Filozofická fakulta Univerzity Palackého v Olomouci

Normalization in Original and Translated Czech: A Corpus Study

(Bakalářská práce)

2024 **Eliška Švecová**

Filozofická fakulta Univerzity Palackého v Olomouci
Katedra anglistiky a amerikanistiky

Normalizace v originální a překladové češtině: Korpusová Studie

(Bakalářská práce)

Normalization in Original and Translated Czech: A Corpus Study

(Bachelor thesis)

Autor: Eliška Švecová

Studijní obor: Angličtina se zaměřením na komunitní tlumočení a překlad

Vedoucí práce: Mgr. Michaela Martinková, Ph.D.

Olomouc 2024

Údaje o práci

Autor práce:	Eliška Švecová
Studijní obor:	Angličtina se zaměřením na komunitní tlumočení a překlad
Název práce česky:	Normalizace v originální a překladové češtině: Korpusová studie
Název práce anglicky:	Normalization in Original and Translated Czech: A Corpus Study
Vedoucí práce:	Mgr. Michaela Martinková, Ph.D.
Počet stran:	61
Počet znaků:	75 774 (včetně mezer)
Počet příloh:	0

Anotace

Cílem této bakalářské práce je výzkum normalizace v českých textech a překladech v korpusu Jerome. Teoretická část práce se zabývá korpusovou a deskriptivní translatologií a překladovými univerzáliemi. Dále je popsána současná stratifikace českého jazyka, obecná čeština a její vybrané rysy na úrovni syntaktické, lexikální, fonologické a morfologické. Metodologická část popisuje korpus Jerome a proces vytváření CQL dotazů pro vyhledávání vybraných rysů obecné češtiny v korpusu. Praktická část zkoumá hypotézu normalizace a představuje analýzu vybraných rysů obecné češtiny.

Klíčová slova: překladové univerzálie, normalizace, obecná čeština, Jerome, korpusová studie

Abstract

The aim of this bachelor thesis is to investigate normalization in Czech texts and translations in the corpus Jerome. The theoretical part of this thesis is concerned with Corpus-based and Descriptive Translation studies and translation universals. Then, the stratification of contemporary Czech, Common Czech and its selected features on syntactic, lexical, phonological and morphological level are mentioned. The methodological part describes the corpus Jerome and the process of creating CQL queries for searching selected features of Common Czech. The practical part of this thesis tests the normalization hypothesis and provides an analysis of selected Common Czech features.

Keywords: translation universals, normalization, Common Czech, Jerome, corpus study

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a uvedla úplný seznam citované a použité literatury.

V Olomouci dne 27. 6. 2024 ……

Eliška Švecová

Poděkování

Ráda bych poděkovala paní Mgr. Michaele Martinkové, Ph.D. za odborné vedení, její nesmírnou trpělivost, ochotu, užitečnou metodickou pomoc a cenné rady, které mi poskytla při zpracování bakalářské práce.

List of Abbreviations

ARA	Arabic
CZE	Czech
DAN	Danish
DUT	Dutch
ENG	English
FIN	Finnish
FRE	French
GER	German
GRA	Ancient Greek
GRN	Modern Greek
HEB	Hebrew
HUN	Hungarian
ICE	Icelandic
ITA	Italian
JAP	Japanese
LAT	Latin
MIX	Mixed
NOR	Norwegian
POL	Polish
POR	Portuguese
ROM	Romanian
RUS	Russian
SER	Serbian
SLK	Slovak
SLV	Slovenian
SPA	Spanish
SWE	Swedish

acc.	accusative
adj.	adjective(s)
dem.	demonstrative
gen.	genitive
ins.	instrumental
masc.	masculine
n.	noun(s)
num.	numeral(s)
pl.	plural
poss.	possessive
pron.	pronoun(s)
sg.	singular

List of Figures and Tables

Table of Contents

1. Introduction

Corpus linguistics plays an important part in translation studies. Through a corpus, a language can be studied from a structural, lexical, or stylistic point of view; parallel and comparable corpora also allow researcher to study languages in contrast. Corpora are also useful in providing sources from which translators can put together terminologies; many translation memories are corpus-based. However, it was only Mona Baker that introduced corpora to translation studies: in 1993 she sets the agenda for what she calls Corpus-based Translation Studies.

According to Mona Baker (1995, 225), "[t]he word *corpus* originally meant any collection of writings, in a processed or unprocessed form, usually by a specific author." However, with a significant growth in this field of study, the word *corpus* is now used as a reference to a collection of texts in digital form, that are capable of being analyzed automatically or semi-automatically rather than manually. Today, it contains not only written texts but also transcribed texts of speeches from a wide range of sources, on different topics, by many writers and speakers and allows a thorough analysis for example of lexical richness or discourse.

Two research approaches can be seen in corpus-based translation studies. One focuses on individual translation styles rather than on the features shared by all translated texts. But the main research side is based on the theory of "translation universals" (Zanettin 2013, 21). The latter is what this bachelor thesis focuses on. The main objective is to investigate original (non-translated) and translated Czech. Specifically, I will compare the features of Common Czech in translated and original (non-translated) Czech in line with research on translation universals (Chesterman 2003). The data will be collected from a monolingual comparable corpus Jerome (Chlumská 2013).

In this thesis, I will first briefly review some general information about Corpus-based Translation Studies, Descriptive Translation Studies and Translation Universals. In the next section, I briefly comment on the stratification of contemporary Czech, Common Czech and its features. In the fourth chapter, the selected corpus and queries for Common Czech features are mentioned. I also pose the following research questions:

i. Are selected Common Czech features more frequent in original (non-translated) texts than in translated texts?
ii. Is there a difference between those Common Czech features?

And to answer them in the practical part, I analyze the findings by doing quantitative analysis.

2. Corpus-based Translation Studies & Descriptive Translation Studies

In 1993 Mona Baker published her seminal paper *Corpus Linguistics and Translation Studies: Implications and Applications,* in which she predicts a turning point in Translation Studies by incorporating large corpora in research (1993, 235). According to her, "[l]arge corpora will provide theorists of translation with a unique opportunity to observe the object of their study and to explore what it is that makes it different from other objects of study" and that it will also allow theorists of translation to explore "the principles that govern translational behaviour and the constraints under which it operates" (Baker 1993, 235). In this seminal paper (1993) she has also directly linked Descriptive Translation Studies (DTS) to Corpus Linguistics and proposed further research in this discipline, specifically on 'translation universals'. However, the research into translation universals goes beyond the scope of this thesis and I will focus only on aspects directly relevant for my thesis.

Descriptive Translation Studies is a discipline whose aim is not to criticize translations or to suggest better solutions to translate certain texts, but rather to define regular language features found in translations (Toury 2012, 16).

In his book *Descriptive Translation Studies – and beyond*, Toury claims that "the cumulative findings of descriptive studies should make it possible to formulate a series of coherent *laws* which would state the inherent relations between all the variables that will have been found relevant for translation" (Toury 2012, 9). He highlights the use of empirical research in descriptive studies and says that conducting such research of translated texts should indicate certain regularities that exist in all of them. Toury has named these regularities 'laws of translational behaviour'. The first is 'the law of growing standardization' defined as: "in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by a target repertoire" (Toury 2012, 304). The second law he posits is 'the law of interference', which states that elements of the source text tend to "force themselves on the translators and be transferred to the target text" (Toury 2012, 310). These laws are said to be interconnected (Toury 2012, 303). But contrary to Toury's 'laws', Baker's term 'universals' is favored by other scholars such as Chesterman 2003 or Zanettin 2014.

2.1 Translation Universals

As was suggested earlier, the term ‘translation universals’ was first introduced by Mona Baker in 1993. She states that all translated texts share the same properties that are linked to the process of translation rather than the linguistic systems of all the languages. She then proceeds to give a list of six translation universals and supports her claim by citing other studies on which her list was based on.

The first universal is called ‘explicitation’ (1993, 243–244), which is the tendency to add information to translated texts. Second is ‘simplification’ (1993, 244), the tendency for translations to be less complex than their source texts. Then there is the preference for conventional ‘grammaticality’ (1993, 244), which is present mainly in interpreting, and ‘the tendency to avoid repetitions’ (1993, 244) in the translated text even though it appears in the source language. The fifth universal is ‘[a] general tendency to exaggerate features of the target language’ (1993, 244). As examples of this universal tendency, Toury (1980, 130) states that ‘binominals composed of synonyms or near-synonyms, which are common feature of Hebrew writing, tend to occur more frequently in translated than in original Hebrew texts and to replace non-binomials in source texts’ (quoted in Baker 1993, 244–245); and Vanderauwera (1985, 11) suggests ‘that translations overrepresent features of their host environment in order to make up for the fact that they were not originally meant to function in that environment’ (quoted in Baker 1993, 245). This particular universal is what this bachelor thesis focuses on and will be discussed in greater detail in the following section. The fifth feature notwithstanding, ‘a specific type of distribution of certain features in translated texts vis-à-vis source texts and original texts in the target language’ (Baker 1993, 245) is mentioned as the last universal feature. As an example, Shamaa (1978, 168–171) reports that common words (such as *day* or *say*) occur with a significantly higher frequency in English texts translated from Arabic than in original English texts (quoted in Baker 1993, 245).

In her next work, Baker (1996) follows up on her above mentioned seminal paper from 1993 and redefines her list of potential translation universals. She works with four of the most recognized universals. First redefined universal is simplification, which she defines as “the idea that translators subconsciously simplify the language or message or both” (quoted in Chlumská 2015, 33). The second one is explicitation, which is “the tendency to spell things out in translation, including, in its simplest form, the practice of adding background information” (quoted in Chlumská 2015, 33). Third is normalization or conservatism, i.e. “the tendency to conform to patterns and practices that are typical of the target language, even to

the point of exagerating them" (quoted in Chlumská 2015, 33). As the last redefined translation universal, Baker mentions levelling-out, which she defines as "the tendency of translated text to gravitate around the centre of any continuum rather than move towards the fringes" (quoted in Chlumská 2015, 33), which implies that translators choose language that is neutral and closer to the target language core to make sure the translated text is more comprehensible.

Later, Andrew Chesterman (2003) sees corpus-based translation studies as representing only the last stage in universalist thinking about translation. Unlike the earlier evaluative stages, which either meant a quest for an ideal translation (2003, 214), or thought of translated texts as deficient (2003, 215), the descriptive universals stage does not perceive translations as deficient but rather as a separate text type or variant with the right to be different from both source and target texts. It "simply accepts that translations will be inevitably influenced by formal features of the source text (and of course by the target language)" (Chesterman 2003, 218). Descriptive universalist hypotheses are divided into two classes based on what the texts are being compared to. S-universals depict universal differences between translations and their source texts. In contrast, T-universals capture differences between translated and non-translated texts written in the same (target) language, "they are claims about the way translators use the target language" (Chesterman 2003, 218).

It is possible to make a connection between Baker's three redefined translation universals (simplification, normalization and levelling-out) and Chesterman's T-universals as these terms are of descriptive nature and operate with common features in translation. The redefined term of explicitation can be linked to S-universals because it means adding some additional information to the target text, which makes the translation different from the source text. It is important to mention that Chesterman's descriptive approach with main focus on T-universals is applied in this thesis.

2.1.1 Normalization

The tendency to normalize translated texts was first mentioned by Gideon Toury in 1980 (Chlumská 2015, 138). Based on his definition of the law of growing standardization (Toury 2012, 304), it can be compared to Baker's normalization. The specific term 'normalization', sometimes also referred to as 'standardization' (Zanettin 2014, 19) or 'conventionalization' (Mauranen 2007, 12), was first introduced by Mona Baker. In her first seminal paper, Baker states the already mentioned 'general tendency to exaggerate features of the target language' (1993, 244) and then three years later uses the term 'normalization' in her redefined list of

translation universals (quoted in Chlumská 2015, 33). Federico Zanettin defines normalization as "the (alleged) tendency of translated texts to conform to target language rather than source language patterns and norms, producing more conventional rather that unusual target strings" (2013, 23). He also states that translators tend to use language that conforms to typical patterns of the target language at the expense of creativity (Zanettin 2014, 19).

Instead of unique features, translations contain clichés, generally unmarked grammar, common lexis and normalized punctuation. Standard language can replace dialect (Mauranen 2007, 13). So, the absence of Common Czech in translation can be seen as an example of normalization if it is replaced by Standard Czech. As a result of normalization, the translated (target) texts may seem more ordinary or "normal" and less creative as translators tend to obey the norm of the target language. As a reason why normalization occurs more often in translations than in their source texts, Toury states that unique textual relationships "are more difficult to reconstruct than institutionalized ones" (2012, 304).

2.2 Criticism of Translation Universals

The term 'translation universals' is also seen as controversial and not all academic scholars agree with it. Some scholars opt for more neutral expressions, such as properties, tendencies or laws (Cvrček and Chlumská 2015, 31; Mauranen 2007, 4). Gideon Toury himself avoided the term 'translation universals' (Toury 2004, 29). He prefers the term 'laws' and has, in fact, come up with two general laws of translation, which are already mentioned in Chapter 2.

Another issue with translation universals is the overlap between concepts. As an example of such problem, Anna Mauranen (2007, 12) points out a certain overlap between normalization and simplification because they "both regard the noticeably high lexical frequencies of certain items as supporting evidence for their hypothesis". She also mentions that translation universals do not necessarily refer to absolute laws, but rather to "general or law-like tendencies, or high probabilities of occurrence" (Mauranen 2007, 6).

Connected with the above-mentioned overlap is also Anthony Pym's (2008, 318) argument that "both explicitation and simplification make texts easier to read, and the line between the two becomes hard to discern." He also points out that Baker's four potential translation universals "seem to elaborate Toury's law of standardization, without touching his proposed law of interference" (Pym 2008, 318) and that explicitation, simplification and normalization "are different aspects of the one underlying universal", i.e. the fourth universal 'levelling-out' (Pym 2008, 318).

Worth mentioning is also the fact that Chesterman (2004, 43) criticizes that the concepts are not well defined and on that account he proposes the distinction between S-universals and T-universals. He sees the term ‘universal’ as valid and useful, “provided it is kept for claims that are actually hypothesized to be universal, not specific to some subset of translations” (Chesterman 2004, 43).

3. Common Czech

The second half of the theoretical part will discuss Common Czech (*obecná čeština*). Firstly, I will briefly comment on the stratification of the Czech language and on what 'Common Czech' means within the scope of this thesis. Then, I will describe typical features of Common Czech on four linguistic levels of language.

3.1 Stratification of Contemporary Czech

Stratification of Contemporary Czech is rather complex. It is dynamic and everchanging, as the varieties of Czech influence each other depending on ongoing social and cultural changes. This is true especially about varieties between the two poles Krčmová (2005, 1)[1] sees as clearly defined, the standard language and the dialect.

Krčmová and Chaloupek (2017) propose to distinguish between structural and non-structural varieties of Czech. Structural varieties have a complete structure on all levels and are acquired in regulated process during schooling. Non-structural varieties have specific lexis but are not structured on other levels of language.

Structural varieties include Standard Czech (*spisovná čeština*), its subcategory Colloquial Czech (*hovorová čeština*), then Common Czech (*obecná čeština*), interdialects and dialects. Standard Czech underwent complete codification, i.e., it is the norm of Czech language with strict rules (Krčmová and Chaloupek 2017; Krčmová 2005, 3). On the other hand, Common Czech, interdialects and dialects are not regulated (Krčmová and Chaloupek 2017). Cvrček et al. (2010, 23) also mention that Standard Czech (*spisovná čeština*) is a prescriptive label.

Slang and argot are listed as non-structural varieties. They are less prestigious, usually used in groups within a common environment (same profession, hobbies etc.). According to Krčmová and Chaloupek (2017), majority of Czechs reject using argot.

In *Mluvnice současné češtiny* (2010), Cvrček et al. further divide Contemporary Czech based on three types of criteria. First criterion is distinguished by medium of communication, where the difference between spoken and written form of language plays a crucial role. The difference between spoken and written Czech is quite significant. In spoken language it is

[1] All Czech references cited in this thesis do not have published translation, therefore all paraphrases and/or quotations are my translations.

frequent to use words that do not appear in written form (such as *lidma, bysme, mladej, vokno, furt*), on the contrary, it is unusual to use words typical for written language in spoken language (such as *titíž, nýbrž, jenž)*. This difference is evident to a native Czech speaker (2015, 27).

Another type of criterion is the region in which the variety is used, we can differentiate between regionally restricted varieties (dialects and interdialects) and the nationally comprehensible variety. According to Cvrček et al. (2010, 24), Common Czech is the nationally comprehensible variety: according to the author, Common Czech is used so widely, that its speakers do not realize they are not in an area, where Common Czech is not used. So, speakers of Common Czech use its specific features as a primary means of expression (*jazykové prostředky*) no matter where they are currently situated. In contrast, speakers of regionally restricted varieties are conscious of this fact and use regionally restricted means of expression only among other speakers of the same regionally restricted variety.

The last criterion proposed by Cvrček et al. concerns the form of communication (written or spoken), the situation (formality) and the text type (genre) in which the variety is used. The author distinguishes between spoken and written form of communication. Then, between formal or informal situation and suggests there are different means of expression suitable for private (intimate) or public setting. Genre can be further divided again into written (academic articles, journalism, literature, screenplays, personal letters) and spoken form (informal conversations or spontaneous monologues in debates). It is stated that language users choose concrete variety according to their knowledge, which is, among other things, based on how frequently a certain variety is used in the specific genre or situation (2010, 25).

3.2 Defining Common Czech

Common Czech is a structural non-standard variety, i.e., it has a complete structure on all levels of language but is not in line with the codification of Czech language (Krčmová and Chaloupek 2017). Moreover, authors differ in its description, classification and delimitation. Some (Krčmová 2000, 63) admit that the concept is “vague”.

In her paper, Krčmová (2000) introduces three possible understandings of Common Czech. First, there is Common Czech as private everyday spoken language in Bohemia (2000, 67). In this case, she proposes that the term ‘Czech Interdialect’ would be more fitting than the term ‘Common Czech’ (Krčmová 2000, 67). Secondly, non-private (public) Common Czech that can be heard in media is mentioned. Here the term ‘Common Czech’ is sufficient

without being specified any further (Krčmová 2000, 68–69). Finally, there is the term 'Stylized Common Czech' (*stylizovaná obecná čeština*), i.e. Common Czech used as means of stylization of spoken language in literature. The term is used to distinguish between real language situation and its literary equivalent (Krčmová 2000, 70). Krčmová (2000, 70) also points out that 'Stylized Common Czech' can be used in journalistic texts, too.

The argument that Common Czech is geographically restricted to Bohemia had been disputed by some linguists, especially by those living in Bohemia. For example, Cvrček et al. reminds that Common Czech is a nationally comprehensible variety of Czech language (2010, 24), and Petr Sgall (2012) even challenges classification of Common Czech as an interdialect. Sgall (2012, 7) also states that thanks to its phonetic and individual features, Common Czech has a stronger connection with Standard Czech than Moravian dialects have with Standard Czech. According to him, Common Czech differs from dialects and interdialects in that it is not disappearing at all; instead, it is slowly penetrating into Moravian cities as well. The result, argues Sgall (2012, 7), is Common Czech slowly turning into what he calls 'a colloquial form of the national language' (*hovorová forma národního jazyka*). However, it is still up to debate whether 'the colloquial form of the national language' is a specific type of language variety or whether it is a fluctuation between Standard and Common Czech (2012, 6).

3.3 Features of Common Czech

This section describes some features typical of Common Czech. In this study, the term refers to Krčmová's (2000, 68–69) second type of Common Czech (non-private Common Czech that can be heard in media) mentioned in the section 3.2 above. I will list and briefly discuss features on syntactic, lexical, phonological and morphological level of language. However, I will not list every feature of Common Czech, I will focus only on the most basic features on each level. The Common Czech examples are always marked as (a) and the gloss "CC" is added for better distinction, their standard equivalents follow in (b).

3.3.1 Syntactic level

Krčmová (2000, 71; 2017) writes that features of Common Czech are associated with the spoken mode, spontaneity and expressiveness. Typical syntactic features of Common Czech on the syntactic level are those that if looked at through the lenses of the written language norm, would be considered as stylistically deviant sentence structures; however in

spontaneous speech they are not just common but typical. Examples of such deviant structures are anacoluthon, zeugmas, independent sentence constituents, repetition and greater occurrence of verb phrases in one sentence. Typical feature of Common Czech on syntactic level is the omission of an auxiliary verb when referring to past in 1st person singular, which can be seen in Krčmová's (2017) example (1):

(1) (a) *Já přijel*
I arrived.PTCP.SG.M.CC
I arrived

(b) *Já jsem přijel*
I AUX.1SG.M arrived.PTCP.SG.M
I arrived

In (1a) the auxiliary (Czech equivalent of *be*) marking 1st person singular agreement is omitted, (1b) presents the Standard Czech version with the auxiliary verb present.

3.3.2 Lexical level

According to Krčmová (2017), monolingual dictionaries mark words as 'common' (*obecněčeská*) when they are not considered to be a part of Standard Czech and are not regionally or socially restricted. Often they are words that do not belong to Standard Czech because they are either too expressive, such as *tutovka* ('a sure thing') or of foreign origin, such as the word *špitál* ('hospital'), which is borrowed from German (*das Spital*) (Krčmová 2017). Formation of new words is usually manifested by means of univerbation (Krčmová 2017). Zdeňka Hladká (2017) gives the following example of an univerbized word in (2a), which formed from the collocation *driver's license* in (2b).

(2) (a) *řidičák*
driver's-license.NOM.SG.M.CC

(b) *řidičský průkaz*
driver's.NOM.SG.M license.NOM.SG.M

3.3.3 Phonological level

According to Sgall and Hronek (2014, 30), the set of phonemes is the same for Standard and Common Czech. Sgall and Hronek (2014, 30–31) propose the following list of basic phonological features that help to differentiate Common Czech from Standard Czech. The features are provided with examples also from Sgall and Hronek (2014).

3.3.3.1 Change of -é into -í/-ý sounds in the suffix of adjectives and pronouns

This feature is common in everyday spoken language in Bohemia, Western and Central Moravia. It is slowly spreading to formal spoken language, so the adjectives with the standard suffix *-é* might give the impression of being bookish (Sgall and Hronek 2014, 31). This change occurs for example in the adjective *malý* (3):

(3) (a) *malý město* ('small town')
small.NOM.SG.N town.NOM.SG.N.CC

(b) *malé město*
small.NOM.SG.N town.NOM.SG.N

In this example, *-ý* in the suffix of the plural adjectival form is a feature of Common Czech. Standard is the suffix *-é* in (3b), but as stated earlier, in everyday spoken language it might feel outdated and unnatural (Sgall and Hronek 2014, 31).

However, this change does not happen in all genders, numbers and cases of adjectives, and in fact it is more common in the singular forms of adjectives than in the plural. Table 1 and Table 2 are based on an overview of Standard Czech declensions of the hard adjective *mladý* ('young')[2] and of what Cvrček et al. (2010, 198–199) call the non-standard forms of adjectives typically occurring in spoken Czech. The forms in the light blue column and italics are non-standard, in addition, the forms in bold are those with the change of *-é* into *-í/-ý* in the suffix, typical in Common Czech. Apart from the change of *-é* into *-í/-ý*, the change of *-é* into *-ej* is displayed as well.

Pl.								
case	**masc. animate**		**masc. inanimate**		**feminine**		**neuter**	
nominative	mladí	*mladý*	mladé	***mladý***	mladé	***mladý***	mladá	*mladý*
genitive	mladých	*mladejch*	mladých	*mladejch*	mladých	*mladejch*	mladých	*mladejch*
dative	mladým	*mladejm*	mladým	*mladejm*	mladým	*mladejm*	mladým	*mladejm*
accusative	mladé	***mladý***	mladé	***mladý***	mladé	***mladý***	mladá	*mladý*
vocative	mladí	*mladý*	mladé	***mladý***	mladé	***mladý***	mladá	*mladý*
locative	mladých	*mladejch*	mladých	*mladejch*	mladých	*mladých*	mladých	*mladejch*
instrumental	mladými	*mladýma/ mladejma*	mladými	*mladýma/ mladejma*	mladými	*mladýma/ mladejma*	mladými	*mladýma/ mladejma*

Table 1: The flection of the hard adjective *mladý* in plural (Cvrček et al. 2010, 198)

In Table 1 we can see that the change of *-é* into *-í/-ý* sounds in the suffix of the plural forms is less frequent than in the singular forms. In masculine animate adjectives, the change of *-é* into

[2] Internet Language Reference Book, accessed June 20, 2024, https://prirucka.ujc.cas.cz/.

-í/-ý happens only in the accusative case, however, the change of *-í* into *-ý* happens in the nominative and vocative. In masculine inanimate and feminine adjectives, the change occurs in the nominative, accusative and vocative case. Lastly, in the suffix of neuter adjectives in plural, this change does not happen at all (because there is no *-é* suffix).

Sg.								
case	**masc. animate**		**masc. inanimate**		**feminine**		**neuter**	
nominative	mladý	*mladej*	mladý	*mladej*	mladá	*mladá*	mladé	***mladý***
genitive	mladého	***mladýho***	mladého	***mladýho***	mladé	***mladý/*** *mladej*	mladého	***mladýho***
dative	mladému	***mladýmu***	mladému	***mladýmu***	mladé	***mladý/*** *mladej*	mladému	***mladýmu***
accusative	mladého	***mladýho***	mladý	*mladej*	mladou	*mladou*	mladé	***mladý***
vocative	mladý	*mladej*	mladý	*mladej*	mladá	*mladá*	mladé	***mladý***
locative	mladém	***mladým/*** *mladym*	mladém	***mladým/*** *mladym*	mladé	***mladý/*** *mladej*	mladém	***mladým/*** *mladym*
instrumental	mladým	*mladým/ mladym*	mladým	*mladým/ mladym*	mladou	*mladou*	mladým	*mladým/ mladym*

Table 2: The flection of the hard adjective *mladý* in singular (Cvrček et al. 2010, 199)

Table 2 shows that in the singular the change of *-é* into *-í/-ý* sounds in the suffix of masculine animate adjectives happens in the genitive, dative, accusative and locative. In the suffix of singular masculine inanimate and feminine adjectives the change happens in genitive, dative and locative case. When it comes to the suffix of singular neuter adjectives, this change happens in all cases except for the instrumental.

The change of *-é* into *-í/-ý* sounds also occurs in pronouns. According to Cvrček et al. (2010), the possessive pronouns *můj* (‘my’), *tvůj* (‘your’), the reflexive possessive *svůj* (‘one’s’) (2010, 215), the demonstrative pronoun *takový* (2010, 219) and the interrogative pronouns *který* (‘which’), *jaký* (‘what’) (2010, 222) have the same declensions as the hard adjective *mladý*, which is displayed in Table 1 and Table 2 above, therefore the change of *-é* into *-í/-ý* occurs in the same grammatical cases in singular and plural.
In example (4), the demonstrative pronouns *takový* in the genitive case is mentioned.

(4) (a) *bez takovýho*
without such.GEN.DEM.SG.M.CC

(b) *bez takového*
without such.GEN.DEM.SG.M

3.3.3.2 Change of -é- into -í-/-ý- sounds in word stems

Change of *-é-* into *-í-/-ý-* sounds is also attested in word stems and is partially similar to the

change in suffixes. This occurs, for example in the Czech equivalent of the verb *to bake* (5) or the adverb meaning *less* (6) (Sgall and Hronek 2014, 31).

(5) (a) *pícť*
bake.INF.CC

(b) *péct*
bake.INF

(6) (a) *míň*
less.ADV.CC

(b) *méně*
less.ADV

According to Sgall and Hronek (2014, 31), the Common Czech verb *pícť* in example (5a) is widely used, but the Standard Czech form is *péct* in (5b). The form *míň* in (6a) is so frequent, that it is now considered to be part of Colloquial Czech, i.e. the standard variety of spoken Czech (Sgall and Hronek 2014, 31).

The change of *-é-* into *-í-/-ý-* also happens after the letter *l* in word stems (Sgall and Hronek 2014, 30). In the following example (7), the change of *-é-* into *-í-* occurs in the noun *milk*:

(7) (a) *mlíko*
milk.NOM.SG.N.CC

(b) *mléko*
milk.NOM.SG.N

Similar to the example (6a) *míň*, the example (7a) *mlíko* is also included in Colloquial Czech (Sgall and Hronek 2014, 31).

An interesting occurrence can be seen in the adverb *dlouho* (‘for a long time’). When used in comparative degree of comparison, the standard form is *déle*, however in Common Czech *-é-* changes into *-ý-*, and the form is *dýl*, which, similarly to *míň* (‘less’), is very frequent in spoken language, but it has not been codified, yet (Sgall and Hronek 2014, 31).

3.3.3.3 Change of -í/-ý into -ej in suffixes of adjectives

According to Sgall and Hronek (2014, 32), the change of *-ý* into *-ej* is very frequent in Bohemia, Western and Southern Moravia. It is advancing in Brno, the second biggest city in Czech Republic, located in the South Moravian Region.

This change occurs in singular animate adjectives in the nominative and vocative

grammatical case, which can be seen in example (8) below. In singular inanimate masculine adjectives, the change occurs in the nominative, vocative and accusative case. However, this change also occurs in plural forms of adjectives in all genders, specifically in the genitive, dative, locative and instrumental grammatical cases (Sgall and Hronek 2014, 39). This feature is also displayed in Table 1 and Table 2 in Section 3.3.3.1 above.

(8) (a) *malej kluk*
little.NOM.SG.M.CC boy.NOM.SG.M
little boy

(b) *malý kluk*
little.NOM.SG.M boy.NOM.SG.M
little boy

3.3.3.4 Change of -í/-ý into -ej in word stems

The vowel *-ý* sometimes changes into *-ej* also in the stems of nouns. This change can occur in words frequently used in spoken language in Bohemia and partly in Western Moravia. Example (9) shows the change of *-ý* into *-ej* in the stem of the Czech equivalent of the noun *soap*.

(9) (a) *mejdlo*
soap.NOM.SG.N.CC

(b) *mýdlo*
soap.NOM.SG.N

This change can also happen in other words used frequently in spontaneous speech, e.g. the verb *cítit* (‘to feel’) becomes *cejtit* or the noun *rýže* (‘rice’) becomes *rejže* (Sgall and Hronek 2014, 32).

However, there are some words that are considered to be bookish even in their Standard form, which means that they are not used frequently in spontaneous speech. As a result, these words do not appear that frequently in their Common Czech form either, for example the verb *nazývat* (‘to call’) appears as *nazejvat* only exceptionally (Sgall and Hronek 2014, 32).

3.3.3.5 Prothetic v-

Prothetic putting *v-* is one at the beginning of a word that starts with the prefix *o-* (10) or a word that has a stem starting with *o-* after a negative prefix (11). It is also frequently used in the third person personal pronouns *on* (‘he’), *ona* (‘she’), *ono* (‘it’), so the pronouns have the following form: *von* (‘he’), *vona* (‘she’), *vono* (‘it’) (Sgall and Hronek 2014, 32).

(10) (a) *vodhodit*
away-throw.INF.CC
to toss

(b) *odhodit*
away-throw.INF
to toss

(11) (a) *nevohrabanej*[3]
clumsy.NOM.SG.M.CC

(b) *neohrabaný*
clumsy.NOM.SG.M

According to Sgall and Hronek (2014, 32), this feature is typically used in Bohemia, Western and Central Moravia.

3.3.4 Morphological level

One of the morphological differences is that some forms present in standard language are not used in Common Czech, for example transgressives, passives, or past conditional verb forms (Sgall and Hronek 2014, 47–48). Some cases of analogical levelling even underwent codification, for example plural masculine and neuter nouns in the locative ending in *-ách*, e.g. *na plechách* ('on metal plates') and *v kolečkách* ('in wheels') (Krčmová 2017). The following list of morphological features contains only the basic morphological features of Common Czech:

3.3.4.1 Plural ending -ma in the instrumental case

This general feature of Common Czech is used in Bohemia and Moravia (Sgall and Hronek 2014, 36). The unified suffix *-ma* is typical for plural forms of nouns (*lidma*, 'people'), adjectives (*dobrýma*, 'good'), demonstrative pronouns (*těma*, 'those'), and numerals (*třema*, 'three') in the instrumental (Sgall and Hronek 2014). Example (12) below includes all these expressions with the unified suffix *-ma*.

(12) (a) *s těma třema dobrýma lidma*
with those.INS.DEM.PL.CC three.INS.CC good.INS.PL.CC people.INS.CC
with those three good people

(b) *s těmi třemi dobrými lidmi*
with those.INS.DEM.PL three.INS good.INS.PL people.INS
with those three good people

[3] This particular example contains two phonological features of Common Czech. First, a prothetic *v-* is inserted before *o-* in the stem of the adjective, then the change of *-ý* into *-ej* in the suffix occurs.

However, it should be noted that the numerals *dva* ('two') and *oba* ('both') in instrumental case have the dual forms of *dvěma* and *oběma*, which are codified as Standard Czech (as opposed to the hypercorrect non-standard form *dvěmi*).

3.3.4.2 Conditional bysme[4]

Czech conditional forms are periphrastic, combining the past participle form of the main verb with the conditional form of the auxiliary verb *být ('to be')*, which is *by*. The conditional auxiliary conjugates, i.e. it agrees with the subject in person and number; the participle is only marked for number and gender. In Czech, there are two types of conditional, the present conditional and the past conditional. However, in present-day Czech, the past conditional is not used as often as the present conditional, so only the example of present conditional is presented here. The following Table 3 shows the conjugation of the Czech conditional form of the verb *to carry* ('nést').[5]

singular		**plural**	
1st person	nesl *bych*	**1st person**	nesli *bychom*
2nd person	nesl *bys*	**2nd person**	nesli *byste*
3rd person	nesl *by*	**3rd person**	nesli *by*

Table 3: Inflection of the Czech auxiliary verb *to be* in conditional mood[6]

In the conditional form of the auxiliary verb *to be* in 1st person plural, Common Czech has the non-standard form (13a) *bysme*, which is frequent in Bohemia as well as Moravia. Its Standard Czech equivalent (13b) *bychom* is considered more bookish than the conditional form *bych* in the 1st person singular (Sgall and Hronek 2014, 48).

(13) (a) *nesli bysme*
carry.PTCP.PL COND.1PL.CC
we would carry

(b) *nesli bychom*
carry.PTCP.PL COND.1PL
we would carry

[4] Due to hypercorrection, another form of the conditional form *to be* in 1st person plural emerges. According to *Akademický slovník cizích slov* (Petráčková and Kraus 2001, 305), hypercorrection is an excessive effort to express oneself as linguistically correctly as possible, which often leads to using incorrect (non-standard) or non-existent words. So, by combining the conditional form of the auxiliary verb *to be* (*by*) and the verb *to be* in 1st person plural (*jsme*), the hypercorrective conditional form *by jsme* is created.

[5] "Podmiňovací způsob (byste, abyste, kdybyste), jakoby a jako by," Internet Language Reference Book, accessed June 10, 2024, https://prirucka.ujc.cas.cz/?id=575.

[6] The table is available online at https://prirucka.ujc.cas.cz/?id=575.

Sgall and Hronek (2014, 48) also mention that the conditional form *by* can be combined with conjunctions, giving the forms (14) *kdyby* ('if') and (15) *aby* ('so that'). In this case, Common Czech has the conditional form of the auxiliary verb *to be* also in 1st person plural (14a, (15a), too.

(14) (a) *kdybysme nesli*
if-would.COND.1PL.CC carry.PTCP.PL
if we would carry

(b) *kdybychom nesli*
if.COND.1PL carry.PTCP.PL
if we would carry

(15) (a) *abysme nesli*
so.that.COND.1PL.CC carry.PTCP.PL
so that we would carry

(b) *abychom nesli*
so.that.COND.1PL carry.PTCP.PL
so that we would carry

4. Methodology

This chapter will introduce the corpus used in this study and the search queries used to find and collect data.

4.1 Corpus Selection

The data comes from the corpus Jerome (Chlumská 2013), a monolingual comparable corpus of Czech. This corpus was specifically designed for analyzing contemporary translated Czech and compare it with non-translated (original) Czech, and therefore its annotation includes additional information relevant for translation studies, such as information about the edition, sex of the author, and sex of the translator. Unfortunately, the annotation does not include information about the translator's place of origin, which could be relevant when analyzing the data in terms of Common Czech (Chlumská and Richterová 2014, 20–23).

To ensure adequate heterogeneity of the corpus, one author is only represented by three texts at most, and similarly, one translator is only represented by three of their translations at most, but each of the translations has to be a translation of a work by a different author. This corpus was created in 2013 with the aim to analyze contemporary translated Czech, therefore only translations from 1992 to 2009 are included in the corpus (Chlumská and Richterová 2014, 20).

The corpus has overall more than 85 million tokens that are evenly divided into translated and original (non-translated) texts. It consists of two types of texts, fiction and non-fiction (Chlumská and Richterová 2014, 21–22). To reflect contemporary situation of Czech translations, source languages are not evenly represented, instead, the number of texts translated from each language corresponds to how much has been translated from that language into Czech in general (according to National Library of the Czech Republic and The Ministry of Education, Youth and Sports). Since most translations are from English to Czech, Jerome has a higher number of translations from English than from any other language (Chlumská and Richterová 2014, 22).

However, to research translation universals, the author also created a subcorpus consisting of 5 million tokens, one consisting of approximately the same number of tokens (circa 100 000) from all included languages. In fiction, there are texts translated from 14 languages, including Romance, Germanic, Slavic and Finno-Ugric languages. In non-fiction, there are texts translated from only six languages, namely English, German, French, Italian,

Polish and Russian (Chlumská and Richterová 2014, 22).

To better visualize the content of the corpus Jerome, I provide Table 4, in which all of the languages are displayed and alphabetically ordered. The table also provides the number of texts and the number of tokens in both fiction and non-fiction for each of the languages. It should be also noted that not all languages are represented in both text types (i.e. fiction and non-fiction).

Language	Number of Texts in Fiction	Number of Tokens (Fiction)	Number of texts in Non-fiction	Number of Tokens (Non-fiction)
ARA	---	---	1	19,628
CZE	394	26,551,540	382	15,949,930
DAN	4	321,388	---	---
DUT	3	201,495	---	---
ENG	283	18,274,340	154	8,748,715
FIN	3	182,722	---	---
FRE	45	2,211,599	23	1,338,413
GER	48	2,161,026	90	3,999,797
GRA	---	---	1	41,819
GRN	1	74,179	---	---
HEB	2	103,399	---	---
HUN	2	98,970	1	57,215
ICE	1	125,594	---	---
ITA	6	309,627	9	231,986
JAP	4	237,073	---	---
LAT	---	---	2	87,904
MIX	---	---	1	88,215
NOR	2	74,565	---	---
POL	6	564,995	11	640,664
POR	2	128,687	---	---
ROM	---	---	1	86,695
RUS	13	729,066	4	269,247
SER	1	22,867	2	29,824
SLK	2	109,237	2	163,887
SLV	1	40,211	---	---
SPA	10	333,009	1	74,106
SWE	5	313,474	1	68,204

Table 4: Languages and the number of texts and tokens in the corpus Jerome (Chlumská 2015, 55–56; Chlumská and Richterová 2014, 22)

4.2 Selected features of Common Czech and Search Queries

Of the features of Common Czech mentioned in Chapter 3, I selected three: two phonological and one morphological. Apart from a brief description, I also provide the query with inserted tags.

The first feature I will analyze is *the change of -é into -í/-ý in the suffix of adjectives and pronouns*. This change occurs in some singular and plural forms of neuter, feminine and (both animate and inanimate) masculine adjectives in various grammatical cases (see Table 1 and Table 2 in section 3.3.3.1). However, a CQL query for this specific Common Czech feature returns too many false positives. For example, [word=".*ý"&tag="A.[FIMN]P.*"] targets plural forms of neuter, feminine and (both animate and inanimate) masculine adjectives. But the part of speech tagger in the corpus tagged adjectives that are supposed to modify the nouns which follow them, but actually the tagged adjectives modify the nouns which proceed them. In example (16), the sentence includes the adjective *přezdívaný* ('nicknamed' / nicknamed.NOM.SG.M), which modifies the preceding proper noun *Digby Parkhurst* and on top of that is also in singular.

(16) Ano žáci, Digby Parkhurst, **přezdívaný** též Silnice král. (Ben Elton, *Totální kolaps*)
Yes class, Digby Parkhurst, also **nicknamed** the King of the Road.

So, the query had to be narrowed down, I resorted to two. The first [tag="A..S[24].*"&word=".*ýho"] targeted singular adjectives ending in *-ýho* in the genitive and accusative case, and it returned e.g. (17), which includes the genitive form of the adjective *Sovětský* ('Soviet' / soviet.GEN.SG.M.CC).

(17) "Panther Generální tajemník Komunistický strany **Sovětskýho** svazu? […]" (Ben Elton, *Totální kolaps*)
"The Panther General Secretary of the Communist Party of the **Soviet** Union. […]"

The second query targeted singular demonstrative and possessive pronouns also ending in *-ýho* in the genitive and accusative case: [tag="P[DS].S[24].*"&word=".*ýho"]. In example (18) below, the genitive form of the possessive pronoun *můj* ('my' / my.GEN.SG.M.CC) is displayed.

(18) "Podle **mýho** názoru jim dělá starosti, aby vůbec bylo kam, pane," prohlásil řidič, […] (Ben Elton, *Totální kolaps*)
"In **my** opinion, they're worried about having anywhere to travel at all, sir," replied the driver […]

The second Common Czech feature I will analyze is *the change of -í/-ý into -ej in suffixes of adjectives*. The search query for this specific feature is [word=".*ej"&tag="A.[IM]S.*"] and it looks for animate and inanimate masculine adjectives in singular with the suffix *-ej*. Below, I provide example (19) from the corpus with the adjective *blbej* (‘stupid’ / stupid.NOM.SG.M.CC).

(19) “Tak přece nejsem **blbej**?” říká již rozpačitě Vítězslav. (Jiří Mlčoušek, *Hajný Vítězslav a fořt Bořivoj*)
“Well, I’m not **stupid**, am I?” says Vítězslav awkwardly.

The last feature I will look at is *the plural ending -ma in the instrumental case*. It was already mentioned that this unified ending is typical for nouns, adjectives, pronouns and numerals. However, I left out the dual forms *dvěma* (‘two’) and *oběma* (‘both’), which are codified as Standard Czech. Also left out was the Common Czech form *voběma* (‘both’); its prothetic *v-* is a feature of Common Czech, but a different one than those targeted in this study[7]. So, after leaving out these numerals, the query looks like this:

[word=".*ma"&tag="[ACNP]..P7.* "&word!="[Oo]běma|[Dd]věma|[Vv]oběma"]

This query also returned several false positives. However, unlike in the case of words with a change of *-é* into *-í/-ý*, their number was not large, and the forms could be identified by simply looking at their list, without checking their contexts. Some of the tokens ending in *-ma* were foreign names of people or places, e.g. *Panama*, other were indeed in the instrumental case but belonged to Standard Czech; this again includes dual forms, such as *očima* (‘eyes’) or *očičkama* (‘little eyes’). I excluded all these expressions by using the following CQL query:

[word=".*ma"&tag="[ACNP]..P7.*"&word!="[Oo]běma|[Dd]věma|[Vv]oběma|[Rr]ukama|[Oo]čima|[Uu]šima|[Nn]ohama|[Oo]čičkama|[Pp]anama|[Mm]azama|[Aa]razima|[Aa]kima|[Mm]adama|[Ss]eriema|[Zz]amama|[Rr]etama|[Jj]uhama|[Oo]uškama|[Aa]krama|[Gg]ama|[Cc]obhama|[Yy]akama|[Ee]lama|[Pp]ýrama|[Ww]illama|[Tt]arama|[Uu]xama|[Ss]aitama|[Ss]atama[Tt]ošima|[Aa]líma|[Vv]alkama|[Dd]oylama|[Mm]ingama|[Tt]akešima|[Bb]ahama|[Rr]iema|[Ss]idama|[Pp]elinama|[Gg]irlandama|[Ff]arama|[Bb]alzama|[Ll]ingama|[Oo]čičkama|[Bb]andama|[Pp]ríma|[Mm]itama|[Nn]ikama|[Pp]alama|[Bb]asama|Uttama|zipsama|vočičkama|Kurama|Pergama|prima"]

[7] I decided not to search for expressions with prothetic *v-* because it is difficult to search for them systematically in the corpus.

As an example from the corpus, I provide sentence (20), which includes the instrumental form of the personal pronoun *nima* (‘them’ / them.INS.PL.CC):

(20) Ty baby jsou všechny stejné. Jak jim něco přeletí přes nos, tak s **nima** není k vydržení. (Jiří Mlčoušek, *Hajný Vítězslav a fořt Bořivoj*)
Women all are the same. Once they are annoyed, it is impossible to stand **them** (literally *with them*).

After running each of the queries, I divided the tokens into those found in translated and non-translated texts; the Frequency tool was used for this and *opus.status* manually selected. The frequencies were then compared in order to confirm the normalization hypothesis (for which the normalized frequency has to be lower in translated texts than in non-translated texts). Then, I used the Corpus Calculator[8] on the Czech National Corpus website to carry out the chi-square test as a statistical significance test and to create a binominal confidence interval comparison. Lastly, I used the Graph Tool[9] on the Lancaster Stats Tools website to analyze the internal variance of data and to see their error plots.

[8] Available online at: https://www.korpus.cz/calc/.
[9] Available online at: http://corpora.lancs.ac.uk/stats/toolbox.php.

5. Data Analysis

In this chapter, I will present an analysis of the selected Common Czech features in original and translated Czech, as documented in the monolingual comparable corpus Jerome. Each feature will be analyzed separately.

5.1 Change of *-é* into *-í/-ý* sounds in the suffix of adjectives and pronouns

As it was already mentioned in the methodological part (in section 4.2) above, two queries were used to download word forms with a change of *-é* into *-í/-ý*: one for adjectives and other for pronouns. My analysis will thus be divided into two parts as well: adjectives ending in *-ýho* in the genitive and accusative case will be analyzed first. Table 5 displays the absolute frequency (AF) and the normalized frequency in instances per million (i.p.m.)[10] of all singular adjectives ending in *-ýho* in the genitive and accusative case in both subcorpora (i.e. translated and non-translated parts of the corpus), and it shows that this specific feature is more frequent in non-translated texts than in translations. Specifically, the average relative frequency of non-translated texts is 86 and the average relative frequency of translated texts is 69.54. To check whether the difference in statistically significant, I carried out the chi-square test. The test proved the difference to be statistically significant at $p<0.05$ ($X^2=74.04$). Based on this frequential analysis, the normalization hypothesis is confirmed.

Subcorpora	**Tokens**	**AF of singular adj. ending in *-ýho* in the gen. and acc. case**	**i.p.m. of singular adj. ending in *-ýho* in the gen. and acc. case**
Non-translated texts	42,501,470	3,655	86
Translated texts	42,563,842	2,960	69.54

Table 5: Frequencies of singular adjectives ending in *-ýho* in the genitive and accusative case in non-translated and translated texts

[10] Since the texts in the corpus Jerome consist of maximum 160,000 words (Chlumská 2015, 58), the relative frequency in instances per million (i.p.m.) is disputable. Calculating the relative frequency in instances per one hundred thousand would be more suitable, however, KonText calculates the relative frequency in instances per million by default.

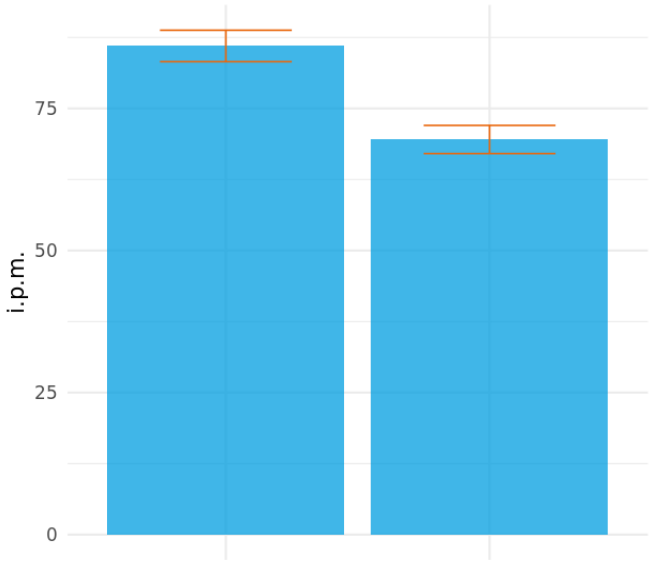

Figure 1: Singular adjectives ending in *-ýho* in the genitive and accusative case in non-translated vs. translated texts

After comparing the absolute and relative frequency of singular adjectives ending in *-ýho* in the genitive and accusative case in both translated and non-translated texts (in Table 5), I decided to check the internal variance of the frequency of singular adjectives ending in *-ýho* in the data by creating a boxplot graph displayed in Figure 2. The boxplot graph below shows a lot of texts with very high frequencies of singular adjectives ending in *-ýho* in the genitive and accusative case, i.e. outliers that skew the statistics and should not be considered.

In Originals, there are two outlier texts that have significantly higher relative frequency (i.p.m.) of singular adjectives ending in *-ýho* in the genitive and accusative case than other texts. Those two texts are *Cirkus Les Mémoires* by Petra Hůlová with the i.p.m of 3,715.14, and *Se srpem v zádech* by Jiří Pilous with the i.p.m. of 3,602.36. However, there are more outlier texts, such as *Zmizení princezny* by Jaroslav Velinský and *Opilé banány* by Petr Šabach,

In Translations, there is one outlier text with a significantly higher relative frequency of singular adjectives ending in *-ýho* in the genitive and accusative case, namely *Dóóóst dobrá schíza* written by Karen McCombie (translated from English), which is one book from the series about a 13-year-old girl.[11] Other outlier texts in Translations are *Prázdné cesty* by Rosa Liksom (translated from Finnish), *Of Mice and Men* (*O myších a lidech*) by John Steinbeck, *The Butcher Boy* (*Řeznickej kluk*) by Patrick McCabe and *On the Road* (*Na cestě*) by Jack

[11] "Dóóóst dobrá schíza," DatabázeKnih.cz, accessed June 26, 2024, https://www.databazeknih.cz/knihy/alice-a-alica-a-dooost-dobra-schiza-4474.

Kerouac. The latter three texts are all translations from English. *Of Mice and Men* is a novel about two poor field workers (one of whom is mentally disabled).[12] *The Butcher Boy* is a novel about Francis Brady, who comes from a broken home and becomes violent.[13] *On the Road* is a novel about two friends who are trying to escape the conventions and constraints of consumer society. This book has become a cult classic of the American Beat Generation.[14]

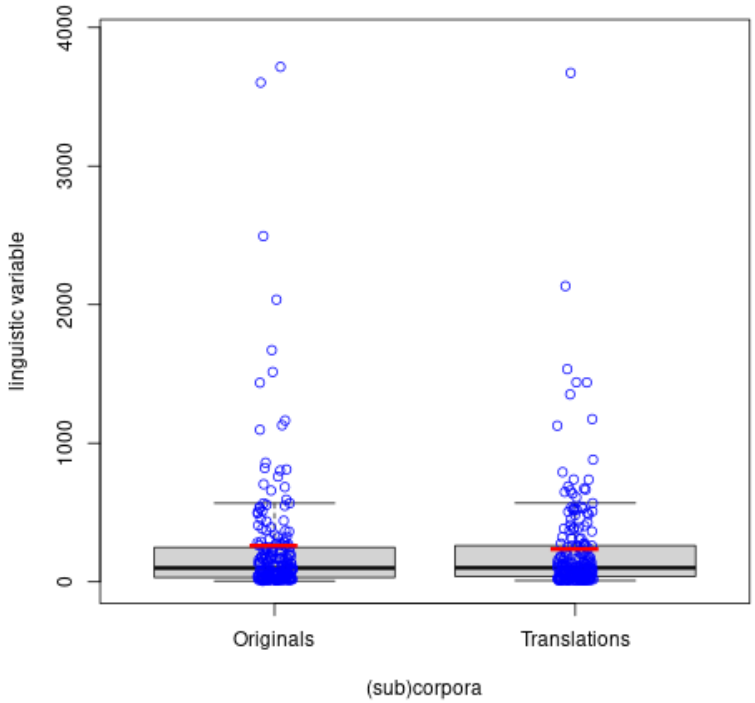

Figure 2: The distribution of singular adjectives ending in *-ýho* in the genitive and accusative case according to non-translated and translated texts in Jerome

By using the Graph Tool, I also created the error plot displayed in Figure 3 below. It shows almost complete overlap between the error bars of Originals and Translations. Therefore the difference in the frequency is not statistically significant and the normalization hypothesis for the change of *-é* into *-í/-ý* sounds in singular adjectives in the genitive and accusative grammatical case cannot be confirmed: the difference observed in the total numbers was only significant due to several outlier texts.

[12] "O myších a lidech," DatabázeKnih.cz, accessed June 26, 2024, https://www.databazeknih.cz/knihy/o-mysich-a-lidech-1962.
[13] "Řeznickej kluk," DatabázeKnih.cz, accessed June 26, 2024, https://www.databazeknih.cz/knihy/reznickej-kluk-19814.
[14] "Na cestě," Městská knihovna v Praze, accessed June 26, 2024,https://search.mlp.cz/cz/titul/na-ceste/3563540/#book-content.

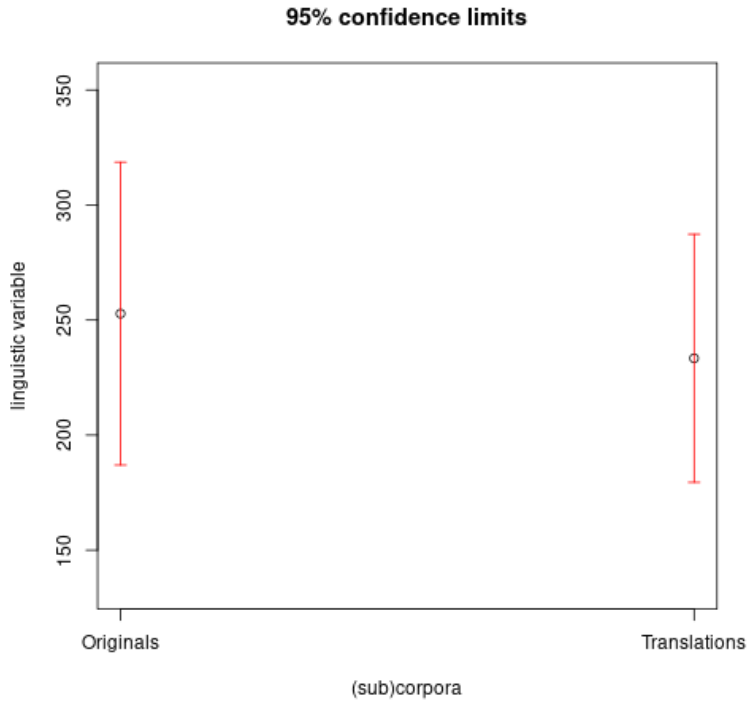

Figure 3: The dispersion of singular adjectives ending in *-ýho* in the genitive and accusative case in Originals (non-translated texts) and Translations

Table 6 provides a list of all source languages in which singular adjectives ending in *-ýho* in the genitive and accusative case appear. The table is sorted by the relative frequency (i.p.m.) of said Common Czech feature. We can see that there are quite significant differences of both types of frequency between various source languages.

Source language	**AF of singular adj. ending in *-ýho* in the gen. and acc. case**	**i.p.m. of singular adj. ending in *-ýho* in the gen. and acc. case**
FIN	67	366.68
JAP	44	185.6
ICE	18	143.32
ENG	2,574	95.25
CZE	3,655	86
GRN	6	80.89
DAN	22	68.45
RUS	63	63.11
DUT	5	24.82
ITA	11	20.31
GER	97	15.75
FRE	43	12.11
POL	8	6.64
SLK	1	3.66
SWE	1	2.62

ARA	0	0
GRA	0	0
HEB	0	0
HUN	0	0
LAT	0	0
MIX	0	0
NOR	0	0
POR	0	0
ROM	0	0
SER	0	0
SLV	0	0
SPA	0	0

Table 6: Frequencies of singular adjectives ending in *-ýho* in the genitive and accusative case in Jerome

Because of the differences between absolute and relative frequencies shown in Table 6 above, I decided to look further at the texts with significantly higher i.p.m., specifically at Finnish, Japanese and Icelandic. The source language with the highest i.p.m. is Finnish with only two texts shown in Table 7.

Name of the text	**AF of singular adj. ending in *-ýho* in the gen. and acc. case**	**i.p.m. of singular adj. ending in *-ýho* in the gen. and acc. case**
Prázdné cesty	65	2,133.67
Možnost ostrova	2	17.74

Table 7: Frequencies of singular adjectives ending in *-ýho* in the genitive and accusative case in texts translated from Finnish

The two translations from Finnish differ a lot. *Možnost ostrova* written by Michel Houellebecq is a novel set just before and after the demise of Western civilization on Earth[15], and the frequency of singular adjectives ending in *-ýho* is well below the mean. The other novel represents the opposite end of the spectrum: the short story collection *Prázdné cesty* by Rosa Liksom depicts the darkest corners of human nature, but also dreams, experiences and hidden undercurrents. The main characters are people living on the margins of society,[16] so this is naturally reflected in their speech. The language is informal, which is reflected in the Czech translation by a high frequency of singular adjectives ending in *-ýho*; in fact, the text was identified as an outlier even in the visualization given in the boxplot graph in Figure 2

[15] "Možnost ostrova," DatabázeKnih.cz., accessed June 20, 2024, https://www.databazeknih.cz/knihy/moznost-ostrova-23064.

[16] "Prázdné cesty," DatabázeKnih.cz., accessed June 10, 2024, https://www.databazeknih.cz/knihy/prazdne-cesty-45391.

above.

In second place come texts translated from Japanese. In Table 8 below, there are two texts containing singular adjectives ending in *-ýho* in the genitive and accusative case, both with a frequency higher than the mean in Czech originals:

Name of the text	AF of singular adj. ending in *-ýho* in the gen. and acc. case	i.p.m. of singular adj. ending in *-ýho* in the gen. and acc. case
Afterdark	39	737.34
Krabí zjevení	5	119.78

Table 8: Frequencies of singular adjectives ending in *-ýho* in the genitive and accusative case in texts translated from Japanese

Afterdark by Haruki Murakami is a novel that takes place in today's Tokyo and its underworld and follows the main character, a 19-year-old Mari[17]. In this case, the setting of the story in the underworld and possibly also a presence of a young protagonist is reflected in the language, reflected again in the Czech translation by the presence this specific Common Czech feature (the change of *-é* into *-í/-ý*). The translation of *Krabí zjevení* by Kotaro Tanaka, a collection of short ghost stories from ancient Japan, full of supernatural creatures and mysterious phenomena,[18] has a much lower frequency of this feature, though still above the mean frequency in Czech originals.

Higher than the mean is the frequency of singular adjectives ending with *-ýho* also in translations from Icelandic, represented by a single text, *Poslední rituál* by Yrsa Sigurdardóttir, a detective story following the murder investigation of a young German student who was obsessed with Iceland's history of torture and witch hunts.[19] The text contains 18 tokens of singular adjectives ending in *-ýho* in the genitive and accusative.

Name of the text	AF of singular adj. ending in *-ýho* in the gen. and acc. case	i.p.m. of singular adj. ending in *-ýho* in the gen. and acc. case
Poslední rituál	18	143.32

Table 9: Frequencies of singular adjectives ending in -ýho in the genitive and accusative case in texts translated from Icelandic

[17] "Afterdark," DatabázeKnih.cz., accessed June 10, 2024, https://www.databazeknih.cz/knihy/afterdark-222378.
[18] "Krabí zjevení," DatabázeKnih.cz, accessed June 20, 2024, https://www.databazeknih.cz/knihy/krabi-zjeveni-podivne-pribehy-ze-stareho-japonska-59816.
[19] "Poslední ritual," DatabázeKnih.cz, accessed June 18, 2024, https://www.databazeknih.cz/knihy/tora-gudmundsdottir-posledni-ritual-9351.

Let me now move to the analysis of frequencies of singular demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative case; these are represented in Table 10, which displays that the mean relative frequency for non-translated texts is 12.56 and for translated texts 14.87. These pronouns are more frequent in translated texts; thus the normalization hypothesis cannot be confirmed even for the change of *-é* into *-í/-ý* sounds in singular demonstrative and possessive pronouns in the genitive and accusative grammatical case. The chi-square test even proved this difference to be statistically significant at $p<0.05$ ($X^2= 8.3$)

Subcorpora	**Tokens**	**AF of sg. dem. and poss. pron. ending in *-ýho* in the gen. and acc. case**	**i.p.m. of sg. dem. and poss. pron. ending in *-ýho* in the gen. and acc. case**
Non-translated texts	42,501,470	534	12.56
Translated texts	42,563,842	633	14.87

Table 10: Frequencies of singular demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative case in non-translated and translated texts

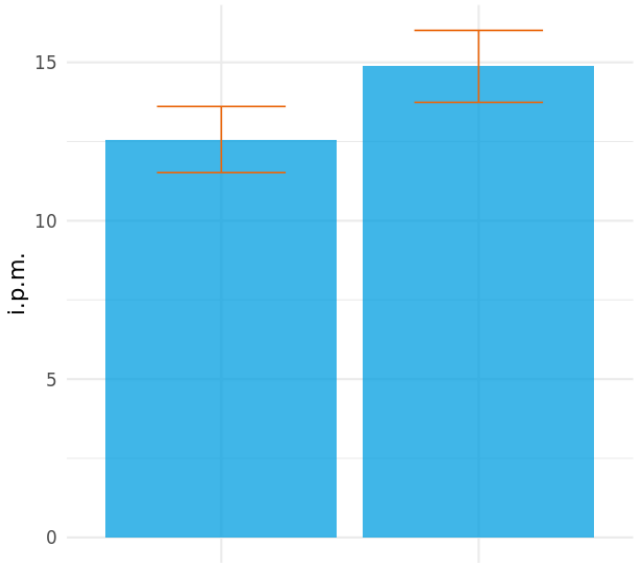

Figure 4: Singular demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative case in non-translated vs. translated texts

To check the internal variance of the relative frequencies of singular demonstrative and possessive pronouns ending in *-ýho* in the data, I created the boxplot graph in Figure 5, which shows four outlier texts in Originals and considerably more outlier texts in Translations.

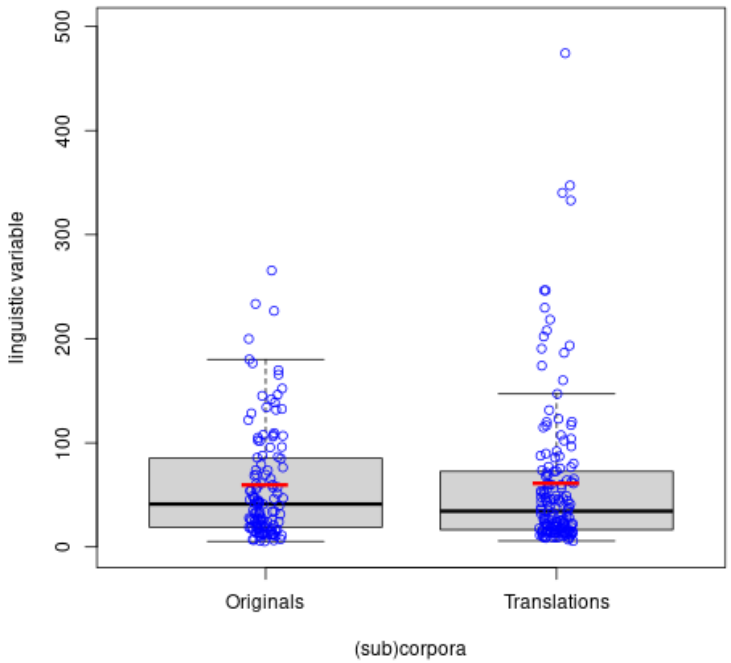

Figure 5: The distribution of singular demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative case according to non-translated and translated texts in Jerome

The four outlier texts in Originals are *Se srpem v zádech* by Jiří Pilous, *Cirkus Les Mémoires* by Petra Hůlová, *Ženy, Havel, hygiena* by Bohuslav Vaněk-Úvalský and *Sedm povídek* by Jan Zábrana. *Se srpem v zádech, Cirkus Les Mémoires* and *Sedm povídek* are already mentioned as outliers in Figure 2 above, in the first part of the analysis of the change of *-é* into *-í* in suffixes of adjectives and pronouns.

In Translations, the text with the highest i.p.m. is the novel *Of Mice and Men* (*O myších a lidech*) by John Steinbeck. Then, *Muž od vedle* by Carol Halston and *The Butcher Boy* (*Řeznickej kluk*) by Patrick McCabe. *Of Mice and Men* (*O myších a lidech*) and *The Butcher Boy* (*Řeznickej kluk*) are also already mentioned as outliers in Figure 2. *Muž od vedle* is a Harlequin romance novel.[20]

Figure 6 below shows almost complete overlap between the error bars of Originals and Translations. Thus, the difference in the frequency is not statistically significant.

[20] "Harlequinky 1993: Víc než matka / Muž od vedle / Rodinné tajemství," DatabázeKnih.cz, accessed June 26, 2024, https://www.databazeknih.cz/prehled-knihy/harlequinky-1993-vic-nez-matka-muz-od-vedle-rodinne-tajemstvi-111188.

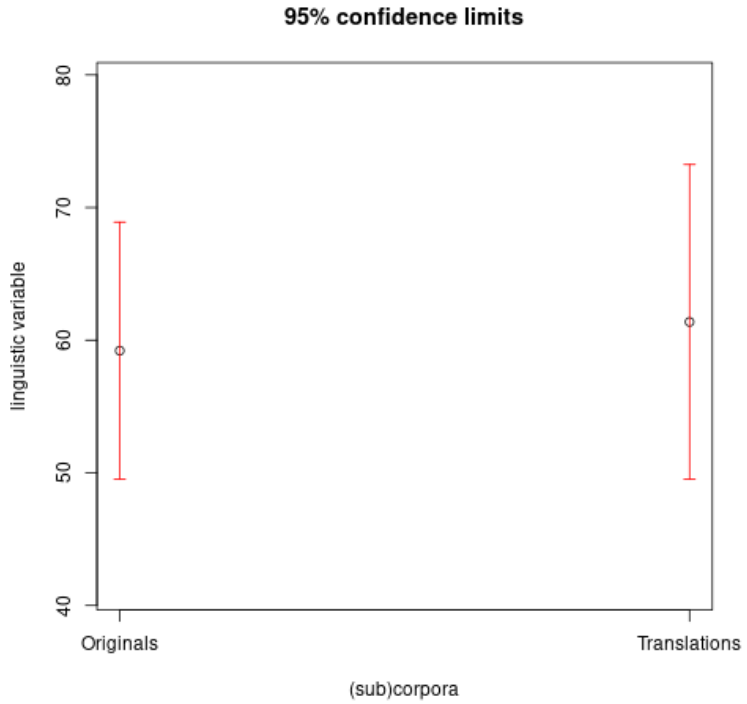

Figure 6: The dispersion of singular demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative case in Originals (non-translated texts) and Translations

Table 11 provides a list of all source languages of texts, in which demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative case occur. The table is sorted by relative frequency (i.p.m.). In the first place with the highest i.p.m. are again texts translated from Finnish and in the second place are texts translated from Japanese (similarly to Table 6, where the frequencies of singular adjectives ending in *-ýho* in the genitive and accusative case are displayed)

Source language	**AF of sg. dem. and poss. pron. ending in *-ýho* in the gen. and acc. case**	**i.p.m. of sg. dem. and poss. pron. ending in *-ýho* in the gen. and acc. case**
FIN	7	38.31
JAP	5	21.09
ENG	558	20.65
DUT	4	19.85
ICE	2	15.92
DAN	5	15.56
CZE	534	12.56
ITA	4	7.39
RUS	6	6.01
GER	25	4.06
FRE	13	3.66
POL	4	3.32

ARA	0	0
GRA	0	0
GRN	0	0
HEB	0	0
HUN	0	0
LAT	0	0
MIX	0	0
NOR	0	0
POR	0	0
ROM	0	0
SER	0	0
SLK	0	0
SLV	0	0
SPA	0	0
SWE	0	0

Table 11: Frequencies of singular demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative case in Jerome

All seven instances of singular demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative case in texts translated from Finnish are in the already mentioned (Table 7) short story collection *Prázdné cesty*, which is depicted in Table 12 below.

Name of the text	**AF of sg. dem. and poss. pron. ending in *-ýho* in the gen. and acc. case**	**i.p.m. of sg. dem. and poss. pron. ending in *-ýho* in the gen. and acc. case**
Prázdné cesty	7	229.78

Table 12: Frequencies of singular demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative case in texts translated from Finnish

Similarly to Table 8 above, which shows frequencies of singular adjectives ending in -ýho in the genitive and accusative case in texts translated from Japanese, Table 13 shows the same texts, the novel *Afterdark* and the collection of short ghost stories from ancient Japan *Krabí zjevení* with higher relative frequency than the average relative frequency in Czech originals.

Name of the text	**AF of sg. dem. and poss. pron. ending in *-ýho* in the gen. and acc. case**	**i.p.m. of sg. dem. and poss. pron. ending in *-ýho* in the gen. and acc. case**
Afterdark	3	56.72
Krabí zjevení	2	47.91

Table 13: Frequencies of singular demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative case in texts translated from Japanese

5.2 Change of *-í/-ý* into *-ej* in suffixes of masculine adjectives

Table 14 shows the absolute frequency (AF) and the normalized frequency in instances per million (i.p.m.) of all animate and inanimate masculine adjectives with the *-ej* suffix in both subcorpora. The mean relative frequency of these adjectives in non-translated texts is 306.01, in translated texts the mean relative frequency is 195.47. Based on this, we can conclude that this Common Czech feature, the *-ej* suffix of masculine adjectives, is indeed lower in translations than in original Czech texts. Also, the chi-square test proved this difference to be statistically significant at $p<0.05$ ($X^2=1036.8$). Therefore, this frequential analysis confirms the normalization hypothesis.

Subcorpus	Tokens	AF of masc. adj. with the suffix *-ej*	i.p.m. of masc. adj. with the suffix *-ej*
Non-translated texts	42,501,470	13,006	306.01
Translated texts	42,563,842	8,320	195.47

Table 14: Frequencies of the *-ej* suffix of masculine adjectives in non-translated and translated texts

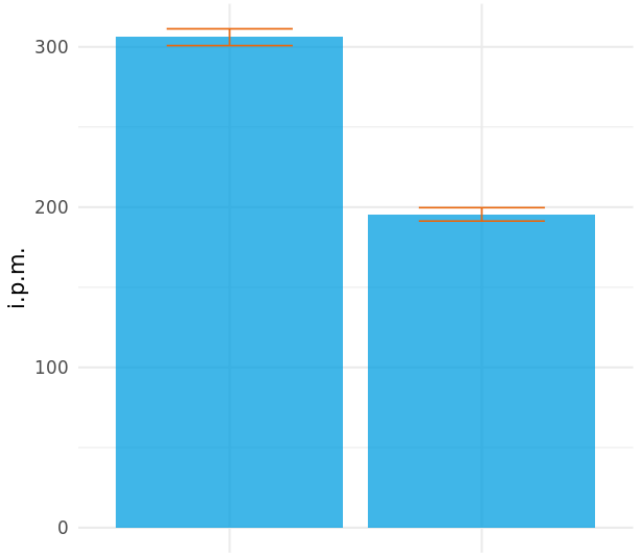

Figure 7: Czech masculine adjectives with the suffix *-ej* in non-translated vs. translated texts

To determine whether the difference in frequencies is due to many outlier texts, it was necessary to check again the internal variance of the relative frequencies of animate and inanimate masculine adjectives with the *-ej* suffix in the data by using the Graph Tool. So, I created a boxplot graph, which is given below in Figure 8.

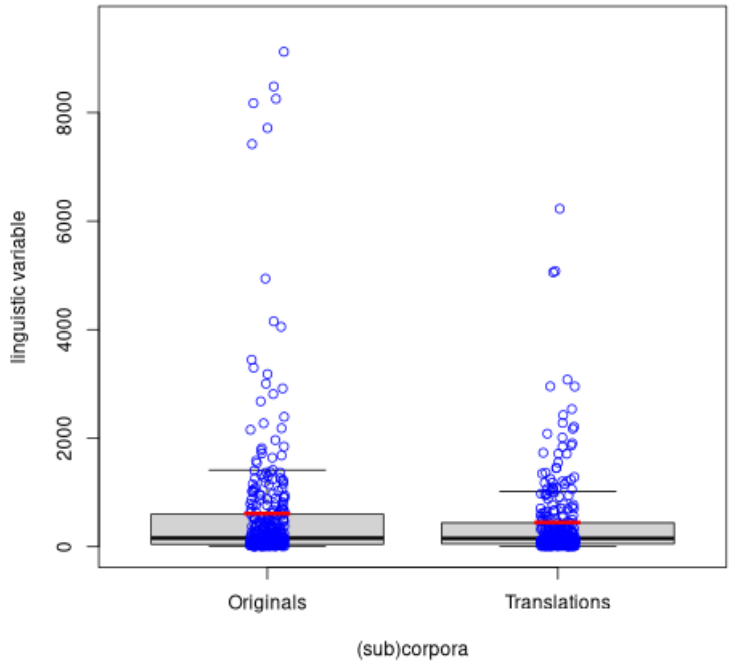

Figure 8: The distribution of the *-ej* suffix in masculine adjectives according to non-translated and translated texts in Jerome

From Figure 8, it is clear that there are a lot of outliers in both original and translated texts. In Originals, there are six outlier texts that have significantly higher relative frequency of masculine adjectives ending in *-ej* than other texts. Those texts are *Sedm povídek* by Jan Zábrana, *Noční práce* by Jáchym Topol, *Zmizení princezny* by Jaroslav Velinský, *Se srpem v zádech* by Jiří Pilous, *Cirkus Les Mémoires* by Petra Hůlová and *Opilé banány* by Petr Šabach. All of these Czech (original) texts are already mentioned as outliers in the first part of the analysis of the change of *-é* into *-í* in suffixes of adjectives and pronouns in Figure 2 (in section 5.1). Then, *Se srpem v zádech* and *Cirkus Les Mémoires* are also mentioned in the second part of the analysis of the change of *-é* into *-í* in suffixes of adjectives and pronouns in Figure 5 (also in section 5.1).

In Translations, there are three texts translated from English with significantly higher relative frequency of masculine adjectives ending in *-ej*, namely *Of Mice and Men* (*O myších a lidech*) by John Steinbeck, *The Butcher Boy* (*Řeznickej kluk*) by Patrick McCabe and *On the Road* (*Na cestě*) by Jack Kerouac. These three translated texts are also already mentioned as outliers in the first part of the analysis of the change of *-é* into *-í* in suffixes of adjectives and pronouns in Figure 2 (in section 5.1)

In Figure 9, the displayed graph shows that the confidence interval in the Originals is wider, more spread out, which makes the data less reliable. In addition, the presence of an

overlap suggests that the difference might not be significant.

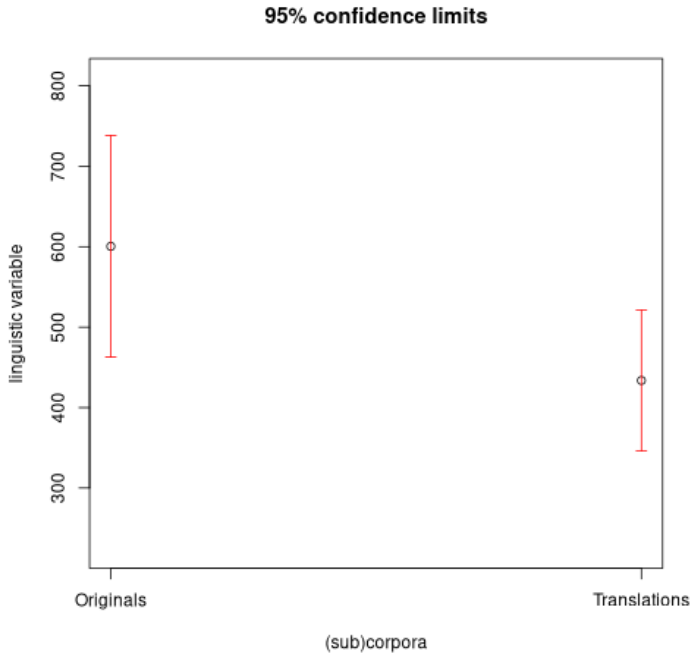

Figure 9: The dispersion of the suffix *-ej* in masculine adjectives in Originals (non-translated texts) and Translations

Table 15 provides a list relative frequencies of masculine adjectives ending in *-ej* in texts by source languages.

Source language	AF of masc. adj. with the suffix *-ej*	i.p.m. of masc. adj. with the suffix *-ej*
FIN	100	547.28
JAP	199	501.96
CZE	13,006	306.01
ENG	7,149	264.55
DUT	48	238.22
RUS	190	190.32
ICE	22	175.17
GRN	12	161.77
DAN	43	133.8
FRE	207	58.31
SER	3	56.94
GER	335	54.38
SLK	13	47.6
POL	51	42.3
SWE	12	31.44
ITA	16	29.54
ARA	0	0
GRA	0	0

HEB	0	0
HUN	0	0
LAT	0	0
MIX	0	0
NOR	0	0
POR	0	0
ROM	0	0
SLV	0	0
SPA	0	0

Table 15: Frequencies of masculine adjectives with the suffix *-ej* in Jerome

Languages with three source texts have a frequency of this feature higher than Czech original texts: texts translated from Finnish and Japanese. The first is Finnish with only one text with a higher frequency than the mean in Czech originals (306.01):

Name of the text	**AF of masc. adj. with the suffix *-ej***	**i.p.m. of masc. adj. with the suffix *-ej***
Prázdné cesty	90	2,954.31
Cizinec přichází	4	101.28
Možnost ostrova	6	53.21

Table 16: Frequencies of masculine adjectives with the suffix *-ej* in texts translated from Finnish

The short story collection *Prázdné cesty* and the novel *Možnost ostrova* have been already mentioned in Table 7, in which *Prázdné cesty* was also in the first place with the highest absolute and relative frequencies of singular adjectives and demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative. *Cizinec přichází* written by Mika Waltari is a psychological novel that takes place in the Finnish countryside[21]. I would just like to point out that these three texts are also the only texts in the corpus Jerome that have been translated from Finnish.

Second is Japanese, also with only three texts. Table 17 displays the three texts translated from Japanese containing masculine adjectives with the suffix *-ej*.

Name of the text	**AF of masc. adj. with the suffix *-ej***	**i.p.m. of masc. adj. with the suffix *-ej***
Afterdark	115	2,174.2
Krabí zjevení	3	71.87
Pivoňková lucerna	1	14.04

Table 17: Frequencies of masculine adjectives with the suffix *-ej* in texts translated from Japanese

[21] "Cizinec přichází," DatabázeKnih.cz, accessed June 20, 2024, https://www.databazeknih.cz/knihy/cizinec-prichazi-3799.

Afterdark and *Krabí zjevení* have also been mentioned above, specifically in Table 8 and Table 13, in which the novel *Afterdark* was in first place with the highest absolute and relative frequencies of singular adjectives and demonstrative and possessive pronouns ending in *-ýho* in the genitive and accusative. *Pivoňková lucerna* by Sanyutei Encho I is a traditional Japanese story of loyalty and betrayal.[22] However, only the novel *Afterdark* has higher relative frequency of masculine adjectives with the suffix *-ej* than the mean frequency of Czech originals (306.01).

5.3 Plural ending *-ma* in the instrumental case

Table 18 shows the absolute frequency (AF) and the normalized frequency in instances per million (i.p.m.) of plural instrumental forms of nouns, adjectives, pronouns and numerals ending in *-ma* in non-translated and translated texts. It is evident that non-translated texts have both higher AF and i.p.m. (150.68 in original and 77.93 in translated Czech) of these forms, and the chi-square test proved this difference to be statistically significant at $p<0.05$ ($X^2=984.95$). So again, this frequential analysis suggests that normalization is taking place.

Subcorpus	Tokens	AF of n., adj., pron. & num. ending in *-ma* in ins. case	i.p.m. of n., adj., pron. & num. ending in *-ma* in ins. case
Non-translated	42,501,470	6,404	150.68
Translated	42,563,842	3,317	77.93

Table 18: Frequencies of nouns, adjectives, pronouns and numerals with the suffix *-ma* in the instrumental case in non-translated and translated texts

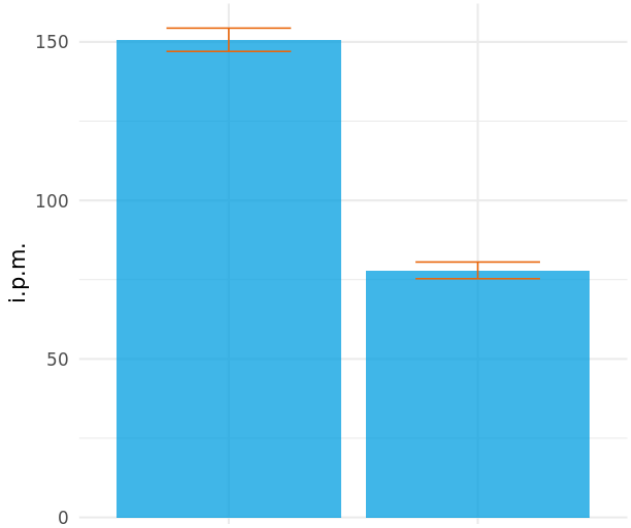

Figure 10: Czech nouns, adjectives, pronouns and numerals with the suffix *-ma* in the instrumental case in non-translated vs. translated texts

[22] "Pivoňková lucerna," Městská knihovna v Praze, accessed June 20, 2024, https://search.mlp.cz/cz/titul/pivonkova-lucerna/2230208/#book-content.

The boxplot in Figure 11 shows internal variance of the data. It shows a lot of outliers in both original and translated texts. However, in Originals, there is one outlier text that has significantly higher relative frequency of nouns, adjectives, pronouns and numerals with the suffix *-ma* in the instrumental case than other texts in both subcorpora. The text is *Cirkus Les Mémoires* by Petra Hůlová with the i.pm. of 8,555.45. Other original Czech texts with a high i.p.m. are *Noční práce* by Jáchym Topol, *Deník - Rumunsko 2007* by Michal Křen, *Se srpem v zádech* by Jiří Pilous, *Sedm povídek* by Jan Zábrana and *Zmizení princezny* by Jaroslav Velinský. Again, there are recurring outlier texts *Cirkus Les Mémoires*, *Noční práce, Se srpem v zádech, Sedm povídek* and *Zmizení princezny* (see Figure 2, Figure 5 and Figure 8).

In Translations, the three outlier texts with significantly higher relative frequency of nouns, adjectives, pronouns and numerals with the suffix *-ma* in the instrumental case are *On the Road* (*Na cestě*) by Jack Kerouac, *The Butcher Boy* (*Řeznickej kluk*) by Patrick McCabe and *Prázdné cesty* by Rosa Liksom. These three texts are also recurring outliers (see Figure 2, Figure 5 and Figure 8).

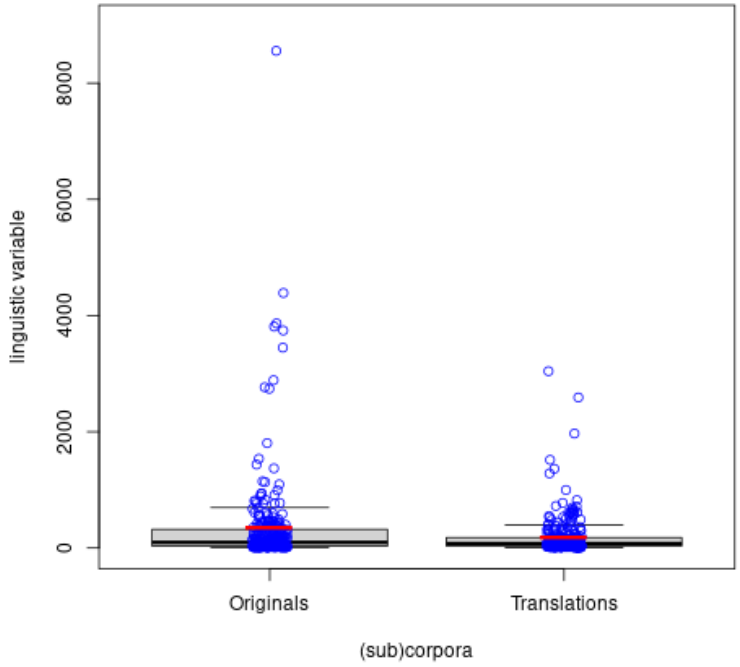

Figure 11: The distribution of nouns, adjectives, pronouns and numerals with the suffix *-ma* in the instrumental case according to non-translated and translated texts in Jerome

A visualization of relative frequencies of plural instrumental forms of nouns, adjectives, pronouns and numerals ending in *-ma* in the form of an error plot in Figure 12 shows more variation in original (non-translated) texts than in translations. There is no overlap between the confidence intervals within which the mean relative frequencies of plural instrumental

forms of nouns, adjectives, pronouns and numerals ending in *-ma* are to be found, so the difference between original and translated texts seems to be statistically significant even when internal variance is taken into consideration.

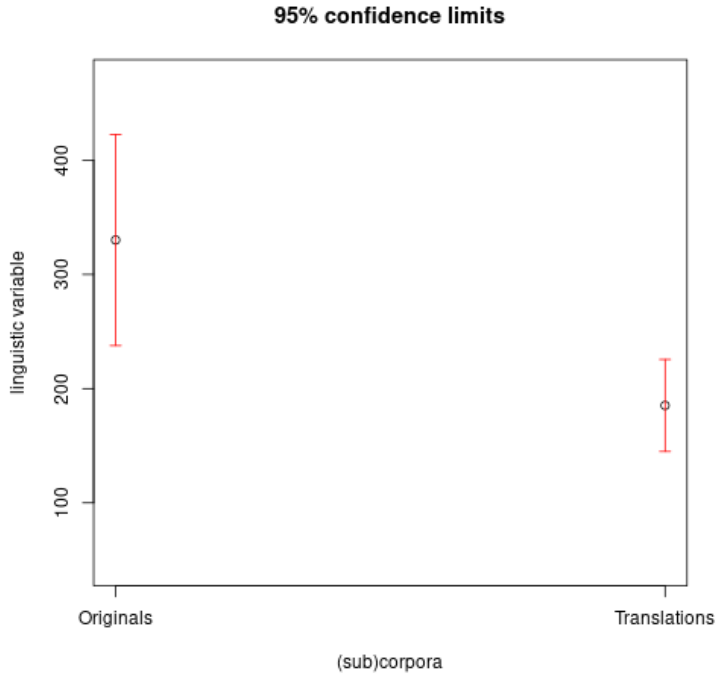

Figure 12: The dispersion of nouns, adjectives, pronouns and numerals with the suffix *-ma* in the instrumental case in Originals (non-translated texts) and Translations

Table 19 provides a list of all source languages in which nouns, adjectives, pronouns and numerals ending in *-ma* in instrumental case appear, sorted by their relative frequencies (i.p.m.). We can see that these expressions are again most frequent in texts translated from Finnish, more frequent than original Czech are they also in translations from Japanese and Icelandic.

Source language	AF of n., adj., pron. & num. ending in *-ma* in ins. case	i.p.m. of n., adj., pron. & num. ending in *-ma* in ins. case
FIN	62	339.31
JAP	44	185.6
ICE	23	183.13
CZE	6,404	150.68
GRN	10	134.81
RUS	105	105.18
ENG	2,745	101.58
DUT	11	54.59
SWE	17	44.54
DAN	10	31.12

GER	190	30.84
ITA	11	20.31
FRE	70	19.72
POL	16	13.27
SLK	2	7.32
SPA	1	2.46
ARA	0	0
GRA	0	0
HEB	0	0
HUN	0	0
LAT	0	0
MIX	0	0
NOR	0	0
POR	0	0
ROM	0	0
SER	0	0
SLV	0	0

Table 19: Frequencies of nouns, adjectives, pronouns and numerals with the suffix *-ma* in the instrumental case in Jerome

There are only two texts translated from Finnish containing nouns, adjectives, pronouns and numerals with the suffix *-ma* in the instrumental case, and only one of them does the frequency of these expressions exceed their average frequencies in Czech original texts: the short story collection *Prázdné cesty*.

Name of the text	**AF of n., adj., pron. & num. ending in *-ma* in ins. case**	**i.p.m. of n., adj., pron. & num. ending in *-ma* in ins. case**
Prázdné cesty	60	1,969.54
Cizinec přichází	2	50.64

Table 20: Frequencies of nouns, adjectives, pronouns and numerals with the suffix *-ma* in the instrumental case in texts translated from Finnish

Next are texts translated from Japanese shown in Table 21 below. The table displays two already mentioned texts, the novel *Afterdark* and the collection of short ghost stories *Krabí zjevení*. Of these, only in the former does the frequency of Common Czech instrumental forms of nouns, adjectives, pronouns and numerals ending in *-ma* exceed their average frequency in Czech original texts.

Name of the text	AF of n., adj., pron. & num. ending in *-ma* in ins. case	i.p.m. of n., adj., pron. & num. ending in *-ma* in ins. case
Afterdark	41	775.15
Krabí zjevení	3	71.87

Table 21: Frequencies of nouns, adjectives, pronouns and numerals with the suffix in the instrumental case in texts translated from Japanese

Lastly in Table 22, a text translated from Icelandic which contains 23 hits of the third selected Common Czech feature is displayed. *Poslední rituál* written by Yrsa Sigurdardóttir is a detective story, which has been already mentioned in Table 9.

Name of the text	AF of n., adj., pron. & num. ending in *-ma* in ins. case	i.p.m. of n., adj., pron. & num. ending in *-ma* in ins. case
Poslední rituál	23	183.13

Table 22: Frequencies of nouns, adjectives, pronouns and numerals with the suffix *-m*a in the instrumental case in a text translated from Icelandic

6. Conclusion

The aim of this thesis was to investigate original (non-translated) and translated Czech. Specifically, I tested the normalization hypothesis by comparing the frequencies of three selected features of Common Czech in translated and non-translated Czech in the monolingual comparable corpus Jerome. The selected features are *the change of -é into -í/-ý in the suffix of adjectives and pronouns*, *the change of -í/-ý into -ej in suffixes of adjectives* and *the plural ending -ma in the instrumental case in nouns, adjectives, pronouns and numerals.* It should be noted that the query targeting *the change of -é into -í/-ý in the suffix of adjectives and pronouns* returned too many false positives, so I created two queries for this specific Common Czech feature, one targeting singular adjectives in the genitive and accusative and the other targeting singular demonstrative and possessive pronouns in the genitive and accusative.

I compared absolute and relative frequencies of above mentioned Common Czech features in the two subcorpora of non-translated and translated texts in the corpus Jerome. The frequential analyses confirmed the normalization hypothesis in *singular adjectives ending in -ýho in the genitive and accusative*, in *animate and inanimate masculine adjectives with the -ej suffix* and in *the plural ending -ma of nouns, adjectives, pronouns and numerals in the instrumental*. The frequential analysis did not confirm the normalization hypothesis in *singular demonstrative and possessive pronouns ending in -ýho in the genitive and accusative*, these turned out to be more frequent in translated than non-translated texts. However, after checking the internal variance of the data by creating boxplot and error plots graphs, only one Common Czech feature, *the plural ending -ma of nouns, adjectives, pronouns and numerals in the instrumental case*, was proved to be statistically significant: it is significantly more frequent in originals than in translations.

By checking the internal variance of the data, I found out that there are several recurring outlier texts in both originals and translations. In originals, they are namely *Cirkus Les Mémoires* by Petra Hůlová, *Se srpem v zádech* by Jiří Pilous, *Zmizení princezny* by Jaroslav Velinský, *Sedm povídek* by Jan Zábrana, *Noční práce* by Jáchym Topol and *Opilé banány* by Petr Šabach. In translations, the outlier texts are mainly translations from English, such as *Of Mice and Men* (*O myších a lidech*) by John Steinbeck, *The Butcher Boy* (*Řeznickej kluk*) by Patrick McCabe and *On the Road* (*Na cestě*) by Jack Kerouac. Another recurring outlier text in translations is *Prázdné cesty* by Rosa Liksom. However, since there are a lot of other texts

translated from English that either do not have such high relative frequencies of the respective Common Czech features (or do not contain those features at all), English texts do not come out as overrepresented. In other source languages, for example Finnish, where there are only three texts in the whole corpus, it is more likely that outlier texts will skew the statistics. So, it is not the source language as a whole that plays a role, but rather the particular style of a particular text.

Next, I created tables, in which I divided the texts according to their source languages and sorted them by relative frequencies (i.p.m.) of the respective Common Czech features. For all selected features, the source languages with the highest i.p.m. were always Finnish in the first place and Japanese in the second place. Then, I looked further at the texts translated from these languages in which the relative frequencies of the Common Czech feature were higher than their frequencies in original Czech texts. The already mentioned short story collection *Prázdné cesty* written by Rosa Liksom (and translated from Finnish to Czech by Vladimír Piskoř) always had a higher relative frequency of the Common Czech feature in question than Czech originals, and this frequency was also much higher than the other texts translated from Finnish, which skewed the statistics if relative frequencies were calculated by the source language only. Similarly, the novel *Afterdark* written by Haruki Murakami (and translated from Japanese to Czech by Tomáš Jurkovič) has also skewed the statistics for texts translated from Japanese.

The difference between the selected Common Czech features is in their overall frequency. The most frequent are *animate and inanimate masculine adjectives with the -ej suffix*, second is *the plural ending -ma of nouns, adjectives, pronouns and numerals in the instrumental*, then come *singular adjectives ending in -ýho in the genitive and accusative*, and in the last place come *singular demonstrative and possessive pronouns ending in -ýho in the genitive and accusative*. This might suggest that *animate and inanimate masculine adjectives with the -ej suffix* are more frequently used because of their geographical distribution (according to Sgall and Hronek (2014, 32) the suffix *-ej* is very frequent in Bohemia, Western and Southern Moravia, and it is advancing in Brno, the second biggest city in the Czech Republic). This might be also due to the fact that the change of *-ý* into *-ej* happens in the nominative (see Table 2), which is one of the most frequent grammatical cases in the Czech language (Čechová 2000, 156).

Since some linguists (Krčmová 2000, 63) admit that the concept of Common Czech is "vague", future research might provide more in depth focus on the differences between the features of Common Czech. Moreover, it is worth looking further at the respective translated

texts and also investigating the role of a translator, their education and place of origin, which might influence the use of Common Czech features in translations.

7. Resumé

Cílem této bakalářské práce je výzkum normalizace v textech původně psaných česky, tedy v textech originálních (nepřekladových), a v textech přeložených do češtiny z různých zdrojových jazyků.

Teoretická část je rozdělena na dvě kapitoly. V první kapitole jsem zmínila korpusovou a deskriptivní translatologii a jejich hlavní předmět výzkumu, překladové univerzálie. Poté jsem podrobněji popsala koncept normalizace. V druhé kapitole jsem se zabývala českým jazykem a jeho současnou stratifikací. Dále jsem popsala obecnou češtinu a její vybrané rysy na úrovni syntaktické, lexikální, fonologické a morfologické. Ke každému rysu jsem uvedla i jeho příklady.

V metodologické části jsem detailně popsala korpus Jerome, který byl speciálně sestavený pro zkoumání originální a překladové češtiny. Dále jsem vybrala tři rysy obecné češtiny a uvedla proces vytváření CQL dotazů pro jejich vyhledávání v korpusu. Mezi vybrané rysy patří změna koncovky *-é* na *-í/-ý* u přídavných jmen v jednotném čísle a u ukazovacích a přivlastňovacích zájmen v jednotném čísle v 2. a 4. pádu, změna koncovky *-í/-ý* na *-ej* u přídavných jmen rodu mužského a koncovka *-ma* u podstatných jmen, přídavných jmen, zájmen a číslovek v 7. pádu. V závěru metodologické části jsem popsala celkový postup analýzy dat.

V praktické části jsem zkoumala hypotézu normalizace, tedy zda jsou vybrané rysy obecné češtiny častější v originálních českých textech než v překladech. Relativní výskyt *přídavných jmen končících na -ýho ve 2. a 4. pádu*, *přídavných jmen rodu mužského životného i neživotného končících na -ej* a *podstatných jmen, přídavných jmen, zájmen a číslovek končících na -ma v 7. pádu množného čísla* naznačil, že by tato hypotéza mohla být potvrzena. Hypotéza normalizace nebyla potvrzena u *ukazovacích a přivlastňovacích zájmen v jednotném čísle končících na -ýho v 2. a 4. pádu*, protože jsou častější v překladech než v originálních českých textech.

Při podrobnější analýze rysů, které byly častější v originálních českých textech, jsem zjistila, že rozdíl u prvního zkoumaného rysu (u *přídavných jmen končících na -ýho ve 2. a 4. pádu*) byl významný pouze kvůli několika odlehlým hodnotám (textům s vysokou relativní frekvencí), a tudíž nemohla být normalizace v tomto případě potvrzena. U druhého zkoumaného rysu (u *přídavných jmen rodu mužského životného i neživotného končících na -ej*) překrytí chybových úseček v grafu 9 na straně 46 naznačilo, že rozdíl mezi výskytem

v nepřekladových a překladových textech nemusí být statisticky významný. Efekt normalizace byl potvrzen jen u posledního zkoumaného rysu, tedy u *podstatných jmen, přídavných jmen, zájmen a číslovek končících na -ma v 7. pádu množného čísla*.

8. References

"Afterdark." DatabázeKnih.cz. Accessed June 10, 2024. https://www.databazeknih.cz/knihy/afterdark-222378.

Baker, Mona. 1993. "Corpus Linguistics and Translation Studies: Implications and Applications." In *Text and Technology*, edited by Mona Baker, Gill Frances, and Elena Tognini-Bonelli, 233–250. Amsterdam and Philadelphia: John Benjamins.

Baker, Mona. 1995. "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research." *Target* 7: 223–243.

"Cizinec přichází." DatabázeKnih.cz. Accessed June 20, 2024. https://www.databazeknih.cz/knihy/cizinec-prichazi-3799.

Cvrček, Václav, and Lucie Chlumská. 2015. "Simplification in translated Czech: a new approach to type-token ratio." *Russ Linguist* 39: 309–325.

Cvrček, Václav, et al. 2010. *Mluvnice současné češtiny 1, Jak se píše a jak se mluví*. Prague: Karolinum.

Čechová, Marie, et al. 2000. *Čeština – řeč a jazyk*. Prague: ISV.

"Doóóšt dobrá schíza." DatabázeKnih.cz. Accessed June 26, 2024. https://www.databazeknih.cz/knihy/alice-a-alica-a-dooost-dobra-schiza-4474.

"Harlequinky 1993: Víc než matka / Muž od vedle / Rodinné tajemství" DatabázeKnih.cz. Accessed June 26, 2024. https://www.databazeknih.cz/prehled-knihy/harlequinky-1993-vic-nez-matka-muz-od-vedle-rodinne-tajemstvi-111188.

Hladká, Zdeňka. 2017. "UNIVERBIZACE." CzechEncy – Nový encyklopedický slovník češtiny, edited by Petr Karlík, Marek Nekula, and Jana Pleskalová. https://www.czechency.org/slovnik/UNIVERBIZACE.

Chesterman, Andrew. 2003. "Contrastive Textlinguistics and Translation Universals." In *Contrastive Analysis in Language*, edited by Timothy Colleman, Bart Defrancq, Dirk Noël, and Dominique Willems, 213–229. New York: Palgrave Macmillan.

Chesterman, Andrew. 2004. "Beyond the particular." In *Translation Universals. Do they exist?*, edited by Anna Mauranen and Pekka Kujamäki, 33–50. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Chlumská, Lucie, and Olga Richterová. 2014. "Jak zkoumat překladovou češtinu: Výzkum simplifikace na korpusu Jerome." *Korpus - gramatika – axiologie* 9: 16–29.

Chlumská, Lucie. 2015. "Překladová čeština a její charakteristiky." PhD diss., Charles

University.

"Krabí zjevení." DatabázeKnih.cz. Accessed June 20, 2024. https://www.databazeknih.cz/knihy/krabi-zjeveni-podivne-pribehy-ze-stareho-japonska-59816.

Krčmová, Marie, and Jan Chaloupek. 2017. "NÁRODNÍ JAZYK." CzechEncy – Nový encyklopedický slovník češtiny, edited by Petr Karlík, Marek Nekula, and Jana Pleskalová. https://www.czechency.org/slovnik/NÁRODNÍ JAZYK.

Krčmová, Marie. 2000. "Termín obecná čeština a různost jeho chápání." In *Čeština - unverzália a specifika 2*, edited by Zdeňka Hladká, and Petr Karlík, 63–77. Brno: Masaryk University.

Krčmová, Marie. 2005. "*Stratifikace současné češtiny*." In *Linguistica ONLINE (LingON)*, edited by Paul Rastall, and Aleš Bičan, 1–5. Brno: Department of Linguistics and Baltic Languages, Masaryk University. Available online at https://www.phil.muni.cz/stylistika/studie/stratifikace.pdf.

Krčmová, Marie. 2017. "OBECNÁ ČEŠTINA." CzechEncy – Nový encyklopedický slovník češtiny, edited by Petr Karlík, Marek Nekula, and Jana Pleskalová. https://www.czechency.org/slovnik/OBECNÁ ČEŠTINA.

Mauranen, Anna. 2007. "Universal Tendencies in Translation." In *Incorporating Corpora: The Linguist and the Translator*, edited by Gunilla Anderman, and Margaret Rogers. Clevedon: Multilingual Matters, 32–48.

"Možnost ostrova." DatabázeKnih.cz. Accessed June 10, 2024. https://www.databazeknih.cz/knihy/moznost-ostrova-23064.

"Na cestě." Městská knihovna v Praze. Accessed June 26, 2024. https://search.mlp.cz/cz/titul/na-ceste/3563540/#book-content.

"O myších a lidech." DatabázeKnih.cz. Accessed June 26, 2024. https://www.databazeknih.cz/knihy/o-mysich-a-lidech-1962.

Petráčková, Věra, and Jiří Kraus. 2001. *Akademický slovník cizích slov*. Prague: Academia. Available online at https://kramerius.lib.cas.cz/uuid/uuid:3a4c798c-7960-4827-bdbd-38ea5b8e9ba7.

"Pivoňková lucerna." Městská knihovna v Praze. Accessed June 20, 2024. https://search.mlp.cz/cz/titul/pivonkova-lucerna/2230208/#book-content.

"Podmiňovací způsob (byste, abyste, kdybyste), jakoby a jako by." Internet Language Reference Book. Accessed June 16, 2024. https://prirucka.ujc.cas.cz/?id=575.

"Poslední rituál." DatabázeKnih.cz. Accessed June 18, 2024.

https://www.databazeknih.cz/knihy/tora-gudmundsdottir-posledni-ritual-9351.

"Prázdné cesty." DatabázeKnih.cz. Accessed June 10, 2024. https://www.databazeknih.cz/knihy/prazdne-cesty-45391.

Pym, Anthony. 2008. "On Toury's laws of how translators translate." In *Beyond Descriptive Translation Studies: Investigations in homage to Gideon Toury*, edited by Anthony Pym, Miriam Shlesinger, and Daniel Simeoni, 311–328. Amsterdam and Philadelphia: John Benjamins.

"Řeznickej kluk." DatabázeKnih.cz. Accessed June 26, 2024. https://www.databazeknih.cz/knihy/reznickej-kluk-19814.

Sgall, Petr, and Jiří Hronek. 2014. *Čeština bez příkras.* 2nd ed. Prague: Karolinum.

Sgall, Petr. 2012. "Obecná čeština." In *Linguistica ONLINE (LingON)*, edited by Paul Rastall, and Aleš Bičan, 1–10. Brno: Department of Linguistics and Baltic Languages, Masaryk University. Available online at https://www.phil.muni.cz/linguistica/art/sgall/sga-001.pdf.

Toury, Gideon. 2004. "Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals?" In *Translation universals: Do they exist*, edited by Anna Mauranen, and Pekka Kujamäki, 15–32. Amsterdam and Philadelphia: John Benjamins.

Toury, Gideon. 2012. *Descriptive Translation Studies – and beyond*. Amsterdam and Philadelphia: John Benjamins.

Zanettin, Federico. 2013. "Corpus Methods for Descriptive Translation Studies." *Procedia – Social and Behavioral Sciences* 95, 20–32.

Zanettin, Federico. 2014. Translation-driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies. New York: Routledge.

Corpus

Chlumská, Lucie. 2013. *JEROME: jednojazyčný srovnatelný korpus pro výzkum překladové češtiny*. Ústav Českého národního korpusu FF UK, Praha. Available online at http://www.korpus.cz.