TIME IS MOTION

A corpus-based study on the quantitative use of cognitive metaphors in translated and non-translated English

Claudia Förster Hegrenæs

Thesis submitted to the Department of Foreign Languages in partial fulfillment of the requirements for the
MA degree

University of Bergen/Norway
Fall 2011

Abstract

Die Übersetzung von kognitiven Metaphern wird oftmals als eine Angelegenheit von konzeptueller Gleichheit oder Ungleichheit zwischen verschiedenen Sprachen und Kulturen betrachtet. Bisherige Studien haben am Beispiel unterschiedlicher Sprachpaare (Ursprungssprache und Zielsprache) eine Reihe von Metaphern erarbeitet, welche den jeweiligen Sprachpaaren gemeinsam sind oder jeweils einer der zwei Kulturen zugeordnet werden können. In der Translationswissenschaft wird im Allgemeinen die übersetzte Varietät einer Sprache als selbstständige Varietät angesehen. Diese Varietät besitzt aber textuelle Merkmale, die sie von vergleichbaren Texten in der gleichen Sprache unterscheidet. Eines dieser Merkmale ist die Übertreibung von texttypischen Charakteristika der Zielsprache in der übersetzten Variante dieser Sprache im Unterschied zur nicht-übersetzten Variante. Übersetzte Sprache weist also quantitativ mehr Fälle dieses textlichen Merkmals auf als die ursprüngliche, nicht-übersetzte Sprache. Mit der Hilfe von drei verschiedenen Textsammlungen (BNC und COCA für originales Englisch und TEC für übersetztes Englisch) wurden quantitative Unterschiede in der Verwendung von metaphorischen Ausdrücken der kognitiven Metapher TIME IS MOTION in übersetztem und nicht-übersetztem Englisch untersucht. Anhand von vier verschiedenen Variablen wurden quantitative Abweichungen in Form von Mehrvorkommen in der übersetzten Varietät von Englisch festgestellt. Diese Ergebnisse deuten darauf hin, dass, unabhängig von Herkunftssprache- und Text, übersetztes Englisch über mehr kognitive metaphorische Ausdrücke verfügt als nicht-übersetztes Englisch. Dies hat Auswirkungen auf zukünftige Forschung in der angewandten Übersetzungswissenschaft, die sich mit Strategien und Prozessen beim Übersetzen von kognitiven Metaphern beschäftigen.

Acknowledgements

First of all, I wish to thank my supervisor Sandra Halverson for her invaluable advice and support during the composition of this thesis. More than once, her guidance was necessary when I was lost in the empirical realm between translation studies and cognitive linguistics.

I also owe a big thank you to my husband and my son who have been very supportive during the last one and a half years. They did not mind spending parts of the summer holidays and a number of weekends by themselves.

Table of contents

List of tables

List of figures

List of abbreviations

BCE	before common era
BNC	British National Corpus
CMT	Cognitive Metaphor Theory
COCA	Corpus of Contemporary American English
DTS	descriptive translation studies
SL	source language
ST	source text
TL	target language
TS	translation studies
TT	target text

1 INTRODUCTION

As a literary device, scholars have been occupied by metaphor for centuries. Originating in the Greek *metapherein*, *metaphor* means ‘to transfer something’ or ‘to carry it over’[1]. As early as ca. 335 BCE, Aristotle brings up the phenomenon of metaphor in his work *Poetics*. In the first half of the twentieth century, the American poet Carl Sandburg wrote in his poem *Fog*[2] that ‘[t]he fog comes on little cat feet’. He generates a mental image of the natural phenomenon fog behaving like an animal – a cat. To oversimplify, metaphor can be described as understanding one thing in terms of another. Sandburg creates a metaphor where the fog is supposed to be understood in terms of an approaching cat. The latter serves as a conceptual source for certain (intended) understanding of the target, the fog. In everyday speech, this comparison might sound rather odd but for the purposes of the poem the line fulfills its duty. More importantly, within the framework of text type (the poem), the line is understood to be a metaphor, a stylistic device which is in no way intended to represent reality. Literary metaphors are, as Ali R. Al-Hasnawi puts it, ‘instances of figurative (as opposite to literal) language’ (2007: unpaginated, introduction).

In 1980, George Lakoff and Mark Johnson proposed a new and different view of metaphor in their book *Metaphors We Live By*. They claim that metaphor is a general cognitive linguistic device that exists in every person’s language use regardless of type of speech (spoken, written, sign), genre (poetry, religion, politics etc.) or even language (English, Chinese, etc.). According to this view, metaphor is a phenomenon which structures human conception and the human thought process and is thus a major component of everyday language. By means of the distinction between language and thought, cognitive metaphor operates on two levels: (1) on the cognitive level within the human mind and (2) on the linguistic level within human speech. In 1997, Frank Boers concludes that ‘abstract thought[s] and their linguistic manifestations are largely metaphorical’ (1997:48). Like Lakoff and Johnson, Boers claims that metaphors are central to all perception and understanding of reality and not only a peripheral phenomenon. Cognitive metaphors aim to explain certain concepts in terms of others (as do literary metaphors) and are thus the means by which people perceive, structure, understand and also communicate reality. Human knowledge is

[1] http://dictionary.reference.com/browse/metaphor
[2] http://www.poetry-archive.com/s/fog.html

structured into conceptual domains which contain certain information (both learnt and self-experienced) and parts of the knowledge of one domain are used to explain parts of another domain.

If, then, cognitive metaphor is a phenomenon governing human cognition and language in general, it also plays an important role in the transmission of discourse from one language into another, i.e. in translation. Nowadays, translation is 'a feature of the internationalization of modern life in almost all spheres of human activity' (Neubert 1997:2) and within the relatively new field of translation studies (TS), an academic discipline closely linked to linguistic studies, cognitive metaphor has experienced growing attention within the last decade. Studies on translational processing and strategies (i.e. how translators deal with linguistic expressions of cognitive metaphors in translations) and on the equivalence of metaphors in source and target expression (Schäffner 2004, Al-Hasnawi 2007, Hegrenæs 2010) have accompanied the field's transformation from a purely normative discipline to an academic discipline. The importance of the translator as a mediator not only between languages but also between cultures has been shown and TS has taken a lift from the purely normative linguistic level to the cognitive linguistic level.

In 1993, Mona Baker claimed that translated language differs from non-translated language and constitutes an independent language variety. Baker introduces what she calls 'universal features of translation' (1993:243). One of them states that translated text is marked by '[a] general tendency to exaggerate features of the target language' (1993:244) which are characterized by quantitative differences between translated and original English texts (245). The majority of studies investigating cognitive metaphors in translation take a qualitative approach often restricted to language pairs (source and target language) to determine translational strategies, translational processing and the potential cognitive workload of the translator (for further reading see Schäffner 2004, Sjørup 2011 and Jääskeläinen 1999). In the course of my own work on cognitive metaphors in translations between English and Norwegian, I came to notice that there are (to some extent large) quantitative differences between the distribution of metaphorical expressions in original English language and English translated from Norwegian. At this point, a few words on the terminology used in this study are necessary. I refer to original English as English which is assumed to have originated in English thought processes and has (presumably) not undergone transmission from another language. In contrast, translated English is assumed to have originated in thought processes of some other language and has later been transferred into English by means of translation or interpreting. To my knowledge, no study investigating quantitative differences (i.e. numeric differences in occurrences) of cognitive metaphors in translated and non-translated English language has been conducted yet. This study, first and foremost, aims to determine such quantitative differences by contrasting lexical expressions of a cognitive metaphor in the target language, i.e. English, and lexical realizations of the same metaphor in translations from a number of source languages into that target language. Thus, the phenomenon of cognitive metaphor in translation is approached from the target language: quantitative differences are investigated ignoring the qualitative meaning of the single tokens by grouping them according to syntactical characteristics, grammatical categories and text varieties. In my view, it is necessary to establish such quantitative differences between translated and non-translated language. It is one thing to determine translational processes and strategies and even to formulate translation rules by thoroughly investigating target texts and their respective source texts. But how do target texts resulting from these strategies and rules fit into the target culture? Are they identifiable as translated texts as Baker claims?

It should be pointed out that this investigation is only one approach to differences in the usage of cognitive metaphors within translated and non-translated English and is by no means intended to be exhaustive. Due to the massive number of cognitive metaphors identified in the literature so far, the methodological approach (corpus studies) and the scope of the paper, the study concentrates on linguistic expressions of the cognitive metaphor TIME IS MOTION as it is represented in metaphorical expressions like *time flies by* or *as the day goes on*.

Following the suggested quantitative approach, this study aims to answer the following research questions:

- Are there quantitative differences between the translated and the non-translated occurrences of metaphorical expressions of the cognitive metaphor TIME IS MOTION in English?
- What do possible differences look like? Are metaphorical expressions over- or underrepresented in translated text compared to non-translated text?

Based on Baker's translation universal and the results of the pilot study, I expect the translated categories to differ quantitatively from the non-translated category. Moreover, I expect the translated categories to be overrepresented compared to the non-translated categories, which conforms to Baker's assumed exaggeration of target text features (1993:244). In other words, regarding the use of metaphorical expressions of the cognitive metaphor TIME IS MOTION, translated English texts are expected to differ from non-translated English texts by using more of these particular expressions.

The structure of this paper is as follows. After introducing theoretical approaches from both translation studies and cognitive linguistics in chapter two, I present the methods and the material I employ to extract, organize and analyze the data in chapter three. In chapter four, the results of the analysis are described according to quantitative distributions between translated and non-translated categories of a number of variables which are introduced in chapter three. Thus, I hope to establish interrelations between the members of these variables and their affiliation to the translated or non-translated category. Chapter five discusses the results of the analysis in chapter four concerning the research questions and the hypothesis and chapter six summarizes the findings, places the study into the prospect of further research and discusses possible problems within the methodology and the analysis.

2 THEORY

The present study is a study conducted within the field of translation studies. The aim is to uncover quantitative differences considering the use of the cognitive metaphor TIME IS MOTION between the translated and the non-translated language varieties of English. Additionally, the study involves two other linguistic disciplines. Firstly, the theory forming the basis for potential quantitative deviations between translated and non-translated English, Cognitive Metaphor Theory (CMT), originates in cognitive linguistics, which is '[a]n approach to the study of language which is based upon human perception and conceptualization of the world' (Trask 2007:40). Secondly, the methodological approach to the study (corpus study) belongs to the relatively young discipline of corpus linguistics and refers to the quantitative and qualitative study of extensive bodies of spoken and written language use.

This chapter on the theoretical groundwork starts by introducing translation studies as an independent empirical discipline before presenting cognitive linguistics and Cognitive Metaphor Theory. In a next step, CMT is connected to TS and particular theoretical approaches are introduced. Corpus studies as a methodological approach is introduced in chapter three.

2.1 Translation studies

Translation studies as an academic discipline is a relatively new field of research which developed out of the need to teach translation (written) and interpreting (oral). What started as an attempt to improve teaching methods and establish guidelines and rules for professional translating and interpreting has long since advanced to a fully evolved scientific field of research which is intertwined with other scientific disciplines like linguistics and psychology.

In its broadest sense, TS deals with 'interhuman communication' (Munday 2001:5) of a special kind: communication between languages and cultures. Since language contact reaches far back in history, writings about translation can be traced back until the first century BCE. Nevertheless, it was not until the second half of the 18th century that foreign language teaching, and thus the teaching of translation, gained scholarly attention when Latin and Greek were taught in schools (ibid.7). From the middle of the 20th century onwards, language teaching and translation experienced an increase in attention and the hitherto preferred strategy of word-to-word translation became criticized. Calls for an alternative translation theory were voiced. Its long journey from a normative discipline to an acknowledged academic discipline may partly be due to its rigid concentration on ‘language teaching and learning’ (Munday 2001:8). Researchers (or in this case often teachers of translation and interpreting) started to recognize the influence of lexical and semantic transfer on the processes and the products of translation. Scientific fields like comparative literature and contrastive analysis (Munday 2001:8-9) began to investigate differences and similarities between source and target texts in order to be able to make statements about semantic and linguistic differences between languages. However, translated texts were widely treated as inferior texts and ‘the study of translation … [was] relegated to the periphery of other disciplines and sub-disciplines’ (Baker 1993:234). Interestingly, Baker also points to an important issue with this demotion of translated texts by stating that ‘this traditional view of translation implies, in itself, an acknowledgement of the fact that translational behaviour is different from other types of linguistic behaviour’ (1993:234). This view is also reflected in the changing focus within research in TS from word-to-word translations to research including the context of the translation situation like source and target culture, the employment situation of the translator and who pays for the translation. The translator assumes not only the role of transferring one language into another but transferring that language from one culture into another. Sense, not only in terms of lexical semantic sense but also in terms of the intended communicative aim of the ST (including author intention), comes to the fore. This also amplifies the notion of the distinction between source and target text as two independent texts with only a semantic connection instead of the assumption that the target text (TT) is a sole copy of the source text (ST) without any individual existence. Theoretical approaches within TS turn towards the fundamental question of what meaning is, how it is incorporated into sentence structure and what consequences this has for both the translational process and the finished product – the target text.

Today, James S. Holmes’s paper *The name and nature of translation studies* from 1988 is widely referred to as the origin of TS as an academic research discipline. Holmes outlines the single components he considers to constitute translation studies and specifies the overall objectives of the field as descriptive and theory building (Munday 2001:11). The latter deals with ‘the establishment of general principles to explain and predict’(2001:11) and the former with the description of target texts, their placement and function within the target culture and last but not least the translational process. Descriptive translation studies (DTS) constitute the empirical means to develop the theoretical framework necessary for the discipline to evolve and progress. Holmes delimits TS as an independent research discipline from other disciplines which up to that point had determined and influenced the research within the field and the main focus shifts from pure translation training to theoretical investigations and reasoning. In other words, 'TS has changed its point of focus from writing manuals on translation to empirical studies of translation situations and processes' (Hegrenæs 2010:4).

In 1995, Gideon Toury 'considers translation to be an activity governed by norms' (Munday 2001:113) which 'determine the (type and extent of) equivalence manifested in actual translations' (Toury, 1995:61). To oversimplify, translating[3] is the process of rendering text from one language into another. The translational process results in text which has to be distinguished from the original text on a variety of levels (Baker 1993: 244-245). In TS, the original text is referred to as source text (ST) and the translated text as target text (TT). Similarly, the language of origin is called source language (SL) and the language to be translated into is the target language (TL). Since translations are no longer carried out on a word-to-word equivalence basis, they do 'not simply … reproduce the formal structures of the source text but also give some thought, and sometimes priority, to how similar meanings and functions are typically expressed in the target language' (Baker 1993:236). Thus, the target text becomes detached from the source text not only in terms of lexis (choice of lexical expressions), syntax (sentence structure) and semantics (meaning) but also as a result of these three by disguising the source culture. Baker states that '[t]he source text is a source of information and … it may be exploited in a variety of ways to meet the expectations of an envisaged audience' (Baker 1993:239) – the audience in the target culture. However, with its own, independent status one would expect the TT to blend into the target culture by imitating cultural, social as well as textual norms. Quite the opposite, Baker claims that 'the need to communicate in translated utterances, operates as a major constraint on translational behavior and gives rise to patterns which are specific to translated texts' (1993:242). According to Baker, these patterns, which are highly specific to translated texts, originate neither in the SL nor in the TL but constitute a new variety of the target language, namely translated language. Baker proposes a number of translation universals on the basis of these specific patterns: 1) an increase in explicitness, 2) disambiguation and simplification and 3) the use of more conventional grammatical forms of the TL (1993: 243-244). She also claims that target texts have '[a] general tendency to exaggerate features of the target language' and that they are marked with 'a specific

[3] The use of the term translating in this chapter includes both written translation and oral interpreting.

type of distribution of certain features in translated texts vis-à-vis source texts and original texts in the target language' (Baker 1993:244-245). Hence, translated texts distinguish themselves from target language texts and it is possible to identify them on the basis of these translational universals. Baker also claims that these distinguishing patterns can be found in the complete translated variety of a language irrespective of the source language and that they differ qualitatively as well as quantitatively from the non-translated variety of that same language (Baker 1993:245). Hence, even though translation is assumed to be profoundly target text and target culture orientated and Munday (2001) refers to the invisibility of translation and the translator (144) to the target text reader, Baker claims that there are significant differences between translated and original texts of one language on a number of levels (lexis, syntax, semantic etc.). Also Gideon Toury (1995) reports on this oppositional behavior of translated texts and claims that translation universals operate on several different levels of the translation process (58). Toury also introduces a methodology for descriptive translation studies (DTS) aiming to develop a general theory of translation studies (1995:70). He suggests text comparison consisting of a three-step model which first situates the text within the larger framework of the culture it is translated into (target culture), secondly compares source and target texts for any possible shifts and thirdly generalizes into theory (or rules) for further translations (ibid.:11).

2.2 Cognitive linguistics

Cognitive linguistics deals with the interrelation between language and the mind. This entails how humans perceive the world, structure reality and interact with and within it. Researchers within cognitive linguistics argue that language is the means with which one both expresses cognitive perceptions of the world as represented in the mind and, in return, shapes new perceptual forms within the mind. Thus, 'language, as representations in the mind and the product of cognitive events, reflects the interaction of cultural, psychological, communicative and functional considerations' (Luchjenbroers, 2006:2). This includes translated language as a variety of any language in question. One of the semantic theories dominating research within the field of cognitive linguistics for the last three decades is the phenomenon of cognitive metaphor.

Until the publication of George Lakoff and Mark Johnson's book *Metaphors We Live By* in 1980, metaphor was considered to be a mere literary feature only to be found in poetry and fiction. Aristotle in his work on literary theory in *Poetics* describes metaphor as 'the application of an alien name by transference either from genus to species, or from species to genus, or from species to species, or by analogy, that is, proportion' (Stanford Encyclopedia of Philosophy)[4]. He was thus the first to assign metaphor the functional purpose of understanding one thing (e.g. species) in terms of another (e.g. genus), though restricting it to literary genres and figurative language. This view was revolutionized by the ideas put forward by Lakoff and Johnson in 1980. Metaphor, they observe, is part of every person's everyday language use and '[o]ur ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature' (Lakoff and Johnson 2003:3). Going back to the central agenda of cognitive linguistics, metaphor, in this view, contributes to perceiving, understanding and structuring reality as well as to communicating about this reality. Metaphor is thus not solely a feature of literary discourse but of every kind of discourse in every possible language there is.

The theory has undergone substantial development and change since its introduction in 1980. The fundamental assumption of CMT that 'the human thought process is regarded to be structured into different metaphorical concepts that, on the one hand, influence our perception of reality and, on the other hand, our linguistic performance' (Hegrenæs 2010:7) is still valid today. Thus, the affiliation to cognitive linguistics becomes even more evident and Lakoff states that 'the locus of metaphor is not in language at all, but in the way we conceptualize one mental domain in terms of another' (Lakoff 1993:203). The latter points towards the basic cognitive principle of CMT, which refers to parts of one conceptual domain, i.e. the source domain denoting parts of another distinct domain, the target domain. Lakoff calls this principle cross-domain mapping. A conceptual domain like TIME contains every conceptual characteristic item (and thus its lexical equivalent) that one connects with TIME, e.g. DAY, MONTH, AGE etc. Such a conceptual inventory of a domain rests on empirical experience, i.e. the experiences one personally has made with TIME including days, months and age on all levels of cognition (e.g. hearing, reading, seeing, touching etc.). Hence, metaphors arise from one's own experience, which does not necessarily mean that people speaking the same language or people speaking different languages share the same cognitive mappings just because they experience the same things through the same cognitive channels.

The cognitive accessibility of domain knowledge in language varies from domain to domain. Abstract domains like TIME, STATES OF BEING or ARGUMENT for example are less

---

[4] http://plato.stanford.edu/entries/aristotle-rhetoric/#8.2

understandable than MOTION, which can be observed and possibly executed on an everyday basis. So how are people able to understand those abstract concepts? According to Lakoff and Johnson, the answer is as simple as this: cross-domain mapping. In 1993, Lakoff states that '[m]etaphor is the main mechanism through which we comprehend abstract concepts and perform abstract reasoning' (1993:244). By taking members of the experiential domain, e.g. MOTION, and mapping them onto the abstract domain, e.g. TIME, people use the known (source domain) to make sense out of the unknown (target domain). In other words, we use the knowledge of the source domain MOTION to reason about the target domain TIME (Lakoff 1993:207). Conceptual metaphors or mappings 'have the form: TARGET-DOMAIN IS SOURCE-DOMAIN, or alternatively, TARGET-DOMAIN AS SOURCE-DOMAIN' (Lakoff 1993:207).

The evidence drawn on to support Lakoff and Johnson's claims about cognitive metaphors consists of linguistic expressions[5] like in the following example:

ARGUMENT IS WAR

Your claims are *indefensible*.
He *attacked every weak point* in my argument.
His criticisms were *right on target*.
I *demolished* his argument.
I've never *won* an argument with him.
You disagree? Okay, *shoot*!
If you use that *strategy*, he'll *wipe you out*.
He *shot down* all of my arguments.

(Lakoff and Johnson 2003:4, their emphasis)

Most of the example sentences above include lexical elements from both source and target domain like in '[h]e *attacked every weak point* in my argument' where the phrase in italics belongs to the source domain WAR and 'my argument' to the target domain ARGUMENT. Stefanowitsch (2006) calls such expressions metaphorical patterns where 'metaphorical expressions contain both source and target domain lexemes' (2006:66). This does not necessarily apply for every metaphorical expression. Sentences like 'If you use that *strategy*, he'll *wipe you out*' from the sample above only contain lexical items from the source domain. Lakoff and Johnson have been extensively

[5] Linguistic expressions are also called metaphorical expressions since they contain lexical elements representing source and/or target domain and thus exemplify the mapping process.

criticized for their use of example sentences because these sentences are constructed and not based on real life language use. Granted, they make sense to a speaker of the English language. But are these sentences (and thus the respective cognitive metaphors) really used recurrently in actual daily English language use? If conceptual metaphors are supposed to be ‘absolutely central to ordinary natural language semantics’ (Lakoff 1993:203), then they have to rest upon naturally occurring language use. This point has recently been raised about the examples used not only by Lakoff and Johnson but also by Zoltán Kövecses (1986). Cognitive linguistic research within the relatively newly established field of corpus linguistics tries to solve this problem by developing methods to retrieve metaphorical expressions from electronic corpora and prove and refine existing conceptual metaphors as well as establish new ones based on naturally occurring discourse. It has been the goal of many scholars in recent years to find proof for the conceptual metaphors proposed by Lakoff, Johnson, Kövecses and others in the earlier years of the theory. I shall not go into more detail here but corpus based-studies (see for example Musolff 2004) have uncovered a wide variety of metaphorical concepts and have thus strengthened Lakoff and Johnson’s assumption that metaphor is an intrinsic part of the way we perceive and handle reality and of how language is structured. Metaphor studies have moved from simply identifying metaphorical expressions and mappings to larger-scale context and discourse studies aiming to answer questions about the communicative aims of cognitive metaphors (see Semino 2008). It is no longer a question of proving whether metaphor is linguistically prevalent or not but that metaphor is a matter of cognition and not language. Thus, language is the means through which metaphor is approached empirically and not its prevailing place of origin. The theory currently strives to prove that, ultimately, metaphorical thinking is the dominant way of all mental activity related to understanding and that it goes beyond language to the mind and, as Lakoff is arguing, even down onto the neurological/physiological level (Lakoff 2008).

2.3 Translation and cognition

As mentioned earlier (section 2.1), the translational process is not simply the transition of a word, a sentence or a whole text (spoken or written) from one language into another. The source text input undergoes a complex and diverse cognitive process before it gets recreated in the TT. The basic data for empirical studies within TS consists of pre-existing discourse which is rendered into another language but is still based on the original discourse (Neubert 1997:2). This makes translation unique to language use. Neubert claims that all sorts of human communication (speaking, listening, reading, writing) are profoundly altered when ‘occur[ing] in the unique languaging context of translation and interpreting’ (1997:2). The main task for TS is thus to identify and classify these changes according to causation and impact on the translational process and product. Regarding cognitive metaphors, the key to the cognitive level is language and to investigate changes within the usage of lexical expressions of cognitive metaphors between source and target texts one has to study translated language.

Also within translation studies, metaphor has undergone a significant change of relevance since 1980. Regarding the establishment and teaching of translational methods, metaphor was treated as a literary device ‘for the purpose of colouring language’ (Newmark 1983:4). Four years prior to the publication of Lakoff and Johnson’s cognitive view, M.B. Dagut stated that metaphor ‘is a phenomenon which is … central to all forms of language use’ and demands that ‘it is high time for translation theory to make a start on a thorough and systematic discussion of the translation implications of “metaphor”’ (1976:21). However, in contrast to CMT and even though Dagut regards metaphor as central to all language forms, he does not understand it to be a part of everyday language use. He rather points to the originality and uniqueness of every single metaphorical verbalization and assumes that ‘every metaphor is an entirely new and unique creation’ which ‘is unpredictable and irreducible to “rules”’ (1976:23).

In recent years, research on metaphor translation within TS has included cognitive metaphors and the question of universal vs. culture-specific metaphors has attracted even more attention. With the introduction of CMT and the scientific development within the theory over the years, the question of translatability ‘becomes linked to the level of conceptual systems in source and target culture’ (Schäffner 2004:1258). Al-Ali and Al-Hasnawi (2006) presuppose that ‘culture influences metaphor in an important way’ and that the cultural influence on the cognition of reality also impacts the linguistic representation of this reality (2006:231). New to this approach to CMT within TS is that cognition as the main contributor to conceptual metaphors is seen to be dependent on culture. Thus, one’s cognitive perception of reality is not unconditionally pure but ‘filtered by the value and belief systems prevailing in the cultural community’ (2006:231). This, obviously, has a great impact on translations from one language into another, from one culture into another. Since perception and understanding of reality differs to various degrees from culture to culture, the translator has to be aware of these differences. However, Al-Ali and Al-Hasnawi also account for culture-overlapping metaphors and refer to the cognitive state of two languages that use similar metaphorical concepts as 'cognitive equivalence' (2006:233).

If, then, the translational process regarding cognitive metaphor is guided by an understanding of culture-specific and culture-overlapping metaphors and the aim of the translational process is to assimilate the TT into the target culture, do translated texts exhibit patterns that differentiate themselves from non-translated texts in the use of cognitive metaphors? Do Baker's (1993) translation universals (e.g. exaggeration of target language features, quantitative differences to target language texts etc., see also section 2.1) apply to the use of metaphorical expressions of cognitive metaphors in translated English texts?

2.4 The positioning of the study in the field of research

Theoretically, the study is motivated by Toury's law of growing standardization and Mona Baker's translation universal on the exaggeration of target text features. The law of growing standardization refers to a high degree of adaptation of the TT to the target culture (Toury 1995:267). Baker proposes so-called comparable corpus studies (i.e. studies of translated and non-translated texts in the same language) to investigate translational universals (1993:237). This paper is taking advantage of this development in as far as several different corpora for the English language are used to extract data for the analysis.

Presupposing the independence of the target text from both the source text and other texts in the target language implies both assimilation into the target culture as well as demarcation from it. On the one hand, target texts aim to fit into the target culture by adapting lexical, semantic as well as cultural norms. On the other hand, the transmission of text from one language into another generates lexical and semantic patterns that are typical for neither the target language nor the source language. Baker claims that these patterns are quantitatively (and qualitatively) detectable by comparing the translated and the non-translated variety of English (1993:245). In this study, lexical expressions of cognitive metaphors are such linguistic patterns that are assumed to deviate between translated and original English. Hence, this study tries to bring together empirical studies on the nature of translations and cognitive studies on the nature of human conceptualization by aiming to establish relationships between cognitive metaphorical expressions on the one hand and the translated and non-translated variety of English on the other.

There have been significant changes within the agenda of translation studies: from sole translation teaching theory to overarching research including translational context and processing and the interdisciplinary incorporation into other academic disciplines like linguistics and philosophy. The present study is an investigation to be situated within descriptive translation studies. Texts translated into English (target texts) as well as original English texts are investigated regarding the quantitative use of lexical expressions of the cognitive metaphor TIME IS MOTION and described in detail with the aim of forming the basis for theoretical accounts on the nature of cognitive metaphors in translation and translations in general.

Halverson (2003) proposes to employ Croft's (2003) methodological approach to language to place empirical studies into the research space of translation studies. Croft states that research within linguistics can lead to generalizations on three different levels: 1) the level of observations, 2) levels of internal generalizations and 3) the level of external generalization 'at which the linguist invokes concepts from psychology, biology and other realms outside the structure of language' (2003:285). Transferring this approach to my study in particular, the investigation into quantitative deviations of lexical (and thus cognitive) representations between translated and non-translated English texts can be placed on the first level which is the level that 'constitutes the basic facts of language' (ibid.: 285). Corpus studies on cognitive metaphors in translation of the type carried out by Christina Schäffner (2004) on political discourse in English and German or by Al-Hasnawi (2007) in English and Arabic can be situated on the first and second level of generalization where the description of translation on the first level leads to internal generalizations on translational strategies (Schäffner) or the existence of culture-overlapping and culture-specific cognitive metaphorical mappings (Al-Hasnawi). Schäffner analyzes different English translation of one and the same German political metaphor, thus aiming to reveal different translation strategies. Al-Hasnawi's analysis of Arabic translations of English metaphorical expressions intends to establish similar mapping conditions (i.e. translation equivalence) on the basis of similar conceptualization of reality within two cultures. Both studies are performed on language pairs (English/German and English/Arabic), thus approaching cognitive metaphors in translation from both the source and the target text/language. In contrast, the study at hand aims to describe quantitative differences within the two varieties of the target language, obscuring the semantic meaning of the single tokens and their source language by including various source languages. No qualitative investigation of the tokens leading into internal generalizations of translation studies or even generalizations into cognitive linguistics is intended.

Olohan (2004) includes a quantitative corpus study of optional syntactic elements in translation starting from the hypothesis that 'translation will render grammatical relations more explicit more often' (2004:104). The study is intended to test Baker's translation universal of explication and exaggeration and concludes by determining that 'there is a clear tendency for the TEC sub-corpus to exhibit higher frequencies of more explicit syntactic constructions' (ibid.107). This conclusion corresponds to my hypothesis about quantitative overrepresentation of metaphorical expressions of TIME IS MOTION in translated English. However, Olohan studied a syntactic linguistic phenomenon while I investigate a cognitive linguistic feature.

Summing up, this study intends to connect empirical investigations into the nature of translations with cognitive linguistics. Previous studies either explored qualitative equivalence and diversity between source and target languages or quantitative equivalences and differences of non-cognitive linguistic phenomena (e.g. syntactic phenomena) between the translated and the non-translated variety of a single target language. However, I aim to study quantitative deviation of a cognitive linguistic phenomenon between the translated and the non-translated variety of English.

3 METHODOLOGY

Investigating cognitive phenomena through actual language use is a relatively new approach to cognitive linguistics and demands careful preparation and a proper methodology. Collecting authentic language data to investigate linguistic phenomena is what Tummers et al. (2005) call ‘usage-based linguistics’ and they elaborate further that ‘corpus linguistics would be an obvious methodology for a usage-based linguistics’ (2005:226). In 2009, Gilquin and Gries conducted an analysis of 81 papers on corpus studies published in three journals on corpus linguistics between 2005 and 2009 and concluded the following:

> [T]he majority of the papers deal with lexis (60) […] especially phraseological issues (collocations, idioms, semantic prosody, etc.). 41% of the papers analyze a syntactic phenomenon. In comparison, morphology, pragmatics, and phonology represent a small proportion of the topics investigated (about 7% each). (2009:10).

There are no studies on cognitive linguistic phenomena. For the present study, I decided on corpus studies as a methodological approach to the question of whether and, if so, how usages of metaphorical expressions of the cognitive metaphor TIME IS MOTION differ between original and translated English. The following chapter first explains this choice of method (in contrast to other methods) before introducing the three corpora used in this study and describing how these corpora were approached from a cognitive metaphorical point of view. The third and fourth sections describe and explain in detail how I went about collecting data from those corpora and how I dealt with the data afterwards in terms of analysis and statistical significance testing.

3.1 Why corpus studies? A usage-based approach to linguistics

Corpora in general and the three corpora used for this project in particular have come a long way. Due to technological development and invention, corpora can easily be compiled and processed with basic computer programs like MS Word©. Through ‘machine-readability, authenticity and representativeness’ (McEnery and Wilson 2001:5), corpora facilitate the scientific investigation of language on a big scale. Large, diversified corpora are available online and are highly suited to investigating any kind of linguistic phenomena in general and cognitive metaphorical language usage in particular because they contain a huge amount of ‘naturally occurring language’

(McEnery 2006:4). Thus, corpora contain language ‘in its most natural form … in the shape of spontaneous, non-elicited language data’ (Tummers et al. 2005:226). Some researchers argue that written corpus data like newspaper and magazine articles but also fictional texts are not as non-elicited as claimed because they are written with a particular intention, contain a certain register (according to the text type or the language variety etc.) or are deliberately edited for publishing (Gilquin and Gries 2009:7). For the purposes of this paper, I consider the written discourse contained in the corpora as authentic and naturally occurring language use. Regarding the results of this study, I acknowledge that the data might be biased by register, variety or other less linguistically motivated procedures. However, in comparison to the introspective intuition-based method used by Lakoff and Johnson in 1980, corpus studies are the more reliable method to investigate metaphorical language use in original and translated English.

Corpora have the advantage of providing users with authentic language use not only in single examples (one or two) but multiple (hundreds or thousands according to the corpus) to generate quantifiable data. The scientific discipline of ‘[c]orpus linguistics should be considered as a methodology with a wide range of applications across many areas and theories of linguistics’ (McEnery 2006:8). One of the main issues of corpus studies and corpora in general is representativeness. To be representative of a certain language or language variety, a corpus has to be compiled of a balanced sample of that language to exemplify language use that is true not only for the sample in question but the whole language or variety. Samples (reduced versions of the whole language) are balanced if they include an adequate variety of text types to fulfill the purpose of the corpus. General corpora like the BNC and the COCA are assumed to be balanced because they contain written as well as spoken language data from different genres and different time periods. Both corpora ‘typically serve as a basis for an overall description of a language or language variety’ (McEnery 2006:15). In contrast, specialized corpora contain language that is typical for a certain variety or a certain text type or concentrate exclusively on written or spoken language. An example of a specialized corpus is the COLT (The Bergen Corpus of London Teenage English)[6], which contains only transcripts of speech of teenagers in London in 1993. The requested balance criteria for COLT are of necessity different than for the BNC or the COCA since they are supposed to serve different purposes. While the latter two give a general overview

[6] http://www.hd.uib.no/colt/

of British and American English in particular time periods, the COLT only represents youth language from London in 1993. Balancing a corpus is also about evening out the proportions of the different text types included. Thus, the ratio of written and spoken texts or the ratio between genres like newspaper, magazine or fiction has to be evened out in order to represent the language use as adequately as possible. This applies primarily to general corpora.

With regard to translation studies, corpora can be used ‘to study the translation process by exploring how an idea in one language is conveyed in another language’ (McEnery 2006:91). This is done by employing multilingual parallel corpora which automatically provide the user with an SL expression and its translations into one or several target languages. Even though this is a very suitable way of investigating translational questions, it might not serve all research purposes equally well for two reasons: 1) the majority of texts have only been translated once and answering one’s hypothesis on the basis of one translation decreases reliability and generalizability considerably and 2) one might not want to restrict one’s research to a single or several target languages but to a translated variety of one or several source languages as well.

Undeniably, corpora as the basis of language studies have a series of advantages as well as disadvantages. Since they represent language produced by actual language users they are doubtlessly more suitable to answer language-related questions or exemplify language use than introspective examples do. In contrast, it has to be kept in mind that findings from corpora are ultimately only true for the very sample they are taken from. Generalizations to the entire language or language variety which is supposed to be represented by the sample have to be done carefully and are dependent on a series of preconditions the corpus has to fulfill (e.g. representativeness, balance etc.) to be a reliable source.

There are several different methods to investigate cognitive linguistic phenomena within translation studies. There is the introspective method used by Lakoff and Johnson, which produces examples by the goodness-of-fit principle without considering actual language use. Then there are surveys (e.g. questionnaires) and experiments (e.g. eye tracking, keystroke-logging)[7] and there are corpus studies (Tummers et al. 2005:229). For this paper, I decided on corpus studies for several reasons. Firstly, the requirements for this paper in terms of length and time did not allow for more demanding research methods and secondly with the BNC and the COCA I had two large general

[7] For more information on keystroke-logging and eye tracking see Jakobsen 2006.

corpora available that are assumed to be representative of their respective language variety (British and American English) to be illustrative of what I call original English in this paper. In many aspects, corpora were my only choice to generate a sufficient amount of data to answer my research questions. Availability, accessibility, electronic processability and representativeness were major factors determining my choice for this method. However, choosing corpus studies for this study does not automatically exclude the other three research methods. Indeed, a combination of two or more different methodological approaches might complement the results and support (or even refute) them, thus strengthening the scientific approach. Hence, the results of the corpus studies might evoke further questions which are possibly best answered by the use of surveys or certain experimental tasks. But, as mentioned before, within the scope of this paper I am only able to conduct a corpus study.

Using corpora to generate a data sample to answer the research questions (and eventually the hypothesis) entails deciding on how to use the data in the course of the study. Corpus-illustrated approaches supplement the study with examples found in the corpus search whereas corpus-based approaches take the data set as a whole to identify and reveal language functions and tendencies (Tummers et al. 2005:234-235). The latter approach can involve quantitative frequency counts and statistical analysis which:

> do not only synthesize the data, but also allow to uncover the significant tendencies underlying the data of actual language use. Furthermore, the use of statistical techniques allows one to tackle research questions that go beyond the analytical scope of traditional introspective linguistics (Tummers et al. 2005:236).

Tummers et al. propose a quantitative corpus-based approach involving two stages:
(1) a descriptive or exploratory stage and (2) an explanatory or hypothesis-testing stage (2005:238). The first stage is supposed to identify and describe the linguistics phenomenon in question while the second stage is used to investigate the behavior of different types of data within the sample individually and mutually. Hence, simple frequency counts conducted in stage one are not sufficient to answer a hypothesis until they are related to each other and subject to various statistical tests to determine their scientific validity.

Summing up, for reasons of availability, restrictions on time and space of the paper and last but not least on the basis of the research questions (i.e. over- and underrepresentation of translated or non-translated language) I decided on corpus studies as the most suitable methodological approach to the hypothesis. Furthermore, the quantitative approach to the data set leads to a corpus-based approach where I count occurrences of certain units of language, describe the data set and conduct statistical tests to determine the empirical relevance of the relations I identify. Having clarified the corpus-based approach to linguistic studies, the next section deals with the cognitive linguistic exploration of corpora.

3.2 The cognitive approach to corpora

Approaching linguistic subjects through corpora and actual language use has become a common practice within many linguistic disciplines (see Gilquin and Gries 2009). But what about cognitive linguistic problems? How can corpora be exploited throughout the level of concrete language use (i.e. syntactical, morphological, semantic and phonological issues)? Andreas Musolff (2004) states that 'metaphors cannot be identified by external features, because they do not belong to the "expression" side of linguistic signs but to their conceptual side' (2004:8). Investigating cognitive metaphors means investigating their linguistic manifestation in actual language use. Thus, one has to distinguish 'between "underlying" metaphorical concepts (domain mappings) and linguistic "surface" text features' (Musolff 2004:8). The latter refers to actual language use as represented in corpora whereas the first refers to the conceptual level behind those expressions. All scientific description and statistical evaluation of corpus data is first and foremost valid for the empirical data drawn from corpora and not for the conceptual level. However, empirical data constitute the basis for the conceptual level and 'any claims about specific metaphorical concepts "underlying", "informing" or "organizing" the discourse and thinking of larger social groups need to be related to empirical discourse data before any significant conclusions can be drawn'(Musolff 2004:9).

Language data extracted from corpora are called tokens (e.g. words, phrases, sentences etc.). Within cognitive linguistic investigations on metaphor employing corpus studies, tokens represent individual manifestations (metaphorical expressions) of cognitive metaphorical mappings (e.g. TIME IS MOTION). Tokens are generated by a particular search string carried out according to the search premises of the specific corpus. To extract tokens of metaphorical expressions for a certain conceptual metaphor from a corpus, the underlying requirement is to assign the particular conceptual mapping certain lexical items that are searchable within the corpus. Semantically, the cognitive metaphorical mapping TIME IS MOTION can presumably be realized in a number of different ways with an even higher number of different nouns and verbs. Musolff proposes the following procedure:

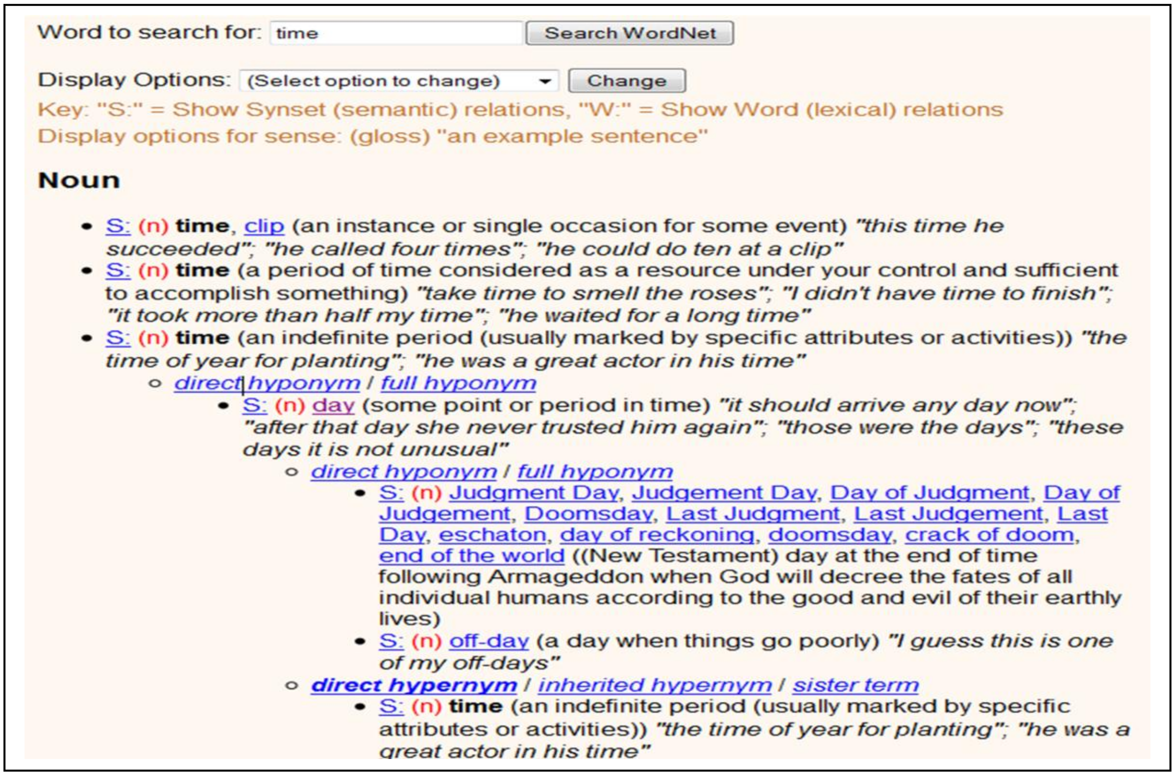
> [grouping] the conceptual elements into source domains by using lexical fields (as exemplified in standard thesaurus categories) as well as patterns of collocation and relative frequencies in the emerging corpus (Musolff 2004:11).

I decided to employ this procedure as a model for my own cognitive approach to corpora. The exact process is described below.

In order to determine lexical items which express the given conceptual metaphor and which can be used as keywords in the corpora search, I started with the existing literature (Master Metaphor List:76[8]; Lakoff 1993:217) as well as my own knowledge and understanding of the English language in general and the metaphor in particular. Following Stefanowitsch's definition for metaphorical patterns (see section 2.2), the metaphor was divided into its two constituting parts: the source domain (TIME) and the target domain (MOTION). To achieve maximum comparability between the corpora and a manageable search string as well as a manageable set of tokens, I decided to exclusively use verbs of motion as lexical items representing the source domain. The target domain TIME consists out of necessity only of nouns. Starting with the target domain, I compiled a list of 19 nouns based on my own understanding of the domain TIME (appendix A). This includes nouns referring to time periods, references to points in time as well as units of time. I acknowledge that this approach is based on my subjective knowledge and is open to criticism but I had to start at some point. Since there has, to my knowledge, not been a project like this before, ready-made lists of lexical items do not exist. The 19 nouns were then entered into the online database WordNet[9] as exemplified below:

[8] http://araw.mede.uic.edu/~alansz/metaphor/METAPHORLIST.pdf
[9] http://wordnet.princeton.edu/

Table 3.1: Extracting lexical items for the source domain TIME as nouns



Word to search for: time [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **time**, clip (an instance or single occasion for some event) *"this time he succeeded"; "he called four times"; "he could do ten at a clip"*
- S: (n) **time** (a period of time considered as a resource under your control and sufficient to accomplish something) *"take time to smell the roses"; "I didn't have time to finish"; "it took more than half my time"; "he waited for a long time"*
- S: (n) **time** (an indefinite period (usually marked by specific attributes or activities)) *"the time of year for planting"; "he was a great actor in his time"*
  - *direct hyponym* / *full hyponym*
    - S: (n) day (some point or period in time) *"it should arrive any day now"; "after that day she never trusted him again"; "those were the days"; "these days it is not unusual"*
      - *direct hyponym* / *full hyponym*
        - S: (n) Judgment Day, Judgement Day, Day of Judgment, Day of Judgement, Doomsday, Last Judgment, Last Judgement, Last Day, eschaton, day of reckoning, doomsday, crack of doom, end of the world ((New Testament) day at the end of time following Armageddon when God will decree the fates of all individual humans according to the good and evil of their earthly lives)
        - S: (n) off-day (a day when things go poorly) *"I guess this is one of my off-days"*
      - ***direct hypernym*** / *inherited hypernym* / *sister term*
        - S: (n) **time** (an indefinite period (usually marked by specific attributes or activities)) *"the time of year for planting"; "he was a great actor in his time"*

Every entry for every noun was examined individually for hyponyms[10], which in turn were chosen if their respective hypernyms[11] were either defined as *time*, *time unit*, *unit of time*, *time period* or *time interval*. In the example above, *time* was entered as the search word. The semantic relation menu (S) was opened for every entry and the option for full hyponym chosen. The first two entries for *time* did not display any hyponyms. The third entry for *time* lists amongst others *day* as a hyponym of time. The direct hypernym of *day* refers back to *time*. Using this procedure for all the 19 nouns from the original list, the list was extended to 51 nouns. The seven days of the week and the 12 months were added, bringing the total to 70 lexical items. The list was arranged alphabetically and supplied with the semantic definitions (according to WordNet) for every single noun and its relation to the domain TIME (appendix B[12]). Some of them include several different semantic definitions and relations to the domain TIME which is given in appendix B but does not

[10] A hyponym is ‘a word of more specific meaning than a general or superordinate term applicable to it’ (http://oxforddictionaries.com/definition/hyponym 16.8.2011).

[11] A hypernym is ‘a word with a broad meaning constituting a category into which words with more specific meanings fall; a superordinate’ (http://oxforddictionaries.com/definition/hypernym 16.8.2011).

[12] The list in appendix B does not include the days of the week and months since they are self-explanatory.

affect the usage of the lexical item in the corpus search. The list could have exceeded 70 nouns easily, but I decided to exclude specific compound terms like *work time* and all the holidays like *Easter* and *Thanksgiving* because they did not refer to any direct TIME relation (e.g. unit of time, time period etc.) in WordNet. It was necessary to restrict the list to a manageable number of lexical items. Also the weekdays and the months do not refer to a specific relation to the domain TIME in WordNet, but I included them because I expected them to be quite frequent in English language use and thus to produce a substantial number of tokens in the corpus search.

In a next step, it was necessary to rank the nouns to decide which are useful in a corpus search and which will most likely not produce any useful tokens. I decided that it was most beneficial to sort them by their frequency according to the two corpora that are representing original English in this study, the COCA and the BNC. Frequency lists for these corpora are available either online (COCA) or in print (BNC). For the COCA, I downloaded the newly published list of the top 500,000 words (2011)[13], converted it into a Word© format and used the search function in Word© to find all the 70 nouns and their respective frequencies. The list was not lemmatized[14], so that I had to search for the different grammatical forms of the nouns (singular and plural) separately. However, the nouns were listed with their respective part of speech (PoS) tags, which made it easier to find the correct entries. This way, I avoided including frequencies for the entry *times* as a general adjective (PoS jj)[15] or the third person form of the verb *time* (PoS vvz). Although the list is from 2011 and thus quite new, it was compiled regarding a total number of 410 million words in the corpus. The number has since changed and claims to have 'more than 425 million words' (Davies 2011). However, the 500_K list is based on a total amount of 410 million words. The final list of TIME nouns from the COCA ranked by frequency is given in appendix C. The frequency numbers for the BNC are taken from Leech, Rayson and Wilson (2001) and searched for manually in the *Alphabetical frequency list for the whole corpus (lemmatized)* (Leech 2001: 25-119). Since the BNC was completed in 1993 and is no longer maintained, this frequency list is the only relevant one. The respective ranked list for the BNC is given in appendix D. Unfortunately, the frequency list for the TEC is not tagged for any part of speech, which makes it

[13] http://www.wordfrequency.info/500k_words.asp
[14] Lemmatization is the summarization of all inflected forms of a word into one lexical item as for example in dictionaries (McEnery 2006:35).
[15] I realize that there is no entry of *times* as an adjective in any version of the OED. However, the frequency list for the COCA includes this entry and it is the frequency count for this entry I did not include in my count.

impossible to determine if, for example, the entry for *time* is a noun or a verb. For this reason, I decided to disregard the frequency information from the TEC since British and American English (i.e. the data from the BNC and the COCA) are assumed to represent the standard use of cognitive linguistic expressions for the mapping TIME IS MOTION in naturally occurring English language and the usage within translated English is supposed to be compared with this. It also became clear that polysemous nouns from the lexical item list (appendix B) like *term* ('a limited period of time' and 'a word or expression') and *present* ('the period of time that is happening now' and 'something presented as a gift') needed to be eliminated, since it is not possible to distinguish the different meanings in the frequency lists of the COCA and the BNC. Thus, polysemous terms (*period*, *second*, *term* and *present*) were disregarded. The raw frequencies from both corpora were added up and normalized frequencies per ten million words were generated for every single noun for the total number of 510 million words contained in both corpora. The nouns were then ranked from highest to lowest. The top two month nouns (May and March) and days of the week (Sunday and Friday) were taken as representatives for their category. The final ranked list of 48 nouns is given in appendix E.

The list of searchable lexical items for the domain MOTION originates in Beth Levin's work on *English Verb Classes and Alternations* (1993) which is an 'investigation of the syntactic and semantic properties of English verbs' under 'the assumption that the behavior of a verb, particularly with respect to the expression and interpretation of its argument, is to a large extent determined by its meaning' (Levin 1993:1). Levin claims that speakers of English (and other languages as well) have an innate ability to judge a verb's syntactic behavior (i.e. its grammatical behavior in association with other lexical units) on the basis of its meaning (ibid.5). Further, Levin proposes classifying verbs into semantically coherent classes by their diathesis alternations, i.e. by their usage in either active or passive voice in relation to their argument:

> Distinctions induced by diathesis alternations help to provide insight into verb meaning, and more generally into the organization of the English verb lexicon, that might not otherwise be apparent … (1993:15).

Levin exemplifies her claims by referring to verbs of motion as a large verb class in English. According to earlier studies by her and others, not all members of this class behave similarly regarding voice and hence the class has to be divided into subclasses like verbs of inherently directed motion and verbs of manner of motion (1993:15). This way, Levin aims to distinguish between semantically (closely) related verbs by investigating the association between the action or state the verb denotes and its argument. This in turn is supposed to reveal information about the syntactical behavior of the verbs in lexical constructions. Hence, Levin approaches a syntactical problem through semantic properties. She lists verbs of motion in seven different categories: 1) verbs of inherently direct motion, 2) *leave* verbs, 3) manner of motion verbs, 4) verbs of motion using a vehicle, 5) w*altz* verbs, 6) *chase* verbs and *accompany* verbs. Manner of motion verbs are further distinguished into *Roll* verbs and *Run* verbs. I incorporated all the verbs Levin categorizes as verbs of inherently directed motion (1993:263) plus all *run* verbs (1993:265) for three reasons. Firstly, they are fairly large categories containing a substantial number of verbs. Secondly, I assume the verbs in these two categories to be basic verbs (in contrast to more specific verbs like *chase* and *waltz* verbs). Thus, I expect these verbs to be highly frequent in English. Thirdly, metaphorization is assumed to be more likely with basic lexical items than with highly specific ones. This way I received a total of 133 verbs. However, it has to be pointed out that I am not interested in Levin's syntactic classification of the verbs. The decision to extract verbs of motion from her book was a clear methodological one and is not grounded in any kind of theoretical consideration concerning her classification of the verbs into verbs of inherently directed motion and *run verbs*. For the purpose of this study, they are plainly verbs of motion – a verb classification based on semantics. I acknowledge that there are other ways to compile a body of motion verbs. However, for reasons of time limitation I decided on this convenient method. I performed the same operations on the frequency lists for COCA and BNC for the verbs as described above: 1) the verbs were searched for in the frequency lists of the COCA and the BNC, 2) the raw frequencies were added up and normalized frequencies generated per ten million words and 3) the verbs were ranked by their normalized frequencies. The complete list of verbs of motion ranked by frequency is given in appendix F. Thus, I ended up with two frequency lists of lexical items for the domains TIME and MOTION to be used as search words to receive tokens from the three corpora.

3.3 Corpus search

It became quite obvious that I could not – at least not in the course of this paper – search the corpora for all 48 nouns in combination with every single one of the 133 verbs in their respective grammatical forms. This would have resulted in an enormous number of different queries which in turn would most likely have produced a huge amount of data. In consequence, I decided to take the 15 most frequent verbs and their respective forms for the present, the past and the infinitive and search for them in combination with the 20 most frequent nouns. This appeared to be the right number of nouns and verbs to generate both a manageable number of queries in the three different corpora and a manageable amount of data resulting from the queries.

3.3.1 The COCA and the BNC

Both corpora are the largest publically available online corpora for the English language. In contrast to the BNC, which has not been continued since 1993, the COCA is still maintained and constantly updated. This results in a total number of 425 million words (as of August 2011) as opposed to 100 million words in the BNC. According to their composition (i.e. different sections/genres and time periods), they are general corpora supposed to reflect the actual everyday language use in the UK and the United States. I decided to include both corpora (and thus the two varieties of the English language) to avoid obtaining quantitative results that might be due to certain distinctive uses within one of the varieties. Of course, I acknowledge that British and American English are just two of a number of different English varieties (e.g. Canadian English, Australian English, African American English etc.). However, because of the availability of the corpora, the time and space limitations of this study and the fact that both varieties are considered to be two of the largest, I decided on them as representatives for original English language use.

Today, the COCA and the BNC are maintained by Mark Davies, a professor of Corpus Linguistics at Brigham Young University in the U.S. Both are equipped with a rather comprehensive search mechanism and additional extensive PoS tagging allows for specific search strings to extract as accurate tokens as possible according to the query. However, since I was looking for lexical expressions of the cognitive metaphor TIME IS MOTION, a pure quantitative search simply including all the tokens resulting from a query was not possible. The tokens had to be examined individually to disregard instances where the noun-verb combination did not fit the conceptual mapping TIME IS MOTION as for example in the phrase *I don't have much time left*, where *left* is not the past tense form of the verb *leave* but an adjective and *time* is not an object or substance in motion but an entity to keep. Another example is the noun *May*, which occurred several times as a proper name instead of referring to the particular month. Anyhow, deciding on nouns as lexical representatives for the domain TIME and verbs for the domain MOTION proved very useful in the corpus search. As an example, the search string for the most frequent noun *time* and the most frequent verb *go* is as follows:

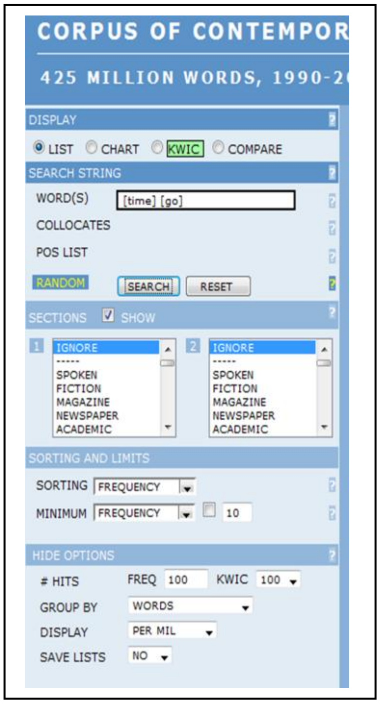
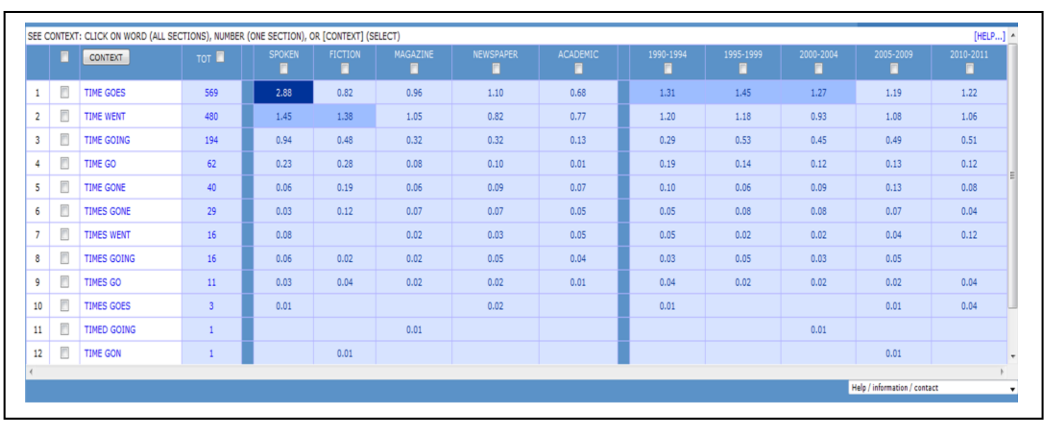


Figure 3.1: Example search string COCA/BNC

I searched for the noun and the verb as lemmatized forms, meaning that the result list displayed any grammatical form of both the noun *time* (i.e. *time*, *times*) and the verb *go* (i.e. *go*, *goes*, *went* etc.). Hence, the result list for the search string above looks as following:


SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT) [HELP...]

	CONTEXT	TOT	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC	1990-1994	1995-1999	2000-2004	2005-2009	2010-2011
1	TIME GOES	569	2.88	0.82	0.96	1.10	0.68	1.31	1.45	1.27	1.19	1.22
2	TIME WENT	480	1.45	1.38	1.05	0.82	0.77	1.20	1.18	0.93	1.08	1.06
3	TIME GOING	194	0.94	0.48	0.32	0.32	0.13	0.29	0.53	0.45	0.49	0.51
4	TIME GO	62	0.23	0.28	0.08	0.10	0.01	0.19	0.14	0.12	0.13	0.12
5	TIME GONE	40	0.06	0.19	0.06	0.09	0.07	0.10	0.06	0.09	0.13	0.08
6	TIMES GONE	29	0.03	0.12	0.07	0.07	0.05	0.05	0.08	0.08	0.07	0.04
7	TIMES WENT	16	0.08		0.02	0.03	0.05	0.05	0.02	0.02	0.04	0.12
8	TIMES GOING	16	0.06	0.02	0.02	0.05	0.04	0.03	0.05	0.03	0.05	
9	TIMES GO	11	0.03	0.04	0.02	0.02	0.01	0.04	0.02	0.02	0.02	0.04
10	TIMES GOES	3	0.01			0.02		0.01			0.01	0.04
11	TIMED GOING	1			0.01					0.01		
12	TIME GON	1		0.01							0.01	

Help / information / contact

Figure 3.2: Example search result COCA/BNC

From the CONTEXT list, I chose the relevant entries (in this case TIME GOES, TIME WENT, TIME GO, TIMES WENT and TIMES GO). The tokens for every section in every entry were then analyzed individually, non-metaphorical uses disregarded and the remaining tokens saved electronically (in MS Word© 2010) and numbered for identification.

3.3.2 The TEC

As mentioned earlier, translational English is treated as an English language variety in this study. The tokens extracted from the Translational English Corpus (TEC) representing translational English originate in quite a number of different source languages. For the purpose of this study however, they are neutralized for their respective source languages plainly constituting translated English. Investigating the influence of the source language of the tokens on a possible quantitative deviation between translated and non-translated language is a step to be taken in another successive study. Such a study is then comparable to Schäffner (2004) and Al-Hasnawi (2007), leading into internal generalization within translation studies (see also section 2.4).

Just like the COCA and the BNC, the TEC belongs to the category of general corpora (see section 3.1) because the corpus is constructed to describe the general aspects of translational English and is not ‘domain (e.g. medicine or law) or genre (e.g. newspaper text or academic prose) specific’ (McEnery 2006: 15). Undeniably, the composition (only written texts) and size (10 million words) of the corpus is small in comparison to the other two corpora. But since the TEC exemplifies translational English, which is a relatively minor variety of English, this is not crucial. For the purposes (and methodology) of this paper, the TEC was the best fitting (as well as the only) option. The corpus is divided into four sub-corpora, namely INFLIGHT MAGAZINES from *Lufthansa* in 1993, NEWSPAPERS from *The Guardian* (1994) and *The European* (1993/1994), BIOGRAPHY with 13 texts and FICTION with 81 texts. Granted, it is not apparent from the structure of the corpus whether the sub-corpus FICTION contains a considerably higher number of words than the other sub-corpora and whether the required balance between the different genres in the TEC is kept. During the statistical analysis of the data set, I make up for this fact by classifying the variable GENRE into fictional and non-fictional texts instead of using the given categories from the corpora. Thus, the sub-corpora INFLIGHT MAGAZINES, NEWSPAPERS and BIOGRAPHY are merged in the variable GENRE, thus (hopefully) achieving balance between the text types in the corpus. This increases comparability. However, I acknowledge the possible imbalance and its probable influence on the results of the analysis of the variable GENRE. But, as mentioned before, the TEC was my only option given the premises of this paper. In accordance with the two previously introduced corpora, the concordancer was employed to search for the noun-verb combinations. Since this search tool is not as sophisticated as the one from the COCA and the BNC, every relevant grammatical possibility (singular/plural, present/past/infinitive) had to be searched for individually. As an example, the search string and the result list for the pattern *time goes* is given below:

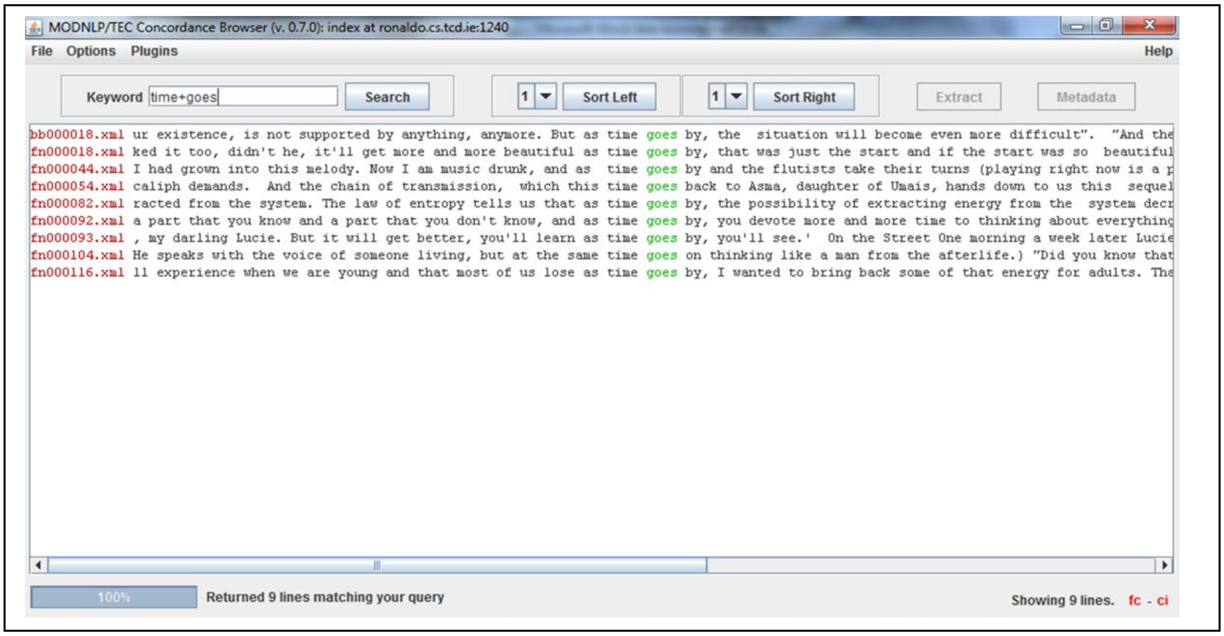


Figure 3.3: Example search string and result list TEC

Again, the tokens were manually scanned for non-metaphorical uses of the noun-verb combination (see section 3.3.1) and the metaphorical tokens saved and numbered.

As a result of the data extracting process from all three corpora, I ended up with n = 6850 tokens, of which 5570 come from the COCA, 709 from the BNC and 571 from the TEC. One can see that the decreasing token count follows the decreasing number of words (410, 100 and 10 million) from one corpus to the next. The following section gives a detailed description of how the data is subject to different statistical methods to gain empirical validity.

3.4 Analysis

The aim of a usage-based study including data is to reveal relations between different parts of the data and thus answering research questions. To determine possible relations between the tokens in my data set, I conducted statistical analyses using the computer program Statistical Package for the

Social Sciences (SPSS). The program is designed to collect, store and process large amounts of data in a time-saving manner. It includes all statistically relevant methods necessary for studies within social science and proves to be equally useful for linguistic studies. As mentioned earlier, all the 6850 tokens collected from the three corpora were numbered. This way, I did not need to enter the whole token (phrase) into SPSS, but only the number for identification. For cross-referencing, the number (and thus the respective token) can easily be found in my sample. In addition, every token was coded according to a number of variables which will be explained in the next section.

3.4.1 The variables

To have as many options as possible to compare tokens, I recorded as much information (variables) as possible about the tokens. A list of the variables is given in the table below:

Table 3.2: Variables in SPSS

1.	noun	10.	year
2.	noun number	11.	translation year
3.	verb	12.	translator gender
4.	verb number	13.	translator employment
5.	verb use	14.	source language
6.	tense	15.	original year
7.	corpus	16.	translation mode
8.	translated	17.	translation place
9.			

Every variable was coded for several different classifications (these are called *values* in SPSS) which are given in appendix G. For the first variable *noun* for example, there are 20 values denoting the 20 TIME nouns used in the corpus search. The number of the nouns was given in either singular or plural. The use of the verb describes whether the noun was used in third person or as an infinitive in questions or in a future construction or as a modal etc. At this point, it is necessary to make a few comments on the different usages of the infinitive form. Within the data set, infinitive forms are part of present tense questions like *Where does the time go?*, past tense questions like *Where did the time go?*, future tense questions like *When will her time come?* or modal constructions like *How far could or should this night go?*. They are also part of modal questions or are used in connection with auxiliary verbs like *do, does, doesn't* etc. and in imperative constructions like *Let the good times come*. Although most of these instances could be counted as either present or past tense, I decided to record them individually to get a better and more detailed description of the data set. In SPSS, the different instances of infinitive uses were coded for their particular occurrence (e.g. past or present tense question, modal construction etc.). Unfortunately, I cannot include this data in the present study due to time and space restrictions. For that reason, all these different instances are conjointly called infinitive. Variable numbers 11 till 17 are only valid for the TEC tokens since they refer to properties of translations. Again, because of the premises of this paper, it became clear that I could not possibly employ all 17 variables with their respective 202 values in the analysis. Regarding the research questions whether and how the usage of metaphorical expressions of the cognitive metaphor TIME IS MOTION differs between translated and non-translated English, five variables were chosen to be investigated further: **noun**, **verb**, **tense**, **genre** and **translated**. Since the variable **translated** is assumed to be the influencing factor (i.e. whether a token originates in original or translated English), it constitutes the independent variable while the other four are dependent variables which are assumed to be influenced by the independent variable. With regard to the twelve variables in the data set which are not used in this particular study, the thesis has an introductory, directing character for the actual scientific problem in question, namely that translated and non-translated English differ in the use of metaphorical expressions not only of the metaphor TIME IS MOTION but also of other conceptual metaphors. This, in turn, might be due to several different reasons like the source language or the employment situation of the translator. I am aware that there are many different options to incorporate more or different variables and values from the data set to answer the same or different research questions possibly supporting or refuting my findings here. I will come back to this issue and future prospects of research in the conclusion.

3.4.2 Descriptive analysis

To start the analysis of the data set and summarize the findings, I describe the data set in terms of frequency distributions according to the different variables. Oakes (1998) states that '[d]escriptive statistics enable one to summarise the most important properties of the observed data' (1998:1). I give the numbers in raw frequencies (number of tokens) as well as normalized frequencies (tokens per ten million words) since I want to describe the data set as accurately as possible, as well as compare frequencies. To compare ranked frequencies between the use of the variables within the data set and within actual English language use (according to the COCA and the BNC) as well as to compare ranked frequencies between the translated and the non-translated category of a variable, the Spearman correlation coefficient is employed. This statistical tool allows for comparing 'data when it is ordinal' (Hinton 2004:300) like the rank positions of the nouns or verbs in this study. The respective value of the correlation coefficient (also called Spearman's rho) lies between -1 and 1 with the negative value indicating no correlation between the ranking of the respective values in a variable and 1 indicating complete correlation (SPSS 19 Help).

In a next step, cross-tabulations are generated for the four dependent variables (**noun**, **verb**, **tense** and **genre**) and **translated** as independent variable. Cross-tabulations (or contingency tables) show relationships between variables, whereby some are dependent (given and assumed to possibly be affected) and some are independent (given or not and assumed to affect). Cross-tabulations are extremely suitable to investigate frequency data because they examine bivariate distribution of data (i.e. the dependent variables vs. the independent variable). A cross-tabulation for the variables **genre** and **translated** for example sets the frequencies of the values of the dependent variable **genre** (fiction, non-fiction) against the frequencies of the values for the variable **translated** (*yes*, *no*). Thus, one can easily extract the absolute frequencies for translated fictional, translated non-fictional, non-translated fictional and non-translated non-fictional tokens within the data set. The numbers can be compared and give useful information about possible over- or underrepresentation of a particular category of tokens in the data set (e.g. if there are more fictional translated than non-translated tokens or if there are fewer non-fictional translated tokens than fictional translated tokens etc.). However, a descriptive analysis of the data set is not sufficient enough to make any claims about its relevance. It is necessary to determine that the distribution of the data is not only a matter of coincidence but an instance of association between at least two relevant factors. In other words, 'we need ways of making sense of the data, and this is the purpose of statistical data' (Butler 1985: vii).

3.4.3 Statistical analysis

The nature of the data in the data set is nominal, i.e. the tokens are assigned different variables and are not graded according to a certain order or rating. Statistical methods and calculations applied to this data are 'standardised procedures to quantitatively estimate and evaluate … relations' (Hannisdal 2007:143) between variables. In statistical analyses, one always assumes that the relevant variables do not have any relation whatsoever and thus 'the … distributions are the same' (Hinton 2004:29). This is the so-called null hypothesis. Significance testing is supposed to prove or refute this null hypothesis. For my data, the null hypothesis states that the distribution of the tokens for every dependent variable (i.e. **noun**, **verb,** their affiliation to a certain **genre** or their grammatical **tense**) is not associated with them being translated from another language or not. There is no reason to believe that the use of metaphorical expressions of the cognitive metaphor TIME IS MOTION in translated English differs from the usage in original, non-translated English. This automatically presupposes that the distribution for every variable between translated and non-translated tokens in the data set is the same, i.e. there are as many translated tokens as there are non-translated tokens for every variable. Quantitative discrepancy (over or underrepresentation) would hint at some kind of association between the dependent variables and the independent variable. In other words, over or underrepresentation of either translated or non-translated tokens of a variable suggests an affiliation of the usage of the tokens according to their classification as translated or non-translated. However, it cannot be excluded that possible over- or underrepresentation of tokens within a variable is due to mere chance of distribution and not subject to an association with the independent variable. Statistical significance testing is a way of investigating whether differences in a data set are random variations in the sample set or whether they are due to an affiliation between dependent and independent variables and thus empirically valid.

Regarding the nature of the data in my data set (nominal data), I employ the chi-square/Fisher's Exact test to determine validity of assumed associations between the dependent variables and the independent variable. Furthermore, adjusted residuals are calculated to assess the difference between observed and expected token counts and finally Cramer's V test and the Phi coefficient are used to evaluate the strength of a possible association between dependent and independent variables. The chi-square test 'examines … proportions and presents the probability of obtaining this pattern when there is no difference in the choices' (Hinton 2004: 275). The result of the chi-square test is presented as the *p*-value and assesses how well the observed data fits the expected results. The latter represent the token distribution if there is no affiliation between the dependent and the independent variable, i.e. the null hypothesis. The *p*-value has to be equal to or less than 0.05 to be statistically significant, thus rejecting the null hypothesis, and is calculated in relation to the number of single values that are included in the dependent variable minus one. This relation is given as degree of freedom (df) in the chi-square table. In a cross-tabulation of the variables **noun** and **translated** for example, the dependent variable **noun** might be represented by 12 different values (12 different nouns). Hence the degree of freedom is 11. It has to be kept in mind though, that statistical significance neither proves nor disproves one's own hypothesis instantly. It rather 'can strengthen it indirectly by rejecting the null hypothesis' (Hannisdal 2007: 142). The chi-square test is not applicable to cells (e.g. the distribution of a certain noun in the **translated** category) that have an expected count of less than five tokens per variable (Kinnear and Gray 2010:427). If for example in the cross-tabulation of the variables **noun** and **translated** the noun *season* only exhibits four expected occurrences amongst the translated tokens in the sample, *season* has to be excluded from the chi-square test. Expected counts for every value in the cross-tabulation are calculated by SPSS automatically. The results of the statistical significance tests are only valid for the included values of the variables. Thus, the result of the chi-square test for the cross-tabulation of the variables **noun** and **translated** for example is not valid for the noun *season* since it had to be excluded for the said reasons. The remaining cells (e.g. all the nouns with five or more expected tokens per translated/non-translated category) are subject to significance testing. Fisher's Exact test resembles the chi-square test but is applicable to 2x2 contingency tables.

The difference between observed and expected counts for every value in a cross-tabulation is assessed by an adjusted residual. The value for the adjusted residual evaluates whether the difference between the two counts is considerably large, implying significant divergence from the null hypothesis, or rather small, neglecting the null hypothesis (Kinnear and Gray 2010:514). A value of two or more indicates significant divergence. The higher the value, the greater the divergence from the expected count and thus from the null hypothesis. Positive values for the adjusted residual designate more observed counts than expected while negative values indicate fewer observed tokens than expected if the null hypothesis was true.

Finally, Cramer's V and Phi coefficient evaluate the strength of the association between the two variables in the cross-tabulation. Simply stating the association between two variables in a data set is not enough. It is also necessary to evaluate this association to determine whether the association is strong or rather weak on the basis of the tokens included in a variable. SPSS calculates the effect size value (Cramer's V or Phi) and Kinnear and Gray (2010) propose to 'transform Cramer's V into the equivalent value of Cohen's index of effect size *w*' to evaluate the effects size value (2010:417). To do this, one simply has to multiply Cramer's V with the square root of the number of rows or columns (whichever is smaller) minus 1.[16]

The problem with large datasets like the present one (n = 6850) is that they 'basically guarantee that even miniscule effects will be highly significant' (Gries 2010:20). In other words, the chi-square tests for these data almost of necessity produce significant results. Effect size tests, on the other hand, help to evaluate these results on a more reliable basis taking into account the size of the data set and are therefore indispensable (Kinnear and Gray 2010:417). Cramer's V value is given on a scale between 0 and 1 with 0 indicating no association and 1 strong association between the variables. The test is employed for cross-tabulations larger than 2x2 while the Phi coefficient evaluates the strength of association 2x2 for contingency tables. The results of the descriptive and the statistical analysis are presented in the following chapter.

[16] The correct equation looks as following: $w = V \sqrt{r-1}$ with V being the Cramer's V value and r the number of rows or columns.

4 RESULTS

This chapter presents the data set and the results of the descriptive and the statistical analysis of the five variables (**noun**, **verb**, **tense**, **genre** and **translated**) as described in section 3.4.1. The chapter is divided into two sections. The first gives a descriptive overview of the whole data set and the interrelation (distribution of the tokens) between the four dependent variables **noun**, **verb**, **tense** and **genre** and the independent variable **translated**. Results are given for the data set as a whole and for each of the four dependent variables individually. All results are displayed in tables and figures for better visualization of interrelations and are generated with SPSS and MS Excel©. The second section describes the results of the statistical tests (chi-square, Cramer's V etc., see section 3.4.3) performed on the data set to determine the significance of the results. These results are given in tables with the respective statistical values (*p*-value, exact value).

For purposes of clarity, the research questions and the hypothesis which form the basis of this paper are repeated below and adjusted to the methodological approach presented in chapter three:

- Are there quantitative differences between the translated and the non-translated category of every dependent variable? If so, are these differences statistically significant, i.e. are the deviations from the null hypothesis sufficient enough to assume significance?
- What do possible differences look like? Is the translated category over- or under-represented compared to the non-translated category of every dependent variable?

Finally, the analysis and the discussion in chapter five addresses my hypothesis based on Toury's translational laws and Baker's translational norms:

- There are quantitative differences within the distribution of tokens between the translated and the non-translated category of a variable indicating that the independent variable **translated** influences the distribution of the tokens for every dependent variable and thus translated language differs from original language in terms of the usage of lexical expressions of the cognitive metaphor TIME IS MOTION.

4.1 Descriptive analysis

The compiled data set is relatively large and consists of n = 6850 tokens. Every token represents a phrase extracted from one of the three corpora, i.e. a phrase containing a metaphorical pattern of a noun representing the target domain TIME and a verb form representing the source domain MOTION. Table 4.1 and figure 4.1 show the distribution of the tokens between the three corpora.

Table 4.1: Observed token distribution per corpus

Corpus	Observed frequencies	Percent
COCA	5570	81.3
BNC	709	10.4
TEC	571	8.3
Total	6850	100.0

The distribution of the tokens between the three corpora follows the total number of words included in the corpora proportionally. The COCA with its 410 million words (April 2011) is represented with the highest number of tokens (n = 5570), followed by the BNC (n = 709) with 100 million words in total and the TEC (n = 571) with 10 million words in total. With 81.3 percent, the COCA provides a clear majority not only of all tokens but also of the tokens representing original English. The remaining tokens are distributed almost equally between the BNC (original English) with 10.4 percent and the TEC with 8.3 percent (translated English).

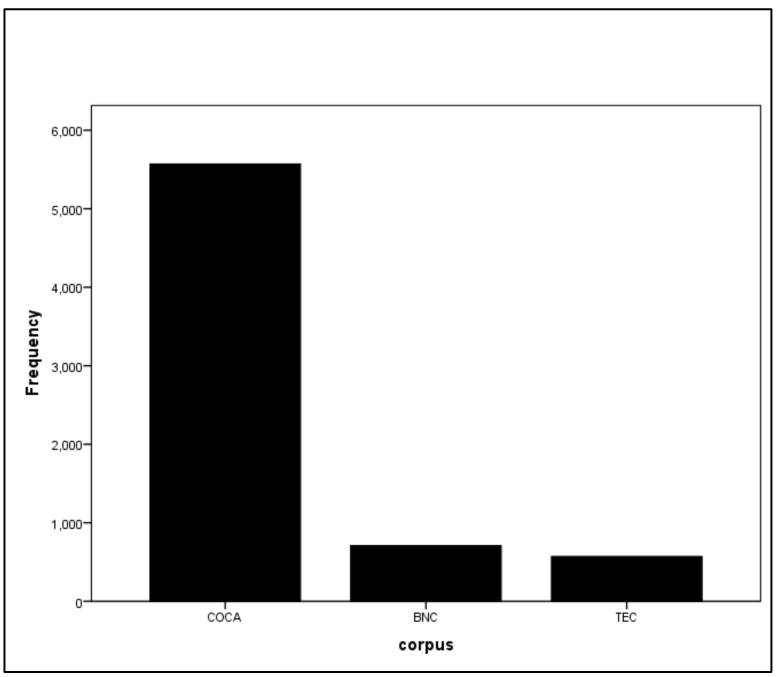


Figure 4.1: Observed token distribution per corpus

Since the tokens from the COCA and the BNC represent original English, they are summarized and represented as the value *no* within the variable **translated**. The tokens from the TEC constitute the translated category within the variable and are labeled by the value *yes*. The distribution between translated and non-translated tokens within the data set is given in figure 4.2. below:

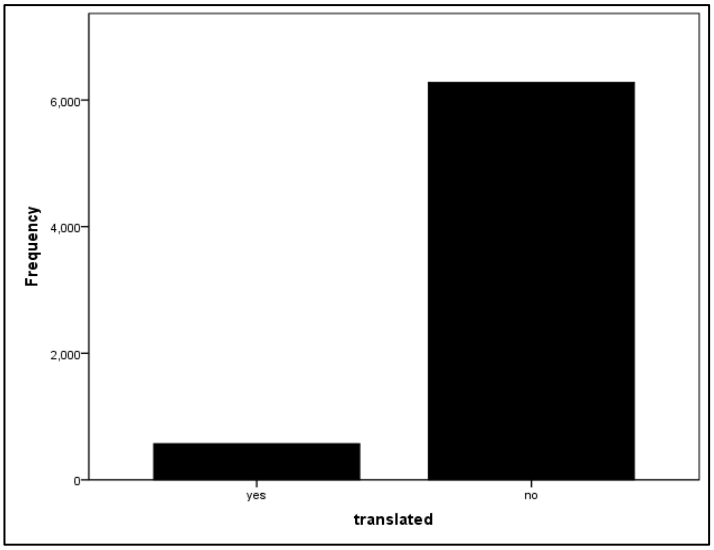


Figure 4.2: Observed token distribution for the variable ***translated***

Without a doubt, the distribution between translated and non-translated tokens within the data set is highly unproportional. The bar for the translated tokens (*yes*) is significantly lower than the bar for the non-translated tokens (*no*). There are 571 translated and 6279 non-translated tokens. This large divergence is mainly due to the aforementioned difference in size between the corpora. Hence, the distribution of the tokens for the dependent variables in the next four sections is given in normalized figures per ten million words to acknowledge this difference and to generate better comparability.

Translated is the independent variable which I consider to influence the usage of metaphorical expressions in one way or another. In other words, I consider quantitative differences between the distributions of the tokens to be due to their affiliation to either the translated or the non-translated category of a variable. Thus, the variable is tested against the four dependent variables. The distribution of these four variables within the data set and the results of their interrelation with the independent variable are described in the following sections.

4.1.1 Noun

As explained in section 3.2, a list of searchable TIME-related nouns was generated and the corpora searched for the 20 most frequent nouns of this list. The distribution of these nouns within the whole data set independent of the corpus of origin is shown below:

Table 4.2: Observed and normalized ranked distribution per noun in the data set

Noun	Oberserved frequencies	Normalized frequencies per ten million words
time	2194	42
day	934	18
life	679	13
night	626	12
year	489	9
moment	445	9
season	247	5
week	244	5
morning	179	3
month	177	3
evening	166	3
hour	156	3
minute	91	2
end	88	2
age	43	1
Sunday	31	1
future	22	0
century	20	0
decade	12	0
May	7	0
Total	6850	

All 20 nouns produced relevant tokens in the corpus search and are thus included in the data set. The table above ranks the nouns by raw frequency (left-hand column) and normalized frequency (right-hand column). This means that the noun *time* occurs 2194 times in the data set, which corresponds to 42 times per ten million words and is thus the most frequent noun. A comparison of this list with the ranked frequency list of original English (appendix E) reveals that the usage of these nouns within expression of the cognitive metaphor TIME IS MOTION differs from the overall use of the same nouns in actual language use. The nouns in the data set (metaphorical uses) do not quite follow the same ranking as the nouns in the COCA and the BNC (metaphorical and non-metaphorical uses). This is illustrated in table 4.3:

Table 4.3: Comparison of ranked normalized frequencies per noun

	Data set	**COCA/BNC**
1	time	time
2	day	year
3	life	day
4	night	life
5	year	week
6	moment	night
7	season	month
8	week	end
9	morning	minute
10	month	morning
11	evening	moment
12	hour	age
13	minute	hour
14	end	season
15	age	century
16	Sunday	future
17	future	decade
18	century	evening
19	decade	May
20	May	Sunday

The table ranks the 20 nouns used in this study by their normalized frequencies. The first column indicates the rank number while the second column contains the ranked order of the nouns from the data set, i.e. the nouns that are used in metaphorical expressions of TIME IS MOTION. The third column gives the ranked frequency of the same nouns within all language use (based on the COCA and the BNC, see appendix E). This includes also the metaphorical uses of the first column. While *year* is the second most frequent noun in everyday language use, it has descended to the fifth position within metaphorical expressions. The same is true for *end*, which has descended from the eighth position to the 14th. On the other hand, *season* has advanced from rank 14 to rank seven. Indeed, only the ranking of *time* in first position concurs between the two columns. All other metaphorically used nouns have either descended or ascended in comparison to the ranking of the nouns in the overall language use. This implies that although highly frequent in everyday language use, the same nouns are not equally available for usage in metaphorical language. For reasons not apparent from the data set, *year* seems to be less suitable to be "cognitively in motion" than *day*, *life* and *night*. On the other hand, *season* seems to be more readily used in metaphorical expressions of motion than in daily language use. However, Spearman's correlation coefficient reveals close correlation between the two rankings of the nouns:

Table 4.4: Spearman's rho for the variable ***noun***

			COCA/BNC	**Data set**
Spearman's rho	**COCA/BNC**	Correlation Coefficient	1.000	.812
		Sig. (2-tailed)		.000
		N	20	20
	Data set	Correlation Coefficient	.812	1.000
		Sig. (2-tailed)	.000	
		N	20	20

The correlation coefficient for the ranking of the nouns in the data set (metaphorical usage) and the ranking within the COCA and the BNC (actual language usage) is close to the value of 1, which represents complete correlation: $r_s = 0.812$, $N = 20$, $p < 0.01$. Hence, the ranking between the different nouns does not differ significantly.

As described in section 3.4.2, contingency tables in the form of cross-tabulations demonstrate the association of the independent variable with a dependent variable. For the variable **noun** the cross-tabulation looks as follows:

Table 4.5: Cross-tabulation normalized distribution per ten million words for ***noun/translated***

noun	translated	
	yes	no
time	157.0	39.9
day	121.0	15.9
life	38.0	12.5
night	70.0	10.9
year	40.0	8.8
moment	23.0	8.3
season	4.0	4.8
week	18.0	4.4
morning	11.0	3.3
month	25.0	3.0
evening	21.0	2.8
hour	27.0	2.5
minute	5.0	1.7
end	5.0	1.6
age	1.0	0.8
Sunday	3.0	0.6
future	0	0.4
century	0	0.4
decade	0	0.2
May	2.0	0.1

The table above represents the normalized distribution[17] of the tokens for every noun between the translated and the non-translated category. The first column gives the nouns, while the second and third give the normalized frequencies for translated (Yes) and non-translated (No) tokens respectively. Thus, *time* as a metaphorical expression of the cognitive metaphor TIME IS MOTION occurs 157 times per ten million words in translated and 39.9 times per ten million words in non-

[17] Since the TEC only consists of 10 million words, the normalized frequencies happen to concur with the observed counts for the translated category (yes).

translated phrases. The distribution between translated and non-translated nouns is demonstrated in figure 4.3.

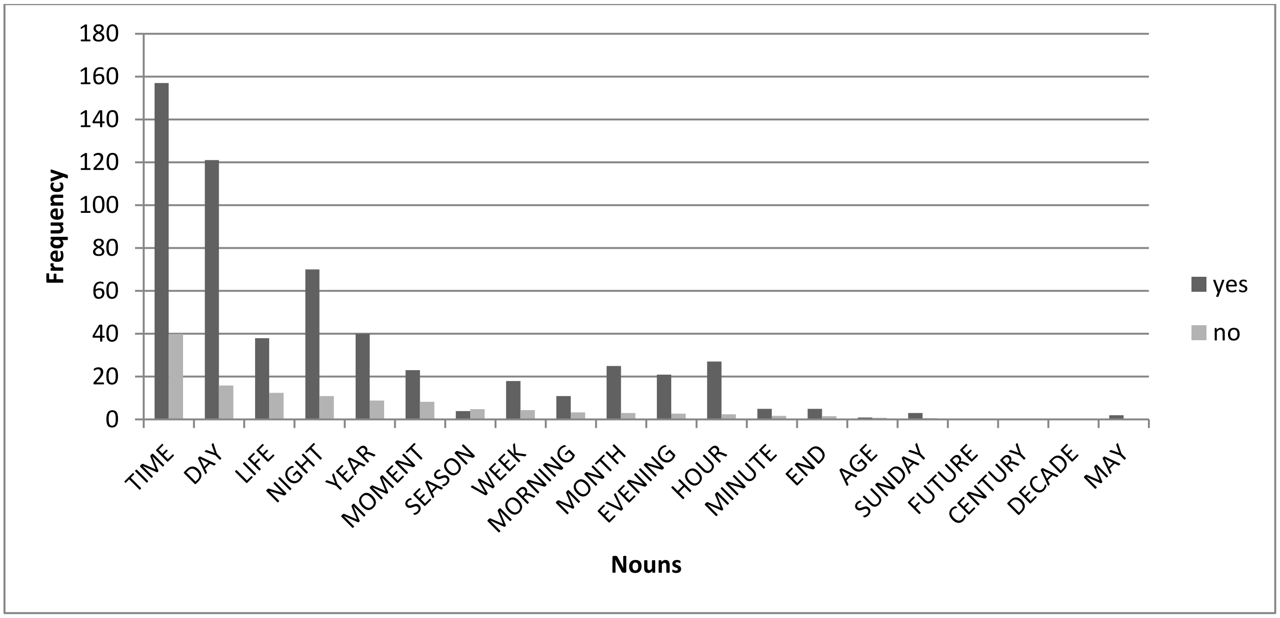


Figure 4.3: Normalized distribution per ten million words ***noun/translated***

The figure graphically describes the normalized distribution per noun. The order of the nouns on the x-axis complies with the ranked order of frequency of the nouns within the whole data set (see table 4.3). The y-axis displays the normalized instances per ten million words. The nouns *future*, *century* and *decade* did not produce any translated tokens (see the normalized value of 0 in table 4.5). It is noticeable that the non-translated tokens follow the ranked frequency distribution for metaphorical expressions in the data set (see table 4.5 and the x-axis in this figure) with *time* being the most frequent noun and *May* the noun with the lowest value. On the other hand, the translated tokens do not follow the same ranked order. The highest value conforms in both categories (translated and non-translated) to the noun *time*. After this point (i.e. the noun *time*) the ranked order deviates considerably. The values for the category of nouns do not follow a quantitative decline from one noun to the next on the x-axis but continue ascending and descending. While *life* (12.5) is more frequently used than *night* (10.9) per ten million words within non-translated English, it is less often used within translated tokens (38 vs. 70). The same applies for the nouns *week*, *month* and *hour*, which are more often used than the respective preceding noun within the translated tokens in contrast to the non-translated tokens. Hence, the ranked distribution between translated and non-translated tokens differs (see table 4.6).

Table 4.6: Comparison of ranked normalized frequencies of ***noun/ translated***

	Non-translated	**Translated**
1	time	time
2	day	day
3	life	night
4	night	year
5	year	life
6	moment	hour
7	season	month
8	week	moment
9	morning	evening
10	month	week
11	evening	morning
12	hour	minute
13	minute	end
14	end	season
15	age	Sunday
16	Sunday	May
17	future	age
18	century	future
19	decade	century
20	May	decade

The most significant difference is the ranking of *season*, which takes up the seventh position within non-translated tokens and descends to the 14th position within translated tokens. It is interesting though that the normalized frequencies for *season* within translated and non-translated tokens are fairly equal with 4 and 4.8 usages respectively per ten million words (see table 4.3). Overall, the ranking between the two categories within the variable **noun** correlates considerably:

*Table 4.7: Spearman's rho for the variables **noun/translated***

			Non-translated	**Translated**
Spearman's rho	**Non-translated**	Correlation Coefficient	1.000	.893
		Sig. (2-tailed)		.000
		N	20	20
	Translated	Correlation Coefficient	.893	1.000
		Sig. (2-tailed)	.000	
		N	20	20

Also the ranking of the nouns within the translated and the non-translated category correlates highly: $r_s = 0.893$, $N = 20$, $p < 0.01$. The correlation coefficient is with a value of 0.893 close to complete correlation.

4.1.2 Verb

In accordance with the ranked frequency list in appendix F, the twenty nouns were also searched for in combination with the first 15 verbs in present tense, past tense and the infinitive form. Table 4.8 gives an overview of the overall distribution of these verbs in the data set:

Table 4.8: Observed and normalized ranked distribution per verb in the data set

Verb	Oberserved frequencies	Normalized frequencies per ten million words
go	2759	53
come	2734	53
fall	481	9
run	240	5
arrive	216	4
roll	170	3
fly	126	2
return	63	1
rise	22	0
enter	17	0
leave	12	0
travel	5	0
jump	5	0
Total	6850	

The table displays the verbs in the infinitive form in the left-hand column. The middle column indicates how many times a verb occurs in the data set and the right-hand column gives the normalized distribution per ten million words. Thus, 2759 (or 53 per ten million words) tokens include a phrase using either *go*, *goes* or *went* and 2734 (or 53 per ten million words) tokens include *come*, *comes* or *came* etc. Again, the ranked frequency list within the data set (i.e. the usage of the verb forms in metaphorical phrases) differs from the ranked frequency list generated from the COCA and the BNC for verbs of motion (see appendix F). Table 4.9 depicts this divergence.

Table 4.9: Comparison of ranked normalized frequencies per verb

	Data set	**COCA/BNC**
1	go	come
2	come	go
3	fall	leave
4	run	run
5	arrive	fall
6	roll	walk
7	fly	return
8	return	arrive
9	rise	rise
10	enter	fly
11	leave	travel
12	travel	enter
13	jump	roll
14		jump
15		hop

Looking at the table above, the usage within metaphorical expressions of TIME IS MOTION does not follow the usage within everyday language use. *Leave* as the third most frequent verb of motion within BNC and COCA (right-hand column) has drastically descended and is only represented in 12 of 6850 tokens in position 11 within the data set (left-hand column). *Come* and *go* still top the list in first and second position respectively on both sides. The verbs *walk* and *hop* did not produce any metaphorical tokens in any of the corpora, which indicates that they are not used to express the cognitive metaphorical concept of TIME IS MOTION. Hence the left-hand column contains only 13 instead of 15 verbs. The test for correlation between the rankings is not significant since the p-value is 0.208, which is >.05.

Table 4.10: Spearman's rho for the variable ***verb***

			COCA/BNC	Data set
Spearman's rho	**COCA/BNC**	Correlation Coefficient	1.000	.345
		Sig. (2-tailed)		.208
		N	15	15
	Data set	Correlation Coefficient	.345	1.000
		Sig. (2-tailed)	.208	
		N	15	15

The correlation coefficient is close to 0: $r_s = 0.345$, $N = 15$, $p < 0.3$.

The distribution of the verbs between translated and non-translated tokens is given in the next table:

Table 4.11: Cross-tabulation normalized distribution per ten million words for ***verb/translated***

verb	**translated**	
	yes	no
go	236.0	49.5
come	208.0	49.5
fall	49.0	8.5
run	6.0	4.6
arrive	30.0	3.6
roll	7.0	3.2
fly	22.0	2.0
return	10.0	1.0
rise	0	0.4
enter	2.0	0.3
leave	1.0	0.2
travel	0	0.1
jump	0	0.1

Table 4.11 above lists the usage of the 13 verbs (infinitive, present tense and past tense form) from the data set (see table 4.9) in the left-hand column and the normalized distribution of tokens between the translated (middle column) and the non-translated (right-hand column) category. In the first row for example, a form of the verb *go* appears 236 times per ten million words in

translated and 49.5 in non-translated tokens. The verbs *rise*, *travel* and *jump* occur in non-translated metaphorical phrases but not in translated phrases in the data set. Figure 4.4 below depicts the relation between the translated and non-translated tokens:

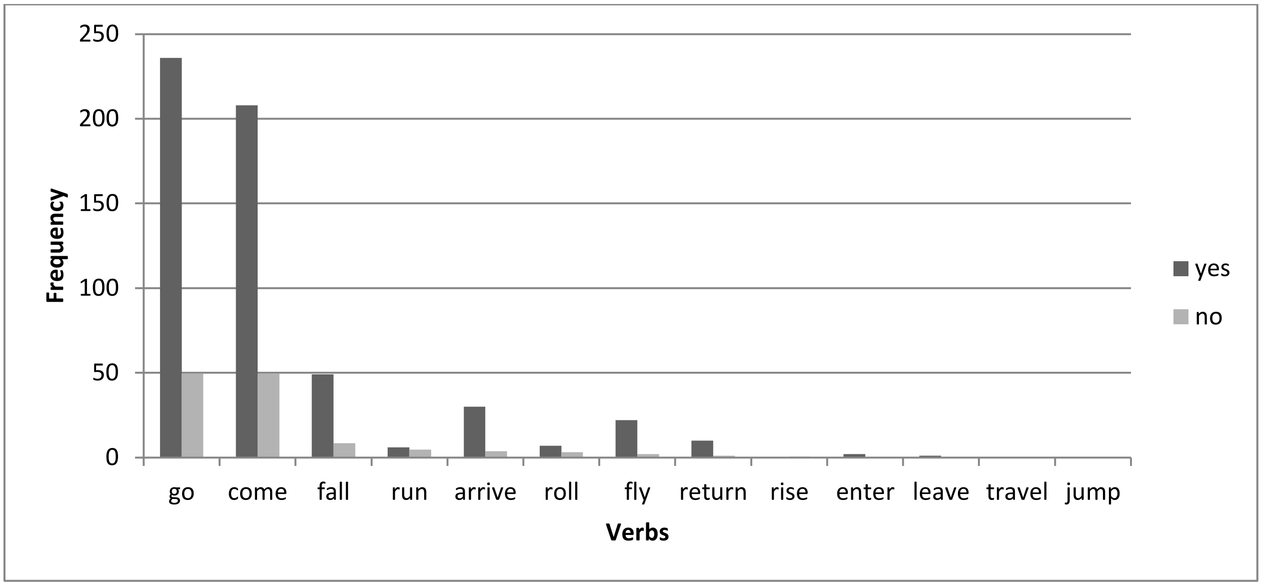


Figure 4.4: Normalized distribution per ten million words ***verb/translated***

The x-axis in the figure above displays the ranked order per verb in the data set (see table 4.9). The y-axis shows instances per ten million words. *Go* and *come* dominate both translated and non-translated tokens. Thus, they are the two most frequently used verbs of motion in metaphorical expressions of TIME IS MOTION as well as in general English language use – both in original and in translated English. While the translated tokens (black bar) exhibit a numeric difference between the two verbs, with *go* (236) being more frequent than *come* (208), the non-translated tokens (gray bar) display equality between the two verbs with 49.5 occurrences per ten million words. Again, the distribution of translated and non-translated tokens for the different verbs differs – in some cases significantly. As mentioned, *go* and *come* are the most frequently used verbs in both translated and non-translated English. Accordingly, both distributions (black for translated and gray for non-translated) display their highest values where *go* and *come* are located on the x-axis. Beginning with the verb *fall*, the bars for both values decrease, indicating a reduction of instances per ten million words for the respective verbs. However, while the descent of the bars for the non-translated verbs is linear (one verb is less frequent than the preceding verb), the bars for the translated verbs display inconsistency. *Run* for example is represented six times per ten million words but *arrive*, which succeeds *run* within non-translated tokens, occurs 30 times per ten million words. Thus the quantitative difference between *run* and *arrive* is oppositional within the translated and the non-translated category. Similarly, *fly* is more frequent than the preceding *roll*. The same divergence has already been described for the variable **noun**.

4.1.3 Tense

The 15 verbs were searched for their respective forms in present tense, past tense and infinitive and categorized accordingly in the data set. The use of the infinitive form in different grammatical constructions of present and past tense was discussed in section 3.4.1. The table below depicts the raw and the normalized distribution between the three verb forms within the whole data set ranked by occurrences:

Table 4.12: Observed and normalized ranked distribution per tense in the data set

verb form	Oberserved frequencies	Normalized frequencies per ten million words
past	3716	71
present	2803	54
infinitve	331	6
Total	6850	

Table 4.12 gives the three different verb forms in the left-hand column, the actually observed count in the data set in the middle column and occurrences per ten million words in the right-hand column. Infinitive constructions are with 331 tokens (six instances per ten million words) rather seldom represented. The majority of tokens (3716 in the data set or 71 per ten million words) contain a verb form in past tense followed by 2803 (54 per ten million) tokens in present tense. The distribution of the tokens between the translated and the non-translated category is displayed in table 4.13 below:

Table 4.13: Cross-tabulation normalized distribution for ***tense/translated***

verb form	translated	
	yes	no
past	389	65
present	161	52
infinitive	21	6

Undeniably, there is a considerable divergence between the translated and the non-translated category. In translated language, there are 389 occurrences per ten million words in past tense, but only 65 usages per ten million words in non-translated language. The same applies for the other two verb forms with 161 to 52 instances per ten million words in present tense and 21 to six occurrences in any of the infinitive forms.

Within this variable, the translated tokens follow the quantitative progression of the non-translated tokens from infinitive constructions at the bottom to present tense phrases and finally past tense tokens at the top of the count (see figure 4.5).

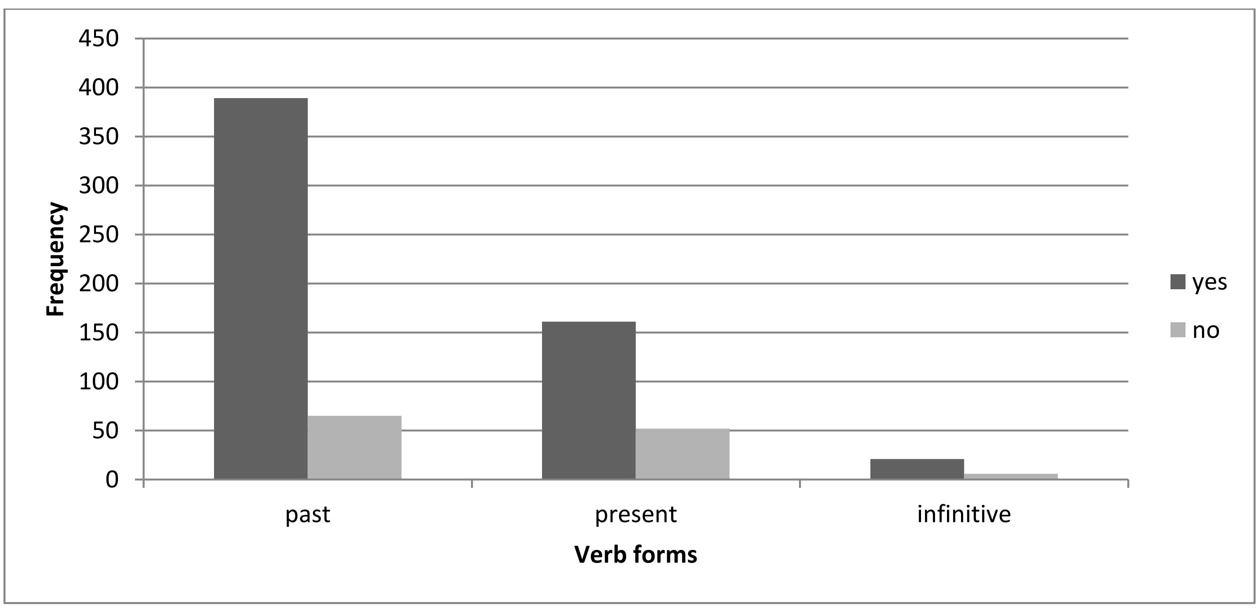


Figure 4.5: Normalized distribution per ten million words ***tense/translated***

The order of the verb forms represented on the x-axis again complies with the ranked order in the data set as demonstrated in table 4.12. The y-axis again shows instances per ten million words. The quantitative progression in instances per ten million words from past tense to infinitive forms is clearly visible in the figure above. Both the translated and the non-translated category of the variable **tense** follow the same pattern, in contrast to the two preceding variables **noun** and **verb** where the distribution of translated and non-translated tokens differs.

4.1.4 Genre

Due to differences in the construction of the three corpora, the categorization of the tokens in a fictional and a non-fictional category in the data set is assumed to be most suitable for the purposes of this study (see also section 3.3.2). Thus, the variable **genre** only contains two values. The next table presents the ranked distribution between those two categories for the whole data set:

Table 4.14: Observed and normalized ranked distribution per genre in the data set

genre	Oberserved frequencies	Normalized frequencies per ten million words
Non-Fiction	3463	67
Fiction	3387	65
Total	6850	

The distribution of fictional and non-fictional tokens within the data set is quite balanced with 3463 tokens (67 instances per ten million words) originating in non-fictional texts and 3387 (65 per ten million words) in fictional texts. Consequently, the quantitative difference between the two categories can be considered marginal. The same does not apply for the distribution of the tokens between the translated and the non-translated category, as table 4.15 below displays:

Table 4.15: Cross-tabulation normalized distribution per ten million words for ***genre/translated***

genre	translated	
	yes	no
Non-Fiction	102	66
Fiction	469	57

The non-fictional part of the translated category (Yes) is considerably smaller than the fictional part. Only 102 tokens per ten million words originate in non-fictional texts while 469 were extracted from fictional texts. In contrast, the non-translated category (No) exhibits a much more balanced distribution of fictional and non-fictional tokens. Sixty-six tokens per ten million words are non-fictional and 57 tokens are fictional. The difference (11 tokens) is considered to be minimal in relation to the 367 tokens per ten million words that distinguish non-fictional translated tokens from fictional translated tokens. The figure below depicts the distribution of the table above in a bar chart:

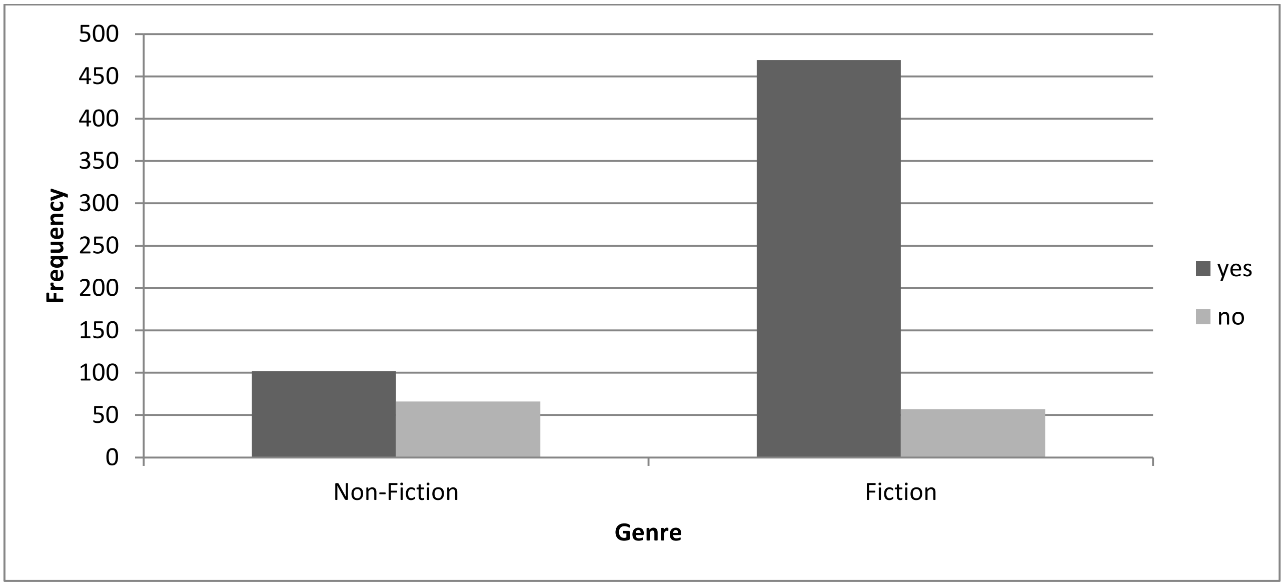


Figure 4.6: Normalized distribution per ten million words ***genre/translated***

Figure 4.6 illustrates the large difference between the translated (black bar) and the non-translated category (gray bar) within fictional tokens. The x-axis displays the two values of the variable **genre**, fiction and non-fiction divided into translated and non-translated instances. The order of the values on the x-axis conforms to the ranked order within the data set (see table 4.14).The y-axis shows instances per ten million words. There are 469 translated tokens per ten million words and 57 non-translated tokens. In contrast, the non-fictional tokens do not exhibit that big a difference. The data set contains 102 translated and 66 non-translated tokens per ten million words (see table 4.13). **Genre** is another variable where the distribution of the values (fictional and non-fictional) within the translated and the non-translated category is oppositional. Whilst the non-translated tokens exhibit a numeral rise from fictional tokens (57) to non-fictional tokens (66), the translated tokens experience the opposite development and decline from 469 fictional tokens to

102 non-fictional ones. Thus, there are more fictional than non-fictional translated tokens but more non-fictional than fictional non-translated tokens.

4.2 Statistical analysis

As described in section 3.4.3, the statistical tests performed on the data are used to determine the empirical significance of the association between the independent variable **translated** and the four dependent variables **noun**, **verb**, **tense** and **genre**. The results of these tests and their implications on the data are presented in the next sections for each dependent variable individually. It has to be pointed out once more that all statistical analysis can only be performed on raw frequencies (observed number of tokens in the data set) and not on normalized frequencies (instances per ten million words).

4.2.1 Noun

The chi-square table below displays whether the observed distribution of the tokens between the translated and the non-translated category (i.e. the distribution within the data set) fits the expected distribution (i.e. the null hypothesis, see also section 3.4.3).

Table 4.16: Chi-Square test ***noun/translated***

	value	df	*p*-value
Pearson Chi-Square	94.398	13	.000
N of Valid Cases	6715		

The test of the null hypothesis stating that the nouns are equally distributed between translated and non-translated tokens shows significance beyond the .05 level: $x^2(13) = 94.398$; $p < 0.001$. Since the chi-square test is not valid for categories with an expected count of less than five tokens per value, the data had to be cleared of these cases in advance (see appendix H for the complete cross-tabulation). The revised cross-tabulation for the variable noun which forms the basis of the chi-square table above is given in table 4.17. The nouns *age*, *May*, *century*, *future*, *Sunday* and *decade* had to be excluded because they have a value equal to or less than five in at least one cell. There remain 14 nouns in the data set, which is given as the degree of freedom (df) in table 4.14 (- 1). These 14 nouns are represented in 6715 tokens of the data set. Thus, the excluded nouns account for 135 tokens, which is a rather small number.

Table 4.17: Revised cross-tabulation ***noun/translated***

noun		**translated**		Total
		yes	no	
time	Count	157.0	20.37	2194.0
	Expected Count	184.6	2009.4	2194.0
	Adjusted Residual	-2.6	2.6	
year	Count	40.0	449.0	489.0
	Expected Count	41.1	447.9	489.0
	Adjusted Residual	-.2	.2	
day	Count	121.0	813.0	934.0
	Expected Count	78.6	855.4	934.0
	Adjusted Residual	5.4	-5.4	
life	Count	38.0	641.0	679.0
	Expected Count	57.1	621.9	679.0
	Adjusted Residual	-2.8	2.8	
night	Count	70.0	556.0	626.0
	Expected Count	52.7	573.3	626.0
	Adjusted Residual	2.6	-2.6	
end	Count	5.0	83.0	88.0
	Expected Count	7.4	80.6	88.0
	Adjusted Residual	-.9	.9	
week	Count	18.0	226.0	244.0
	Expected Count	20.5	223.5	244.0
	Adjusted Residual	-.6	.6	
month	Count	25.0	152.0	177.0
	Expected Count	14.9	162.1	177.0
	Adjusted Residual	2.8	-2.8	
moment	Count	23.0	422.0	445.0
	Expected Count	37.4	407.6	445.0
	Adjusted Residual	-2.6	2.6	
morning	Count	11.0	168.0	179.0
	Expected Count	15.1	163.9	179.0
	Adjusted Residual	-1.1	1.1	
hour	Count	27.0	129.0	156.0
	Expected Count	13.1	142.9	156.0
	Adjusted Residual	4.0	-4.0	
minute	Count	5.0	86.0	91.0
	Expected Count	7.7	83.3	91.0
	Adjusted Residual	-1.0	1.0	
evening	Count	21.0	145.0	166.0
	Expected Count	14.0	152.0	166.0
	Adjusted Residual	2.0	-2.0	
season	Count	4.0	243.0	247.0
	Expected Count	20.8	226.2	147.0
	Adjusted Residual	-3.9	3.9	
Total	Count	565.0	6150.0	6715.0
	Expected Count	565.0	6150.0	6715.0

Table 4.17 describes the raw frequency distribution of the nouns between translated and non-translated tokens including the statistically relevant variables *observed count*, *expected count* and *adjusted residuals* (see section 3.4.3) for every cell (i.e. noun). The noun *time* in the first row for example displays an observed count of 157 tokens for the translated and 2037 tokens for the non-translated category as well as an expected count of 184.6 tokens for the translated and 2009.4 tokens for the non-translated category. This means that if the null hypothesis was true (no association between the variables **noun** and **translated**), the data set would have exhibited the expected values. Instead, the observed count differs from these values.

The size of the deviation between observed and expected counts is assessed by the adjusted residual. A quantitative value equal to or greater than 2/-2[18] refers to a considerable deviance between the two counts, meaning that the number of tokens is either significantly larger or smaller than the expected count. In the case of *time*, the adjusted residual of -2.6 for the translated category indicates that the deviation is significant and that there are fewer tokens in the data set than would be expected if the two variables were unrelated. In contrast, the positive value of 2.6 for the non-translated category indicates that there are more tokens than expected. In like manner, there are positive or negative deviances for the nouns *day*, *life*, *night*, *month*, *moment*, *hour*, *evening* and *season*. In consequence, the null hypothesis can be rejected for these cases. The remaining five nouns (*year*, *end*, *week*, *morning* and *minute*) do not exhibit enough deviation to be significantly different from the null hypothesis. It has to be pointed out though, that negative deviation does not necessarily occur for the translated category and positive deviation for the non-translated category (as in the example above). In fact, *day*, *night*, *month*, *hour* and *evening* display the opposite in the table above. There are more translated and fewer non-translated tokens for those nouns. The distribution between significant and insignificant deviations of the nouns between the two categories is displayed once more in the table below:

[18] The value is actually given as 1.96 in the literature. However, by convention and since I am not working with more than one decimal, 2 is adequate for this study.

*Table 4.18: Distribution of significant and insignificant deviations for the variable **noun** between the categories*

Significant deviation in the translated category		**Insignificant deviation in the translated and non-translated category**
overrepresentation	*underrepresentation*	
day	time	year
night	life	end
month	moment	week
hour	season	morning
evening		minute

Table 4.18 gives an overview of the distribution of the nouns displaying significant deviation (left-hand column) and lists the nouns that do not show significant deviation in the right-hand column. The left-hand column is divided into the nouns that display significant positive deviation (overrepresentation) from the expected count in the translated category and the nouns that show significant negative deviation (underrepresentation) from the expected count. Simultaneously, overrepresentation in the translated category means underrepresentation in the non-translated category and vice versa. Underrepresentation in the translated category means overrepresentation in the non-translated category. The noun *day* for example exhibits significantly more counts than expected in the translated category and thus fewer counts than expected in the non-translated category. In contrast, the noun *life* displays fewer counts than expected in the translated category and is thus overrepresented in the non-translated category. According to this table, there are five nouns which display insignificant deviation from the null hypothesis, five nouns which are significantly overrepresented in the translated category and four nouns which are significantly underrepresented in the translated category.

However, it is not sufficient to statistically reject the null hypothesis (chi-square value) and determine the significance of the differences between observed and expected counts (adjusted residuals). It is also necessary to identify the strength of the association between the two variables. For cross-tabulations larger than 2 x 2, Cramer’s V is a suitable test to measure association related to the size of the sample, with a value of 0 indicating no association and 1 denoting strong association. The test result for the given variables is displayed in table 4.19.

Table 4.19: Cramer's V test for ***noun/translated***

	value
Cramer's V	.119
N of valid cases	6715

The observed significance level value is 0.119. The continuative evaluation of Cramer's V by transferring it into Cohen's *w* (see section 3.4.3) is not necessary since both values are the same[19]. Cohen's effect size index (Kinnear and Gray 2010:414) evaluates the association between the variables **noun** and **translated** as weak since *w* <.3.

4.2.2 Verb

The revised cross-tabulation for expected counts equal to or larger than five is given in table 4.22 on the next page[20]. The chi-square test results resting upon that table are given in table 4.20:

Table 4.20: Chi-square test ***verb/translated***

	value	df	*p*-value
Pearson Chi-Square	45.849	7	.000
N of Valid Cases	6789		

The test of the null hypothesis that the verbs are equally distributed between the translated and the non-translated category shows significance beyond the .05 level: $x^2(7) = 45.849$; $p < 0.001$. Hence, there is a relation between the verbs being translated or not. The test of strength of association reveals a value of $w = 0.082$, suggesting trivial association ($w < 1$):

Table 4.21: Cramer's V test for ***verb/translated***

	value
Cramer's V	.082
N of valid cases	6789

[19] The equation for Cohen's index in this case is as following: $w = 0.119 \sqrt{2-1}$. Accordingly, $w = 0.119$. Since the result of the equation for the square root always will be 1 in my study, $w = V$ applies for all effect size measures in this study.

[20] The cross-tabulation including all verbs is given in appendix I.

Table 4.22: Revised cross-tabulation verb/translated

verb		translated		Total
		yes	no	
go	Count	236.0	2523.0	2759.0
	Expected Count	230.8	2528.2	2795.0
	Adjusted Residual	.5	-.5	
come	Count	208.0	2526.0	2734.0
	Expected Count	228.7	2505.3	2734.0
	Adjusted Residual	-1.9	1.9	
run	Count	6.0	234.0	240.0
	Expected Count	20.1	219.9	240.0
	Adjusted Residual	-3.3	3.3	
fall	Count	49.0	432.0	481.0
	Expected Count	40.2	440.8	481.0
	Adjusted Residual	1.5	-1.5	
return	Count	10.0	53.0	63.0
	Expected Count	5.3	57.7	63.0
	Adjusted Residual	2.2	-2.2	
arrive	Count	30.0	186.0	216.0
	Expected Count	18.1	197.9	216.0
	Adjusted Residual	3.0	-3.0	
fly	Count	22.0	104.0	126.0
	Expected Count	10.5	115.5	126.0
	Adjusted Residual	3.7	-3.7	
roll	Count	7.0	163.0	170.0
	Expected Count	14.2	155.8	170.0
	Adjusted Residual	-2.0	2.0	
Total	Count	568.0	6221.0	6789.0
	Expected Count	568.0	6221.0	6789.0

Table 4.22 above summarizes the distribution of the verbs that are included in the statistical analysis, i.e. that have an expected count of five or higher. The verbs *leave*, *rise*, *enter*, *travel* and *jump* had to be excluded because they did not meet this requirement (see complete cross-tabulation in appendix I). Thus, there remain eight verbs to be statistically evaluated. These eight verbs produce the results given in tables 4.20 and 4.21 above. The adjusted residuals for the verbs *run*, *return*, *arrive*, *fly* and *roll* suggest significant positive or negative deviation from the expected count. *Run* and *roll* exhibit fewer translated and more non-translated tokens whereas *return*, *arrive* and *fly* display more translated and fewer non-translated tokens than the expected count. For the verbs *go*, *come* and *fall* the adjusted residual value does not exceed 2/-2 and does thus not suggest significant deviation from the null hypothesis (see table 4.23).

*Table 4.23: Distribution of significant and insignificant deviations for the variable **verb** between the categories*

Significant deviation in the translated category		**Insignificant deviation in the translated and non-translated category**
overrepresentation	*underrepresentation*	
return	run	go
arrive	roll	come
fly		fall

The table above clarifies once more what was just described about the significant deviations of the verbs from the respective expected counts. The verbs *go*, *come* and *fall* do not display enough deviation to differ significantly from the distribution if the null hypothesis was true. The verbs *return*, *arrive* and *fly* are significantly overrepresented in the translated category and thus significantly underrepresented in the non-translated category. Finally, the two verbs *run* and *roll* are significantly underrepresented in the translated category and thus overrepresented in the non-translated category. Of the eight verbs that could be included in the statistical evaluation, three are insignificantly different from the null hypothesis, three are overrepresented in the translated category and two are underrepresented.

4.2.3 Tense

For the variable **tense**, it was not necessary to revise the cross-tabulation for expected counts less than five since there are only three verb forms included in this study and thus the observed as well as the expected token counts are fairly high:

*Table 4.24: Cross-tabulation **tense/translated***

verb form		**translated**		Total
		yes	no	
past	Count	389.0	3327.0	3716.0
	Expected Count	309.8	3406.2	3716.0
	Adjusted Residual	7.0	-7.0	
present	Count	161.0	2642.0	2803.0
	Expected Count	233.7	2569.3	2803.0
	Adjusted Residual	-6.5	6.5	
infinitive	Count	21.0	310.0	331.0
	Expected Count	27.6	303.4	331.0
	Adjusted Residual	-1.3	1.3	
Total	Count	571.0	6279.0	6850.0
	Expected Count	571.0	6279.0	6850.0

The deviation between observed and expected counts is not significant for the category infinitive (adjusted residual of -1.3/1.3). In contrast, the adjusted residuals for past and present reflect considerable deviation of -6.5/6.5 for the present tense and 7/-7 for the past tense. The present tense tokens exhibit fewer translated and more non-translated tokens while the past tense tokens contain more translated and fewer non-translated tokens than expected. The chi-square test for the two variables tense and translated results in statistical significance as demonstrated in table 4.25:

Table 4.25: Chi-square test ***tense/translated***

	value	df	*p*-value
Pearson Chi-Square	48.478	2	.000
N of valid cases	6850		

The test of the null hypothesis that the tenses of the verb forms are equally distributed between the translated and the non-translated category shows significance beyond the .05 level: $x^2(2) = 48.478$; $p < 0.001$. The strength of the association between the two variables is statistically evaluated to be trivial since the value is less than .1 (0.084):

Table 4.26: Cramer's V test for ***tense/translated***

	value
Cramer's V	.084
N of valid cases	6850

4.2.4 Genre

Since the variable **genre** only contains two classifications (fictional and non-fictional), the cross-tabulation with the variable **translated**, which consists of two categories (translated and non-translated), results in a 2 x 2 table. In general, the chi-square test is not considered to be reliable for such tables and Fisher's Exact test recommended instead. The strength of associations for such tables is evaluated by the Phi coefficient. Table 4.27 displays the observed and expected distribution of the tokens and the significance of potential deviances. Table 4.28 gives the results of Fisher's Exact test and Phi in table 4.29 evaluates the strength of the association.

Table 4.27: Cross-tabulation ***genre/translated***

genre		**translated** yes	no	Total
Non-Fiction	Count	102.0	3361.0	3463.0
	Expected Count	288.7	3174.3	3463.0
	Adjusted Residual	-16.3	16.3	
Fiction	Count	469.0	2918.0	3387.0
	Expected Count	282.3	3104.7	3387.0
	Adjusted Residual	16.3	-16.3	
Total	Count	571.0	6279.0	6850.0
	Expected Count	571.0	6279.0	6850.0

In 2 x 2 contingency tables all values for the adjusted residual 'will have the same absolute value, but exactly 2 of them will be negative'[21]. In the table above all four cells display -16.3/16.3 which indicates statistical significance. Fictional tokens exhibit a negative deviation from the expected count within the non-translated category while non-fictional tokens exhibit the negative deviation within the translated category.

Table 4.28: Fisher's Exact test for ***genre/translated***

	value	df	*p*-value	*p*-value
Pearson Chi-Square	266.327	1	.000	
Fisher's Exact Test				
N of valid cases	6850			.000

Also Fisher's Exact test evaluates the possibility that the two variables are associated as highly significant. The exact *p*-value is less than 0.001.

Table 4.29: Phi test for ***genre/translated***

	value
Phi	.197
N of valid cases	6850

As with the three preceding variables, the strength of association is evaluated as weak and the observed statistical value (0.197) given as close to 0.

[21] https://www304.ibm.com/support/docview.wss?uid=swg21479605 7.11.2011

5 DISCUSSION

This chapter aims to bring together the results of the analyses presented in chapter four and the question of whether or not the quantitative distribution of the values of every dependent variable (**noun**, **verb**, **tense** and **genre**) between translated and non-translated tokens is affected by the variable **translated**. If so, what does this influence look like? Is the translated category over- or underrepresented compared to the respective non-translated category of the variables, i.e. are metaphorical expressions of the cognitive metaphor TIME IS MOTION more often used within translated or within non-translated English? Or are they equally distributed? At the end of the chapter, I use these results to evaluate my hypothesis in how far translational language differs from original language in terms of the usage of lexical expressions of TIME IS MOTION. The chapter discusses every variable individually before summarizing them in a fifth section and relating them to the hypothesis.

5.1 Noun

The descriptive analysis of the variable **noun** in the preceding chapter reveals that the usage of nouns in metaphorical expressions of TIME IS MOTION differs significantly between the translated and the non-translated category. Translated tokens outnumber the respective non-translated tokens of a noun in almost all of the cases (see table 4.5 and figure 4.3). The noun *time* for example occurs 157 times per ten million words in translated texts but only 39.9 times in non-translated texts. Phrases containing a metaphorical usage of the noun *night* are used approximately seven times more per ten million words in translated (70) than in non-translated texts (10.9). Only the nouns *season, future, century* and *decade* are less frequent in translated than in non-translated language. In addition to quantitative deviations between the single nouns in the translated and the non-translated category, the quantitative distribution of the nouns within each of the categories also differs though the correlation coefficient is high indicating high correlation (see table 4.7). This results in deviant ranked orders of the nouns within translated and non-translated language (see table 4.6). Only the nouns *time* and *day* in first and second position respectively in both translated and non-translated language concur. Afterwards, the order of the nouns in translated and non-translated metaphorical language differs. Consequently, some of the nouns (as used in metaphorical expressions of TIME IS MOTION) are used more often in translated language than others while the same nouns are used less often within non-translated language or vice versa. Within translated language for example, the noun *hour* is more frequently used than *month,* while it is less frequent within non-translated language. Conversely, *life* is less frequent within translated language than *night* but more frequent within non-translated language. An additional quantitative analysis involving more variables might shed more light on the deviation of the nouns between translated and non-translated language as well as the differing ranked orders. Adding, for example, the variable **verb** to the variable **noun** (that is to say studying the nouns in combination with the verbs) might reveal more information about the usage of the nouns in translated language. There might be some verbs that are quantitatively dominating metaphorical expressions with certain nouns and these collocational patterns might give a further insight into the nature of over- and underrepresented nouns in translated language.

Statistically, the association between the variables **noun** and **translated** is significant and the divergence between translated and non-translated usages of the nouns not random. For 13 of the 20 nouns, there is an either positive or negative (statistically relevant) divergence from the expected count (if the null hypothesis was true that there is no relation between the two variables). However, for five of the 13 nouns the deviation is not large enough to be statistically relevant. There are only nine nouns which are estimated to differ significantly: five are overrepresented in translated language and four are underrepresented. Related to the sample size (the data set), the association between the two variables is evaluated as rather weak (see table 4.18). However, the descriptive and the statistical analysis reveal a relation between the usage of the nouns in metaphorical expressions and their origin in translated or original language. In conclusion, it can be stated that in relation to the usage of at least 13 of the 20 nouns in metaphorical expression of TIME IS MOTION translated language differs from non-translated language, some more and some less. Some are overrepresented and some are underrepresented in translated language.

5.2 Verb

Also for the variable **verb**, the descriptive analysis reveals divergence between the translated and the non-translated category. First of all, ten of the thirteen verbs in the data set produce more metaphorical instances per ten million words in translated than in original language. The remaining three verbs (*rise, travel* and *jump*) did not produce any translated tokens in the corpus search of the TEC and the respective normalized frequencies for the non-translated tokens are close to zero (see table 4.11). This might be due to different reasons. Firstly, the TEC might not contain the necessary amount of language (words) to include instances of these three verbs used in metaphorical expressions in combination with one of the twenty nouns. Secondly, these verbs might not be used in respective metaphorical expressions in translated language. The latter reason is interesting in so far as an additional qualitative analysis of the tokens including source languages might reveal interesting relations between the source language and English regarding culture-overlapping and culture-specific cognitive metaphorical mappings. Furthermore, both possibilities could be studied through the inclusion of more and different instances of translated language. In the present data set, the majority of verbs are over-represented in translated language compared to the corresponding non-translated verbs (see figure 4.4). Moreover, the usage of the translated verbs follows a different ranked order than the non-translated verbs. While the non-translated verbs follow the same ranked frequency distribution as the data set (see table 4.9), the translated tokens diverge from this order. That the ranked order of the non-translated tokens is similar to the ranked frequency order of the data set is not surprising since non-translated tokens constitute the majority of the tokens in the data set. However, that the translated tokens diverge from this order is another indicator of different usage of TIME IS MOTION within translated and non-translated language. The descriptive analysis reveals that some of the verbs are more frequently used within the translated category than within the non-translated category in comparison to other verbs that are precedent in the ranked order. This indicates that some of the translated verbs are metaphorically preferred over others in contrast to their non-translated counterparts. Of course, there is no apparent reason in the data to generalize into all verbal use in expressions of TIME IS MOTION and further research is necessary to find out more about this tendency. The facts stated above are only true for my data set and the methodology applied to generate this data.

Statistically, the association of the variables **verb** and **translated** in the data set is significant, meaning that the distribution of the tokens between the translated and the non-translated category is not due to random dissemination but to their affiliation to one of the categories. For at least eight of the 13 verbs that produced metaphorical tokens in the corpus search, there is a significant deviation (either positive or negative) from what would have been the distribution if it would have been random (i.e. the null hypothesis). Three of the verbs are significantly over-represented in translated language, two are significantly underrepresented and three more verbs do not differ enough from the expected count to be statistically significant. The test for strength of association evaluates the association as rather weak but does not reject it. Hence, also the variable **verb** is subject to some kind of influence of the variable **translated**, i.e. the usage within translated and/or non-translated language.

5.3 Tense

In accordance with the two preceding variables **noun** and **verb**, the variable **tense** also exhibits a considerable divergence of normalized instances between translated and non-translated tokens. In all three verb forms (infinitive, present and past tense), translated language contains far more metaphorical instances of TIME IS MOTION per ten million words than non-translated language (see table 4.13). Interestingly, both varieties – translated and original English – exhibit the same ranked order of the three forms, with past tense being quantitatively most frequent followed by present tense and infinitive forms. This suggests that metaphorical expressions of TIME IS MOTION, both in translated and in non-translated English, are preferably produced in past tense and least often in any form of infinitive construction. This is an interesting finding in the data set and it is necessary (but impossible in the course of this particular paper) to investigate whether this preference for past tense verb forms can also be found in connection with other cognitive metaphors.

The chi-square analysis conducted in section 4.2.3 reveals that the association between the variable **tense** and the variable **translated** is statistically significant. In other words, the distribution of past tense, present tense and infinitive verb forms between translated and non-translated cases is not random but somehow influenced by the fact that the cases are either translated or not. This is first and foremost evident for the past tense and the present tense form, where the divergence between the expected count and the actual observed count is assessed with an adjusted residual of 7.0/-7.0 and -6.5/6.5 respectively. To recall, an adjusted residual equal to or higher than -2/2 indicates significant deviation. For contingency tables larger than 2 x 2 applies the rule the higher the value, the more significant is the deviation. Interestingly, translated cases are overrepresented in past tense forms but underrepresented in present tense forms related to the respective non-translated categories. Thus, the deviation from the null hypothesis for both categories is still largest within past tense, but translated language exhibits positive deviation while non-translated language displays negative deviation. In concordance with the previous two variables, the divergence between translated and non-translated cases within the data set indicates a difference between translated and original English in relation to the use of the cognitive metaphor TIME IS MOTION.

5.4 Genre

The variable **genre** contains only two classifications, i.e. fiction and non-fiction. The distribution of the tokens between those two within the data set is fairly balanced. However, this does not apply for the distribution between the translated and the non-translated category.

Translated fictional and non-fictional tokens are more frequent than their non-translated counterparts. Thus, the translated category outnumbers the non-translated category.

It is also noticeable that the distribution of non-translated tokens between fictional and non-fictional texts is almost balanced (see table 4.15) with 66 non-fictional and 57 fictional tokens, whereas there is a significant difference within the translated tokens with 102 non-fictional and 469 fictional cases. There is a considerable overrepresentation of fictional tokens in translated language. This suggests that within translated English, fictional texts are more likely to contain metaphorical expressions of TIME IS MOTION than in original English language. Still, this finding has to be treated with caution since this divergence might be due to the quantitative overrepresentation of fictional tokens within the translated category of the data set (see figure 4.7). Fictional tokens (469) dominate the translated category in relation to the non-fictional tokens (102). The question remains whether this is due to an imbalance between fictional and non-fictional texts within the TEC or actually points towards different usage of metaphorical expressions within translated and non-translated texts. Unfortunately, the TEC does not provide any information as to the frequency distribution of words between the different sub-corpora so that I cannot make any claims about a possible imbalance in the corpus. I can only refer to the findings that result from the data set and acknowledge that these findings are highly dependent on the composition of the corpus, which I am not satisfactorily familiar with. More information about the TEC or further studies on the same or other data material have to take a closer look at the influence of the variable **genre** on the distribution of metaphorical expressions within translated language.

Statistically, the association between the variable **genre** and **translated** is significant and the divergence of the tokens from the expected distribution considerably large. Hence, the distribution is noticeably different from purely random distribution and determined by the variable **translated**, i.e. dependent on the fact whether the token is translated or not. However, in accordance with the preceding three dependent variables, the association between **genre** and **translated** is statistically evaluated as weak. Notwithstanding this weak association, the two variables are related and the descriptive analysis reveals overrepresentation in the translated category for both fictional and non-fictional texts indicating that in connection with cognitive metaphorical expressions of TIME IS MOTION translated language differs from non-translated language regarding genre as well.

5.5 Summary

Concluding, all the four dependent variables (**noun**, **verb**, **tense** and **genre**) are statistically associated to the independent variable **translated** and hence translated language differs quantitatively from non-translated language. Firstly, within all the four dependent variables, the translated category is quantitatively overrepresented in relation to the respective non-translated category per ten million words. Thus, according to the analyses, translated language does employ more metaphorical expressions of TIME IS MOTION than non-translated language. Secondly, within three of the four variables the usage of the different constituents (the 20 nouns, the 15 verbs and the two genres) differ regarding the ranking within the translated and the non-translated category. In other words, in translated language some nouns and verbs occur more often than others compared to in non-translated language. Additionally, translated language employs more metaphorical expressions in fictional texts while non-translated language appears to be marked by a slight overrepresentation in non-fictional texts. Only in relation to the tense of the verbs does translated and non-translated language concur, preferring past tense over present tense and infinitive forms. However, quantitative overrepresentation within translated language dominates the analyses and it can thus be stated that translated language differs from non-translated language by employing more expressions of the cognitive metaphor TIME IS MOTION.

6 CONCLUSION

Metaphor is a cognitive phenomenon assumed to be closely related to all human conceptualization and languaging including translation. My claim in this study has been that there are quantitative differences concerning the usage of metaphorical expressions of cognitive metaphors in texts that are originally produced in English and texts that are translated from languages into English. This claim was investigated on metaphorical expressions representing the cognitive metaphor TIME IS MOTION. Researchers within translation studies (e.g. Baker 1993) have long established the fact that translated texts (i.e. target texts) distinguish themselves from comparable non-translated texts through a number of features such as the over-use of conventional grammar and exaggeration of target language features (Baker 1993:244). Originally, these features applied to traditional linguistics categories like lexis, syntax, semantics etc. In recent years, investigations into translation norms have also included cognitive linguistic phenomena like cognitive metaphors. Previous studies on cognitive metaphors in translation have been occupied with the qualitative analysis of source and target text expressions mainly investigating semantic equivalence or deviation (Schäffner 2004, Al-Hasnawi 2007). Studies of this kind aim to investigate translational processes and strategies and draw conclusions about the nature of translation or even about empirical theories of other scientific disciplines (linguistics, psychology etc.). However, there has, to my knowledge, not yet been a study investigating possible quantitative deviation between original and translated English texts concerning the usage of cognitive metaphorical expressions. This is the aim of this study.

Approaching translations quantitatively using the comparable corpus method (as done in this study) helps to establish that there indeed is a divergence between the translated and the non-translated variety of English. Regarding four different variables (i.e. **noun**, **verb**, **tense** and **genre**) there is a significant deviation between the translated and the non-translated tokens per ten million words. Thus, regarding my research questions, I can conclude that:

1. Translated language differs quantitatively from non-translated language regarding the use of metaphorical expressions of the cognitive metaphor TIME IS MOTION.
2. Translated text exhibits overrepresentation of metaphorical expressions of the cognitive metaphor TIME IS MOTION per ten million words regarding the four variables investigated in this study.

a. There are more instances of metaphorical expressions employing certain nouns and verbs per ten million words.
b. The usage of the verbs in the three investigated verb forms (infinitive, present and past tense) within the translated tokens differs quantitatively from the respective non-translated tokens. In other words, the quantitative overrepresentation of the variable **verb** continues throughout the syntactically related variable **tense**.
c. The fourth variable **genre** also exhibits overrepresentation of the translated tokens. Interestingly, the translated category displays an inner quantitative deviation of the distribution of translated tokens between fictional and non-fictional tokens which the non-translated category does not.

Hence, just as I expected, the quantitative usage of metaphorical expressions of the cognitive metaphor TIME IS MOTION in translated English texts differs from non-translated text by over-representation. In other words, this study comes to the conclusion that with regard to Mona Baker's translation universal of exaggeration (1993), metaphorical expressions of the cognitive metaphor TIME IS MOTION appear in translated English language and are exaggerated by over-use. However, the statistical analysis reveals differences in deviations from the null hypothesis for the different values of the four dependent variables and association between each dependent variable and the independent variable **translated** is evaluated as rather weak. Further studies and a refined methodological approach might shed more light on the problem.

Concerning the methodology employed in this study, I acknowledge that there are problems with the representativeness of at least one of the corpora employed to extract tokens representing the different varieties. Firstly, the TEC is relatively small compared to the other two corpora representing the non-translated variety of English. This in itself does not pose an overarching problem since the figures are compared on the basis of normalized frequencies per ten million words. Additionally however, the TEC does not provide enough information to assess whether the distribution of words between the different sub-corpora is balanced as a corpus should be to be representative of a certain language or language variety. This way, there might be a considerable imbalance between the fictional and the non-fictional texts already in the corpus causing the respective deviation within the data set. I also acknowledge the fact that the corpora chosen to represent original English in this study might also contain translated texts. The sub-corpus newspaper within the COCA and the BNC for example might contain newspaper articles that are translated into English from other languages. Unfortunately, the corpora do not include any information about this possibility so I have to treat the language population represented in the two corpora as non-translated. Further research on the subject is necessary to eliminate such potentially compromising factors, providing more empirical control over the setting of the investigation. Compiling one's own corpora of translated and non-translated English instead of the usage of pre-prepared online corpora appears to be one solution to the problems and is recommended for future studies on the topic.

Further research on the subject can take a number of different directions to test the findings of this study. First of all, continuing with the four variables included in this investigation, a combination of two, three or even all four of them can be tested against the expected influencing variable **translated**. This way, one can examine how the distribution of the tokens evolves when they are treated as phrases including a noun and a verb (e.g. *time flies*) and not only phrases containing either a noun (e.g. time) or a verb (e.g. *flies*). Further on, these phrases can be studied regarding the verb forms and the genre of the texts they are extracted from. Thus, it can be investigated if the quantitative deviation of the tokens between translated and non-translated language changes when the phrase *time flies* is investigated as an instance of a present tense phrase from a non-fictional text. Are there significant differences between the two varieties? Is the translated category still overrepresented or does the merging of variables cause a different distribution? Since the data in the data set is coded for additional information (e.g. number of the noun/verb, corpus), several variables could be included in the analysis to investigate whether any of them might change the prevailing deviation between the distributions. Secondly, different cognitive metaphors including their respective metaphorical expressions have to be examined to confirm the fact that the imbalance between translated and non-translated texts is not only due to the metaphor TIME IS MOTION but to cognitive metaphors in general. Thirdly, when establishing a broader and more controlled source of data, other text types should be included. Since this study only includes written texts in both the translated and the non-translated category, the inclusion of spoken texts (i.e. translated texts that originate in interpreting) is a possibility. This might for example be realized by the application of other research methodologies, like experimental settings, where interpreters are asked to translate speeches prepared by the researcher. Furthermore, translated tokens representing particular values of the variables (certain nouns and verbs etc.) that differ significantly from the non-translated variety can be investigated regarding the information about the nature of the translation, that is to say the source language, the translation mode, the translator employment etc. Are there certain source languages that produce more over-representation than others or is the employment of the translator causing these deviations? Last but not least, a qualitative analysis of the tokens concerning the employment of particular translation strategies from the source language to the target language might be investigated where possible. There is one more interesting finding that emerged from this study, namely that metaphorical expressions of TIME IS MOTION seem to be used more often in past tense in both translated and non-translated English. This is more a subject of metaphor studies within cognitive linguistics than within translation studies. However, a study on different cognitive metaphors might reveal whether this is a phenomenon alluding to metaphorical expressions of cognitive metaphors in general or only to the investigation of TIME IS MOTION in the present study. In both cases, further research can shed more light on possible reasons for this preference of certain verb forms.

As this last paragraph on further research has pointed out, this investigation is in no way meant to be complete or finished. Firstly, the methodological approach of quantitatively investigating differences between the translated and the non-translated variety of English is considered to be an introductory study approaching the phenomenon in question from a wider perspective, providing an opening for further, more detailed studies. Secondly, due to limitations on time, space and the source of the data (at least for the translated variety), the results of the analysis are first and foremost valid for this particular data set. Further and more detailed studies, as described above, are necessary before generalizations into the whole variety of translated language are possible. The findings of this particular study show a quantitative difference between translated and non-translated language for the cognitive metaphor TIME IS MOTION.

7 APPENDICES

Appendix A

time

future

past

day

week

month

year

decade

Christmas

Thanksgiving

Easter

today

yesterday

tomorrow

moment

second

minute

hour

period

Appendix B

Table 7.1: List of nouns including semantic definition and time relation

	term	definition	time relation
1	afternoon	the part of the day between noon and evening	daytime
2	age	a time of life (usually defined in years)	time of life
3	century	a period of 100 years	time period
4	dark	the time after sunset and before sunrise while it is dark outside	time period
5	dawn	the first light of day	time of day
		an opening time period	time period
6	date	a particular but unspecified point in time	point in time
7	day	time for earth to make a complete rotation on its axis	time unit
		some point or period in time	time
		the time after sunrise and before sunset while it is light outside	time period
8	daylight	the time after sunrise and before sunset while it is light outside	time period

9	daytime	the time after sunrise and before sunset while it is light outside	time period
10	decade	a period of 10 years	time period
11	early days	an early period of time	time period
12	end	the point in time at which something ends	point in time
13	epoch	a period marked by distinctive character or reckoned from a fixed point or event	time period
14	era	a period marked by distinctive character or reckoned from a fixed point or event	time period
15	eve	the period immediately before something	time period
		the latter part of the day (the period of decreasing daylight from late afternoon until nightfall)	daytime

16	evening	the latter part of the day (the period of decreasing daylight from late afternoon until nightfall)	daytime
		a later concluding time period	time period
		the early part of night (from dinner until bedtime) spent in a special way	time period
17	future	the time yet to come	time
18	history	the aggregate of past events	past times
19	hour	a period of time equal to 1/24th of a day	time unit
		clock time	time
		a special and memorable period	time period
20	instant	a particular point in time	point in time
21	life	the period during which something is functional (as between birth and death)	time period
		the period between birth and the present time	time period
		the period between the present until death	time period

22	maturity	the period of time in your life after your physical growth has stopped and you are fully developed	time of life
23	midday	the middle of the day	time of day
24	midnight	12 o'clock at night; the middle of the night	time of day
25	minute	a unit of time equal to 60 seconds of 1/60th of an hour	time unit
		an indefinetly short time	time
		a particular point in time	point in time
26	moment	a particular point in time	point in time
		an indefinetly short time	time
27	morn	the time period between dawn and noon	time period
28	morning	the time period between dawn and noon	time period
		the first light of day	time period

29	month	one of the twelve divisions of the calendar year	time period
		a time unit of approximately 30 days	time unit
30	night	the time after sunset and before sunrise while it is dark outside	time period
		a period of ignorance or backwardness or gloom	time period
		the period spent sleeping	time period
		the dark part of the diurnal cycle considered a time unit	time unit
		the time between sunset and midnight	time unit
31	nighttime	the time after sunset and before sunrise while it is dark outside	time period

32	noon	the middle of the day	time of the day
33	overtime	playing time beyond regulation, to break a tie	time period
34	past	the time that has elapsed	time
		an earlier period in someone's life (especially one that they have reason to keep secret)	time period
35	past times	the time that has elapsed	time
36	period	the interval taken to complete one cycle of a regularly repeating phenomenon	time interval
37	phase	any distinct time period in a sequence of events	time period
38	present	the period of time that is happening now; any continuous stretch of time including the moment of speech	time

39	season	a period of the year marked by special events or activities in some field	time period
		one of the natural periods into which the year is divided by the equinoxes and solstices or atmospheric conditions	time period
		a recurrent time marked by major holidays	time period
40	second	1/60th of a minute; the basic unit of time adopted under the Systeme International d'Unites	time unit
		an indefinetly short time	time
		a particular point in time	point in time
41	semester	half a year; a period of 6 months	time period
42	term	a limited period of time	time period

43	time	a period of time considered as a resource under your control and sufficient to accomplish something	time period
		an indefinite period (usually marked by specific attributes or activities)	time period
44	tomorrow	the near future	time to come
45	trimester	a period of three months; especially one of three three-month periods into which human pregnancy is divided	time period
46	week	any period of seven consecutive days	time period
		a period of seven consecutive days starting on Sunday	time period

47	weekend	a time period usually extending from Friday night through Sunday; more loosely defined as any period of successive days including one and only one Sunday	time period
48	while	a period of intermediate lenght (usually short) marked by some action or condition	time
49	year	a period of time containing 365 (366) days	time period
		a period of time occupying a regular part of a calendar year that is used for some particular activity	time period
		the period of time that it takes for a planet to make a complete revolution around the sun	time period
50	yesterday	the recent past	past times
51	youth	the time of life between childhood and maturity	time of life
		an early period of development	time period

Appendix C

Table 7.2: Rank normalized frequencies per ten million words TIME nouns COCA (lemmatized)

nn1 = singular common noun
nn2 = plural common noun
nnt1 = temporal noun, singular
nnt2 = temporal noun, plural
npd1 = singular weekday noun
npd2 = plural weekday noun
npm1 = singular month noun
npm2 = plural month noun

noun	PoS	Normalized frequencies per ten million words	Observed frequency
year		**19602**	**803707**
year	*nn1/nnt1*	*7891*	*323541*
years	*nnt2*	*11711*	*480166*
time		**19549**	**801513**
time	*nn1/nnt1*	*15980*	*655190*
times	*nn1/nnt2*	*3568*	*146323*
day		**11042**	**452741**
day	*nn1/nnt1*	*7307*	*299593*
days	*nn1/nnt2*	*3735*	*153148*
life		**8483**	**347833**
life	*nn1*	*7123*	*292045*
lives	*nn1/nn2*	*1360*	*55788*
week		**5109**	**209505**
week	*nn1/nnt1*	*3411*	*139887*
weeks	*nn1/nnt2*	*1698*	*69618*
night		**4685**	**192122**
night	*nn1/nnt1*	*4341*	*178011*
nights	*nnt2*	*344*	*14111*
month		**4162**	**170657**
month	*nn1/nnt1*	*1518*	*62268*
months	*nn1/nnt2*	*2570*	*105389*
end		**3419**	**140183**
end	*nn1*	*3147*	*129052*
ends	*nn1/nn2*	*271*	*11131*
minute		**3281**	**134540**
minute	*nn1/nnt1*	*808*	*33166*
minutes	*nn1/nnt2*	*2472*	*101374*
morning		**2924**	**119905**
morning	*nn1/nnt1*	*2843*	*116568*
mornings	*nnt2*	*81*	*3337*
history		**2826**	**115897**
history	*nn1*	*2826*	*115897*

moment		**2798**	**114741**
moment	*nn1/nnt1*	*2305*	*94543*
moments	*nn1/nn2*	*492*	*20198*
age		**2635**	**108057**
age	*nn1*	*2266*	*92911*
ages	*nn2*	*369*	*15146*
hour		**2336**	**95809**
hour	*nn1/nnt1*	*1353*	*55476*
hours	*nn1/nnt2*	*2201*	*90253*
season		**2159**	**88523**
season	*nn1/nnt1*	*1867*	*76579*
term		**1659**	**68059**
term	*nn1/npd1*	*832*	*34128*
terms	*nn1/nn2*	*819*	*33619*
century		**1658**	**68014**
century	*nn1/nnt1*	*1380*	*56589*
centuries	*nnt2*	*278*	*11425*
period		**1632**	**66918**
period	*nn1*	*1386*	*56839*
periods	*nn2*	*245*	*10079*
second		**1444**	**59225**
second	*nnt1*	*851*	*34901*
seconds	*nnt2*	*593*	*24324*
future		**1400**	**57403**
future	*nn1*	*1400*	*57403*
decade		**1382**	**56684**
decade	*nn1/nnt1*	*646*	*26488*
decades	*nn1/nnt2*	*736*	*30196*
evening		**1042**	**42760**
evening	*nn1/nnt1*	*967*	*39654*
evenings	*nnt2*	*75*	*3106*
past		**1006**	**41262**
past	*nn1*	*996*	*40847*
pasts	*nn2*	*10*	*415*
Sunday		**917**	**37608**
Sunday	*nn1/npd1*	*832*	*34128*
Sundays	*npd2*	*84*	*3480*
July		**889**	**36470**
July	*npm1*	*889*	*36457*
Julies	*npm2*	*0*	*13*
June		**879**	**36047**
June	*npm1*	*879*	*36042*
Junes	*npm2*	*0*	*5*
May		**878**	**36031**
May	*nn1*	*874*	*35874*
Mays	*nn2/npm2*	*3*	*157*

March		**876**	**35948**
March	*nn1/npm1*	*876*	*35948*
weekend		**865**	**35482**
weekend	*nnt1*	*731*	*29993*
weekends	*nnt2*	*133*	*5489*
afternoon		**842**	**34539**
afternoon	*nn1/nnt1*	*791*	*32440*
afternoons	*nnt2*	*51*	*2099*
Friday		**823**	**33765**
Friday	*nn1/npd1*	*786*	*32262*
Fridays	*npd2*	*36*	*1503*
April		**812**	**33323**
April	*nn1/npm1*	*812*	*33318*
Aprils	*npm2*	*0*	*5*
Saturday		**803**	**32923**
Saturday	*nn1/npd1*	*742*	*30461*
Saturdays	*npd2*	*60*	*2462*
date		**791**	**32468**
date	*nn1*	*649*	*26634*
dates	*nn2*	*142*	*5834*
while		**756**	**31022**
while	*nn1/nnt1*	*756*	*31007*
whiles	*nnt2*	*0*	*15*
September		**737**	**30223**
September	*nn1/npm1*	*736*	*30200*
Septembers	*npm2*	*0*	*23*
youth		**725**	**29763**
youth	*nn1*	*611*	*25070*
youths	*nn2*	*114*	*4693*
January		**669**	**27457**
January	*nn1/npm1*	*670*	*27471*
Januaries	*npm2*	*0*	*4*
October		**616**	**25294**
October	*nn1/npm1*	*616*	*25267*
Octobers	*npm2*	*0*	*27*
November		**611**	**25055**
November	*nn1/npm1*	*610*	*25037*
Novembers	*npm2*	*0*	*18*
Monday		**609**	**25000**
Monday	*nn1/npd1*	*585*	*23996*
Mondays	*npd2*	*24*	*1004*
August		**605**	**24823**
August	*npm1*	*604*	*24801*
Augusts	*npm2*	*0*	*22*

December		**581**	**23835**
December	*nn1/npm1*	*580*	*23806*
Decembers	*npm2*	*0*	*29*
Tuesday		**549**	**22525**
Tuesday	*nn1/npd1*	*519*	*21300*
Tuesdays	*npd2*	*29*	*1225*
Thursday		**491**	**20152**
Thursday	*nn1/npd1*	*460*	*18895*
Thursdays	*npd2*	*30*	*1257*
era		**490**	**20104**
era	*nn1*	*470*	*19288*
eras	*nn2*	*19*	*816*
February		**480**	**19705**
February	*nn1/npm1*	*480*	*19705*
Wednesday		**454**	**18620**
Wednesday	*nn1/npd1*	*429*	*17625*
Wednesdays	*npd2*	*24*	*995*
dark		**418**	**17163**
dark	*nn1*	*418*	*17163*
phase		**407**	**16694**
phase	*nn1*	*333*	*13675*
phases	*nn2*	*73*	*3019*
present		**332**	**13622**
present	*nn1/np1*	*332*	*13622*
midnight		**213**	**8749**
midnight	*nnt1*	*213*	*8749*
noon		**192**	**7884**
noon	*nn1/nnt1*	*191*	*7871*
noons	*nnt2*	*0*	*13*
dawn		**173**	**7124**
dawn	*nnt1*	*173*	*7124*
instant		**146**	**5989**
instant	*nnt1*	*143*	*5899*
instants	*nnt2*	*2*	*90*
semester		**115**	**4729**
semester	*nn1*	*104*	*4277*
semesters	*nn2*	*11*	*452*
daylight		**88**	**3613**
daylight	*nn1*	*88*	*3613*
maturity		**78**	**3200**
maturity	*nn1*	*78*	*3200*
eve		**75**	**3108**
eve	*nn1/nnt1*	*73*	*3003*
eves	*nnt2*	*2*	*105*
daytime		**65**	**2684**
daytime	*nnt1*	*65*	*2684*

overtime		**55**	**2278**
overtime	*nn1*	*55*	*2278*
midday		**40**	**1675**
midday	*nnt1*	*40*	*1675*
nighttime		**37**	**1547**
nighttime	*nnt1*	*37*	*1547*
epoch		**19**	**787**
epoch	*nn1*	*19*	*787*
trimester		**12**	**518**
trimester	*nn1*	*11*	*479*
trimesters	*nn2*	*0*	*39*
morn		**7**	**304**
morn	*nn1*	*6*	*280*
morns	*nn2*	*0*	*24*
tomorrow		**3**	**123**
tomorrow	*nn1*	*0*	*20*
tomorrows	*nn2*	*2*	*103*
today		**1**	**81**
today	*nn1*	*1*	*81*
yesterday		**0**	**14**
yesterday	*nn1*	*0*	*14*

Appendix D

Table 7.3: Rank normalized frequencies per ten million words for TIME nouns BNC (lemmatized)

NoC = common noun
NoP = proper noun

noun	PoS	normalized frequencies per ten million words
time	NoC	**18330**
time		*15420*
times		*2920*
year	NoC	**16390**
year		*7370*
years		*9020*
day	NoC	**9400**
day		*6100*
days		*3310*
life	NoC	**6450**
life		*5660*
lifes		*10*
lives		*780*
week	NoC	**4760**
week		*3220*
weeks		*1540*
end	NoC	**4580**
end		*4290*
ends		*290*
month	NoC	**3980**
month		*1500*
months		*2480*
night	NoC	**3930**
night		*3650*
nights		*280*
hour	NoC	**3020**
hour		*1130*
hours		*1890*
term	NoC	**2880**
term		*1230*
terms		*1650*
period	NoC	**2830**
period		*2430*
periods		*400*
minute	NoC	**2660**
minute		*820*
minutes		*1830*

moment *moment* *moments*	NoC	**2540** *2210* *320*
age *age* *ages*	NoC	**2520** *2160* *360*
century *century* *centuries*	NoC	**2330** *1970* *360*
morning *morning* *mornings*	NoC	**2190** *2110* *80*
history *history* *histories*	NoC	**2010** *1930* *1930*
date *date* *dates*	NoC	**1770** *1580* *190*
future *future* *futures*	NoC	**1560** *1420* *140*
evening *evening* *evenings*	NoC	**1530** *1380* *150*
May	NoP	**1500**
April	NoP	**1470**
June	NoP	**1460**
March	NoP	**1450**
season *season* *seasons*	NoC	**1220** *1090* *120*
July	NoP	**1190**
October	NoP	**1060**
September	NoP	**1040**
dark	NoC	**1040**
January	NoP	**1020**
Sunday *Sunday* *Sundays*	NoP	**1010** *930* *80*
second *second* *seconds*	NoC	**980** *560* *420*
December	NoP	**940**
November	NoP	**940**

afternoon *afternoon* *afternoons*	NoC	**890** *840* *50*
Saturday *Saturday* *Saturdays*	NoP	**870** *830* *40*
past *past*	NoC	**860** *860*
February	NoP	**840**
August	NoP	**790**
weekend *weekend* *weekends*	NoC	**730** *630* *100*
while	NoC	**630**
youth	NoC	**630**
decade *decade* *decades*	NoC	**620** *370* *250*
Friday *Friday* *Fridays*	NoP	**580** *550* *30*
Monday *Monday* *Mondays*	NoP	**560** *530* *30*
phase *phase* *phases*	NoC	**560** *460* *100*
youth *youths*		**540** *90*
present *present* *presents*	NoC	**500** *410* *90*
Wednesday *Wednesday* *Wednesdays*	NoP	**460** *440* *20*
Thursday *Thursday* *Thursdays*	NoP	**390** *370* *20*
Tuesday *Tuesday* *Tuesdays*	NoP	**370** *360* *20*
era *era* *eras*	NoC	**220** *210* *10*
midnight	NoC	**190**
dawn	NoC	**150**

maturity maturity *maturities*	NoC	**150** *140* *10*
daylight	NoC	**110**

Appendix E

Table 7.4: Ranked normalized frequencies per ten million words for TIME nouns COCA and BNC (lemmatized)

	noun	total normalized frequencies per ten million words COCA/BNC
1	time	1931
2	year	1897
3	day	1072
4	life	808
5	week	504
6	night	453
7	month	412
8	end	364
9	minute	315
10	morning	278
11	moment	274
12	age	261
13	hour	247
14	season	197
15	century	179
16	future	143
17	decade	123
18	evening	113
19	May	100
20	June	99
21	date	98
22	past	97
23	Sunday	93
24	afternoon	85
25	weekend	83
26	Saturday	81
27	youth	70
28	dark	54
29	phase	43
30	era	43
31	midnight	20
32	dawn	16
33	noon	15
34	instant	11
35	maturity	9
36	daylight	9
37	semester	9
38	eve	6

39	daytime	5
40	overtime	4
41	midday	3
42	nighttime	3
43	epoch	1
44	trimester	1
45	morn	0
46	tomorrow	0
47	today	0
48	yesterday	0

Appendix F

Table 7.5: Rank normalized frequencies per ten million words for MOTION verbs COCA and BNC (lemmatized)

	verb	normalized frequencies per ten million words
1	go	24970
2	come	22030
3	leave	9430
4	run	6770
5	fall	4270
6	walk	3740
7	return	3510
8	arrive	2190
9	rise	2180
10	enter	1950
11	fly	1590
12	travel	1560
13	jump	1040
14	roll	1030
15	charge	960
16	cross	930
17	escape	860
18	rush	620
19	slide	590
20	climb	580
21	race	580
22	hop	530
23	sweep	500
24	advance	490
25	float	470
26	swim	450
27	tear	440
28	wander	430
29	bound	390
30	file	390
31	flee	380
32	hurry	380
33	drift	360
34	leap	340
35	march	330
36	speed	320
37	descend	280
38	bounce	280
39	crawl	250
40	stumble	250

41	plunge	240
42	depart	230
43	creep	220
44	stride	150
45	hike	100
46	exit	80
47	tumble	80
48	roam	70
49	glide	60
50	skip	60
51	recede	50
52	dart	50
52	sneak	50
54	stroll	50
55	dash	40
56	jog	40
57	stagger	40
58	stomp	40
59	wade	40
60	zoom	40
61	hasten	30
62	bolt	30
63	inch	30
64	limp	30
65	lurch	30
66	stray	30
67	trudge	30
68	ascend	20
69	clamber	20
70	hobble	20
71	hurtle	20
72	meander	20
73	parade	20
74	prowl	20
75	rove	20
76	shuffle	20
77	streak	20
78	tack	20
79	amble	10
80	bowl	10
81	coast	10
82	gallop	10
83	journey	10
84	lope	10
85	lumber	10

86	mince	10
87	pad	10
88	ramble	10
89	sidle	10
90	skitter	10
91	slither	10
92	slouch	10
93	strut	10
94	stump	10
95	tiptoe	10
96	trek	10
97	trot	10
98	vault	10
99	waddle	10
100	backpack	0
101	canter	0
102	carom	0
103	cavort	0
104	clump	0
105	dodder	0
106	flit	0
107	frolic	0
108	gambol	0
109	goosestep	0
110	lollop	0
111	mosey	0
112	nip	0
113	perambulate	0
114	plod	0
115	prance	0
116	promenade	0
117	romp	0
118	sashay	0
119	skedaddle	0
120	skulk	0
121	sleepwalk	0
122	slink	0
123	slog	0
124	somersault	0
125	swagger	0
126	toddle	0
127	totter	0
128	traipse	0
129	tramp	0
130	troop	0

131	trundle	0
132	whiz	0
133	zagzag	0

Appendix G

Table 7.6: List of variables SPSS

variable	values
numberNOUN	1 = singular 2 = plural
tense	1 = infinitive 2 = present 3 = past
corpus	1 = COCA 2 = BNC 3 = TEC
translated	1 = yes 2 = no
genre	1 = fiction 2 = non-fiction
numberVERB	1 = singular 2 = plural
TranslationGender	1 = Male 2 = Female
TranslatorEmployment	1 = unknown 2 = lecturer 3 = professor 4 = writer 5 = translator 6 = british Council Officer 7 = teacher
SourceLanguage	1 = unknown 2 = German 3 = Arabic 4 = Spanish 5 = Italian 6 = Welsh 7 = Russian 8 = Finish 9 = Modern Greek 10 = French 11 = Serbian 12 = Norwegian 13 = Hebrew 14 = Chinese 15 = Tamil 16 = Polish 17 = Brazilian Portuguese 18 = Hungarian 19 = Slovene 20 = Portuguese 21 = Hopi 22 = Japanese 23 = Thai 24 = Turkish 25 = Swedish 26 = Serbo-Croat 27 = Greek

TranslationMode	1 = unknown	4 = into 2nd language
	2 = into mother tongue	5 = into foreign language
	3 = into language of habitual use	6 = into native language
TranslationPlace	1 = unknown	6 = Belgium
	2 = UK	7 = Madras
	3 = USA	8 = Finland
	4 = Cyprus	
	5 = India	
VerbUse	3 = 3rd person	Time goes by. Time went by.
	4 = question present	Where does the time go?
	5 = question past	Where did the time go?
	6 = future will	Not only will life go on
	7 = make +	make, makes, made, making
	8 = let +	let, lets, let, letting
	11 = do+	do, don't, does, doesn't,did,didn't
	12 = other verb +	watching time go by; help the time go by
	13 = question future	When will her time come?
	14 = modal +	Only then would the time come to …
	16 = question modal	How far could or should this night go?

Appendix H

Table 7.7: Complete cross-tabulation ***noun/translated***

noun		translated		Total
		yes	no	
time	Count	157.0	2037.0	2194.0
	Expected Count	182.9	2011.1	2194.0
	Adjusted Residual	-2.4	2.4	
year	Count	40.0	449.0	489.0
	Expected Count	40.8	448.2	489.0
	Adjusted Residual	-.1	.1	
day	Count	121.0	813.0	934.0
	Expected Count	77.9	856.1	934.0
	Adjusted Residual	5.5	-5.5	
life	Count	38.0	641.0	679.0
	Expected Count	56.6	622.4	679.0
	Adjusted Residual	-2.7	2.7	
night	Count	70.0	556.0	626.0
	Expected Count	52.2	573.8	626.0
	Adjusted Residual	-2.7	-2.7	
end	Count	5.0	83.0	88.0
	Expected Count	7.3	80.7	88.0
	Adjusted Residual	-.9	.9	
week	Count	18.0	226.0	244.0
	Expected Count	20.3	223.7	244.0
	Adjusted Residual	-.6	.6	
month	Count	25.0	152.0	177.0
	Expected Count	14.8	162.2	177.0
	Adjusted Residual	2.8	-2.8	
moment	Count	23.0	422.0	445.0
	Expected Count	37.1	407.9	445.0
	Adjusted Residual	-2.5	2.5	
morning	Count	11.0	168.0	179.0
	Expected Count	14.9	164.1	179.0
	Adjusted Residual	-1.1	1.1	
hour	Count	27.0	129.0	156.0
	Expected Count	13.0	143.0	156.0
	Adjusted Residual	4.1	-4.1	
minute	Count	5.0	86.0	91.0
	Expected Count	7.6	83.4	91.0
	Adjusted Residual	-1.0	1.0	
age	Count	1.0	42.0	43.0
	Expected Count	3.6	39.4	43.0
	Adjusted Residual	-1.4	1.4	

May	Count	2.0	5.0	7.0
	Expected Count	.6	6.4	7.0
	Adjusted Residual	1.9	-1.9	
evening	Count	21.0	145.0	166.0
	Expected Count	13.8	152.2	166.0
	Adjusted Residual	2.0	-2.0	
century	Count	0	20.0	20.0
	Expected Count	1.7	18.3	20.0
	Adjusted Residual	-1.4	1.4	
future	Count	0	22.0	22.0
	Expected Count	1.8	20.2	22.0
	Adjusted Residual	-1.4	1.4	
season	Count	4.0	243.0	247.0
	Expected Count	20.6	226.4	247.0
	Adjusted Residual	-3.9	3.9	
Sunday	Count	3.0	28.0	31.0
	Expected Count	2.6	28.4	31.0
	AdjustedResidual	.3	-.3	
decade	Count	0	12.0	12.0
	Expected Count	1.0	11.0	12.0
	Adjusted Residual	-1.0	1.0	
Total	Count	571.0	6279.0	6850.0
	Expected Count	571.0	6297.0	6850.0

Appendix I

Table 7.8: Complete cross-tabulation ***verb/translated***

verb		translated		Total
		yes	no	
go	Count	236.0	2523.0	2759.0
	Expected Count	230.0	2529.0	2759.0
	Adjusted Residual	.5	-.5	
come	Count	208.0	2526.0	2734.0
	Expected Count	227.9	2506.1	2734.0
	Adjusted Residual	-1.8	1.8	
leave	Count	1.0	11.0	12.0
	Expected Count	1.0	11.0	12.0
	Adjusted Residual	.0	.0	
run	Count	6.0	234.0	240.0
	Expected Count	20.0	220.0	240.0
	Adjusted Residual	-3.3	3.3	
fall	Count	49.0	432.0	481.0
	Expected Count	40.1	440.9	481.0
	Adjusted Residual	-1.5	-1.5	
return	Count	10.0	53.0	63.0
	Expected Count	5.3	57.7	63.0
	Adjusted Residual	2.2	-2.2	
arrive	Count	30.0	186.0	216.0
	Expected Count	18.0	198.0	216.0
	Adjusted Residual	3.0	-3.0	
rise	Count	0	22.0	22.0
	Expected Count	1.8	20.2	22.0
	Adjusted Residual	-1.4	1.4	
enter	Count	2.0	15.0	17-0
	Expected Count	1.4	15.6	17.0
	Adjusted Residual	.5	-.5	
fly	Count	22.0	104.0	126.0
	Expected Count	10.5	115.5	126.0
	Adjusted Residual	3.7	-3.7	
travel	Count	0	5.0	5.0
	Expected Count	.4	4.6	5.0
	Adjusted Residual	-.7	.7	
jump	Count	0	5.0	5.0
	Expected Count	.4	4.6	5.0
	Adjusted Residual	-.7	.7	

roll	Count	7.0	163.0	170.0
	Expected Count	14.2	155.8	170.0
	Adjusted Residual	-2.0	2.0	
Total	Count	571.0	6279.0	6850.0
	Expected Count	571.0	6279.0	6850.0

BIBLIOGRAPHY

Al-Hasnawi, Ali R. 2007. A Cognitive Approach to Translating Metaphors. *Translation Journal* 11 (3).

Al-Ali, Mohammed N., Ali R. Al-Hasnawi&Mohammad Q. Al-Zoubi. 2006. Cogno-cultural issues in translating metaphors. *Perspectives: Studies in Translatology* 14 (3):230-239.

Aristotle. *Poetics*. Available from http://plato.stanford.edu/entries/aristotle-rhetoric/.

Baker, Mona, Gill Francis, John Sinclair, and Elena Tognini-Bonelli. 1993. *Text and technology: in honour of John Sinclair*. Philadelphia: J. Benjamins.

Boers, Frank. 1997. When a bodily source domain becomes prominent. The joy og counting metaphors in the socio-economic domain. In *Metaphor in cognitive linguistics. Selected papers from the fifth international cognitve linguistic conference*, edited by R. W. J. a. G. J. S. Gibbs. amsterdam/Philadelphia: John Benjamins Publishing Company.

Butler, Christopher. 1985. *Statistics in linguistics*. Oxford: Blackwell.

Croft, William. 2003. *Typology and universals*. 2nd ed, *Cambridge textbooks in linguistics*. Cambridge: Cambridge University Press.

Dagut, Menachem B. 1976. Can "metaphor" be translated? *Babel* 22 (1):21-33.

Gilquin, Gaëtanelle and Stefab Th. Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and LinguisticTheory* 5 (1):1–26.

Gries, Stefan Thomas. 2010. Useful statistics for corpus linguistics. In *A mosaic of corpus linguistics: selected approaches*, edited by A. S. M. Almela. Frankfurt am Main: Peter Lang.

Halverson, Sandra. 2003. The cognitive basis of translation universals. *Target* 15 (2):197-241.

Hannisdal, Bente. 2007. Variability and change in received pronunciation: a study of six phonological variables in the speech of television newsreaders. Dissertation for the degree of doctor artium., Department of English, University of Bergen, Bergen.

Hegrenæs, Claudia Förster. 2010. Cognitive Metaphors in Translations:
A case study on English and Norwegian. Bergen/Norway: University of Bergen.

Hinton, Perry R. 2004. *SPSS explained*. London: Routledge.

Holmes, James S. 1988/2000. The name and nature of translation studies. In *The translation studies reader*, edited by L. Venuti. London and New York: Routledge.

Jakobsen, A.L. 2006. Research methods in translation - TRANSLOG. In *Computer keystroke logging: methods and applications*, edited by K. P. H. S. a. E. Lindgren. Oxford: Elsevier.

Jääskeläinen, Riitta. 1999. Tapping the process: an explorative study of the cognitive and affective factors involved in translating, Joensuun yliopisto, Joensuu.

Kinnear, Paul R., and Colin D. Gray. 2011. *IBM SPSS statistics 18 made simple*. Hove: Psychology Press.

Kövecses, Zoltán. 1986. *Metaphors of anger, pride, and love: a lexical approach to the structure of concepts*. Amsterdam: John Benjamins.

Lakoff, George. 2008. The neural theory of metaphor. In *The Cambridge Handbook of Metaphor and Thought*, edited by J. Raymond W. Gibbs. Cambridge: Cambridge University Press.

Lakoff, George, and Mark Johnson. 2003. *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, George. 1993. The contemporary theory of metaphor. In *Metaphor and Thought*, edited by A. Ortony. Cambridge: CUP.

Leech, Geoffrey N., Paul Rayson, and Andrew Wilson. 2001. *Word frequencies in written and spoken English: based on the British National Corpus*. Harlow: Longman.

Levin, Beth. 1993. *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago Press.

Luchjenbroers, June and Australian Linguistics Institute. 2006. *Cognitive linguistics investigations: across languages, fields and philosophical boundaries*. Amsterdam: J. Benjamins.

McEnery, Tony, and Andrew Wilson. 2001. *Corpus linguistics: an introduction*. Edinburgh: Edinburgh University Press.

McEnery, Tony, Richard Xiao, and Yukio Tono. 2006. *Corpus-based language studies: an advanced resource book*. London: Routledge..

Munday, Jeremy. 2001. *Introducing translation studies: Theories and appliactions*. London: Routledge.

Musolff, Andreas. 2004. *Metaphor and political discourse: analogical reasoning in debates about Europe*. Basingstoke: Palgrave Macmillan.

Neubert, Albrecht. 1997. Postulates for a theory of Translatio. In *Cognitive Processes in Translation and Interpreting.*, edited by J. H. Danks. Thousand Oaks, Calif.: Sage Publications.

Newmark, Peter. 1983. *The translation of metaphor*, *Paper, Series A*. Trier: L.A.U.T.

Oakes, Michael P. 1998. *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Olohan, Maeve. 2004. *Introducing corpora in translation studies*. London: Routledge.

Schäffner, Christina. 2004. Metaphor and translation: some implications of a cognitive approach. *Journal of Pragmatics* 36:1253-1269.

Semino, Elena. 2008. *Metaphor in discourse*. Cambridge: Cambridge University Press.

Sjørup, Anette C. 2011. Cognitive Effort in Translation: An Eye-tracking Study. In *Cognitive Explorations of Translation*, edited by S. O'Brien. London: Continuum International Publishing Group.

Stefanowitsch, Anatol. 2006. Corpus-based approaches to metaphor and metonymy. In *Corpus-based approaches to metaphor and metonymy*, edited by A. S. a. S. T. Gries. Berlin: Mouton de Gruyter.

Toury, Gideon. 1995. *Descriptive translation studies and beyond*. Amsterdam: John Benjamins.

Trask, R. L., and Peter Stockwell. 2007. *Language and linguistics: the key concepts*. Abingdon: Routledge.

Tummers, Jose, Kris Heylen, and Dirk Geeraerts. 2005. Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1 (2):225–261.