

# **Normalisation in translation: a corpus-based study of *take***

**By Jørgen Vik Bergstad**

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

**University of Bergen  
Spring term 2009**



## **Sammendrag:**

Denne oppgaven gjør et forsøk på å undersøke hvorvidt normalisering er et trekk ved oversetterhandlingen ved hjelp av corpus-teknologi. Normalisering har lenge vært foreslått som et fenomen som oversettere underlegger seg i sin overføring av ord og begreper fra et språk til et annet. Likevel er dette et svært lite utforsket område. Noen spredte undersøkelser har blitt utført tidligere uten å ha kunnet fremdrive utvetydige bevis for (eller imot) at dette er et reelt trekk ved oversetterhandlingen. Dette fenomenet består av at typiske trekk ved mål språket blir overholdt, og til og med overdrevet, i oversettelser. I denne undersøkelsen har det engelske verbet *take* (ta) blitt valgt som testgjenstand. En prototypestruktur av det semantiske nettverket til *take* er blitt utviklet, der de prototypiske meningene av verbet blir ansett som typiske trekk ved språket. Deretter har setninger som inneholder dette verbet blitt trukket ut av samlinger med tekster (corpora) som består av oversatte tekster på engelsk og ikke-oversatte tekster på samme språk. I en analyse av disse setningene har meningene av verbet blitt identifisert og registrert i forskjellige meningskategorier, oversatte og ikke-oversatte setninger hver for seg. Dermed sitter man igjen med en rekke medgjørige tall som man kan bruke til å sammenligne distribusjonen av de forskjellige meningene av *take* i oversatte tekster med distribusjonen av de forskjellige meningene av *take* i ikke-oversatte tekster.

## **Table of contents**

Sammendrag (abstract).....	ii
List of tables.....	v
List of abbreviations.....	vi
<b>Chapter 1: Introduction</b> .....	<b>1</b>
<b>Chapter 2: Background and theory</b>	
2. 1. Translation Studies	
2. 1. 1. Descriptive translation studies (DTS).....	4
2. 1. 2. Corpora in DTS.....	7
2. 1. 3. Universals.....	12
2. 2. Related research	
2. 2. 1. Prototypicality.....	21
2. 2. 2. Gravitational pull.....	24
2. 2. 3. Basic verbs.....	26
<b>Chapter 3: Material and method</b>	
3. 1. Material	
3. 1. 1. English-Norwegian Parallel Corpus (ENPC).....	29
3. 1. 2. Translational English Corpus (TEC).....	31
3. 1. 3. British National Corpus (BNC).....	32
3. 1. 4. Comparing and adjusting the corpora.....	33
3. 2. Method	
3. 2. 1. What to include.....	35
3. 2. 2. Deriving a hypothesis for category structure .....	35
3. 2. 3. Treatment of the data.....	41
<b>Chapter 4: Results and discussion</b>	
4. 1. Results	
4. 1. 1. General overview of sense categories.....	45
4. 1. 2. General patterns.....	47
4. 1. 3. Text type.....	51

4. 1. 4. Verb form.....	53
4. 1. 5. Concrete/abstract object.....	59
4. 2. Discussion	
4. 2. 1. General overview of sense categories.....	62
4. 2. 2. General patterns.....	64
4. 2. 3. Text type.....	67
4. 2. 4. Verb form.....	69
4. 2. 5. Concrete/abstract object.....	71
4. 2. 6. Summary.....	72
<b>Chapter 5: Conclusion.....</b>	<b>74</b>
References.....	76

## **List of tables:**

3. 1. Dictionary listings.....	36
3. 2. Dictionary listings revised.....	38
4. 1. Senses distributed across the corpora.....	46
4. 2. Senses distributed in non-translations and translations.....	47
4. 3. ENPC translations and the TEC.....	49
4. 4. ENPC non-translations and the BNC.....	50
4. 5. Sense categories distributed across text types (fiction/non-fiction).....	52
4. 6. Verb form in translated texts.....	54
4. 7. Verb form in non-translated texts.....	55
4. 8. Verb form: ‘move’, ‘remove’.....	56
4. 9. Verb form: ‘bring’, ‘carry’.....	57
4. 10. Verb form: ‘hold’, ‘get hold of’.....	57
4. 11. Verb form: ‘require’.....	58
4. 12. Verb form: ‘occupy’.....	59
4. 13. Concrete/abstract object.....	60
4. 14. Concrete/abstract object across non-translations and translations.....	61

**List of abbreviations:**

BNC – British National Corpus

DTS – Descriptive Translation Studies

ENPC – English-Norwegian Parallel Corpus

ESPC – English-Swedish Parallel Corpus

L1 – Language 1

L2 – Language 2

NON-TEC – Non-Translational English Corpus

SL – Source Language

ST – Source Text

TEC – Translational English Corpus

TEI – Text Encoding Initiative

TL – Target Language

TT – Target Text

## **CHAPTER 1: INTRODUCTION**

There has been a widespread interest among scholars within the field of Translation Studies the last three decades in locating and investigating so-called ‘universals’, or ‘features of translation’. The notion that translations differ in some respects from non-translated language is well established; however, the debate circles around where the difference lies and to which degree. This is evidenced by the variety of theories put forward by a host of scholars. The most prominent of these are Mona Baker (1993) and Gideon Toury (1995), who both have different views on how the issue of translational features should be approached. Baker (1993) proposed four features of translation which she claimed could be found to be universal. Among Baker’s four universals we find normalisation, which can be described as a tendency for translators to adhere to typical linguistic patterns of the target language, even to the point of exaggeration (Baker 1996: 183). This idea is also reflected in Toury’s law of growing standardization, as he claims that the translators tend to choose elements that are more habitually used in the target language to replace the elements in the source texts (Toury 1995: 268). In other words, Toury claims that translators tend to choose items and structures in their translations which are habitually, and thus typically, used in the target language. Some research has been done to investigate the suggestion that translators tend to normalise in their language in translations (e.g. Kenny (2001), Olohan (2004)), but the conclusions drawn on the basis of most of these studies have not made it possible to establish normalisation as a feature of translation, much less a translation universal. In order to establish normalisation as a universal feature of translation, a cognitive explanation must be found. If the feature occurs because of cognitive factors, it is much more likely that it will occur in translations from every language and to every language, and the feature would be universal.

In Halverson (2003) we might find a possible cognitive explanation for the tendency for translators to adhere to typical features of the target language, even to the point of exaggeration. This possible explanation relates to the notion of gravitational pull. According to Halverson, we would expect to find an over-representation of prototypical items and high level schemas in translations in comparison with non-translated language. The idea is that the prototypical members of a category and higher level schemas in the schematic network are cognitively more easily accessible for a producer of language than more peripheral members and lower level schemas, and that these therefore would be more frequent than other members across all languages. Furthermore, a translator would connect peripheral members and lower level schemas in the source language more easily with prototypical structures and higher level

schemas in the target language, and thus, the translations would have a higher frequency of these structures than non-translations. To investigate normalisation in translation, therefore, we would be well advised to go searching for a category with a rich schematic network with a clearly defined prototypical structure.

Several cases of prototypical structure in categories are described by Viberg (2002a, 2002b). These relate to the notion of basic verbs and the structure of their inherent senses. According to Viberg, basic verbs are verbs which have the same reference points in terms of conceptual domains across languages. These verbs are divided into two categories: Areally specific basic verbs (*can, must, have*, etc) and nuclear verbs (*take, give, see*, etc). Areally specific basic verbs share conceptual features across some languages, while nuclear verbs are thought to share conceptual domains across *all* languages. When it comes to nuclear verbs, language-specific words are mapped onto universal semantic units. Viberg (2002b) also explores the rich semantic structure of the verb *get*, and identifies a sense which he calls the ‘basic meaning’. In other words, basic verbs can be said to have a rich schematic network with the possibility of establishing a prototypical structure.

Virtually every word in every language has multiple senses, which come into use depending on the context. This is also true for *take*. Since this verb is identified by Viberg (2002a) as being a basic verb, I found it interesting to locate these senses and to see if there are any differences in the selection of these senses between texts translated into English and non-translated texts in English. The purpose of this is to use the notion of prototypicality to investigate whether normalisation actually happens during translation, as proposed by Baker and Toury. The different senses of *take* will be organised in a list reflecting the frequency of which they are used and their centrality. The list will function as a working definition of prototypicality concerning these senses (which senses are viewed as prototypical vs. peripheral). Earlier research investigating features of translation has shown that other factors, such as text type, might play a part in how translators behave. Therefore, factors such as text type (fiction vs. non-fiction), verb form, and whether the verb takes a concrete or an abstract object will be taken into consideration in the analysis.

My hypothesis is that the prototypical (most frequent) senses of the verb *take* will be *over-represented* in translations in comparison to non-translations. Similarly, I hypothesise that the fringe (least frequent) senses of the verb *take* will be *under-represented* in translations in comparison to non-translations. An additional research question I will deal with in this thesis is whether other factors, such as text type, verb form and whether the verb takes a



concrete or an abstract object, play a part in how translators behave and, if so, what differences this results in between translations and non-translations.

An overview of the theoretical background for this thesis is given in Chapter 2 which is called 'background and theory'. Here, the development and structure of the field of Translation Studies is presented with a focus on the period from the 1980's and onwards. In addition, some of the most important terms that have emerged in the field are elaborated and explained. A second part of this chapter deals with the issues of basic verbs, prototypicality and cognitive aspects of translation.

Chapter 3 is called 'material and method' and is, as the title suggests, divided into two parts. The first part is dedicated to presenting the material that I have used for the purpose of this thesis. The reader is given a comprehensive account of the corpora I have extracted my data from and how this has been done. The corpora presented here, and which I have used, are ENPC, the BNC and the TEC. The second part of this chapter describes the methodological aspects of the thesis, giving the reader an account of how the methodological challenges I met, along with the solutions for these challenges.

Chapter 4, called 'results and discussion', is the main chapter of the thesis. This is where I present my results, using tables showing the percentages and raw figures in relation to the variables I have opted to include. In the second section of this chapter, follows a discussion of these results, attempting to link my findings to the theories presented in Chapter 2.

Finally, Chapter 5 presents the reader with a conclusion where generalisations are attempted along with a couple of suggestions for further research.

## **CHAPTER 2: BACKGROUND AND THEORY**

The first section of this chapter will give an overview of the field of Translation Studies, with special emphasis on the use of corpora in descriptive translation studies and the issue of translation universals. The second section is dedicated to research which lies outside of, or on the borders of, the field of Translation Studies, but that nevertheless is relevant for this thesis.

### **2. 1. Translation Studies**

Translation Studies is the term used for ‘the complex of problems clustered round the phenomenon of translating and translations’ (Holmes, 1988; p 67). This section will look into some of the main issues within this field of study, with special focus on the matters that are important for my thesis.

#### **2. 1. 1. Descriptive translation studies (DTS)**

For a long time, the main focus of scholars doing research on translations was prescriptive. They opted to make statements about how translations *should be* rather than how they *actually are*. Also, an essentialist view that ‘meanings are objective and stable, that the translator’s job is to find and transfer these and hence to remain as invisible as possible’ (Chesterman and Arrojo 2000: 151) was predominant. Later, scholars came to realize that meanings are not stable and objective across languages, but rather that a culture has a set of meanings which do not necessarily coincide with the set of meanings in another culture. For example, there is not necessarily a one-to-one relationship between the meaning of *to walk* in one language and its closest counterpart in another language. This realization is evident in the move of focus from the notion of equivalence to the notion of shifts. Shifts refer to the situation where a component in the source text is represented in the target text by a component which is not its formal correspondent (Hatim & Munday 2004: 28). Perhaps the most notable scholar when it comes to this move of focus is Gideon Toury, who as early as 1980 called for a more scientific and organized approach to the phenomenon of translating and translations (1980). Later, he brings to mind another scholar who even earlier had ideas about the study of translations, namely James S. Holmes (Toury 1995: 7-10). In the Third International Congress of Applied Linguistics (Holmes 1972), Holmes gave support to the term ‘Translation Studies’ and furthermore presented his model of how this field should be organised.

In this model, Holmes proposes a separation of the applied branch and the 'pure' branch. The former deals with aspects like translation training, aids, and criticism, while the latter has a more scientific edge to it as it opens for empirical testing. The idea is that the nodes on the 'pure' branch are mutually complimentary. The data collected in the descriptive node help in forming theories which go into the theoretical node. These theories will then bring up new issues and hypotheses that again are tested in the descriptive node. In addition, the results and theories from the 'pure' branch are meant to be brought into the applied branch, so that translators are given a truer picture of what a translation can look like.

However, Holmes' model has not been accepted by everyone. Anthony Pym (1998) is one scholar who expresses scepticism towards the general acceptance of this model. He is of the opinion that the model does not encompass diachronic considerations. In other words, he cannot find a place for history in Holmes' model. He admits that translations of the past can be studied under Holmes' product-oriented descriptive branch, and also that the historical functions of past translations can be studied under the function-oriented descriptive branch. What Pym brings to attention, however, is the map's inability to account for areas of translation studies that are not descriptive from a historical point of view, like translation criticism and the theoretical branch. He claims that a historian would have to leap from one branch to another during his/her investigation instead of neatly following the paths that a model designed to encompass the historical point of view would provide:

Translation historians of any but the narrowest variety would seem condemned to jump from one patch to another, describing products here, analyzing functions there, and finding themselves marginally implicated in a metadescription of the whole lot (Pym 1998: 2).

Also, Pym warns that if everything that is done in translation studies is based on this map and all the work that is done is adapted to this one model, some perspectives may be lost. He claims that Holmes' map is not all-inclusive in that 'it delineates no ground for any specific theory of translation history, nor for historiography as a way of applying and testing theories' (Pym 1998: 2).

Pym points to the authority and power that a map of this kind may hold. 'Maps are peculiar instruments of power. They tend to make you look in certain directions; they make you overlook other directions' (Pym 1998: 3). If every scholar within a field conducts his/her research with basis on only one map, this map has absolute power, and everything that falls

outside of the map is neglected. As an example, Pym compares Holmes's map with Lawrence Murphy's map from the 16<sup>th</sup> century. Pym notes that the translator has disappeared from the 'pure' branch entirely. The focus is now on the product, the process or the function of the product. This is in sharp contrast to Murphy's map. Here, the main division was between the translator and the product/process. Pym suggests that a re-introduction of the 'people' (translators and researchers) back into the branch of 'pure' research may be called for.

Nevertheless, Andrew Chesterman (2000) outlines three models in which descriptive research can take form which corresponds to Holmes' notion of process- function- and product-oriented ways of approaching translations; the process, causal and comparative models. The process model will be employed by scholars attempting to discover the dynamics of the translation process. The research a scholar with this goal would be involved in includes 'what happens during translation, i.e. of the mental steps taken by translators between, and including, reception of the source text and production of the target text' (Olohan 2004: 38). Since the object of investigation for this type of research (the human brain) is not easily accessible, one has found other ways of gathering information about the translation process. Introspective approaches have been highly preferred, although these methods have been met with some scepticism due to their methodological weaknesses (Olohan 2004: 38).

According to Chesterman, the causal model is used more in a socio-linguistic approach to translation studies. It deals with what effects a translation has in a social, textual or cognitive domain, and what causes these effects. The causal model also relates to what conditions that lay ground for the decision to produce a translation. Often, research following the causal model relies on extra-textual material, like author's and translator's prefaces, editor's notes, interviews with translators, etc. This model also unites translation studies with cultural studies since part of the idea is to gain knowledge about how translations affect a culture.

In Chesterman's view, the comparative model is focussed on comparing texts and on drawing some conclusions or general tendencies from the comparison. There are a number of different comparisons that can be made using this model within the field of translation studies. We can compare source-texts with their translations, original texts in one language with translated texts in the same language, translations in one language with translations in another language, concentrate on the translations made by one translator or the texts written by one author and the translations of these texts, etc. A great part of the research done in translation studies has been made with the comparative model as a basis.

One of the developments that have been crucial for the employment of the comparative model is the developing of electronic corpora.

### 2. 1. 2. Corpora in DTS

This section will account for the history and aspirations of the introduction of electronic corpora to the area of Descriptive Translation Studies. In addition it will serve as a foundation for the next section (Section 2. 1. 3.), which will be dedicated to the notion of translation universals. When using the comparative model in translation studies, electronic corpora become a very important tool. Electronic corpora make it possible to gather sizable collections of texts and also make it easy to extract suitable data. One of the first to call for a merging of the discipline of translation studies and to point out the advantages that electronic corpora offer was Mona Baker (1993). In her discussion, she pointed out that it was the extensive developments within computer science and the ability to store large amounts of data which allowed for large collections of texts to be gathered in order to extract data in itself or in comparison with other large collections of texts. In the beginning, corpora were mostly used within the field of linguistics, but Baker drew attention to the impact this method could have on the field of translation studies as well. At the time that Baker's article was written, 'the vast majority of research carried out in this...discipline, [was]...concerned exclusively with the relationship between specific source and target texts, rather than with the nature of translated texts as such' (Baker 1993: 234). Clearly, Baker not only hoped for a new method of research being put to the use, but also a shift of focus in the whole field of translation studies from merely investigating with the goal of aiding translators in their training to trying to describe the very nature of translations. She frowns upon the notion that translations are perceived to be inferior to original texts (Baker 1993: 233), which again suggests that the translator should strive to make the translation as similar as possible to the source texts. According to Baker, priority should not be given to close formal equivalence between the source text and the translation, and thereby making a 'copy' of the source text in another language, but rather the focus should be on 'how similar meanings and functions are typically expressed in the target language' (Baker 1993: 236). Following this line of thought, 'the need to study authentic instances of similar discourse in the two languages becomes obvious' (Baker 1993: 236).

Baker also describes other developments that encourage the introduction of computerized corpora into the field of translation studies. For example, in her view, the notion of equivalence has undergone some changes. It is not necessarily the expression that is most

closely related to the meaning of the source text that should be chosen in the translation, but perhaps rather the expression most closely related in actual use of the expression found in the source text. In other words, equivalence has become more situational and dependent on the context, and this usage-based form of approach is best investigated by computerized corpora. Also, Baker discusses Toury's 'tripartite model in which norms represent an intermediate level between competence and performance' (Baker 1993: 239). Competence represents an 'inventory of all the options that are available for the translators in a given context, and performance...the subset of options that are actually selected by the translators from this inventory' (Baker 1993: 239). In relation to this, norms are the alternatives that occur more often than others in particular situations, i.e. the alternatives most likely to be chosen by the translator 'at a given time and in a given socio-cultural situation' (Baker 1993: 239). In Baker's view, in order to determine what would qualify as a norm, we would have to scan through a massive amount of data, and therefore we would be dependent on a sizeable computerized corpus of translated texts. Norms do not derive from neither source texts or the target language, rather 'they are a product of a tradition of translating in specific ways, a tradition which can only be observed and elaborated through the analysis of a representative body of translated texts in a given language or culture' (Baker 1993: 240).

Baker predicts the potential effect that computerized corpora may have on the field of translation studies, both on the applied and the descriptive and theoretical branches. In the former branch it will certainly enhance the performance of the translators, and eventually it may even aid in the development of machine translation. In the latter branches it will help us outline certain features of translation that are thought to be universal, or as Baker calls them; 'features which typically occur in translated text rather than original utterances and which are not the result of interference from specific language systems' (Baker 1993: 243). With the use of a computerized corpus, we can compare a collection of original texts in one language with a collection of translated texts in the same language, preferably texts translated from a number of different languages so that the issue of cultural or language specificity is eliminated. Since the aim of any scientific discipline is to make general claims about the phenomena in question, hopefully we can detect some features of translated language that differ from original language and that these features exist across all cultural and linguistic borders, so that we can call them universals.

Another potential effect that computerized corpora can have on the field of translation studies is the establishment of translational norms and the separation of these from the translation universals. Translational norms are quite similar to universals, but they do not

necessarily cross cultural or linguistic borders, they are language/culture specific or specific for a period of time.

A corpus can be put together in a number of different ways. The parallel corpus is one type of corpus that is of great use in the field of translation studies. A parallel corpus can be unidirectional with source texts in language A and translations in language B. It can also be bidirectional with source texts in language A and translations in language B + source texts in language B and translations in language A. These kinds of corpus may be bilingual or multilingual, the former referring to the examples above, the latter meaning source texts in language A and translations in languages A, B, C and D, or vice versa. These types of corpora are favoured by scholars wishing to perform contrastive analyses, i.e. comparing different languages to explore similarities or differences between them. The focus is often on a single word's behaviour or on grammatical constructions, and the parallel corpus allows us to search for the word or grammatical construction in question in large quantities of texts and to extract the data that is needed in order to perform the analysis. The corpus gives us all the instances of the item in question that exist in it and supplies us with the item's context so that we can be sure we have found what we needed in case the item's environment is a factor.

An example of parallel corpus put to use by translation scholars is Josef Schmieid's (1998) study of English *with* and German *mit*. In his study he compared original texts in English with their translations in German and sought to investigate why translators come to make different translation solutions even when dealing with such a simple lexical item which also are quite similar between the two languages. While acknowledging that one cannot determine why the translator has made the choices he/she has made, Schmieid still gives an account of what choices have been made in stead of *mit* in the variety of situations, and hypothesize on what factors *might* have driven the translator to make his/her choice. Also, Schmieid sets up a bilingual lexicographical resource on the basis of his results that might make up for some of the inadequacies of the traditional bilingual dictionary which often fails to give all the options of meanings that are available.

Another scholar who has been occupied with how a parallel corpus can be used to carry out contrastive studies is Stig Johansson. For example, in a paper from 2007, he used the ENPC (English-Norwegian Parallel Corpus) and the ESPC (English-Swedish Parallel Corpus) to explore the potential overuse of the Swedish verb *tillbringa* (*spend*) and the Norwegian counterpart *tilbringe* in translations into Swedish and Norwegian from English. Since this word is found to occur more often in translations into these two languages than in non-translated texts, he assumes that this is due to influence from the source language's (English)

relatively more frequent occurrence of *spend*. This leads to Johansson's question: 'what alternatives are there for conveying the English notion of spending time [in Swedish and Norwegian]?' (Johansson 2007a: 1). He finds that there are numerous alternatives to the above mentioned Swedish and Norwegian version of the English verb *spend*, and that there is, therefore, no reason for translators to overuse them. What he did to discover this was to reverse the direction of translation and to see what words occurred in the Swedish and Norwegian non-translations where *spend* had been used in the translations. Johansson has also collected a range of his studies in a book (2007b) which aims to show how corpora can be used in contrastive studies.

The comparable corpus is another type of corpus that is of importance when it comes to translation studies. It is compiled of original texts in one language and translations in the same language. It was Baker who, in her paper from 1995, suggested that this form of corpus should be put to use in translation studies. She wanted a shift away from the heavy focus on comparing translations with their source texts and language A with language B. In stead, the focus should be on comparing translations with original texts in the same language, so that one could see how translated language differs from language that has not been constrained by the process of translating. Baker not only was the one to suggest this new type of corpora, she was also the main force behind the construction of the first corpus of this kind, namely the Translational English Corpus (TEC) (<http://ronaldo.cs.>), together with another prominent scholar in the field, Sara Laviosa. TEC consists of translations into English from a variety of European and non-European source languages. In order for this corpus to be used as a comparable corpus, it needs a corpus of original texts to be compared with. The British National Corpus (BNC) serves this purpose perfectly, as it contains the same kind of sectioning (fiction, biography, etc.), approximately the same number of words (approximately ten million) and uses the same time-frame in which the texts which the samples are collected from are written.

While the parallel corpus, with its inherent focus on the source text, was found to be somewhat inadequate in trying to learn more about the translation process, it is thought that the comparable corpus, which focuses on the translation product, may be more useful in this respect. By focussing on the translation product, it is believed that 'scholars are prioritizing the activity and factors that influence it, of which the source text is but one such influencing factor' (Olohan, 2004; p. 39). In other words, the source text is just one factor that has bearing on the translation process, so when using parallel corpora, it is argued that one is only covering a fraction of the 'problem'. It is not thereby said that it is necessarily enough only to



consider the product in an analysis hoping to gain information about the translation process. This is why the translational context is given in TEC's online concordance browser, ensuring easy access to all necessary information surrounding the translation situation of each individual text.

Both parallel and comparable corpora are useful within the field of translation studies. However, there are problems that have to do with comparability. When creating a sub-corpus in, for example, TEC, one has a number of alternatives to choose from, ranging from text type, full text/extract, synchronic/diachronic to gender of translator/author, professional/untrained translator and degree of accessibility of the texts. While some of these criteria are easy to adjust between the corpus of translated texts and the corpus of original texts, others are more difficult to fulfil in both corpora.

There are also restrictions to the two types of corpora's applicability. The parallel corpus will not supply the right data in order to answer questions about the process of translating as it will rely on the data from the source texts. At the same time, the comparable corpus may be somewhat lacking in the source text-area as its main focus is on the translation product itself. Dorothy Kenny (2005: 157) suggests a merging of these two types of corpora, since the comparable corpus can not inherently say anything about the effect that the source language may have on the translations. Kenny applies the findings of a comparable study carried out by Maeve Olohan and Mona Baker (2000) to a parallel corpus. Olohan and Baker found that the reporting *that* occurs more often in the corpus of translations than in the corpus of original texts, and hypothesized that this might be a sign that translators tend to make their language more explicit than writers of original texts. Kenny claims that Olohan and Baker have failed to take into consideration the effect the source texts and source language may have on these results. Therefore, she tests these findings in a parallel corpus consisting of German original texts and their translations into English. She gets similar results to what Olohan and Baker got when she searches the corpus of translated texts for the phenomenon. When comparing this with the source texts in German, she finds that when *that* occurs in the translated texts, the German equivalent *dass* usually occurs in the original texts also; likewise when *that* has been omitted from the translations it is also usually omitted from the source texts. When differences occur it is for the most part because the reporting *that* has been added in the translations (Kenny 2005: 161). This backs up the findings of Olohan and Baker and also deals with the factor of language specificity. The comparable corpus can not be relied upon to deal with language specific issues in the source texts. This is why it might be a good idea to combine the two types of corpora.

### 2. 1. 3. Universals

Early work on the notion of translation universals and other similar concepts is dominated by Mona Baker (1993) and Gideon Toury (1995). As was briefly mentioned above, translation universals are central in recent descriptive translation research. Mona Baker (1993) was the first to call for the use of corpora to investigate whether any of the claimed features of translation could be said to be universal, i.e. typical for all translations, regardless of source language and target language. As Olohan (2004) puts it; ‘since [universals] are a product of constraints inherent in the translation process, they would not vary across cultures, unlike norms of translation, which are products of social, cultural and historical contexts’ (92). Dorothy Kenny (2001: 53) asserts that it is not in the establishment of these features that we can say if they are universals or norms, but rather in the explanation of them. For example, if we find a feature that looks to be a feature of translation and we find that the explanation why this feature occurs in translations is of a social or cultural nature, we are dealing with a norm. If, however, the explanation is of a cognitive nature, we may safely assume that the feature is universal.

Baker (1993) initially proposes four features of translation that she considers may be universals; simplification, explicitation, levelling out and normalisation. Simplification is ‘the idea that translators subconsciously simplify the language or message or both’ (Baker 1996: 181). In other words, the translator is thought to make the text more accessible for the reader by simplifying the language and/or message. Explicitation relates to ‘the tendency to spell things out in translation, including, in its simplest form, the practice of adding background information’ (Baker 1996: 180). The translator makes the text more accessible to the reader by paraphrasing, adding vital information or excessive explanation. Levelling out adheres to ‘the tendency of translated text to gravitate around the centre of any continuum rather than towards the fringes’ (Baker 1996: 184). This entails lower lexis rate, use of more standard sentence structures and generally sticking to the standards of the target language.

Normalisation can be described as ‘the tendency to conform to patterns and practices that are typical of the target language, even to the point of exaggeration’ (Baker 1996: 183). Examples of normalisation may be replacing a source text metaphor with a metaphor that is canonized in the target language, preferably a metaphor that carries much the same meaning. Or if the punctuation in the source text is odd, or maybe perfectly natural for the source culture, but odd for the target culture, the translator may ‘decide’ to adapt these to the target language norms.

Anna Mauranen (2007) also lists a selection of hypothesized translation universals. New entries in this list are interference which deals with transfers of elements from the source text; untypical collocations which is the tendency for translators to favour collocations that are rarely or never found in original texts in the target language; and unique items which refers to a potential under- (or over-) representation of target language unique items in translation.

In what follows in the next chapters of this thesis there will be a focus on normalisation as a universal. Therefore, it would make sense to take a closer look at the more recent issues and studies that have been done on this subject. However, some of the other proposed universals also deserve to be elaborated since they relate to normalisation. One proposed universal which is not mentioned by Baker is the one concerning unique items. The idea that under-representation of unique items may be a translation universal was first proposed by Sonja Tirkkonen-Condit (2002). She got the idea from an earlier discussion made by Katharina Reiss (1971) who claimed that ‘translations may not fully exploit the linguistic resources of the target language’ (Tirkkonen-Condit 2002: 208). In this lies the thought that because translation production is triggered by elements in the source texts, some elements specific to the target language may not appear as often in translated texts as in texts originally written in that target language. These elements, Tirkkonen-Condit has decided to call unique items. In a 2004 paper she further elaborates on how she defines a unique item:

Every language has linguistic elements that are unique in the sense that they lack straightforward linguistic counterparts in other languages. These elements may be lexical, phrasal, syntactic or textual, and they need not be in any sense untranslatable; they are simply not similarly manifested (e.g. lexicalized) in other languages.  
(Tirkkonen-Condit 2004: 177)

Chesterman (2007) admits that investigation into whether under-representation of unique items in translation is a universal tendency may prove fruitful, but puts forward some questions regarding the conceptualization of these unique items. The first question he asks is whether the items need to be unique in relation to all other languages, or just the source language at hand. Tirkkonen-Condit is somewhat unclear on this matter in her papers, but in an email to Chesterman, she explains that the item only needs to be unique in relation to the source language at hand. The second question deals with absolute uniqueness as opposed to degree of uniqueness: does the item have to be absolutely unique? Chesterman draws from

Tirkkonen-Condit formulation which includes the concept of (lack of) similarity that we are dealing with degrees of uniqueness.

Translationally equivalent items in two languages can be more or less similar, and moreover, more or less similar in an infinite number of different ways. The less similar they are, the more unique a given target item is said to be; the degree of uniqueness depends inversely on the degree of similarity. (Chesterman 2007: 5)

The third question relates to the identification of uniqueness. How does one locate and determine that an item is unique for the target language? He claims that Tirkkonen-Condit's explanation is too 'slippery' when she says that 'linguistic elements...are unique in the sense that they lack straightforward linguistic counterparts in other languages' (Tirkkonen-Condit 2004: 177). Chesterman's answer is that 'an item counts as unique if it cannot be readily translated back into a given source language without a unit shift' (Chesterman 2007: 7). By a unit shift, he means a shift from morpheme to word, word to phrase, phrase to sentence, etc. The fourth question that Chesterman asks is whether the items should be linguistically unique or perceptually unique. In other words, should the items be proven to be unique by contrastive evidence or would it suffice that the given translator perceives the item to be unique? His answer is that they should be linguistically unique, because 'the whole point of the hypothesis is that the target unique items are not perceived...they are not even triggered' (Chesterman 2007: 9). He then moves on to ask if under-representation of unique items occur in texts other than those which are translated. He finds evidence from earlier research that suggest that this might be the case. This feature is not specific for translations, but for texts which are produced with constraining factors present, such as producing a text in a non-native language. In light of these considerations, Chesterman finds the term 'unique items' to be somewhat misleading. He feels the emphasis is too great on the uniqueness of the items, since he has found that the

[u]niqueness is relative, rather than absolute; the hypothesis refers to particular sets of source and target languages rather than to all languages; uniqueness can be defined only rather loosely in the contrastive terms of relative formal difference; and the phenomenon in question may not be unique to translation but may also be typical of second-language usage (Chesterman 2007: 11).

Finally, he states that the methodology used to investigate under-representation of unique items in translations needs to be revised. Instead of starting with an intuition about what might constitute a unique item, and then go from there, one should take steps to make sure that the item is unique before starting the investigation. He sets up a four-point strategy to ensure that we are dealing with a unique item. The main elements in this strategy are to use a contrastive corpus to find items that differ in frequency between translations and non-translations, focus on the items that occur more often in non-translations than translations, and then find explanations for the under-representation. If this explanation is that the item is formally very different from its source language counterpart, there is a good chance that we have found a unique item. Tirkkonen-Condit proposes a slightly different approach. She wants to investigate language pairs in order to find items that have the same basic meaning but are formally different in the two languages. Then she moves on to comparing the frequencies of these items in translations and non-translations to find if any of them are under-represented in translated language.

Granted, these are elements that need to be sorted out if any real research into under-representation of unique items in translation is to be carried out. Nevertheless, Tirkkonen-Condit has done some important preliminary work to establish that this possible universal is worth looking further into. In her paper from 2004, she uses the Corpus of Translated Finnish, a comparable corpus, to get an overview of the frequencies of Finnish verbs of sufficiency and clitic particles *-kin* and *-han*. It is, of course, her hypothesis that these verbs and clitic particles, which in her view are unique for the Finnish language, will appear with a higher frequency in non-translated texts than in translations because the source texts do not trigger them in the translators mind. The results of this study strongly support her hypothesis as the frequency of these items is markedly lower in translated texts. She explains this by referring to her initial hypothesis:

The most obvious explanation for the relative scarcity of the verbs of sufficiency in Translated language is the explanation suggested by the Unique Items Hypothesis itself, namely that translators dismiss these verbs because they are not obvious equivalents for any particular items in the source text. (Tirkkonen- Condit 2004: 181)

What makes the Unique Items Hypothesis special in relation to the other proposed universals is that it might be cognitively determined. In her paper from 2008, Kirsten Malmkjær explores what lies in the terms 'norm' and 'universal'. She finds that while norms are socially

constrained, absolute universals are cognitively determined. By this she is saying that ‘norms...regulate behaviour; but the behaviour survives, though it may be considered deviant, even if the norms are not adhered to’ (Malmkjær 2008: 52). Universals, on the other hand are not subject to social influence, only cognitive influence. She separates non-absolute universals from absolute universals, saying that while the former are cognitively constrained, the latter are cognitively determined. Non-absolute universals are subject to choice-making by the mind, but absolute universals are not. Using this framework for identifying translation universals, Malmkjær is only able to identify one potentially absolute universal, namely the under-representation of unique items:

The phenomenon of under-representation in translation of features unique to the target language arises because such features are under-represented in a translator’s mental lexicon while he or she is translating. Nothing in the source text is likely to trigger them. This is an excellent candidate for the status of a universal: The phenomenon receives a cognitive explanation. (Malmkjær 2008: 55)

Sara Laviosa (2002) begins her section on simplification by warning about the mistake of considering simplification as a feature inherent in the translation process itself, when it easily could be accounted for by other explanations concerning social and cultural constraints. She gives a study made by Blum-Kulka and Levenston (1983) as an example of this. They claimed to have found through their research simplification strategies of translation that are of universal nature, but Laviosa repudiates these claims by providing social and cultural explanations for these strategies. She goes on to critically examine the main research that has been done since then but before the introduction of computerized corpora, research that she claims does not ‘give a clear picture of the nature and universality of simplification’ (Laviosa 2002: 47). Included in this review are Vanderauwera’s (1985) ‘analysis of the strategies and manipulations occurring in the English translations of 50 Dutch novels’ (Laviosa: 47), Baker’s (1992) discussion of ‘the different strategies used by professional translators for dealing with non-equivalence at word level’ (Laviosa 2002: 48) and Toury’s (1995) discussion of translation-specific lexical items and their lexicographical treatment, among others. What she finds to be common to all these efforts is the inadequacy of the methodological starting points. The corpora that they have used for their analysis are, with the exception of Vanderauwera’s, too small to give any clear statements about their findings.

Also, the selection of text types is too narrow and the studies, more often than not, are based on only translations from one language to another, and can therefore not say anything about the universality of the feature of simplification. Thus, her conclusion is that these studies serve better as guidelines for further research than as a complete theoretical account of the universal nature of this feature.

After showing us what is lacking in the earlier research in this area, Laviosa moves on to put forward her own work concerning simplification as a translation universal. As mentioned above, Laviosa was involved in the making and designing of the TEC, as part of the bipartite corpus named the English Comparable Corpus (ECC). The other part of this corpus (NON-TEC) consists of original texts in English and the two sub-corpora are evenly matched regarding text genre, time of publication, distribution of male and female authors, distribution of single and team authorship, overall size, and target-audience age, gender and level. She uses this corpus to test three hypotheses concerning simplification; less lexical variety, lower information load and use of shorter sentences in translations as compared to original texts. She found three types of evidence that supports her hypothesis regarding less lexical variety in translated language;

[f]irst of all the proportion of high frequency words versus low frequency words is relatively higher in translated texts. Secondly the list head of the corpus of translated texts...accounts for a larger area of the corpus, which means that the words most frequently used are repeated more often. Thirdly, the list head of the corpus of translated texts contains fewer lemmas. (Laviosa 2002: 62)

She also finds support for her second hypothesis concerning lower information load in translated language, as the evidence shows a lower percentage of content words versus grammatical words in translated texts than in original texts. However, the findings on her third hypothesis, use of shorter sentences in translated language, were inconclusive. What is most interesting with this study, however, is the methodology she has used; computerised corpora and computerised analysis tools, and all focussed on the translation products themselves instead of their relations to the source texts.

The universal of levelling out is mentioned by Baker (1996: 184) as one of the least investigated feature of translation. As mentioned above, levelling out adheres to ‘the tendency of translated text to gravitate around the centre of any continuum rather than towards the fringes’ (Baker 1996: 184). This is taken to mean that we would expect less variance in

textual features in translated language than in non-translated language (Olohan 2004: 100). Since little research has been done on levelling out as a translation universal, not much evidence has been provided for its viability. One study that does shed some light on the issue, however, is Laviosa's (2002) aforementioned investigation into lexical variety, information load and sentence length. The results here are conflicting, however, and only lend partial support to the suggestion that this feature is typical for translations. Pym (2008) suggests that this proposed universal contradicts the other proposed universals. He claims that if levelling out is a feature of translation, any extreme explicitation, simplification and normalisation would not occur since this would be to stray from the centre of a continuum (<http://www.tinet.org/>: 11).

As mentioned above, the focus of this paper will be on Baker's proposed universal of normalisation. While the Unique Items Hypothesis claims that features of the target language are neglected in translations, both simplification and normalisation are tendencies to adhere to typical features of the target language. Baker describes normalisation as 'the tendency to conform to patterns and practices that are typical of the target language, even to the point of exaggeration' (Baker 1996: 183).

Toury (1995) also takes up the issue of normalisation, though, perhaps, in a more indirect way. He identifies two laws of translational behaviour; the law of interference and the law of growing standardization. While the law of interference deals with influence from a source text, the law of growing standardization deals with the conversion of a source language item into a target language item. '[I]n translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by a target repertoire' (Toury 1995: 268). The deconstruction of the components in the source texts is not temporary, since the structures that appear in the target texts are made up of components that are more typical of the target language. Furthermore, Toury claims that the items chosen for the target texts tend to be selected on a lower level than the items in the source texts. This is very close to Mona Baker's proposals that translators tend to simplify and normalise the language in their translations. By choosing items or structures that are more general in nature than the items or structures in the source texts, the translators make the translations more easily accessible for the reader, and hence, more simple.

Some research has been done on normalisation in translation. Perhaps the most prominent study on this subject was made by Dorothy Kenny (2001). She investigated what happens to creative lexis in source texts during the process of translation. One way she



identified creative lexis, was to list the hapax legomena (word forms which occur only once) she could find in her specifically constructed corpus of German source texts. Then she excluded the words that fell under the categories of technical terminology, German verbs with separable prefixes, non-standard orthographic variations, words that appeared in standard dictionaries and words that had been used by other writers. She also identified items that were particular to a specific writer. Lastly, she identified unusual collocates of the German word *auge* (eye). These creative items were then compared with their translations into English. She found that in 44 % of these cases of creative hapax legomena, the translators chose to normalise in their translations. However, the author-specific creative items were not normalised by the translators, and only 25 % of the unusual collocates of *auge* were normalised. Hence, Kenny found evidence supporting the claim that translators tend to normalise, but also evidence supporting creativity by translators.

A study by Olohan (2004) compares the distribution of colour synonyms in non-translated English and translated English. In a previous study, she had found that *-ish* words in relation to colours were three times as frequent in translations as in non-translations. This led her to believe that translators make use of fewer synonyms for colours (and to a lesser degree) than writers of non-translated language. She found that her assumption was correct and that translators use significantly fewer colour synonyms than writers of non-translations. She takes this to mean that there is less variation in translated texts and, hence, that translators normalise their language.

In a critique of Toury's latter part of his book *Descriptive Translation Studies and Beyond*, Anthony Pym takes a closer look at Toury's law of growing standardization and how it pertains to Baker's proposed universals. He proposes that Baker has taken Toury's law of growing standardization and divided it into four parts (explicitation, simplification, normalisation and levelling out), and disregarded Toury's second law (of interference). Furthermore, Pym questions the separateness of Baker's universals, claiming that explicitation and simplification overlap when it comes to preference for finite structures, lower lexical density and lower type-token ratios in translations. These are all factors that make the text easier to read; which, certainly, is the purpose for both explicitation and simplification. Similarly, he also claims that simplification and normalisation overlap, since Baker lists standardization of punctuation under both of these universals. Finally, if these three universals overlap, and if they can be found in all translations, this would lead to a realisation of the fourth universal, levelling out. 'The norm is theoretically in the center of a bell curve, after all, and this fourth universal refers to the same linguistic variables as the

previous ones (lexical density, type-token ratio, sentence length and, in Shlesinger, explicitation and normalization)' (<http://www.tinet.org/>: 10).

Pym also attacks the very concept of 'universals', saying that if the proposed universals contradict one another, they can hardly be said to be viable at all times. He gives the example of sentence length. Simplification makes for shorter sentences, while explicitation calls for longer sentences. It can even be said that producing longer sentences in order to make the message more explicit is a type of simplification. In addition, finishing incomplete sentences is listed under normalization. Consequently, the term 'universals' does not give an accurate picture of what Baker's ideas contain.

Pym arrives at the conclusion that Baker's universals and Toury's law of growing standardization basically say the same thing (perhaps with the exception of explicitation, which is not mentioned explicitly by Toury). However, while Baker looks to universal (cognitive) explanations for her ideas, Toury moves in the direction of socio-cultural explanations. He claims that standardization is more likely to happen when the act of translation has low status in the target culture, while interference, Toury's second law, is more likely to happen if the source language or culture has a high status in the target culture.

According to Pym, in Baker, the socio-cultural aspect of explaining translation features seems more like an afterthought than something truly incorporated in her theory (<http://www.tinet.org/>: 13-14). Pym speculates whether the reason for this is that she is too focused on using comparable corpora to investigate possible explanations for her universals. Baker is of the opinion that we can determine how translations differ from non-translations, and find explanations for these differences, exclusively by using comparable corpora. When comparing translated texts with non-translated texts in the same language, however, the socio-cultural aspect can not be accounted for, especially since the source texts of the translations are likely to come from a variety of different languages. According to Pym, this might also be why Baker has not taken interference from source texts into consideration. Obviously, using comparable corpora to investigate the behaviour of translators does not provide for opportunities to see whether interference has happened, since the source texts are not given in these types of corpora (<http://www.tinet.org/>: 15).

This section has given the reader an account of the most important developments in the field of Translation Studies. For the purpose of this paper, the focus has been on the introduction of computerised corpora to the field and on the notion of translation universals. We have seen that corpora technology has been essential for the development of the field in that it opens up a whole range of new possibilities for research. The newfound ability to

investigate linguistic phenomena in large collections of texts has made it possible to make generalisations about translated language and, consequently, on cognitive processes during translation on a safer ground than before. It has made the investigation of proposed translation universals possible, since substantial generalisations such as these need to be tested against a huge amount of data. The next section will provide the reader with some complementary information on essential issues pertaining to this thesis, such as prototypicality, gravitational pull and basic verbs.

## **2. 2. Related research**

### 2. 2. 1. Prototypicality

From the era of the ancient Greeks and up to a couple of decades ago, one way of approaching the matter of categorisation has dominated. This way has been loosely called ‘the Platonic view’ and defines categories in terms of necessary and sufficient properties inherent in the single object. A ‘bachelor’ has the properties of [+ male] and [+ single]. Any object that can be said to have these properties is regarded as equal realisations of a ‘bachelor’. However, a new approach to categorisation has emerged from the field of cognitive psychology. This approach, called prototypicality, ‘claim[s] that natural categories contain good and less good examples, which possess a larger or smaller number of characteristic properties’ (Gilquin 2006: 160-161). Working with this approach, the prototype of a ‘bachelor’ would be ‘a 30-year-old single male who has not yet married, but [the category] includes other, more marginal members (e.g. a baby boy, a pope or a divorced man’ (Gilquin 2006: 161). As reflected by the above example, the notion of prototypicality was initially meant to be applied to the semantics area of (psycho-)linguistics, but it was soon also found useful in other areas. The emergence of large electronic collections of texts (corpora) made it possible to look at different linguistic phenomena in a wider scope. The discovery was that linguistic categories were not as discrete and clear cut as thought; rather, more or less good examples in a category were found, making the category boundaries fuzzy (see Taylor 1989).

Gilquin (2006) goes on to discuss the link between usage frequency and prototypicality. Many scholars are inclined to believe that the variant that is found to be most frequent in corpora is also the variant that is the most salient and, therefore, the prototype. However, there are other aspects that need to be taken into account. The prototype may be area-specific, in other words, only viable in a certain geographical or cultural area. Gilquin brings to mind an experiment carried out by Rosch (1975) which was later commented upon

by Aitchison (1998). Rosch found that the inhabitants of California regarded nectarines and boysenberries as more representative of the fruit-category than mangoes and kumquats. (One can, of course, discuss whether boysenberries are fruits at all, and not berries.) Aitchison reflects that this is not all that surprising since ‘nectarines and boysenberries are more common in California than mangoes and kumquats’ and that ‘[n]o doubt the results would have been different if the experiments had taken place, say, on the African or Asian continent’ (Gilquin 2006: 168).

Another aspect that seems to go against the viability of using frequency tests in corpora to establish the prototype within a category, is that results from elicitation tests meant to draw out what comes to a person’s mind first, often differ from frequency results from corpora. Gilquin refers to a comparison made by Shortall (2007) of the realisation of *there*-constructions between elicited data and corpus data. ‘Whereas in elicitation tasks people tend to produce sentences with concrete nouns...(about 60 % of the cases), in the British spoken section of the Bank of English, abstract nouns are predominant (59 %)’ (Gilquin 2006: 169).

Since sufficient evidence has been found to suggest discrepancies between prototypes found using elicitation tasks and frequency in corpora, Gilquin moves on to conducting a comparison between cognitive models of prototypical causation and the frequency results found from data taken from the BNC (British National Corpus). He finds that the causation structures thought to be prototypical ‘account for an astonishingly small proportion of the corpus data’ (Gilquin 2006: 175). Although he narrows the gap between the two somewhat by explaining how the corpus data may show some distorted figures, he still finds differences that need to be explained.

In the end, he suggests that

prototypicality is perhaps best described as a multifaceted concept, bringing together (1) theoretical constructs found in the cognitive literature and relying on deeply-rooted neurological principles as the primacy of the concrete over the abstract, (2) frequently occurring patterns of (authentic) linguistic usage, as evidenced in corpus data, (3) first-come-to-mind manifestations of abstract thought, as revealed through elicitation tests and (4) possibly other aspects that contribute to the cognitive salience of a prototype. (Gilquin 2006: 180)

We understand that the issue of prototypicality is not as clear cut as one might think. It might be that prototypicality is governed by many different factors, and it is difficult to establish which and to which degree they should be taken notice of.

On a similar note, Dawn Nordquist (2004) also questions the previous perception that prototypicality can be established by looking at frequencies of actual use (corpus frequency). She points to discrepancies between elicited data and frequent corpora patterns. She finds three possible explanations for this discrepancy. Firstly, that ‘different discourse pressures exist for each data type’ (Nordquist 2004: 212). It has been proposed that certain factors affect the features of elicited data, and that these factors are absent (or different) when it comes to corpus data. Secondly, she proposes that the two types of data have different registers. While elicited data can be said to be influenced by a certain degree of formality deriving from the setting of which it is retrieved, corpus data is of a more informal nature. Thirdly, Nordquist suggests that the forms of data retrieval may have different effects on the data. Elicited data is retrieved in an experimental setting, and she suggests that the speaker ‘attempt[s] to reduce the processing strain in an experimental context’ (Nordquist 2004: 212). This might produce more simplistic variants of the linguistic object in question in elicited data than in corpus data.

Nordquist goes on to investigate the third explanation in more detail, since she found that the first two did not affect her results:

While the discourse and sociolinguistic pressures remained constant throughout the experiment, the mismatch between elicitation and corpora did not. This suggests that factors other than discursive or sociolinguistic ones led to the patterns found in this study’s elicited data (Nordquist 2004: 212).

Furthermore, she finds another reason to investigate the psycholinguistic aspect of elicitation. Theoretically, infrequently used structures in conversation should also be infrequent in elicited data since their mental representation is weaker than other, more used, structures. ‘Frequent structures, on the other hand, should be easily accessed because their resting state of activation is high, facilitating usage’ (Nordquist 2004: 213).

Nordquist’s experiment involved an elicitation test where the subjects were asked to provide three utterances using each word that was presented to them on

index cards. She then compared the results from the elicitation tests with frequencies in conversational corpora. She found some differences between the elicited data and the corpus data. She explains these differences by suggesting that ‘lexically-specific, highly entrenched units will not be reproduced in elicitation...because of their autonomous mental representations and the higher likelihood of open choice processing in the context of elicitation’ (Nordquist 2004: 221). Rather, the mind looks to fill each open slot with more general representations of what they wish to express, since these representations are more easily accessible.

### 2. 2. 2. Gravitational pull

Sandra Halverson (2003) has explored how features of translation can be linked to human cognition. As discussed in Section 2. 1. 3., some scholars are of the belief that some features of translation may be proved to be universal. For a feature to be universal it is plausible to suggest that the feature appears because of cognitive characteristics in the human brain. If not, they can hardly be universal because other factors, which are more likely to be culture specific, govern whether the features appear or not. Therefore, if one is to say that a feature of translation is universal, it is important to connect it with a feature of cognition.

To relate the whole article by Halverson here would be too extensive for the purpose of the subject raised in this paper, since she spends a great deal of time introducing Langacker’s theory on cognitive grammar (for an overview of this theory, see Halverson 2003 or Langacker 1987, 1993, 1999). Hence, the following will be a short summary of the cognitive theories which are relevant for the purpose of explaining features of translations. It is thought that a symbolic unit has two poles; one semantic pole (concepts) and one phonological pole (lexis). On the semantic pole a network of similar or related senses is linked to a phonological unit. The phonological unit serves as a trigger for the semantic or conceptual network. Furthermore, it is assumed by bilingualism researchers that ‘bilinguals have one knowledge store, with various access routes, either via L1 or L2’ (Halverson 2003: 215). It is not necessarily so that a bilingual finds a one-to-one relationship between semantic or conceptual networks and their phonological representations in the two languages. A word may share all, some or none of the nodes in a conceptual network. This might lead to a bilingual finding no phonological counterpart in the other language of a conceptual network that has its phonological representation in the first language. As Halverson points out, ‘these could be exemplified, for instance, by so-called “culture-specific concepts”’ (Halverson 2003: 215).

At the other end of the scale, ‘there may be networks in which the two phonological nodes share all conceptual nodes, e.g. in words with highly concrete meanings’ (Halverson 2003: 215). A phonological access point in one language follows the route to a concept network which again follows a route to a phonological representation in the other language. When a routine such as this has been repeated often enough, the links become stronger and the connections are made quicker. Also, the more nodes two phonological representations share on the conceptual level, the quicker the connections are made. The words that are language specific are accounted for by assigning metalinguistic knowledge to the networks. This is knowledge a bilingual person needs in order to cope with the words that are culture-specific and, hence, not directly translatable.

Halverson goes on to explain that some elements in the networks are more central than others. These are the nodes that are ‘linked to domains of space, vision, and sensory experience, those at a certain level of schematicity (basic-level categories) and those that are most deeply entrenched (Halverson 2003: 216). These nodes are made more salient by the category prototype principle and the highest level schema in a gravitational pull situation. Halverson refers to de Groot (1992a, 1992b, 1993) who has shown that translators translate faster and at a higher degree of precision when it comes to words of a high degree of concreteness. De Groot links this to the suggestions that concrete words are more likely to be represented in much the same way across languages than abstract words. However, de Groot does not attribute this to the concreteness of the words, but rather for the quality of concrete words to conceptually overlap across languages because of their centrality for human beings. This pertains to the notion of basic-level, since ‘the basic level is defined as that category level that is most cognitively significant to humans on the basis of perceptual and functional (behavioural) characteristics’ (Halverson 1993: 217).

The nodes that are “most deeply entrenched” are the nodes that are most frequently activated. Entrenchment is a major factor in the determination of prototypicality and also for establishing higher level schemas. In other words, it is crucial in determining what element comes to mind first when a concept network is activated. When it comes to translation, the words that are highly entrenched (basic level) in L1 are also typically highly entrenched in L2, and this strengthens the link in both languages to the conceptual network they connect to.

Halverson claims that the prominence of deeply entrenched basic level categories leads to an over-representation of the highly salient words and structures that are connected to them in translations.

In a translation task, a semantic network is activated by lexical and grammatical structures in the ST. Within this activated network, which also includes nodes for TL words and grammatical structures, highly salient structures will exert a gravitational pull, resulting in an overrepresentation in translation of the specific TL lexical and grammatical structures that correspond to those salient nodes and configurations in the schematic network (Halverson 2003: 218).

She backs this statement up by referring to empirical studies made by a score of scholars investigating the translational features of simplification, generalization, normalization, sanitization, conventionalization, exaggeration of TL features and the law of growing standardization (Toury 1995). However, she takes care to make a point that “the gravitational pull posited here is in no way meant to function in a deterministic way” (Halverson 2003: 220). Even though gravitational pull is in effect, it can be overridden by stronger motivations.

While concrete words are typically highly salient, abstract words are not. De Groot showed that translators made more errors, worked more slowly and made more omissions when dealing with abstract words. Also, Halverson predicts greater variation in translations when it comes to culture-specific words (words that share no nodes between two languages on the conceptual level). She claims that this might be the cause for what is found in many empirical studies, namely that translators tend to under-represent words of low salience. For example, the Unique Items Hypothesis claims that items that are culture-specific for the target language have a strong tendency to be left out of translations into that language (see Tirkonnen-Condit 2004).

### 2. 2. 3. Basic verbs

In a paper from 2002, Åke Viberg puts focus on second language (L2) learning, and the problems that lie therein. Some of these are ‘the isolation of word forms...in more or less continuous strings of sounds with no clear boundaries’ (Viberg 2002a: 52); ‘the identification of word meanings with the help of linguistic cues...or cues from observations of the situational context’ (Viberg 2002a: 52). To look into these problems, Viberg uses basic verbs as examples. These are the verbs that are most frequently in use in any given language, like *vara* (*be*), *ha* (*have*), *kunna* (*can*) and *ska* (*shall*) in Swedish. He divides these verbs into further sub groupings; nuclear verbs are verb meanings that achieve the status of basic verbs in all languages (*take*, *give*, *see*, etc); areally specific basic verbs are verb meanings that achieve basic verb status in only some languages (*can*, *must*, *have*, etc). Viberg observes that



acquisition of nuclear verbs for L2 learners corresponds to Levinson's (2001) degree 1 of mapping word forms unto concepts. This means that we are dealing with a mapping of language-specific word forms unto universal semantic units. The acquisition of areally specific basic verbs for L2 learners, however, corresponds to Levinson's degree 2 of mapping word forms unto concepts. In other words, areally specific basic verbs are mapped 'unto language-specific meanings constructed from universal concepts' (Viberg 2002a: 52).

However, Viberg notes that even if a verb is nuclear, this does not mean that there is a one-to-one relationship for the verb between languages. The core meaning may be the same, but one language may either assign it additional meaning so that it covers a larger domain of meaning or a language may retract meaning from it so that it covers a smaller domain of meaning (Viberg 2002a: 55). Viberg refers to a series of studies that showed that 'intertranslatability for nuclear verbs tended to range between 35% and 45%' (Viberg 2002a: 55). This means that only between one in two and one in three of the times that a translator is to translate a nuclear verb from one language to another he/she opts for the basic verb equivalent in the target language. Due to the variance of meaning extensions of nuclear verbs across languages, it would be plausible to assume that L2 learners would (at least at first) not master these extensions fully and therefore we could detect less use of the peripheral meanings of the verb in their language use.

Viberg also notes that early L2 learners have a clear tendency to over-represent nuclear verbs in their language production. 'The favouring of nuclear verbs to a greater extent than by native speakers is characteristic of L2 learner speech for an extended period of time' (Viberg 2002a: 61). He goes on to explain this preference of nuclear verbs over other verbs that could yield the same meaning.

The favouring of a small number of verbs can best be explained with reference to processing capacity. L2 learners...favour nuclear verbs even more than L1 speakers due to a general high load on processing capacity. There is an additional reason for this type of speakers especially at early stages. Nuclear verbs are often the best choice when the lexical repertoire is very restricted due to their semantic coverage and the choice of a nuclear verb can in many cases be interpreted as a communication strategy. (Viberg 2002a: 66)

In another paper (Viberg 2002b), Viberg takes a closer look at one of these basic verbs and how it differs in use and behaviour across languages. The verb he investigates is *get* and its

Swedish counterpart *få*, and his data are retrieved from the English-Swedish Parallel Corpus (ESPC) which is a corpus compiled of original texts and their translations in both English and Swedish. His data consists of the occurrences of *få* in Swedish original texts in ESPC and a random sample of occurrences of *få* in the Stockholm-Umeå Corpus (SUC), along with their translations into English, French and Finnish. He lists the various meanings the verb has in Swedish and compares them with their counterparts in English, French and Finnish translations. He finds that in the cases where *få* is used in what Viberg calls its ‘basic meaning’ (possession) the most common translation into English is the basic verb *get*. In the other less prototypical senses of the verb *få* (modality, causative, etc), however, the translations vary greatly and *get* is seldom used. In general, Viberg’s findings show that when it comes to the peripheral senses (less prototypical senses) of the verb *få*, translations like *may*, *be allowed*, *can* and *make* are preferred instead of the basic verb correspondent *get*. He notes that ‘in spite of a strong universality at the conceptual level, the lexicalization patterns are very language-specific at a more detailed level’ (Viberg 2002b: 147). Viberg is of the opinion that the lack of correspondence between basic verb-pairs in different languages calls for ‘a detailed contrastive analysis, which can be used for applied purposes such as translation and language teaching’ (Viberg 2002b: 147).

This section has presented issues which are essential for this thesis, but which, although related to it, lie outside the field of Translation Studies. The notion of prototypicality, gravitational pull and basic verbs are all elements that can be implemented in research pertaining to Translations Studies, which is exemplified by the present thesis. The aim of this thesis is to investigate Baker’s (1993) proposed universal of normalisation. The idea is to establish prototypicality in the semantic structure of the basic verb *take* and to see if gravitational pull might be a contributing factor to the tendency for translators to adhere to, and exaggerate, typical features of the target language. The prototypical senses of *take* will be seen as typical features of the target language, and if these are over-represented in translations in relation to non-translations, this will be seen as support for the claim that translators tend to normalise in translations.

## **CHAPTER 3: MATERIAL AND METHOD**

This chapter will have a closer look at the main elements and data that are essential for my analysis. First, the corpora that will be used are presented and elaborated upon, and then a description of how the data is extracted will follow. Lastly, the method that will be applied is thoroughly accounted for.

### **3. 1. Material**

The data collected are examples of the verb *take* used in written texts by authors of non-translated texts and translated texts. The selection of *take* was done on the basis of Viberg's (2002a, 2002b) papers on basic verbs, where he identified *take* as one of these (see Section 2. 2. 1.). Such a verb is ideal for an investigation of normalisation in translation where the object is to search for over-representation of prototypical members in a category. Halverson (2009) has made a similar selection in a similar study focusing on another basic verb:

In looking for a test case for the current purpose, a useful starting point is a lexical item for which the schematic network might be expected to be relatively rich, and for which there is posited a schema or prototype. One such set of items includes the so-called "basic" verbs, which is the set of most frequent verbs in a given language. (Halverson 2009: 92)

#### **3. 1. 1. English-Norwegian Parallel Corpus (ENPC)**

The English-Norwegian Parallel Corpus (ENPC) was originally a research project at the Department of British and American Studies at the University of Oslo in cooperation with Norwegian Computing Centre for the Humanities at the University of Bergen. It consists of a set of English texts and their translations into Norwegian, and also a set of Norwegian texts and their translations into English. The initial thought was to use it in contrastive linguistic studies and since the designers realized the problem with comparing texts with their translations for this purpose, they found it necessary to include texts and translations in both

languages. (Olohan 2004: 57) A consequence of this decision was that the ENPC also became interesting for scholars in the field of translation studies. The size of the corpus is approximately 2.6 million words, distributed evenly in the four sub corpora (Norwegian original texts, Norwegian translations, English original texts, and English translations). Thus, in each of the four sub corpora, there are approximately 650 000 words. The texts can also be divided into fiction and non-fiction; however, the texts containing fiction dominate the corpus with its 60 percent in relation to non-fiction which only occupies 40 percent of the total corpus. The fiction texts include detective novels, children's novels and general fiction, with the latter dominating the collection with nineteen out of thirty novel extracts. The non-fiction texts include works on religion, social sciences, law, natural sciences, medicine, arts and geography and history (<http://www.hf.uio.no/ilos/forskning/>).

The texts that have been included in the ENPC are extracts of the complete works of authors and translators. The decision to include extracts instead of full texts was governed by two main considerations: firstly, the problem of obtaining copyright permissions is avoided; and secondly, it opens up for a greater variety of authors/translators in a corpus of this relatively small size. The extracts range from between 10 000 to 15 000 words in size and all of them start from the beginning of the full texts, but 'front matter - prefaces, forewords, list of contents, etc. - is not included in the extracts. In some cases, introductions have been left out as well, e.g. introductions by scholars to works of fiction' (<http://www.hf.uio.no/ilos/forskning/>). The variation in size of the extracts relates to the reluctance to cut off an extract in the middle of a chapter. Furthermore, the ENPC was 'scanned and encoded with bibliographical information and a text classification, following the TEI guidelines' (Olohan 2004: 57).

Scholars have made use of this corpus (and others like it) in a number of different ways. Because it consists of original texts and translations in both languages (English and Norwegian), several distinct comparisons may be made. One can compare original language in one language with original texts in the other language, translations into one language with translations into the other language, original texts in one language with translations into the other language, and original texts in one language with translations into the same language. It is clear that a corpus designed in this way becomes a valuable tool for scholars within translations studies as well as scholars of general linguistics. Another corpus that is an invaluable resource for any scholar connected to the field of translation studies is the Translational English Corpus.

### 3. 1. 2. Translational English Corpus (TEC)

The Translational English Corpus was designed mainly by Mona Baker and Sara Laviosa. Baker was the one to suggest this new type of corpus (Olohan 2004: 59). The TEC consists of translations into English from a variety of European and non-European source languages and is divided into four main sections; fiction, biography, newspaper texts and in-flight magazines. However, the majority of the corpus is made up of fiction and biography texts (together approximately 95 percent of the corpus). According to the TEC web page (<http://www.monabaker.com/>), the total size of the TEC is approximately 10 million words. However, a count made by the present writer showed that the fiction section contained almost 5 million words while the non-fiction section contained almost 500 000 words, bringing the total close to 5.5 million words. It was important to sort these figures out for two reasons; first, it was necessary for a preliminary test to establish the frequency of the verb *take* in the corpus, a test which compares the total number of occurrences with the total number of words in the corpus; second, it was important to establish the size relation between the section of the corpus containing fiction texts and the section containing non-fiction in order to be able to compare the corpus with other corpora.

To arrive at these figures, I found it necessary to exclude thirteen of the text extracts in the non-fiction section due to lack of information on the size of these texts. These texts will also be excluded from the pending analysis. The exclusion of some texts from my count will certainly explain some of the deviance between the information about corpus size found on the above web site and my own count. But it is unlikely that the excluded texts would amount to 4.5 million words in order to reach the total of 10 million words which the web site claims. Therefore, we can only assume that an unknown reason (copyright issues, temporary reconstructions of the corpus, etc) has forced the administrators to remove some texts from the corpus (indeed, it seems like the TEC was under some kind of reconstruction or update period at the time I collected my data, since a new version has substituted the one I used). In any case, the exclusion of the thirteen non-fiction texts means that the fiction section amounts to over 90 percent of the total corpus. When it comes to the issue of sampling the texts, the designers of the TEC have avoided this issue by prohibiting direct access to the texts by the users. The user only gains access to the immediate co-text of the examples he/she has searched for, and in this way, copyright issues have been sidestepped.

Extra material like footnotes, picture captions, endnotes and so on are automatically excluded so that we are left with the main texts to search in, but the extra material can be

included if one wishes. There is also the possibility to define your search in terms of source authors, source language, name, nationality, gender or sexual orientation of the translator. This is made possible by the online TEC Concordance Browser which sorts the texts by information that has been collected through questionnaires from publishers and translators and added to the texts' header files. This metadata 'do not adhere to any particular annotation guidelines (such as TEI) or metadata scheme (such as Dublin Core)' (<http://www ldc.upenn.edu/>). Nevertheless, it provides us with vital information and options that correspond to these types of guidelines.

The TEC has mainly been used for two kinds of research: comparing translated language (into English) with original language (English) and uncovering stylistic variance between translators. Of course, in order for this corpus to be used as a comparable corpus, it needs a corpus of original texts to be compared with. The British National Corpus (BNC) may serve this purpose, as it contains the same kind of sectioning (fiction, biography, etc.), although it is a corpus of a considerably larger size than the TEC. The BNC is the next corpus we will have a look at.

### 3. 1. 3. British National Corpus (BNC)

The British National Corpus (BNC) was created under the management of the BNC Consortium which, in turn, is led by The Oxford Press. When designed it was thought that this corpus could be applied in areas of reference book publishing, academic linguistic research, language teaching, artificial intelligence, natural language processing, speech processing and information retrieval in addition to the obvious linguistics research advantages gained by such a tool. It is a monolingual corpus consisting of original texts written in the English language. The total number of words in this corpus is in excess of 111 million words. It can be divided into fiction and non-fiction with nearly 20 million words in the fiction section and about 91 million words in the non-fiction section. The BNC is compiled of both written (90 percent) and spoken (10 percent) language. The texts are written in the latter part of the 20<sup>th</sup> century and collected in the period between 1991 and 1994 (<http://www.natcorp.ox.ac.uk/>).

In order to achieve a wide representation of texts and to avoid over-representing idiosyncratic texts, the size of each text has been limited to 45 000 words, which means that sampling has been involved for the longer works. The target size for each sample is 40 000 words. Furthermore, the samples begin randomly from the beginning, middle or end of the

texts, and natural cut off points, like the end of a chapter, are preferred where possible. For copyright reasons, texts that are smaller than 45,000 words are further reduced by 10 percent. A wide variety of texts is crucial for this corpus, as it is meant to represent British language as a whole, not just the most typical version. The BNC's header files are using marking and tagging in line with the TEI guidelines, which makes it easy to compare our findings with findings from other corpora, e.g. the TEC. Finally, it is important to recognize that I have accessed the BNC through the Sketch Engine (<http://www.sketchengine.co.uk/>), also called Word Sketch Engine. Sketch Engine is a tool that provides a full account of a word's grammatical and collocational behaviour, in addition to normal corpus information (a word in its context).

The next section compares the corpora I have selected to work with.

#### 3. 1. 4. Comparing and adjusting the corpora

As seen above, there are some differences between the three corpora that have to be taken into consideration if one is to compare data taken from them. The most notable difference is perhaps the size. The ENPC (2. 6 million words) and the TEC (5. 5 million words) are relatively small corpora in relation to the BNC (111 million words). Also, there is the issue of representation of fiction vs. non-fiction, which also has to be dealt with since there exists great variance between the corpora. Luckily, these are problems that can be overcome.

The difference between the corpora regarding total size is of least concern, since a random selection of examples of the verb *take* will be taken in order to decrease the number of occurrences. In other words, data samples will be made from each corpus, each containing 250 instances, which is a more manageable amount of data. The total number of occurrences of *take* in a corpus will be divided by a number that yields 250. Then, the number that the total number of occurrences is divided by will be used to select instances, e. g. every sixth instance is drawn out of the corpus and used in the sample. In this way, we eliminate the problem of different corpus sizes and we avoid any questions concerning biased and unrepresentative sets of data since the selection is systematic, but still random. When it comes to the BNC, the sampling process will happen in a slightly different way. The Word Sketch Engine provides us with the possibility to select samples automatically. This is also perfectly random, so the questions of biased and unrepresentative sets of data are avoided yet again.

However, there is one aspect of data sampling that complicates matters a bit. This is the factor of verb form. The number of occurrences of each verb form (*take, took, taken, taking* and *takes*) in the sample should represent the total number of occurrences of each verb form in the full corpus. This is no problem for the BNC, since it is a lemmatized corpus, which means that all the verb forms are included in the search, and because the sampling is stratified. In the other corpora, on the other hand, the different verb forms must be searched for separately. This means that in order to get the correct representation of verb forms in the sample, we need to do the above procedure for each form. Then we get the same percentage of each tense in the sub-corpus as in the full corpus.

The distribution of text types (fiction/non-fiction) is also not quite comparable between the corpora. In ENPC, about 60 percent of the corpus is compiled of fiction texts, and the rest (40 percent) are non-fiction texts. By contrast, over 90 percent of the TEC is compiled of fiction texts, while only ten percent is non-fiction texts. In the BNC it is the other way around with a vast majority of the corpus consisting of non-fiction texts. This creates obvious problems since a quick search in the three different corpora shows that there are notable differences in the distribution of the verb *take* in relation to the total number of words in the sub corpora (fiction and non-fiction). This means that steps have to be taken in order to achieve the same relation between fiction and non-fiction in the three corpora. In the ENPC and the BNC this seems to be an easy task, as the ENPC has a 60-40 percent relation, which corresponds nicely with part of the BNC when we exclude all the written-to-be-spoken and all unpublished texts, and only include leisure non-fiction (which corresponds best with the non-fiction found in the other two corpora). So now we have 60 percent fiction and 40 percent non-fiction in both the ENPC and the BNC. In order to get the same in the TEC, we need to exclude a large number of the fiction texts. The most logical thing to do is to begin the excluding process from the end of the text list which has been ordered alphabetically. It has also been found necessary to exclude two texts in the middle of the list due to the variety in text size, in order to achieve the number of words that is wanted. Now we have a 60-40 percent relation between fiction and non-fiction in all three corpora.



## **3. 2. Method**

### **3. 2. 1. What to include**

Now that we have established the material from which to extract our data, we can move on to the considerations concerning whether any data should be excluded from a linguistic point of view. Since what we are interested in here is the verb *take*, it goes without saying that any instances of *take* as a noun-form should be disregarded in this context.

However, a selection of data will do us no good if we do not have a framework for the analysis of this data. In a project such as this it is crucial that the tools for analysis are as precise and inclusive as possible. After an extensive search of the internet, including various linguistic search engines (JSTOR, BIBSYS, Linguistics Abstracts Online, Translation Studies Bibliography, etc) I found that a thorough semantic analysis on the verb *take* had not yet been done. Therefore, I had to find a way to do this by myself. A semantic analysis like this could probably have been a thesis in itself and the time frame assigned to finish this thesis does not allow me to make a very thorough and wholly scientific list of semantic categories of the verb. However, I believe the approach to the question of the structure of *take* is quite satisfactory for the purpose at hand. The framework will be employed in my analysis of the examples I have extracted from the corpora described in 3. 1. This will be elaborated upon later in this chapter.

### **3. 2. 2. Deriving a hypothesis for category structure**

Operationalization is achieved through careful use of lexicographical resources. In some respects, semantic analyses of words are available for everybody. What I speak of here is the existence of comprehensive dictionaries. Especially the most reputable publishers, like Oxford University Press (<http://www.askoxford.com>), Pearson Longman (<http://www.ldoceonline.com/>) and Collins Cobuild (<http://www.collinslanguage.com>) have lexicographers employed in developing lexical resources. In addition to dictionaries, we have other linguistic tools that are quite helpful in establishing semantic qualities of a word. One of these is Princeton University's Wordnet. (<http://wordnet.princeton.edu/>). Wordnet is a lexical database consisting of 117,659 sets of synonyms. In some ways it is like a dictionary, however it deals only with words belonging to the open word classes (nouns, verbs, adjectives and adverbs).

I have now introduced the online dictionaries by Oxford, Longman and Collins Cobuild and Wordnet's lexical database. These are the aids I have used in order to arrive at a list of senses for the verb *take*. The dictionaries have all used some kind of criteria for ordering the senses listed for a word. They all used frequency of occurrence as a factor in deciding the order, however some to a higher degree than others. Longman and Collins Cobuild, for example, use only frequency as criteria, while Wordnet use frequency combined with common sense. Oxford use frequency of occurrence combined with a notion of centrality presumably derived from common sense (information about the criteria used ordering the senses in the dictionaries was obtained through correspondence by e-mail with the publishers). I searched for *take* in the four linguistic tools and then, as shown in Table 3. 1. below, I took the first sixteen entries in each of these four and placed them in the table. By doing this, I got an impression of what semantic variants of *take* existed and also, an impression of what the most frequent variants were.

Table 3. 1. Dictionary listings

	<b>Wordnet</b>	<b>Longman</b>	<b>Oxford</b>	<b>Cobuild</b>
<b>1</b>	Take action	Take action	Hold, get hold of	Move, remove, hold, get hold of
<b>2</b>	Occupy, use up	Move	Occupy	Bring, carry
<b>3</b>	Direct, lead	Remove	Capture, gain possession	Accept
<b>4</b>	Hold, get hold of	Require, need	Bring, carry, convey	Travel
<b>5</b>	Assume, adopt	Accept	Remove	Suit, fit
<b>6</b>	Interpret	Hold, get hold of	Subtract	Consume
<b>7</b>	Bring	Travel	Consume	Consider
<b>8</b>	Take into possession	Study	Bring into a specified state	Require
<b>9</b>	Travel	Do (a test)	Experience, feel	Participate
<b>10</b>	Choose, select	Suit, fit	Travel	Teach
<b>11</b>	Accept, have	Collect	Accept, receive	Blank
<b>12</b>	Fill, occupy	Consider, react	Acquire, assume	Blank
<b>13</b>	Consider	Experience, feel	Require, use up	Blank
<b>14</b>	Require, demand	Get possession	Hold, accommodate	Blank
<b>15</b>	Experience, feel	Consume	Act on	Blank
<b>16</b>	Film, shoot	Elevate, follow up	Regard, view, deal with	Blank

Some issues have to be taken up here. In some cases (especially in the case of the sense of *change of possession*), some of the dictionaries elected to divide the senses into definitions

that I found very similar to each other. A good example of this is that of *capture*, *steal*, *get hold of*, etc. Here, I made a category ('hold', 'get hold of') that encompassed all these versions of change of possession and listed it where the first instance of this sense appeared. Another issue that arose was that the dictionaries did not always agree on the boundaries of each definition. An example of this is Cobuild, which merged the sense of *movement* with the sense of *possession*, which none of the other dictionaries did. I decided that this was no real problem and listed these cases as they were. One last issue is that Cobuild did not provide enough definitions to fill all the slots. This is the reason why I have entered *blank* in the last six slots.

As we see in the table, the four sources do not agree on what categories are the most frequent. For example, Wordnet and Longman agree that the meaning 'take action' is the most frequent semantic variant of *take*, but Oxford have decided to put this variant considerably further down on the list and Cobuild has decided to keep it out of their list of definitions entirely. In fact, Oxford also very nearly excludes it entirely from the top sixteen, as it appear in a slightly different form only in slot fifteen. This meant that if I were to construct a list of senses for *take* where I start out with the most frequent sense, then the second most frequent sense and so on, I had to take steps to ensure that this would be possible. I solved this problem by giving each sense from each source a value depending on where on the individual list it appeared. By this method, the category of 'take action' from the Wordnet would get the value one. Similarly, the same sense from Oxford would get the value fifteen. I then added the values a category had received in each dictionary/resource and created a new list. The category with the lowest total value was placed as number one; the category with the second lowest total value was placed as number two and so on. In Table 3. 2. we can see how this was done.

One problem that arose in this process was that sometimes a sense was not represented in every dictionary/resource. In these cases, I could not give this sense a zero value, because that would move it higher up in the final list than it deserved to be. Therefore, I gave these cases a value of sixteen to make sure that the sense as a whole did not climb into a place that was unjustifiably high. Another challenge I had to face in making this list was that some of these senses were so semantically similar that I decided to merge them. This was the case with the two senses 'move' and 'remove'. This resulted in two things: firstly, the situation where one sense in the final list had two values from the same dictionary/resource. The sense of 'move', 'remove' in the final list brought with it the values 2 and 3 from the Longman

dictionary and the values 3 and 9 from Cobuild. In these cases I decided to use the value of the sense that was closest to the top of the list, because merging should not take importance away from a sense. Secondly, it resulted in the situation where new senses entered the list. One example of this is Wordnet's 'contain', 'hold', which entered the list at 14<sup>th</sup> place, since some of the higher-ranking senses merged. Using these criteria to establish the order in which the categories should be listed, I ended up with a list that I could use to make a pilot test analysis of my data. This list is shown in Table 3. 2. below.

Table 3. 2. Dictionary listings revised

No.	Sense	Wordnet	Longman	Oxford	Cobuild	Total
1.	Remove, move	3	2	4	1	10
2.	Bring, carry	3	2	3	2	10
3.	Hold, get hold of	4	5	1	1	12
4.	Require	2	3	10	9	24
5.	Accept	8	4	8	4	24
6.	Travel	7	6	7	5	25
7.	Carry out	1	1	11	16	29
8.	Consider	6	9	12	8	35
9.	Consume	12	11	6	7	36
10.	Occupy	9	16	2	16	43
11.	Participate	13	7	13	10	43
12.	Experience	10	10	8	16	44
13.	Suit, fit, use	14	8	16	6	44
14.	Assume, acquire	5	16	9	16	46
15.	Subtract	16	12	5	16	49
16.	Film, shoot	11	16	16	16	59

A pilot test analysis was needed in order to find out if my list of semantic categories of *take* was good enough. After a couple of revisions following the pilot test analysis, I ended up with this list of categories:

1. Move, remove (something or someone)  
**He took his feet off the table.**
2. Bring, carry, convey (take with)  
**She took the knife with her.**
3. Hold, get hold of, get possession, collect, steal, capture (change of possession)  
**Take this book as a present.**
4. Require, demand, call for

- It takes courage to listen to heavy metal.**
5. Accept, select, choose  
**I can't take it anymore.**
6. Travel  
**It's easier to take the bus.**
7. Carry out, perform (action)  
**She is taking steps to ensure that will not happen.**
8. Consider, regard, view, *interpret*  
**I'll take that as a no.**
9. Consume  
**You should take your medicine.**
10. Occupy (place, time or position)  
**The event took place at the park.**
11. Participate, *attend*  
**Not everybody took part in the event.**
12. Experience, submit to, *feel*  
**I take envy in your ability to stay on top of things.**
13. Contain, suit, *fit, use*  
**I take size 43 in shoes.**
14. Assume, acquire, adopt  
**He takes after his father in that respect.**
15. Subtract  
**If you take three from five, you are left with two.**
16. Film, shoot  
**This picture was taken in the Rocky Mountains.**
20. Idiom  
**You should learn to take it easy.**

As we can see, some modifications have been made with regards to the replacement verbs that I have used to classify my examples according to categories. The replacement verbs in italics in the list above are the ones that have been added after the pilot test. These were not listed in the dictionaries, but included in order to expand the categories so that similar senses fit into them. If the example of *take* in my data can be replaced by one of the replacement verbs, this is a good indication as to which category the example falls into. For example, in 'she *took* the

knife with her', *took* can be replaced with 'brought' without losing its meaning; therefore this example of *take* belongs to sense category 2. As I analysed the examples from my data, I found the need for more replacement verbs. In other words, I have expanded the categories somewhat to encompass a greater variety of meaning in each category. I do, however, believe that this will not affect the core semantic values of the different categories. The replacement verbs written in italics are the ones that were not in the dictionaries, but that I nevertheless thought necessary to include after the pilot run.

The observant reader has no doubt noticed another change from Table 3. 2. An additional sense category has been added at the bottom, namely that consisting of idioms. John Saeed (2003) defines idioms as 'expressions where the individual words have ceased to have independent meanings' (60). He also compares them to 'collocations [that] undergo a fossilization process until they become fixed expressions' (Saeed 2003: 60). An example of idioms found in my data is shown in [1] below:

[1] "You'll have to *take* it easy," he said, turning away. (ENPC, DL1)

This example is taken from ENPC non-translations. We can see that the words do not denote any rational meaning individually; they need to be taken in as a whole in order to yield any meaning.

As mentioned at the beginning of this section, this is a simplified structure of a prototype category. This is down to the approach used to determine prototypicality in the category, using dictionary listings of senses of *take* as sources. Ideally, the list of senses should have been a product of careful investigation, including corpus frequencies, elicitation tests and studying cognitive literature as proposed by Gilquin (2006) (see Section 2. 2. 1.). However, the list of senses has some value, at least for the purpose of this thesis. It gives us an idea of which senses of *take* are central and which senses are more peripheral. For instance, the senses dealing with 'movement' (categories 1 and 2) and 'possession' (category 3) can be said to be more central than the senses denoting 'subtraction' (category 15) and 'filming' (category 16). It provides us with a working definition of the prototypical structure of *take*.

Now that the data and tools for analysis are settled and done with, we may safely proceed to the analysis itself and how this will take form.

### 3. 2. 2. Treatment of the data

If we are to analyse a set of data, we need a framework in which this will be done. First of all, the programme that I have used for the analysis is Microsoft Excel. Excel provides us with a suitable tool for processing data, and in addition, it gives us the possibility to organize the processed data in tables that the desired information may be extracted from. The second thing that should be accounted for is the categorization that the data will be subjected to. First, it is crucial to record which corpus the individual observations have been extracted from.

Therefore, each corpus was identified by number, and each observation was coded for the source corpus. The coding is as follows:

#### **Corpus**

1. ENPC non-translations
2. ENPC translations
3. BNC
4. TEC

By marking each example by what corpus it stems from, I may easily divide the results into the different corpora and then extract the information that is needed.

Also, it may prove useful to mark each example with an identification tag. This makes it easy to locate the example in my list of examples if any problems should occur, or if I need to integrate an example in this thesis. This will be done by giving each example, both in the data collection and in the analysis, a number according to the following:

#### **Identification**

1. 1001-1250 (ENPC original)
2. 2001-2250 (ENPC translations)
3. 3001-3250 (BNC)
4. 4001-4250 (TEC)

In addition to coding the examples according to which corpus they are extracted from, it might be useful to code them according to which *kind* of corpus they are extracted from; in

other words, whether the example came from a corpus containing non-translated or translated texts. This is undoubtedly an invaluable shortcut when it comes to analyzing my results. The coding is as follows:

### **Non-translated/translated**

1. Non-translated
2. Translated

The data will also be classified by the semantic categories listed above, although with some more categories added. As mentioned above in Section 3. 2. 2., I have used replacement verbs as a substitution test as a means of classifying the examples in my data according to sense categories. However, some of the examples from my data were ambiguous and impossible to classify in terms of one single category. Although these instances were few, they still needed to be accounted for. Therefore, I created three more categories (31, 32 and 33) in order to deal with them in a structured manner. The full list of sense categories the data will be put in is as follows:

### **Senses**

1. Move, remove (something or someone)
2. Bring, carry, convey (take with)
3. Hold, get hold of, get possession, collect, steal, capture (change of possession)
4. Require, demand, call for
5. Accept, select, choose
6. Travel
7. Carry out, perform (action)
8. Consider, regard, view, *interpret*
9. Consume
10. Occupy (place, time or position)
11. Participate, *attend*
12. Experience, submit to, *feel*
13. Contain, suit, *fit, use*



14. Assume, acquire, adopt
15. Subtract
16. Film, shoot
20. Idiom
31. Ambiguous movement/possession (1/2)
32. Ambiguous movement/take with (1/10)
33. Ambiguous carry out/idiom (3/20)

Also, I wish to code the data according to text type. The preliminary testing of the data showed differences in the distribution of *take* between fiction texts and non-fiction texts. As a result, I deem it necessary to make a distinction between the two by marking them. The coding is as follows:

### **Text type**

1. Fiction
2. Non-fiction

Another type of categorization is that which deals with verb forms. If I should want to separate my results in terms of verb form in order to find differences between them, it will be easy to do so when I code for that in my analysis. The coding is as follows:

### **Verb form**

1. -∅
2. past (took)
3. -n
4. -ing
5. -s

Some of the sense categories above have the potential to include examples that have a concrete or an abstract object. This is the case with categories 1 ('move', 'remove'), 2 ('bring', 'carry'), 3 ('hold', 'get hold of'), 5 ('accept', 'select', 'choose'), 6 ('travel') and 10 ('occupy'). In this light, I find it interesting to mark each example in these sense categories according to this quality of the object. The codes are as follows:

## Object

1. Concrete
2. Abstract

Having coded my data, tables will be made to reflect the results. The chi-square test will be performed where feasible in order to test for statistical significance. However, in order to do so, some sense categories need to be collapsed. This means that we will lose information about the significance of the more peripheral sense categories with the lowest frequencies. However, the more prototypical categories will not be affected.

Chapter 4 will go more into depth on the semantic categories of *take* in relation to the results of my analysis.

## **CHAPTER 4: RESULTS AND DISCUSSION**

While the last chapter was dedicated to preparing for the analysis, this chapter will give an account of the results of the analysis. After the results have been presented a discussion will follow, where an attempt will be made to consider possible explanations for the aforementioned results.

### **4. 1. Results**

This section will be devoted to the presentation of the results of my analysis. In this analysis, I have coded a selection of data (1,003 examples in all from four different corpora) for sense category, corpus, text type, verb form and concrete/abstract object. During the next few pages, I will present the results I obtained, starting with a general overview of the sense categories and the distribution of *take* within them in each corpus. Next, I will show some generic patterns, before moving on to presenting the way *take* is distributed between the two text types (fiction and non-fiction). Towards the end I will touch on how the distribution of *take* is realised across verb forms, until, finally, I conclude by presenting the results from the concrete/abstract object analysis.

#### **4. 1. 1. General overview of sense categories**

As mentioned above, this section will give an account of the results of my analysis. The obvious starting point is to see whether the ordering of sense categories by frequency is compatible with my results. The order in which the sense categories are presented should reflect which senses of the verb *take* are most frequent in use, the ones on the top of the list being the most frequent. The rightmost column in Table 4. 1. shows us the frequency of each sense in the data.

Table 4. 1. Senses distributed across the corpora

Sense	ENPC non-translations		ENPC translations		BNC		TEC		Total
	n	%	n	%	n	%	n	%	n
1. Move, remove	25	10	32	13	25	10	38	15	120
2. Bring, carry	23	9	26	10	28	11	26	10	103
3. Hold, get hold of	64	26	76	30	95	38	73	29	308
4. Require	20	8	13	5	19	8	13	5	65
5. Accept, choose	12	3	4	2	8	3	13	5	37
6. Travel	5	2	8	3	6	2	5	2	24
7. Carry out	28	11	4	2	5	2	7	3	44
8. Consider	8	3	3	1	8	3	2	1	21
9. Consume	9	4	10	4	7	3	7	3	33
10. Occupy	10	4	18	7	16	6	14	6	58
11. Participate	4	2	14	6	4	2	7	3	29
12. Experience	4	2	3	1	2	1	6	2	15
13. Suit, fit, use	1	0	0	0	0	0	0	0	1
14. Assume, acquire	9	4	9	4	1	0	9	4	28
15. Subtract	1	0	0	0	1	0	0	0	2
16. Film, shoot	1	0	5	2	3	1	4	2	13
20. Idiom	21	8	20	8	19	8	19	8	79
31. Movement/possession	0	0	0	0	0	0	5	2	5
32. Movement/bring	3	1	3	1	2	1	4	2	12
33. Carry out/idiom	2	1	2	1	1	0	1	0	6
<b>Total</b>	<b>250</b>	<b>98</b>	<b>250</b>	<b>100</b>	<b>250</b>	<b>99</b>	<b>253</b>	<b>102</b>	<b>1003</b>

Before I comment on the figures in the table, notice that the cells showing the total percentages in the corpora (in the bottom row) do not always show 100 %. This is due to the fact that the figures are rounded to the nearest whole number, which creates some inaccuracies. This will be the same for all the following tables with percentages. When we look at the figures showing the totals, we see that the first six categories follow the hierarchy of frequency, except category 3 ('hold', 'get hold of') (308), which by far exceeds the frequency of categories 1 ('remove', 'move') (120) and 2 ('bring', 'carry') (103). Another anomaly in this respect is category 7 ('carry out'), which has a higher frequency (44) than the two categories directly above it (category 5: 37; category 6: 24). If we take a look at the distribution of category 7 in the different corpora, we notice that this is due to an unexpected surge in ENPC original texts (28). Two other notable disturbances in the frequency hierarchy are that categories 10 ('occupy') (58) and 14 ('assume', 'acquire') (28) have higher frequencies than expected.

Performing a chi-square test on this table presents some methodological issues. For the chi-square test to be considered valid, each cell in the table must have a count of 5 or higher. A quick look at the above table tells us that this is not the case here. There are several cells that show frequencies lower than 5. This means, for this table, that the least frequent senses

will be collapsed into one category, and by doing this, the count in every cell of each category now meets the demands required for performing the chi-square test. The chi-square test result for Table 4. 1. is  $\chi^2 = 32.817$ ,  $df = 24$ ,  $p > 0.05$ . The distribution across sense categories is not significantly different across the corpora.

This section showed that the frequency of sense categories, which are ordered by anticipated frequencies according to dictionary sources, in large follows the hierarchy of frequency, although with a couple of exceptions, at least one of which is striking.

#### 4. 1. 2. General patterns

My hypothesis claims that the senses of the verb *take* that are generally most frequent (the sense categories at the top of the list) should have a higher occurrence rate in translated texts than in non-translated texts. This would be logical since a translator, according to Baker's (1996) proposed universal of normalisation, would tend 'to conform to patterns and practices that are typical of the target language, even to the point of exaggeration' (Baker 1996: 183). Table 4. 2. below shows us the distribution of *take* between the sense categories in non-translated texts and translated texts.

Table 4. 2. Senses distributed in non-translations and translations

Sense	Non-translations		Translations		Total
	N	%	N	%	n
1. Move, remove	50	10	70	14	120
2. Bring, carry	51	10	52	10	103
3. Hold, get hold of	159	32	149	30	308
4. Require	39	8	26	5	65
5. Accept, choose	20	4	17	3	37
6. Travel	11	2	13	3	24
7. Carry out	33	7	11	2	44
8. Consider	16	3	5	1	21
9. Consume	16	3	17	3	33
10. Occupy	26	5	32	6	58
11. Participate	8	2	21	4	29
12. Experience	6	1	9	2	15
13. Suit, fit, use	1	0	0	0	1
14. Assume, acquire	10	2	18	4	28
15. Subtract	2	0	0	0	2
16. Film, shoot	4	1	9	2	13
20. Idioms	40	8	39	8	79
31. Movement/possession	0	0	5	1	5
32. Movement/bring	5	1	7	1	12
33. Carry out/idiom	3	1	3	1	6
<b>Total</b>	<b>500</b>	<b>100</b>	<b>503</b>	<b>100</b>	<b>1003</b>

If we take a look at the sense frequencies between the corpora that contain non-translated texts (ENPC originals and the BNC) and the corpora containing translated texts (ENPC translations and the TEC), we get some conflicting results. In the case of category 1 ('move', 'remove') there is a tendency that this sense occurs more often in translated texts than in original texts. The figures show a frequency of 50 (10 % of the occurrences of *take*) in non-translated texts and 70 (14 % of the occurrences of *take*) in translated texts. The only other category in which this is a clear tendency is category 11 ('participate'), where the numbers reach 8 (2 %) in non-translated texts and 21 (4 %) in translated texts. Category 14 ('assume', 'acquire') also has a higher occurrence frequency in translated texts (4 %) than in non-translated texts (2 %), however, this is due to a rather strange deficit in the BNC as seen in table 4. 1.

My hypothesis also claims that senses of the verb *take* that are generally least frequent in use would have a higher frequency in non-translated texts than in translated texts. Categories 4 ('require') (non-translations: 39 (8 %); translations: 26 (5 %)), 7 ('carry out') (non-translations: 33 (7 %); translations: 11 (2 %)) and 8 ('consider') (non-translations: 16 (3 %); translations: 5 (1 %)) are examples of this. In the majority of cases, though, the senses have similar occurrence frequencies in both types of corpora (with a difference margin of 1 %).

When it comes to performing the chi-square test on Table 4. 2., I had to do the same thing as with the previous table, namely merge sense categories which contained cells with a lower frequency than 5 into a joined category. The figures in Table 4. 2. show a high level of significance  $\chi^2 = 34.956$ ,  $df = 15$ ,  $p < 0.01$ .

We have now seen how different sense categories behave in translated (ENPC translations and the TEC) and non-translated (ENPC non-translations and the BNC) texts. Since these two collections of texts are made up of two corpora each, it may prove worthwhile to see if there are any major differences in the distribution of the sense categories between the two corpora in each collection. First we compare the corpora containing translations, as shown in Table 4. 3.

Table 4. 3. ENPC translations and the TEC.

Sense	ENPC translations		TEC	
	N	%	n	%
1. Move, remove	32	13	38	15
2. Bring, carry	26	10	26	10
3. Hold, get hold of	76	30	73	29
4. Require	13	5	13	5
5. Accept, choose	4	2	13	5
6. Travel	8	3	5	2
7. Carry out	4	2	7	3
8. Consider	3	1	2	1
9. Consume	10	4	7	3
10. Occupy	18	7	14	6
11. Participate	14	6	7	3
12. Experience	3	1	6	2
13. Suit, fit, use	0	0	0	0
14. Assume, acquire	9	4	9	4
15. Subtract	0	0	0	0
16. Film, shoot	5	2	4	2
20. Idiom	20	8	19	8
31. Movement/possession	0	0	5	2
32. Movement/bring	3	1	4	2
33. Carry out/idiom	2	1	1	0
<b>Total</b>	<b>250</b>	<b>100</b>	<b>253</b>	<b>102</b>

When looking at the percentages in the rightmost column for each category, we see that there are not many major differences in the distribution of the senses between ENPC translations and the TEC. In fact, all the senses are of a frequency are equally represented in both corpora with a difference margin of maximum 2 %, except in sense categories 5 ('accept', 'choose') (ENPC translations: 2 %; TEC: 5 %) and 11 ('participate') (ENPC translations: 6 %; TEC: 3 %) which have a difference margin of 3 %.

According to the chi-square test, the figures in Table 4. 3. are not statistically significant:  $\chi^2 = 9.547$ ,  $df = 10$ ,  $p > 0.05$ .

The relationship between the corpora containing non-translations is shown in Table 4. 4.

Table 4. 4. ENPC non-translations and the BNC

Sense	ENPC non-translations		BNC	
	N	%	n	%
1. Move, remove	25	10	25	10
2. Bring, carry	23	9	28	11
3. Hold, get hold of	64	26	95	38
4. Require	20	8	19	8
5. Accept, choose	12	3	8	3
6. Travel	5	2	6	2
7. Carry out	28	11	5	2
8. Consider	8	3	8	3
9. Consume	9	4	7	3
10. Occupy	10	4	16	6
11. Participate	4	2	4	2
12. Experience	4	2	2	1
13. Suit, fit, use	1	0	0	0
14. Assume, acquire	9	4	1	0
15. Subtract	1	0	1	0
16. Film, shoot	1	0	3	1
20. Idiom	21	8	19	8
31. Movement/possession	0	0	0	0
32. Movement/bring	3	1	2	1
33. Carry out/idiom	2	1	1	0
<b>Total</b>	<b>250</b>	<b>98</b>	<b>250</b>	<b>99</b>

Similar to the comparison of the corpora containing translations, not many major differences in the distribution of the senses are found between ENPC non-translations and the BNC. Most of the sense categories are equally distributed between the two, again with a difference margin of 2 %. However, the exceptions found here show a greater difference margin than we found between the corpora containing translations. The sense category with the greatest variance between ENPC non-translations and the BNC is category 3 ('hold', 'get hold of'), which accounts for 26 % of the occurrences in the former and 38 % in the latter, bringing the difference to 12 %. Another sense of the verb *take* that has a rather large difference margin between the two corpora is category 7 ('carry out') (ENPC non-translations: 11 %; BNC: 2 %). Lastly, category 14 ('assume', 'acquire') has a relatively large difference margin (ENPC non-translations: 4 %; BNC: 0 %).

The chi-square test showed that the figures in Table 4. 4. are highly significant  $\chi^2 = 28.318$ ,  $df = 11$ ,  $p < 0.01$ .

To summarize, we found that my results do to some degree support my hypothesis that the senses of the verb *take* that are generally most frequent should have a higher occurrence rate in translated texts than in non-translated texts and vice versa. We found that translators tend to over-represent category 1 ('move', 'remove'). We also found instances where



translators tended to under-represent the more peripheral senses. Also, the chi-square test showed that the differences between translations and non-translations (Table 4. 2.) are very statistically significant. Furthermore, we noticed that the corpora containing translations showed very similar distribution of the senses of *take*, while the corpora containing non-translations had some notable differences between them. In other words, the language in the corpora containing translations is more homogenous than the language in the corpora containing non-translations.

#### 4. 1. 3. Text type

When performing my analysis, I catalogued the examples from my data according to text type so that it would be possible to see what effect this factor could have on my results. In addition to showing the actual numbers, I opted to show the percentage of each sense category for non-translated texts and translated texts. The percentages reflect the occurrence of each meaning in the different text types. To present the numbers by percentage was necessary since the texts of fiction made up 60 % of the corpora, while the non-fiction texts only accounted for 40 % of the corpora. This way, we have made up for the difference in size between the text types. Organized in this way, we get a rather different picture of how the meaning categories are distributed in the corpora (seen in Table 4. 5. below).

Table 4. 5. Sense categories distributed across text types (fiction/non-fiction)

Sense	Non-translations				Translations			
	Fiction		Non-fiction		Fiction		Non-fiction	
	N	%	n	%	n	%	n	%
1. Move, remove	38	12	12	6	58	17	12	7
2. Bring, carry	43	14	8	4	34	10	18	11
3. Hold, get hold of	97	32	62	32	99	30	50	30
4. Require	30	10	9	5	15	5	11	7
5. Accept, choose	14	5	6	3	11	3	6	4
6. Travel	4	1	7	4	11	3	2	1
7. Carry out	9	3	24	13	7	2	4	2
8. Consider	9	3	7	4	3	1	2	1
9. Consume	10	3	6	3	13	4	4	2
10. Occupy	12	4	14	7	12	4	20	12
11. Participate	2	1	6	3	11	3	10	6
12. Experience	5	2	1	1	8	2	1	1
13. Suit, fit, use	0	0	1	1	0	0	0	0
14. Assume, acquire	5	2	5	3	9	3	9	5
15. Subtract	2	1	0	0	0	0	0	0
16. Film, shoot	2	1	2	1	6	2	3	2
20. Idiom	20	7	20	10	26	8	13	8
31. Movement/possession	0	0	0	0	5	2	0	0
32. Movement/bring	4	1	1	1	3	1	4	2
33. Carry out/idiom	2	1	1	1	3	1	0	0
<b>Total</b>	<b>308</b>	<b>103</b>	<b>192</b>	<b>102</b>	<b>334</b>	<b>101</b>	<b>169</b>	<b>101</b>

For instance, categories 2 ('bring', 'carry') and 6 ('travel'), which in Table 4. 2. had almost the same frequencies in non-translations as in translations, now show slightly different figures. In category 2, there are higher frequencies in the non-translated fiction texts (14 %) than in the fiction translated texts (10 %), but the opposite in the non-fiction texts for the same category (non-translations: 4 %; translations: 11 %). This means that the non-fiction texts follow my hypothesis, while the fiction texts contradict it. When it comes to category 6 ('travel'), it is the other way around. Here, the fiction texts support my hypothesis (non-translations: 1 %; translations: 3 %), while the non-fiction texts goes against it (non-translations: 4 %; translations: 1 %). Another thing to make note of is that only eight of the categories behave more or less the same way regardless of text type, i.e. they have higher (or lower) frequencies in both text types. This shows that the factor of text type is a highly influential one.

We also find that some senses of the verb *take* are more frequent in use in non-fiction texts than in fiction texts. The most notable here are categories 10 ('occupy'), 11 ('participate') and 14 ('assume', 'acquire'). In other words, senses of the verb that are thought to be generally less frequent. Category 10 has an occurrence rate of 4 % (non-translations)

and 4 % (translations) in fiction texts, and 7 % (non-translations) and 12 % (translations) in the non-fiction texts. Similarly, category 11 has an occurrence rate of 1 % (non-translations) and 3 % (translations) in the section containing fiction, and 3 % (non-translations) and 6 % (translations) in the non-fiction section. Finally, category 14 has an occurrence rate of 2 % (non-translations) and 3 % (translations) in fiction texts, and 3 % (non-translations) and 5 % (translations) in the section containing non-fiction. A note of caution is required here. The sense categories mentioned in this paragraph do all have small totals of frequencies (with the exception of category 10). This means that absolute conclusions cannot be drawn from these figures, since the possibility of coincidental distribution is too high.

#### 4. 1. 4. Verb form

Another distinction I made in my analysis was that of verb form. There are five different forms of the verb *take*: present tense forms plus infinitive form (*take/s*), past tense form (*took*), past participle form (*taken*) and present participle form (*taking*). It might prove useful to make this distinction to see if there are any irregularities between the corpora in this respect. However, since there are five verb forms to keep from each other, it might not be possible to read anything out of the results from the sense categories with very low frequencies if we include all the corpora. Therefore, I will first only focus on separating non-translated texts from translated texts. Afterwards, I will have a closer look at the sense categories which are most frequent in relation to all the four corpora. Table 4. 6. shows how the different senses are distributed in relation to the verb forms in translated texts and non-translated texts.

Another consequence of the number of verb forms to take into account is that there is too much information to be included in one table. The solution here is to present the numbers in two separate tables, one for translations (Table 4. 6.) and another for non-translations (Table 4. 7.). Table 4. 6. contains figures showing the distribution of the senses of the verb *take* in translations (raw figures and percentages).

Table 4. 6. Verb form in translated texts

Sense	Translations										Total
	Take		Took		Taken		Taking		Takes		
	N	%	n	%	n	%	n	%	N	%	
1. Move, remove	20	11	29	20	10	12	2	4	9	21	70
2. Bring, Carry	28	15	10	7	2	2	7	16	5	12	52
3. Hold, get hold of	45	25	52	35	29	34	14	31	9	21	149
4. Require	8	4	6	4	6	7	0	0	6	14	26
5. Accept, choose	11	6	1	1	4	5	1	2	0	0	17
6. Travel	6	3	1	1	3	4	2	4	1	2	13
7. Carry out	7	4	3	2	0	0	1	2	0	0	11
8. Consider	2	1	0	0	3	4	0	0	0	0	5
9. Consume	5	3	7	5	0	0	4	9	1	2	17
10. Occupy	11	6	10	7	9	11	2	4	0	0	32
11. Participate	13	7	4	3	1	1	2	4	1	2	21
12. Experience	1	1	6	4	1	1	0	0	1	2	9
13. Suit, fit, use	0	0	0	0	0	0	0	0	0	0	0
14. Assume, acquire	4	2	7	5	1	1	5	11	1	2	18
15. Subtract	0	0	0	0	0	0	0	0	0	0	0
16. Film, shoot	4	2	0	0	2	2	2	4	1	2	9
20. Idioms	12	7	9	6	12	14	2	4	4	9	39
31. Movement/possession	1	1	0	0	0	0	0	0	4	9	5
32. Movement/bring	4	2	1	1	2	2	0	0	0	0	7
33. Carry out/idiom	1	1	1	1	0	0	1	2	0	0	3
<b>Total</b>	<b>183</b>	<b>101</b>	<b>147</b>	<b>102</b>	<b>85</b>	<b>100</b>	<b>45</b>	<b>97</b>	<b>43</b>	<b>98</b>	<b>503</b>

These figures are easily comparable with the figures showing the distribution of the senses of the verb *take* in non-translations, which are given in the table below.

Table 4. 7. Verb form in non-translated texts

Sense	Non-translations										Total
	Take		Took		Taken		Taking		Takes		
	N	%	n	%	n	%	n	%	n	%	
1. Move, remove	16	9	14	10	11	13	8	12	1	4	50
2. Bring, Carry	18	10	17	12	8	9	6	9	2	7	51
3. Hold, get hold of	54	30	52	36	21	25	23	35	9	32	159
4. Require	8	4	23	16	3	4	2	3	3	11	39
5. Accept, choose	8	4	4	3	4	5	4	6	0	0	20
6. Travel	9	5	1	1	0	0	0	0	1	4	11
7. Carry out	16	9	4	3	8	9	4	6	1	4	33
8. Consider	6	3	2	1	5	6	3	5	0	0	16
9. Consume	6	3	5	3	2	2	3	5	0	0	16
10. Occupy	9	5	5	3	7	8	2	3	3	11	26
11. Participate	1	1	1	1	3	4	3	5	0	0	8
12. Experience	1	1	3	2	0	0	1	2	1	4	6
13. Suit, fit, use	0	0	0	0	0	0	1	2	0	0	1
14. Assume, acquire	3	2	0	0	0	0	1	2	6	21	10
15. Subtract	2	1	0	0	0	0	0	0	0	0	2
16. Film, shoot	1	1	1	1	2	2	0	0	0	0	4
20. Idioms	15	8	10	7	10	12	4	6	1	4	40
31. Movement/possession	0	0	0	0	0	0	0	0	0	0	0
32. Movement/bring	3	2	1	1	0	0	1	2	0	0	5
33. Carry out/idiom	2	1	0	0	1	1	0	0	0	0	3
<b>Total</b>	<b>178</b>	<b>99</b>	<b>143</b>	<b>100</b>	<b>85</b>	<b>100</b>	<b>66</b>	<b>103</b>	<b>28</b>	<b>102</b>	<b>500</b>

The first thing we notice in the tables above is that there are over twice as many occurrences of *took* in translated texts than in non-translated texts when it comes to category 1 ('move', 'remove'). When we convert the number into percentages, we find that category 1 accounts for 20 % of the occurrences of *took* in the translated texts for this category, and only 10 % of *took* in the non-translated texts. Also the verb form *takes* follows this patterns when it comes to category 1 (translations: 21 %; non-translations: 4 %). *Taking*, on the other hand, goes the other way with only 4 % in the translated texts and 12 % in the non-translated texts.

In category 2 ('bring', 'carry'), *take* (translations: 15 %; non-translations: 10 %) has a higher frequency in translations, while *took* (translations: 7 %; non-translations: 12 %) and *taken* (translations: 2 %; non-translations: 9 %) have higher frequencies in non-translated texts. Also category 3 ('hold', 'get hold of') shows some differences in distribution across the verb forms between translations and non-translations, especially when it comes to *taken* (translations: 34 %; non-translations: 25) and *takes* (translations: 32 %; non-translations: 21 %).

One of the major puzzlements in this table, however, is found in category 4 ('require'), namely in the case of *took*. In the translated texts, *took* accounts for 4 % of the total, while in

the non-translated texts *took* accounts for 16 % of the total. In other words, category 4 accounts for four times as many of the total number of occurrences of *took* in non-translations than in translations.

As mentioned above, we will now have a closer look at the categories with highest frequencies, namely categories 1 ('move', 'remove'), 2 ('bring', 'carry'), 3 ('hold', 'get hold of'), 4 ('require') and 10 ('occupy'). In Table 4. 8. below we can see the results from category 1.

Table 4. 8. Verb form: 'move', 'remove'

Verb form	ENPC non-translations		ENPC translations		BNC		TEC		Total
	n	%	n	%	n	%	n	%	
<b>Take</b>	8	32	6	19	8	32	14	37	36
<b>Took</b>	6	24	13	41	8	32	16	42	43
<b>Taken</b>	8	32	5	16	3	12	5	13	21
<b>Taking</b>	3	12	1	3	5	20	1	3	10
<b>Takes</b>	0	0	7	22	1	4	2	5	10
<b>Total</b>	<b>25</b>	<b>100</b>	<b>32</b>	<b>101</b>	<b>25</b>	<b>100</b>	<b>38</b>	<b>100</b>	<b>120</b>

The biggest (and clearest) difference between the non-translations corpora and translations corpora we find when dealing with past tense *took*. There is a higher frequency of *took* in ENPC translations (41 % of the occurrences of *took* within this sense category) and TEC (42 %) than in the corpora containing non-translated texts (ENPC non-translations: 24 %; BNC: 32 %). Also, *takes* is more represented in translations (ENPC translations: 22 %; TEC: 5 %) than in non-translations (ENPC non-translations: 0 %; BNC: 4 %). *Taking*, on the other hand, occurs more often in non-translated texts (ENPC non-translations: 12 %; BNC: 20 %) than in translated texts (ENPC translations: 3 %; TEC: 3 %). When it comes to *take*, we find an even distribution across the corpora containing non-translations (32 %), while there is great variation between the corpora containing translations (ENPC non-translations: 19 %; TEC: 37 %).

If the chi-square test is to be applied on Table 4. 8. the figures in the verb forms *taken*, *taking* and *takes* need to be collapsed within each corpus in order to make the test valid (each cell must have a count of 5 or higher). The test shows that the figures above are not statistically significant  $\chi^2 = 6.747$ ,  $df = 6$ ,  $p > 0,5$ .

When it comes to category 2 ('bring', 'carry') (Table 4. 9.), the clearest results are found in the first three verb forms.

Table 4. 9. Verb form: ‘bring’, ‘carry’

Verb form	ENPC non-translations		ENPC translations		BNC		TEC		Total n
	n	%	n	%	n	%	n	%	
<b>Take</b>	7	30	12	46	11	39	16	62	46
<b>Took</b>	9	39	6	23	8	29	4	15	27
<b>Taken</b>	3	13	1	4	5	18	1	4	10
<b>Taking</b>	3	13	3	12	3	11	4	15	13
<b>Takes</b>	1	4	4	15	1	4	1	4	7
<b>Total</b>	<b>23</b>	<b>99</b>	<b>26</b>	<b>98</b>	<b>28</b>	<b>101</b>	<b>26</b>	<b>100</b>	<b>103</b>

It seems that *take* has a higher frequency in translated texts (ENPC translations: 46 %; TEC: 62 %) than in non-translated texts (ENPC non-translations: 30 %; BNC: 39 %). On the other hand, *took* occurs more often in non-translated texts ENPC non-translations: 39 %; BNC: 29 %) than in translated texts (ENPC translations: 23 %; TEC: 15 %). *Taken* also follows this pattern with ENPC non-translations (13 %) and BNC (18 %) dominating in relation to ENPC translations and TEC (4 % each).

The same precautions need to be met as with regards to Table 4. 8. if we are to perform the chi-square test on Table 4. 9. Here, however it is necessary to include the verb form *took* in the collapsed category, since the count in the TEC only reaches 4. The test showed that the figures are not statistically significant  $\chi^2 = 5.231$ ,  $df = 3$ ,  $p > 0.5$ .

The distribution of the different verb forms of *take* in sense category 3 (‘hold’, ‘get hold of’) is outlined in Table 4. 10.

Table 4. 10. Verb form: ‘hold’, ‘get hold of’

Verb form	ENPC non-translations		ENPC translations		BNC		TEC		Total n
	n	%	n	%	n	%	n	%	
<b>Take</b>	20	31	25	33	34	36	20	27	99
<b>Took</b>	22	34	24	32	30	32	28	38	104
<b>Taken</b>	8	13	17	22	13	14	12	16	50
<b>Taking</b>	10	16	5	7	13	14	9	12	37
<b>Takes</b>	4	6	5	7	5	5	4	5	18
<b>Total</b>	<b>64</b>	<b>100</b>	<b>76</b>	<b>101</b>	<b>95</b>	<b>101</b>	<b>73</b>	<b>98</b>	<b>308</b>

The most significant result in this table is that *taking* seems to be more represented in corpora containing non-translated texts (ENPC non-translations: 16 %; BNC: 14 %) than corpora containing translated texts (ENPC translations: 7 %; TEC: 12 %). Actually, there are twice as many occurrences in ENPC originals (10) as in ENPC translations (5). There are also more occurrences in BNC (13) than in TEC (9). In the other verb forms, the tendency is that the greatest variations are found between the corpora containing translations. While the verb

forms are practically equal in distribution in the non-translated corpora, the corpora containing translations show very different distributions in every verb form. There is a minimum of 5 % difference between these corpora in the first four verb forms, going either way. In *take* the distribution is higher in ENPC translations (33 %) than in the TEC (27 %) and this is also the case with *taken* (ENPC translations: 22 %; TEC: 16 %). When it comes to *took*, on the other hand, the frequency is higher in TEC (38 %) than in ENPC translations (32 %). *Taking* also follows this pattern (ENPC translations: 7 %; TEC: 12 %). One final thing to make note of is the massive difference between ENPC non-translations (13 %) and ENPC translations (22) when it comes to the distribution of *taken*, and also the difference in frequency of *take* between BNC (36 %) and TEC (27 %).

In order to successfully performing the chi-square test on Table 4. 10., I merged the last two verb forms (*taking* and *takes*) in order to bring the count higher than 5 in every cell. The test showed that the figures are not statistically significant  $\chi^2 = 5.914$ ,  $df = 9.714$ ,  $p > 0.5$ .

Table 4. 11. below shows the distribution of the various verb forms of *take* with the meaning of ‘require’. Since the frequency number adhering to this category is relatively small, one might argue that the results we get in a table like this are invalid. However, I still believe that two significant (and valid) points can be made from this table.

Table 4. 11. Verb form: ‘require’

Verb form	ENPC non-translations		ENPC translations		BNC		TEC		Total n
	n	%	n	%	n	%	n	%	
<b>Take</b>	4	20	5	38	4	21	3	23	16
<b>Took</b>	13	65	4	31	10	53	2	15	29
<b>Taken</b>	0	0	2	15	3	16	4	31	9
<b>Taking</b>	1	5	0	0	1	5	0	0	2
<b>Takes</b>	2	10	2	15	1	5	4	31	9
<b>Total</b>	<b>20</b>	<b>100</b>	<b>13</b>	<b>99</b>	<b>19</b>	<b>100</b>	<b>13</b>	<b>100</b>	<b>65</b>

The first thing is that the past tense *took* has a higher frequency in the non-translated texts (ENPC non-translations: 65 %; BNC: 53 %) than in the translated texts (ENPC translations: 31 %; TEC: 15 %). A second thing to notice is the differences between ENPC non-translations and ENPC translations, which especially is noticeable in the first two verb forms. In *take*, the frequency is much higher in ENPC translations (38 %) than in ENPC non-translations (20 %), while in *took* the frequency is over twice as high in ENPC non-translations (65 %) than in ENPC translations (31 %). One final thing relates to how much more often *took* is used with this meaning than the other tenses, as we see in the rightmost



column. Almost 50 % of all occurrences of *take* with the meaning of ‘require’ are represented by past tense *took*. The chi-square test is not fit to give reliable results for this table since there are too many cells with a count less than 5.

The last table dealing with verb form is Table 4. 7., which shows how *take* meaning ‘occupy’ is distributed between the corpora.

Table 4. 12. Verb form: ‘occupy’

Verb form	ENPC non-translations		ENPC translations		BNC		TEC		Total n
	n	%	n	%	n	%	n	%	
<b>Take</b>	5	50	4	22	4	25	7	50	25
<b>Took</b>	2	20	7	39	3	19	3	21	15
<b>Taken</b>	2	20	5	28	5	31	4	29	16
<b>Taking</b>	1	10	2	11	1	6	0	0	4
<b>Takes</b>	0	0	0	0	3	19	0	0	3
<b>Total</b>	<b>10</b>	<b>100</b>	<b>18</b>	<b>100</b>	<b>16</b>	<b>100</b>	<b>14</b>	<b>100</b>	<b>58</b>

It may be worthwhile to comment on the relatively high frequency of past tense *took* in ENPC translated texts (39 %) in relation to the other corpora. However, this sense category suffers under the same disadvantage as the former; the total frequency is too low to yield any significant and valid results. The chi-square test is not fit to give reliable results for this table since there are too many cells with a count less than 5.

This section has been concerned with how the figures are distributed across verb forms. Due to the number of variables for the figures to be divided into, some sense categories were not able to provide reliable results, but the ones that did showed some interesting outcomes. Most notable was the figures in category 1 (‘move’, ‘remove’) and category 2 (‘bring’, ‘carry’), which showed many differences when comparing translations with non-translations. As mentioned, very few sense categories had total frequencies high enough to be suitable for closer examination. The ones that did were put in separate tables and given extra attention. This made it possible to detect further differences in the distribution of the five verb forms in the four corpora and to perform tests for statistical significance in three of the sense categories. However, we found that the figures were not statistically significant.

#### 4. 1. 5. Concrete/abstract object

In my analysis I separated examples with concrete objects from examples with abstract objects. This is a factor that potentially could shed some light on the distribution of *take* across translations and non-translations. If the figures should turn out to be significant or

follow a certain pattern, this factor should indeed be included in future investigations similar to this one. Of course, not all the sense categories were suited for such a distinction because some would only take abstract objects, some only concrete objects, while some would be intransitive and take no object at all. But the ones that were yielded some interesting results. To get a clearer picture of how the distribution between the corpora takes shape, I have set up the percentages for each sense in each corpus along with the raw data.

Table 4. 13. Concrete/abstract object

Sense	ENPC non-translations				ENPC translations				BNC				TEC			
	Concrete		Abstract		Concrete		Abstract		Concrete		Abstract		Concrete		Abstract	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
1. Move, remove	20	80	5	20	30	94	2	6	22	88	3	12	33	87	5	13
2. Bring, carry	20	87	3	13	21	81	5	19	27	96	1	4	22	85	4	15
3. Hold, get hold of	34	53	30	47	25	33	51	67	29	31	66	69	20	27	53	73
5. Accept, choose	4	33	8	67	0	0	4	100	3	38	5	62	3	23	10	77
6. Travel	1	20	3	80	7	87	1	13	6	100	0	0	5	100	0	0
10. Occupy	5	50	5	50	3	17	15	83	3	19	13	81	8	57	6	43
31. Movement/possession	0	0	0	0	0	0	0	0	0	0	0	0	5	100	0	0
32. Movement/bring	3	100	0	0	3	100	0	0	1	50	1	50	4	100	0	0

We can see from Table 4. 13. that senses that involve *movement* (categories 1, 2, 31 and 32) are inclined to take concrete objects, while senses that involve (*change of*) *possession* (categories 3 and 5) are relatively more inclined to take abstract objects. To give an example, consider category 1 ('move', 'remove') which clearly is a sense denoting movement. ENPC non-translations (concrete: 80 %; abstract: 20 %), the BNC (concrete: 88 %; abstract: 12 %), ENPC translations (concrete: 94 %; abstract: 6 %) and the TEC (concrete: 87 %; abstract: 13 %) all favour concrete objects over abstract objects when it comes to category 1. Category 3 ('hold', 'get hold of') denotes a sense of possession and, as we can see in the BNC (concrete: 31 %; abstract 69 %), ENPC translations (concrete: 19 %; abstract: 81 %) and the TEC (concrete: 13 %; abstract 87 %), it is considerably more inclined to take abstract objects. Also in ENPC non-translations (concrete: 53 %; abstract: 47 %) this sense show a stronger inclination to take abstract objects than the senses denoting movement, although there is a majority of occurrences with concrete objects.

When it comes to differences between non-translated texts and translated texts, it seems that translations allow *take* to take more abstract objects than non-translations in the cases of category 2 ('bring', 'carry') (ENPC non-translations: 13 %; BNC: 4 %; ENPC

translations: 19 %; TEC: 15 %), category 3 ('hold', 'get hold of') (ENPC non-translations: 47 %; BNC: 69 %; ENPC translations: 67 %; TEC: 73 %) and category 5 ('accept', 'choose') (ENPC non-translations: 64 %; BNC: 62 %; ENPC translations: 100 %; TEC: 77 %). In category 1 ('move', 'remove') (ENPC non-translations: 20 %; BNC: 12 %; ENPC translations: 6 %; TEC: 13 %), on the other hand, the opposite happens. Granted, the TEC (13 %) has a higher frequency of abstract objects than the BNC (12 %). But if we compare the corpora containing non-translations against the corpora containing translations we find that *take* meaning *move* or *remove* allows abstract objects in 16 % of the occurrences in non-translations and only in 10 % of the occurrences in translations (seen in Table 4. 14. below).

Lastly, we find that the verb *take* is as likely to take a concrete object in a translated text as in a non-translated text. In Table 4. 14. below I have put together the numbers from ENPC non-translations and the BNC, as well as the numbers from ENPC translations and the TEC.

Table 4. 14. Concrete/abstract object across non-translations and translations

Sense	Non-translations				Translations				Total
	Concrete		Abstract		Concrete		Abstract		
	N	%	n	%	n	%	N	%	
1. Move, remove	42	84	8	16	63	90	7	10	<b>120</b>
2. Bring, carry	47	92	4	8	43	83	9	17	<b>103</b>
3. Hold, get hold of	63	40	96	60	45	30	104	70	<b>308</b>
5. Accept, choose	7	35	13	65	3	18	14	82	<b>37</b>
6. Travel	7	70	3	30	12	92	1	8	<b>23</b>
10. Occupy	8	31	18	69	11	34	21	66	<b>58</b>
31. Movement/possession	0	0	0	0	5	100	0	0	<b>5</b>
32. Movement/bring	4	80	1	20	7	100	0	0	<b>12</b>
<b>Total</b>	<b>178</b>	<b>55</b>	<b>143</b>	<b>45</b>	<b>189</b>	<b>55</b>	<b>156</b>	<b>45</b>	<b>666</b>

Direct your attention to the bottom row showing the total amount of occurrences of *take* with concrete or abstract objects in non-translations and translations. Here, we find that in both non-translations and translations 55 % of the occurrences of *take* take concrete objects. This is somewhat surprising considering the variation between the corpora showed above.

To summarise this section, we found that the sense categories denoting *movement* were inclined to take concrete objects while the sense categories denoting (*change of*) *possession* were relatively more inclined to take abstract objects. We also found that there is considerable variation within senses and across corpora.

## 4. 2. Discussion

In this section I will discuss the implications of my results. I will consider the tables one by one, or sometimes compare tables, in order to draw some conclusions. However, before I plunge into the discussion, it could be useful to place this paper within the ever expanding field of Translation Studies. Taking Holmes' map of this field (which is useful despite its controversy advocated by Pym (1998)) as a starting point, this paper clearly places itself under the 'pure' branch which consists of a descriptive branch and a theoretical branch. The aim is to describe the nature of translated language compared to original (non-translated) language in order to investigate the validity of existing theories. This might also have indirect implications for the applied branch of Translation Studies (translation training in particular), since some of the findings here might make people more aware of how translations differ from non-translations. However, it is not my goal to make judgments on how translations should look like, only to describe how translations *actually* look like. Looking at the subcategories in Holmes' map, we see that a descriptive study can be focused on different aspects of translation. The present study is mainly product-oriented as it is focused on the finished products of translators. On the other hand, it is not unthinkable that the results from these products, in comparison to non-translations, might tell us something about how a translator's mind works, which would lead us into the area of process investigation. This paper is mainly occupied with exploring the claim made by Toury (1995) and Baker (1996) that normalisation is a feature that consistently appears in translated language.

### 4. 2. 1. General overview of sense categories

One part of my hypothesis was that the frequencies of the sense categories in the corpora should follow the list, from top to bottom, that I had set up on the basis of dictionary sources. This is based on the assumption made by many scholars that prototypicality is a reflection of frequency of actual language use. The most used concept of a variant is thought to be the prototype. In this respect, my list of sense categories is meant to reflect the degree of prototypicality inherent in the senses of the verb *take*. As seen in Table 4. 1., this is not the case since some sense categories situated down on the list are found to have a higher frequency than some categories towards the top of the list. The main upset of the hierarchy was seen in category 3 ('hold', 'get hold of') where the frequency was substantially higher than the first two categories. However, this can be explained by the fact that it is a broad category, encompassing senses that were listed separately in the dictionaries, but that I found

logical to unite into one single category. These unified senses include *get hold of*, *take into one's possession*, *take by force* and *obtain by winning*, all of which are listed separately in Wordnet alone. In addition come all the abstract and metaphorical senses like *take advantage*, *take credit*, etc., which also often were listed separately. Table 4. 13. shows us that within this category, *take* has been very susceptible to taking an abstract object (over 50 % of the instances). It is evident that these abstract and metaphorical senses played some part in making the frequency count unexpectedly high. The fact that this category was made to contain this wide variety of sub-senses may account for the surge in frequency it obtained.

Category 7 ('carry out') also fails to adhere to the expected frequency list, but this has nothing to do with the makeup of the category like with category 3. Rather, this anomaly can be explained by having a look at the frequency in ENPC non-translation in relation to the other corpora. While this sense category accounts for between 2 % and 3 % of the occurrences of *take* in the other corpora, it accounts for as much as 11 % in ENPC non-translations. If we assume that the raw frequency in ENPC non-translations was 6 (which is the mean frequency in the other corpora), the total count of this category would reach 24 and it would fit nicely into the hierarchy. So why does *take* behave so differently in ENPC non-translations than in the rest of the corpora in this respect? When coding the data, I noticed that in the non-fiction part of this corpus there was an excessive number of examples of *take* attaching to the nouns *measure*, *step* and *decision*, all stemming from two documents, written by the European Union and the European Council. These are clearly official documents and they describe a new directive and an agreement between parties, and therefore these types of phrases are to be expected to dominate them. My observations are confirmed when we have a look at Table 4. 5. showing how the sense categories are distributed between text types in non-translations and translations. We find that the frequency is higher in the non-fiction texts of the non-translations (24) for this category than in all the other sub-divisions put together (20).

Another significant disruption of the hierarchy of sense categories is category 10 ('occupy'). It reaches an overall frequency of 58, which would move the category 5 places up the presupposed list. It is difficult to establish an explanation to why this sense category reached such a high frequency. However, if we take a look at the frequencies in the four corpora separately, we see that the greatest variations are found between the two ENPC corpora. These corpora are relatively small, which might lead to some skewed results.

There are two categories that show a significant deficit in frequencies, even in relation to their modest positions on the presupposed list. These are categories 13 ('suit', 'fit', 'use') and 15 ('subtract'), which only occur once and twice, respectively, in my data consisting of

1,003 examples. The reason why these categories failed to reach high frequencies might be that they overlapped other categories. Category 13 might be said to share some qualities with categories 4 ('require') and 10 ('occupy'). It is conceivable that this might be part of the reason why category 10 reached such a high frequency (as discussed above). It may have been attributed qualities that the dictionaries meant to reserve for category 13. When it comes to category 15 ('subtract') it is not unexpected that it should receive a low frequency given its position on the list. However, it may be the case that this category overlapped category 1 ('move', 'remove'), since the concept of something being removed is closely related with the concept of subtracting.

Given the discussion on whether prototypicality is reflected in corpus frequency or not, it is not surprising to find that my presupposed list which is made up of listings in dictionaries does not entirely coincide with the frequencies found in the corpora I have used. One reason for this is that the dictionaries I used to make the presupposed list used a mix of corpus frequency, concepts of centrality and common sense to decide on the order in which they presented the word senses. Due to the different criteria used by the dictionaries it was expected that the results from my data would differ somewhat from my presupposed list. This is because there is no one-to-one relationship between concepts of centrality (meant here as concepts that first come to mind) and corpus frequency, at least according to research by Gilquin (2006) and Nordquist (2004). They find that the variants that first come to mind in elicitation tests do not necessarily correspond to the most frequent variants found in electronic corpora. My results give support to this view, since my presupposed list constructed on the basis of both corpora frequency and the notion of 'central concepts' (first comes to mind) does not entirely coincide with the frequencies found in the corpora I have used.

#### 4. 2. 2. General patterns

According to Viberg (2002a, 2002b), second language learners have a strong preference for using basic verbs in their language rather than more narrow verbs denoting more specified concepts. This may be transferred to the use of a particular basic verb, and a claim that second language learners have a strong preference for using the verb in its more prototypical senses, while the more peripheral senses of the verb are under-represented. This again may be transferred to patterns of verb use by translators, which leads to my hypothesis that translators tend to over-represent the prototypical senses of the verb and to under-represent the fringe senses. This hypothesis, if supported by my data, would strengthen the claim made by Toury

(1995) and Baker (1993, 1996) that translators have a tendency to normalise their language, over-representing common aspects of the target language even to the point of exaggeration.

As it turns out, my data does seem to lend some support to the claim that translators tend to over-represent typical features of the target language. This will be manifested by higher frequencies of the prototypical senses of the verb *take* in translated texts than in non-translated texts. Here, it might prove useful to establish which sense categories can be seen as prototypical and which sense categories can be seen as fringe categories. My suggestion is that only the first three categories on my list can be defined as more or less prototypical. What I base this suggestion on is the actual frequencies with which the categories occur in my data. The first three categories have frequencies that are almost twice as high as the next category on the list (category 4 ('require')). Category 3 ('hold', 'get hold of') actually reaches a frequency that is more than four times as high as category 4 ('require'). It should be noted that this division is just a rough one which is sufficient to serve the purpose of this paper. The consequence of this division between prototypical categories and peripheral categories is that over-representation in translations is expected to be found in the three first categories, while under-representation is expected to occur in the remaining categories.

In Table 4. 2., which shows the distribution of the various senses of *take* in non-translations and translations, we see that there are not many substantial differences between the two. However, when it comes to the sense category which in the presupposed list is found to be the most prototypical sense of the verb *take*, namely category 1 ('move', 'remove'), we find that the frequency with which it occurs in translations (14 %) is significantly higher than in non-translations (10 %). It would seem that translators tend to over-represent this sense in their works.

When we turn to the sense category that is actually most frequently used in both non-translations and translations, however, we see a different tendency. According to our claim of normalisation, we would expect that this sense (category 3 ('hold', 'get hold of')) also is over-represented in translations. However, this is not the case. On the contrary, it actually occurs more often in non-translations (32 %) than in translations (30 %).

In the other end of the scale we would expect to find higher frequencies in non-translations than in translations since my hypothesis claims that translators, as well as over-representing prototypical senses, tend to under-represent peripheral senses of a verb. When examining my data, however, we find that this categorically is not the case. From sense category 9 ('consume') and downwards, the frequencies in translations are either equal to or higher than the frequencies in non-translations. Also, the overall differences between the

distributions of *take* in non-translations and translations are not very significant. This means that any asymmetries between non-translated language and translated language in relation to the verb *take* must lie elsewhere than merely in the distribution patterns between the two. Here it becomes useful to look at inherent factors such as text types, verb forms and whether the verb takes a concrete or an abstract object. But first, one thing needs to be cleared up. Since we are dealing with four different corpora, two containing non-translations and two containing translations, it may be useful to examine whether there are any differences between the corpora in each of the two modes. A comparison of ENPC translations and the TEC and of ENPC non-translations and the BNC seems imminent.

As is reflected by the chi-square test done on Table 4. 3. ( $p > 0.05$ ), there are few, if any, significant differences in the distribution of *take* between ENPC translations and the TEC. This means that the figures in Table 4. 2. showing the frequencies of the senses of *take* in non-translations and translations, are not disturbed by great variation of frequencies in the two corpora containing translations. The same can hardly be said by the two corpora containing non-translations. As the chi-square test done on Table 4. 4. reflects ( $p < 0.01$ ), great variations can be found in the distribution of *take* between ENPC non-translations and the BNC. The fact that practically no variations were found between the two corpora containing translations, while great variations were found between the corpora containing non-translations might be taken to mean that translated language is more homogenous than non-translated language. The tendency for translators to conform to typical features of the target language is so consistent that the variation that exists in non-translated language is less likely to occur in translated language.

The most significant differences between ENPC non-translations and the BNC can be found in categories 3 ('hold', 'get hold of') and 7 ('carry out'). Category 3 constitutes 38 % of the examples in BNC, while it only represents 26 % of the examples in ENPC non-translations. If we assume that the results in the latter are most representative of the actual distribution of this sense of *take* in non-translations, we find that translators *over*-represent this sense in translations (30 %). If, on the other hand, we assume that the results in the former are most representative of the actual distribution of this sense, we would find that translators dramatically *under*-represent this sense in translations. But which version is most likely to be true?

Perhaps the answer may be found if we take into account the other major difference between the two corpora containing non-translations, namely the difference found in category 7 ('carry out'). As earlier pointed out, this category was skewed in ENPC non-translations by



one author (or group of authors) writing on the behalf of the European Union and the European Council, which used *take* in this sense excessively due to the nature of the article. Since the frequency of category 7 in the BNC (2 %) is quite similar to the frequencies in the corpora containing translations (ENPC translations: 2 %; TEC: 3 %), we could assume that the surge in frequency in ENPC non-translations (11 %) is abnormal, and furthermore may have caused the relatively low frequency of category 3 in this corpus. In other words, since category 7 occupies a larger part of this corpus than in the other corpora, one (or many) of the other categories will inevitably suffer in terms of lower frequencies. It would be tempting to ascribe the relatively low frequency in category 3 in this corpus to the unexpected surge of frequency in category 7. If we go into the numbers, however, this theory is only partly supported. The excessive frequency of category 7 ('carry out') in ENPC non-translations was found to be due to a surge in the section containing non-fiction. It would then be expected that the relatively low frequency of category 3 ('hold', 'get hold of') in ENPC non-translations should be due to a deficit in the section containing non-fiction in relation to the frequencies found in the BNC. However, despite this being the case (ENPC non-translations: 26 %; BNC: 39 %), there is also a clear deficit in the section containing fiction (ENPC non-translations: 26 %; BNC: 38 %). Therefore, we cannot solely 'blame' the surge of frequency of category 7 in the section containing non-fiction in ENPC non-translations for the low frequency found in category 3 in the same corpus. There is thus a possibility that this low frequency did not appear by chance and that translators over-represent this sense in their translations. However, nothing conclusive can be said on the basis of these results.

#### 4. 2. 3. Text type

When doing an analysis such as this one, comparing translated language with non-translated language, it is useful to explore whether other factors may have an influence on one's results. This is why I decided to code my data according to text type, along with other possibly useful factors which will be dealt with in the next sections of this chapter. In fact, it is shown in Table 4. 5. that whether the text is fiction or non-fiction has an impact on the distribution of the verb *take*. The figures for category 1 ('move', 'remove') reveal that it is mostly in fiction texts that translators tend to over-represent *take* in this sense. This means that the difference is even greater than was shown in Table 4. 2. In non-translations, category 1 amounts to 12 % of the instances in fiction texts, while in translations it amounts to 17 % of the instances in fiction texts. In non-fiction texts, however, the frequency of use of category 1 is practically even across non-translations and translations.

As mentioned above in Section 4. 1. 3., the impact that text types have on the distribution of *take* is especially evident in category 2 ('bring', 'carry'). In the table showing the distribution of *take* according to corpus type and sense category (Table 4. 2), we saw that the frequencies in this category were the same in both non-translations and translations. When we include text type in the equation, on the other hand, we find that this sense is under-represented in fiction translations, while it is over-represented in non-fiction translations. Because non-fiction texts are often linked to more formal language, it can be speculated that perhaps translators deem this form of expressing conveyance too informal to be used in this context. It could be useful in future studies to explore if other expression forms denoting this meaning (e.g. *bring*, *carry*, *convey*, *direct*, etc) are used excessively in non-fiction texts by translators.

In 3. 1. 4. I also mentioned category 6 ('travel') as one of the categories prone to the influence of text type on my results. But in this case, the influence is of the opposite nature, meaning that translators tend to over-represent *take* meaning *travel* in fiction texts and under-represent it in non-fiction texts. Clearly, translators find this word form inappropriate to express concepts relating to the act of travelling in formal language, something which does not coincide with the attitudes belonging to producers of non-translated texts. I presume that translators find the word form *travel* more appropriate to use in formal situations, but this again is something that would make for interesting future research. However, translators make up for this prudence by over-using it in less formal language found in texts of fiction.

One final sense category which is illuminated when dividing the results into fiction and non-fiction is category 10 ('occupy'). In Table 4. 2., this category was found to have a slightly higher frequency in translations (6 %) than in non-translations (5 %). This finding is in itself not very significant. When we introduce the new factor of text type, however, we find that *take* meaning *occupy* is evenly distributed between non-translations and translations when it comes to fiction (4 % in both). In non-fiction texts the story is rather different. There is a significantly higher frequency in translations (12 %) than in non-translations (7 %). This finding is not coherent with my hypothesis which predicts a lower frequency in translations than in non-translations in peripheral senses. Listed as number 10, this category must be said to be one of the peripheral senses of *take*. Therefore, we should seek to find alternative explanations. One possible explanation is that there might be some kind of interference from the source texts. A similar explanation might be that there is interference from the existing norms in the source language(s). However, if we look at the total number of occurrences of this sense in the corpora (Table 4. 1.) we see that this category is actually the fifth largest

category in terms of frequency. While still considered to be a peripheral sense of *take*, this sense is perhaps frequent enough to be said to be typical of the target language. The over-representation of this sense in non-fictional texts might, seen in this light, be said to be an example of normalisation in translation.

Finally, the reason why there are more occurrences of *take* meaning *travel* in translated non-fiction than in non-translated non-fiction might be an issue of normalisation. As we can see from the figures relating to this category, it is more common to use this word when talking about travelling in non-fiction texts than in fiction texts. My claim here will be that the translator picks up on this, presumably subconsciously, and tries to do his best (also subconsciously, perhaps) to adhere to the norms of the target language, which leads to an over-representation of the common English feature.

#### 4. 2. 4. Verb form

My data contains all five forms of the verb *take*, and each of them is represented in my four data samples at the same rates as in the four full corpora (see Section 3. 1. 4). This makes for a great opportunity to explore whether there are any differences in the distribution of the verb forms between non-translations and translations. In section 4. 1. 4. there were two tables; one that shows the distribution of the senses of *take* in each verb form in non-translations and one that shows the distribution of the senses of *take* in each verb form in translations. The tables show the percentages of the total number of occurrences of each verb form in each sense category. From these tables we were able to detect several differences between non-translations and translations.

One of these differences was that category 1 ('move', 'remove') had twice as many occurrences of past tense *took* in translations (20 %) as in non-translations (10 %). An even greater difference was also found in the same category when it came to *takes* (translations: 21 %; non-translations: 4 %). These findings suggest that writers of non-translations perhaps turn to other solutions than *take* when the task is to denote *movement* in past tense or present participle forms. Therefore we might expect to find the same patterns in category 2 ('bring', 'carry') which also is a *movement* category. However, here we find that the frequencies of *took* are quite similar in translations (10 %) and non-translations (12 %), and even a bit higher in non-translations. *Takes*, on the other hand, follows the same pattern in this category as in the previous with a frequency of 12 % in translations and only 7 % in non-translations. However, in both of these senses, the frequencies of *takes* are too low to make any conclusions about their distribution across the corpora.

Another significant difference in category 1 between translations and non-translations is the case of *taking*. A total of 12 % of all the occurrences of this verb form in non-translations denote this sense, while only 4 % denote this sense in translations. Again, an explanation for this difference may be found in interference from the source texts. Perhaps English *taking* is a verb form that does not correspond well with the verb forms found in other languages in this sense. This will result in source texts with alternative verb forms in it, and thus, the translator does not have the textual triggers that would normally result in the verb form *taking* in the target text.

On another note, my hypothesis claims that translators normalise their language in translations; they adapt their language to the most common features of the target language, even to the point of exaggeration. In this context, we would then assume that the verb form that is most in use in non-translations in each sense category is used to the same degree (if not more) in translations. In Tables 4. 8. – 4. 12. we saw the distribution of the verb forms in each of the four corpora for the five most frequent sense categories. In the first table, showing the figures for category 1 (‘move’, ‘remove’), the verb form that is most in use overall in the four corpora is *took*. The corpora that have the highest frequencies here are the ones containing translations, together amounting to 67 % of the total number. However, the verb form that is most in use in non-translated English for this sense category is *take*. Even here, we detect an over-representation of the verb form in the translation corpora, which together account for 56 % of the total occurrences. The less used verb forms have more occurrences in non-translations, except *takes* which, together with *taking* have too few occurrences across the corpora to provide reliable results.

The next sense category (‘bring’, ‘carry’) shows similar results. The most frequent verb form is *take* and 61 % of the overall occurrences were found in the translation corpora. The other verb forms that could provide reliable results in terms of sufficient overall occurrences found, have higher frequencies in non-translations than in translations, except for *taking*, where the occurrences were fairly evenly distributed across the corpora. When it comes to category 3 (‘hold’, ‘get hold of’), the occurrences of the most frequent verb forms, *takes* and *took*, and *takes* are more or less evenly distributed across the corpora, while *taken* has a slightly higher frequency in translations than in non-translations. The interesting result here, however, is the one concerning *taking*. We find that the vast majority of occurrences are found in non-translations (68 %), something that gives fuel to the indication above that this verb form does not correspond well with the alternative verb form in other languages. Furthermore, this might be an equivalence problem that especially exists between English and

Norwegian, since the least occurrences of this form were found in ENPC translations, which contains English translations from Norwegian source texts. Unfortunately, none of the other sense categories were found to have high enough frequencies for this verb form for the results to be deemed reliable, but a glance at the total number of occurrences in each of the corpora showed that ENPC translations had by far the lowest frequency of *taking*, something which also supports the above indications.

The next sense category that had an overall frequency high enough to look at in more detail is category 4 ('require'). The most frequent verb form in this category is *took*. In contrast to the first two categories, the majority of the occurrences of the most frequent verb form were in this category found in the corpora containing non-translations (79 %). In other words, there is a massive under-representation of *took* meaning *require* in translations. This might be in line with my hypothesis, since this sense category could be perceived to be a peripheral category in that it only appears in fourth position, with a relatively low frequency in comparison with the first three categories (see Table 4. 1.). In this case, we would expect the translators to under-represent the feature.

The last category to consider here is category 10 ('occupy'), which ended up with an unexpected number of occurrences in my data, which allows us to look more closely at it. In this category, *take* is the most frequent verb form, and there is a slightly higher frequency in translations (55 %) than in non-translations (45 %). This again suggests that translators tend to over-represent common target language features in their translations. There are also higher frequencies in translations than in non-translations in the second most (*taken*) and third most (*took*) frequent verb forms.

#### 4. 2. 5. Concrete/abstract object

The last aspect I wanted to see the verb *take* in the light of is whether it takes a concrete or an abstract object. The results found here were rather interesting as the sense of the verb seemed to have a clear effect on what type of object it took. Senses involving *movement*, such as 'move', 'remove', 'bring' and 'carry', were inclined to take concrete objects, whereas senses involving (*change of*) *possession*, such as 'hold', 'get hold of', 'accept' and 'choose', were more inclined to take abstract objects. However, the latter is perhaps not all that surprising, considering the several and frequent metaphorical uses that exist which derive from the concept of possession change (e.g. 'take advantage', 'take control', 'take the opportunity', etc.).

Given, then, that the senses denoting *movement* tend to take concrete objects and senses denoting (*change of*) *possession* tend to take abstract objects, we would then expect that translations would follow this pattern even stronger than non-translations. This is again based on the hypothesis that translators have a tendency to conform to, and even exaggerate, common features of the target language. In Table 4. 14. above we can see that this indeed is the case in category 1 ('move', 'remove'). Here, 84 % of the instances in non-translations take concrete objects while the same is true for 90 % of the instances in translations. Similarly, 60 % of the instances of category 3 ('hold', 'get hold of') take abstract objects in non-translations, while 70 % of the instances take abstract objects in translations. Category 5 ('accept', 'choose') also follows this pattern, with 65 % of the instances taking abstract objects in non-translations and 82 % of the instances take abstract objects in translations. It is safe to say that my results provide support for my hypothesis.

#### 4. 2. 6. Summary

It is quite plausible to attribute the findings from my results to the tendency for translators to normalise their language in terms of preferring frequent variants of the target language to less used variants. However, it is also shown above that other factors need to be taken into account in order to get a clear picture of how translators tend to behave. In Section 4. 2. 2. we saw that sense category 1 ('move', 'remove') was over-represented in translations in relation to non-translations. However, we later found that this was only true for fictional texts, as the frequencies in non-fictional texts were practically equal between the two. Thus, translators only tend to normalise their use of *take* in this sense in fictional texts. We also found that the factor of verb forms had implications for the distribution of *take* meaning *move* or *remove*. The two verb forms *took* and *takes* were the only ones which were heavily over-represented in translations. *Takes* had a total frequency which were too small to provide reliable results, but *took* was the most frequent verb form overall in this sense category. Thus, the over-representation of this verb form in translations might be a manifestation of the tendency for translators to conform to, and exaggerate, typical features of the target language.

The need to take other factors into account when comparing translations with non-translations becomes even more evident in the case of sense category 2 ('bring', 'carry'). Initially it seemed like *take* in this sense was equally distributed in both translations and non-translations. Bringing the factor of text type into the equation, however, we found that this sense was heavily over-represented in non-fictional texts, leading to the suggestion that normalisation, in this case, depends on text type. Also, we find that the most frequent verb

form for this sense, *take*, is markedly over-represented in translations in relation to non-translations. Thus, normalisation is surely an issue when it comes to category 2 as well, despite the initial discovery that it is equally distributed between translations and non-translations.

Sense category 3 ('hold', 'get hold of') showed a slight under-representation in translations. However, there was some indication of normalisation in this category too. The frequencies in the corpora containing translations were practically equal (see Table 4. 3.), while the frequencies in the corpora containing non-translations varied greatly (see Table 4. 4.). This might indicate that writers of non-translated language vary their language to a greater extent than translators do. Translators normalise their language by adhering to the average use by writers of non-translators.

When it comes to sense category 4 ('require') we found that it was slightly under-represented in translations. Since category 4 is defined above as a peripheral category (not a prototypical category), this was expected since it is thought that translators use fringe senses at a lower frequency than writers of non-translated language. However, when considering the factor of text type, we found an even greater under-representation of this sense in translations in fiction texts, but actually a slight over-representation in translations in non-fiction texts. This shows that normalisation is manifested by the under-representation in translation of peripheral categories, but that it depends on text type.

Another interesting sign of normalisation in translation was the uniformity of the two corpora containing translations which was a sharp contrast to the relatively significant differences between the two corpora containing non-translations. What I would like to suggest is that while writers of non-translations have great variation in their language, translators tend to conform to the centre of the axis, which result in little variation of language use.

Finally, we found that whether the verb takes a concrete or an abstract object is a relevant factor to include when comparing translations and non-translations. Interestingly, we saw that when the verb *take* denotes *movement* it was inclined to take concrete objects, while when it denotes (*change of*) *possession* it was relatively more inclined to take abstract objects. Furthermore, we found that this tendency is even greater in translations than in non-translations, which means that normalisation in translation takes place here too. The translators seem to adhere to, and exaggerate, the tendency displayed by writers of non-translated language.

## **CHAPTER 5: CONCLUSION**

Before plunging into the conclusions of this thesis, it is important to acknowledge the limitations of the present study. First of all, some general observations on choice of corpus type need to be addressed. The decision to make use of comparable corpora as material presents some issues when the objective is to make statements on translational behaviour. One of these issues is the lack of meta-linguistic information provided for by this type of corpora. Of course, they provide some information of this nature, like gender and nationality of translator/author and source text language, but aspects like status of source language in the target culture, status of translated texts in the target culture and the purpose of the translation are not included (nor are they expected to be). This makes it difficult for a researcher to provide socio-cultural explanations for her/his results. Another issue which arises when using comparable corpora as material is that the source texts are absent. The only types of texts that are given are the translations and non-translations in the same language. This means that the foundation for making explanations related to interference from source texts is non-existent.

Another issue is related to the more specific choice of corpora I have made. While the BNC and, arguably, the TEC are fairly large corpora, ENPC is considerably smaller (approximately 600,000 words per corpus). The size of ENPC might create some irregularities in the results, since the total amount of words may not be large enough to make up for translator-specific tendencies. A good example of this being the case is the frequency found in ENPC non-translations for sense category 7 (carry out) in non-fiction texts. The excessive use of phrases like “take measure” and “take steps” by one or two translators had considerable impact on my results, making this sense category’s frequency much higher than anticipated. A larger corpus would have evened out this irregularity.

A third issue presenting a limitation on this study is related to the notion of prototypicality and my list of sense categories. The definition of prototypicality is still debated. Still, one seems to agree that the establishment of prototypicality within a category is dependent on a number of factors, including corpus frequency, results from elicitation tests (first-comes-to-mind-tests) and cognitive constructs. Thus, creating a list reflecting the prototypical hierarchy of the different senses of *take* would be an enormous task, possibly able to fill a thesis on its own. However, it is my belief that the alternative method I employed, using dictionary listings, is satisfactory for the purpose of this paper.



Having gone through the formalities of stating the limitations of this study, we can now proceed to making some conclusions about my results and following discussion. To recapitulate, the aim of this thesis has been to investigate the notion of normalisation in translation in relation to the nuclear verb *take* and the different senses thereof. My claim was that translators tend to over-represent the prototypical senses of *take*, while they under-represent the peripheral senses of the verb. Furthermore, I suggested that factors such as text type, verb form and whether the verb takes a concrete or an abstract object play a part in translational behaviour.

My results supported my hypothesis to a certain degree. It was, indeed, found that translators tend to over-represent the prototypical senses of the verb *take*, however with some modifications. In some cases, translators over-represented a sense category in fiction texts and not in non-fiction texts, while in other cases it was the other way around. Moreover, we saw that verb form also was a decisive factor for translators' tendency to normalise their language, subconsciously, I assume. Another interesting finding was that whether the verb took a concrete or an abstract object was highly dependent on which sense of the verb that was employed. Senses denoting *movement* were inclined to take concrete objects while the opposite was true for senses denoting (*change of*) *possession*. The interesting thing about this, however, was that translators, almost without exceptions, exaggerated this tendency. In short, translators do have a general tendency to conform their language to typical features of the target language, but this tendency is also governed by other factors, like the ones described above.

Of course, this study does not provide conclusive evidence for the suggestion that translators tend to normalise their language. It does, however, give an idea about what future research into translational features, like normalisation, should take into consideration. Previous studies (like Kenny, 2001) have, as a rule, not included factors like text type, verb form, etc. It is my firm belief that factors such as these should be included in order to be able to get a clear picture of how translators behave. Also, the fact that the source texts are left out of the picture when dealing with comparable corpora, leaves the full story untold in that we have no way of determining whether our results are affected by interference from the source texts or source language. Thus, I would like to suggest that future research should make use of *both* comparable corpora and parallel corpora in order to bring all factors into the equation.

## References:

Aitchison, Jean. 1998. Bad birds and better birds: prototype theories. In Virginia P. Clark, Paul A. Eschholz, and Alfred S. Rosa (eds.) *Language. Readings in language and culture*, 6<sup>th</sup> edition. Boston: Bedford/St. Martin's, 225-239.

Baker, Mona. 1992. *In other words: a coursebook on translation*. London/New York: Routledge.

Baker, Mona. 1993. Corpus linguistics and translation studies: implications and applications. In Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds.) *Text and technology: in honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, 233-250.

Baker, Mona. 1995. Corpora in translation studies: an overview and some suggestions for further research. In *Target* 7 (2), 223-243.

Baker, Mona. 1996. Corpus-based translation studies: the challenges that lie ahead. In Harold Somers (ed.) *Terminology, LSP and Translation. Studies in language engineering in honour of Juan C. Sager*. Amsterdam/Philadelphia: John Benjamins. 175-186.

Blum-Kulka, S. and Levenston, E. A. 1983. Universals of lexical simplification. In C. Faerch and G. Kasper (eds.) *Strategies in interlanguage communication*. London: Longman. 119-139.

Chesterman, Andrew and Arrojo, Rosemary. 2000. Shared ground in translation studies. In *Target* 12, 151-160.

Chesterman, Andrew. 2000. A causal model for translation studies. In Olohan, Maeve (ed.) *Intercultural faultlines. Research methods in Translation Studies 1. Textual and cognitive aspects*. Manchester: St Jerome. 15-27.

Chesterman, Andrew. 2007. What is a unique item? In Gambier, Yves, Miriam Shlesinger and Radegundis Stolze (eds.) *Doubts and directions in Translation Studies*. Amsterdam/Philadelphia: John Benjamins, 3-13.

De Groot, Anette M. B. 1992a. Bilingual lexical representation: a closer look at conceptual representations. In Ram Frost and Leonard Katz (eds.) *Orthography, phonology, morphology, and meaning*. Amsterdam: North Holland, 389-412.

De Groot, Anette M. B. 1992b. Determinants of word translation. In *Journal of experimental psychology: learning, memory, and cognition* 18:5, 1001-1018.

De Groot, Anette M. B. 1993. Word-type effects in bilingual processing tasks: Support for a mixed representational system. In Robert Schreuder and Bert Weltens (eds.) *The bilingual lexicon*. Amsterdam/Philadelphia: John Benjamins, 27-51.

Gilquin, Gaëtanelle. 2006. The place of prototypicality in corpus linguistics: Causation in the hot seat. In Stefan Th. Gries and Anatol Stefanowitsch (eds.) *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*. Berlin/New York: Mouton de Gruyter, 159-191.

Gilquin, Gaëtanelle and Shortall, Terry. 2007. Reconciling corpus data and elicitation data in FLT. In *Proceedings of the Fourth Corpus Linguistics Conference*, University of Birmingham, 27-30. (<http://ucrel.lancs.ac.uk/publications/CL2007/>)

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

Halverson, Sandra. 2009. Elements of doctoral training: The logic of the research process, research design, and the evaluation of research quality. In *The interpreter and translator trainer (ITT)* 3 (1), 79-106.

Hatim, Basil & Munday, Jeremy. 2004. *Translation. An advanced resource book*. London/New York: Routledge.

Holmes, James S. 1972. The name and nature of Translation Studies. In *3<sup>rd</sup> International Congress of Applied Linguistics: abstracts*. Copenhagen, 88.

Holmes, James S. 1988. The name and nature of Translation Studies. In James S. Holmes (ed.) *Translated! Papers on literary translation and Translation Studies*. Amsterdam/Atlanta: Rodopi, 67-80.

Jakobsen, Arnt Lykke. 2006. Research methods in translation – translog. In Gert Rijlaarsdam (Series Ed) and Kirk Sullivan & Eva Lindgren (Vol. Eds), *Studies in writing, vol 18, computer keystroke logging: methods and applications*. Oxford: Elsevier, 95-105.

Johansson, Stig. 2007a. Spending time in English and Swedish. In *Nordic journal of English studies* 6/1. <http://ojs.ub.gu.se/ojs/index.php/njes/article/view/15/17>

Johansson, Stig. 2007b. *Seeing through multilingual corpora: on the use of corpora in contrastive studies*. Amsterdam: John Benjamins.

Kenny, Dorothy. 2001. *Lexis and creativity in translation. A corpus-based study*. Manchester: St. Jerome

Kenny, Dorothy. 2005. Parallel corpora and Translation Studies: old questions, new perspectives? Reporting *that* in Gepcolt. A case study. In Pernilla Danielson and Michaela Mahlberg (eds) *Meaningful texts: the extraction of semantic information from monolingual and multilingual corpora*. London: Continuum, 154-165.

Klaudy, Kinga. 1997. *The theory and practice of translation*.

Langacker, Ronald. 1987. *Foundations of cognitive grammar 1*. Stanford, California: Stanford University Press.

Langacker, Ronald. 1991. *Concept, image, and symbol*. Berlin/ New York: Mouton de Gruyter.

Langacker, Ronald. 1999. *Grammar and conceptualization*. Berlin/New York: Mouton de Gruyter.

Laviosa, Sara. 2002. *Corpus-based Translation Studies: theory, findings, applications*. Amsterdam/ New York: Rodopi.

Levinson, Stephen C. 2001. Covariation between spatial language and cognition, and its implications for language learning. In Melissa Bowerman & Stephen C. Levinson, eds. *Language acquisition and conceptual development*. Cambridge: Cambridge University Press, 566-588.

Malmkjær, Kirsten. 2008. Norms and nature in Translation Studies. In Gunilla Anderman and Margaret Rogers (eds) *Incorporating corpora - corpora and the translator*. Clevedon: Multilingual Matters, 49-59.

Mauranen, Anna. 2007. Universal tendencies in translation. In M. Rogers and G. Anderman (ed.) *Incorporating corpora: the linguist and the translator*. Clevedon: Multilingual Matters, 33-48.

Nordquist, Dawn. 2004. Comparing elicited data and corpora. In Michel Achard and Suzanne Kemmer (eds.) *Language, culture, and mind*. Stanford: CSLI Publications, 211-23.

Olohan, Maeve. 2004. *Introducing corpora in Translation Studies*. Oxfordshire/New York: Routledge.

Pym, Anthony. 1998. *Method in translation history*. Manchester: St Jerome Publishing.

Pym, Anthony. 2008. On Toury's laws of how translators translate. In Anthony Pym, Miriam Shlesinger and Daniel Simeoni (eds.) *Beyond Descriptive Translation Studies: investigations in homage to Gideon Toury*. Amsterdam: John Benjamins, 311-328.  
[http://www.tinet.org/~apym/on-line/translation/2007\\_toury\\_laws.pdf](http://www.tinet.org/~apym/on-line/translation/2007_toury_laws.pdf)

Rosch, Eleanor. 1975. Cognitive representations of semantic categories. In *Journal of experimental psychology, general* 104 (3), 192-233.

Reiss, Katharina. 1971. *Möglichkeiten und grenzen der übersetzungskritik: Kategorien und kriterien für eine sachgerechte beurteilung von übersetzungen*. München: Max Heuber Verlag.

Saeed, John. 2003. *Semantics*. 2<sup>nd</sup> edition. Oxford: Blackwell

Schmied, Josef. 1998. To choose or not to choose the prototypical equivalent. In Rainer Schulze (ed.) *Making meaningful choices in English. On dimensions, perspectives, methodology, and evidence*. Tübingen: Gunter Narr, 207-222.

Taylor, John R. 1989. *Linguistic categorization. Prototypes in linguistic theory*. Oxford: Clarendon Press.

Tirkkonen-Condit, Sonja. 2002. 'Translationese' – a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target* 14 (2), 207-220.

Tirkkonen-Condit, Sonja. 2004. Unique items – over- or under-represented in translated language? In Anna Mauranen and Pekka Kujamäki (eds.) *Translation universals: do they exist?* Amsterdam/Philadelphia: John Benjamins, 177-184.

Toury, Gideon. 1980. *In search of a theory of translation*. Tel Aviv: The Porter Institute for Poetics and Semiotics.

Toury, Gideon. 1995. *Descriptive Translation Studies and beyond*. Amsterdam/Philadelphia: John Benjamins.

Vanderauwera. R. 1985. *Dutch novels translated into English: the transformation of a 'minority' literature*. Amsterdam: Rodopi.

#### Web sites

Wordnet, 15. 05. 09.

<http://wordnet.princeton.edu/>

Wordnet search results, 15. 05. 09.

<http://wordnetweb.princeton.edu/perl/webwn?s=take&sub=Search+WordNet&o2=&o0=1&o7=&o5=&o1=1&o6=&o4=&o3=&h=>

Oxford online dictionary, 15. 05. 09.

<http://www.askoxford.com/?view=uk>

Oxford search results, 15. 05. 09.

[http://www.askoxford.com/concise\\_oed/take?view=uk](http://www.askoxford.com/concise_oed/take?view=uk)

Collins Cobuild, 15. 05. 09.

<http://www.collinslanguage.com/>

Collins Cobuild search results, 15. 05. 09.

<http://www.collinslanguage.com/results.aspx?context=4&reversed=False&action=define&homonym=2&text=take>

The ENPC homepage, 15. 05. 09.

<http://www.hf.uio.no/ilos/forskning/forskningsprosjekter/enpc/>

The ENPC manual, 15. 05. 09.

[http://www.hf.uio.no/ilos/forskning/forskningsprosjekter/enpc/ENPCmanual.html#\\_Toc445194138](http://www.hf.uio.no/ilos/forskning/forskningsprosjekter/enpc/ENPCmanual.html#_Toc445194138)

Oslo Multilingual Corpus (access ENPC), 15. 05. 09.

<http://www.hf.uio.no/ilos/OMC/>

TEC: a toolkit and application..., 15. 05. 09.

<http://www ldc.upenn.edu/exploration/expl2000/papers/luz/>

Longman online dictionary, 15. 05. 09.

<http://www.ldoceonline.com/>

Longman search results, 15. 05. 09.

[http://www.ldoceonline.com/dictionary/take\\_1](http://www.ldoceonline.com/dictionary/take_1)

The TEC homepage, 15. 05. 09.

<http://www.monabaker.com/tsresources/TranslationalEnglishCorpus.htm>

The BNC homepage, 15. 05. 09.

<http://www.natcorp.ox.ac.uk/>

The TEC (access), 15. 05. 09.

<http://ronaldo.cs.tcd.ie/tec2/jnlp/>

Sketch Engine (access BNC), 15. 05. 09.

<http://www.sketchengine.co.uk/>