

Univerzita Palackého v Olomouci
Filozofická fakulta
Katedra anglistiky a amerikanistiky



**Parallel Corpus in Sketch Engine: Creation and Data
Mining**

(Diplomová práce)

Parallel Corpus in Sketch Engine: Creation and Data Mining
(Diplomová práce)

Autor: **Bc. Magdaléna Vyvíjalová**

Studijní obor: Anglická filologie – Španělská filologie

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

Počet stran: 67

Počet znaků: 104 973

Olomouc 2018

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

V Olomouci dne

Děkuji vedoucí, Mgr. Michaele Martinkové, Ph.D., za vstřícnost, trpělivost, užitečnou metodickou pomoc a cenné rady při zpracování mé diplomové práce.

Abstract

Sketch Engine can be either a software for creating, installing and managing one's own corpora, or the web service for exploring pre-loaded or users' corpora. The present thesis employs both its uses. First, Sketch Engine as a corpus architect is used in order to create a parallel corpus that would contain original Spanish fiction and the corresponding English translation, since such a corpus does not seem available so far. In the second part of the thesis, Sketch Engine functions as a concordancer in a case study investigating lexicalization patterns of Spanish and English Motion events based on Talmy's typology dividing the languages into two broad categories, i.e. verb-framed and satellite-framed. The case study examines whether the characteristic rhetorical style of each language is retained or adapted. Simultaneously, also the Czech translation of the selected Spanish source text is considered to confirm or disprove possible intratypological differences between English and Czech.

Key words

parallel corpus, Sketch Engine, InterCorp, Talmy's typology, rhetorical style, Motion verbs, Path of Motion, Manner of Motion

Anotace

Sketch Engine může plnit dvě základní funkce, slouží jako software tvorby, instalaci a upravování korpusů, nebo jako server umožňující zadávání dotazů pro vyhledávání daných jazykových jevů ve vybraném korpusu. Tato diplomová práce využívá postupně obou těchto možných uplatnění. Nejprve je pomocí Sketch Engine vytvořen paralelní korpus obsahující originální španělskou beletrii a její překlad do angličtiny, jelikož se takovýto paralelní korpus prozatím nezdá být jinde dostupný. Ve druhé části této diplomové práce je pak Sketch Engine využit jako prostředek zkoumání sémantických jednotek týkajících se sloves pohybu ve španělštině a v angličtině, jakožto ve “verbálně rámcovaném” a “satelitně rámcovaném” jazyce, podle typologie amerického lingvisty Leonarda Talmyho. V rámci této případové studie je sledováno, zda mají jednotlivé jazyky tendenci uchovat si svůj “rétorický styl” nebo jej spíše přizpůsobit. Zohledněn je také český překlad španělského zdrojového textu, za účelem potvrzení či vyvrácení možných rozdílů v rámci stejné typologické skupiny, do které patří čeština společně s angličtinou.

Klíčová slova

paralelní korpus, Sketch Engine, InterCorp, Talmyho typologie, “rétorický styl,” slovesa pohybu, směr pohybu, způsob pohybu

Obsah	
Parallel Corpus in Sketch Engine: Creation and Data Mining	1
Introduction	7
1 Parallel corpus	10
1.1 Ambiguity of the term <i>parallel corpus</i>	10
1.2 Origins of parallel corpus	12
2 Sketch Engine	16
2.1 Core functions	16
2.2 Languages	20
2.3 Corpora and use	21
3 Creating a Parallel corpus in Sketch Engine	24
3.1 The availability of parallel corpora containing original Spanish fiction translated into English	24
3.2 Preparing the data: Translation Memory Exchange format	25
3.3 Uploading the data	28
4 Case study	30
4.1 Introduction	30
4.1.1 Lexicalization patterns of Motion events	30
4.1.2 Motion events in translation	39
4.2 Research questions	44
4.3 Methods: the corpus investigation	45
4.3.1 The query	46
4.3.2 Sorting the data	47
4.4 Data analysis	51
Conclusions	58
Works Cited	60
Appendix I – Path strategies	64
Appendix II – Strategies for neutral Motion events	68
Appendix III – Manner strategies	68

Introduction

While investigating different meanings of the polysemous English degree modifier *quite*, it was found that, according to the Park interface, there were not available any texts of original Spanish fiction directly translated into English in the InterCorp (version 7) (Vyvihalová 2015, 30). One of the main objectives of the thesis will be to create such a parallel corpus in Sketch Engine that would contain original Spanish fiction and its English translation. Therefore, the first two chapters of the thesis will deal with a theoretical background concerning parallel corpora as such, and the Sketch Engine, i.e. the corpus architect used during the process of creating the parallel corpus. Before preparing the data for the compilation of the parallel corpus, the availability of the original Spanish fiction and its English translation in the Sketch Engine and in the current version of InterCorp (version 10) will be checked, subsequently, the process of aligning, formatting and uploading the data to create the parallel corpus will be described. The next main part of the thesis will focus on exploring the newly-created parallel corpus. Namely, a case study on lexicalization patterns used when describing a Motion event will be conducted.

Talmy (1985) seems to be the first to categorize the world languages into two main typological groups, i.e. *verb-framed* and *satellite-framed* languages. The division is based on where users of a particular language encode the semantic element of Path while describing Motion events, since the Path of motion is considered to be “the basis for a major typological system” (1985, 75). The verb-framed languages (e.g. Spanish, French, Turkish or Portuguese) are characterized by employing the Path in the verb roots, whereas for the satellite-framed group (e.g. English, Russian, German, Dutch or Czech) it is typical to use so called *satellites* in order to express the Path. Talmy (1991, 486) defines the term *satellites* as “the grammatical category of any constituent other than a nominal complement that is in a sister relation to the verb root,” e.g. separate verb particles in Germanic languages, or prefixal morphemes in Slavic languages.

Several studies based on Talmy’s typology show that both typological groups follow different *rhetorical styles* (Slobin 1991, 1996, 2005) and there are also *intratypological differences* within the same typological group (Slobin 2004,

Ibarretxe-Antuñano 2009). The main differences regarding the rhetorical styles (which can be present intratypologically as well) are that the verb-framed languages provide a more static description of Motion events which is less detailed as for the Manner and directionality, while the satellite-framed languages are more dynamic and express the Manner and directionality of the Motion events in a more specific way (Slobin 1996, 205).

When we take Spanish (a verb-framed language) and English (a satellite-framed language) into consideration, Spanish speakers “cannot compactly express manner and directionality in compound expressions” (Slobin 1991, 18) as the English speakers can. This is caused by the fact that the Spanish verbs of motion encode either directionality, i.e. Path (e.g. *entrar* [enter], *salir* [exit]) or Manner (e.g. *volar* [fly], *correr* [run]). Another reason is that the Spanish prepositions, as opposed to English, do not express such a detailed locative specification but they are rather directional, e.g. Spanish preposition *de* can have several meanings such as “of, out of, in” and the meaning mostly must be inferred. This systematic difference means that “English tends to assert trajectories, leaving resultant locative states to be inferred; Spanish tends to assert locations and directions, leaving trajectories to be inferred” (1991, 18).

Slobin (1996) was the first to examine whether the Path and Manner depictions of Motion events were retained in translations from English into Spanish and vice versa. His study shows that the English translators mostly follow the Spanish original, sometimes they tend to add the Manner as well as the Path component. Spanish translators, on the other hand, are prone to reduce the full Path-Ground description of English original (1996, 210) and they omit the Manner in about half of the cases (1996, 212). Another study by Slobin (2005), focusing on typological factors which could be considered shaping the rhetorical styles of narratives in translations of Chapter 6 of *The Hobbit* (Tolkien 1937), confirms that the verb-framed languages are less concerned with the conceptualization of the Manner of the motion than the satellite-framed languages. As Ibarretxe-Antuñano and Filipović (2013, 251) point out, these typological differences are very important in Translation Studies, as the rhetorical style of the source language can have a significant impact on the target language. Their study was based on rating of how violent an event appears to be when

described in both Spanish and English and there was a considerable difference in the rating, implying that the event seems less violent when depicted in Spanish than in English (2013, 272).

The aim of the case study is to verify whether the rhetorical style of the verb-framed source language, Spanish, influences the satellite-framed target languages, English and Czech, or the target language rhetorical styles are retained. The newly-created corpus will be used in order to examine lexicalization patterns describing Motion events when translated from Spanish into English, while InterCorp (version 10) will be used to observe the tendencies in translation of Motion events from Spanish into Czech. In pursuance of the case study also possible intratypological differences between English and Czech translations will be discussed.

1 Parallel corpus

A corpus is generally defined as “a collection of written or spoken material in machine-readable form, assembled for the purpose of linguistic research.”¹ Corpora can vary in the size, in the number of languages involved, in the purpose of their compilation, or they can consist of diverse types of texts etc. Therefore, one can come across many types of corpora, such as for example *general language (reference) corpora* consisting of “representative” texts reflecting a particular language or language variety (e.g. The Brown Corpus, The British National Corpus, BNC), *historical corpora* representing a certain stage or stages of a language (e.g. A Representative Corpus of Historical English Registers, ARCHER), or *learner corpora* including text produced by learners of a language (e.g. International Corpus of Learner English, ICLE) (Nesselhauf 2011, 2–3). There are also *domain specific corpora*² which are prepared from specific domains, e.g. CAJA (academic journal articles), COMPAS (newspaper dailies related to immigration), Medical Web Corpus (medical), e-flux (art), etc.

This chapter will focus on one of such types, namely on *parallel corpora*, in the sense that it will discuss the ambiguity of the parallel corpus definition since the terminology is not consistent across various linguistic fields (see Section 1.1). Subsequently, its development will be described (Section 1.2).

1.1 Ambiguity of the term *parallel corpus*

As Sylviane Granger (2003, 19) points out, the fact that corpora appear to be a common resource for different disciplines, such as Contrastive Linguistics (CL) and Translational Studies (TS), is closely related to the ambiguity in the terminology concerning the types of corpora.

According to Johansson and Hasselgård (1999, cited in Granger 2003, 19), contrastive linguists differentiate two basic types of corpora used in cross-linguistic research, that is *comparable corpus* (usually a multilingual corpus consisting of the texts of the same genre, time of composition, or target audience) and *translation corpus* (involving original texts and their translations). The mayor

¹ <https://en.oxforddictionaries.com/definition/corpus>

² <https://www.sketchengine.eu/domain-specific-corpora/>

advantages and disadvantages (2003, 19–20) of these two groups are that even though comparable corpora represent original texts, i.e. the texts which are not influenced by other languages, it is difficult to establish the comparability of the texts as they can be culture-specific, or they can have no equivalent in other language. In case of translation corpora, the same semantic content of compared texts is to a certain extent guaranteed, however, the translated texts may not be available, or they can be affected by the source texts. Besides these two terms labelling the types of corpora, contrastive linguists also use the term *parallel corpus*, however, it is employed inconsistently either to refer to a comparable corpus or a translation corpus or both.

TS researchers, on the other hand, consider the parallel corpus a separate category of corpora, therefore, they distinguish three types of corpora: *comparable*, *translation*, and *parallel* (2003, 20). Contrary to contrastive linguists, the comparable corpus in TS tends to be monolingual, thus it incorporates original texts and translations in the same language. An example of such a corpus is *The Jerome Corpus*³ which includes both translated and non-translated Czech texts, where the non-translated part serves as a reference corpus. As for the translation corpus, it is monolingual as well, and it consists of translated texts, e.g. *The Translational English Corpus (TEC)*⁴ is comprised of written texts translated into English from a variety of source languages. The term parallel corpus in TS corresponds to the translation corpus in CL, i.e. a corpus containing a source text aligned with its corresponding translation (2003, 20). The terminological diversity is summarized in Table 1.

Contrastive Linguistics	<i>comparable</i> (usually multilingual) <i>translation</i>
Translation Studies	<i>comparable</i> (usually monolingual) <i>translation(al)</i> <i>parallel</i> (= <i>translation corpus</i> in CL)

Table 1: Types of corpora distinguished in Contrastive Linguistics and Translation Studies

³ <https://wiki.korpus.cz/doku.php/en:cnk:jerome>

⁴ <https://www.alc.manchester.ac.uk/translation-and-intercultural-studies/research/projects/translational-english-corpus-tec/>

In order to eliminate the ambiguity, Granger (2003, 21) suggests a unified typology (see Figure 1), where “the term *parallel corpus* is used unambiguously to refer to corpora of source texts and their translations” which can be unidirectional (in other words, from language X to language Y) or bi/multidirectional (i.e. from language X to Y, or from Y to X, or it involves even more languages) (2003, 20). Therefore, the term should not be confused with *comparable corpora* composed of non-translated or translated texts matched by genre, neither it should be mistaken for a generic term to refer to any type of multilingual corpus. Likewise, McEnery et al. (2006, 47) see *parallel corpus* as a corpus consisting of source texts aligned with their translations as opposed to *comparable corpus* which is, according to them, also “composed of L1 data collected from different languages using the same sampling techniques.”

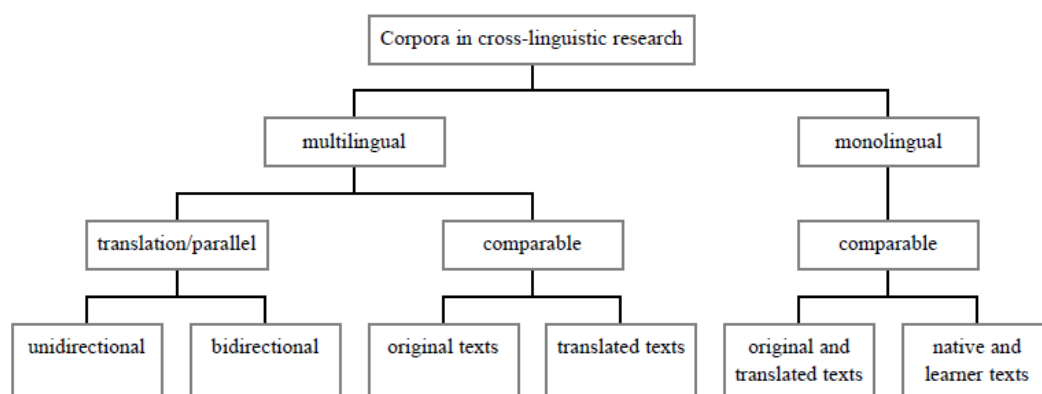


Figure 1: Corpora in cross-linguistic research (Granger 2003, 21)

Overall, the term *parallel corpus* can be explained as a kind of multilingual corpus including two monolingual corpora of the same texts where one is the translation of the other and the corresponding segments are aligned. It should not be confused with the term *comparable corpus* nor with the general term *multilingual corpora*.

1.2 Origins of parallel corpus

Early developments of corpus linguistics date from 1960s when the first two machine-readable corpora emerged; it was the Brown corpus followed by the Lancaster-Oslo/Bergen (LOB) corpus. These monolingual corpora enabled progress in quantitative and variation studies. Additionally, the same design of the

two corpora encouraged comparative studies (Hasselgård 2010, 2). In the 60s there were also the first attempts to carry out a contrastive analysis of Serbo-Croatian and English using a translation method based on a corpus of examples (Filipović 1969, 38). The plan was to take parts of the Brown corpus and match them to corresponding Serbo-Croatian translations, the next key step would be to create a corpus of equal size consisting of few Serbo-Croatian novels translated into English (1969, 43). It could be considered the first attempt to create a multilingual corpus, however, the project was not realized, and the corpora remained monolingual until 1990s when the English-Norwegian Parallel Corpus (ENPC) was developed.

The ENPC project was started by Stig Johansson and Knut Hofland in 1994 (see Johansson and Hofland 1994) and according to the website of University of Oslo⁵ it was completed in 1997. The corpus contains original texts in both English and Norwegian and the translations into the other language (i.e. English to Norwegian and Norwegian to English), the possibilities of comparisons are illustrated in Figure 2. The structure of the corpus can be used for contrastive studies either based on parallel original texts (as indicated by the solid diagonal line in Figure 2) or based on original texts and their translations in both directions: from the source text to the translation and vice versa (see the solid horizontal lines in Figure 2). While translation studies researchers use the corpus to investigate translation problems in view of one of the languages (see the solid horizontal lines in Figure 2), deviations of translated texts as compared with original texts in the same language (see the vertical lines in Figure 2), and general features of translated texts (see the broken diagonal line in Figure 2) (Johansson et al. 2002, 4).

⁵ <http://www.hf.uio.no/ilos/english/services/omc/enpc/>

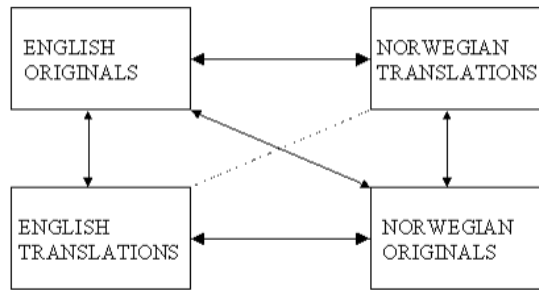


Figure 2: The structure of the English-Norwegian Parallel Corpus (Johansson et al. 2002, 4)

ENPC includes fictional as well as non-fictional texts which are limited to passages ranging from 10,000 to 15,000 words to provide material by various authors and translators (Johansson et al. 2002, 3).

While building the parallel corpus, the invention of a new software for the alignment and the system for parallel corpus concordancing were fundamental. It was Knut Hofland who developed *The Translation Corpus Aligner* which aligns the original and translated texts at sentence level, i.e. “each s-unit in the original and the translation is provided with a unique identifier (‘id’ attribute) and a ‘corresp’ attribute pointing to the corresponding s-unit(s) in the parallel text.” In order to allow the user to search and browse the corpus, *The Translation Corpus Explorer* was developed by Jarle Ebeling. The program produces concordances based on the ‘id’ and ‘corresp’ attributes of s-units (Johansson et al. 2002, 2).

As Hasselgård (2010, 2) mentions, the ENPC was developed in close cooperation with the Swedish linguists, Karin Aijmer and Bengt Altenberg (see Aijmer et al. 1996), who compiled the English-Swedish Parallel Corpus (ESPC) using the same design criteria of the corpus and partly the same English original texts. Also, the English-Finnish parallel corpus (EFPC) was created in collaboration with the ENPC and ESPC teams. Since the 90s many other examples of parallel corpora compiled according to the ENPC model has been created, e.g. the PLECI corpus of English and French (Poitiers-Louvain Échange de Corpus Informatisés), the English/German translation corpus developed at Chemnitz University, or the ACTRES parallel corpus of English and Spanish compiled at the University of León (Hasselgård 2010, 4). Another example is the

parallel corpus InterCorp which started in 2005 as a part of the project Czech National Corpus.⁶

To sum up, Corpus linguistics seems to develop rapidly over last decades since it presents a more reliable source of empirical data for contrastive linguists, whose research was previously rather intuition-based, as well as it provides translation studies researchers with verification of their theories (Granger 2003, 17). The key role in the development of multilingual corpora plays the ENCP and its sister projects, the ESCP and the EFPC, which are considered the models for creating other parallel corpora.

⁶ <http://wiki.korpus.cz/doku.php/en:cnk:intercorp>

2 Sketch Engine

In the previous chapter the term *parallel corpus* was defined, now the concordancer named Sketch Engine serving, besides other things, to create such a corpus will be described. As Kilgarriff et al. (2014, 8) state, “Sketch Engine” refers to two different things: it is either the software for creating, installing and managing one’s own corpora, or the web service for exploring pre-loaded or users’ corpora. The Sketch Engine is defined on its web⁷ as a tool which can be used to see how the language works as it helps linguists, lexicographers, students, teachers and translators to identify what is a typical or a rare phenomenon in a language or what phenomenon is coming to usage. It consists of “400 ready-to-use corpora in 90+ languages, each having a size of up to 20 billion words to provide a truly representative sample of language.”

This chapter therefore deals with the core functions of the Sketch Engine (in Section 2.1), it also focuses on languages included in the Sketch Engine (see Section 2.2), as well as it describes the basic types of corpora occurring in the Sketch Engine and what they are used for (see Section 2.3).

2.1 Core functions

The Sketch Engine software has three core functions: word sketches, concordancing and thesaurus. The software’s name is derived from the first function, *word sketch*, which is “a one-page summary of a word’s grammatical and collocational behaviour” (Kilgarriff et al. 2014, 9). An example of a word sketch can be seen in Figure 3: when searching for a word sketch of the verb *fall*, the system gives you a list of the most frequent modifiers or subjects of the verb, one can also get the idea about wh-words or adjectives which most often follow the verb. The word sketch shows also prepositional phrases, particles, pronominal subjects of the verb as well as -ing objects.

⁷ <https://www.sketchengine.co.uk>



Figure 3: Word sketch for English verb *fall* from the *Gabriel García Márquez, English corpus*

The second core function of Sketch Engine, *concordance*, shows the raw data which are the basis of any analysis (Kilgarriff et al. 2014, 10). As opposed to word sketches, concordances give you a list of instances of the searched item in context (as illustrated in Figure 4). There are several ways of getting a concordance. One can click on the number of hits appearing next to an item in a word sketch, e.g. the number of hits of “*fall*” *into*... in the *fall* word sketch (resulting concordance can be seen in Figure 4). Other way to get the concordance is by using a simple query (see Figure 5) which is case-insensitive, and it searches either for a word form or a lemma or a sequence. There are also various query types (as in Figure 6) specifying a lemma, a phrase, or a word form. For languages that do not put spaces between words, it is useful to employ a “character search.” Subsequently, a Corpus Query Language (CQL) option is a basis of all other query types that are automatically transformed into the CQL by the software (2014, 11–12).

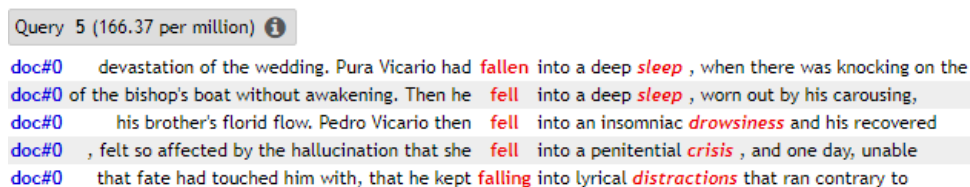


Figure 4: Concordance for *fall into* from the *Gabriel García Márquez, English corpus*

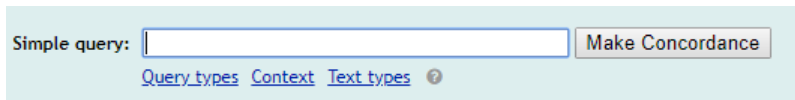


Figure 5: Simple query

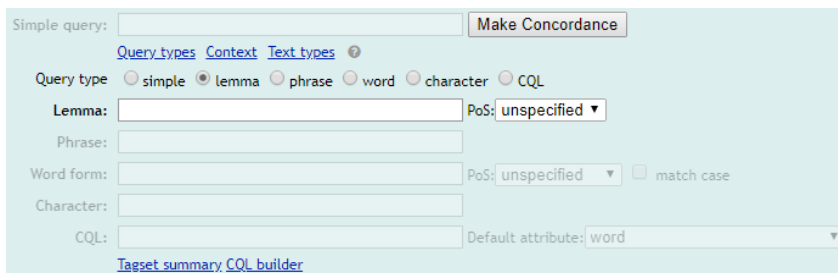


Figure 6: Query types

Concretely, a CQL query is “a pattern which may match a token or series of tokens in the corpus. Each token is assigned a set of *attributes* (word form, lemma, part-of-speech tag etc.) and each corpus might be assigned a set of *structures*. Structures may identify any sequence of tokens and are typically used to mark up documents, paragraphs, sentences, utterances, syntactic phrases of various kinds and named entities” (Jakubiček et al. 2010, 742). E.g. when searching for a noun which most frequently follows two adjacent words such as *mind* and *your*, the CQL query would be: [word="mind"] [word="your"] [tag="NN"] (a sample of resulting concordance is shown in Figure 7). Another way to examine a particular pattern of use is to specify the context by using lemma filter or PoS filter (as in Figure 8). Finally, a concordance can be specified by the selection of text types, depending on whether the texts are written or spoken; the texts can be sorted also according to the date or place of publication, author, etc. (see Figure 9).

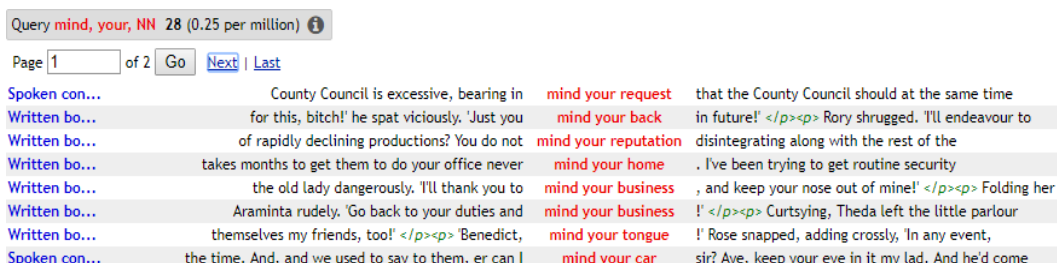


Figure 7: Sample of the CQL query “mind your + noun” searched in the BNC corpus

Context

Lemma filter PoS filter

Window: left 1 tokens. Window: left 1 tokens.

Lemma(s): all of these items. PoS: adjective adverb conjunction determiner noun all of these items.

Figure 8: Context

Text types

Subcorpus: None (whole corpus) info create new

<p>TEXT TYPE</p> <input type="checkbox"/> Spoken context-governed <input type="checkbox"/> Spoken demographic <input type="checkbox"/> Written books and periodicals <input type="checkbox"/> Written miscellaneous <input type="checkbox"/> Written-to-be-spoken <input type="button" value="Select All"/>	<p>PUBLICATION DATE</p> <input type="checkbox"/> 1960-1974 <input type="checkbox"/> 1975-1984 <input type="checkbox"/> 1985-1993 <input type="checkbox"/> Unknown <input type="button" value="Select All"/>	<p>NAME OF AUTHOR</p> <input type="text"/>
<p>PLACE OF PUBLICATION</p> <input type="checkbox"/> Ireland <input type="checkbox"/> UK (unspecific) <input type="checkbox"/> UK: Midlands (north of Bristol Channel-Wash line) <input type="checkbox"/> UK: North (north of Mersey-Humber line) <input type="checkbox"/> UK: South (south of Bristol Channel-Wash line) <input type="checkbox"/> United States <input type="checkbox"/> Unknown <input type="button" value="Select All"/>	<p>REGION WHERE SPOKEN TEXT CAPTURED</p> <input type="checkbox"/> Midlands <input type="checkbox"/> North <input type="checkbox"/> South <input type="checkbox"/> Unknown <input type="button" value="Select All"/>	<p>DOMAIN FOR CONTEXT-GOVERNED SPOKEN MATERIAL</p> <input type="checkbox"/> Business <input type="checkbox"/> Educational/Informative <input type="checkbox"/> Leisure <input type="checkbox"/> Public/Institutional <input type="button" value="Select All"/>

Figure 9: Text type selection for BNC

The third core function of Sketch Engine enables to create a “distributional thesaurus” for a corpus based on a common collocation (Kilgarriff et. al. 2014, 14). Such a thesaurus is created by taking a corpus and identifying contexts for each word, then the words which share the most contexts are identified and they appear in the same thesaurus entry. The larger the corpus is, the more precise entry appears (Rychlý et al. 2007, 41). E.g. Figure 10 shows the thesaurus entry for *tea* in the *enTenTen12* corpus: the bigger a word in the word cloud is, the more similar it is to *tea* (Kilgarriff et al. 2014, 14).

tea ^(noun)
enTenTen12 freq = **614,038** (47.3 per million)

Lemma	Score	Freq
<u>coffee</u>	0.544	783,395
<u>drink</u>	0.431	674,357
<u>juice</u>	0.416	389,262
<u>chocolate</u>	0.405	409,061
<u>wine</u>	0.39	1,048,054
<u>fruit</u>	0.378	858,835
<u>beer</u>	0.37	465,565
<u>beverage</u>	0.369	191,818
<u>milk</u>	0.368	516,439
<u>cream</u>	0.354	543,585
<u>herb</u>	0.344	306,003
<u>honey</u>	0.328	158,877
<u>bean</u>	0.327	263,509
<u>vegetable</u>	0.325	516,918
<u>soup</u>	0.322	188,722

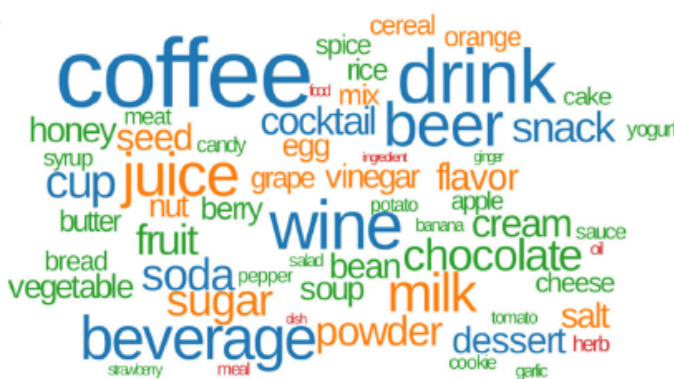


Figure 10: Thesaurus entry for *tea* (Kilgarriff et al. 2014, 14)

2.2 Languages

According to Kilgarriff et al. (2014, 17), Sketch Engine covers 19 out of 24 “large”⁸ languages of the world (see Table 2) as well as many smaller ones. Four of the large languages have basic resources in Sketch Engine, i.e. they have only a corpus, however, regarding languages that do not have spaces between the words, also a *segmentation tool* is included. While fifteen of the large languages have high-level resources in Sketch Engine, i.e. more tools are available, namely, individual words of these languages can be identified thanks to a *tokenizer*, lemmas are identified by means of a *lemmatizer*, as well as parts of speech by means of a *part-of-speech tagger*, and the structure can be determined thanks to a parser or a “sketch grammar.” The tools by it selves do not have to be totally reliable when processing the data, therefore, a person who speaks the language, a *collaborator*, is crucial in order to identify problems, so that he helps the Sketch Engine Team to improve the quality of the corpora (2014, 18).

⁸ By the word *large* it is meant that the language is used by over 50 million speakers.

Rank	Language	Primary Country	Total countries	Speakers (millions)	SkE status
1	Chinese [zho]	China	33	1,197	Good
2	Spanish [spa]	Spain	31	406	Good
3	English [eng]	United Kingdom	101	335	Good
4	Hindi [hin]	India	4	260	Good
5	Arabic [ara]	Saudi Arabia	59	223	Good
6	Portuguese [por]	Portugal	11	202	Good
7	Bengali [ben]	Bangladesh	4	193	Basic
8	Russian [rus]	Russian Federation	16	162	Good
9	Japanese [jpn]	Japan	3	122	Good
10	Javanese [jav]	Indonesia	3	84.3	No
11	German [deu]	Germany	18	83.8	Good
12	Lahnda [lah] ^a	Pakistan	7	82.7	No
13	Telugu [tel]	India	2	74	Basic
14	Marathi [mar]	India	1	71.8	No
15	Tamil [tam]	India	6	68.8	Basic
16	French [fra]	France	51	68.5	Good
17	Vietnamese [vie]	Viet Nam	3	67.8	Good
18	Korean [kor]	South Korea	6	66.4	Good
19	Urdu [urd]	Pakistan	6	63.4	No
20	Italian [ita]	Italy	10	61.1	Good
21	Malay [msa]	Malaysia	13	59.4	Basic
22	Persian [fas]	Iran	29	56.6	Good
23	Turkish [tur]	Turkey	8	50.7	Good
24	Oriya [ori]	India	3	50.1	No

Table 2: All the world languages with over 50 million speakers (Kilgarriff et al. 2014, 17)

2.3 Corpora and use

Corpora in the Sketch Engine can be divided into two primary groups: those owned and managed by the Sketch Engine Team, i.e. “preloaded corpora,” or user corpora, owned and managed by the user⁹ (Kilgarriff et al. 2014, 23). The general types of corpora and their examples were outlined in Chapter 1, now the focus is on their use in the Sketch Engine.

Initially, the Sketch Engine was used by lexicographers who employed the word sketches at first,¹⁰ then they made use especially of *general language corpora* which provide a sufficient amount of recent language data, usually based on web corpora (2014, 23).

The Sketch Engine is also widely used in the academic research, mainly for teaching and research in linguistic and language departments, as well as for discourse analysis, or teaching translation, whereas in computing departments the

⁹ How a user can upload and install a corpus via the Sketch Engine interface will be described in detail in Chapter 3.

¹⁰ Macmillan was the first to use word sketches while preparing the first edition of the *Macmillan English Dictionary* (Kilgarriff et al. 2014, 14).

teaching and research is related to *Natural Language Processing(NPL)*¹¹ (2014, 15). Thus, for the purposes of academic research *general language corpora* and *parallel corpora* seem to be mostly employed, additionally, those linguists who study development and change of a language use *historical corpora*.

Another field in which the Sketch Engine plays a significant part is Language Teaching (LT). It is primarily used for English Language Teaching (ELT), then it is also used while teaching Chinese, Japanese and Arabic (2014, 16). In 1994 the “Teaching and Language Corpora” (TALC) community developed based on Tim Johns’ work who was the first to promote using corpora in LT (Kilgarriff et. al 2015, 63). As Kilgarriff et al. (2015, 64) point out, there are two kinds of use of corpora in LT: *direct use* includes students looking at the concordances in the classroom, and the *indirect use* involves preparing coursebooks, dictionaries, syllabi and other teaching materials. *Learner corpora* enables to study the process of language learning, i.e. it helps to identify what difficulties a learner at a particular level can get into, which is useful, for example, when creating curricula or tests etc., i.e. learner corpora can be considered a part of indirect use. Then, *general language corpora* provide practical examples of the learned language in use, i.e. general language corpora can be a part of both direct and indirect use (Kilgarriff et al. 2014, 24). Even though, Sketch Engine seems to be an ideal tool for LT, as Kilgarriff et al. (2015, 65) state, according to feedback, the complexity of the software puts the students off. It led to the creation of SKELL,¹² which is “a language learning website in which all the reports are corpus-based, using fully automated methods, and are designed to avoid scaring the students. It is currently available only for English. It offers three reports: word sketch, examples and ‘similar words.’”

The Sketch Engine is also very useful for identifying the terminology and phraseology in order to achieve a consistent use of the same term within each language (Kilgarriff et al. 2014, 16). Namely, *general language corpora* and *domain specific corpora* are crucial in maintaining the terminological consistency.

¹¹ It is “an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things. NLP researchers aim to gather knowledge on how human beings understand and use language so that appropriate tools and techniques can be developed to make computer systems understand and manipulate natural languages to perform desired tasks” (Chowdhury 2005, 51).

¹² <https://skell.sketchengine.co.uk>

The corpora in the Sketch Engine, regardless of its type, can be also used as a central resource for language technology applications, i.e. the software can produce word lists¹³ which can serve for speech recognition, text prediction, or spelling correction (2014, 16).

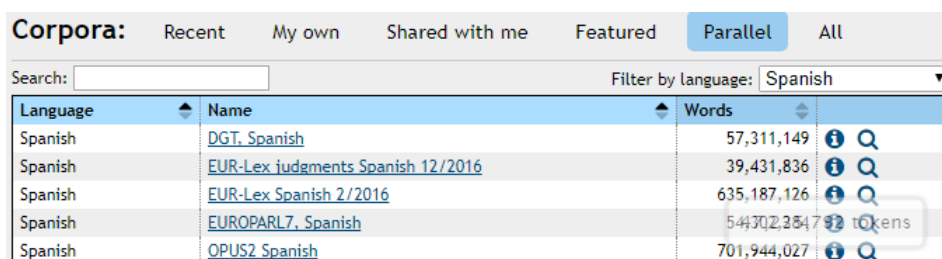
¹³ Except for word lists, it can be also lists of n-grams, keywords, lemmas, terms etc.

3 Creating a Parallel corpus in Sketch Engine

Before creating a new parallel corpus in Sketch Engine, it will be verified whether there is a parallel corpus which would contain original Spanish fiction and its English translation in Sketch engine or in InterCorp (version 10). Then the process of creating a parallel corpus will be described in Section 3.2, i.e. choosing the source text, aligning the text to the target text and formatting into a suitable TMX format. Finally, uploading of the data and finishing the compilation of the new parallel corpora will be illustrated in Section 3.3.

3.1 The availability of parallel corpora containing original Spanish fiction translated into English

To check the availability of a parallel corpus of original Spanish fiction and its English translation, I looked up all parallel corpora containing Spanish language in Sketch Engine and the results (see Figure 11) showed corpora composed of legislative documents (*DTG, Spanish*), European Union's judicial documents (*EUR-Lex judgments Spanish 12/2016*), also a corpus covering a vast area of subjects concerning European Union law (*EUR-Lex Spanish 2/2016*), as well as a corpus created from the European Parliament Proceedings (*EUROPARL7, Spanish*). Thus, these four corpora do not contain any examples of original Spanish fiction. Another parallel corpus available through Sketch Engine, *OPUS2 Spanish*, contains automatically pre-processed free online data, as it is said on the OPUS website.¹⁴ Although *OPUS2* includes also subcorpora of translated copyright-free books, however, there were not found any examples of original Spanish fiction translated into English.



Language	Name	Words	
Spanish	DGT, Spanish	57,311,149	🔍
Spanish	EUR-Lex judgments Spanish 12/2016	39,431,836	🔍
Spanish	EUR-Lex Spanish 2/2016	635,187,126	🔍
Spanish	EUROPARL7, Spanish	544,302,354,732 tokens	🔍
Spanish	OPUS2 Spanish	701,944,027	🔍

Figure 11: Parallel corpora in Sketch Engine containing Spanish language

¹⁴ <http://opus.lingfil.uu.se/>

In my previous research concerning English and Spanish fiction (Vyvjalová 2015, 30) it was found that, according to the Park interface, there were not available any texts of original Spanish fiction directly translated into English in InterCorp (version 7). Thus, before compiling the new parallel corpus, it was also checked whether any new texts were added to InterCorp version 10. Since the interface Park no longer exists, it was necessary to specify the query in InterCorp, as exemplified in Figure 12. First, *InterCorp v10 – Spanish* was selected and *InterCorp v10 - English* was aligned. Then the *core* text group was chosen as well as the *fiction* text type, and *Spanish* as a source language was selected. Finally, the option of the texts in the original language was ticked. After refining the selection of the query no results were found.

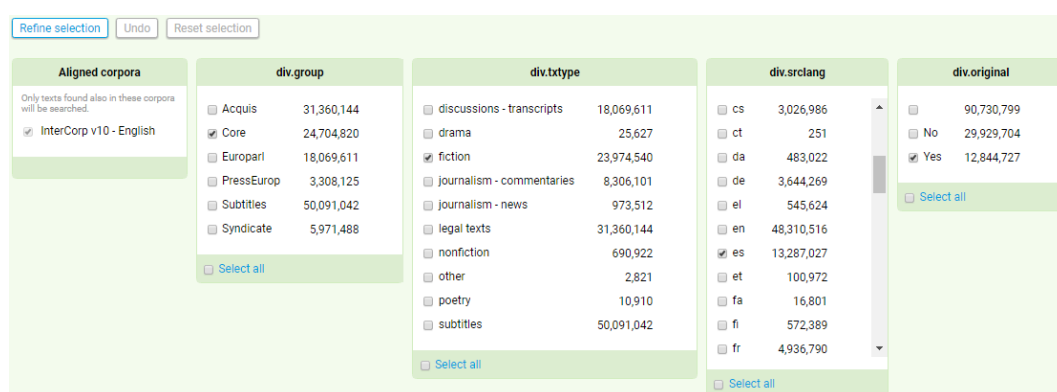


Figure 12: Refining the selection of the query in the InterCorp v10

To conclude, neither Sketch Engine nor InterCorp version 10 appear to contain a parallel corpus of original Spanish fiction and its English translation, thus, it was convenient to create such a corpus.

3.2 Preparing the data: Translation Memory Exchange format

In order to explore whether the rhetorical style of a verb-framed language, Spanish, influences the translation into satellite-framed languages English and Czech, I created a parallel corpus containing a novella by Gabriel García Márquez called *Crónica de una muerte anunciada* (1981) and its translation *Chronicle of a Death Foretold* by Gregory Rabassa. Both the original and the English translation are available online, one in PDF format and the other in DOC format,

additionally, the original and the Czech version are available in *InterCorp v10*. The choice of the novella does not follow any specific criteria.

Before uploading data to Sketch Engine, it was necessary to adjust the data. Firstly, parts of the text such as information about the author' s life or prologue were excluded since they were mentioned only in the Spanish version of the book. Then the Spanish text was converted to a DOC file in order to have two same DOC files which could be subsequently transferred to a single file in a TMX format.¹⁵

To create the TMX file which would contain the Spanish text aligned with its English version, *+Tools* and *Wordfast classic* programs must be added to the Word program, i.e. they need to be saved as Templates in the Word program. *+Tools* enables the alignment of the two texts and *Wordfast classic* is a Translation Memory software which helps to convert the file into a TMX format. Initially, I opened the two prepared DOC files, then I used *+Tools* to start the alignment, in the process, the Spanish original was selected as the source document and the English version as the target text (see Figure 13).

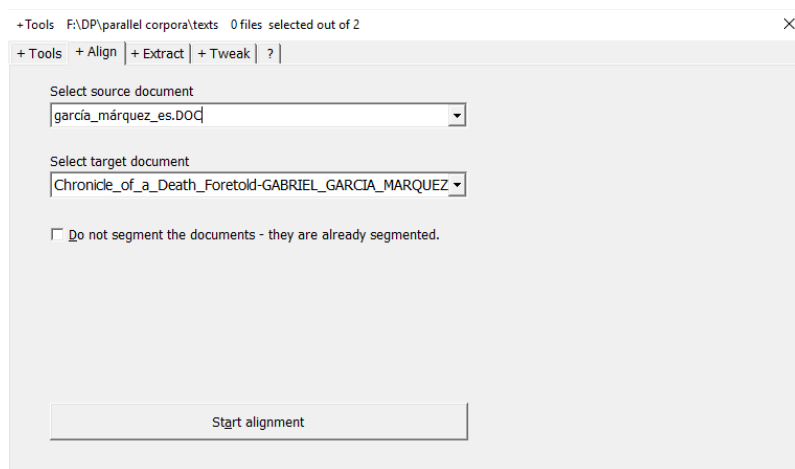


Figure 13: Starting the alignment

Thanks to *+Tools* the two texts were automatically aligned, however, the manual correction was needed. Mostly, the cells of the Spanish text were merged (as illustrated in Figure 14 and Figure 15) or split (as in Figure 16 and Figure 17) to correspond to the translated version of the text. During the manual correction, also

¹⁵ The main supported formats are TMX or XLS but formats such as XLIFF, XML, TSV, TAB, or xlsx can be used as well.

misprinted characters were edited and numbers of pages and numbers of chapters were eliminated since they are not relevant to the further research.

El día en que lo iban a matar, Santiago Nasar se levantó a las 5.30 de la mañana para	ON THE DAY THEY WERE GOING TO KILL him, Santiago Nasar got up at five-thirty in the morning to wait for the boat the bishop was coming on.
esperar el buque en que llegaba el obispo.	He'd dreamed he was going through a grove of timber trees where a gentle drizzle was falling, and for an instant he was happy in his dream, but when he awoke he felt completely spattered with bird shit.

Figure 14: Process of manual correction before merging the cells

El día en que lo iban a matar, Santiago Nasar se levantó a las 5.30 de la mañana para esperar el buque en que llegaba el obispo.	ON THE DAY THEY WERE GOING TO KILL him, Santiago Nasar got up at five-thirty in the morning to wait for the boat the bishop was coming on.
Había soñado que atravesaba un bosque de higueros donde caía una llovizna tierna, y por un instante fue feliz en el sueño, pero al despertar se sintió por completo salpicado de cagada de pájaros.	He'd dreamed he was going through a grove of timber trees where a gentle drizzle was falling, and for an instant he was happy in his dream, but when he awoke he felt completely spattered with bird shit.

Figure 15: Process of manual correction after merging the cells

No se alarmó, porque había otros caminos para nuestra casa. Próspera Arango, la cachaca, le suplicó que hiciera algo por su padre que estaba agonizando en el sardinel de su casa, inmune a la bendición fugaz del obispo.	He wasn't alarmed, because there were other ways to get to our house.
«Yo lo había visto al pasar —me dijo mi hermana Margot—, y ya tenía cara de muerto».	Próspera Arango, the uplander, begged him to do something for her father, who was in his death throes on the stoop of his house, immune to the bishop's fleeting blessing.

Figure 16: Process of manual correction before splitting the cells

No se alarmó, porque había otros caminos para nuestra casa.	He wasn't alarmed, because there were other ways to get to our house.
Próspera Arango, la cachaca, le suplicó que hiciera algo por su padre que estaba agonizando en el sardinel de su casa, inmune a la bendición fugaz del obispo.	Próspera Arango, the uplander, begged him to do something for her father, who was in his death throes on the stoop of his house, immune to the bishop's fleeting blessing.
«Yo lo había visto al pasar —me dijo mi hermana Margot—, y ya tenía cara de muerto».	"I'd seen him when I passed," my sister Margot told me, "and he already had the face of a dead man."

Figure 17: Process of manual correction after splitting the cells

As a result of the alignment, a TM document emerged which was saved as a TXT file and it was exported as a TMX file through the *Wordfast classic* Data Editor afterward, as it is shown in Figure 18.

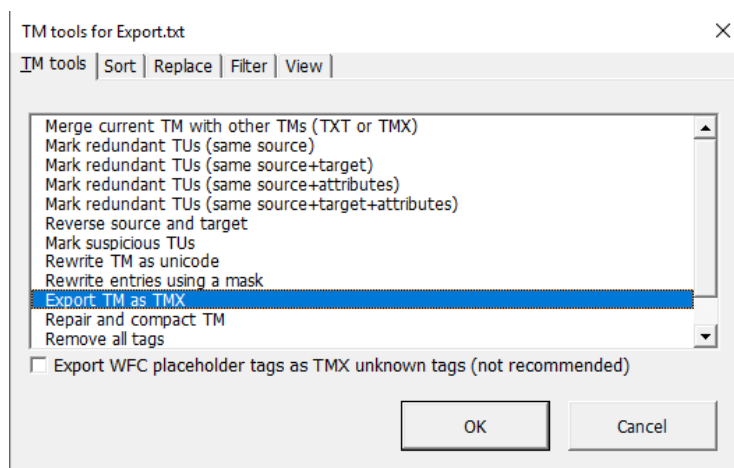


Figure 18: Exporting a TM file as a TMX file

3.3 Uploading the data

The newly-created TMX file could be uploaded to Sketch Engine by clicking on “Upload TMX of XSL,” specifying the name of the corpus and choosing the source file, which is demonstrated in Figure 19. Then the names and languages were set, as in Figure 20, and the parallel corpora could be created. It was successfully processed, as Figure 21 indicates.

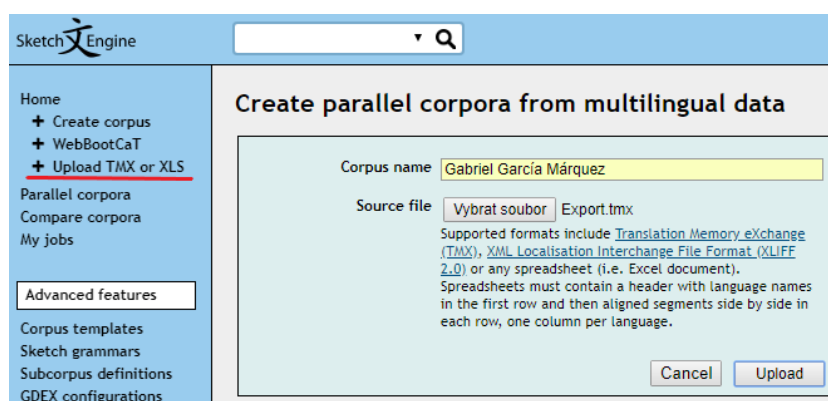


Figure 19: Uploading the TMX file

Create parallel corpora from multilingual data

Corpus name (en)

Corpus language (en)

Corpus name (es)

Corpus language (es)

Figure 20: Finishing the compilation of the corpus

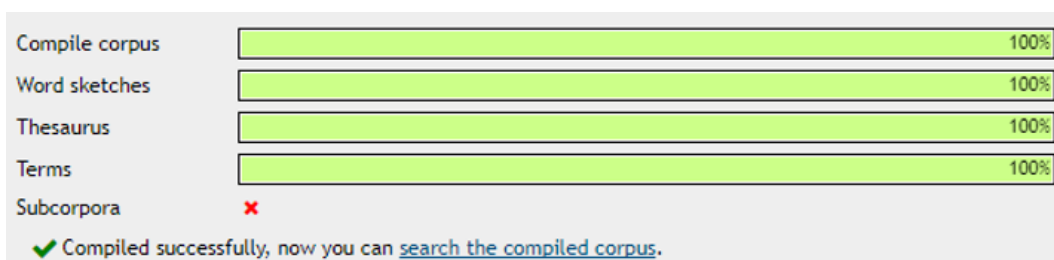


Figure 21: The corpus is finished

The size of the newly created parallel corpora is illustrated in Table 3. The Spanish corpus (*Gabriel García Márquez, Spanish*) contains 26,846 words, which includes 28,455 tokens, while the English corpus (*Gabriel García Márquez, English*) is comprised of 26,291 words and 30,053 tokens.

Name	Words	Tokens
<i>Gabriel García Márquez, Spanish</i>	26,846	28,455
<i>Gabriel García Márquez, English</i>	26,291	30,053

Table 3: Size of the corpora

4 Case study

To explore the newly-created parallel corpus, a case study focusing on lexicalization patterns of Motion events is conducted. The present chapter provides a theoretical background summarized in Subchapter 4.1, based on the theory and previous researches, the research questions of the case study are defined in Section 4.2, and the methodology used in the research is described in Section 4.3. Finally, the data analysis is presented in Section 4.4.

4.1 Introduction

4.1.1 Lexicalization patterns of Motion events

Talmy (1985, 60) defines Motion event as a situation where an entity makes a movement (or is moved) or maintains its location. According to him, it consists of four basic *internal* semantic elements (1985, 61), i.e. **Figure** (the object of moving or maintaining the location), **Ground** (the reference object), **Motion** (the presence of motion or location), and **Path** (the direction of the movement or the occupied location). The components are demonstrated in (1) where *the keg* represents the Figure, *the storeroom* is the Ground, the Motion is expressed by *rolled* or *pushed*, and *into* functions as the Path. According to Talmy (2000, 25–26), the Motion component typically refers either to *translational motion*, i.e. to the motion when the Figure changes its location, or *self-contained motion*, i.e. the Figure preserves its location, e.g. rotation, oscillation, etc.¹⁶

- (1) a) *I rolled the keg into the storeroom.*
b) *I pushed the keg into the storeroom.*
(Talmy 2000, 28)

Additionally, Talmy (2000, 26) associates Motion event also with *external* semantic elements, i.e. *co-event* including usually **Manner** or **Cause**. The difference between these two components lies in what the Figure does or what the Agent (or Instrument) does (2000, 28–29). The example can be seen in (1a) where *rolled* refers to what *the keg* did, therefore, it expresses the Manner or the way in

¹⁶ Since the thesis focuses on the *translational motion*, the *self-contained motion* will not be further taken into consideration.

which *keg* was moving, while in (1b) *pushed* refers to what *I* did, thus it shows the Cause of the event or the origin of the motion.

Talmy (1985, 75) considers the Path to be the defining component of a Motion even, therefore, he divides the world languages into two main typological groups, i.e. *verb-framed* and *satellite-framed* languages,¹⁷ based on where this semantic element is characteristically employed while describing Motion events. Spanish is classified as a verb-framed language, since the Path is typically expressed in the main verb (e.g. *subir* [ascend], *bajar* [descend]), whereas English and Czech belong to satellite-framed languages, as they express the Path by using *satellites* (e.g. English: *go up*, *go down*; Czech: *výjít* [up-go], *sejít* [down-go]). In the verb root of satellite-framed languages the Manner/Cause is typically encoded as opposed to verb-framed languages which optionally express the Manner/Cause by adjunct, see (2).

- (2) a) *La botella se fue de la orilla (flotando).*
the bottle moved-away from the bank (floating)
“The bottle **floated** away from the bank.”
(Talmy 1985, 70)

As opposed to Talmy who regards the Path as the major component determining the typological differences, Cifuentes-Férez et al. (2006, 456) provide a distinct perspective. They suggest that “the difference in semantic encoding of motion events between verb-framed and satellite-framed languages may reside chiefly in the Manner component, and not in the path component.” Their study was based on the *novel word mapping technique* (see Naggy and Gentner 1990), namely, English and Spanish speakers were supposed to infer the meanings of novel motion verbs and novel nouns from eight short passages¹⁸ describing unusual events such as “rolling a device designed to remove burrs over one’s clothes” (2006, 447). Results showed that all the participants covered the Path component when interpreting the meanings of novel Motion events, the only

¹⁷ In the thesis also terms *V-languages* and *S-languages* are used to refer to *verb-framed* and *satellite-framed* languages.

¹⁸ Each passage had a verb version and a noun version differing only in the part of speech of the novel word, the context was the same (Cifuentes-Férez et al. 2006, 448).

difference was in *where* the speakers put the component. Correspondingly to the Talmy's theory mentioned above, Spanish speakers put the information about Path in the verb, while English speakers in the preposition or in the particle. The Manner, on the other hand, was included only by the English speakers (in the verb) and completely omitted by the Spanish speakers, which implies that the main distinction in encoding of semantic components in Motion events lies not in the different sensitivity to Path but in *the different sensitivity to Manner*, since the Manner seems to be usually encoded only in satellite-framed languages (2006, 457).

As Slobin (2004, 7) points out, the use of Manner verbs in verb-framed languages is limited due to the fact that V-languages tend not to express the Manner when the Path expression involves *boundary crossing*. This statement follows from Slobin and Hoiting's "boundary crossing constraint" (1994, 494) based on Aske's division of *path phrases* into two types (1989, 6), i.e. a **locative path phrase** (e.g. *Lou ran through the park.*) which describes only a location, thus, no boundary is crossed, and a **telic path phrase** (e.g. *Pat swam into the cave.*) that, besides the path of motion, includes also an "end-of-path state," therefore, the Motion event indicates a crossing of a spatial boundary. Regarding V-languages, using the locative phrase is completely fine (see (3a), while the telic path phrase would not sound natural (as illustrated in (3b) because of "the preference to mark a *change of state* with a verb, rather than by some other device" (Slobin 2004, 7). To express the Manner in boundary crossing events, a subordinate construction is required (as in (3c), but the overall tendency is to omit the Manner to avoid "heavy" constructions.

- (3) a) *El hombre corrió hasta la casa.*
 "The man ran up to the house."
 b) **El hombre corrió en la casa.*
 "The man ran into the house."
 c) *El hombre entró corriendo a la casa.*
 "The man entered running to the house."

Not only that Spanish speakers tend to omit the Manner, they also seem to pay less attention to the Path details as opposed to English speakers as shows study by Slobin (1996, 201). He used so called “frog stories” from children of preschool age to adults depicting a wordless picture story-book *Frog, Where Are You* (Mayer, 1969).¹⁹ The results of the study imply that English narrators focus on descriptions of movement using a greater number of motion verbs and having a richer means for path description, while Spanish narrators rather elaborate descriptions of settings, leaving the Path details to be inferred (1996, 204). Slobin (1996, 205) explains the differences by “a distinct contrast in *rhetorical style* between English and Spanish.” It confirms Slobin’s *thinking for speaking hypothesis* (1991, 12) saying that “in acquiring a native language, the child learns particular ways of thinking for speaking...and...such [language-specific] patterns have implications for the development of *rhetorical style* in each of the languages.” Therefore, for the English rhetorical style it is typical to assert actions, implying the results (exemplified in (4a), as opposed to the Spanish rhetorical style for which it is characteristic to assert results, implying actions (demonstrated in (4b)).

- (4) a) *The boy climbed the tree.*
 b) *El niño está subido en el árbol.*
 the boy is climb-PST-PTCP in/on the tree
 “The boy is in state of having climbed the tree.”
 (Slobin 1991, 19)

While exploring the influences of the typology of languages on their rhetorical style, Slobin (2004, 24) came across evident limitations of Talmy’s binary typology, e.g. regarding *serial-verb languages*,²⁰ it is not always obvious

¹⁹ The book works as a model for stories in many languages, therefore, it appears to be a convenient source of data for several studies, since these “frog stories” follow the same event line but they are not translations of the original.

²⁰ It is a language in which each verb in the series is morphologically unmarked and monosyllabic, e.g. Mandarin Chinese. It is regarded as a satellite-framed language, since “Talmy considers the manner verb to be the main verb and the path verb to be a satellite, because path verbs often do not function as full verbs and because there is a small, closed set of path verbs” (Slobin 2004, 8).

which verb in a series is considered the main verb (if any). Then there are *bipartite verb languages* that have verbs consisting of two morphemes of equal status, one expressing Manner and the other Path. Another case where neither Manner nor Path is without a doubt the main element in clause are *generic verb languages* which have a verb lexicon limited to only 24 “function verbs” and in order to express Motion events a combination of a generic verb (encoding a deictic or aspectual function) and satellite-like elements, “coverbs,” (encoding both Path and Manner) is used. Therefore, it seemed convenient to introduce a new language type, i.e. *equipollently-framed languages* which express both the Path and the Manner by equivalent grammatical forms (2004, 25). This group is further divided into three groups according to the typical construction type, i.e. *serial verb languages* (verb_{MANNER} + verb_{PATH}), *bipartite verb languages* ([manner + path]_{VERB}), and *generic verb languages* (coverb_{MANNER} + coverb_{PATH} + coverb_{GENERIC}). Talmy’s typology and Slobin’s newly proposed language type as well as their examples are summarized in Table 4.

Language type	Preferred means of expression	Typical construction type	Examples
<i>verb-framed</i>	path expressed by finite verb, with subordinate manner expression	verb _{PATH} + subordinate verb _{MANNER}	Romance, Semitic, Turkic, Basque, Japanese, Korean
<i>satellite-framed</i>	path expressed by non-verb element associated with verb	verb _{MANNER} + satellite _{PATH}	Germanic, Slavic, Finno-Ugric
<i>equipollently-framed</i>	path and manner expressed by equivalent grammatical forms	<i>serial verb:</i> verb _{MANNER} + verb _{PATH}	Niger-Congo, Hmong-Mien, Sino-Tibetan, Tai-Kadai, Mon-Khmer, Austronesian
		<i>bipartite verb:</i> [manner + path] _{VERB}	Algonquian, Athabaskan, Hokan, Klamath-Takelman
		<i>generic verb:</i> coverb _{MANNER} + coverb _{PATH} + verb _{GENERIC}	Jaminjungan

Table 4: Tripartite Typology of Motion-Event Constructions (Slobin 2006, 6 after Slobin 2004, 25)

However, not even the addition of the new language type was found satisfactory enough for discourse analysis, since it appears more profitable to lay out factors that influence habitual expressions of Manner (see Section 4.1.1.1) and Path (see Section 4.1.1.2) across languages and that together contribute to particular

rhetorical styles, rather than to divide languages into typological categories (2004, 25).

4.1.1.1 *Manner salience cline*

Regarding habitual expressions of Manner, Slobin’s study (2004, 6) focuses on the *emergence of the owl*²¹ scene in five languages categorized by Talmy as V-languages (Spanish, French, Italian, Hebrew, Turkish) and five categorized as S-languages (English, German, Dutch, Mandarin, Russian).²² It confirms Talmy’s categorization to the extent that V-language narrators almost always use a single path verb, meaning “exit,” in order to describe the appearance of the owl and they mostly do not express the Manner (as shown in Figure 22), while S-languages employ a kind of Manner verb together with a Path satellite, however, Figure 22 also illustrates that “there are obvious differences between the relatively low use of manner verbs by the three Germanic languages, the higher use in Mandarin [Tsou] and Thai, and the much higher use in Russian” (2004, 6).

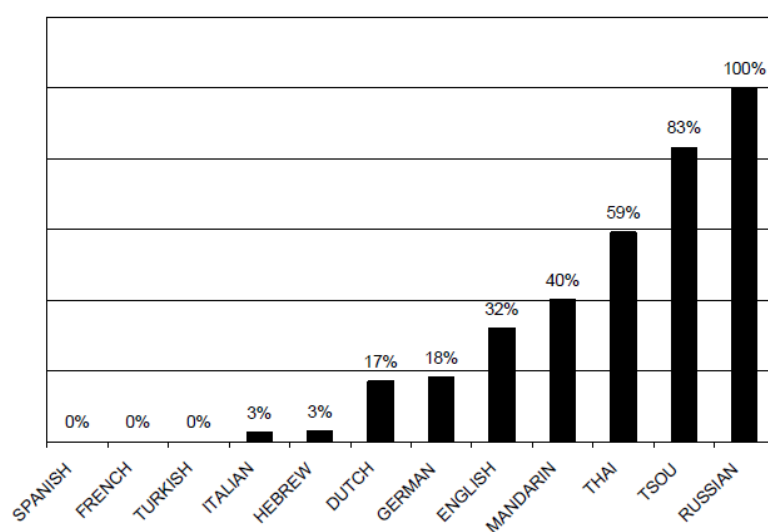


Figure 22: Percentage of Narrators Using a Manner-of-Motion Verb for the Owl’s Emergence (Slobin 2006, 8 after Slobin 2004, 7)

Based on the **intra-typological differences** in degree of Manner description within the group of S-languages, Slobin (2004, 26) proposes that, with respect to Manner of motion, it seems more useful to rank languages on *a cline of manner*

²¹ It is a Motion event selected from *Frog, Where Are You* (Mayer, 1969).

²² A bipartite-verb language Tsou and a serial-verb language Thai were later added to the S-language group.

salience rather than dividing them according to Talmy's two-way or the tripartite typologies regarding the Path expressions. Slobin (2004, 26) distinguishes *High-manner-salient languages* (e.g. Russian) that have an accessible slot for Manner in the language,²³ and *Low-manner-salient languages* (e.g. Spanish) where Manner is subordinated to Path.

According to this approach Germanic and Slavic languages do not seem to belong to the same group, as Germanic languages appear to have more in common with low-manner-salient V-languages than with high-manner-salient Russian. When describing the *emergence of the owl*, speakers of the Germanic languages had a tendency to choose between two options: a deictic option *come out*, or a Manner option *fly/pop/jump out*. Slobin (2004, 8) suggests that the relatively infrequent use of the Manner option was caused by the fact that the Germanic languages, similarly to V-languages, preferred to focus on the exit of the owl rather than the Manner, therefore, the expression *come out* was the most frequent. In case the narrator would like to describe the Manner as well as the aspect of the emergence, a heavier construction, e.g. *come flying out*, would be required (just as in V-languages). In Russian, on the other hand, there is no independent verb equivalent to the verb *come*. In order to express the motion towards the narrator's perspective a deictic prefix *pri-* needs to be added to a motion verb (e.g. *pri-letet'* [come-fly]), while when expressing the owl's emergence, the prefix *vy-* is added to a motion verb (e.g. *vy-letet'* [out-fly]). Since the prefixes cannot be stacked, the combination "come" + "out" is impossible (e.g. **pri-vy-letet'*). As opposed to speakers of Germanic languages, Russian speakers most frequently used constructions consisting of the prefix *vy-* and a Manner verb, e.g. *vy-skocit'* [out-jump], *vy-letet'* [out-fly], *vy-lezit'* [out-crawl]. These findings imply that not only typological character of the language but also morphosyntactic structure is important when determining rhetorical style (Slobin 2004, 8).

4.1.1.2 Path salience cline

When focusing on expressions of Path, languages cannot be compared according to whether they have an accessible slot for Path or not, since it is an obligatory

²³ That is a main verb in S-languages, a Manner verb in serial-verb languages, a Manner morpheme in bipartite verbs, a Manner preverb in Jaminjung languages, etc.

component of Motion events. In other words, when there is no Path component, there is no Motion event. However, Slobin (2004, 17) points out that languages can be compared with respect to their *degree of granularity* of an event description, and elaboration of Path constructions.

The degree of granularity stems from “how many sub-trajectories combine into an overall trajectory” (2004, 17). Slobin studied the degrees of granularity based on the *fall from the cliff*²⁴ scene, which is an example of the overall trajectory that can be divided into various sub-trajectories illustrated in ((5)).

- (5) *moving to the cliff*
stopping at the cliff
throwing the boy and dog down
falling of the boy and dog into the water (Slobin 2004, 18)

Based on the results of his study shown in Table 5, Slobin (2004, 18) claims that speakers of S-languages tend to segment the overall trajectory into more sub-trajectories than speakers of V-languages, therefore, he attributes differences in Path segmentation to the type of lexicalization pattern.

S-LANGUAGES			V-LANGUAGES		
Languages	Average number of event segments	Percentage of narrators mentioning 3 segments	Languages	Average number of event segments	Percentage of narrators mentioning 3 segments
Germanic (Dutch, English, German, Icelandic, Swedish)	3.0	86%	Romance (French, Portuguese, Spanish)	2.1	30%
Slavic (Polish, Russian, Serbo-Croatian)	2.8	76%	Semitic (Hebrew)	2.0	30%

Table 5: Path Segmentation in the Scene of the Fall from the Cliff (Slobin 2004, 18)

²⁴ It is a Motion event selected from *Frog, Where Are You* (Mayer, 1969).

The elaboration of the Path components also seems to depend on the typological character of the language. As mentioned previously, S-languages tend to pay more attention to the Path details resulting in a more dynamic depiction of a Motion event, while V-languages focus rather on scene-setting descriptions, leaving the trajectories to be inferred (Slobin 1996, 204). This is also caused by the fact that in S-languages, e.g. in English, more components of Path, i.e. more Path satellites, can be attached to one verb (as in (6a), while in V-languages, in this case Spanish, each information about Path requires a separate Path verb (as in (6b) because the boundary crossing is involved (Slobin 2004, 17). Consequently, there is a tendency to omit some details about the Path, such as the Path component *pass the trees* mentioned in ((6a) which is omitted in (6b) (Ibarretxe-Antuñano et al. 2013, 257).

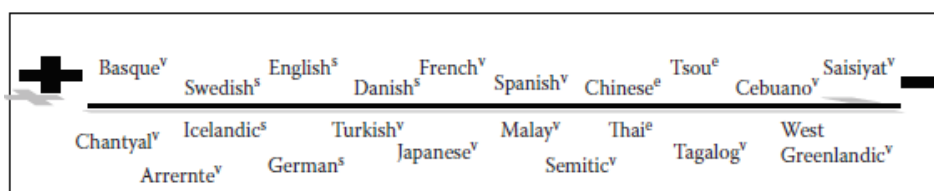
(6) a) *boy **ran** out of the house, over the fence, down the path, pass the trees, into the cave*

b) *el chico **salió** de la casa y tras **cruzar** la valla, **siguió** por el camino y **terminó** en la cueva*

“the boy **exited** of the house and after **crossing** the fence, **followed** over the path and **ended** in the cave”

(Ibarretxe-Antuñano et al. 2013, 257)

However, Ibarretxe-Antuñano (2009, 410) argues against the Slobin’s statement that a degree of elaboration of Path is closely related to the typological group of a particular language. Based on a study of 24 languages,²⁵ she proposes the *path salience cline* that categorizes languages along a continuum between *high-path* and *low-path saliency* as shown in Figure 23.



²⁵ All these verb-framed, satellite-framed and equipollently-framed languages are listed on the Path salience cline (Figure 23) except for Ewe (equipollently-framed language), Hebrew (verb-framed language), and Squiliq (verb-framed language).

Figure 23: Path salience cline (v: verb-framed, s: satellite-framed, c: equipollently-framed) (Ibarretxe-Antuñano 2009, 410)

To conclude, Talmy's typology (1985, 1991, 2000) based on characterization of lexicalization patterns has provided crucial findings about typical tendencies regarding particular language types, however, further researches showed that his theory seems to be "only a part of a complex system of interacting factors" (Slobin 2004, 5). Namely, Slobin (2004) proposing the *manner salience cline* and Ibarretxe-Antuñano (2009) with her suggestion of the *path salience cline* pointed out that also the intratypological differences play a key role when categorizing the languages.

4.1.2 Motion events in translation

Besides of comparing descriptions of Motion events in different languages based on oral narratives stemming from a wordless picture book *Frog, Where Are You* (Mayer, 1969), Slobin also introduced the possibility of using translation as a means of exploring typological as well as intratypological differences. In the first study applying the theory of lexicalization patterns to translation, Slobin (1996, 206) compared five English²⁶ and five Spanish²⁷ novels and their corresponding translations. He chose twenty random Motion events from every novel and evaluated their translations in terms of fidelity to Manner and Path-Ground depictions. The results of his study show that the English translators in both cases mostly follow the Spanish original, sometimes they even add a bit (as demonstrated in (7), the Manner is added in almost a quarter of instances (1996, 212). As opposed to Spanish translators who tend to reduce some components of the full Path-Ground description of English source texts (1996, 210), as well as they omit the Manner information about half of the time (as exemplified in **Chyba! Nenalezen zdroj odkazů.**) (1996, 212).

²⁶ Daphne Du Maurier: *Rebecca* (1938), John Fowles: *The French Lieutenant's Woman* (1969), Ernest Hemingway: *For Whom the Bell Tolls* (1941), Doris Lessing: *A Proper Marriage* (1952), James Michener: *Chesapeake* (1978)

²⁷ Isabel Allende: *La Casa de los Espíritus* [The House of the Spirits] (1982), José Donoso: *Coronación* [Coronation] (1983), Gabriel García Márquez: *Cien Años de Soledad* [One Hundred Years of Solitude] (1967), Ernesto Sabato: *El Túnel* [The Tunnel] (1988), Mario Vargas Llosa: *La Tía Julia y el Escribidor* [Aunt Julia and the Scriptwriter] (1977)

- (7) Don Federico avanzó sin apresurarse...
“Don Federico advanced without hurrying...”
Don Federico walked unhurriedly towards her... (Slobin 1996, 213)

Therefore, English source texts lose more in translation than Spanish ones, regarding both the Manner and the Path-Ground information, since the translators mostly choose not to use a “heavy” constructions in Spanish target texts as in example **Chyba! Nenalezen zdroj odkazů.**)

- (8) He stomped from the trim house...
Salió de la pulcra casa...
“He exited from the trim house...” (Slobin 1996, 212)

- (9) She rustled out of the room...
Salió de cuarto, acompañada del susurro siseante de sus ropas...
“She exited from the room, accompanied by the swishing rustle of her clothing.” (Slobin 1996, 213)

Slobin (1996, 218) suggests that the sparse character of Spanish target texts can be caused not only by the typological features of verb-framed languages, but it can be also related to boundary crossing constraints and consequent using of separate Path clauses for each segment of a complex Motion event, altogether contributing to a particular rhetorical style. The rhetorical style appears to be maintained in both target texts, i.e. in English by adding the semantic components, and in Spanish by reducing and omitting the semantic components.

A bigger project by Slobin (2005), examining factors influencing shaping of rhetorical styles, focused only on a single original text translated into eleven languages: S-languages: English (the source text), Dutch, German, Russian, Serbo-Croatian, and V-languages: French, Spanish, Portuguese, Italian, Hebrew, Turkish. The Chapter 6 of *The Hobbit* (Tolkien 1937) was chosen as the source text, since it is widely translated and, additionally, there is are plenty of vivid Motion events described (2005, 3). The study confirmed the overall tendency of

verb-framed languages not to express Manner and to reduce the number of Path segments or to break Path components into different sorts of segments by inserting a new verb (2005, 11). Regarding S-languages, Slobin (2005, 6) again²⁸ points out some intratypological differences between Germanic and Slavic languages. Concretely, Slavic languages, just like V-languages, are constrained to use separate verbs for path segments based on the fact that Path satellites in Russian and Serbo-Croatian are verb prefixes and no more than one prefix can be combined with a verb (contrary to separable verb particles in Germanic languages). However, unlike V-languages, a rich lexicon of Manner verbs of Slavic languages enables free combinations with Path prefixes. Thus, in the *Hobbit* translations Slavic languages maintain both the Manner and the full Path depiction.

Another study focusing on expressing and translation Motion events by Ibarretxe-Antuñano and Filipović (2013, 251) point out that according to current approaches to Translation Studies, translations are no longer considered only a transfer from a source into a target language requiring a high fidelity to the source text, but they are regarded as “manipulations, retextualization conducted by a translator, who determines what is functionally suitable in the target language.” In other words, overall tendency is to retain the rhetorical style of the target language rather than adapt it to the source language. However, their study shows that in certain contexts the choice of omitting or adding information in order to maintain the rhetorical style can be crucial in contexts, such as witness reports, since an improperly translated report may result in unfair outcome of a trial (2013, 276). The study was based on rating of how violent an event depicted in witness reports appears to be when described in both Spanish and English and the results indicate that the event seems less violent when depicted in Spanish than in English (2013, 272). Omitting the Manner components in Spanish target texts and adding the Manner expressions in English target texts again confirmed the tendency to maintain the rhetorical style of each language.

On the other hand, Alonso’s study (2011, 9) implies a preference to use the L1 lexicalization pattern when describing a Motion event in translations from Spanish into English, i.e. the Path was mostly expressed in the main verb and the

²⁸ See Section 4.1.1.1 *Manner salience cline*.

Manner was expressed in a separate component. The study was based on translation of a series of ten Spanish sentences²⁹ into English. The participants were twenty Spanish speaking translation students who specialize in translation from Spanish/English, English/Spanish (2011, 6). The results showed that the informants produced four types of constructions (as demonstrated in Table 6), i.e. they translated a sentence without using a Motion verb (as in (10a)); the second most frequent construction was to employ a Manner verb as well as Path expression (as in (10b), thus, the lexicalization pattern characteristic for English; the most frequent option was to use a Path verb and express the Manner by a separate expression (see (10c), therefore, L1 lexicalization pattern; and there were also few instances when the informant used a non-Manner verb and expressed the Path by a separate constituent (see (10d).

Category	Percentages
Non motion	3.6%
Manner verb+Path expression	36.4%
Non Manner verb+Manner expression	57.9%
Non Manner verb+Path in a separate constituent	2.1%

Table 6: The frequency of occurrence of each construction used by the translators (Alonso 2011, 8)

- (10) a) El futbolista se deslizó en la portería.
 “The footballer slid into the net.”
[The footballer] finished inside the goal (Alonso 2011, 10)
- b) Se metió en el coche como una flecha.
 “He darted into the car.”
He rushed into the car. (Alonso 2011, 7)
- c) El profesor salió del aula caminando
 “The teacher walked out of the classroom.”
The teacher left the classroom walking. (Alonso 2011, 7)
- d) El futbolista se deslizó en la portería.
 “The footballer slid into the net.”
The footballer fell down into the goal door. (Alonso 2011, 8)

²⁹ The sentences included seven examples of Motion verbs and three distractors.

Alonso (2011, 11) explains the most frequent misuse of the English pattern appearing in the translations as a possible example of *conceptual transfer*,³⁰ consequently, non-target-like translations are produced.

One can say that the resulting non-target like translations in Alonso's study (2011) may be caused by an insufficient level of proficiency of the students taking part in the study. Cifuentes-Férez (2015a, 2), besides other things, investigated whether the level of English of Spanish translators in training has any effect in the production of more English-like descriptions of Motion events. The main goal of her study was to examine, whether the acquisition of English lexicalisation patterns could be facilitated by instructing the students to *think-for-translating*³¹ into English. For the purpose of the study, forty-three participants studying Translation and Interpreting at the University of Murcia were chosen, with English level ranging from A2 to C2 according to the Common European Framework of Reference for Languages. The participants were supposed to translate five narrative fragments³² from Spanish into English. Additionally, they were asked "to perform the multiple-choice grammar part of the Oxford Placement Test which consisted of a total of 100 questions" (2015a, 14). Once they completed the task, they were instructed to think-for-translating into English, and one week later they were told to translate other five fragments from Spanish into English (2015a, 14). The results (2015a, 26) shows that giving the instructions contributes to a better performance in the production of English lexicalization pattern constructions, however, the participants still appear to have problems using satellites and prepositional phrases. As for the level of proficiency, surprisingly, no correlation between the translators in training's level of proficiency and use of English lexicalization pattern was found, suggesting that

³⁰ *Conceptual transfer* can be characterized as "the hypothesis that certain instances of crosslinguistic influence in a person's use of one language originate from the conceptual knowledge and patterns of thought that the person has acquired as a speaker of another language" (Jarvis 2007, 44).

³¹ *Thinking-for-translating theory* stems from previously mentioned Slobin's thinking for speaking hypothesis (1991) (see Section 4.1.1), it is a process-oriented methodology which "discusses the consequences that differing attention to Manner may have in the translation process between languages that are typologically different or similar" (Cifuentes-Férez et al. 2015, 276).

³² The fragments were chosen from Leonor Sáez Méndez's oral narrative of Mercer Mayer's (1969) picture book, *Frog, Where are you?* (Cifuentes-Férez 2015, 14).

a higher level of proficiency in English does not predict a better performance in the translation of Motion events from English into Spanish.

So far, several studies on Motion events in translation from Spanish into English and vice versa were described. Contrary to Spanish and English, there are not any studies that would compare English and Czech encoding of Motion events except for Šimoníková's thesis (2016, 33–34) stemming from students' translations of the English story script *Frog, where are you?* (Salt 2009, Mayer 1969). Fourteen Czech students of Translation and Interpreting Studies at Palacky University in Olomouc were given the source text together with the illustrations, and they were supposed to translate it into Czech not knowing the study was focusing on expressions of Motion events. The results indicate that the lexical meaning of the source text is retained as well as the rhetorical style of the target language is preserved. Thus, the Czech Manner verbs express the same degree of detail with respect to the Manner of motion as the English Manner verbs, and the satellites do not seem to cause any restriction in Czech (2016, 53), as opposed to other Slavic satellite-framed languages, e.g. Serbo-Croatian, in which the satellites restrict the coding of Manner in the verb (2016, 51).

To sum up, most of the selected studies on Motion events in translations imply that there is a strong tendency to retain the rhetorical styles of the target texts (Slobin 1996, 2005, Ibarretxe-Antuñano and Filipović 2013). However, there is also a study indicating a preference to follow the source language rhetorical style in the target text while translating Motion events from Spanish into English (Alonso 2011). Even though in the Alonso's study, the participants are students specializing in translation, which could influence the non-target like translations as they are not professional translators, the study by Cifuentes-Férez (2015) shows that the level of proficiency in English does not seem to be a determinant factor in translating Motion events from Spanish into English.

4.2 Research questions

Since the previous studies on Motion Events in translation from Spanish into English and vice versa showed both the tendency to employ rhetorical styles of the target languages, i.e. adding the Manner in translations into English, or omitting the Manner and some aspects of Path in translations into Spanish (Slobin

1996, 2005, Ibarretxe-Antuñano and Filipović 2013), as well as the tendency to follow the source language lexicalization pattern in translations from Spanish into English (Alonso 2011, Cifuentes-Férez 2015), the first research question of the case study is following: Does the rhetorical style of the source language, Spanish, influence the rhetorical styles of the target languages, English and Czech, or are their rhetorical styles retained?

Subsequently, possible intratypological differences between English and Czech translations will be discussed. As previously indicated, Slavic and Germanic languages, namely Russian and English, in spite of both being satellite-framed languages, they (according to Slobin's manner-saliency cline) appear to belong to distinct groups because of different degree of Manner expressed in Motion events. There seems to be only one study, comparing Czech and English intratypological differences (Šimoníková 2016) showing no restrictions in following the source language as well as target language rhetorical styles while translating from English into Czech. The second research question therefore is: Will the two satellite-framed languages, English and Czech, show any intratypological differences when being translated from a verb-framed language, Spanish?

4.3 Methods: the corpus investigation

The unit of data analysis is a Motion event which is for the purpose of the case study defined as a **non-causative translational motion** where a **protagonist** or an **animal** or a **vehicle** of the movement is the Figure. As a source of the data, the Spanish novella *Crónica de una muerte anunciada* (García Márquez 1981) and its English and Czech translations, *Chronicle of a Death Foretold* (Rabassa 1982) and *Kronika ohlášené smrti* (Hodoušek 1984), were used. In case of the English translation, the newly created parallel corpus in Sketch Engine *Gabriel García Márquez, Spanish* was explored. While for examining the Czech translation InterCorp version 10 was searched, concretely, Spanish as a language resource was aligned with Czech, refined only to div.title *Crónica de una muerte anunciada*. In the thesis this small subcorpus will be called *ICv10 es-cz: crónica*, its creation is demonstrated in

Figure 24. The size of each corpus is shown in Table 7.

Name	Tokens
<i>Gabriel García Márquez, Spanish</i>	28,455
<i>ICv10 es-cz: crónica</i>	31,609

Table 7: Size of the corpora
New subcorpus

Corpus: ★

New subcorpus name:

Specify subcorpus using: ▼

Aligned corpora	div.group	div.txtype	div.srclang	div.original
<input type="checkbox"/> InterCorp v10 - Arabic <input type="checkbox"/> InterCorp v10 - Belarusian <input type="checkbox"/> InterCorp v10 - Bulgarian <input type="checkbox"/> InterCorp v10 - Catalan <input checked="" type="checkbox"/> InterCorp v10 - Czech <input type="checkbox"/> InterCorp v10 - Danish <input type="checkbox"/> InterCorp v10 - German <input type="checkbox"/> InterCorp v10 - Greek <input type="checkbox"/> InterCorp v10 - English <input type="checkbox"/> InterCorp v10 - Estonian	<input type="checkbox"/> Acquis 31,360,144 <input checked="" type="checkbox"/> Core 24,704,820 <input type="checkbox"/> Europarl 18,069,611 <input type="checkbox"/> PressEurop 3,308,125 <input type="checkbox"/> Subtitles 50,091,042 <input type="checkbox"/> Syndicate 5,971,488 <input type="button" value="Select all"/>	<input type="checkbox"/> discussions - transcripts 18,069,611 <input type="checkbox"/> drama 25,627 <input checked="" type="checkbox"/> fiction 23,974,540 <input type="checkbox"/> journalism - commentaries 8,306,101 <input type="checkbox"/> journalism - news 973,512 <input type="checkbox"/> legal texts 31,360,144 <input type="checkbox"/> nonfiction 690,922 <input type="checkbox"/> other 2,821 <input type="checkbox"/> poetry 10,910 <input type="checkbox"/> subtitles 50,091,042 <input type="button" value="Select all"/>	<input type="checkbox"/> ct 251 <input type="checkbox"/> de 483,022 <input type="checkbox"/> de 3,644,269 <input type="checkbox"/> el 545,624 <input type="checkbox"/> en 48,310,516 <input checked="" type="checkbox"/> es 13,287,027 <input type="checkbox"/> et 100,972 <input type="checkbox"/> fa 16,801 <input type="checkbox"/> fi 572,389 <input type="checkbox"/> fr 1,000,000 <input type="button" value="Select all"/>	<input type="checkbox"/> 90,730,799 <input type="checkbox"/> No 29,929,704 <input checked="" type="checkbox"/> Yes 12,844,727 <input type="button" value="Select all"/>

div.translator	div.author	div.title
<input type="checkbox"/> 120,889,435 <input type="checkbox"/> Abella, Rafael 33,109	<input type="text" value="Gabriel García Márquez"/>	<input checked="" type="checkbox"/> Crónica de una muerte anunciada <input type="text" value="Start writing..."/>

Figure 24: Creating the subcorpus *ICv10 es-cz: crónica*

4.3.1 The query

In the case all Motion verbs out of 100 most frequent verbs appearing in the Spanish source text will be examined. In order to find them, firstly, the subcorpus *ICv10 es-cz: crónica* was used and a CQL query, namely `[tag="V.*"]`, was employed. Then the Frequency list including Lemmas was selected.

Table 8 shows the resulting Frequency list of the 100 most frequent verbs in *ICv10 es-cz: crónica* where the lemmas that possibly include Motion verbs are marked. Next step was to look up the selected lemmas in the *Gabriel García Márquez, Spanish* and in *ICv10 es-cz: crónica* and sort them out with respect to whether they describe a Motion event.

lemma	freq	lemma	freq	lemma	freq
1 ser	442	40 venir	23	79 caer	11
2 haber	379	41 buscar	23	80 subir	10
3 decir	262	42 contestar	23	81 vestir	10
4 estar	218	43 crear	23	82 envolver	10
5 tener	114	44 querer	23	83 prevenir	10
6 ver	112	45 despertar	22	84 erar ser	9
7 hacer	100	46 tomar	21	85 andar	9
8 saber	77	47 recibir	20	86 comer	9
9 ir	73	48 mandar	19	87 declarar	9
10 pensar	63	49 alcanzar	18	88 convencer	9
11 parecer	62	50 escribir	18	89 bajar	9
12 poder	61	51 gritar	18	90 ordenar	9
13 matar	60	52 ocurrir	17	91 mirar	9
14 llevar	58	53 pedir	17	92 ayudar	9
15 pasar	56	54 deber	17	93 regresar	8
16 volver	53	55 terminar	17	94 costar	8
17 poner	49	56 entender	16	95 explicar	8
18 entrar	46	57 parrar	16	96 revelar	8
19 dar	43	58 llorar	16	97 colgar	8
20 encontrar	43	59 perder	16	98 servir	8
21 contar	41	60 cumplir	15	99 causar	8
22 salir	40	61 correr	15	100 ver vi	8
23 seguir	39	62 cantar	15		
24 esperar	35	63 quitar	14		
25 empezar	34	64 llamar	14		
26 llegar	34	65 comprar	14		
27 quedar	33	66 atravesar	14		
28 acabar	30	67 olvidar	13		
29 dejar	29	68 morir	13		
30 dormir	29	69 levantar	13		
31 sentir	27	70 aparecer	12		
32 recordar	27	71 abrir	12		
33 tratar	26	72 tocar	12		
34 preguntar	26	73 sentar	12		
35 conocer	26	74 asustar	11		
36 hablar	26	75 aprender	11		
37 ir ser	25	76 casarse	11		
38 casar	25	77 beber	11		
39 oír	25	78 cerrar	11		
		79 caer	11		

Table 8: Frequency list of verbal lemmas in *ICv10 es-cz: crónica*

4.3.2 Sorting the data

There was a problem with the real number of instances of the individual verbs as the lemmatisation of the Spanish verbs in the *ICv10 es-cz: crónica* is not precise. The Frequency list of lemmas showed 984 items, while 572 hits out of the total

number of so called lemmas appeared only once, 313 of the items emerged twice and 120 instances were used three times. It is most probably caused by the inflectional character of Spanish language which makes it difficult for the software to properly recognize the inflectional morphemes added to a verb root, e.g. the verb *beber* [to drink] is mentioned in the list more than once: firstly, as a lemma *beber* [to drink] with the occurrence of 11 instances, then as another lemma *beberás* [you will drink] showing 1 instance.

The same problem appears also the other way around: when searching for a verb as a lemma in *Gabriel García Márquez, Spanish*, one cannot be sure that the concordance is complete. E.g. when searching for lemma *salir* [exit] in the *ICv10 es-cz: crónica*, one gets 40 instances, while when searching the lemma in *Gabriel García Márquez, Spanish*, it shows only 39 hits. The same happens with *llegar* [arrive] which occurs 34 times in *ICv10 es-cz: crónica*, and only 30 times in *Gabriel García Márquez, Spanish*.

Therefore, I chose 13 Spanish Motion verbs out of 100 most frequent Spanish verbs found in *ICv10 es-cz: crónica* (illustrated in Table 10), and I manually checked whether there are other inflectional forms of a particular verb which could appear separately in the frequency list due to inaccurate lemmatization of Spanish verbs. Then I saved the results of Spanish and Czech concordances including the “separately counted” instances of the selected verbs found in *ICv10 es-cz: crónica* as an XSLX file, separately for each lemma. Subsequently, I used *Gabriel García Márquez, Spanish* to look up corresponding lemmas and I manually aligned the concrete solutions of English translations from *Gabriel García Márquez, Spanish* to the original and Czech version in the relevant XSLX file (as in

Table 9).

SALIR				
Spanish		Czech		English
con su razón de madre : " Mi hijo no	salía	nunca por la puerta de atrás cuando estaba bien vestido	" Syn nikdy nevycházel zadními vrátky , když byl dobře oblečen . "	In reality, the only valid explanation seemed to be that of Placida Linero, who answered the question with her mother wisdom: " My son never went out the back door when he was dressed up."
6.58 . " De pronto pensé que había vuelto a	salir	armado " , me dijo Cristo Bedoya . Pero encontró	" Vtom mě napadlo , že už zase odešel se zbrání , " řekl mi Cristo Bedoya .	"Suddenly I thought that he'd come back so that he could go out armed," Cristo Bedoya told me.

Table 9: Example of the manual alignment of English translation with the Spanish and Czech concordance from *ICv10 es-cz: crónica*

Once all the three language versions were aligned in a table, I started to sort out the verbs according to whether they include a Motion event or not. Then the Motion events that concern the previously determined Figure were chosen. The total number of instances of selected Spanish Motion verb are listed in Table 10.

Type of Spanish Motion verb	Spanish original	Total number of hits	English equivalent	Czech equivalent
Path verb	<i>Salir</i>	37	[leave]	[vyjít, odejít]
	<i>irse</i>	15	[leave]	[odejít]
	<i>entrar</i>	48	[enter]	[vstoupit]
	<i>bajar</i>	9	[descend]	[sestoupit]
	<i>subir</i>	11	[ascend]	[vystoupit, vystoupat]
	<i>llegar</i>	24	[arrive]	[přijít, přijet]
	<i>venir</i>	23	[come]	[přijít, přijet, dorazit]
	<i>alcanzar</i>	6	[reach]	[dostihnout]
	<i>atraversar</i>	9	[cross]	[přejít, přejet]
Neutral verb	<i>ir</i>	30	[go]	[jít]
Manner verb	<i>caer</i>	4	[fall]	[padat]
	<i>correr</i>	10	[run]	[běžet]
	<i>andar</i>	5	[walk]	[chodit]

Table 10: Selected Spanish Motion verbs and their English and Czech equivalents

As for the Spanish verb *ir* [to go], the procedure was more complicated due to the fact that the past tense of the verb coincides with the past tense of the verb *ser* [to be], therefore, I downloaded all the instances of the verb *ser* [to be] and *ir* [to go] occurring in *ICv10 es-cz: crónica*, included in lemmas “ser,” “ir,” and “ir|ser,” and I eliminated the cases when it appeared in sense of *to be*. During the process of elimination, I also came across instances of the verb *irse* [leave] which did not appear in the Frequency list as a separate lemma, thus the instances were gathered to create a separate group. There were also many cases where *ir* [to go] functions as a part of verbal periphrasis consisting of a conjugated form of *ir* + *a* preposition + an infinitive expressing periphrastic future (as exemplified in (11)). These instances were removed from the data as well.

- (11) *Vamos a matar a Santiago Nasar.*
go-PRS.1PL to kill-INF Santiago Nasar-ACC.SG.M
“We’ re going to kill Santiago Nasar.”
(TN:11879, file:6254896, *Gabriel García Márquez, Spanish*)

After all these eliminations, the corresponding English chunks from *Gabriel García Márquez, Spanish* were aligned with the remaining Spanish instances and their Czech translations, and the resulting data was further sorted with respect to Motion character of the verb and its Figure as indicated before.

4.4 Data analysis

In order to evaluate the translations of Motion events with respect to the original, the Spanish verbs were divided into groups according to the prevalent semantic element included in the verb root, i.e. Path verbs, Manner verbs, and neutral verbs (as can be seen in Table 10). Depending on this categorization of the verbs, groups of *translation strategies*, adapted from Cifuentes-Férez (2015b, 9–10), were established, i.e. *Path strategies*, *Manner strategies*, and *strategies for neutral Motion events* (as demonstrated in Table 11).

PATH STRATEGIES	
Strategy P-1	<p>Omission of the Path verb but retention of the Path expression.</p> <p>e.g. <i>para que él pudiera entrar otra vez</i> “so that he could get back in” → <i>aby...mohl zase doynitř</i> “so that he could inside again” (es:Garcia Marquez-kronika:0:24:8)</p>
Strategy P-2	<p>Complete omission of the Motion event.</p> <p>e.g. <i>el buque de vapor en que llegaba el obispo</i> “the steamboat in which the bishop arrived” → <i>the bishop’s steamboat</i> (TN:1898, file:6254896, Gabriel García Márquez, Spanish)</p>
Strategy P-3	<p>Translation of a Path verb by using a Path verb.</p> <p>e.g. <i>salió del cuarto</i> “left the room” → <i>left the room</i> (TN:1422, file:6254895, Gabriel García Márquez, Spanish)</p>
Strategy P-4	<p>Translation of a Path verb by using a Path verb + Path expression.</p> <p>e.g. <i>entró en el dormitorio de Santiago Nasar</i> “entered the santiago Nasar’s room” → <i>proceeded into Santiago Nasar’s bedroom</i> (TN:24817, file:6254896, Gabriel García Márquez, Spanish)</p>
Strategy P-5	<p>Translation of a Path verb by using a neutral verb.</p> <p>e.g. <i>Ahí viene...</i> “There he comes...” → <i>Tamhle jde...</i> “There he goes...” (es:Garcia_Marquez-kronika:0:280:1)</p>
Strategy P-6	<p>Translation of a Path verb by using a neutral verb + Path expression.</p> <p>e.g. <i>no salía a la calle</i> “→ <i>she hadn’t gone out into the streets</i> (TN:4304, file:6254896, Gabriel García Márquez, Spanish)</p>

Strategy P-7	<p>Inclusion of a different type of Path in translation.</p> <p>e.g. <i>El último que <u>salió</u>...</i> “the last to leave...” → <i>The last to <u>come</u>...</i> (TN:26589, file:6254896, <i>Gabriel García Márquez, Spanish</i>)</p>
Strategy P-8	<p>Substitution of a Path verb for a Manner verb.</p> <p>e.g. <i>la monja <u>entró</u> en el dormitorio</i> “the nun entered the bedroom” → <i>[jeptiška] <u>vrazila</u> do ložnice</i> “the nun tore into the room” (es:Garcia_Marquez-kronika:0:173:7)</p>
Strategy P-9	<p>Substitution of a Path verb for a Manner verb + Path expression.</p> <p>e.g. <i>la monja <u>entró</u> en el dormitorio</i> “the nun entered the bedroom” → <i>the nun <u>ushed into</u> the bedroom</i> (TN:16474, file:6254896, <i>Gabriel García Márquez, Spanish</i>)</p>
Strategy P-10	<p>Translation of a Path verb + Manner expression by using Path verb + Manner expression.</p> <p>e.g. <i>las mujeres <u>salían corriendo</u> de los patios</i> “the women ran out of the yards” → <i>women <u>came running</u> out of their yards</i> (TN:4574, file:6254896, <i>Gabriel García Márquez, Spanish</i>)</p>
Strategy P-11	<p>Substitution of a Path verb for a non-motion verb.</p> <p>e.g. <i><u>entraban</u> a tomar el primer café</i> “they entered to have the first cup of coffee” → <i>they would <u>drop in</u> to have their first cup of coffee</i> (TN:14329, file:6254896, <i>Gabriel García Márquez, Spanish</i>)</p>
STRATEGIES FOR NEUTRAL MOTION EVENTS	
Strategy N-1	<p>Complete omission of the Motion event.</p> <p>e.g. <i>cuando <u>iban</u> para el mercado</i> “when they were going to the market” → <i>cestou na trh</i> “along the way to the market” (es:Garcia_Marquez-kronika:0:154:8)</p>
Strategy N-2	<p>Translation of a neutral verb by using a neutral verb.</p> <p>e.g. <i><u>Fue</u> al baño...</i> “he went to the bathroom...” → <i>He <u>went</u> to the bathroom...</i> (TN:16338, file:6254896, <i>Gabriel García Márquez, Spanish</i>)</p>
Strategy N-3	<p>Translation of a neutral verb by using a Manner verb.</p> <p>e.g. <i><u>Iban</u> tan contentos...</i> “they were going along so contentedly...” → <i><u>Vykračovali</u> si tak spokojeně...</i> “they were walking so contentedly” (es:Garcia_Marquez-kronika:0:223:3)</p>
Strategy N-4	<p>Translation of a neutral verb by using a Path verb.</p>

	e.g. <i>Iba otra vez hacia el puerto...</i> “he went again to the docks...” → <i>He was heading toward the docks again...</i> (TN:25186, file:6254896, <i>Gabriel García Márquez, Spanish</i>)
Strategy N-5	Substitution of a neutral Motion event for a non-motion verb. e.g. <i>iba solo en un avión</i> “he went by plane alone” → <i>sedí sám v letadle</i> “he is sitting alone in a plane” (es:Garcia_Marquez-kronika:0:3:4)
MANNER STRATEGIES	
Strategy M-1	Omission of the Manner verb but retention of the Manner expression. There were <u>no examples</u> of such a strategy when translating a Manner verb.
Strategy M-2	Complete omission of the Motion event. There were <u>no examples</u> of such a strategy when translating a Manner verb.
Strategy M-3	Translation of a Manner verb by using a Manner verb. e.g. <i>corrió hacia la puerta principal</i> “he ran to the main door” → <i>he ran to the main door</i> (TN:27208, file:6254896, <i>Gabriel García Márquez, Spanish</i>)
Strategy M-4	Translation of a Manner verb by using a neutral verb. e.g. <i>Andaba de pueblo en pueblo...</i> “I’ve been walking from town to town...” → <i>I’ve been going from town to town...</i> (TN:5555, file:6254896, <i>Gabriel García Márquez, Spanish</i>)
Strategy M-5	Translation of a Manner verb by using a Manner verb + Path expression. e.g. <i>se echó a andar</i> “he started to walk” → <i>vykročil</i> “set off walking” (es:Garcia_Marquez-kronika:0:290:4)
Strategy M-6	Inclusion of a different type of Manner in translation. e.g. <i>andaba solo</i> “he went about alone” → <i>jezdil sám</i> “he rode alone” (es:Garcia_Marquez-kronika:0:203:14)
Strategy M-7	Substitution of a Manner verb for a non-motion verb. e.g. <i>andaba reclamando lo suyo</i> “was going about demanding what was hers” → <i>[duše manželky]se dožaduje svých práv</i> “wife’s soul is demanding her rights” (es:Garcia_Marquez-kronika:0:191:7)

Table 11: Translation strategies for Motion events (adapted from Cifuentes-Férez 2015b)

Since Spanish is a verb-framed language, which means that the Path component of a Motion event tend to be encoded in the main verb, it is not surprising that the majority of Motion events found in both corpora, *ICv10 es-cz: crónica* as well as *Gabriel García Márquez, Spanish*, include Path verbs. There are 182 instances of Spanish Path verbs, as opposed to 30 neutral verbs, and 19 Manner verbs.

Results of the Path strategies (illustrated in Table 12) show that the most common tendency in the English target text (TT) is to simply use a Path verb as a Path strategy, e.g. *irse* [leave] → *leave, depart*. This strategy was chosen in more than 40% of the cases, contrary to the Czech target text in which this option of translation was not found at all. It implies that in the Czech target text there is a stronger tendency to rather add some information about the Path or the Manner than use only a Path verb itself. The most common strategy in the Czech target text, as well as most frequent strategy in total, is to use a neutral verb and complete it with a Path prefix (i.e. P-6 strategy), e.g. *salir* [leave] → *ode-jít, odcházet, vy-jít, vy-cházet* [out-go].

PATH STRATEGIES		English	Czech	Total
P-1	Omission of the Path verb but retention of the Path expression.	0	1	1
P-2	Complete omission of the Motion event.	4	2	6
P-3	Path verb	76	0	76
P-4	Path verb + Path expression	33	19	52
P-5	Neutral verb	7	14	21
P-6	Neutral verb + Path expression	51	90	141
P-7	Different type of Path	2	10	12
P-8	Manner verb	1	4	5
P-9	Manner verb + Path expression	4	39	43
P-10	Path verb + Manner expression	1	0	1
P-11	Non-motion verb	3	3	6

Table 12: Results of the Path strategies

Regarding the Different type of Path Strategy, in both TTs there are instances where even though the Spanish Path verb is translated by another Path verb, the Paths do not correspond. For example, in case of *salir* [leave], the verb expresses a movement “out of something,” “out of someone,” or “from the inside to the outside,” however, examples (12b, c) show that the target verbs are

concerning rather movement “into something,” “towards someone.” Therefore, a change of the deictic centre seems to be implied.

- (12) a) *Prudencia Cotes salió a la cocina*
“Prudencia Cotes went out to the kitchen”
- b) *Prudencia Cotes came into the kitchen*
(TN:14438, file:6254896, *Gabriel García Márquez, Spanish*)
- c) *vešla do kuchyně Prudencia Cotesová*
“Prudencia Cotes came into the kitchen”
(es:Garcia_Marquez-kronika:0:157:2)

In the theoretical introduction of the case study, phenomenon of a boundary crossing constraint was mentioned. Because of the constraint, the Manner of Motion usually must be inferred, or it is rarely expressed by a combination of a Path verb and a separate Manner expression. That seems to be the reason why there are only 19 instances of Manner verb and only 4 instances of Path verbs accompanied by a Manner expression. An example of the constraint can be seen in (13a) where a spatial boundary of “los patios” [the yards] is crossed, the Manner is therefore expressed by a separate constituent, i.e. a semipredicative construction, otherwise the sentence would not sound natural. The boundary constraint appears to reveal the use of the source text rhetorical style in the English TT as the “heavy” construction is used as well (see (13b)). The Czech TT, on the other hand, seem to follow its own rhetorical style, since a telic Path phrase is used (see (13c)).

- (13) a) *las mujeres salían corriendo de los patios*
“the women came running out of the yards”
- b) *women came running out of their yards*
(TN:4574, file:6254896, *Gabriel García Márquez, Spanish*)
- c) *ženy vybíhaly ze dvorků*
“the women ran out of the yards”
(es:Garcia_Marquez-kronika:0:41:3)

However, the three remaining examples of the Spanish Path verb accompanied by a Manner expression are translated into English as Manner verbs, such as *run out* (translation of another instance of *salir corriendo* [come running]), *stagger out* (*salir dando tumbos* [come staggering]), or *run* (*venir corriendo* [come running]). As for the Czech TT, the Manner verbs are used as well, i.e. *vy-běhl* [out-run], *vy-potáčet se* [out-stagger], *běžet* [run]. Another instance of a Spanish verb complemented by a semipredicative Manner expression is also included in the strategies for neutral Motion, as it is shown in (14).

- (14) a) *Me sentía como cuando uno va corriendo en un caballo...*
 “I felt the way you do when you are running on a horseback...”
- b) “I felt the way you do when you’re galloping on horseback...”
 (TN: 27860, file:6254896, *Gabriel García Márquez, Spanish*)
- c) *Měl jsem pocit, jako když člověk uhání na koni...*
 “I had a feeling such as when you ride at a gallop on a horseback...”
 (es:Garcia_Marquez-kronika:0:289:12)

In translations from satellite-framed languages into verb-framed languages omission of the Manner appears to be a characteristic feature, while adding the Manner is expected in inverse translations. The results indicate that the Manner is omitted neither in English nor in Czech target text (see Table 13) and only one Spanish Manner verb seems to be translated by a non-motion verb (as illustrated in Table 11, Strategy M-7) but adding the Manner component is not that common for both texts. In the English TT the Manner was added, i.e. expressed in situations where the Manner was not included in the source text (ST), only twice, whereas in the Czech TT, the Manner was added in 31 cases, as demonstrated in (15c).

MANNER STRATEGIES		English	Czech	Total
M-1	Omission of the Manner verb but retention of the Manner expression	0	0	0
M-2	Complete omission of the Motion event	0	0	0
M-3	Manner verb	8	12	20
M-4	Neutral verb	4	1	5
M-5	Manner verb + Path expression	7	3	10
M-6	Different Manner verb	0	2	2

M-7	Non-motion verb	0	1	1
-----	-----------------	---	---	---

Table 13: Results of the Manner strategies

- (15) a) *[los perros] salieron a su encuentro*
 “the dogs came out to meet him”
- b) *[the dogs] came out to meet him*
 (TN:23486, file:6254896, *Gabriel García Márquez, Spanish*)
- c) *přiběhli mu naproti [psi]*
 (es:Garcia_Marquez-kronika:0:233:2)

The strategies for neutral Motion events regard only one type of a neutral verb, i.e. *ir* [go]. Except for translating the Spanish neutral verb by an English or Czech equivalent, which is apparently the most frequent strategy, the Czech TT indicates well developed lexicon of Manner verbs, e.g. *vy-kračovat si* [march], *chodit* [walk], *pospíchat* [rush], and *jezdit* [ride]. Quite a considerable number of instances, in comparison to other strategies, occurred regarding the N-5 strategy, i.e. substitution for a non-motion verb phrase, such as *be in a hurry*, or *zastavit se* [drop by], *sedět* [sit], and *stát* [stand].

STRATEGIES FOR NEUTRAL MOTION EVENTS		English	Czech	Total
N-1	Complete omission of the Motion event	1	2	3
N-2	Neutral verb	19	13	32
N-3	Manner verb	1	9	10
N-4	Path verb	5	1	6
N-5	Non-motion verb	4	5	9

Table 14: Results of the strategies for neutral Motion events

Overall results show that the Czech target text appears to be more abundant in diverse types of verbs, since in almost all the cases there are more types of translated verbs in the Czech TT than in the English TT, except for verbs *bajar* [descend], *alcanzar* [reach], and *caer* [fall]. It can indicate either a more developed lexicon of Motion verbs, or it can mean a greater effort to retain the Czech rhetorical style. As opposed to the English TT that mostly follows the lexicalization patterns of the source language and the Manner is not added to the Motion event so frequently as in the Czech target text is.

Conclusions

The thesis had two main objectives, the first was to create parallel corpora that would include original Spanish fiction and its equivalent English translation, since such corpora do not seem available. Therefore, a novella by Gabriel García Márquez, *Crónica de una muerte anunciada* (1981), was chosen and as well as its English translation, *Chronicle of a Death Foretold* (Rabassa 1982). The texts were aligned and formatted into the TMX file so that they could be uploaded to Sketch Engine.

Once the new parallel corpora *Gabriel García Márquez, Spanish* and *Gabriel García Márquez, English* were compiled, the next main purpose of the thesis, the case study on lexicalization patterns of Motion events, could be done. However, the newly-created parallel corpus was not that helpful, as the lemmatization of Spanish verbs was very imprecise, due to the inflectional character of the language. Thus, the data had to be sorted mostly manually.

The case study was based on Talmy's typology of verb-framed and satellite-framed languages that characteristically encode the Path component of Motion event either in the main verb or in the satellite (Talmy 1985, 75). However, Talmy's binary typology seems to have its limitations, since there appear to be intratypological differences regarding Manner salience cline (Slobin 2004, 26) and Path salience cline (2009, 410). In the previous research it was found that the Manner tends to be omitted in the Spanish target texts, contrary to English target texts which usually add a Manner or a Path expression in order to follow its rhetorical style (Slobin 1996, 2005, Ibarretxe-Antuñano and Filipović 2013). Therefore, verb-framed language Spanish and its translation into satellite-framed languages English and Czech were examined with respect to whether they retain its rhetorical style or not and to show whether there appear any intratypological differences between English and Czech.

As for English target text tendencies, it mostly follows the source language rhetorical style, since mainly Path and neutral verbs are used, while the Czech target text shows more abundant lexicon of Motion verbs and the Manner component is quite frequently added. Thus, it seems that the Czech target text is inclined to keep its rhetorical style. Following seemingly different rhetorical

styles may indicate a different sensitivity to the Manner component, which can be considered an intratypological difference.

Concerning the fact that the novella chosen for the parallel corpora compilation was not originally meant to be used for a Motion event case study, the text is not exactly full of Motion verbs and it does not describe any complex, vivid Motion events. Another inconvenient aspect is that only one speaker of particular languages is providing the source or target text, thus the data are not quite objective. Even though the study reveals some interesting translation strategies such as a change of a deictic centre, or adding the Manner information to the verb, further research would be needed in order to make more generalizable conclusions.

Works Cited

- Aske, John. 1989. "Path predicates in English and Spanish: A Closer look." In *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society*, edited by Kira Hall, Michael Meacham, and Richard Shapiro, 1–14. Berkeley: Berkeley Linguistics Society.
- Aijmer, Karin and Bengt Altenberg. 1996. "Introduction." In *Languages in Contrast*, edited by K. Aijmer, B. Altenberg, M. Johansson, 11–16. Lund: Lund University Press.
- Alonso, Rosa. 2011. "The translation of motion events from Spanish into English: a cross-linguistic perspective." *Perspectives* 19(4):353–366(1–14). Accessed March 7, 2018. doi:10.1080/0907676X.2011.607238.
- Chowdhury, Gobinda. 2005. "Natural language processing." *Information Science and Technology* 37(1):51–89. Accessed April 3, 2018. <https://doi.org/10.1002/aris.1440370103>.
- Cifuentes-Férez, Paula, and Dedre Gentner. 2006. "Naming motion events in Spanish and English." *Cognitive Linguistic* 17(4):443–462. Accessed March 1, 2018. doi:10.1515/COG.2006.014.
- Cifuentes-Férez, Paula. 2015a. "Thinking-for-translating: Acquisition of English physical motion constructions by Spanish translators in training." *Cognitive Linguistic Studies* 2(2):302–329 (1–40). Accessed April 30, 2018. 10.1075/cogls.2.2.05cif.
- Cifuentes-Férez, Paula. 2015b. "Language, culture and translation: motion events in English and Spanish." In *Studies in Philology: Linguistics, Literature and Cultural Studies in Modern Languages*, edited by Gema Alcaraz Mármol, Maria del Mar Jiménez-Cervantes Arnao, 1–17. Cambridge: Cambridge Scholars Publishing. https://www.researchgate.net/publication/274068337_Language_culture_and_translation_motion_events_in_English_and_Spanish
- Cifuentes-Férez, Paula, and Ana Rojo. 2015. "Thinking for translating: A think-aloud protocol on the translation of manner-of-motion verbs." *Target* 27(2):273–300. Accessed April 30, 2018. <https://doi.org/10.1075/target.27.2.05cif>.
- García Márquez, Gabriel. 1981. *Crónica de una muerte anunciada*. Colombia: La Oveja Negra.
- Granger, Sylviane. 2003. "The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies." In *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, edited by Jacques Lerot and Stephanie Petch-Tyson, 17–29. Amsterdam: Rodopi. https://www.researchgate.net/publication/236262589_The_corpus_approach_a_common_way_forward_for_Contrastive_Linguistics_and_Translation_Studies.
- Filipović, Rudolf. 1969. "The Yugoslav Serbo-Croatian-English Contrastive Project." Zagreb University, Institute of Linguistics. Accessed March 26, 2018. www.eric.ed.gov/PDFS/ED121083.pdf.

- Hasselgård, Hilde. 2010. "Parallel corpora and contrastive studies." Paper presented at The International Symposium on Using Corpora in Contrastive and Translation Studies, Edge Hill University, Ormskirk, July 27–29.
- Hodoušek, Edurard, trans. 1984. *Kronika ohlášené smrti*. Odeon.
- Ibarretxe-Antuñano, Iraide. 2009. "Path Salience in Motion Events." In *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaacs Slobin*, edited by Guo Jiansheng, Nancy Budwig, Susan Ervin-Tripp, Kei Nakamura, and Seyda Özçalışkan, 403–414. New York: Psychology Press.
- Ibarretxe-Antuñano, Iraide, and Luna Filipović. 2013. "Lexicalization patterns and translation." In *Cognitive Linguistics and Translation: Advances in Some Theoretical Models and Applications*, edited by Ana Rojo, and Iraide Ibarretxe-Antuñano, 251–281. Berlin/New York: Mouton de Gruyter.
- Jakubiček, Miloš, Adam Kilgarriff, Diana McCarthy, and Pavel Rychlý. 2010. "Fast Syntactic Searching in Very Large Corpora for Many Languages." In *PACLIC 24 Proceedings*, 741–747.
- Jarvis, Scott. 2007. "Theoretical and methodological issues in the investigation of conceptual transfer." *Vigo International Journal of Applied Linguistics* 4:43–72. Accessed April 30, 2018. https://www.researchgate.net/publication/242535284_Theoretical_and_Methodological_Issues_in_the_Investigation_of_Conceptual_Transfer.
- Johansson, Stig, and Hilde Hasselgård. 1999. "Corpora and cross-linguistic research in the Nordic countries." In *Contrastive Linguistics and Translation*, edited by Granger et al. 145–162.
- Johansson, Stig, and Knut Hofland. 1994. "Towards an English-Norwegian parallel corpus." In *Creating and Using English Language Corpora*, edited by U. Fries, G. Tottie and P. Schneider, 25–37. Amsterdam: Rodopi.
- Johansson Stig, Jarle Ebeling, and Signe Oksefjell. (1999) 2002. "English-Norwegian Parallel Corpus: Manual." University of Oslo, Department of British and American Studies. Accessed March 27, 2018. http://www.hf.uio.no/ilos/english/services/omc/enpc/ENPC_manual.pdf.
- Kilgarriff, Adam, Fredrik Marcowitz, Simon Smith, and James Thomas. 2015. "Corpora and Language Learning with the Sketch Engine and SKELL." *Revue Française de Linguistique Appliquée* 20(1):61–80. Accessed April 3, 2018. https://www.researchgate.net/publication/282751109_Corpora_and_Language_Learning_with_the_Sketch_Engine_and_SKELL.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel. 2014. "The Sketch Engine: ten years on." *Lexicography* 1:7–36. doi: 10.1007/s40607-014-0009-9.
- Mayer, Mercer. 1969. *Frog, Where Are You?* New York: Dial Press.
- McEnery, T., Xiao, R., and Tono, Y. 2006. *Corpus-based language studies: An advanced resource book*. London: Routledge.

- Nesselhauf, Nadja. (2005) 2011. "Corpus Linguistics: A Practical Introduction." Heidelberg University, Department of English. Accessed March 26, 2018. <http://www.as.uni-heidelberg.de/personen/Nesselhauf/files/Corpus%20Linguistics%20Practical%20Introduction.pdf>.
- Rabassa, Gregory, trans. 1982. *Chronicle of a Death Foretold*. London: Pan Books.
- Rychlý, Pavel, and Adam Kilgarriff. 2007. "An efficient algorithm for building a distributional thesaurus (and other Sketch Engine developments)." Paper presented at the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, June 23–30.
- Salt software. 2009. *Frog, where are you? By Mercel Mayer*. Accessed February 27, 2016, http://www.saltsoftware.com/coursefiles/1202/FWAY_English.pdf.
- Sketch Engine. 2017. Accessed August 15. <https://www.sketchengine.co.uk/user-guide/user-manual/corpora/corpus-types/>.
- Slobin, Dan I. 1991. "Learning to Think for Speaking: Native Language, Cognition, and Rhetorical Style." *Pragmatics* 1(1):7–26. Accessed November 11, 2017. <https://journals.linguisticsociety.org/elanguage/pragmatics/article/view/330.html>.
- Slobin, Dan I. 1996. "Two ways to travel: Verbs of motion in English and Spanish." In *Grammatical Construction: Their form and meaning*, edited by Masayoshi Shibatani, and Sandra A. Thompson, 195–219. Oxford: Clarendon Press.
- Slobin, Dan I. 2004. "The many ways to search for a frog: Linguistic typology and the expression of motion events." In *Relating events in narrative: Vol. 2. Typological and contextual perspectives*, edited by S. Strömquist, and L. Verhoeven. 219–257(1–32). Mahwah, NJ: Lawrence Erlbaum Associates. <http://verbisintagmatici.caissa.it/Slobin%202004%20-%20The%20many%20ways%20to%20search%20for%20a%20frog.pdf>.
- Slobin, Dan I. 2005. "Relating Narrative Events in Translation." *Institute of Human Development*. University of California, Berkeley, 1–12. Accessed November 11, 2017. https://www.researchgate.net/publication/226468533_Relating_Narrative_Events_in_Translation.
- Slobin, Dan I. 2006. "What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition." In *Space in languages: Linguistic systems and cognitive categories*, edited by M. Hickmann, and S. Robert. 59–81 (1–20). Amsterdam/Philadelphia: John Benjamins. https://www.researchgate.net/publication/313051429_What_makes_manner_of_motion_salient_Explorations_in_linguistic_typology_discourse_and_cognition.
- Slobin, Dan I., and Nini Hoiting. 1994. "Reference to movement in spoken and signed languages: Typological considerations." In *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistic Society*, 487–505. Berkeley: Berkeley Linguistic Society.
- Šimoníková, Petra. 2016. "The Expression of Motion Events in English and Czech." Bc., Palacký University in Olomouc.

- Talmy, Leonard. 1985. "Lexicalization Patterns: Semantic structure in lexical forms." In *Language typology and lexical descriptions*, edited by Tim Shopen, 36–149. Cambridge: Cambridge University Press.
- Talmy, Leonard. 1991. "Path to realization: A typology of event conflation." In *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on The Grammar of Event Structure*, 48–519.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics. Vol. II: Typology and Process in Concept Structuring*. Cambridge, MA: The MIT Press.
- Tolkien, J. R. R. 1937. *The Hobbit or there and back again*. London: George Allen & Unwin.
- Vyvijalová, Magdaléna. 2015. "Intensifiers in English and in Spanish: A Contrastive Corpus-based Study." Bc., Palacký University in Olomouc.

Corpora

Czech National Corpus – Intercorp. Institute of the Czech National Corpus FF UK, Praha.
Accessible at <http://ucnk.ff.cuni.cz/Intercorp/>.

Appendix I – Path strategies

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits						
P-1	SALIR	37	leave	10		1						
P-2												
P-3							come <u>out</u>	9	<u>vystoupit</u>	1		
P-4							step <u>out</u>	1				
P-5									jít	1		
P-6							go <u>out</u>	12	<u>odejít, odcházet</u> <u>vyjít, vycházet</u>	5 20		
P-7							come	1	<u>přijít, přicházet</u>	2		
							come <u>into</u>	1	<u>vejít</u> <u>projít</u>	1 2		
P-8							salir corriendo salir dando tumbos		run <u>out</u>	1	<u>vyběhnout, vybíhat</u>	2
P-9									stagger <u>out</u>	1	<u>vypotácet se</u> <u>přiběhnout</u>	1 1
P-10							salir corriendo		come <u>running</u>	1		
P-11												

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits				
P-1	IRSE	15	depart	1						
P-2										
P-3							leave	8		
P-4										
P-5							go	4	jít	3
P-6							go back	1	odejít	7
							go on	1		
P-7										
P-8										
P-9									<u>odstěhovat se</u> <u>odjet</u> <u>odplout</u>	1 1 2
P-10										
P-11			vzít si	1						

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits
P-1	ENTRAR	48				1
P-2				2		
P-3			enter	8		
			come	1		
P-4			come <u>in</u>	11	<u>v</u> ystoupit, <u>v</u> stupovat	9
			come <u>into</u>	2		
			proceed <u>into</u>	1		
P-5			go	1	jít	3
P-6			go <u>in</u>	8	<u>v</u> ejít, <u>v</u> cházet	17
			go <u>into</u>	9	<u>p</u> řijít, <u>p</u> řicházet	9
			get <u>in</u>	2	dostat se <u>d</u> ovnitř	1
P-7			jít <u>d</u> ál	1		
			<u>p</u> rojit	1		
			<u>z</u> ajít	2		
P-8			vrátit se	1		
P-9			v razit	1		
P-10			r ush into	1	o dskočit si <u>d</u> ovnitř	1
P-11			stop off	1	zastavit se	1
			drop in	1		

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits
P-1	BAJAR	9				
P-2				1		
P-3			drop	1	<u>s</u> nést	1
P-4			bring <u>d</u> own	1	<u>s</u> pustit	1
			come <u>d</u> own	2	<u>v</u> ystoupit	4
P-5						
P-6			get <u>o</u> ut	2	<u>s</u> ejít	2
			get <u>o</u> ff	2	<u>p</u> řijít	1
			go <u>d</u> own	1		
P-7						
P-8						
P-9						
P-10						
P-11						

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits				
P-1	SUBIR	11								
P-2										
P-3										
P-4							come <u>up</u>	1	<u>vystoupit</u>	2
P-5							go	1		
P-6							go <u>up</u>	8	<u>vyjít</u>	2
									<u>jít nahoru</u>	1
									<u>přijít nahoru</u>	1
									<u>vydat se</u>	1
P-7									<u>odejít nahoru</u>	1
P-8									chodit	1
P-9	climb <u>up</u>	1	<u>vyběhnout</u>	1						
P-10			<u>vyjet</u>	1						
P-11										

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits				
P-1	LLEGAR	24								
P-2										
P-3							reach	3		
							come	11		
							arrive	6		
P-4							come <u>in</u>	1		
							come <u>on</u>	1		
P-5							get	1	dostat se	4
									jít	1
									<u>přijít, přicházet</u>	8
			<u>příjet, přijíždět</u>	7						
			<u>přivážet</u>	1						
			<u>příplouvat</u>	2						
P-10										
P-11										

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits					
P-1	VENIR	23	come arrive	1 17 3	jít <u>přijít</u> , přicházet <u>přinášet</u>	1 5 1					
P-2											
P-3											
P-4											
P-5											
P-6											
P-7							venir corriendo	run	1	běžet jet <u>přijet</u>	1 1 13
P-8											
P-9											
P-10											
P-11											

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits
P-1	ALCANZAR	6	catch catch <u>up</u> get <u>ahead</u> get <u>to</u>	2 2 1 1	<u>dostihnout</u> <u>dohonit</u>	1 5
P-2						
P-3						
P-4						
P-5						
P-6						
P-7						
P-8						
P-9						
P-10						
P-11						

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits
P-1	ATRAVERSAR	9	cross pass <u>through</u> go <u>through</u>	5 1 3	jít projít <u>přejít</u> , přecházet, jít přes	1 4 4
P-2						
P-3						
P-4						
P-5						
P-6						
P-7						
P-8						
P-9						
P-10						
P-11						

Appendix II – Strategies for neutral Motion events

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits
N-1	IR ir corriendo	30		1		2
N-2			go	18	jít	13
N-3			attend	1		
			gallop	1	uhánět	1
					vykračovat si	1
N-4				chodit	4	
				pospíchat	2	
				jezdit	1	
N-5			come	3	přijít	1
			heading toward	1		
			go out	1		
			be in a hurry	2	zastavit se	1
			be	2	sedět	1
				stát	1	
				chystat se	2	

Appendix III – Manner strategies

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits		
M-1	CAER	4	fall	1	padnout	2		
M-2							upadnout	2
M-3								
M-4			fall <u>into</u>	2				
M-5								
M-6			fall <u>on</u>	1				
M-7								

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits		
M-1	CORRER	10	run	6	běžet	7		
M-2							utíkat	1
M-3								
M-4			run <u>after</u>	1				
M-5								
M-6			run <u>toward</u>	3	<u>vyběhnout</u>	1		
M-7					<u>doběhnout</u>	1		

Strategy	Spanish original	N. of hits	English translation	N. of hits	Czech translation	N. of hits
M-1	ANDAR	5	walk go go about	1 3 1	putovat vykročit jezdit dožadovat se	1 1 2 1
M-2						
M-3						
M-4						
M-5						
M-6						
M-7						