences can be traced in the area of field reflected in experiential meanings, due to the fact that the samples deal with completely distinct topics, namely with matters of political, religious and artistic nature. On the other hand, the samples share numerous characteristics in mode reflected in the textual

¹⁵ Here we could employ the term of the Prague Linguistic Circle, the poetic functional style comprising the poetic or literary function.

meanings, for in all cases, the texts display both, features of literacy as well as orality, but their channel is always graphic. It should be noted, nevertheless, that different proportions of these two zones may be traced. For example, the novel displays the greatest measure of orality in this respect, as it frequently employs direct speech. Regarding the tenor reflected in the interpersonal meanings, the text types are rather or entirely formal (the novel can be classified as "normal" in this respect, while the Bible as rigid or at least very formal). The samples include primarily interaction between people or participants in the situation of communication, but secondarily, they also reflect the relationship between the prospective reader and the author (or else the editor).

Furthermore, the sub-corpora were analysed as "mirror" or parallel translations, in English¹⁶ and in French. Therefore, the corpus embraces interlingual translation exclusively. These texts were chosen also because there were different expectations concerning the faithfulness of the translation. It was assumed that the biblical text would be the closest to its original (thus verging on the literal translation), the translation of a novel would be the most liberal and free, and finally, the political sample would be marked by communicative and pragmatic features. The corpus was assembled in view of the balance between the distinct approaches to translation.

Given the fact that English is the primary borrowing language, the basic length was always established for the English version of the samples, and was set at 10 000 words. (Since the translations often vary in length, the size of the French samples may be slightly different.)

All the texts assembled in the specialized corpus are relatively recent, for none is older than 30 years. Moreover, all of the samples are available online¹⁷:

⁵ We might add that the varieties of English analysed in this dissertation are as follows: *British* in case of the political and religious subcorpora, and *American* in case of the artistic subcorpus. This choice was made in order to provide an even representation as far as varieties are concerned.

The Bible : the Darby version (Matthew's gospel)

<http://unbound.biola.edu/>¹⁸

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

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

a selection of negotiations of the European Parliament (1996 – 2011)

<http://www.statmt.org/euoparl/>¹⁹

the novel *The unbearable lightness of being* by Kundera²⁰

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

KUNDERA, M., *The Unbearable Lightness of Being*, New York City: Harper Perennial, 1999, 166 pages.

These samples may be regarded as complex texts, for all of them comprise diverse passages (sequences, see Adam, 1987), which alternate in them, for ex. in the literary text, the primary function seems to be the expressive one, although narrative, contemplative or descriptive passages may also be traced. Similarly, the Bible subcorpus features above all the conative function, but it is possible to encounter narrative, descriptive or argumentative passages in it. Finally, the political register serves chiefly the referential function, but includes argumentative, descriptive, instructive and phatic sequences.

All of the chosen texts exhibit traits of both, the written and spoken language (all of them comprise direct speech), which is also why they can be viewed as complex. Due to the absence of flaws, the texts were apparently edited so that the final product could comply with the norms governing literacy.

¹⁸ <http://unbound.biola.edu/> The website was accessed on the 14th February 2017.

¹⁹ <http://www.statmt.org/euoparl/>. The website was accessed on the 10th February 2017.

²⁰ The translations of the novel *L'insoutenable légèreté de l'être* (Kundera) are available in the pdf format.

The corpus employed in this thesis can be described as synchronic and recent (A2), specialised – assembled for the purposes of the present analysis (B2), written (C2), bilingual (D2) and closed (E2).

Moreover, it represents a compiled corpus, composed of samples of authentic texts – parallel translations – which are all available online. Two of the subcorpora exhibit some features of terminological corpora, whereas the literary subcorpus does not (Laviosa 2002). The size of the subcorpora meets the criterion of a uniform length which in this project was set at 10, 000 words for the English versions.

The text samples under analysis were chosen as representatives of diverse sociolinguistic domains, which, according to secondary sources, were exposed to French-English inter-language and intercultural contacts. These domains include among others precisely religion, administration and art.

Finally, it should be mentioned that all the texts were analysed using the AntConc 3.4 2014 corpus tool which seemed convenient, for it was designed to handle the samples in the two languages in question.

3.2 Method

At the outset it has to be pointed out that this study does not adopt a diachronic view; it focusses exclusively on the synchronic perspective. Therefore, it was assumed that its results might be also of use for foreign language teaching methodology.

Furthermore, the phenomenon was studied from the parole viewpoint exclusively. Thus the dissertation aims at observing the actual use of the lexical elements in question in various registers (the text typology being based on text-internal criteria – see Esser (2009), although correlating meaning configurations systematically with features of the context of situation).

Given the fact that it is especially the pronunciation which can affect (and at times even preclude) the comprehension, the present corpus contains exclusively written (published, printed) texts, traditionally associated with rather formal standard language.

The texts find themselves at the boundary between orality (spokenness) and literacy (writtenness), as they exhibit characteristic patterns of both (for instance in the direct speech passages), the medium, however, being always graphic. Therefore, they were all presumably adjusted to the norms of writing. All the samples involved in the corpus were assumed to represent well-formulated, official translations. All of the text samples are relatively recent, as none of them was published more than thirty years ago. Moreover, they are all available online.

For the purposes of this dissertation, three different text samples were selected in order to illustrate various registers: the religious (the Darby Bible, Matthew's gospel), the political (a collection of the European Parliament's negotiations, taking place between 1996 and 2011) and the fiction/artistic (prosaic) one (the novel *The Unbearable Lightness of Being* by Kundera, 1990 and 1999). All three text samples were analysed as parallel (mirroring), English and French translations, with the aid of the relevant corpus tool, the programme AntConc 3.4 2014, which seemed easy to operate. Moreover, this programme is suitable for analyzing both English and French.

Based on secondary literature, a hypothesis was formulated that English would take the role of a primarily borrowing language. Therefore, at the beginning of the research, a decision was made to take the English versions as default ones and to establish the size of the individual text samples consistently on the English versions. Hence all of the English translations comprise 10, 000 words, while for the French counterparts, such size is only approximate. In fact, the length of the French translations does not have to correspond perfectly and precisely with the word number of the English versions. For the sake of transparency, the size of the samples is specified below:

Bible :

the English translation : 10, 000 words,

the French translation : 9, 857 words.

Parliament :

the English translation : 10, 000 words,

the French translation : 10, 547 words.

Novel :

the English translation : 10, 000 words,

the French translation : 10, 729 words.²¹

Using the AntConc corpus tool, particularly its “word list” function, a list of all the words employed in the text was generated. Subsequently, all the grammatical (function) words were eliminated. In the next step, from among the lexical word lists, the first 100 relatively most frequent cases of the true friends in the two lists were collected. In so doing, the order of words (true friends) suggested by AntConc was naturally retained, and hence the words were again arranged with gradually decreasing frequency.

Finally, two sets of results accomplished for English and for French were mutually compared in view of several factors: first, regarding their relative frequency in the register samples; secondly, their translation in concrete passages of the texts was examined (thus verifying the correspondence in the particular co(n)texts). Last of all, account was only taken of instances which featured the same word class in both translations in the same co-texts.

In the following stage, only those words whose frequency or presence in the English version amounted to, or exceeded, 5 tokens were scrutinized further (the English version was held to be default; for the purposes of this dissertation, words with lesser frequency in the English version were disregarded). Admittedly, this frequency of five tokens was established arbitrarily; nevertheless it was introduced in order to ensure better understanding of the tendencies, for the number of 5 tokens in fact corresponds to at least 0.05% of the entire (English) register subcorpus.

²¹ Due to the fact that the English text samples feature exactly 10, 000 words, a decision was made to take the English versions as the default ones.

This analysis is based on the assumption that if two words related in form and meaning exist in the two languages, it should be natural for the translator to employ them in the same co-text of the same register. Since the dissertation investigates the actual usage, it deals exclusively with those instances where it was possible to find the matching cognate in the particular contexts (the cases where it was not possible to find such a match were disregarded for the purposes of the present project). Hence wherever such counterparts were actually employed by the professional translator, this was seen as confirmation of a perfect match of such items in the co(n)text of the respective registers of the two distinct languages.

As a result, two categories of true friends were generated, deemed relevant to this work: the category of units with *identical* frequency (comprising true friends employed with exactly the same frequency in the two translations), and the category of *similar* frequency (referring to true friends employed in approximately the same way, for there was a slight difference of up to two tokens in the parallel translations (i.e., the number of tokens was always five or more in the English samples, whereas the French versions displayed +/- 1 or 2 tokens).

In the research, two categories had to be disregarded. First, the item did not fulfill the minimal frequency criterion (i.e., the English samples failed to include at least five tokens); secondly, the number of the English tokens did meet the criterion, but the number of the French counterparts exceeded the moderate difference in frequency of up to +/-2 tokens. For these instances, it seemed difficult or even impossible to establish valid tendencies especially due to the fact that the frequency values were either very low, or rather imbalanced. Hence it was concluded that such items would deserve a specialized investigation which exceeded the objectives of the present dissertation.

It is perhaps unnecessary to add that all of the registers underwent the same analysis. The following pages are thus concerned exclusively with those instances in which the English register subcorpora exhibit the frequency of five or more tokens and simultaneously, where the frequency of the items drawn from the two translations differs by up to 2 tokens. For clarity reasons, the cases corresponding with the above-mentioned frequency criteria will be further on referred to as “the central set”.

It has been stated above that even with the true friends, disproportions in frequency may occasionally arise, which need not always suggest solely the differences in the translators' idiolects. For example, in the particular passages a lexical word may be replaced with a pronoun or zero means (in grammatical ellipsis), etc., and in other cases, a completely new construction may be employed, for example featuring passive voice, paraphrase, etc.

In practice, the French items could have been either somewhat more frequent or rarer than 5 tokens (the latter case proved to be rather exceptional, as only 4 such cases were detected in the literary register. Mostly even the French cognates amounted to 5 or more tokens).

Furthermore, other true friends (i.e., words which had their matching counterparts in the other translation) whose frequency ranged between 2 and 4 tokens in the English register samples were also generated. For completion sake, a list of these words is also provided, even though their analysis exceeds the objectives of the present dissertation. As has been mentioned, their frequency values were considered rather low and insufficient for the formulation of any tendencies. Therefore, to draw some valid conclusions even for this group of items, a considerably more sizable corpus would have been necessary.

Based on this study, it seems possible to arrange the true friends on a scale ranging from absolute identity (the minority of instances) to a closer or more remote similarity. The distribution/placement of items along this scale is established on the interplay of values in several criteria, namely, pronunciation, spelling, morphology and frequency, which can give rise to various configurations. Understandably, these criteria are mostly formal, as they concern primarily the *signifiant*. In fact the sense, i.e., the *signifié*, was taken for granted as identical, for the unit was employed in the same passage in the same register sample by a professional translator. Moreover, this identity of meaning was verified in the same particular co(n)text and further indirectly corroborated by an identical (or very similar) frequency in the two translations of a particular register sample. It should be born in mind that every text sample (a sub-corpus) is always limited. Naturally, it cannot comprise and feature all the senses a (polysemic) word.

All the criteria proposed above are consistently sub-divided into the following zones: *identical*, *similar* and *different*. The *identical* category suggests absolute identity or identity with minor and negligible modifications (for ex. “accent” in the French spelling or the different pronunciation of “r” in the two languages): in such cases, it is assumed that identity characteristics clearly prevail.

The *similarity* category is defined more loosely, as a conspicuous correspondance with a slight difference, in one respect per a criterion only, for ex. a difference in a single letter, or else difference in letter order, i.e., swapping of two letters in a written form.

Finally, the *difference* category applies to those true friends which differ or depart from each other more pronouncedly, i.e., are distinguished from one another by two or more features (for ex. 3 different letters in written form).

3.2.1 Criteria

3.2.1.1 Pronunciation

In the present dissertation, pronunciation is classified as (almost) *identical* (for ex. Madam [mæd.əm] (E) – Madame [ma.dam] (F)), as *similar* (for ex. place [pleis] (E) – place [plas] (F)), or as *different* (for ex. cohesion [kəʊ'hi:.ʒən] (E) – cohésion [kœzjɔ̃] (F)). It should be noted that strictly speaking the pronunciation of “true friends” is never perfectly identical: after all, the two related words come from two different languages and only their typical pronunciation features determine which language they belong to.

3.2.1.2 Spelling

Spelling can also be divided into *identical*, *similar* (or slightly different) and *different* zones.

Identical spelling is marked by a perfect agreement in written forms (for ex. compassion (E, F), long (E, F), question (E, F)), while *similar* spelling shows some language-specific peculiarities, as for ex. swapping of some letters (member (E) – membre (F)), substitution of letters (fund (E) – fond (F)), or other minor modifications, such as insertion or elimination of letters (second (E) – seconde (F), homage (E) –

hommage (F), sabbath (E) – sabbat (F)) – in the last two instances, the label of the resulting modification depends on the direction of comparison (English vs. French/ French vs. English). In order to qualify the spelling as *similar*, in the present project, the criterion of one different letter or a single instance of swapping was established. *Different* spelling then involves more complex changes, for ex. a reduction, a substitution and an addition of letters all at the same time (mountain (E) – montagne (F)). To meet the criterion for the *different* spelling, at least two different letters in the true cognates had to be detected.

Note : The French “accents” (*aigu*, *grave* and *circonflexe*) seem to represent but a minor modification of the written word form, somewhat more negligible compared to other differences, such as additions or reductions of letters (since from the perspective of non-native speakers, in most cases a word without the accents still remains recognizable). Therefore, in the present dissertation, providing two items differed only in the French accents, they were always assigned to the category of *identical* spelling.

3.2.1.3 Morphology

In this dissertation, a distinction between the *type* (in this case word type) and its *token* (concrete instance of use/realization) has to be made. That is why in this analysis, the frequency of a noun is established on both, its singular and plural forms, and the frequency of a verb is based a variety of forms, including the infinitive and all the forms featuring inflections, etc. Thanks to the abundance of the French grammatical endings (especially with verbs), the variety of French tokens is usually richer than it is in English. Naturally, the lists and tables incorporated in this project only feature the words in their base forms (*lemmas*) – similar to dictionaries (by convention, infinitive for a verb, singular for a noun). The number provided next to the particular type thus corresponds to all the tokens (concrete manifestations) employed in the register sample; in practice, then, the tokens detected (realizations) may be either the same, or different from, those given in the lists, for ex. the figure for the French verb *adopter* is established on such tokens as (*nous*) *adoptons*, *adoptez!*, *en adoptant*, etc. ; the figure for the French noun *question* corresponds to tokens *question*, *questions*.

In this dissertation, suffixes are also regarded as morphological features. Again, three categories may be distinguished: *identical*, *similar* (or nearly identical) and *different* suffixes.

An *identical* suffix can occur even if the true cognates differ in spelling, for ex. judgement (E) – jugement (F). In most instances, the words displaying the same suffix were borrowed as ready-made units, including the suffixes in question.

The *similar* (or nearly identical) suffix can be marked by some distinction in pronunciation or stress or by a modification in spelling, for ex. economic (E) – économique (F), development (E) – développement (F).

We include also the « domestic » suffixes, characteristic for the borrowing language, for ex. divinely (E) – divinement (F).

Considering the suffixes, a special group of morphological differences is constituted by the *zero suffix*, frequently found especially in English – as against the “explicit marker” of the wordclass (for ex. concern (E) – concern (F)). This is apparent especially with verbs. In fact, words have to function in the neat grammatical systems of the particular languages. In English, conversion is easily made possible, frequently without the slightest formal changes (for ex. from noun to verb), which is not always the case in French.

Since no finding represented a compound and only a single finding fell into the minor category of clippings (telephone (E)- téléphone (F)), these processes will not be given further attention.

The true friends are mostly perceived as whole words (and not solely as roots of such words), which naturally involves also their membership in a particular word class (part of speech). If the translation process does not preserve the word class, statistical differences may occur (following from such instances as divinement (F) – in a divine way (E)). There might be various reasons for this, for ex. a distinct syntactic construction of a sentence, the type of translation (eg. literal vs. pragmatic), the idiolect of the translator, etc. (It should be noted that there are also studies examining the degree of constancy of the word classes in translation).

3.2.1.4 Frequency

The frequency of true friends may be different in the two languages (and translations) when the word in one of the languages imposes collocation restrictions or bears distinct stylistic characteristics (for ex. it can be more or less formal in comparison with its “friend” from the other language; it can be also regarded as marginal, specialized, archaic, etc.), or it may be employed only in certain registers in one of the two languages, which then leads to its replacement by a different type. Moreover, with regard to the frequency of individual words, the idiolect of the translator undoubtedly plays an important role as well (for ex. translator’s pet words, etc.)

The frequency (statistics) is influenced by both, the grammatical and textual (or more precisely, text-formation) parameters: sometimes, in the process of translation lexical words may be replaced by grammatical words (for ex. pronominalisation: Jesus – he); at other times, it is possible to omit them completely (through ellipsis, such as in the case of a second mention, or when employing a completely different grammatical structure, for ex. passive transformation).

3.2.1.5 Meaning

A common misconception is that the meaning of true cognates should be the same in the two languages, which might be problematic and confusing. If for one meaning/sense of the words there is a corresponding match, but in a different meaning/sense, the two words diverge, it is necessary to perceive them as partial true friends (*faux amis partiels*). In extreme cases, the identical meaning can be restricted to a specific (closed) register, while all the other meanings may feature in different (e.g., more open) registers. For this reason, in this dissertation the true friends were analysed always in the context of a concrete register and furthermore, they were verified in the particular co-texts. This procedure was deemed to be sufficient to establish identity of meaning.

The related words can also fail to comprise all the shades of meaning of their counterpart from the other languages (for ex. demand (E) – demander (F, ask). For this reason we make also observations and statements exclusively based on the meaning/sense of a word in a concrete register.)

It is a well-known fact that analytical languages (including English and French) are characterized by their remarkable rate of polysemy and homonymy (see for ex. Skalička, 2004). It is thus probable that certain words display a number of shades of meaning, employed with both different probability and frequency in various registers (for ex. an angel outside the religious context refers to someone very kind or innocent). In this dissertation, all the results are based on the analyses of the particular subcorpora.

Another potential difficulty is constituted by the conversion (zero affixal derivation), as a rather frequent phenomenon in English (for ex. in English *face* functions as a noun and as a verb). It is for clarity and simplicity reasons that in the lists, solely *one* prevalent wordclass is always indicated next to the word in question.

3.2.2 Illustrative Analysis of Text Samples (extracts from sub-corpora)²²

NOTATION:

abc = proper nouns/antroponyms

abc = geographical nouns/ toponyms

abc = true friends

1. THE BIBLE

1.1. English version (extract)

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

²² In the present dissertation, the true friend lists feature exclusively those true friends the frequency of which amounted to 5 or more tokens in the English subcorpora and moreover, where the frequency of tokens gained from the two parallel translations was the same or differed by 2 tokens at most. It should be noted that in the present dissertation, proper nouns did not represent the primary object of analysis. However, their use is commented on briefly in section 4.4.

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. French version (extract)

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>

2. NEGOTIATIONS OF THE EUROPEAN PARLIAMENT

2.1. English version (extract)

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/>

2.2. French version (extract)

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/>

3. THE NOVEL

3.1. English version (extract)

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.

3.2. French version (extract)

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 Results

4.1 The Bible

The results regarding the religious register are provided below in view of the above criteria:

CENTRAL TRUE FRIENDS – BIBLE

English word list:						French word list:	
1*	S	prophet	~	22	N	prophète	22
1*	S	disciple	°	21	N	disciple	21
4*	S	fruit	*	15	N	fruit	16
3*	S	judgement	*	11	N	jugement	12
1*	S	parable	*	11	N	parabole	11
4*	S	generation	*	10	N	génération	9
2*	D	angel	~	9	N	ange	10
rare	D	pharisee	*	9	N	pharisien	9
1*	S	homage	*	7	N	hommage	6
5*	D	nation	*	7	N	nation	8
rare	S	sabbath	*	7	N	sabbat	8
4*	S	secret	*	7	N	secret	7
rare	S	baptise	~	6	V	baptiser	4
3*	I	enemy	~	6	N	ennemi	6
5*	S	people	*	6	N	peuple	6
rare	S	scribe	*	6	N	scribe	6
1*	S	synagogue	*	6	N	synagogue	8
4*	D	commit	°	5	V	commettre	5

KEY:

N = noun, ADJ = adjective, V = verb

abc = different spelling, abc = similar spelling, abc = identical spelling

similar frequency identical frequency

I = identical pronunciation, S = similar pronunciation, D = different pronunciation

* = identical suffix ~ = similar suffix ° = different suffix

4.1.1 Pronunciation

In the list of the true friends extracted from the religious register sample, all three pronunciation groups are featured: words with (almost) *identical*, *similar* and *different* pronunciation. The (almost) *identical* pronunciation was traced only once (enemy [enəmi] (E) – ennemi [ɛnmi] (F)), *similar* pronunciation was represented by 12 cases (for ex. demon [di:mən] (E) – démon [demɔ̃] (F), fruit [fru:t], [frɥi] (E, F), judgement [dʒʌdʒmənt] (E) – jugement [ʒyʒmɑ̃] (F)) and *different* pronunciation was exhibited by 5 true friends (for ex. angel [em.dʒəl] (E) – ange [ɑ̃:ʒ] (F), enter [entə(r)] (E) – entrer [ɑ̃tre] (F), pharisee [fær.i.si:] (E) – pharisen [farizjɛ̃] (F)).

4.1.2 Spelling

Concerning the graphic shape, seven true friends were marked by *identical* spelling (for ex. disciple (E, F), generation (E) – génération (F), nation (E, F), secret (E, F)), 9 true friends were displayed *similar* spelling (for ex. angel (E) – ange (F); homage (E) – hommage (F); prophet (E) – prophète (F)) and finally, there were only two specimens of true friends *different* in spelling (commit (E) – commettre (F), pharisee (E) – pharisien (F)).

4.1.3 Morphology

Considering the word classes found in the list of the “central set”, the subcorpus included 17 nouns and one verb. (It should be noted that no adjective or adverb satisfied the criteria of the “central set”).

Furthermore, the list features diverse suffixes, embodying all the three categories. The *identical* suffixes were found 4 times (for ex. judgement (E) – jugement (F), generation (E) – génération (F), homage (E) – hommage (F), nation (E, F)), the *similar* suffixes were registered twice (for ex. prophet (E) – prophète (F)), and there was only one finding of a *different* suffix (pharisee (E) – pharisien (F)).

4.1.4 Frequency

As far as the frequency is concerned, this section will be opened by a number of terms constituting the “central set”, whose frequency corresponds with, or exceeds, five

tokens in the English version and which were employed rather evenly in both the parallel translations (with a maximal difference in frequency of plus or minus two tokens). Such criteria were satisfied exactly by 18 types. The total frequency (i.e., the sum of all the concrete instances employed in the text) amounts to 171 tokens, and thus corresponds to 1.71% of the 10 000 words of the entire text sample.

It seems also important to examine the proportion of the true friends marked by identical frequencies in the parallel translations and to set them off from such pairs where the frequency was only similar (differing by up to two instances). Considering the total number of words displaying identical frequency in the two translations, 33 words (types) were detected, constituting 56% all in all (for ex. secret (E, F)), out of which only 9 words belonged to the “central set”, delimited by the frequency 5 or more tokens. There were also 26 words (i.e., 44%) featuring a similar frequency (plus or minus two cases), for ex. generation (E) – génération (F), out of which only 9 words belonged to the central list. All in all, 56% of the specimens displayed identical token frequencies in the two parallel translations; when taking into consideration only the “central set”, the proportion dropped to solely 15.25% of all.

Finally, in order to counterbalance the limited size of the sample corpus, it may be of interest to provide the frequency bands of the terms in the entire English lexicon. It should be noted that only the “central set” delimited by the frequency of 5 or more tokens will be analysed. In the Collins COBUILD dictionary (2001), two cases of the “central set” are classified as very frequent (5*), 4 word types are marked as rather frequent (4*), 3 cases as frequent (3*), one unit is considered as less frequent (2*), 5 cases as infrequent (1*), with 4 words being rare. This seems to indicate that a greater proportion of the loanwords are of a rather specialized nature (and may be regarded as marked), or exclusively religious within the entire English lexicon.

4.2 Parliament

The results regarding the political/parliamentary register are provided below in view of the above criteria:

CENTRAL TRUE FRIENDS – PARLIAMENT

English word list:					French word list:		
4*	S	<u>parliament</u>	*	49	N	parlement	51
1*	I	madam	~	35	N	madame	34
Rare	S	<u>report</u>	*	29	N	rapport	29
5*	I	programme	*	27	N	programme	29
5*	S	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*	S	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*	S	development	*	7	N	développement	7
5*	I	party	*	7	N	parti	5
4*	S	session	*	7	N	session	6
1*	S	coordination	*	6	N	coordination	6
5*	I	<u>economic</u>	°	6	ADJ	économique	6
3*	S	importance	*	6	N	importance	5
4*	D	<u>necessary</u>	°	6	ADJ	nécessaire	8
5*	S	<u>process</u>	°	6	N	procéssus	6
4*	D	propose	~	9	V	proposer	8
2*	I	tourism	~	6	N	tourisme	7
4*	S	activity	~	5	N	activité	7
4*	S	annual	~	5	ADJ	annuel	5
5*	S	place	*	5	N	place	5
3*	D	procedure	*	5	N	procédure	6
4*	D	<u>sector</u>	°	5	N	secteur	6

KEY:

N = noun, ADJ = adjective, V = verb

abc = different spelling, abc = similar spelling, abc = identical spelling

similar frequency

identical frequency

I = identical pronunciation, S = similar pronunciation, D = different pronunciation

* = identical suffix

~ = similar suffix

° = different suffix

4.2.1 Pronunciation

In the list of the true friends based on the political register (and particularly in the “central set” delimited by the frequency of 5 words or more), the *identical* pronunciation features 7 times (for ex. conference [kɒnfərəns] (E) – conférence [kɔ̃ fɛʁɑ̃ ːs] (F), party [pa:ti] (E) – parti [parti] (F), programme [prəʊgræm], [prɔ̃grɑm] (E, F). The *similar* pronunciation is represented by 12 specimens (for ex. development [di'veləpmənt] (E) – développement [devləpmɑ̃] (F), fund [fʌnd] (E) – fond [fɔ̃] (F), parliament [pa:ləmənt] (E) – parlement [parləmɑ̃] (F)), and the *different* pronunciation may be traced in 14 true friends (for ex. case [keɪs] (E) – cas [kɑ] (F), necessary [nesəsəri] (E) – nécessaire [nesesɛ:r] (F), particular [pə'tɪkjələ(r)] (E) – particulier [partikylje] (F)).

4.2.2 Spelling

As for spelling, all three categories were detected, namely *identical* spelling, which was represented by 11 items (for ex. programme (E, F), point (E, F), possible (E, F)), *similar* spelling exhibited by 13 friends (for ex. Madam (E) – Madame (F), case (E) – cas (F), period (E) – période (F)) and finally, *different* spelling displayed by 9 true friends (for ex. council (E) – conseil (F), parliament (E) – parlement (F), necessary (E) – nécessaire (F)).

4.2.3 Morphology

As for morphological findings, first the word classes incorporated in the “central set” will be provided. The subcorpus embraced 25 nouns, 6 adjectives and 2 verbs (with no adverb satisfying the criteria of the “central set”).

Furthermore, the “central set” featured a number of affixes. *Identical* suffixes were found in 9 true friends (for ex. coordination (E, F), importance (E, F), parliament (E) – parlement (F), possible (E, F)), *similar* suffixes were discovered 5 times (for ex. structural (E) – structurel (F), economic (E) – économique (F)), and there were only 2

specimens marked by different suffixes (for ex. **necessary** (E) – **nécessaire** (F), **sector** (E) – **secteur** (F)).

4.2.4 Frequency

Regarding frequency, initially, attention will be given to the items in the “central set”. Strikingly, as many as 33 true friends (i.e., types) were detected which satisfied the “central set” criteria (i.e., where the number of English tokens had to be at least five and the frequency of tokens had to be comparable).

The sum total of all the tokens employed in the text sample corresponds to 344, and thus constitutes 3.44% of the entire English register sample of 10,000 words. It seems also appropriate to give the number of words which show *identical* frequency in the two translations, and to set them off from items with similar frequency only: There were 10 items in the *identical* group (for ex. **economic** (E) – **économique** (F), **place** (E, F), **report** (E) – **rapport** (F)), and 23 words exhibiting *similar* frequency in the parallel translations (with the number of tokens differing by plus or minus two cases): 23 (for ex. **importance** (E, F), **possible** (E, F), **programme** (E, F)).

Finally, the Collins COBUILD dictionary (2001) was consulted in order to classify the words in terms of frequency bands in the entire English lexicon, as English was the primarily borrowing language. Altogether 14 items of the “central set” were assigned to the very frequent band (5*), 10 cases to rather frequent words (4*), 2 cases were considered frequent (3*), 2 words belonged to less frequent wordstock (2*), 4 types were labeled as infrequent (1*), and finally, there was one rare word only. This finding seems to indicate that the loanwords analysed are mostly already well-integrated in the English lexicon.

4.3 Novel

The results regarding the artistic/literary discourse are provided below in view of the above criteria:

CENTRAL TRUE FRIENDS – NOVEL

English word list:						French word list:	
2*	S	compassion	*	19	N	compassion	19
2*	D	mistresse	*	11	N	maîtresse	9
4*	S	table	*	8	N	table	8
2*	S	eternal	~	7	ADJ	éternel	8
5*	S	place	*	7	N	place	8
4*	I	positive	~	6	ADJ	positif	5
5*	S	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*	S	press	°	5	V	presser	4

KEY:

N = noun, ADJ = adjective, V = verb

abc = different spelling, abc = similar spelling, abc = identical spelling

similar frequency

identical frequency

I = identical pronunciation, S = similar pronunciation, D = different pronunciation

* = identical suffix

~ = similar suffix

° = different suffix

4.3.1 Pronunciation

Even the list of the true friends based on the novel exhibited all three categories of pronunciation. *Identical* pronunciation was detected three times (erotic [ɪˈrɒt.ɪk] (E) – érotique [ɛʁ ɔtik] (F), letter [lɛtə(r)] (E) – lettre [lɛtʁ] (F), positive [pɒzətɪv] (E) – positif [pozitif] (F)), *similar* pronunciation was found in 6 true friends (for ex. compassion [kəmˈpæʃ.ən], [kɔ̃ pasjɔ̃] (E, F), eternal [ɪˈtɜː.nəl] (E) – éternel [etɛrnɛl] (F), place [pleɪs], [plas] (E, F)), and *different* pronunciation was traced in 4 specimens (for ex. change [tʃeɪndʒ] (E) – changer [ʃ ɑ̃ʒe] (F), decision [dɪˈsɪʒn] (E) – décision [desizjɔ̃] (F), mistresse [mɪs.trəs] (E) – maîtresse [metʁɛs] (F)).

4.3.2 Spelling

As far as spelling is concerned, 4 true friends were *identical* in spelling (for ex. compassion (E, F), decision (E) – décision (F), place (E, F)) and 4 specimens displayed *similar* spelling (for ex. eternal (E) – éternel (F), letter (E) – lettre (F), second (E) – seconde (F)). Finally, 5 items were spelled in a *different* way (for ex. erotic (E) – érotique (F), mistress (E) – maîtresse (F), positive (E) – positif (F)).

4.3.3 Morphology

At the onset of the morphological section, it should be pointed out that the “central set” included altogether three different parts of speech (word classes), namely 7 nouns, 3 adjectives and 3 verbs (with no adverb meeting the criteria of the “central set”).

Furthermore, examining the suffixes featured by the true friends in question, the list embraced 3 words with *identical* suffixes (for ex. compassion (E, F), decision (E) – décision (F), mistress (E) – maîtresse (F)), 3 items with *similar* suffixes (par ex. eternal (E) – éternel (F), erotic (E) – érotique (F), positive (E) – positif (F)), with not a single instance of true friends marked by *different* suffixes.

4.3.4 Frequency

Regarding frequency, the number of the “central set” items in this subcorpus was found to be the most constrained of all, for only 13 true friend types satisfied the frequency condition (and displayed 5 or more tokens).

The total number of tokens employed in the register sample was 94, which corresponds to 0.94% of the English text sample of 10,000 words. Attention was also given to the proportion of words which show absolutely *identical* frequency in the two parallel translations. In fact, only 4 true friend types with *identical* frequency were traced (for ex. compassion (E, F), letter (E) – lettre (F), table (E, F)), whereas words exhibiting *similar* frequency (differing by up to two instances) prevailed, for such were 9 true friends (for ex. decision (E) – décision (F), place (E, F), second (E) – seconde (F)).

Finally, to counterbalance the size of the corpus at least to some extent, it seems necessary to give the frequency bands for the English items based on the Collins

COBUILD dictionary (2001). As regards the “central set” findings, the subcorpus comprised 7 cases assigned to the very frequent word band (5*), 2 specimens of rather frequent category (4*), no frequent words (3*), 4 tokens corresponding to the less frequent word band (2*) and no instances of infrequent (1*) or rare words. That suggests that most loanwords generated are already well-integrated in the English lexicon and do not constitute a specialized layer of vocabulary; rather than that, they seem to belong to the layer of the commonly used or even spoken language items.

4.4 Proper nouns

In addition to the aforementioned, the chapter providing the results would not be complete without a brief survey of a marginal type of true friends, i.e., proper nouns – precisely anthroponyms and toponyms²³, and hence some attention will be also given to them. Concerning their word classes, they represent either nouns (for ex. Tomas (E, F), Prague (E, F)), or adjectives (for ex. European (E) – européen (F)). In certain samples, the nouns were very frequent (for ex. in the religious register, Jesus (E) 47 – Jésus (F) 49). Even this group nevertheless shows that languages adjust the pronunciation and spelling of the loanwords, in this case proper nouns, with regard to their respective systems.

To illustrate pronunciation categories, the following may serve as suitable examples of *identical* pronunciation: Babylon (E) – Babylone (F), those of *similar* pronunciation: David (E) – David (F) and finally, those of *different* pronunciation: Jordan (E) – Jourdain (F).

Moreover, all the spelling categories were also traced, i.e., true friends with *identical* spelling illustrated by Anna (E, F), those with *similar* spelling, such as Egypt (E) – Égypte (F) and others with *different* spelling, for instance Czech (E) – tchèque (F).

It can be remarked that certain proper nouns seem to be register-specific. For example, Jesus (E) – Jésus (F) and Herod (E) – Hérode (F) associate almost exclusively with the religious register, being very frequent or indispensable in its discourse.

²³ In this dissertation, proper nouns will be understood exclusively as items that cannot be employed as common nouns (and the corresponding derived items).

5 Conclusion

The present dissertation aims at investigating true friends between English and French, two neighbouring languages, interrelated in many ways, chiefly culturally and historically.

To recapitulate, for the purposes of this dissertation, a specialized corpus was designed and assembled, containing approximately 60,000 words. The corpus comprised parallel “mirror” translations of three text samples. Since the two parallel versions need not exhibit exactly the same number of words, a decision was made to designate the English versions as the default ones. This decision followed from secondary literature sources, for in view of the phenomenon investigated, English represents the primarily “borrowing” language. Therefore, whereas all the English subcorpora comprise exactly 10,000 words, the number of words in the respective French versions differ moderately.

The three text samples embody different registers: religious register (the Bible), political register (a collection of negotiations of the European Parliament) and the register of artistic prose (fiction, a novel). These registers were chosen, among others, with regard to the intensive linguistic and cultural contacts between the two languages in these domains. The corpus on which the dissertation was established ranks among the specialized ones, i.e., it was assembled for the purposes of the present analysis. Apart from being specialized, it may be further described as synchronic, recent, written, bilingual, and closed. With reference to Laviosa’s framework (2002), all the subcorpora are composed of samples of parallel translations, with two of them displaying some features of terminological corpora, while the literary subcorpus does not.

What is more, all the three texts constituting the corpus were meant to illustrate three potentially distinct types of translation. It was assumed that the biblical text would be the closest to the original (thus approaching the literal translation), the translation of a novel would be the most artistic (and thus relatively free) and the translation of the

political discourse would show communicative and pragmatic characteristics.²⁴ By assembling the corpus in this way an endeavour was made to achieve a balanced representation of the aforementioned approached to translation.

The present analysis was carried out with the help of a corpus tool, namely the programme AntConc 3.4 2014, available for free online. Among other considerations, this tool was selected for it was applicable to the research of the two languages in question.

The procedure adopted consistently was as follows. First, two tentative sets of results (for English and for French samples) were generated and then compared in view of their frequency in the two languages (thus providing the statistical perspective). Furthermore, the findings were verified in the particular text passages (thus providing the co-textual perspective). Since the two languages are primarily analytic, attention was given only to items sharing the word class (part of speech) in both the translations, all the other cases being disregarded.

Based on the frequency findings (verified in the particular co-texts), a list of true friends displaying the same or similar frequency (differing by up to two tokens) was generated for each of the registers. When the three register-based lists were put together, the number of true friends detected in the entire corpus amounted altogether to 135 words (attached below). Nevertheless, it is necessary to keep in mind that the findings provided in this dissertation are always pertinent to the particular registers (register subcorpora) explored (see below).

Among the true friend pairs generated, two categories were deemed relevant to this dissertation: those displaying *identical* frequency (comprising true friends with the same number of tokens in both the translations), and those exhibiting *similar* frequency (comprising true friends employed with a small difference of plus or minus 1 or 2 tokens).

As has already been mentioned, analysis was primarily focused on the words whose frequency in the English list amounted to 5 or more tokens and the French

²⁴ It is possible that the frequency data in the concrete sub-corpora were influenced by the style and type of translation.

findings differed by up to 2 tokens (i.e., the items with lesser incidence were enlisted, but were not scrutinized further).

For the purposes of this dissertation, the true friends, satisfying the frequency conditions and verified co-textually, were then designated as “central” and were subjected to a multiaspectual scrutiny. The analysis itself was based on several criteria: pronunciation, spelling, morphology and frequency. The stylistic criterion could be disregarded due to the fact stated above that the results were deemed to concern the concrete registers (subcorpora) and co(n)texts exclusively, the underlying hypothesis being that the stylistic status in both the parallel translations would be the same. The semantic criterion was in fact satisfied by comparing and verifying the items in the particular co-texts, and moreover, it was indirectly corroborated by (almost) identical statistical findings.

It is perhaps needless to add that all the registers underwent the same analysis. More specifically, three zones were delimited beforehand and were determined within the majority of the criteria: the identity zone, the similarity zone and the difference zone, in order to disclose relevant tendencies.

Attached to this dissertation is a complete inventory of the generated true friends. The appendix features the lists of true friends according to the individual registers as well as the true friends generated from the entire corpus analysed (i.e., amounting to 30, 000 words). The register-specific findings are arranged with declining frequency, with the “central set” always featured initially, whereas the overall findings are ordered alphabetically. Hence, the results concern both, the entire corpus and the individual register-based sub-corpora. What is more, the central as well as the marginal instances are provided.

5.1 PRONUNCIATION

Regarding pronunciation, the generally least common case proved to be *identical* pronunciation. In contrast, *similar* pronunciation constituted the dominant finding in all the three register subcorpora. – Out of 64 central true friends (i.e., types), 11 cases were of *identical* pronunciation, 30 cases displayed *similar* pronunciation and 23 cases represented *different* pronunciation. Therefore, it can be stated that the pronunciation oscillated between *similar* and *different* categories.

The *identical* pronunciation instances were found to be the most common in the artistic register (novel). In the religious register, on the other hand, the smallest number of true friends exhibiting *identical* pronunciation were discovered.

The greatest discrepancies in the pronunciation categories of the true friends (i.e., identical, similar, different) were noticed in the religious subcorpus, while in the political register, the shares of the three pronunciation categories proved to be the most balanced of all.

The tendencies traced for the true friend types, mostly correspond also with the token findings. Among the tokens, similarity dominated, followed by difference, with the least common being identity. What is more, this tendency holds for all the investigated subcorpora.

Admittedly, the pronunciation of the “true friends” is never perfectly identical: after all, the two related words come from two different languages and only their typical pronunciation features determine which particular language they belong to.

Given the fact that English and French pronunciations seem to differ rather significantly and that the identical pronunciation proved to be very rare, the true friends seem to be much easier to recognize in the written language.

5.2 SPELLING

Out of the total of 64 central true friends (i.e., types), the number of cognates marked by *similar* spelling (26 cases) and by *identical* spelling (22 cases) were found to be almost balanced, with the former moderately prevailing. When these two groups were put together, they composed a clear majority. In contrast, the true friends with *different* spelling formed a minority (16 cases in all), and moreover, they were marginal in two subcorpora. This finding seems to suggest that the true friends are more easy to identify in the written medium.

The artistic register achieved almost balanced results in spelling, for it featured 4 cases of *identically* spelled true friends, 4 cases of items spelled in a *similar way* and 5 cases of *differently* spelled true friends. Thus considering the true friend types, the prevailing spelling categories proved to be similarity and *identity*.

Similarly, the parliamentary register exhibited relatively even shares between the three spelling categories. In particular, there were 11 friends marked by *identical* spelling, 13 cases exhibiting *similar* spelling and 9 cases characterized by *different* spelling.

In the biblical register, the majority of true friends displayed *similar* spelling (9 cases). Apart from these, there were 7 true friends with *identical* spelling, and finally, only two true friend pairs *differed* in spelling.

What is more, the same tendency can be uncovered for the tokens. Examining the individual subcorpora, however, distinct tendencies may surface. Regarding spelling, the token findings for the religious and political registers were dominated by *similarity* (closeness) of spelling, while the token findings for the literary subcorpus were marked mostly by *identity* of spelling.

Hence the dominant category proved to be *similar* spelling, the second most common was found to be *identical* spelling, whereas *different* spelling among true friends was traced the least frequently of all.

5.3 MORPHOLOGY

As far as word classes are concerned, it can be stated that out of the 64 central true friends types, the nouns constituted a majority (49 cases), the rest being composed of adjectives (9 cases) and the least prominent verbs (6 cases). Furthermore, concerning the distribution of word classes in individual registers, the nouns prevailed in all the tree subcorpora.

The religious subcorpus did not feature a single adjective among the true friends detected. In contrast, the true friends drawn from the political register included more adjectives than verbs.

Hence the majority of findings (regarding types as well as tokens) was constituted by nouns, which were followed by less common adjectives and finally, by rarer verbs. Comparing the three subcorpora, the highest proportion of nouns among the findings was attested in the religious register, while adjectives and verbs were most readily used in the literary subcorpus, where the shares of the word classes were relatively the most balanced of all the subcorpora. Presumably, the significant decrease of nouns and simultaneous increase of verbs in the literary subcorpus may be explained by the narrative text-type, i.e., by the increasing dynamicity of literary discourse.

As for suffixes, out of the 64 true friends (i.e., types) identified in this project, a lesser half of all the types (i.e., 29 in all) were coined by suffixal derivation (45.2%), of which only two cases featured prefixes as well. Suffixation thus clearly outnumbered prefixation. It should be noted that the leading suffixes proved to be those forming nouns, such as -ation, -ion, -ment, etc.

The greatest representation of elements coined by derivation was detected in the parliamentary register, since 16 out of 33 words (48.48%) proved to be derived. The literary register then showed 46.14% of derived units, and in the biblical subcorpus, 38.89% of true friends employed affixation. These results indicate that the true friends generated were frequently long words with a stylistically rather elevated status in English.

Considering the tokens, the suffixation process turned out to be even more noticeable, for it could be traced in over a half of true-friend tokens (52.83%). Nevertheless, the register corpora differed in this respect. Unlike in the religious register, in the parliamentary and artistic registers derivation prevailed. Hence a sizable proportion of the findings in the registers under investigation proved to be long words, for they constituted altogether more than a half of all tokens.

As for the word origin of the true friends detected, they were mostly borrowings from a Romanic language (usually Latin or Old French) to English. In the morphological section, this study focuses on determining the proportion of borrowings comprising derivation. Sometimes the borrowings did not feature any affixes, while sometimes they were borrowed as ready-made units, together with the original suffixes.

If we examine the individual subcorpora findings from the standpoint of suffix categories, the most widespread proved to be *identity* of suffixes, followed by suffixal *similarity* between true friends, whereas *differences* in suffixes between true friends turned out to be rather rare. This tendency was attested with types as well as with tokens. Therefore, it seems more comfortable for the borrowing language to borrow whole words including their affixes. In addition to that, it also appears that preserving the written shape (with the affixation) may be easier than retaining the original pronunciation.

5.4 FREQUENCY

Attention was also given to the actual proportion of true friend tokens in each of the subcorpora. From this perspective, remarkable discrepancies surfaced. In fact, the study gained the highest share of central true friends from the parliamentary register, for the tokens composed 3.44% of the entire text sample (i.e., 10,000 graphic words, including the grammatical ones). In this respect, the biblical register occupied the second place, for the true friend tokens constituted 1.71% of all the words in the subcorpus and in the artistic register, the tokens of true friends were found to be less common (0.94%). Consequently, it seems that the more closed the register, the more

likely true friend findings seem to be. (However, it would be necessary to verify this tendency using a much larger corpus).

Examining the “central set” (i.e., specimens satisfying the condition of 5 or more English tokens and similar frequency in French) from the standpoint of the register sources, the highest number of the items (as types) was drawn from the political register (33 cases). True friends attested in the religious sample were almost twice rarer (18 items in all) and literary lists included only 13 specimens.

The highest number of true friend types as well as tokens with identical frequency values was found in the religious subcorpus; conversely, literary subcorpus, and even more conspicuously the political subcorpus were dominated by similar frequency findings for both, types and tokens. These results may be related to the distinct translation approaches in question.

Considering the true friend types, in the biblical register an interesting chasm was uncovered: on the one hand, more than a half of the true friends were infrequent or rare, whereas the lesser half was classified as frequent or even very frequent in the English lexicon according to Collins COBUILD dictionary (2001).

In contrast, the other two registers explored, namely parliamentary and literary discourse, exhibited a considerable share of words belonging to the very high frequency band within the entire English lexicon.

Considering the individual tokens, the leading proportion of items in the parliamentary subcorpus belonged to the two top frequency bands (70%), whereas in the literary subcorpus the share of the two top categories declined (55.31%) at the expense of a rarer category (2*, 44.69%). By comparison, in the religious subcorpus, the two top frequency bands were underrepresented (29.24%) and it was the lowest bands (classified as (1*) or rare) that grew remarkably (55.56%). These findings seem to confirm major stylistic differences between the subcorpora, i.e., the above-mentioned stereotypicality of the units employed in the political subcorpus, the tendency to enliven the style by less common vocabulary in creative fiction, and the relatively marked character of numerous expressions featured in the biblical discourse, which appear almost confined to this register.

Furthermore, in the entire corpus three true friends proved to reach top frequencies of tokens of all. The highest number of tokens was collected for the item parliament (E, 49 instances) – parlement (F, 51 cases); the second highest for the word (E, 35) – Madame (F, 34) and the third highest frequency was exhibited by the word report (E, 27) – rapport (F, 29). It is perhaps no coincidence that all these specimens were drawn from the political register.

In order to examine the proportion of very common true friends, the frequency criterion was arbitrarily raised twice. As a result, the criterion of a minimum of 10 tokens was satisfied by 9 words belonging to the political register (parliament (E) – parlement (F), Madam (E) – Madame (F), programme (E, F), fund (E) – fond (F), report (E) – rapport (F), point (E, F), council (E) – conseil (F), structural (E) – structurel (F), case (E) – cas (F)), by 6 words drawn from the biblical subcorpus (disciple (E, F), generation (E) – génération (F), fruit (E, F), judgement (E) – jugement (F), parable (E) – parabole (F), prophet (E) – prophète (F)) and only by 2 words from the literary text (compassion (E, F), mistress (E) – maîtresse (F)). These findings also reaffirm the rather closed character of the parliamentary register sample, and the rather open character of the novel providing enough space for creativity, diversity and variability.

For completion sake, the entire list of findings will be briefly discussed as well (i.e., where the frequency findings in parallel translations differed by up to 2 tokens per each true friend pair). The religious register contained the highest number of true friends (59) of all. Nevertheless, only a lesser third of them fell into the “central set” (i.e., they attained the frequency of 5 or more tokens). In fact, for the most part, those words did not constitute the center of the English lexicon; quite the opposite, they tended to be associated exclusively with the religious discourse, which is why they were assigned to relatively low frequency bands within the English lexicon by Collins COBUILD dictionary (2001).

The parliamentary text sample included 38 cases of true friends. Interestingly, a clear majority (98.84%) of the findings simultaneously satisfied the condition of 5 or more tokens and were integrated in the “central set”. This suggests that the items contribute to the stereotyped, patterned formulation, typical of the register.

As for the literary subcorpus, it comprised the same number of true friends as its parliamentary counterpart, i.e., 38 cases, but only one third reached the frequency of 5 or more tokens. Moreover, according to Collins COBUILD dictionary (2001), many items form the center of the lexicon, for they are assigned either to the high frequency band (5*) or rather high frequency (4*) band by the dictionary. However, the remaining items are used less commonly (2*), which suggests that the author also strived to employ a rich, variegated language.

To sum up, to counterbalance the limited size of the corpus at least to some extent, the Collins COBUILD dictionary (2001) was systematically consulted to investigate the frequency of the “central set” items in the English lexicon. In view of the frequency bands the true friends drawn from the three subcorpora are generally assigned to by the dictionary, the two top categories (5* and 4*) prevail in terms of types (61%) as well as tokens (57%). This corroborates their important status and mostly full integration in the English wordstock.

5.5 MEANING

The meaning of true friends featuring in this study was verified in the co(n)text and moreover, the items were always analysed within the particular register subcorpus. The results generated are thus relevant to a particular register only. In case a given word is polysemic, and it carries but one concrete meaning in the analysed text samples, this project will not address the remaining meanings. Therefore it exceeds the objectives of the present dissertation to examine the senses the particular polysemic words might display outside the register subcorpora examined. It follows that it is not possible to make any conclusions concerning *other* registers which may feature the same true friends as listed in this project.

It was verified in an etymological dictionary that all the 64 cases of the central true friends, whose frequency was five or more tokens, were of Romanic origin, borrowed by English. Even though the individual terms may differ in their particular origin (some came directly from Latin, some originally Latin words were introduced

into English through via Old French, and in still other cases, the actual origin is ambiguous, identified by the Etymological dictionary as Latin or Old French), this may be disregarded, for in all the examined cases English always took on the role of a borrowing language, and never that of the providing one.

In the following tables, the concrete results are produced in percentage terms :

KEY:

the percentage of true friends within the frame of the registers of the Bible, the novel, the parliament

the most remarkable value/proportion among the three registers

the highest percentage within the entire corpus

WORD TYPES²⁵ (lexemes) :

			BIBLE		NOVEL		PARLIAMENT		(en. corpus)
			(entire corpus)		(entire corpus)		(entire corpus)		
1	Pronunciation	I	5.56%	1.56%	23.07%	4.69%	21.21%	10.94%	17.19%
		S	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	Spelling	I	38.88%	10.93%	30.77%	6.25%	34.33%	17.19%	34.37%
		S	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%
		S	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%
		_	61.11%	17.19%	53.86%	10.94%	51.52%	26.56%	54.69%
	(word classes)	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	Frequency	I	50%	14.06%	30.77%	6.25%	30.3%	15.63%	35.94%
		S	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%

²⁵ On type vs. token, see Soutet, 2005 : 66 or Čermák, 2001 : 26.

WORD TOKENS ²⁶:

			BIBLE		NOVEL		PARLIAMENT		(en. corpus)
			(entire corpus)		(entire corpus)		(entire corpus)		
1	Pronunciation	I	3.51%	0.94%	17.02%	2.5%	26%	15.2%	18.64%
		S	75.44%	21.16%	55.32%	8.15%	42.36%	24.76%	54.07%
		D	21.05%	4.7%	27.66%	4.07%	31.64%	18.52%	27.29%
2	Spelling	I	42.11%	11.29%	41.49%	6.11%	27.35%	15.99%	33.39%
		S	49.70%	13.32%	24.47%	3.61%	38.34%	22.41%	39.34%
		D	8.19%	2.19%	34.04%	5.02%	34.31%	20.06%	27.27%
3	(suffixes)	I	29.91%	5.49%	37.23%	5.49%	45.31%	26.49%	37.47%
		S	16.37%	4.39%	19.15%	2.82%	8.58%	5.02%	12.23%
		D	5.26%	1.41%	0%	0%	2.95%	1.72%	3.13%
		–	57.89%	15.52%	43.62%	6.43%	43.16%	25.24%	47.17%
	(word classes)	N	97.08%	26%	64.89%	9.56%	83.38%	48.75%	84.33%
		ADJ	0%	0%	19.15%	2.82%	11.8%	6.9%	9.72%
		V	2.92%	0.78%	15.97%	2.35%	4.82%	2.82%	9.95%
4	Frequency	I	54.38%	14.58%	39.36%	5.8%	23.32%	13.63%	34.02%
		S	45.62%	12.24%	60.64%	8.93%	76.68%	44.82%	65.98%
		5*	7.6%	2.04%	40.42%	5.96%	40.21%	23.51%	31.51%
		4*	21.64%	5.8%	14.89%	2.19%	29.76%	17.4%	25.39%
		3*	9.94%	2.66%	0%	0%	2.94%	1.72%	4.38%
		2*	5.26%	1.41%	44.69%	6.58%	4.29%	2.5%	10.49%
		1*	39.18%	10.5%	0%	0%	15.03%	8.78%	19.28%
		Rare	16.38%	4.39%	0%	0%	7.77%	4.56%	8.95%

²⁶ On type vs. token, see Soutet, 2005 : 66 or Čermák, 2001 : 26.

For the majority of the criteria, three zones were established, namely identity, similarity and difference. In addition to that based on the interplay of these values, it is also possible to arrange the true friends themselves along a tentative scale. In fact, the continuum comprises true friend cases which are the most recognizable (i.e., those which have displayed identity in most of the criteria) via those that are very similar (showing similarity in most of the criteria) to those that have drifted relatively the “furthest apart” in the two languages (nevertheless, they are still recognizable and recognized as *true friends*).

As examples of true friends which are identical in almost every respect, *synagogue* (E, F) or *long* (E, F) might be provided. Similar true friends may be illustrated by *period* (E) – *période* or *development* – *développement* (F). The most dissimilar true friends include *mistresse* (E) – *maîtresse* or *counsil* (E) – *conseil* (F), for they are marked by more conspicuous differences in various factors.²⁷ It is self-evident that even the true friends can be arranged according to the theory of center, transition and periphery (see Daneš, 1966). It should be noted that this conception of the centre, transition and periphery has been applied from the standpoint of the user of one of the two languages in question.

Naturally, the same theory may be employed to address the differences between the factors investigated. For example, in this respect, among the suffixes, *identity* clearly prevailed, constituting over 55% of suffixes in true friend types and 71% in the respective tokens. On the other hand, in terms of pronunciation and spelling, the cognates tended to be mostly *similar*. This indicates that they were usually borrowed as entire units, but having existed in two disparate languages and having been subjected to distinct patterns of use, the true friends have gradually become more distant, dissociated and estranged one from another, both in the written and in the spoken medium.

²⁷ It is interesting that our classification of true friends resembles the one proposed by Marcheteau et Dahan : “certain, (clear, evident)” true friends, “similar” true friends and true friends that “become misleading when pronounced”, (translation mine:) “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)

Nevertheless, as was shown above, they seem to have moved further apart in pronunciation than in spelling.

The results of the present analysis show that the “central set” of true friends comprises 64 types which correspond to 638 tokens; this means that on average, one type corresponds to 9.97 tokens. In contrast, the peripheral zone includes 71 types amounting to 214 tokens. Hence, in the periphery, the average frequency decreases remarkably, corresponding to 3.01 tokens per a type. All in all, however, since the research identified 135 types in 852 corresponding tokens, the overall average frequency corresponded to 6.31 tokens per type.

Although the corpus size was limited, the results also show that there was only a single item among the central true friends²⁸ from the “central set” that was shared by two registers. More specifically, the only true friend pair shared by the parliamentary and literary subcorpora was *place* (E, F). This statement seems to confirm the assumption that the phenomenon explored, i.e., inter-language true friends, is largely register-specific, and that the items may not attain the same characteristics/senses, frequencies, collocability patterns, etc. across a range of distinct registers.

To conclude, unlike the majority of the existing publications on the topic of English-French vocabulary relationships which mostly address the differences in meaning between the lexical counterparts (false friends), this dissertation gave attention to similarities, pointing out how much the two languages have *in common*. Thus, it is believed that this study could have a motivating effect on students of English and/or French.

Moreover, even though this *parole* study is established solely on a corpus of limited length (60, 000 words), some of the true friend findings it produced enriched the lists available in the existing lexicology publications dealing with this subject. Naturally, more extensive research based on more sizable corpora would be needed to verify the tendencies suggested and particularly, to yield additional true friends.

²⁸ Outside the “central set” items, there were 5 other cases of this sort detected in the entire list: compassion (E, F), condem (E) – condamner (F), letter (E) – lettre (F), point (E, F), table (E, F).

6 LITERATURE

6.1. SECONDARY SOURCES

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 pp.

ATKINS, S., CLEAR, J., and OSTLER, N. Corpus Design Criteria. *Literary and Linguistic Computing*, 7.1, 1992, 1-16.

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

BELL, A. *The Guidebook to Sociolinguistics*. Oxford: Wiley Blackwell, 2014, 367pp.

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., and REPPEN, R., *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge: Cambridge University Press, 1998, 300 pp.

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

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

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 pp.

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

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

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

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

ČERMÁK, F. *Jazyk a slovník. Vybrané lingvistické studie*, Praha: Karolinum, 2014, 524 pp.

CHISS, J.L., and BALLY, C. (1865 - 1947) *Historicité des débats linguistiques et didactiques*, Stylistique, Énonciation, Crise du français Peeters, 2006, 257 pp.

CROFT, W. *Explaining Language Change*. Harlow: Longman. 287pp.

CRYSTAL, D. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press, 1995, 489 pp.

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 pp.

GUILHAMOU, J., HABERT, B., NAZARENKO, A. and SALEM, A., Les linguistiques de corpus. In: *Histoire & Mesure*, 1998 volume 13 - n°3-4. Varia. p. 414-416.

HALLIDAY, M.A.K and HASAN, R., *Language, Context and Text: Aspects of Language in a Social-semantic Perspective*, Oxford: Oxford University Press, 3 ed, 1991, 125 pp.

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 pp.

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, p. 209-248.
- JAKOBSON, R. Closing Statement. Linguistics and Poetics. In : T.A. Sebeok (ed) *Style in Language*. Cambridge Massachusetts : MIT Press, 1960, pp. 350–377.
- KNITTLOVÁ, D. et al., *Překlad a překládání*, Olomouc: Univerzita Palackého v Olomouci, FF, 2010, 291 pp.
- KOHNEN, T. and MUKHERJEE, J. (Eds.) *English Corpus Linguistics*. Frankfurt am Main: Peter Lang, 2011, 267pp.
- LAVIOSA, S. *Corpus-Based Translation Studies. Theory, Findings, Applications*. Amsterdam and New York: Rodopi. 2002, 138pp.
- MARCHETEAU, M. and DAHAN, L., *Vrais et faux amis en anglais*, Paris: CPI, 2009, 95 pp.
- MARTINET, A., *La Linguistique, guide alphabétique*, Paris: Denoel, 1969.
- OGDEN, C.K. and RICHARDS, I.A., *The Meaning of Meaning*. Cambridge: University of Cambridge, 1923.
- PEPRNÍK, J. *English Lexicology*, Olomouc: Univerzita Palackého v Olomouci, 2001, 186 pp.
- PETITJEAN, A., Les typologies textuelles, *Pratiques* 62, 1989 , p. 86-125.
- PICOCHÉ, J. *Précis de lexicologie française*. Paris: Nathan, 1992, 190 pp.
- QUIRK, R., GREENBAUM, S., LEECH, G., and SVARTVIK, J. *A Comprehensive Grammar of the English Language*. London: Longman, 1985.
- RIZZO, C.R. Getting on with Corpus Compilation: From Theory to Practice. *ESP World*, 27/1, vol. 9, 2010, 1-23.
- SAUSSURE, F. de, *Cours de linguistique générale. Edition critique*, ed. & comm. Tullio de Mauro, Paris: Payot, 1972/1997, 509 pp.
- 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, 219pp.

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, 71pp.

VACHEK, J. *Chapters from Modern English Lexicology and Stylistics*. SPN : Praha, 1991, 221pp.

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.

6.2. DICTIONARIES

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

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

HABOURY, F. (ed.), *Le Dictionnaire érudit de la langue française*, Paris: Larousse, 2009, 2109 pp.

HLADKÝ, J. *Zrádná slova v angličtině*. SNP: Praha, 1990, 187pp.

KOL., *Francouzsko-český, Česko-francouzský slovník*, Olomouc: Fin Publishing, 1998, 1488 pp.

MATTHEWS, P.H. *Oxford Concise Dictionary of Linguistics*. Oxford and New York: Oxford University Press, 1997.

ONIONS, C.T. (ed.), *The Oxford Dictionary of English Etymology*, Oxford: Oxford University Press, 1966, 1024 pp.

RADINA, O., *Zrádná slova ve francouzštině*, Praha: Státní pedagogické nakladatelství, 1988, 184 pp.

REY, A. and REY-DEBOVE, J. (eds.) *Le Nouveau Petit Robert 2008. Dictionnaire alphabétique et analogique de la langue française*. Paris: Le Robert, 2008, 2837pp.

SCRIVEN, R. (ed.), *French Dictionary*, New York: HarperCollins Publishers, 2013, 618 pp.

SINCLAIR, J. (ed.) *Collins Cobuild English Dictionary for Advanced Learners*, Glasgow: HarperCollins Publishers, 2001, 1824 pp. (COLLINS COBUILD)

6.3. PRIMARY SOURCES

KUNDERA, M., *L'insoutenable légèreté de l'être*, KÉRÉL, F. (translation), Paris: Gallimard, 1990, 476 pp.

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

<http://unbound.biola.edu/index.cfm?method=searchResults.doSearch> – English version of the Bible : Darby

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

<http://www.statmt.org/europarl/> (the source of the subcorpus based on political negotiations)

6.4. ELECTRONIC RESOURCES:

EDWARDS, Michael, *Can a member of the French Academy use anglicisms?* (translation mine)²⁹, video-interview by 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

²⁹ *Un membre de l'Académie française peut-il employer des anglicismes?*

IGOT, Pierre, <http://fauxamis.fr> – website dedicated to false friends between English and French

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

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

Résumé rigorózní práce:

Analysis of English and French True Friends (Vrais Amis)

in a Corpus of Authentic Text Samples

Analýza anglo-francouzských pravých protějšků v korpusu autentických vzorků textů

Tato rigorózní 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.

Rigorózní 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 rigorózní 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 slovnědruhovému zařazení, (a vedle jednoznačného směru vypůjčování z francouzštiny do angličtiny) 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 právě protějšky přiřadit do tří pásem od identity přes blízkost až po relativní nepodobnost. Tyto zóny jsou jasně vymezeny a 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 résumé, soupis použité bibliografie, a dále celá řada příloh, včetně tabulek zjištěných pravých protějšků pro jednotlivé subkorpusy uspořádaných frekvenčně a seznamu všech centrálních pravých protějšků zastoupených v celém zkoumaném korpusu uspořádaného abecedně.

Pro účely této práce byl sestaven speciální korpus paralelních anglických a 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 přijímajícím jazykem bývá většinou angličtina, rozsah 10 000 slov byl stanoven jako kritérium jen pro anglickou

verzi, přičemž francouzské překlady byly 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 (tj. 0,05% subkorpusu) 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, shodných slovních druhů v konkrétních kontextech, 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 slovnědruhové příslušnosti hlavně substantiva, okrajově adjektiva. Také ta podléhala modifikaci formy, neboť lze vysledovat případy výslovnostní i pravopisné identity, blízkosti, ale i vzdálené podobnosti.

Délka slov je v angličtině podstatná, protože nezřídka bývá ukazatelem stylové příslušnosti. I když se jednalo ve sto procentech o výpůjčky, navíc výhradně o výpůjčky do anglického jazyka (prověřené dle Etymologického slovníku), bylo zajímavé zjistit, jaký podíl v těchto výpůjčkách měl proces derivace. Výzkum ukázal, že slovovorný postup derivace, v drtivé většině případů vlastně sufixace, se vyskytuje takřka u poloviny protějšků. Jednalo se však o prvky značně frekventované, neboť z hlediska počtu tokenů derivované prvky tvoří mírnou většinu. Vzhledem k časté blízkosti a někdy dokonce identitě sufixů je zřejmé, že se přejímala celá slova včetně sufixů, velmi odlišných sufixů je minimum. Lze doplnit, že zejména angličtina pak může přejatá slova dále snadno konvertovat a přepisovat jim jiné slovnědruhové charakteristiky mnohdy i bez formálních změn.

Okruh centrálních pravých protějšků zahrnoval celkem 64 typů, jímž v subkorpusech odpovídalo 638 tokenů. To znamená, že průměrný počet tokenů na jeden typ dosahoval takřka deseti výskytů (9,97). Naproti tomu prvky periferní z hlediska využití v analyzovaných subkorpusech (které nebyly podrobeny multiaspektuální analýze, nicméně jsou pro úplnost uvedeny v celkovém závěrečném přehledu v tabulce), čítaly celkem 71 typů. Jím odpovídalo v subkorpusech 214 tokenů. V tomto periferním okruhu tak průměrná četnost dosahuje pouze zhruba tří (3,01) tokenů na jeden typ. Tudíž z tohoto okruhu nebylo možno dovodit žádné podložené tendence. Pokud pro úplnost výsledky obou frekvenčních zón sloučíme, v celém korpusu bylo zjištěno 135 pravých anglo-francouzských protějšků-typů, jímž odpovídalo 852 tokenů. Průměrně tak byl každý pravý protějšek zaznamenán více než šestkrát (četnost 6,31). Pokud jde o porovnání tendencí, které se týkají typů a tokenů, nejpřesněji je shrnují obě připojené tabulky v samotném závěru 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 devět 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 vykazoval celkově nejnižší počet *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 mluvčím 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 (centrálním) pouze formální blízkost pravých protějšků, neboť prvky již delší dobu existují v obou jazycích na sobě nezávisle.

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.

7 APPENDICES

Attached to this dissertation is a complete inventory of the generated true friends. The appendix features the lists of true friends according to the individual registers as well as the true friends generated from the entire corpus analysed (i.e., amounting to 30,000 words per language). The register-specific findings are arranged with declining frequency, with the “central set” always featured initially, whereas the overall findings are ordered alphabetically.

THE CENTRAL TRUE FRIENDS AS AGAINST THE PERIPHERAL ONES

1. TRUE FRIENDS IN INDIVIDUAL REGISTER SUBCORPORA

1.1. BIBLE

TRUE FRIENDS - BIBLE				FRENCH list:	
ENGLISH list:					
prophet	22			prophète	22
disciple	21			disciple	21
fruit	15			fruit	16
judgement	11			jugement	12
parable	11			parabole	11
generation	10			génération	9
angel	9			ange	10
pharisee	9			pharisien	9
homage	7			hommage	6
nation	7			nation	8
sabbath	7			sabbat	8
secret	7			secret	7
baptise	6			baptiser	4
enemy	6			ennemi	6
people	6			peuple	6
scribe	6			scribe	6
synagogue	6			synagogue	8
commit	5			commettre	5
adultery	4			adultère	6
condemn	4			condamner	2
judge	4			juge	3

mage	4	mage	4
mountain	4	montagne	4
paralytic	4	paralytique	3
sign	4	signe	4
voice	4	voix	4
command	3	commander	1
glory	3	gloire	2
joy	3	joie	4
lamp	3	lampe	3
leper	3	lépreux	3
pearl	3	perle	1
return	3	retourner	5
table	3	table	2
temple	3	temple	3
abundance	2	abondance	3
ancient	2	ancien	2
authority	2	autorité	3
confess	2	confesser	3
courage	2	courage	2
divinely	2	divinement	2
justify	2	justifier	2
long	2	long	3
member	2	membre	2
perfect	2	parfait	1
prepare	2	préparer	2
prison	2	prison	2
sacrifice	2	sacrifice	2
salute	2	saluer	2
agitate	1	agiter	1
assemble	1	assembler	1
blaspheme	1	blasphème	3
calm	1	calme	1
chamber	1	chambre	2
compassion	1	compassion	1
condition	1	condition	1
debt	1	dette	1
divorce	1	divorce	1

1.2. NOVEL

VRAIS AMIS - NOVEL ENGLISH list:			FRENCH list:	
compassion	19		compassion	19
mistresse	11		maîtresse	9

table	8	table	8
eternal	7	éternel	8
place	7	place	8
positive	6	positif	5
second	6	seconde	4
change	5	changer	5
decision	5	décision	6
erotic	5	érotique	4
letter	5	lettre	5
offer	5	offrir	3
press	5	presser	4
arrive	4	arriver	5
condemn	4	condamner	6
invite	4	inviter	4
melancholy	4	mélancolie	4
possibility	4	possibilité	3
question	4	question	3
telephone	4	téléphone	5
alcohol	3	alcool	3
allusion	3	allusion	3
clearly	3	clairement	3
colleague	3	colleague	4
compromise	3	compromis	4
consider	3	considérer	3
director	3	directeur	5
hour	3	heure	3
invasion	3	invasion	3
large	3	large	4
movement	3	mouvement	3
normal	3	normal	3
occupation	3	occupation	5
part	3	partie	4
point	3	point	5
reason	3	raison	5
train	3	train	4
advance	2	avance	3
circumstance	2	circonstance	3

1.3. PARLIAMENT

TRUE FRIENDS			FRENCH list:	
---------------------	--	--	---------------------	--

PARLIAMENT			
ENGLISH list:			
parliament	49	parlement	51
madam	35	madame	34
report	29	rapport	29
programme	27	programme	29
fund	25	fond	26
point	15	point	14
council	10	conseil	11
structural	10	structurel	11
case	10	cas	11
concern	9	concerner	10
possible	9	possible	8
principle	9	principe	7
propose	9	proposer	8
harmonisation	8	harmonisation	7
particular	8	particulier	8
period	8	période	8
cohesion	7	cohésion	7
conference	7	conférence	6
course	7	cours	6
development	7	développement	7
party	7	parti	5
session	7	session	6
coordination	6	coordination	6
economic	6	économique	6
importance	6	importance	5
necessary	6	nécessaire	8
process	6	procéssus	6
tourism	6	tourisme	7
activity	5	activité	7
annual	5	annuel	5
place	5	place	5
procedure	5	procédure	6
sector	5	secteur	6
certain	4	certain	5
creation	4	création	5
effort	4	effort	5
extremely	4	extrêmement	6
letter	4	lettre	5

2. TRUE FRIENDS IN THE ENTIRE CORPUS

ENGLISH list:		FRENCH list:	
abundance	2	abondance	3
activity	5	activité	7
adultery	4	adultère	6
advance	2	avance	3
agitate	1	agiter	1
alcohol	3	alcool	3
allusion	3	allusion	3
ancient	2	ancien	2
angel	9	ange	10
annual	5	annuel	5
arrive	4	arriver	5
assemble	1	assembler	1
authority	2	autorité	3
baptise	6	baptiser	4
blaspheme	1	blasphème	3
calm	1	calme	1
case	10	cas	11
certain	4	certain	5
circumstance	2	circonstance	3
clearly	3	clairement	3
cohesion	7	cohésion	7
colleague	3	collegue	4
commit	5	commettre	5
command	3	commander	1
compassion (B)	1	compassion (B)	1
compassion (N)	19	compassion (N)	19
compromise	3	compromis	4
concern	9	concerner	10
condemn (B)	4	condamner (B)	2
condemn (N)	4	condemn (N)	6
condition	1	condition	1
confess	2	confesser	3
conference	7	conférence	6
consider	3	considérer	3
coordination	6	coordination	6
council	10	conseil	11
courage	2	courage	2
course	7	cours	6
creation	4	création	5

debt	1	dette	1
decision	5	décision	6
development	7	développement	7
director	3	directeur	5
disciple	21	disciple	21
divinely	2	divinement	2
divorce	1	divorce	1
economic	6	économique	6
effort	4	effort	5
enemy	6	ennemi	6
erotic	5	érotique	4
eternal	7	éternel	8
extremely	4	extrêmement	6
fruit	15	fruit	16
fund	25	fond	26
generation	10	génération	9
glory	3	gloire	2
harmonisation	8	harmonisation	7
homage	7	hommage	6
hour	3	heure	3
chamber	1	chambre	2
change	5	changer	5
importance	6	importance	5
invasion	3	invasion	3
invite	4	inviter	4
joy	3	joie	4
judgement	11	jugement	12
judge	4	juger	3
justify	2	justifier	2
lamp	3	lampe	3
large	3	large	4
leper	3	lépreux	3
letter (P)	4	lettre (P)	5
letter (N)	5	lettre (N)	5
long	2	long	3
madam	35	madame	34
mage	4	mage	4
melancholy	4	mélancolie	4
member	2	membre	2
mistress	11	maîtresse	9
mountain	4	montagne	4
movement	3	mouvement	3
nation	7	nation	8
necessary	6	nécessaire	8

normal	3	normal	3
occupation	3	occupation	5
offer	5	offrir	3
parable	11	parabole	11
paralytic	4	paralytique	3
parliament	49	parlement	51
particular	8	particulier	8
party	7	parti	5
part	3	partie	4
pearl	3	perle	1
people	6	peuple	6
perfect	2	parfait	1
period	8	période	8
pharisee	9	pharisien	9
place (P)	5	place (P)	5
place (N)	7	place (N)	8
point (P)	15	point (P)	14
point (N)	3	point (N)	5
positive	6	positif	5
possible	9	possible	8
possibility	4	possibilité	3
prepare	2	préparer	2
press	5	presser	4
principle	9	principe	7
prison	2	prison	2
process	6	procéssus	6
procedure	5	procédure	6
programme	27	programme	29
prophet	22	prophète	22
propose	9	proposer	8
question	4	question	3
reason	3	raison	5
report	29	rapport	29
return	3	retourner	5
sabbath	7	sabbat	8
sacrifice	2	sacrifice	2
salute	2	saluer	2
scribe	6	scribe	6
second	6	seconde	4
secret	7	secret	7
sector	5	secteur	6
session	7	session	6
sign	4	signe	4
structural	10	structurel	11

synagogue	6		synagogue	8
table (B)	3		table (B)	2
table (N)	8		table (N)	8
telephone	4		téléphone	5
temple	3		temple	3
tourism	6		tourisme	7
train	3		train	4
voice	4		voix	4

Univerzita Karlova, Pedagogická fakulta

M. Rettigové 4, 116 39 Praha 1

Evidenční list žadatelů o nahlédnutí do listinné podoby práce

Jsem si vědom/a, že závěrečná práce je autorským dílem a že informace získané nahlédnutím do zveřejněné závěrečné práce nemohou být použity k výdělečným účelům, ani nemohou být vydávány za studijní, vědeckou nebo jinou tvůrčí činnost jiné osoby než autora.

Byl/a jsem seznámen/a se skutečností, že si mohu pořizovat výpisy, opisy nebo rozmnoženiny závěrečné práce, jsem však povinen/povinna s nimi nakládat jako s autorským dílem a zachovávat pravidla uvedená v předchozím odstavci tohoto prohlášení.

Poř. č.	Datum	Jméno a příjmení	Adresa trvalého bydliště	Podpis
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				