Annual Reports in Translation: A Comparative Study

Monika Nidrlová

Bachelor Thesis
2012

Tomas Bata University in Zlín
Faculty of Humanities

Univerzita Tomáše Bati ve Zlíně
Fakulta humanitních studií
Ústav anglistiky a amerikanistiky
akademický rok: 2011/2012

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení:	**Monika NIDRLOVÁ**
Osobní číslo:	**H09532**
Studijní program:	**B 7310 Filologie**
Studijní obor:	**Anglický jazyk pro manažerskou praxi**
Téma práce:	**Výroční zprávy v překladu: Srovnávací studie**

Zásady pro vypracování:

Studium odborné literatury
Shromáždění výzkumného materiálu
Formulace hypotézy
Analýza materiálu a testování hypotézy
Vyvození závěrů

Rozsah bakalářské práce:

Rozsah příloh:

Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam odborné literatury:

Baker, Mona. 1992. In other words: A coursebook on translation. London: Routledge.
Friedlob, George Thomas, and Ralph E. Welton. 2008. Keys to reading an annual report. 4th ed. Barron's Business Keys. Hauppauge, N.Y.: Barron 's Educational Series.
Hatim, Basil and Jeremy Munday. 2004. Translation: An advanced resource book. New York: Routledge.
Munday, Jeremy. 2001. Introducing translation studies: Theories and applications. New York: Routledge.
Robinson, Douglas. 1997. Becoming a translator: An accelerated course. London: Routledge.

Vedoucí bakalářské práce: **PhDr. Katarína Nemčoková**
Ústav anglistiky a amerikanistiky

Datum zadání bakalářské práce: **30. listopadu 2011**

Termín odevzdání bakalářské práce: **4. května 2012**

Ve Zlíně dne 6. února 2012

doc. Ing. Anežka Lengálová, Ph.D.
děkanka

UNIVERZITA TOMÁŠE BATI VE ZLÍNĚ
Fakulta humanitních studií
L.S.
①

doc. Ing. Anežka Lengálová, Ph.D.
ředitelka ústavu

PROHLÁŠENÍ AUTORA BAKALÁŘSKÉ PRÁCE

Beru na vědomí, že

- odevzdáním bakalářské práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby [1];
- beru na vědomí, že bakalářská práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k nahlédnutí;
- na moji bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3 [2];
- podle § 60 [3] odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- podle § 60 [3] odst. 2 a 3 mohu užít své dílo – bakalářskou práci - nebo poskytnout licenci k jejímu využití jen s předchozím písemným souhlasem Univerzity Tomáše Bati ve Zlíně, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše);
- pokud bylo k vypracování bakalářské práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tj. k nekomerčnímu využití), nelze výsledky bakalářské práce využít ke komerčním účelům.

Prohlašuji, že

- elektronická a tištěná verze bakalářské práce jsou totožné;
- na bakalářské práci jsem pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.

Ve Zlíně 4.5.2012 ..... Michelova

1) zákon č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, § 47b Zveřejňování závěrečných prací:

(1) Vysoká škola nevýdělečně zveřejňuje disertační, diplomové, bakalářské a rigorózní práce, u kterých proběhla obhajoba, včetně posudků oponentů a výsledku obhajoby prostřednictvím databáze kvalifikačních prací, kterou spravuje. Způsob zveřejnění stanoví vnitřní předpis vysoké školy.

(2) Disertační, diplomové, bakalářské a rigorózní práce odevzdané uchazečem k obhajobě musí být též nejméně pět pracovních dnů před konáním obhajoby zveřejněny k nahlížení veřejnosti v místě určeném vnitřním předpisem vysoké školy nebo není-li tak určeno, v místě pracoviště vysoké školy, kde se má konat obhajoba práce. Každý si může ze zveřejněné práce pořizovat na své náklady výpisy, opisy nebo rozmnoženiny.

(3) Platí, že odevzdáním práce autor souhlasí se zveřejněním své práce podle tohoto zákona, bez ohledu na výsledek obhajoby.

2) zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, § 35 odst. 3:

(3) Do práva autorského také nezasahuje škola nebo školské či vzdělávací zařízení, užije-li nikoli za účelem přímého nebo nepřímého hospodářského nebo obchodního prospěchu k výuce nebo k vlastní potřebě dílo vytvořené žákem nebo studentem ke splnění školních nebo studijních povinností vyplývajících z jeho právního vztahu ke škole nebo školskému či vzdělávacího zařízení (školní dílo).

3) zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, § 60 Školní dílo:

(1) Škola nebo školské či vzdělávací zařízení mají za obvyklých podmínek právo na uzavření licenční smlouvy o užití školního díla (§ 35 odst. 3). Odpírá-li autor takového díla udělit svolení bez vážného důvodu, mohou se tyto osoby domáhat nahrazení chybějícího projevu jeho vůle u soudu. Ustanovení § 35 odst. 3 zůstává nedotčeno.

(2) Není-li sjednáno jinak, může autor školního díla své dílo užít či poskytnout jinému licenci, není-li to v rozporu s oprávněnými zájmy školy nebo školského či vzdělávacího zařízení.

(3) Škola nebo školské či vzdělávací zařízení jsou oprávněny požadovat, aby jim autor školního díla z výdělku jím dosaženého v souvislosti s užitím díla či poskytnutím licence podle odstavce 2 přiměřeně přispěl na úhradu nákladů, které na vytvoření díla vynaložily, a to podle okolností až do jejich skutečné výše; přitom se přihlédne k výši výdělku dosaženého školou nebo školským či vzdělávacím zařízením z užití školního díla podle odstavce 1.

ABSTRAKT

Tato bakalářská práce zkoumá naturálnost anglického jazyka přeložených výročních zpráv. Teoretická část obecně popisuje vývoj této oblasti zkoumání, metodologii a příklady jiných výzkumných prací s podobným zaměřením. Praktická část pak konkrétně studuje jazyk ve vybraných výročních zprávách přeložených do anglického jazyka srovnáváním s výročními zprávami originálně napsaných v anglickém jazyce.

Klíčová slova: překlad, translatologie, univerzálie v překladu, metodologie, analýza korpusu, žánr, výroční zprávy.

ABSTRACT

This bachelor thesis investigates the naturalness of the English language in the translated annual reports. Theoretical part generally describes the development of this area of study, the methodology and examples of other works with similar focus. Practical part then concretely studies the language of chosen annual reports translated into English language by comparing them with the annual reports originally written in English language.

Keywords: translation, Translation Studies, universals of translation, methodology, corpus analysis, genre, annual reports.

ACKNOWLEDGEMENTS

I would like to take the opportunity to thank my advisor, PhDr. Katarína Nemčoková, for her support and patience. I much appreciate the time which was devoted to my bachelor thesis and advices which were given to me. I also want to thank to my family who was supporting me during my studies.

CONTENTS

INTRODUCTION

The divergence of the language of translations and features which typically occur in them have been an object of many debates. However, with the development of electronic corpora and linguistic research software for processing collected data, many scholars confirmed the idea that translation is a distinct type of language which shares some features typical for all translations regardless of the source languages involved. Along with the investigation to the nature of language of translations four universal features of translation were defined, i.e. explicitation, simplification, normalization and levelling-out (Ulrych and Anselmi 2008, 261).

The present study focuses on the universal features occurring in the Czech-English translations of chosen annual reports. The evidence will be generated with the help of WordSmith Tools Software by comparing two corpora. The theoretical part will give the necessary background information for the latter analysis. The first chapter will describe the view on translation within the Translation Studies and the notion of translation universals. The second chapter will discuss the term corpus, what types of corpora can be compiled and also what are the criteria for building corpora in general. This chapter will also describe the WordSmith Tools software and its functions. Finally, the third theoretical chapter will be concerned with the genre of annual report, its characteristic features and the sections they consist of the mentioned sections will play a big role in compiling the corpus of the translated texts, since some Czech companies do not translate some sections of the annual report but write them originally in English.

The analysis itself implements the methodology from corpus-based research, i.e. is based on compiled corpora which are processed with the help of computer-assisted analysis techniques. The analysis compares basic statistics of each corpus as type-token ratio, standardized type/token ratio, mean sentence length and further studies frequency of pronouns and standard deviance of given features. These mentioned features are considered to be good indicators of translation universals. The analysis will generate the values of chosen features of translated annual reports and will compare them with the values of the non-translated texts. The main goal is to test the nature of translated annual reports by giving the empirical evidence. The hypothesis for the present study is that translated annual reports tend to have simpler style than the non-translated annual reports in the target language regarding the occurrence of chosen lexical items. Other hypothesis assumes that the translated texts tend to be more explicit. The last hypothesis assumes that translated

texts are more similar to each other than the non-translated texts. The supposed result is that translated texts have lower lexical variety, use less pronouns and are relatively homogenous texts in comparison to non-translated texts.

I. THEORY

1 TRANSLATION

The thesis studies translations by comparing them with non-translated texts. This chapter will introduce the theoretical framework of translation within the development of Translation Studies and will describe the so called translation universals which will be an object of study in practical part.

1.1 Translation and Translation Studies

It is generally acknowledged that the discipline of Translation Studies emerged in the 1960s. 'Translation Studies' is a field of study which describes, analyzes and theorizes the processes, contexts and products of the act of translation (Williams and Chesterman 2002, 2). Before the 1960s the focus was mainly on the source text and translation was rarely an object of study, and if so, it had a theoretical and prescriptive character stressing fidelity to the source text and translations were perceived as 'equivalents' of their source texts. However, equivalent is not possible in some instances of translation process and this fact led to searching for the causes of this phenomenon (Anderson 2011, 1028-1029).

Understanding that translantions create a distinct type of language different from that of their source texts and also original texts in the target language was significant shift from prescriptiveness (Eskola 2004, 83). William Frawley was one of the first who introduced the idea that the language of translation is a distinct type of language. He says that "the confrontation between source language and target language during the translation process results in creating a 'third code'" (Kruger 2002, 80). Baker explains this as "a kind of compromise between the norms or patterns of the source language and those of the target language". Thus, the language of translation is viewed as "a separate sub-language of the same language". This sub-language is also called 'the third language' by some scholars, or 'translationese' by others (Yajun and Zaixin 2008, 26-27).

With the development of corpus linguistics and scientific branch of Translation Studies, translations started to be described empirically and the focus was on finding objective statements and laws about translations (Anderson 2011, 1027 and Cyrus 2009, 88-89). According to Toury, this new descriptive translation studies approach helps to study translations within the target culture and considering translation as a target language utterance supporting the idea that translations belong to the target textual system. Thus, this target-oriented approach wants to study translation in its own right rather than give primacy to the source text and to the notion of equivalence (Masubelele 2002, 29).

This new descriptive approach was extensively developed with the recent production of large machine-readable linguistic corpora in the 1990s which gave rise to a new area of study known as Corpus-based Translation Studies. (Anderson 2011, 1027; Kruger 2002, 70) It was Baker who came with the idea that corpora can be used for studying translations. Baker also predicted that with the development of large-machine readable corpora and implementation of methodology from corpus linguistics will be possible to study the nature of translations (Eskola 2004, 84).

1.1.1 Corpus-based Translation Studies

Corpus-based Translation Studies (CTS) were introduced in 1990s with the progression of descriptive translation studies and draws up from descriptive translation studies and corpus linguistics (Masubelele 2002, 36-37). Corpus-based translation scholars consider translation as a specific kind of text production (Olohan 2002, 422) and examine translations with the help of theory, diverse data and flexible methodology from corpus linguistics (Kruger 2002, 70, 79). With the growing number of available corpora in many languages, corpus-based studies redirected its focus from previous descriptive research into translation universals (Laviosa 2008, 225).

1.2 Translation universals

Baker proposes various types of corpus-based research, where she emphasizes especially the importance to understand "the nature of translated text as a mediated communicative event." This type of research tries to identify the so called 'universal features of translation', "i.e. features which typically occur in translated texts rather than in original texts and which are independent of the influence of the specific language pairs involved in the translation" (Masubelele 2002, 42-43). Thus, translations are here "specific kinds of texts that are not only different from their original source language (SL) texts, but also from comparable original texts in the same language as the target language (TL)" (Hansen and Teich, 2001). Investigating translation universals, the idea was that translated texts are generally "simpler and more conventional than both their source texts and comparable texts originally produced in the target language" (Jantunen 2004, 101; Benardini and Zanettin 2004, 53).

According to Chesteman, the prime benefit of investigating universals has been methodological. Corpus-based research into translation universals led to adopting "standard scientific methods of hypothesis generation and testing." Other benefits are

certainly bringing new knowledge about translations, testing new hypothesis and a fact, that universals substantiated by empirical evidence can have explanatory force regarding the occurrence of a given feature in a given translation (Chesterman 2004, 46; Laviosa 2008, 224).

Chesterman also distinguished two types of universals, i.e. *S-universals* and *T-universals*. S-universals refer to "universal differences between translations and their source texts" (Laviosa 2008, 224) and this type of universals is investigated by comparing translations with their source texts (Pym 2011, 93). T-universals, i.e. "universal differences between translations and comparable non-translated texts" (Laviosa 2008, 224) are investigated by comparing translations with non-translated texts in the same language (Pym 2011, 96).

A lot of work has been done on comparing corpora of translated and non-translated texts in the target language so far, and that is valid mainly for translated and non-translated texts in English. The patterns of English translations have been largely investigated with the help of *Translational English Corpus* (TEC) found by Mona Baker at the University of Manchester. TEC serves as a large database of translated texts from many source languages into English and enables to investigate translated English on a large-scale corpora (Xiao, He and Yue 2008, 1-2). On the other hand, *British National Corpus* (BNC) is an example of huge amounts of non-translated texts written in English (British National Corpus 2012).

With the investigation four main categories of translation universals were defined, and these are *explicitation, simplification, normalization* and *levelling-out.* Each of these will be described separately in the following subchapters (Ulrych and Anselmi 2008, 261).

1.2.1 Explicitation

One of the features which is regarded as a typical pattern of translated texts is explicitation. Explicitation is a "tendency of translators to 'spell things out' rather than leaving them implicit, including the practice of adding background information" (Ulrych and Anselmi 2008, 261). In the translation process, explicitation is seen as a translation technique using shifts from the source texts related to content and structure. When studying translation as a product, explicitation is understood as a text feature, which contributes to a "higher level of explicitness in comparison with non-translated texts." In practice, these linguistic features can be manifested by their higher frequency than in non-translated texts and also in added linguistic information (Puurtinen 2004, 145). The tendency of explicitness can be caused

by the stylistic preferences and systemic differences between source and target languages involved as well as by the translation tradition involved (Laviosa 2008, 227).

Explicitation appears in translated texts on syntactical and lexical level. Possible measure is a "sentence and text length, explanatory vocabulary, optional words and subordinators" (Yajun and Zaixin 2008, 28). Olohan and Baker for instance uses concordances from the TEC and the BNC corpora to compare the use of *that*-connective with reporting verbs *say* and *tell*. The results showed a higher usage of the *that* -connective in the translated texts of TEC, meanwhile in the BNC, there was a higher occurrence of zero-connective. These findings strongly support the syntactic explicitation. An example of lexical explicitation can be Olohan's study of intensifiers *quite, rather, pretty* and *fairly* in translated and non-translated texts in English. The approach was to investigate the relationship "between collocation and moderation". Results showed lower usage of the mentioned intensifiers in translated texts, moreover, the intensifiers are used in more variations and the common collocates are less repeated in comparison to non-translated English texts (Xiao, He and Yue 2008, 3).

1.2.2 Simplification

Explicitation is related to simplification. Simplification appears in translated texts as a "tendency to simplify the language used in translation". That means that language of translations tends to be simpler than non-translated language and it can be described "lexically, syntactically or stylistically" (Xiao, He and Yue 2008, 7). Baker formulates for instance this simplification hypothesis:

"Translations tend to use simpler language than original texts in the same language as the TL, most likely to optimize the readability of the target language text. Possible measures for simplification are average sentence length, lexical density and type-token ratio, the latter being a standard measure for the vocabulary variation in a text" (Hansen and Teich 2001).

Translators appear to use less variety of words and therefore "use a relatively higher proportion of high-frequency lexical items" and that is valid for translations independent on their source language (Laviosa 2008, 226). Translators also seem to break up long sentences to more and shorter sentences (Yajun and Zaixin 2008, 29). According to Yajun and Zaixin, vocabulary richness or vocabulary variety is considered to be a good indicator of simplification in translations (2008, 29).

Laviosa (1999) for example studies translated newspaper articles into English in comparison with non-translated English newspaper articles to investigate distinctive features between them. Laviosa revealed several patterns of translated texts independent on their source text, as that translated articles have "a relatively lower proportion of lexical or content words versus grammatical words", and "a higher proportion of frequent words versus less frequent words." Other finding was that most frequent words appeared in translated texts more often and their nucleus was less varied, the mean sentence length was lower and the present tense of the auxiliary verbs *to be* and to *have* appeared also more often in the translated texts. By saying all this, the work of Laviosa significantly supported the simplification hypothesis.

In her later study of translated and non-translated texts of narrative prose, Laviosa revealed other typical patterns of translated texts which she calls 'core patterns' of lexical use. The core patterns which Laviosa found in her study are follows:

- "Translated texts have a relatively lower percentage of content words versus grammatical words (i.e. their lexical density is lower);
- The proportion of high frequency words versus low frequency words is relatively higher in translated texts;
- The list head of a corpus of translated texts accounts for a larger area of the corpus (i.e. the most frequent words are repeated more often);
- The list head of translated texts contain fewer lemmas" (Xiao, He and Yue 2008, 92).

1.2.3 Normalization

Normalization is defined as "the tendency to exaggerate features of the target language and to conform to its typical patterns". Toury says that some linguistic forms and structures occur in translations, which are present in non-translated texts very rarely. Normalization is represented by the use of typical grammatical structures, collocational patterns and punctuation of the target language (Xiao, He and Yue 2008,8-9) Lexical normalization has been investigated for instance by Toury who chose the monolingual contrastive approach to study features of translated and non-translated Hebrew. It was suggested that the

monolingual comparable corpora can be used for investigating normalization at the level of collocation (Kenny 1998, 516).

1.2.4 Levelling-out

Another feature of translation universals is levelling-out. The term levelling-out stands for expression, that translated texts tend to be closer related to each other in terms of lexical density, type-token ratio and mean sentence length in comparison with non-translated texts in the target language (Hansen and Teich 2001).

In conclusion, translation is no longer understood as a mere copy of the original stressing the importance of fidelity, but is viewed as a special kind of text production. With the help of methodology and knowledge from corpus linguistics, translation scholars have been investigating the nature of translated text and unveiled some typical features occurring in the translated texts which were later labeled as universals of translation. These are basic background facts which are important to mention since the practical part will deal with the language of translation as a special kind of text production and will focus on universals features in the translated annual reports.

2 CORPORA IN TRANSLATION STUDIES

Practical part will study translated and non-translated texts by their comparison. This will be done with the help of corpora and by corpus linguistic tools. Therefore this chapter will provide general information about corpora, describe types of corpora and criteria applied when building corpora. This chapter will also speak about WordSmith Tools, which is the software used for processing corpora in practical part.

Traditionally, a *corpus* was understood as "a collection of naturally occurring language data" (McEnery 2003, 449) and the term was used for "any collection of writings by a specific author" (Kruger 2002, 71). With the development of computers the term *corpus* has widen its meaning to a "collection of texts held in electronic form which can be read and analyzed automatically or semi-automatically rather than manually" (Fernandes 2006, 88).

In Corpus-based Translation Studies a corpus is connected with these attributes: "*electronic form, size*, and *representativeness*" (Fernandes 2006, 88). Biber emphasizes that "a corpus in not simply a collection of texts. Rather, a corpus seeks to represent a language or some part of a language" (Pastor and Seghiri 2009, 87). However, from the concept of representativeness there is yet not a clear description that distinguishes a corpus from any other kind of collection, nor does it say what the ideal size of the corpus is (Seghiri 2012, 377).

Representativeness is tightly related to size of the corpus (Rea Rizzo 2010, 6). Corpus has been connected with "vast quantities of data extracted from large collections of text" (Fernandes 2006, 88) to ensure the reliability of the results in the corpus-based studies. Nevertheless, in some studies relatively small corpora were used to test some hypothesis. Some scholars argue that there should not be a limit on corpus size, since there is a limit in available texts for research and it can also be difficult to compile these texts which would fulfill established selection criteria (Rea Rizzo 2010, 6).

With the greater availability and use of corpora in 1990s, "automatic techniques for *annotating* language data with information about its linguistic properties" were developed (Ide 2004, 290). There is a disagreement in the issue of building corpora. Some linguists say that to consider a collection of texts as a corpus they have to be properly annotated. But others argue that sometimes it is unnecessary for their purposes or even impossibility to demand annotation (Maia 2008, 60).

2.1 Types of corpora

According to terminology of Baker , there are three types of corpora for translation research and these are *parallel, multilingual* and *comparable* corpora.

Parallel corpora consist of "original, source language-texts in language A and their translated versions in language B." *Multilingual corpora* are described as "sets of two or more monolingual corpora in different languages, built up either in the same or different institutions on the basis of similar design criteria" (Fernandes 2006, 90).

Comparable corpora are corpora which "consist of two separate collections of texts in the same language - one corpus consists of original texts in the language in question and the other consists of translations in that language from a given language or languages" (Fernandes 2006,90). In practice, large bodies of translations in a given language are compared to large bodies of non-translated texts in the same language and this is how the differences between translations in a given language and non-translated texts in the same language and the naturalness of the translations can be examined. This type of quantitative research investigates relative distinctions regarding occurrence of specific features in a text (Williams and Chesterman 2002, 7).

By comparing large bodies of texts we are able to test the particular phenomena of translations suggested by some scholars (Yajun and Zaixin 2008, 27). Baker states that comparable corpora are very important tool for identifying patterning of translated texts which is not depend on the source and target languages involved. The texts are usually comparable with respect to genre, time, when the texts were published but also with respect to text type and length of the text (Jantunen 2004, 106).

According to Laviosa, these corpora are believed to be a "fruitful resource for the systematic study of the product and the process of translation". For these purposes several corpora were built and explored to test the differences between translated and non-translated texts from the point of view of lexical, syntactic and other features. Number of studies using these corpora confirmed the notion of the third code of the translated language (Yajun and Zaixin 2008, 27).

2.1.1 Subject area

We can distinguish two types of corpora according to the aims of study in which they are used: *linguistic* and *translational*. *Linguistic* corpora are used in study of language, while

translational corpora are those by which we investigate the translation products and processes (Fernandes 2006, 92).

2.1.2 Domain

When speaking about corpora in respect of its investigation, we distinguish *general* and *specialized* corpora. General corpora are compiled to investigate translated language as a whole (Fernandes 2006, 92-93). A general corpus is usually compiled to serve as a reference corpus for contrastive analysis or to describe language in general. Therefore a general corpus usually includes texts from many genres and topic areas, as well as spoken and written English (Rea Rizzo 2010, 3). An example of such corpus can be the BNC consisting of over than 100-million words and which size is still growing (Mohammadi 2007).

Specialized corpora are used for more specific researches, who study the translated language of specific genres or text-types (Fernandes 2006, 92-93). This kind of investigation is useful for creating dictionaries, studying development of language, or for analyzing language of specific subject domain. Specialized corprora have to be compiled in a way that the researchers get enough samples of whatever they are studying to make the research reliable and not just illustrative (Rea Rizzo 2010, 3). An example of this type of corpora is a 5-million word Cambridge and Nottingham Corpus of Discourse in English (CANCODE) or the Michigan Corpus of Academic Spoken English (MISCASE) (Mohammadi 2007).

2.1.3 Mode

As to the composition of the corpus, we have to decide whether the corpus will consist of written or spoken texts or both (Fernandes 2006, 93).

2.1.4 Temporal restrictions

Depending on restrictions of a time period, we distinguish *synchronic* and *diachronic* corpus. Synchronic corpus is a static collection of texts, which are written at one particular point in time to represent the language within this time span (Fernandes 2006, 93; Rea Rizzo 2010, 7). Diachronic corpus, on the other hand, is dynamic collection of texts from different periods of time used for study of language changes and development (Fernandes 2006, 93; Rea Rizzo 2010, 7).

2.1.5 Number of languages

We can also distinguish corpus according to the number of languages covered in the corpus. *Monolingual* corpus consists only from one language, *bilingual* from two languages and *multilingual* corpus consists of more than three languages (Fernandes 2006, 93).

2.2 Corpus linguistic tools

Corpora are considered to be a very useful tool for linguistic analysis (Kruger 2002, 79). Translation Studies has also made use of corpus linguistic tools for studying translated texts and their prototypical translation patterns (Alves and Vale 2011, 105).

The study in a practical part uses the software WordSmith Tools (version 6) created by Mike Scott, and therefore it is important to describe the tools provided by this software. WordSmith Tools is a suite of programs for finding and studying word patterns and frequencies in texts or whole corpora (Berber Sardinha 1999, 1; Scott 2011). The tools of the WordSmith software are *WordList, Concord* and *KeyWords*. These tools will be described in detail in the following subchapters.

2.2.1 WordList

WordList is a program for creating word lists. This tool provides a list of all words or word-clusters in chosen text or texts, where the words can be ordered by frequency or alphabetically (Scott 2011). "Word frequency information helps with identifying characteristics of a text or of a genre" (Scott 2001, 47).

This tool enables to study simply the type of vocabulary used, to identify word clusters, to compare the frequency of a word in different text files or across genres, or to get a concordance of one or more of the words in a list (Scott 2011). By this tool it is also possible to compare two word lists.

WordList also provides basic statistics as number of tokens (running words) in a text, types (distinct words), type/token ratio (TTR), mean word length (in characters), how many words there are in each text (one-letter, two-letter words and so on), number of sentences and mean sentence length, and other statistics (Scott 2012).

Saved word-lists can be also used as input to the KeyWords program, which analyses the words in a given text and compares frequencies with a reference corpus for the purpose to generate lists of "key-words" or "key-key-words" (Scott 2011).

2.2.1.1 The type/token ratio

The type/token ratio (TTR) indicates a lexical complexity on the surface of a text or in other words, relative richness of vocabulary. The term *token* means the total number of running words in a text, and the term *type* refers to a number of distinct word-forms (Pâpai 2004, 157; Pym 2011, 96). The type/token ratio is calculated by dividing the number of tokens by the number of types. The higher the percentage the more varied the vocabulary and that means that there is a little repetition of words. It is important to mention, that the type/token ratio is very sensitive to length of the text or corpus. It is probable that the words will be repeated more often in the longer text and thus lowering the ratio (Kruger 2002, 74). Therefore there is also another strategy how to measure the lexical complexity, and that is the *standardized type/token ratio* (STTR). STTR is computed every **n** words (n=1000 by default) and the STTR will be an average type-token ratio based on 1000-word chunks of text. The STTR is a percentage of types for every n tokens. This strategy allows comparing type/token ratios across texts of different lengths (Scott 2011).

2.2.2 KeyWords

KeyWords program is a program which helps to identify the "key" words in one or more texts. The term *key word* is not defined in Linguistics, but for the program purposes it is a word with unusually high frequency in comparison with some reference corpus. Computed words can have *positive* or *negative* keyness. When the key word is positive, it means that the word occurs more often than would be expected when comparing with the reference corpus. Negative key words on the other hand occur less often than would be expected by the comparison with the reference corpus. Key words are useful way to characterize a text or genre (Scott 2011).

2.2.3 Concord

The WordSmith Tools program *Concord* locates all references to any search word or phrase within given corpus. The references are displayed in standard concordance lines showing the search word in the centre and within its context. With a help of this tool we can further examine collocates of the studied search word, dispersion plots, which show where in the texts the search word occurs, and also lists of recurring clusters or phrases (Scott 2001, 47).

To conclude this chapter, first it was introduced what are corpora and what features they have within Translation Studies. It was also described what types of corpora can be compiled and used for translation research and some other criteria which are important to consider when building corpora according to a purpose of study. Other focus of this chapter was on corpus linguistic tools for processing data, the WordSmith Tools software in particular. The functions were described for each of three tools to make easier understanding of the practical part where the compiled corpora will be processed by these programs.

3 ANNUAL REPORTS

The practical part will study the texts of annual accounting narratives in the genre of annual report. For this reason this chapter will briefly describe the genre generally and also the genre of annual reports in particular. The goal of this chapter is mainly to mention the sections of the narrative part of annual reports which will be used as texts for the analysis.

The notion of genre was first used in humanities for a literary criticism. For linguists and social theorists of communication nowadays the term genre constitutes "a class of communicative events, the members of which share some set of communicative purposes" (Rutherford 2005, 352).

A genre comprises traditional literary genres such as "drama, poetry and prose fiction". In translation studies, genre is also connected to clearly defined text types intended for translation such as tourism texts, legal documents or business texts (Williams and Chesterman 2002, 9; Munday 2009, 193) where each type has its goal and also some norms which determine what can and what should not be used in the given genre settings (Hatim and Munday 2004, 88).

3.1 Genre of annual reports

The annual report is perceived as a "highly specialized text" (Zanola 2009, 2). They belong to a reporting genre, particularly business reports where the typical goal is "reporting on overall perspective on an organization" (Bhatia 2004, 81-82). Annual reports are made up of several sections, each of which has specific features and creates thus other subgenre within the genre of annual report (Zanola 2009, 6).

Annual report is divided into narrative and financial part (Deloitte 2009, 8). The first narrative part which will be an object of the analysis includes all or some of these sections:

- "Chairman's statement
- Chief executive's statement
- Business review
- Financial review
- Corporate social responsibility statement (CSR)
- Directors' report
- Corporate governance statement

- Directors' remuneration report
- Statement of directors' responsibilities"(Deloitte 2009, 8).

The accounting narratives included in annual report have to make a connection between the narrative part and the financial statements and also to follow some regulations required (Rutherford 2005, 51). The form of accounting narratives is also influenced by lexical word choice and its frequency of use (Rutherford 2005, 51). These patterns of annual report narratives represent the genre of business communication (Rutherford 2005, 50).

In conclusion, annual reports belong to the genre of business reports and are divided to two parts- i.e. narrative and financial part. The narrative part consists of other sections each of which has its specific features. Since this thesis studies the language of the annual report rather than the genre itself, the information given in this chapter only gave necessary information about how the narrative part of annual reports are divided for better understanding of the compilation of the corpus for analysis. More detailed information about the narrative part of both, Czech and UK annual reports will be given in the analysis itself.

II. ANALYSIS

The theoretical part gave necessary information needed for better understanding of the practical part. The practical part will analyze the language of English translated annual reports by comparing them with non-translated texts of the annual reports written originally in English. The purpose is to unveil differences between these two regarding the occurrence of chosen typical features with the help of WordSmith Tools 6. Firstly, however, the process of building the corpus is outlined.

4 BUILDING THE CORPORA

This chapter will describe criteria applied for the compiling corpora of translated and non-translated annual reports.

4.1 Type of corpora

The present study uses *comparable corpora* for the analysis, i.e. two corpora, where one consists of annual reports translated into English and the second of annual reports originally written in English. The advantage of comparable corpora is their availability. It is much easier to find original texts and comparable translated texts than original texts and their translations (Maia 2003, 2).

4.2 Subject area

The aim of this analysis is to study the translation as a product. Therefore the used corpora are considered as *translational* rather than *linguistic*.

4.3 Domain

Since the analysis will study the translations of specific genre, i.e. annual reports, and the number of the annual reports used for the investigation will be of lower size, the used corpora can be marked as a *specialized corpora*.

4.4 Mode

The used corpora are composed only from written texts. Sometimes it is difficult to decide whether the texts are written or spoken, since some texts are written to be spoken or vice versa. Nevertheless, this study assumes that the annual reports are produced with the aim to be read.

4.5 Temporal restrictions

The corpora are *synchronic*, consisting only of annual reports of the year 2010. Thus, it is a static collection of texts representing the language at one particular point in time.

4.6 Number of languages

The annual reports consist only of English texts, i.e. they are monolingual corpora.

4.7 WordSmith Tools

The functions of the WordSmith Tools were described in the theoretical part. For the analytical section it is necessary to mention that the program has to be downloaded to the computer. To have an unrestricted access to the functions of the program a registration code is needed. Without the code, the WordSmith works only as a demo version which not shows full information about the investigated corpora.

In conclusion, the present study uses specialized monolingual comparable corpora, which are focused on the translation product written at one particular time period.

5 ANNUAL REPORTS

Each corpus consists of nine annual reports, i.e. nine annual reports translated into English and nine annual reports originally written in English. The detailed description of the annual reports and the chosen texts from them will be given in this chapter.

5.1 Choosing the annual reports

All of the used annual reports were found on the internet as the readable PDF files. The annual reports were taken from many sources, mainly from the official websites of the companies but the selection was also supported by a selection from special databases of the annual reports.

Translated annual reports:	Non-translated annual reports:
Telefónica O2 Czech Republic, a. s.	Bisichi Mining plc
AL INVEST Břidličná, a. s.	Iomart Group plc
CETELEM ČR, a. s.	James Fisher and Sons plc
Československá obchodní banka, a. s.	London & Associated Properties plc
Česká exportní banka, a.s.	London Capital Group Holdings plc
České aerolinie a. s.	Oxford BioMedica plc
HOCHTIEF CZ a. s.	Premier Foods plc
Komerční banka, a. s.	UBM plc
UNIPETROL, a. s.	Wilmington Group plc

The selection was mostly random; however, some obstacles appeared when searching the texts. First, it was important to consider, that some Czech companies do not translate some parts of the annual report but write them originally in English. That is valid mainly for the financial part, where there is a lot of special terminology and it is easier for the companies to write them initially in English and then translate them into the Czech language. To eliminate the error of considering original English as an English translation and thus misinterpreting the results, the financial part was not covered in the chosen texts. Other obstacle regarding the selection of the translated annual reports was the fact that many Czech companies do not state that the English version of the annual report is a translation. In this case an assumption was applied that if the annual reports are written according to the Czech law and standards, their initial language of the narrative part is the Czech language.

Other aim when compiling the corpora was to built corpora of similar size. That is due to the fact that with balanced corpora it will be possible to study the differences between them.

The present analysis works with annual reports from many fields of business. The annual reports are not only from one field of business, for instance *Telecommunication*, because then it could cause that some words would appear more frequently than it would be normal in a corpora of mixed annual reports and thus, giving a distorted picture.

5.2 Extracting the texts

The WordSmith Tools works with texts in a format of a plain text (.txt). Therefore it was necessary to extract the texts from the readable PDF files. This was done manually by copying the chosen text to the Word document and after small editing stored as a plain text.

5.2.1 Chosen texts

As pointed out earlier, the texts are built only from the narrative parts of the annual reports. Other aspect applied when selecting the texts was the form. The texts for the analysis consist only of whole sentences. The *sentence* is here understood as a grammatical unit consisting of words, which starts with a capital letter and ends with a punctuation mark. Thus, any other form of a text such as headlines, the text after a bullet or any other text not meeting the above definition of a sentence, was not comprised in the corpora.

5.3 Storing the texts

After choosing the texts and extracting them from the PDF files to the plain texts they had to be stored. Translated and non-translated texts were each stored to a separate folder. The text of each annual report was also stored as a separate file. The name of the file consists of the name of the company, the language or its variety, and the year of the annual report. For the translated texts it is this form *ALINVEST_CZ_EN_2010*. *CZ_EN* indicates that the annual report of Alinvest for the year 2010 is written originally in Czech and translated into English. In the case of non-translated text the file name looks as following *BISICHI_EN_2010*. Again, from the file name it is clear that it is an annual report of the company Bisichi for the year 2010, originally written in English.

The information provided in the name of the file is an important factor when working with the texts. It helps to have a better orientation in the text files and to identify the properties of the text files. Moreover, the corpora may be used in some later studies and therefore the

provided information of the file name would make the selection of the texts easier. Table 1 shows the name of each text file and its characteristics of the file size and number of tokens.

Table 1. List of the text files, their file size and number of tokens in each text

Text file (_CZ_EN)	File size	Tokens (running words)	Text file (_EN)	File size	Tokens (running words)
02	212,667	34,559	BISICHI	45,876	7,495
ALINVEST	18,336	2,916	IOMART	36,348	5,608
CETELEM	63,267	10,107	JFS	80,821	12,113
ČSOB	74,686	11,733	LAP	65,690	10,037
ČEB	59, 043	9,492	LCG	54,441	8,228
ČSA	54, 268	8,642	OXFORD BIOMETRICA	132,324	18,947
HOCHTIEF	21,129	3,154	PREMIER FOODS	153,992	23,741
KB	210,363	30,828	UBM	172,067	25,594
UNIPETROL	172,997	27,995	WILMINGTON	92,475	13,064
Overall	886,756	139,426	Overall	834,034	125,466

Source: WordSmith 6, Wordlist

The overall number of tokens serves here as the indicator of the size. When comparing the size of the two types of corpora, i.e. 137,426 tokens for the translated texts and 125,466 tokens for the non-translated texts, it can be considered as a relatively same size and therefore a balanced corpus.

One could wonder why there is not the same number of tokens for each text file and thus having perfectly balanced corpora. For this study the important factor is that the text file covers all the narrative part of the annual report rather than only the extracts. By covering all the narrative parts, the texts will be better balanced regarding their contents and thus each text file will have the same opportunity to share some lexical patterning which can be specific for a given section in the narrative part.

In conclusion, the annual reports were collected on the internet in the form of PDF files. Then according to the criterion of the sentence form within the narrative part the text was being chosen and copied to the Word document and stored as a plain text with a file name showing the properties of the text. This is the procedure of preparing the corpora for the analysis.

6 ANALYSIS OF THE TEXTS

The analysis will compare translated and non translated annual reports by WordSmith tools with the aim to investigate the translation universals. For this purpose a methodology of hypotheses of Laviosa shown in the Table 2 was implemented, where hypotheses for specific translation universal features are described and a suitable tool suggested.

Table 2. Methodology of hypotheses to investigate the so-called universal translation features

Translation feature	**Hypothesis**	**Tool**
Simplification	Translated English texts have a significantly lower type-token ratio than original language English texts.	Wordlist statistics
	Translated English texts have higher frequency of most frequent words and the most frequent words cover larger area of a corpus.	Wordlist
	Translated English texts have lower average sentence length than original language English texts.	Wordlist statistics
Explicitation	Translated English texts have a significantly lower frequency of pronouns than original language English texts.	Wordlist
Normalization	Translated English texts have lower Standard Deviation than non-translated texts.	Wordlist statistics

Source: Laviosa, Sara. 1998. Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta* 43, no. 4: 557-570.

6.1 Type-token ratio

The hypothesis of type-token ratio says that the type-token ratio of translated English is lower than the ratio of non-translated English. In practice it means that the translated English has less varied vocabulary and it is thus simpler. It will be investigated by the WordSmith program Wordlist, where the type/token ratio is displayed.

Table 3. Statistics of type/token ratio

	TTR ratio of translated texts	TTR of non-translated texts
Tokens (running words)	139,426	125,466
Tokens used for word list	133,039	119, 259
Types (distinct words)	11,442	14,157
Type/token ratio	8.60	11.87

Source: WordSmith 6, Wordlist

The type-token ratio for translated texts is 8.60 and for the non-translated texts 11.87. These results confirm the simplification hypothesis of the lower ratio for the translated texts.

Nevertheless, since the type/token ratio is very sensitive to the length of the text and the corpus in the analysis consists of short texts of different lengths, the type/token ratio can give a distorted result. Table 4 shows more detailed statistics of each text file in the corpus.

Table 4. Statistics of translated (_CZ_EN) and non-translated texts (_EN)

Translated texts				Non-translated texts			
Text file (_CZ_EN)	Tokens used for word list	Types	Type/ token ratio	Text file (_EN)	Tokens used for word list	Types	Type/ token ratio
O2	32,532	3,591	11.04	BISICHI	7,213	1,324	18.36
ALINVEST	2,800	914	32.64	IOMART	5,294	1,464	27.65
CETELEM	9,806	1,647	16.80	JFS	11,594	2,773	23.92
ČEB	9,054	1,818	20.08	LAP	9,433	2,490	26.40
ČSA	8,340	1,737	20.83	LCG	7,843	2,196	28.00
ČSOB	11,109	2,339	21.06	OX.BIO.	18,114	4,193	23.15
HOCHTIEF	3,041	1,236	19.69	PR.F.	22,583	4,260	18.86
KB	29,534	5,962	40.64	UBM	24,121	4,952	20.53
UNIPETROL	27,995	3,419	20.19	WILMIN.	13,064	2,925	22.39

Overall	133,039	11,442	8.60	Overall	119,259	14,157	11.87

*Source:*WordSmith 6, Wordlist

Looking closer at the statistics of the text files, it is recognizable that the text files consisting of shorter texts have higher type/token than the longer texts. The reason can be the fact, that in longer texts the words appear more often a thus lowering the ratio.

Saying this, a better strategy how to measure the vocabulary variety would be to use the standardized type token ratio (STTR), which is the ratio computed every 1000-word chunk of text. Table 5 shows the overall STTR for both, translated and non-translated texts.

Table 5. Statistics of standardized type/token ratio

	Translated texts	Non-translated texts
STTR	38.55	43.12
STTR basis	1,000	1,000

Source: WordSmith 6, Wordlist

Table 5 shows that the translated texts have the ratio 35.55 and the non-translated 43.12. These results where the ratio is lower in the translated texts also support the hypothesis that the translated texts are simpler, i.e. their lexical variety is lower.

6.2 Frequency and proportion of high frequency words

Other hypothesis concerning the simplification says that there will be a higher frequency of most frequent words in the translated texts and that the most frequent words will cover larger area of the corpus.

The most frequent words are characterized here as words in a word list which have higher or the same frequency percentage in the word list as the value 0.1. (For the raw data from word lists see appendices I and II). The overall frequency of most frequent words is calculated as a sum of all frequencies of the chosen words. First, the word lists were created and then stored as an Excel document, where it was easier to calculate the value.

The proportion of the most frequent words in comparison to less frequent words was calculated by summing the percentage values of all chosen words. The percentage value indicates the frequency of a word as a percent of the running words in the text the word list was made from (Scott 2012). Again, the overall value for the proportion of high frequency words was calculated with the help of Excel program. Table 10 shows the results.

Table 10. Frequency and proportion of high frequency words

	Translated texts	Non-translated texts
Frequency of words	73,164	63,014
Proportion	52.48	50.22

Source: WordSmith 6, Wordlist

The table above shows that the overall number of most frequent words in translated texts is 73,164. This number is higher than in the non-translated texts, where there are 63,014 most frequent words. This result supports the hypothesis that the translated texts have more high frequent words. The result can be explained as a tendency of the translated English to use the frequent words more often, and on the other hand, using the less frequent words less often. This can be interpreted as that the words in translated texts are less varied.

Another result regarding the proportion of the most frequent words in the corpus shows that the most frequent words of the translated texts cover larger area of the corpus with the percent 52.48. For the non-translated texts the value is 50.22. These results also support the hypothesis that the proportion of the high frequent words is higher than in the non-translated texts.

6.3 Mean sentence length

The second hypothesis concerning simplification says that translated texts have lower mean sentence length. It assumes that the lower the mean sentence length, the simpler is the text. This fact can be caused by the translator's preference to break up long sentences to shorter sentences (Laviosa 1999, 310).

Table 6 shows the mean sentence length for both collections of texts.

Table 6. Mean sentence length

	Translated texts	Non-translated texts
Number of sentences	6,157	5,880
Mean sentence length (in words)	21.61	20.28

Source: WordSmith 6, Wordlist

The results do not confirm the second hypothesis. The mean sentence length is lower in non-translated texts with the value 20.28. Translated texts have a higher value; however the difference is not of big significance. Nevertheless the results can be interpreted as that the translated texts do not show tendency to simplify texts in this way. It can be caused by the natural tendency of business texts to be rather complicated in any language.

6.4 Pronouns

Other hypothesis concerning another feature of translation universal – explicitation, is that translated texts have a significantly lower frequency of pronouns than the original texts. Table 7 shows the pronouns found in a word list, their frequency and a percentage they represent from all the running words in a word list.

Table 7. Pronouns in translated and non-translated texts

Translated texts				Non-translated texts			
Place	Pronoun	Frequency	Percentage	Place	Pronoun	Frequency	Percentage
16	ITS	807	0.58	16	OUR	806	0.64
24	HE	533	0.38	18	WE	775	0.62
34	IT	365	0.26	37	ITS	359	0.29
53	THEIR	271	019	43	IT	289	0.23
68	OUR	204	0.15	50	THEIR	224	0.18
84	WE	159	0.11	63	HE	185	0.15

Source: WordSmith 6, Wordlist

The pronouns were selected from the word list manually. Only high frequency pronouns were chosen from the word lists, with the rule that pronoun with lower percentage of occurrence than 0.1 was not covered in the Table 6.

For both collections of texts, the forms of pronouns *its, he, it, their, our* and *we* were present in the word lists as more frequent items. Also the forms of the pronouns are the same for both collections. This can be caused by the specific genre of the annual report which is the same for all the texts.

To test the hypothesis of pronouns and explicitation, Table 8 gives overall results of the number of pronouns and their percentage of the frequency in the texts.

Table 8. Overall number and percentage of the frequency of the pronouns

	Number of pronouns	Percentage
Translated texts	2,339	1.67
Non-translated texts	2,638	2.11

Source: WordSmith 6, Wordlist

Table 8 shows that 2,339 pronouns occur in the translated texts and 2,638 in non-translated. These results do confirm the hypothesis that pronouns in translated texts have a lower frequency. This result can be caused by the fact that the translated texts tend to be more explicit and do not use the pronouns as often.

6.5 Standard Deviation

Standard deviation (SD) is a measure of how tightly all the results of each text are near to the mean value. In other words, SD measures the homogeneity of the texts. In practice, the lower the value, the more homogenous the group is and "the higher its level of textual conventionality". According to Laviosa, conventionality can be considered as a normalization feature (1999, 310-311).

It is assumed here that the translated texts will be more homogenous. The homogeneity will be measured by the SD of STTR and mean sentence length. Table 9 shows the results.

Table 9. Statistics of standard deviation of type/token ratio and mean sentence length

	Translated texts	Non-translated texts
SD of STTR	61.91	57.16
SD of mean sentence length	12.30	10.10

Source: WordSmith, Wordlist

The results show that the SD of both, STTR and mean sentence length is lower in the non-translated texts. This is against the hypothesis that translated texts are more homogenous. Thus the normalization feature was not confirmed in the translated texts.

7 RESULTS

The analysis studied the English language with regard to test some hypotheses of Laviosa (1998) concerning the universal features of translation, namely simplification, explicitation and normalization. Result of each investigated hypothesis will be described in detail in this chapter.

The first simplification hypothesis assumes that the translated texts will have a significantly lower type-token ratio than the non-translated texts. The two values for the type-token ratio were generated and the result showing the lower value of the type/token ratio thus confirmed the first hypothesis. To verify the correctness of the first simplification hypothesis another test was done with the more sophisticated strategy where the standardized type/token ratio was implemented to calculate the result. This second way of investigation between the translated and non-translated texts also confirmed the hypothesis. Thus, it can be concluded that the translated texts have lower lexical variety than the non-translated texts, i.e. is simplified.

The second simplification hypothesis assumes that translated texts have a higher frequency of most frequent words and that the most frequent words cover larger area of a corpus. The values for both, the frequency and the frequency percentage were obtained by summing the frequency and the percentage of the frequency of the chosen running words separately for the translated and non-translated texts. The results showing a higher number of most frequent words and also the higher percentage of frequency of the most frequent words confirm the second simplification hypothesis. Thus, these results can be interpreted as other instance of lower lexical variety in the translated texts supporting the result of the hypothesis of the lower type-token ratio.

The third simplification hypothesis assumes that the translated texts will have lower mean sentence length than the non-translated texts. The results showed that there was a little difference between the two values. Nevertheless, the value of the mean sentence length of the translated texts was relatively higher and thus the third hypothesis concerning the lower mean sentence length of the translated texts is refuted. Therefore it cannot be confirmed the idea that translators break up long sentences to shorter ones to make easier reading of the texts.

Another investigation studied the feature of explicitation. A hypothesis has been established that translated texts have a significantly lower frequency of pronouns than non-translated texts. First, the pronouns were extracted from the word lists. Then their overall

frequency was calculated. The values of the frequency showed that the pronouns occur in translated texts less often. This result thus confirms the hypothesis of explicitation saying that the translated texts tend to explain things in more detail then explain them more abstractly by the use of pronouns.

The last hypothesis concerns the normalization feature. The set hypothesis assumes that the translated texts have lower standard deviation value of STTR and mean sentence length than the non-translated texts. The values were obtained from the WordSmith program Wordlist. The results show that the standard deviation for both, STTR and the mean sentence length is higher in the translated texts, i.e. do not confirmed the hypothesis of the investigated normalization feature. Thus, it cannot be said that the translated texts are more homogenous than the non-translated texts.

In conclusion, the analysis investigated the features of translation where five hypotheses were set for testing. The results confirmed the simplification hypotheses and also the explicitation one. Thus, the translated texts of annual reports can be considered as texts of lower lexical variety using pronouns less frequently. On the other hand, the hypothesis of the normalization feature regarding the homogeneity of the translated texts was refuted and thus it can be said that translated annual reports were not normalized in this way.

CONCLUSION

The aim of this bachelor thesis was to study the language and the naturalness of translations. The theoretical part gave the necessary background information of the features occurring in translations and also the type of their investigation.

The present work focused on the comparison of translated and non-translated annual reports. The analytical part tried to test if some of the chosen typical translation features would occur also in the present corpora. Five hypotheses concerning the typical translation features were implemented from the study of translation universals of Sara Laviosa (1998).

By investigating each of the hypotheses the notion of the specific language of translations was confirmed. The results of the analysis showed that the translated annual reports have lower lexical variety and thus not using so many different types of words and also that the translated texts tend to be more explicit. On the other hand, the analysis found that the translated texts are not as similar to each other as it was assumed.

The aim of this thesis was not only to investigate the specific features of translation but also to faithfully describe the procedure of the whole process of the investigation. Author believes that this study can be used as a guide for similar studies for people who are not very familiar with this topic.

Nothing from the analysis could be done without using the specialized linguistic tool which is able to process the texts. The analysis used the WordSmith Tools. Without the software it would be very time consuming or even impossible to generate the results. The WordSmith was found very useful and also very easy to use. The present study was working only with one of the programs provided, i.e. Wordlist and this still not in full use. The author suggests further studying of the translation language by making use of more programs by which more concrete studies can be carried out. Giving more detailed empirical evidence would give deeper insights into the translation language and thus would bring better understanding of the language of translations.

The author believes that studying the language of translation can help not only to unveil the typical patterns occurring in them but this kind of study can make use in education and translating. Translators aware of the typical features may implement these facts to the translation process and by minimizing them they could make more natural translations. Students, on the other hand, could make use of the linguistic tools processing data for language learning.

REFERENCES

Alves, Fabio, and Daniel Couto Vale. 2011. "On Drafting and Revision in Translation: A Corpus Linguistics Oriented Analysis of Translation Process Data." *Translation Computation, Corpora Cognition* 1, no. 1: 105-22.

Anderson, Kristine J. 2011. "Translation Studies." *Choice* 48, no. 6: 1027-36.

Berber Sardinha, Tony. 1999. "Using Key Words in Text Analysis: Practical Aspects." *DIRECT Papers* 42, LAEL, Catholic University of São Paulo.

Bernardini, Silvia, and Federico Zanettin. 2004. "When Is a Universal Not a Universal? Some Limits of Current Corpus-based Methodologies for the Investigation of Translation Universals." In *Translation Universals*: *Do They Exist?,* edited by Anna Mauranen and Pekka Kujamäki, 51-62. Amsterdam: John Benjamins.

Bhatia, Vijay K. 2004. *Worlds of Written Discourse: A Genre-based View*. London: Continuum.

British National Corpus. 2012. "What is the BNC?" Accessed May 13. http://www.natcorp.ox.ac.uk/corpus/index.xml?ID=intro

Chesterman, Andrew. 2004. "Beyond the Particular." In *Translation Universals*: *Do They Exist?,* edited by Anna Mauranen and Pekka Kujamäki, 33-49. Amsterdam: John Benjamins.

Cyrus, Lea. 2009. "Old Concepts, New Ideas: Approaches to Translation Shifts." *Monografías de Traducción e Interpretación* 1: 87-106.

Deloitte. 2009. "A Telling Performance: Surveying Narrative Reporting in Annual Reports."

Eskola, Sari. 2004. "Untypical Frequencies in Translated Language: A Corpus-based Study on a Literary Corpus of Translated and Non-translated Finnish." In *Translation Universals*: *Do They Exist?,* edited by Anna Mauranen and Pekka Kujamäki, 83-100. Amsterdam: John Benjamins.

Fernandes, Lincoln. 2006. "Corpora in Translation Studies: Revisiting Baker's Typology." *Fragmentos*, no. 30: 87-95.

Hansen, Silvia, and Elke Teich. 2001. "Introduction to Multi-layer Analysis for Translation Corpora: Methodological Issues and Practical Implications." In *Proceedings of Workshop on Multi-Layer Corpora,* Iasi, Romania.

Hatim, Basil, and Jeremy Munday. 2004. *Translation: An Advanced Resource Book.* London and New York: Routledge.

Ide, Nancy. 2004. “Preparation and Analysis of Linguistic Corpora.” In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 289-305. Malden: Blackwell.

Jantunen, Jarmo Harri. 2004. “Untypical Patterns in Translations: Issues on Corpus Methodology and Synonymity.” In *Translation Universals*: *Do They Exist?*, edited by Anna Mauranen and Pekka Kujamäki, 101-26. Amsterdam: John Benjamins.

Kenny, Dorothy. 1998. “Creatures of Habit? What Translators Usually Do With Words.” *Meta: Translators' Journal* 43, no. 4: 515-23.

Kruger, A. 2002. “Corpus-based Translation Research: Its Development and Implications for General, Literary and Bible Translation.” *Acta Theologica: Supplementum* 2: 70-106.

Laviosa, Sara. 1998. “Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose.” *Meta* 43, no. 4: 557-70.

———. 1999. “The Design and Analysis of a Comparable Corpus of English Newspaper Articles.” U.M.I.S.T. Manchester.

———. 2008. “Corpus Studies of Translation Universals: A Critical Appraisal.” In *Investigating English with Corpora. Studies in Honour of Maria Teresa Prat*, edited by A. Martelli and V. Pulcini, 223-38. Italy, Monza: Polimetrica Publisher.

Maia, Belinda. 2003. “What Are Comparable Corpora?” Faculdade de Letras da Universidade do Porto.

———. 2008. “Tools for Educating Translators.” In *Topics in Language Resources for Translation and Localisation*, edited by Elia Yuste Rodrigo, 57-71. Amsterdam: John Benjamins.

Masubelele, Mthikazi Roselina. 2002. “Role of Bible Translation in the Development of Written Zulu: A Corpus-based Study.” D.Litt. et Phil. diss., University of South Africa.

McEnery, A. 2003. “Corpus Linguistics.” In *The Oxford Handbook of Computational Linguistics*, edited by Ruslan Mitkov, 448-63. Oxford: Oxford University Press.

Mohammadi, Maryam. 2007. “Specialized Monolingual Corpora in Translation.” *Translation Journal* 11, no. 2. Accessed May 14, 2012. http://translationjournal.net/journal/40corpus.htm (accessed May 14, 2012).

Munday, Jeremy ed. 2009. *The Routledge Companion to Translation Studies*. Rev. ed. London: Routledge.

Olohan, Maeve. 2002. “Corpus Linguistics and Translation Studies: Interaction and Reaction.” *Linguistica Antverpiensia*, no. 1: 419-29.

Pâpai, Vilma. 2004. "Explicitation? A Universal of Translated Text?" In *Translation Universals*: *Do They Exist?,* edited by Anna Mauranen and Pekka Kujamäki, 143-64. Amsterdam: John Benjamins.

Pastor, Corpas Gloria, and Miriam Seghiri. 2009. "Virtual Corpora as Documentation Resources: Translating Travel Insurance Documents (English-Spanish)." In *Corpus Use for Learning to Translate and Learning Corpus Use to Translate,* edited by Allison Beeby, Patricia Rodriguez Inés, and Pilar Sánchez-Gijón, 75-107. Amsterdam: Benjamins.

Puurtinen, Tiina. 2004. "Explicitation of Clausal relations: A Corpus-based Analysis of Clause Connectives in Translated and Non-translated Finnish Children's Literature." In *Translation Universals*: *Do They Exist?,* edited by Anna Mauranen and Pekka Kujamäki, 165-76. Amsterdam: John Benjamins.

Pym, Anthony ed. 2011. "Translation Research Terms: A Tentative Glossary for Moments of Perplexity and Dispute." *Translation Research Projects* 3: 75-110.

Rea Rizzo, Camino. 2010. "Getting on with Corpus Compilation: From Theory to Practice." *ESP World* 9, no. 1. Spain: Technical University of Cartagena.

Rutherford, Brian A. 2005. "Genre Analysis of Corporate Annual Report Narratives." *Journal of Business Communication* 42, no. 4 (October): 349-78.

Seghiri, Miriam. 2012. "Creating Electronic Corpora: Design, Compilation Protocol and Representativeness." In *Aspects of Literary Translation? Building Linguistic and Cultural Bridge in Past and Present*, edited by Eva Parra-Membrives, Miguel Ángel García Peinado, and Albrecht Classen, 373-81. Tübingen: Narr Verlag.

Scott, Mike. 2001. "Comparing Corpora and Identifying Key Words, Collocations, and Frequency Distributions through the WordSmith Tools Suite of Computer Programs." In *Small Corpus Studies and ELT:Theory and Practice*, edited by M. Ghadessy, A. Henry, and R.L. Roseberry, 47-67. Amsterdam: Benjamins.

———. 2011. "Lexical Analysis Software. WordSmith Tools: Software for Finding Word Patterns." http://www.lexically.net/downloads/version6/HTML/?wshell.htm

———. 2012. "WordSmith Tools." Liverpool: Lexical Analysis Software.

Ulrych, Margherita, and Simona Anselmi. 2008. "Towards a Corpus-based Distinction between Language-specific and Universal Features of Mediated Discourse." In *Investigating English with Corpora. Studies in Honour of Maria Teresa Prat,* edited by A. Martelli and V. Pulcini, 257-73. Italy: Polimetrica Publisher.

Williams, Jenny and Andrew Chesterman. 2002. *The Map: A Beginner's Guide to Doing Research in Translation Studies.* Manchester: St. Jerome Publishing.

Yajun, Jiang, and Ren Zaixin. 2008. "Translated English and Universals of Translation." *English Today* 24, no. 4 (December): 26-33.

Xiao, Richard, Lianzhen He, and Ming Yue. 2008. “In Pursuit of the Third Code: Using the ZJU Corpus of Translational Chinese in Translation Studies.” Zhejiang University.

Zanola. Annalisa. 2009. “The Annual Report: An Interdisciplinary Approach to a 'Contaminated' New Genre.” Paper presented at the international meeting *Genre on the Move: Hybridization and Discourse Change in Specialized Communication*, December 9, Napoli, Italy.

APPENDICES

P I Wordlist of translated texts.

P II Wordlist of non-translated texts

APPENDIX P I: WORDLIST OF TRANSLATED TEXTS

Place	Word	Frequency	Percentage
1	THE	11440,00	8,21
2	OF	7163,00	5,14
3	#	6387,00	4,58
4	AND	4470,00	3,21
5	IN	4261,00	3,06
6	TO	2797,00	2,01
7	A	2482,00	1,78
8	FOR	1605,00	1,15
9	S	1352,00	0,97
10	ON	1300,00	0,93
11	AS	1106,00	0,79
12	WITH	1035,00	0,74
13	BY	1004,00	0,72
14	IS	965,00	0,69
15	WAS	815,00	0,58
16	ITS	807,00	0,58
17	COMPANY	759,00	0,54
18	FROM	727,00	0,52
19	AT	724,00	0,52
20	BOARD	715,00	0,51
21	YEAR	649,00	0,47
22	CZECH	636,00	0,46
23	WHICH	535,00	0,38
24	HE	533,00	0,38
25	ARE	504,00	0,36
26	GROUP	494,00	0,35
27	CZK	474,00	0,34
28	ALSO	460,00	0,33
29	AN	434,00	0,31
30	TELEFÓNICA	422,00	0,30
31	THAT	415,00	0,30
32	BANK	401,00	0,29
33	HAS	399,00	0,29
34	IT	365,00	0,26
35	DIRECTORS	358,00	0,26
36	SUPERVISORY	353,00	0,25
37	THIS	343,00	0,25
38	SERVICES	339,00	0,24
39	UNIPETROL	337,00	0,24
40	WERE	336,00	0,24
41	FINANCIAL	329,00	0,24
42	OR	329,00	0,24
43	NEW	327,00	0,23
44	ALL	325,00	0,23
45	MARKET	325,00	0,23

46	KB	305,00	0,22
47	BUSINESS	299,00	0,21
48	REPUBLIC	298,00	0,21
49	MANAGEMENT	286,00	0,21
50	BE	284,00	0,20
51	MEMBERS	282,00	0,20
52	OTHER	279,00	0,20
53	THEIR	271,00	0,19
54	ČSOB	267,00	0,19
55	MEMBER	253,00	0,18
56	EMPLOYEES	248,00	0,18
57	GENERAL	248,00	0,18
58	NOT	247,00	0,18
59	RISK	247,00	0,18
60	CUSTOMERS	243,00	0,17
61	MILLION	220,00	0,16
62	COMMITTEE	219,00	0,16
63	MEETING	210,00	0,15
64	NUMBER	210,00	0,15
65	PRODUCTS	210,00	0,15
66	BEEN	209,00	0,15
67	CLIENTS	206,00	0,15
68	OUR	204,00	0,15
69	AUDIT	200,00	0,14
70	DEVELOPMENT	197,00	0,14
71	MOBILE	197,00	0,14
72	ACTIVITIES	188,00	0,13
73	WILL	182,00	0,13
74	BANKA	173,00	0,12
75	CREDIT	172,00	0,12
76	ONE	171,00	0,12
77	SINCE	171,00	0,12
78	CORPORATE	169,00	0,12
79	POSITION	169,00	0,12
80	MORE	166,00	0,12
81	KOMERČNÍ	165,00	0,12
82	AIRLINES	164,00	0,12
83	NETWORK	159,00	0,11
84	WE	159,00	0,11
85	DIRECTOR	155,00	0,11
86	DECEMBER	153,00	0,11
87	CHAIRMAN	152,00	0,11
88	WORKED	151,00	0,11
89	TOTAL	149,00	0,11
90	COMPANIES	143,00	0,10
91	NO	142,00	0,10
92	PRAGUE	141,00	0,10
93	PART	140,00	0,10
94	UNDER	139,00	0,10

95	END	138,00	0,10
96	WITHIN	137,00	0,10
97	INTO	136,00	0,10
98	THESE	136,00	0,10
99	INTEREST	135,00	0,10
100	BILLION	134,00	0,10

APPENDIX P II: WORDLIST OF NON-TRANSLATED TEXTS

Place	Word	Frequency	Percentage
1	THE	8685,00	6,92
2	#	6207,00	4,95
3	OF	4451,00	3,55
4	AND	3816,00	3,04
5	TO	3465,00	2,76
6	IN	2950,00	2,35
7	A	1995,00	1,59
8	IS	1315,00	1,05
9	FOR	1276,00	1,02
10	ON	997,00	0,79
11	ARE	908,00	0,72
12	WITH	893,00	0,71
13	AS	848,00	0,68
14	GROUP	846,00	0,67
15	S	808,00	0,64
16	OUR	806,00	0,64
17	BY	782,00	0,62
18	WE	775,00	0,62
19	COMPANY	716,00	0,57
20	AT	681,00	0,54
21	THAT	659,00	0,53
22	DIRECTORS	642,00	0,51
23	YEAR	608,00	0,48
24	WHICH	575,00	0,46
25	THIS	571,00	0,46
26	HAS	562,00	0,45
27	BE	494,00	0,39
28	HAVE	490,00	0,39
29	BOARD	464,00	0,37
30	FROM	461,00	0,37
31	WAS	450,00	0,36
32	AN	448,00	0,36
33	EXECUTIVE	425,00	0,34
34	WILL	373,00	0,30
35	FINANCIAL	364,00	0,29
36	OR	363,00	0,29
37	ITS	359,00	0,29
38	COMMITTEE	354,00	0,28
39	BUSINESS	338,00	0,27
40	SHARE	327,00	0,26
41	ALL	309,00	0,25
42	SHARES	292,00	0,23
43	IT	289,00	0,23
44	ALSO	275,00	0,22
45	NON	269,00	0,21

46	PERFORMANCE	259,00	0,21
47	BEEN	242,00	0,19
49	NOT	238,00	0,19
50	THEIR	224,00	0,18
51	MARKET	222,00	0,18
52	DIRECTOR	219,00	0,17
53	NEW	218,00	0,17
54	WERE	212,00	0,17
55	RISK	210,00	0,17
56	OTHER	202,00	0,16
57	THESE	202,00	0,16
58	REMUNERATION	199,00	0,16
59	UNDER	197,00	0,16
60	CHAIRMAN	194,00	0,15
61	MANAGEMENT	189,00	0,15
62	REPORT	187,00	0,15
63	HE	185,00	0,15
64	THERE	185,00	0,15
65	UK	184,00	0,15
66	NO	183,00	0,15
67	ANY	179,00	0,14
68	OVER	179,00	0,14
69	ANNUAL	176,00	0,14
70	MAY	174,00	0,14
71	CASH	172,00	0,14
72	UBM	170,00	0,14
73	GROWTH	169,00	0,13
74	MILLION	161,00	0,13
75	OUT	161,00	0,13
76	SHAREHOLDERS	157,00	0,13
77	SERVICES	153,00	0,12
78	AUDIT	150,00	0,12
79	YEARS	149,00	0,12
80	EACH	147,00	0,12
81	SET	146,00	0,12
82	DEVELOPMENT	141,00	0,11
83	PRODUCTS	138,00	0,11
84	REVIEW	137,00	0,11
85	DECEMBER	134,00	0,11
86	STATEMENTS	134,00	0,11
87	ORDINARY	132,00	0,11
88	VALUE	130,00	0,10
89	EMPLOYEES	129,00	0,10
90	MADE	129,00	0,10
91	MORE	126,00	0,10
92	OXFORD	126,00	0,10
93	US	126,00	0,10
94	FURTHER	125,00	0,10
95	UP	125,00	0,10

96	THAN	124,00	0,10
97	THROUGH	123,00	0,10
98	GENERAL	121,00	0,10