Tracing Translation Universals and Translator Development by Word Aligning a Harry Potter Corpus

Sofia Helgegren

2005-05-24

LIU-KOGVET-D--05/09--SE

Magisteruppsats i kognitionsvetenskap
Instutionen för datavetenskap
Linköpings universitet

Handledare och examinator: Magnus Merkel
Abstract

For the purpose of this descriptive translation study, a translation corpus was built from roughly the first 20,000 words of each of the first four Harry Potter books by J.K. Rowling, and their respective translations into Swedish. I*Link, a new type of word alignment tool, was used to align the samples on a word level and to investigate and analyse the aligned corpus. The purpose of the study was threefold: to investigate manifestations of translation universals, to search for evidence of translator development and to study the efficiency of different strategies for using the alignment tools.

The results show that all three translation universals were manifested in the corpus, both on a general pattern level and on a more specific lexical level. Additionally, a clear pattern of translator development was discovered, showing that there are differences between the four different samples. The tendency is that the translations become further removed from the original texts, and this difference occurs homogeneously and sequentially. In the word alignment, four different ways of using the tools were tested, and one strategy was found to be more efficient than the others. This strategy uses dynamic resources from previous alignment sessions as input to I*Trix, an automatic alignment tool, and the output file is manually post-edited in I*Link.

In conclusion, the study shows how new tools and methods can be used in descriptive translation studies to extract information that is not readily obtainable with traditional tools and methods.
Acknowledgements

First and foremost, I would like to thank Magnus Merkel, my tutor, for all his patience, not to mention the invaluable support and feedback. Michael Petterstedt was a great help, especially in the early stages, by assisting in the setup and guiding in the use of the alignment tools. I am also grateful to Helge Dyvik for creating semantic mirrors from the resources built up during the alignment.

A big thanks to my family, who have supported and encouraged me throughout this sometimes slightly overwhelming process. Lastly, I am forever indebted to Oscar, who always listens and cares when I need it the most.
## Contents

1 Introduction 1

2 Background 5
   2.1 A Brief Introduction to the Harry Potter Series . . . . . . . . . . 5
      2.1.1 The Harry Potter Series and Culture . . . . . . . . . . . . 6
   2.2 The HP Series from a Translation Studies Perspective . . . . . . 7
      2.2.1 A Note on the Translator . . . . . . . . . . . . . . . . . . 7
      2.2.2 The Harry Potter Books as Novels . . . . . . . . . . . . 7
   2.3 Previous Studies of the Harry Potter Books . . . . . . . . . . . . 8

3 Translation Theory 9
   3.1 Translation and Culture . . . . . . . . . . . . . . . . . . . . . . 9
   3.2 Descriptive Translation Studies . . . . . . . . . . . . . . . . . . 10
   3.3 The Effect of the Translator . . . . . . . . . . . . . . . . . . . . 11
   3.4 Translation Universals . . . . . . . . . . . . . . . . . . . . . . . 11
      3.4.1 Explicitation . . . . . . . . . . . . . . . . . . . . . . . . 12
      3.4.2 Simplification . . . . . . . . . . . . . . . . . . . . . . . . 12
      3.4.3 Normalisation . . . . . . . . . . . . . . . . . . . . . . . . 13
   3.5 Translation of Fiction . . . . . . . . . . . . . . . . . . . . . . . 14
   3.6 Children’s Literature in Translation . . . . . . . . . . . . . . . . 15
   3.7 Constraints on Translation of Children’s Literature . . . . . . . . 16

4 Studying Translations 19
   4.1 Parallel Corpora . . . . . . . . . . . . . . . . . . . . . . . . . . 19
   4.2 Sentence Alignment . . . . . . . . . . . . . . . . . . . . . . . . . 20
   4.3 Part-of-speech Tagging . . . . . . . . . . . . . . . . . . . . . . . 20
   4.4 Word Alignment . . . . . . . . . . . . . . . . . . . . . . . . . 21
      4.4.1 Guidelines for Manual Word Alignment . . . . . . . . . . 21
   4.5 Non-1-to-1-operations . . . . . . . . . . . . . . . . . . . . . . . 22
   4.6 Lexical Shifts . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
      4.6.1 Strategy for Lexical Shifts . . . . . . . . . . . . . . . . . 25
   4.7 Paraphrasing and Lexical Choice . . . . . . . . . . . . . . . . . 25
## CONTENTS

4.7.1 Examples of Rejected Lexical Choices in the HP-corpus . 26

5 Methodology 27
   5.1 The Sequence of Work .................................. 27
   5.2 A Presentation of the Tools .............................. 28
      5.2.1 I*Link ............................................. 28
      5.2.2 I*Trix ............................................ 30
      5.2.3 New Tools, New Possibilities ....................... 30

6 The Making of the HP-corpus 33
   6.1 The Corpus .............................................. 33
   6.2 Word Aligning the Corpus ................................ 34
      6.2.1 Aligning HP1 ...................................... 35
      6.2.2 Aligning HP2 ...................................... 36
      6.2.3 Aligning HP3 ...................................... 37
      6.2.4 Aligning HP4 ...................................... 37
   6.3 Comments on the Alignment Process ..................... 37
      6.3.1 Problems Common to the Samples .................... 38
   6.4 Post-editing HP1 ........................................ 38

7 Results 41
   7.1 The HP-Corpus .......................................... 41
   7.2 Translational Results ................................... 41
      7.2.1 Additions and Deletions ............................ 42
      7.2.2 Translation Universals ............................. 49
      7.2.3 Investigating Translational Choices ................ 53
   7.3 Methodological Results ................................ 58
      7.3.1 Evaluation of the Different Strategies .......... 59

8 Discussion 61
   8.1 Discussion on the Translational Results ................ 61
      8.1.1 FDG Imperfections ................................ 61
      8.1.2 The Relationship between Additions, Deletions and Lexical Shifts ............................................. 62
      8.1.3 Translation Universals ............................. 63
      8.1.4 The Development of the Translator ................ 64
      8.1.5 Sources of Error for the Translational Results .... 64
   8.2 Discussion on Tools and Methodological Results ........ 65
      8.2.1 Using the Alignment Tools .......................... 65
      8.2.2 Advantages and Disadvantages of Using I*Link .... 65
      8.2.3 Specifics of I*Link as Sources of Error ........... 66
      8.2.4 Suggestions for Improvements of the Tools ........ 67
   8.3 Suggestions for Further Research ....................... 68

Bibliography 71
Chapter 1

Introduction

Translations and the original text they are supposed to be the equivalent of are not exactly the same. They differ in many ways, some changes occur naturally as the basic structure of different languages is not the same, and some are due to choices made by the translator. However, the differences between the two texts might not be caused just in the process of translation, but also by the process of translation.

Studies have shown that there are structural differences not only between a specific original and its translation, but also between translations and other texts written in the same language in general (Baker 1996). Translated text often has certain characteristics that sets it apart from other texts written in the same language. These characteristics are claimed to be the result of a subconscious process in translators to ensure that the text is understandable to the new readers in the new context.

Translation is all about context. It is about taking one text out of its cultural context and making it available to a whole new readership that is not a part of that cultural context, and therefore cannot have the same vantage point as a reader from the source culture reading the source text. Because of this, in translating the words of the text, the translator must also take the foreignness of the text into consideration, and decide whether that is something worth preserving for the foreign feel, or if it should be adapted to the target culture readers. Over the years, the translation studies community has shifted from favouring the source-oriented approach and its very close rendering of the original text, to the target-oriented approach that focuses on readability and achieving an equivalent effect in the target culture (Tabbert 2002). It is a shift from smaller segments to larger and from closeness to ease of understanding.

The objects chosen for this study are the first four books in the astoundingly successful Harry Potter series written by J.K. Rowling, and their translations into Swedish. There are many reasons behind this choice, but a fact that makes them so interesting to study is that they belong to the genre of children’s literature but have not been treated exclusively as such. They have attracted readers both older and younger than the intended one, and through their success, they
have gained a unique status in children’s literature. Moreover, the Harry Potter books belong to a sub-genre, namely fantasy.

There are two additional reasons to why the Harry Potter books were chosen for this study. Firstly, they are all translated into Swedish by the same translator, Lena Fries-Gedin. This fact makes it possible to study the books contrastively, and see if there are any structural differences between the translations. The translations may have changed over time and therefore, it is interesting to analyse the samples sequentially. With a formal description of actual changes, it would be possible to ascertain if the translator’s work is consistent over time or if changes can be detected.

Secondly, the four books were written, translated and published within a relatively short period of time, which makes it more likely that any contrastive differences between the samples are actually due to changes in the approach of the translator, and not any of the other possible sources of change, such as a change in the cultural climate due to long periods of time passing between the publication of the original text and the translation. This makes it possible to pose questions concerning whether the translator in some way develops over time and whether that is traceable in the produced texts.

In order to study the translations, the samples of source and target texts were aligned. Alignment is a method in which each sentence in the source text is paired with the corresponding sentence in the target text. This method was also used on a word-level, i.e. each word or cluster of words, depending on the nature of the text segments, was paired with the corresponding units. This allows for all translated segments to be linked together in units of source and target words, and makes visible the words in the source texts that were omitted in the translation process, as well as any words that were added to it, i.e. exists in the target text but not in the original. Consequently, changes in the text that occur in translation can be studied, which is why alignment is the chosen method for the study. This study is data-driven, and the hypotheses took a preliminary shape during the manual sentence alignment.

The purpose of this study is threefold. In the field of translation studies, one purpose is to investigate whether the so called translation universals are manifested in these texts, and if they are, what form do they take? The second purpose is to contrastively study the samples to discover whether there are detectable differences between them that could indicate that the translator’s approach has in some way changed from the first book to the fourth.

The methodological purpose is to investigate alignment and to evaluate the different alignment strategies used. Aligning as a method for studying translations is also evaluated, especially in relation to the new kind of information the new type of alignment tools used in this study can provide in comparison with traditional tools.

The hypotheses that will be investigated are:

- The translation universals are in some way manifested in the samples.
- The translator’s style in translating the texts will have changed over time,
primarily measured in the number of additions and deletions made to the
texts.

- Different strategies in using the alignment tool should affect the efficiency
  of the process of using the tools in significantly detectable ways.

The set of alignment tools that is used in this study is new and unexplored,
which means that this is a new type of study. Therefore, there are no existing
frameworks for analysis available, and a great deal of effort has gone into the
analysis of the material. The lack of a framework also means that a critical
approach must be taken concerning the usefulness of the tools. Consequently,
the advantages and disadvantages of the tools and the associated methods will
be discussed.

This is a first attempt at a new way of studying translations, and it must be
seen as such. No similar studies have been done, to my knowledge, neither using
the same type of tools, nor attempting to investigate the change over time in
the translations of one particular translator, of books from one particular genre
written by the same author. The point of this study is not to uncover universal
truths about translations, but to study one particular type of translation made
by one translator, and to present a way to systematically investigate translations
using new tools and methods.
Chapter 2

Background

It is difficult to find words to describe the success of the Harry Potter books, and considering the number of copies sold in both English and various translations worldwide, perhaps an introduction seems superfluous. Nevertheless, the books present a rather specific mix of two different worlds which presents difficulties to both translators of the series, and to readers of this thesis that are unfamiliar with the books. Therefore, a brief summary of some important aspects of the series is provided below. In addition, an explanation is given as to why the Harry Potter books were chosen for this study.

2.1 A Brief Introduction to the Harry Potter Series

To date, five books have been published in the Harry Potter series. Each book is set in two different worlds, one being the suburban boredom of number four, Privet Drive, in some fictitious town in middle England. The other is the exciting and action-packed magic world, predominantly set at Hogwarts School of Witchcraft and Wizardry.

The protagonist is the young Harry Potter, a lonely, unloved, bespectacled, friendless 11-year-old orphan who lives in Privet Drive with the Dursley family, consisting of his aunt Petunia, her husband Vernon, and their son Dudley. As the reader soon learns, Harry’s parents were a witch and wizard, and they were killed by an evil wizard named Voldemort when Harry was a baby. For some reason, Harry survived the attack with only a lightning-shaped scar on his forehead. Friends of his parents brought him to the Dursleys, who unwillingly accepted to bring him up.

The Dursleys want nothing to do with the magical world, and by ignoring it they hope to eradicate Harry’s potential magical powers. This proves to be fruitless, and as Harry is repeatedly invited to come to Hogwarts, they are in the end forced to give up. The game keeper and keeper of the keys at Hogwarts, Rubeus Hagrid, simply comes and takes Harry with him to go shopping for his
CHAPTER 2. BACKGROUND

Apart from this introduction to the magical world in the first book, the books all follow more or less the same format. They start in Privet Drive, in the summer holidays, with a bored and lonely Harry harassed by Dudley. As the school year starts, Harry by some means, usually the chartered Hogwarts Express leaving from platform 9 \(\frac{3}{4}\) at King’s Cross station, goes to the magical world of Hogwarts where great adventures of different sorts happen. The books end with a crisis and a sometimes bitter-sweet triumph for Harry in a fight in which he defeats the Dark Side, i.e. Lord Voldemort or some of his followers.

2.1.1 The Harry Potter Series and Culture

There are essentially three layers of culture in the Harry Potter books. The first is of course the image of normality, or life as the reader knows it, portrayed in the life in Privet Drive.

The second is the British public school system that the stories are so dependent on, particularly the fact that Hogwarts is a boarding school (Davies 2003). The Hogwarts culture is vividly described by Rowling through the use of boarding school elements, which contribute greatly to the very explicit Britishness of the books. A few examples of this is the Hogwarts Express, the chartered train that takes the students directly to Hogwarts, the school houses, dormitories and the Head Boys and Girls. Through the boarding school setting, the books portray a very British world, and one question the translator needs to ask him- or herself is if this should or should not be retained in translation.

The third layer is the magical world, which places the books in the fantasy genre. This layer is very much woven into the boarding school setting, and separating the two is perhaps not necessary for the purpose of this study. Suffice it to say that the basis in the books is always the normality and boredom of Privet Drive, skilfully contrasted with the other layers that serve to trigger the imagination and capture the interest of the reader. The complex interaction of the different layers of culture present an interesting challenge in the translation process.

Specifics of the Magical World

Apart from the Britishness of the books, they are very much characterised by the magical elements. Shopping for magical things (such as cloaks, spell books, pewters, potion ingredients and wands) is done in Diagon Alley, a street in a magical, parallel part of London unreachable for the non-magic people, or Muggles, i.e. those without magic power. The students at Hogwarts School of Witchcraft and Wizardry take classes such as Transfiguration, Defence Against the Dark Arts, Potions, and Care of Magical Creatures.

In the names of the professors and the rest of the characters, Rowling has used a lot of imagery and cultural references. This would normally pose a problem to translators and could be an interesting area of research, but the
2.2. THE HP SERIES FROM A TRANSLATION STUDIES PERSPECTIVE

The plots in the samples are of little interest as this is not a literary study, but some aspects of the magical world need explaining. This is because both the world of magic, witches and wizards, and the boarding school setting of Hogwarts closely woven into the magic world, pose a problem to translators. There is a vast terminology related to magic, which is, in effect, a subculture, and the usage of some terms differs between English and Swedish. In addition, Rowling frequently coins new terms and invents new concepts that are not normally associated with magic, for example the game *Quidditch*. These new concepts increase the complexity of the magical world, and are perhaps even more difficult to transfer in translation because they are completely new. Sometimes, these will carry certain connotations and cultural references that the translator must both recognise and succeed in translating.

Translating such a complex mix of worlds is neither easy nor straightforward. That is why this study focuses on the general patterns of choices made by the translator, rather than isolated mistakes or successful translational choices. This is also why the study focuses on a contrastive investigation of the samples; it would not be unreasonable to expect some sort of development in the translations, because of the large amount of text written by the same author in the same genre translated by the same translator.

2.2.1 A Note on the Translator

As mentioned above, the HP books are all translated by the same translator, Lena Fries-Gedin. She has been translating for nearly fifty years, starting when she was a student, and continuing parallel to her teaching career, but increasing heavily after her retirement. Fries-Gedin has mainly translated literature for adults, but because she had translated some books about a princess and dragons, she was offered to take on the first Harry Potter book, *Harry Potter and the Philosopher’s Stone* (Bergius 2003).

2.2.2 The Harry Potter Books as Novels

Placing the Harry Potter series in a genre is not as straightforward as could be expected. The obvious solution would be to state that they are children’s books, but I argue that this is not the whole truth. As O’Connell points out (1999), all children’s books are, to some extent, written at least in part for adults (see section 3.7), and for a number of reasons, this is even more so with the Harry Potter books.
As is obvious to any reader of the series, the length of the books has increased with every new published piece. Particularly the later books that span 636 pages for the fourth book (Rowling 2001a) and 766 pages for the fifth book, *Harry Potter and the Order of the Phoenix* (which is not included in the HP-corpus), demand very much more of a young reader than ordinary children’s fiction does. The length alone suggests that the books are meant to be read by fairly accomplished readers with a certain amount of patience and stamina, and for children perhaps even more so since no pictures or illustrations are used. Moreover, as Harry Potter grows older (as he does with every book, because each book describes the event of one school year), the plot becomes more complicated and the demands on the reader therefore increase. Consequently, at least the later books in the series merit discussion as novels, in my opinion, at least from a purely literary perspective.

The conclusion of the discussion above is that first and foremost, the books are fiction, as they portray fictitious events. Secondly, they contain many elements from the fantasy-genre. Thirdly, and naturally, they are children’s books. In general, however, I state that they can be seen as novels targeted on both adults and children. From a translational perspective, however, it is important to consider the fact that at least one part of the targeted audience is children, which will be explained in section 3.7.

### 2.3 Previous Studies of the Harry Potter Books

There are a few published studies on the Harry Potter books from a translation studies perspective. Eirlys E. Davies, for example, has studied the treatment of culture-specific items, or CSIs as she calls them, in *Harry Potter and the Philosopher’s Stone* and several of its translations (2003). This article is a very interesting read for anyone with a scholarly interest in the Harry Potter books, although most of what it covers is beyond the scope of this study.

The process of translating Harry Potter into Brazilian Portuguese is recounted in an article by professional translator Lia Wyler (2003). Though it is a reflection of her personal experience, it discusses the books from an insiders point of view, as well as gives an interesting peek into the Harry Potter phenomenon and its reception in Brazil.
Chapter 3

Translation Theory

In this chapter, relevant theory from the translation studies field is presented. The particular research questions investigated in this study are explained in connection with the corresponding background theories.

3.1 Translation and Culture

Translation is, in the words of Peter Newmark, “rendering the meaning of a text into another language in the way that the author intended the text” (1988, p. 5). The text to be rendered, the original, is commonly referred to as the source text, or ST. The text that the translator produces is the target text, or TT. Some words, phrases and concepts in the source language, or SL, have one-to-one correspondences in the target language, or TL, and are fairly simple to render in the new language.

However, “since no two languages are identical...it stands to reason that there can be no absolute correspondence between languages. Hence there can be no fully exact translations” (Nida 2000, p. 126). Languages are not identical, because a language and the culture in which it is used are very intimately connected, and any text that is produced in a certain language is an artifact of the accompanying culture. Naturally, this has implications when a text is to be translated, because “translation is a kind of activity which inevitably involves at least two languages and two cultural traditions, i.e., at least two sets of norm-systems” (Toury 1995, p. 56). Translating is taking a text out of its cultural context and bringing it into another, foreign context.

Because there can be no absolute correspondence between languages, translations must be closer to either the source or the target language. The source-oriented approach is literal translation, in which closeness to the original text is pivotal, whereas free translation favours the target language and culture (Newmark 1988). The distinction between the two is by no means absolute, and most translations are not fully, but to some degree, oriented towards either the SL or the TL.
In the history of translation studies, much discussion has pivoted around the concept of free and literal translation, and which one is to be preferred. Until the beginning of the nineteenth century, a free style that emphasised the spirit and sense of the text was favoured. After this, the study of cultural anthropology dictated that language “was entirely the product of culture”, which brought with it the idea that translation was nearly impossible, and that it at any rate needed to be as literal as possible (ibid., p. 45). This rather extreme point of view was gradually abandoned, however, and today, translations tend to be more target oriented (Baker 1996). Moreover, in translation studies, the prescriptive approach saying what a translation should be like has been replaced by a descriptive approach, aiming instead to explain what a translation is really like (Tabbert 2002).

3.2 Descriptive Translation Studies

According to Toury (1995), translation studies can be divided into sub-genres on different levels. On the first level it is a question of “pure” or applied translation studies. The latter concerns translator training, translation aids and translation criticism, which is beyond the scope of this study. The interest here is in pure translation studies, which can be either theoretical or descriptive. In turn, the descriptive branch is focused on either the product, i.e. the text itself, the process of translation, or the function of the text. Toury claims that the three are not as separate as the division implies, but that they are in fact to some degree interdependent on each other.

This study focuses on the product of the process, that is the text in itself, and the possible differences between the source and target versions. It is not a study of the process of translation, as the only artifact that is available for study is the text, and the text says very little about the process. The process is cognitive, and as with all cognitive processes there is a black box problem, in that processes that take place in the human brain cannot be studied in a simple way (Holmes 2000). However, with the help of the alignment tools used in the study, certain aspects of the process can be investigated through the linguistic patterns the translator produces, as the tools allow consistent differences between the source and target texts to be discovered. Patterns are, naturally, not inconclusive evidence of the translation process, but if strong and general patterns can be detected, this is in the very least an indication that the linguistic choices that are the basis of the patterns are indeed part of the process, and not just coincidence. What lies behind the specific choices made by the translator is, however, impossible to determine simply through studying the text and is beyond the scope of this study.

The received opinion, nowadays, is that the source text is just one factor of many that come into play in the translation process (Newmark 1988). Translations are instead seen as the product of a situational process, where elements like the translator in question, the target culture and the particular constraints on the situation (such as deadlines, payment etc.) interact and influence the
3.3 The Effect of the Translator

Traditionally, translation has not been seen as a creative activity, and translators are not supposed to have a style of writing of their own that is visible in the target text (Baker 2000). However, it is a truth universally acknowledged in the field of translation studies that if a number of translators were all given the same source text to translate into the same language, not many sentences would be translated in exactly the same way. If there is so much variation in the way different people translate, there must be an effect of the translator. The question is how, and indeed if, such an effect can be studied.

A small-scale study made by Mona Baker suggests that it is possible to “identify patterns of choice which together form a particular thumb-print or style of an individual literary translator” (ibid., p. 260). The focus in such studies, Baker emphasises, must be on the patterns the translator produces, rather than on the specific cases that could be brought up in order to prove a certain point. These patterns can be investigated using a corpus made up of large parts of the translator’s production.

In investigating the style of a translator, his or her background and what is known about it must be taken into consideration, and “whatever we manage to establish as attributable to the translator’s own linguistic choices must be placed in the context of what we know about the translator in question” (ibid., p. 258). In addition, the relationship between the cultures involved is significant, specifically whether they are closely related or disparate.

The HP-corpus only contains texts from one genre, written by the same author, and is in no way representative of the scope of Lena Fries-Gedin’s work. Therefore, it is natural that anything that can be said here about her translating style is limited to the material used in this study. It is specific to this genre of text, written by this author. However, the four samples can be compared and contrasted sequentially, in order to reveal whether the specific style in this context has changed from the first to the fourth book.

3.4 Translation Universals

Translations have certain universal features that separate them from original texts, and these features are caused in and by the process of translation. Mona Baker has given this issue a lot of attention, and states that the universal features come natural, since “the nature and pressures of the translation process must leave traces in the language that translators produce” (Baker 1996, p. 177).

One challenge that faces scholars interested in the universals of translation is that they are rather vague notions and studying them is by no means straightforward or easy. The first question to ask is in what way each feature might be manifested in a particular text, and how these manifestations can be located.
When this is done properly, a computerised corpus should be able to provide a lot of information and is the proposed basis for a study of the translation universals.

Baker focuses on three universals of translation, namely explicitation, simplification and normalisation, and the combined effect of the three is that translations are usually less complex than their original texts. This is particularly interesting in a study of the translation of children’s literature, since the strategies that are less faithful to the original but serve to adapt the text to the target language are used more freely for this genre in order to achieve texts that are easy to read (O’Connell 2003).

3.4.1 Explicitation

The theory of explicitation concerns the tendency in translations to “spell things out rather than leave them implicit” (Baker 1996, p. 180). Explicitation can be expressed syntactically or lexically. For example, translated texts tend to have a higher degree of conjunctions than original texts. Lexical explicitation can be made through various means, but oftentimes it is made by adding nouns in order to explain some piece of information that needs to be explained to a target culture reader.

Another possible manifestation of explicitation is the fact that translations tend to be longer than their original texts. When translations become longer, the additions to the ST are often made to explain features in the ST that might not be known to readers in a TT-culture. Thus the translation becomes more understandable than a more faithful rendering. This manifestation has the advantage of being relatively easy to examine.

In this study, explicitation is thought to manifest itself in two ways. Firstly, that the TTs are longer than the STs was evident on a very early stage. Secondly, if more information has been added to the target texts than removed from the source texts, this also indicates that they have been explicitated.

3.4.2 Simplification

Simplification is the tendency of translated texts to contain simplified language compared to the original text (ibid.). For example, long sentences are often divided into several shorter ones.

One indicator of simplification is a relatively low lexical density, meaning that the number of function words or grammatical words is high, in proportion to the number of lexical words. Lexical words contain more information than grammatical words, and using fewer lexical words means that the reader will have to keep track of less information. Using less varied vocabulary is also one manifestation of simplification.

Another possible sign of simplification is that punctuation tends to change in translations. According to Malmkjær (1997), punctuation is rateable on a scale from weak to strong in the order comma, semicolon and full stop. In translations, punctuation usually becomes stronger, in that commas are often translated into
semicolons or full stops, and semicolons are translated into full stops. If the punctuation is stronger, it is highly likely that there are more sentences in the TT than in the ST, which indicates that long and complex sentences have been divided into several shorter ones, and thereby the complexity of the text has been decreased.

In the HP-corpus, simplification is assumed to be manifested in long sentences being divided into several shorter ones, stronger punctuation and the removal of the regional dialects that some characters speak in (see discussion below).

### 3.4.3 Normalisation

*Normalisation or conservatism* is what Baker calls the “tendency to exaggerate features of the target language and to conform to its typical patterns” (1996, p. 183). This can take the shape of the translator over-using clichés or typical grammatical structures of the TL, often grammaticising elements of texts that are ungrammatical in the source.

Normalising also involves adapting the punctuation to the typical usage of the TL. For example, commas are used much more in English than in Swedish. Ingo states that a Swedish reader is much disturbed by an overuse of commas, and strongly recommends that the amount of commas is adapted to the usage of the target language (1991). One of the ways in which normalisation will be investigated in the HP-corpus is through the treatment of punctuation, and whether or not any evidence can be found of it being adapted to fit Swedish usage.

Another element of the Harry Potter books in which normalisation might be manifested is in the treatment of the different dialects used for certain characters in the source texts dialogues. Dialect “differs from person to person primarily in the phonetic medium” and “has to do with the user in a particular language event: who (or what) the speaker/writer is” (Hatim & Mason 1990, p. 39). The effect of changing a character’s dialect can be considerable, as in the French version of the first Harry Potter book, where the dialect of Rubeus Hagrid has been normalised and grammaticised (Davies 2003). In the English versions of the books, Hagrid’s speech casts him as a “down-to-earth, simple, uneducated and in some ways childlike character” but in the French version, his utterances are “characterized by impeccable grammar and standard, even somewhat formal vocabulary” (ibid., p. 82).

Dialect is a language variation that is dependent on the user, and Hatim and Mason distinguish between idiolectical, geographical, temporal, social and standard/non-standard variation (Hatim & Mason 1990). For the purpose of this study, the main interest in dialect is the use of different geographical dialects, or accents. Accent is the variation in language that roughly corresponds to the geographical origin of the speaker. Accents can carry ideological and political implications that translators must be aware of, and because of this translation of accent is problematic (ibid.).
In the Harry Potter series, accent is used actively in the depiction of different characters, not only for Rubeus Hagrid, but also for Stan Shunpike, the conductor on the Knight Bus in *Harry Potter and the Prisoner of Azkaban* (Rowling 2000a). Through alternative spelling in the utterances of Hagrid and Stan Shunpike, that clearly deviates from standard English spelling, Rowling represents the phonic qualities specific to two very different geographical dialects.

Example of Hagrid’s dialect (Rowling 1998, p. 48):

'It’s gettin’ late and we’ve got lots ter do tomorrow,’ said Hagrid loudly. ‘Gotta get up ter town, get all yer books an’ that.’

Example of Stan Shunpike’s dialect (Rowling 2000a, p. 31):

'Can’t do nuffink underwater. 'Ere,’ he said, looking suspici ous again, 'you did flag us down, dincha? Stuck out your wand 'and, dincha?'

Both dialects are to certain extents ungrammatical, and it could prove interesting to see if the translator has chosen to grammaticise the utterances, or adapted them to Swedish in some other way. Significantly, the dialects are very different, and should this difference not have been retained in the target texts, this is not only an instance of normalisation, but also of simplification, since it decreases the complexity of the texts.

3.5 Translation of Fiction

The books in the HP series belong to the fantasy genre, which also entails that they are fiction. In translation theory, it is very difficult to find theorists that speak about fiction with any interest at all. The focus tends to be on literary texts, which are considered to be serious and artistic, and neither fiction nor children’s literature is usually included in this category. However, due to the reasons stated in section 2.2.2, I argue that the Harry Potter books have many of the elements that characterise serious literature, and are therefore subject to some of the same constraints.

Bearing this in mind, there are a number of issues particular to the translation of literary texts that put constraints of the translator, demanding a lot of effort. Newmark distinguishes between three functions a translation must meet, namely the expressive, the informative and the vocative functions (1988). There is no strict division between these, and elements of all three can usually be found in most texts, although to different degrees. Fiction, in the form of novels, is placed among the serious imaginative literature as having mainly expressive and vocative functions.

Prominent for the expressive function is the mind of the writer, who “uses [the] utterance to express his feelings irrespective of any response” (ibid., p. 39). This is reflected in the writer’s personal use of language, and Newmark
emphasises that those personal, expressive, elements must not be normalised in translation. Examples of expressive elements can be “untranslatable’ words”, unconventional syntax, neologisms and uses of dialect (ibid., p. 40).

The vocative function concerns the readership, and the intended effect of the text to make the reader “act, think or feel, or indeed ’react’ in the way intended by the text” (ibid., p. 41). One factor in these texts is the relationship between the writer and the readership. Another is the fact that “these texts must be written in a language that is immediately comprehensible to the readership” (ibid., p. 41-42). It can be argued that in the case of children’s literature, this is especially important due to an assumed difference in linguistic skills and world knowledge between the translator and the readership (see discussion below).

3.6 Children’s Literature in Translation

Toury claims that translations usually occupy peripheral positions in the target literary system (1995). The more peripheral a text or its genre seems to be to the target culture, the more adjustments of the text will the translator tend to make in order to adapt it to the norms of the receiving culture.

Children’s books and translations of children’s literature tend to be seen as peripheral in most systems, something that can affect the process of translation. Shavit (1981) argues that translators of children’s literature have a much greater degree of freedom in relation to the source text, and “can permit himself great liberties regarding the text because of the peripheral position [of] children’s literature” (p. 171). However, this generalisation does not hold for the Harry Potter series, as it cannot be rightfully described as peripheral. Nevertheless, Shavit’s argument is still valid for the first book, which was still peripheral at the time of translation into Swedish.

Stolze (2003) has indicated that it can be questioned whether or not the translation of children’s literature is indeed different from the translation of adult literature, since the original of a translation for children was also written for children, as adult novels in translation are originally written for adults. Stolze’s opinion is that this is dependent on the way children are seen in different cultures, and that they should not be looked down upon as not being able to understand many things. However, translating takes place in the publishing industry, in which children are indeed marginalised (O’Connell 1999).

Consequently, the translation of children’s literature is subject to certain constraints that sets it apart from translating for adults. O’Connell (2003) points out that children have their own culture into which adults, among them the translator, have limited insight. Moreover, there is a significant difference “between the knowledge and linguistic skills of the translating adult and the children who make up the target language audience” and in translating for adults the translator can “expect the target readership to have approximately corresponding levels of linguistic skills, general knowledge and world experience” (ibid., p. 229). The knowledge level of the receiving audience is indeed a constraint in the translation of the Harry Potter series, because of the fact that
it is set in such a British environment and contains so many concepts that are completely foreign to Swedish children.

3.7 Constraints on Translation of Children’s Literature

Children’s literature was for a long time a neglected area in translation studies (O’Connell 1999). Today, it enjoys much more attention and both descriptive and theoretical studies on the subject abound (Tabbert 2002).

Eithne O’Connell points out four features specific to this genre, indicating some issues that separate translating children’s literature from translating adult literature (1999). Children’s literature:

1. has two specific audiences, namely children and adults.
2. has ambivalent texts, with both literal meaning and a deeper, interpretable meaning.
3. is written and purchased by others than the primary readership, i.e. adults.
4. has many functions and cultural constraints, in that they are intended to both entertain and educate.

The fact that the genre has two audiences has some interesting implications. In the relationship between adults and children, the power is with the former group, which is very much reflected in the area of children’s literature. Adults write, edit, publish, market and buy the books that are intended for children, which means that the primary audience is more or less without say when it comes to what they read. Parents decide what is suitable for their child, but children and adults are not likely to have the same taste in literature (ibid.).

Number two above, although worth investigating, is not something that will be pursued further in this study, as it is more interesting to do so from a literary angle.

Because works of this genre are produced in a more or less exclusively adult environment, it is important for the adults in that environment to be very much in touch with current children’s culture. In all literary production, the writers, publishers, editors and indeed, translators, have to be aware of the current trends in the culture for which they produce, which is not a trivial matter, and in children’s literature, it is complicated by the fact that adults cannot be equal members of the child community. Still, they must know and understand the culture, in terms of what children find interesting, how they speak and think, current vocabulary, and so on. Otherwise, the style of the language used in the translation risks being dated, and the readers will notice this. As Eirlys E. Davies points out, “translating for children may present more of a challenge than translating for adults; young readers are perhaps less likely to be tolerant of the occasional obscurity, awkwardness or unnatural-sounding
phrasing which adults, conscious that they are dealing with a translation, may be more accepting of” (2003, p. 66).

Due to the educational goal of children’s literature, studying explicitation, simplification and normalisation might be of particular relevance, as there is an even greater need to make texts understandable for the readership in order to meet with the goal to educate. One important part of the purpose to educate is, as Puurtinen (1998) points out, that adults expect children’s literature to help in the development of the child’s linguistic skills. Therefore, there might be a stronger tendency for translators of children’s literature to normalise the texts by grammaticising them, in order to avoid the readership learning faulty grammar from the books.
Chapter 4

Studying Translations

This chapter gives background information on corpora and how alignment of corpora can be used in studies of translations. It also explains how some complex changes that translators make in translating texts are treated in the alignment.

4.1 Parallel Corpora

Originally, the word *corpus* was used for a collection of writings, usually written by the same author. In modern corpus linguistics, it has come to mean “a collection of texts held in machine-readable form and capable of being analysed automatically or semi-automatically in a variety of ways” (Baker 1995, p. 225). Corpora are created for specific purposes, and can be of different types depending on the intended use.

*Parallel corpora* consist of texts that in some way are parallel. The typical parallel corpus contains original texts written in one language or language variety, and one or more translations of this text into one or more target languages, or language varieties (Borin 2002). The relationship between the text(s) and its translation(s) is one of *translation equivalence* (ibid.).

With parallel corpora, translated text can be studied in a number of ways, but in this study, the point is to discover *translation effects*. The basic idea behind this concept is that translated text can be linguistically and structurally different from original text, and in what way they differ can be discovered comparing STs with their TTs through the use of parallel corpora.

When starting a parallel corpus project, the first step is to select the texts to be included and create electronic versions of them. This can often be quite time-consuming, as it usually involves a great deal of manual work, such as typing, scanning and proofreading the material (ibid.). Borin also points out that the use that can be made of parallel corpora depends heavily on which type of tools are available to the researcher. However, the next step in the process can be done manually without the use of specialised tools.
4.2 Sentence Alignment

Alignment of the corpus texts is a process performed on parallel corpora. Aligning a corpus is “the process of identifying and pairing up corresponding units in the two (or more) languages making up the parallel corpus” (ibid., p. 20). This can be done on different levels, for example sentence alignment and word alignment.

In sentence alignment pairs of more or less equivalent source and target sentences are by some means put next to each other, which can be done by using simple tables. This is done to discover the most obvious changes to the text, such as elements of meaning being transferred to another sentence in the TT, long sentences being translated as several short ones, and extensive omissions and additions. An excerpt of the sentence aligned HP-corpus is shown in table 4.1, in which the second sentence pair is an example of how the translator has chosen to translate one sentence as two sentences.

<table>
<thead>
<tr>
<th>Source sentence</th>
<th>Target sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>He sat up and Hagrid’s heavy coat fell off him.</td>
<td>Han satte sig upp och Hagrids tunga rock f&quot; oll av honom.</td>
</tr>
<tr>
<td>The hut was full of sunlight, the storm was over, Hagrid himself was asleep on</td>
<td>Rucklet var fyllt av solljus, stormen var ˚ over, Hagrid sj¨ alv sov på den</td>
</tr>
<tr>
<td>the collapsed sofa, and there was an owl rapping its claw on the window, a</td>
<td>nersjunkna soffan och det var en uggla som knackade med klon p˚ a f¨ onstret. I</td>
</tr>
<tr>
<td>newspaper held in its beak.</td>
<td>n¨ abben h¨ oll den en tidning.</td>
</tr>
</tbody>
</table>

Table 4.1: An excerpt of the sentence aligned corpus.

For small corpora like the HP-corpus, sentence alignment can be done quite easily using basic word processing software such as Microsoft Word. For larger collections of text, automatic tools are necessary.

4.3 Part-of-speech Tagging

Once the sentence alignment is done, the corpus can be classified on a more fine-grained level. Part-of-speech tagging (henceforth POS-tagging) of the words in the corpus is one way to proceed that is often used in translation studies. POS-tagging is done because keeping track of the structural information of words and other text components is relevant. In translation, words and segments of a source text will sometimes change word class or have another function in the target text. The voice can also change, from passive to active or vice versa. These small linguistic changes can be indicators of more wide-spanning changes done to the text, which makes them liable for further investigation. Modern language processing tools such as the Machinese Syntax by Connexor
4.4 Word Alignment

To be able to discover when the corresponding word is of a different word class in the TT than in the ST, the texts must be aligned on the word level. The ST word (or words) must be linked with the corresponding TT word (or words), and for this task, specialised software tools are required.

Traditionally, word alignment is done automatically and the performance of the software that is used is evaluated on both precision and recall. Precision is “the accurateness of the links relations” and recall is “the number of possible links that are retrieved” (Sägvall Hein 2002, p. 68). The automated systems tend to have precision figures ranging from 80 to 95 percent (Merkel, Petterstedt & Ahrenberg 2003). As for recall, automatic alignment systems do well if the texts contain only one-to-one correspondences, but have severe difficulties in identifying multi-word units, “especially those that are discontinuous or have a low frequency; it is more or less impossible to know exactly how many multi-word units there are in a text” (Ahrenberg, Merkel, Sägvall Hein & Tiedemann 2000, p. 2). This causes problems for the recall measure, which can “therefore in practice only be made on samples of a bitext” (ibid., p. 2). Since very few texts contain only one-to-one correspondences, the performance of automated systems is simply not good enough if a full investigation of all words and tokens in a corpus is to be carried out. However, for very large corpora, manual alignment is not an option because of the workload involved, and in such cases, it is necessary to use an automated system.

4.4.1 Guidelines for Manual Word Alignment

When aligning a corpus manually, it is important to link the material as consistently as possible, which is difficult to achieve when several annotators work together on one project (Merkel 1999). But even with just one annotator, it is important to work with consistency in mind. In my opinion, the task of achieving consistency becomes more complex the larger the corpus. Not only specific terms, names and other lexical items need to be consistently aligned, but also syntactic structures, and remembering exactly how one treated a word or construction 1000 sentence pairs ago is not always easy. In this sense, the annotator’s job is reminiscent of the translators, and the same challenges face both the one producing the target text, and the one studying that very translation.

The general guidelines used in the annotation of the HP-corpus were via Merkel (1999) adopted from Véronis:

1. Mark as many words as necessary on both the target and source side.
2. Mark as few words as necessary on both the target and source side.
Following the guidelines is supposed to ensure that all links have a two-way equivalence between the source and target segments.

### 4.5 Non-1-to-1-operations

When aligning a corpus it becomes evident that some segments of the ST do not have a one-to-one correspondence with a TT segment, and the annotator is forced to link together segments in (usually) larger chunks. These non-1-to-1-operations include additions, deletions, convergences and divergences (Merkel 1999).

The focus of this study is on the segments of both source and target texts that do not have a corresponding segment in the other language, namely additions and deletions. These are significant changes to the text made by the translator, and in the aligning process, they lead to the annotator marking the segments as NULL-links, i.e. segments without corresponding segments. This does not apply to divergence and convergence, and they will only be mentioned briefly below for completeness. All examples below are taken from the Harry Potter corpus.

**Divergence and Convergence**

*Divergence* is when a construction spans more segments in the target text than in the source text. *Remember* in the example below corresponds to *komma ihåg*, and a one word construction in the source text has become a two word construction in the target text.

Example:

He rolled onto his back and tried to remember the dream he had been having.

Han rullade över på rygg och försökte komma ihåg drömmen han hade haft.

*Convergence* is the opposite, when the TT equivalent of an ST-expression spans fewer segments. In this example, the two word construction in the source text corresponds to the one word construction in the target text.

Example:

At last.

Äntligen.

Divergence and convergence are oftentimes necessary operations that are motivated by differences between the languages that need to be accommodated for. Additions and deletions, however, are rarely completely motivated by differences in the languages, but rely more heavily on the choices of the particular translator.
Additions

Translators sometimes add information to the text, and those additions are elements of the TT that are not present in the ST. The effect an addition has on a text is to a great extent dependent on the linguistic nature of the addition. It is reasonable to expect that added verbs, nouns and adjectives add actual information, whereas added pronouns can indicate that the translator has in fact grammaticised the text. In the ideal case, the translator only makes additions when it is absolutely necessary. However, this is not always the case, as can be seen in the example below, where Fries-Gedin has added the equivalent of long, a piece of information that is not motivated by the meaning of the source word cloaks.

Example:
People in cloaks.
Folk i långa mantlar.

Deletions

Deletions occur in the aligned material when the translator has chosen not to include some piece of information from the ST. The effect of a deletion is usually that the text has been simplified. In the example below, around has been deleted.

Example:
He looked around at Harry and Hermione.
Han såg på Harry and Hermione.

Should the source sentence contain a deletion and the target sentence an addition, it can be reasonable to suspect that there might be a relationship between the two.

Studying Additions and Deletions

Deletions and additions are structural changes that are easily detectable with the tools and methods used in this study. Thus the interest in these particular changes is twofold, partly motivated by the ease of structuring and studying them with the available tools, and partly by the fact that they are rarely completely necessary operations. Additions and deletions tend to a great extent to be based on subjective judgements made by the isolated translator, and therefore depend heavily on the individual translator.

As a general recommendation for translators, Newmark emphasises the naturalness of the target text (1988). Accuracy, however, is even more important and “you have no licence to change words that have plain one-to-one translations just because you think they sound better than the original, though there
is nothing wrong with it” (ibid., p. 36). Specifically, “mind particularly your descriptive words: adjectives, adverbs, nouns and verbs of quality” (ibid., p. 36). Consequently, the use a translator makes of adding or deleting descriptive words and segments to or from the text can be seen as a part of his or her style of translating, and will be the focus of the investigation into how Fries-Gedin uses addition and deletion in the samples.

4.6 Lexical Shifts

In translations, the meaning of some segments is sometimes changed between the source and target texts. These lexical shifts can be of three different types, according to Merkel (1999). The translated lexical item can be:

1. less specific, i.e. more general, than the source item.
2. more specific than the source item.
3. neither less nor more specific and not equivalent, i.e. it has a different meaning than the source item.

These definitions can also be termed a less specific shift, a more specific shift, and an other lexical shift (ibid.). Examples of the different types of lexical shifts are given below. The bold faced words are the source item and its chosen translation. Gloss translations of the actual meaning of the chosen target segments are given in the square brackets in the English sentences, illustrating the lexical shifts (whelk has in Swedish been generalised into [seafood], it has been specified as [The stench], and darkly has been changed into [quietly]).

Example of a less specific lexical shift:
Ate a funny whelk [seafood].
Åt nät konstigt skaldjur.

Example of a more specific lexical shift:
It [The stench] seemed to be coming from a large metal tub in the sink.
Stanken verkade komma från en stor plåtbalja i diskhon.

Example of an other lexical shift:
The giant chuckled darkly [quietly].
Jätten skrockade tyst.

Like additions and deletions, lexical shifts are significant changes made to the text, and they are rarely necessary to make. Consequently, analysing translations in terms of lexical shifts can illustrate the influence of the translator on the text.
4.6.1 Strategy for Lexical Shifts

In the word alignment system used in this study, it is not possible to mark segments where lexical shifts have occurred as lexical shifts. As a result, the choice is either to accept lexical shifts as regular translations, or to mark the segments as additions and deletions.

In this study, where the influence of the translator is measured in significant changes made to the text, it is important to solve the dilemma of how to mark lexical shifts. Some lexical shifts are perhaps necessary to make, due to differences in vocabulary between the source and target languages. Such necessary shifts do not depend as heavily on the choices of the particular translator, and can in this study therefore be linked as regular translations.

For the lexical shifts that the translator has made voluntarily, my solution is to focus on the degree of change each specific lexical shift implies. If a small change has been made, like when a pronoun has been changed into the noun it refers to, as in the example showing a more specific lexical shift on the opposite page, I have chosen to somewhat reluctantly accept the segments as a regular translation. This is because although these lexical shifts do imply that the meaning of the segment has been changed voluntarily, they do not change the meaning of the reference, they only make it more explicit. Above all, they are not as significant changes as additions and deletions.

Where the target segment is farther removed from the meaning of the source segment, however, as for most lexical shifts, I have opted to mark these segments as additions and deletions. This includes the examples for less specific lexical shift and other lexical shift on the opposite page.

The advantage of the chosen strategy is that it at least makes small changes distinguishable from significant changes, that depend more heavily on the choices of the translator. The disadvantage is that smaller more and less specific lexical shifts cannot be distinguished from regular translations, and more significant more and less specific lexical shifts, as well as other lexical shifts, cannot be distinguished from additions and deletions. The implications of this will be further dealt with in section 8.1.2 in the discussion chapter.

4.7 Paraphrasing and Lexical Choice

When aligning a corpus, the passages that are the most problematic tend to be those that paraphrase the meaning of the source words. It is very difficult indeed to draw a line between what is a working paraphrase and what is too far from the original sentence to be accepted as a natural and accurate translation.

Paraphrases sometimes border on errors in lexical choice, and it is not always easy to determine whether the translator has made a mistake or not. In such cases, the annotator must trust his or her own resources, both in the form of personal knowledge about a word, concept or activity, as well as dictionaries and other sources of linguistic information. The annotator must, in the end, make a choice and either accept or reject the choice of the translator.
4.7.1 Examples of Rejected Lexical Choices in the HP-corpus

The focus of this study is on patterns that differ between the source and target texts, but in order to briefly explain my strategy in the alignment process, a few examples of dubious lexical choices and how I chose to treat them are needed. In any case where I was reluctant to accept the choice made by the translator, I consulted one or more dictionaries, both English/Swedish and English/English.

One example of a lexical choice I rejected was the choice the translator had made for sherbet lemons, the Muggle sweet that Albus Dumbledore eats in the first chapter of the first Harry Potter book, *Harry Potter and the Philosopher’s Stone* (Rowling 1998). This is translated as citronisglassar, and the literal translation in English of this is lemon ice lollies. This particular lexical choice is not equivalent to the source segment, i.e. it is an instance of an other lexical shift. Furthermore, it is semantically impossible even within the context of the story, as Dumbledore explicitly states that they are a Muggle sweet kept in a bag in his pocket, which is an impossible way to store an ice lolly. I chose to not accept this link and thus treated the former as a deletion and the latter as an addition.

One other recurring lexical choice that was treated as a deletion/addition pair was don’t ask questions and its chosen translation, kom inte med några frågor. This is because kom inte med appears dated and is not common Swedish usage, whereas the English equivalent is common usage in the source language. Consequently, kom inte med was marked as an addition, and don’t ask as a deletion.
Chapter 5

Methodology

This chapter outlines how the HP-project was carried out, and describes the specialised software tools that were used in the process. In addition, some advantages of using these new types of alignment tools are explained.

5.1 The Sequence of Work

The sequence of work in this study can be summarised as follows.

1. The texts to be included in the corpus were chosen and read, and a decision was made on a suitable size of the samples.

2. The sample texts were transferred to electronic form, in this case by scanning, and the texts were proof-read.

3. The samples were aligned manually on the sentence level.

4. Machinese Syntax by Connexor was used to supply part-of-speech-tags to all tokens in the samples.

5. The POS-tagged samples were word aligned using two different tools, I*Link and I*Trix. The two tools were combined in four different strategies. Each strategy was used for one sample only, in order to enable an evaluation of the chosen tools and strategies.

6. The word aligned samples were studied using LinkInspector and LinkReporter, tools included in I*Link, and the results were analysed.

7. A small scale case study was performed on the treatment of the dialects of the characters Rubeus Hagrid and Stan Shunpike.

8. A close investigation of the last 150 sentence pairs of HP4 was made in order to investigate possible relationships between additions and deletions and to search for manifestations of the translation universals in more detail.
5.2 A Presentation of the Tools

In the word alignment, two different tools were used, I*Link and I*Trix. Both were developed at NLPLAB, the natural language processing division of the computer department of Linköping University.

5.2.1 I*Link

The word alignment system used in this study, I*Link, is interactive in that it is used in collaboration with a human annotator in order to increase the efficiency and performance of the tool. In collaboration with a human annotator, the precision figure of I*Link is more or less 100 percent, which is necessary in this study. In order to study the entire samples and search for patterns, the entire samples including complex structures that are sometimes very difficult to align must be as fully aligned as possible.

I*Link is a semi-manual alignment tool that uses information from bilingual resources and built-in heuristics to suggest correspondence candidates for alignment, which the user accepts, revises or rejects (Merkel et al. 2003). Any element the tool cannot suggest a match for, the user chooses a match for manually by clicking on the matching word, should one exist, and then presses the “Match”-button. If no matching word exists, the user marks the element as a NULL-link. I*Link uses machine learning techniques to store the choices of the user in dynamic resources that are built during and used directly in the linking process. Thus “the accuracy of the proposed word links is continuously improved during and across word alignment sessions, which in turn means increased efficiency” (ibid., p. 2). This is, however, dependent on the ability of the user to be consistent in his or her chosen links. If the choices are inconsistent, it will harm the learning effect and I*Link will not perform optimally.

In addition to the built-in resources, I*Link can be fed with user-specific dynamic resources. If the user has worked with the tool previously, the resources collected from those sessions can be used as an additional knowledge base for the system, which should enhance the performance of the system. I*Link automatically collects statistical data on the performed translational actions.

The Graphical Interface of I*Link

The graphical interface of I*Link consists of four windows: the Link Panel, the Link Table Panel, the Resource Panel and the Settings Panel. The Link Panel in figure 5.1 is the window in which the current sentence pair is presented, the source sentence in the upper half and the target sentence in the lower half. It is in this window that the user can accept or reject the automatic proposals and select links manually. Chosen links are marked using corresponding colours, and are also shown in the Link Table Panel in figure 5.2. Additions and deletions can be marked as NULL-links by right-clicking with the mouse on the word or words, and choosing NULL.
5.2. A PRESENTATION OF THE TOOLS

Figure 5.1: Screenshot of the Link Panel in I*Link.

Figure 5.2: Screenshot of the Link Table Panel in I*Link.
In the centre of the Link Panel, directly below the windows where the source and target sentences are shown, some important pieces of information are displayed. The box in the middle contains the number of the current sentence pair, in this case number 1258. The green pieces of text on both sides of this box say “Source completed” and “Target completed” when both sentences are fully aligned and the “Done”-button is pressed. This is significant since the advantage of this system is that full and complete alignment can be achieved, and it is thus important to be able to verify that all tokens in each sentence have been aligned before moving on to the next sentence.

The eight fields in the lower left corner of the Link Panel window show linguistic data on the current link on four levels: word form, base form, POS and the function the word or words have in the sentence.

The Resource Panel and the Settings Panel were not used actively in this project. Descriptions of these panels are available in Merkel et al. (2003).

Tools Included in the I*Link System

I*Link also features two possibilities to search the corpus material, in the rather similar tools LinkInspector and LinkReporter. Both can be used to search for, among other things, occurrences of different word classes, constructions, words, aligned pairs and added and deleted elements.

5.2.2 I*Trix

Another word alignment tool that was used in the study is I*Trix, which differs from I*Link by being a tool with which fully automatic alignment can be done. The sample to be aligned is run through I*Trix, which links whatever it can in the sample. The output can then be manually post-edited in I*Link by the user, in order to correct mistakes and achieve a complete alignment where all tokens in the sample are aligned. Like I*Link, I*Trix can be fed with user-specific resources built up in previous sessions using I*Link in order to enhance the performance of the tool.

5.2.3 New Tools, New Possibilities

The big difference between the tools used in this project and traditional word alignment tools is the possibility for interaction between I*Link and I*Trix. Traditional tools tend to include only the automatic part, corresponding to I*Trix. With I*Link, it is possible to align samples manually or semi-manually, thereby creating user-specific resources that in turn can be used to train either I*Link or I*Trix. This training will increase the performance of the tools. Traditional tools are usually not possible to train, meaning that the researcher cannot affect the performance of the tool or the quality of the output.

Using I*Link and I*Trix in combination represents a new way of studying translations, and they comprise a more powerful resource in comparison with traditional corpus tools. However, the fact that this is a new set of tools means
that the old framework for analysis is less useful, which entails that this study differs somewhat from traditional studies.

Traditionally, other measurements were used, such as type-token ratio and lexical density (Baker 1996). The main purpose of these measures is to investigate translation in a broader perspective and to describe general principles that can be found in translations. In contrast, the tools used in this study makes it possible to systematically analyse particular translations in a more powerful way than was possible with traditional tools. Consequently, the methods used in this study are not suitable for investigating translations in general, but are very well suited for making a more thorough investigation of one or more translations.
Chapter 6

The Making of the HP-corpus

In this chapter, the building and aligning of the HP-corpus are described. The four different alignment processes are described and discussed in some detail, so as to explain how the different strategies affect the process.

6.1 The Corpus


<table>
<thead>
<tr>
<th>Book</th>
<th>Sample name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harry Potter and the Philosopher’s Stone</td>
<td>HP1</td>
</tr>
<tr>
<td>Harry Potter and the Chamber of Secrets</td>
<td>HP2</td>
</tr>
<tr>
<td>Harry Potter and the Prisoner of Azkaban</td>
<td>HP3</td>
</tr>
<tr>
<td>Harry Potter and the Goblet of Fire</td>
<td>HP4</td>
</tr>
</tbody>
</table>

Table 6.1: The names of the books the samples are taken from, and the names of the corresponding samples.

Samples of the first 20000 words in each ST were chosen, rounded to the nearest chapter. There were several reasons as to why only whole chapters were used in the samples, the perhaps most important one being that in order to study the translations contrastively, the semantic integrity of the texts needed to be preserved. This was also why the samples all contain the beginnings of
the four books, as it was deemed more difficult to track the translator’s change unless the same part of the different books were being studied. In addition, the beginning and ending of chapters have specific characteristics. The extent of the resulting samples in the number of tokens and chapters included can be seen in table 6.2. The total token count in the corpus is 189116 tokens.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample size ST</th>
<th>Sample size TT</th>
<th>Chapters</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>26121</td>
<td>26642</td>
<td>5</td>
</tr>
<tr>
<td>HP2</td>
<td>24780</td>
<td>25359</td>
<td>5</td>
</tr>
<tr>
<td>HP3</td>
<td>20036</td>
<td>20671</td>
<td>4</td>
</tr>
<tr>
<td>HP4</td>
<td>22546</td>
<td>22961</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 6.2: The respective sizes (in number of tokens) of the samples in the HP-corpus, and the number of chapters in each sample.

The selected samples were transferred to electronic versions by scanning, proofread, and sentence aligned manually using Microsoft Word. This manual process enabled the author to form an initial idea of what translational phenomena might be interesting to investigate further. The production of the sentence aligned corpus required 7 weeks. The corpus contains in total 5816 sentence pairs, and how these are divided over the four samples can be seen in table 6.3.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number of sentence pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>1768</td>
</tr>
<tr>
<td>HP2</td>
<td>1545</td>
</tr>
<tr>
<td>HP3</td>
<td>1233</td>
</tr>
<tr>
<td>HP4</td>
<td>1270</td>
</tr>
</tbody>
</table>

Table 6.3: The number of sentence pairs in the samples.

After sentence aligning the samples, they were part-of-speech-tagged automatically using Machinese Syntax by Connexor.

### 6.2 Word Aligning the Corpus

The POS-tagged data was used as input to I*Link and I*Trix, in which the word alignment took place using different strategies for each sample. All four samples were aligned sequentially, starting with the first sentence of the first chapter, and finishing with the last sentence in the sample. Table 6.4 shows a summary of the different strategies and resources used.

Before starting the alignment of HP1, I spent 4 hours learning how to use I*Link. This was necessary because otherwise, the aligning of HP1 would take an unproportional amount of time compared to the other samples, because too much time would have been devoted to understanding the system and how to use
Table 6.4: The different strategies used in the alignment of the samples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Tool(s) used</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>I*Link</td>
<td>Built-in</td>
</tr>
<tr>
<td>HP2</td>
<td>I<em>Trix, I</em>Link</td>
<td>Built-in</td>
</tr>
<tr>
<td>HP3</td>
<td>I<em>Trix, I</em>Link</td>
<td>Dynamic from HP1 and 2</td>
</tr>
<tr>
<td>HP4</td>
<td>I*Link</td>
<td>Dynamic from HP1, 2 and 3</td>
</tr>
</tbody>
</table>

it. In addition, trying to learn the system before starting to align the samples was, of course, positive both regarding the attempt to be consistent and the overall quality of the chosen links.

### 6.2.1 Aligning HP1

The strategy for the first sample was to semi-manually align HP1 in sequence, using only the built-in resources of I*Link. The system suggested links, which were continuously accepted or rejected. The alignment of HP1 took 18 hours to complete.

When I first started aligning HP1, it became clear that I was not entirely familiar with the system, and I was very preoccupied with trying to be consistent, although that failed somewhat. In retrospect, more time should have been spent familiarising myself with the system.

Unfortunately, two sentence pairs were flawed in HP1, namely 1042 and 1043. The base of the problem is that 1042 was completely deleted, i.e. had no corresponding sentence in the target text, which means that the opposite cell in the sentence aligned corpus was marked \&\&NO\_CORRESPONDENCE&&. This was done in the corresponding empty cell of all completely added and deleted sentences in order to prevent the empty cells from disappearing in the POS-tagging. For some unknown reason, this failed to work in this instance, and in the output from Machinese Syntax the first sentence of 1043 had migrated to 1042. This meant that neither of these sentence pairs were possible to link.

However, the sample includes 1768 sentence pairs, and though it is unfortunate that two sentence pairs were marred, it is in no way catastrophic as they were such a minute part of the sample. Therefore, the fact that the two sentences were left unaligned has been disregarded, and they have not been subtracted from the corpus, neither in the number of sentence pairs, nor in the number of tokens.

It also became evident during the aligning of HP1 that despite the meticulous proof-reading of the scanned texts, some small imperfections prevailed in the samples. These were few and far between, however, and any effect they might have had would have been marginal. They were therefore not corrected, as it was judged that correcting them might simply take too much time in proportion

\(^1\)This number does not include the 3 hours required for the subsequent post-editing of HP1, see section 6.4.
to the marginal effect it would produce. Such imperfections also exist in the other samples, and were treated in the same way.

6.2.2 Aligning HP2

HP2 was automatically aligned using the built-in resources of I*Trix, and the data from this second session was post-edited using I*Link. The alignment of HP2 required 16.5 hours.

The difference in the strategies used for HP1 and HP2 had some interesting implications. When only using the buttons Match, Accept, Reject and Done in I*Link, as in HP1, the links are presented in turn, and with enough pushing of the buttons, either a match can be found or the word is treated as a NULL link. This means that the links get coloured one by one, and it is therefore easy for the annotator to keep track of the segments that are linked together. HP2, however, was aligned automatically using I*Trix, which means that in the post-editing session using I*Link, the links already matched by I*Trix were already coloured. Sometimes, words that stand next to each other in the sentences but are not a part of the same link can have very similar colours. This means that there is a risk of accepting links made by I*Trix that should not be accepted, because the eye does not pick up on the slight difference between the colour of the matched links. To me, this meant that in aligning HP2, I had to be very careful, and quite a few sentence pairs were aligned before I discovered this, and so the sentences that I aligned while still ignorant of this had to be post-edited. This was done in the same session, as soon as it was discovered, and the required time is included in the 16.5 hours it took to align the whole sample.

Another consequence of the colour scheme was that in linking HP2, I learned to use the Link Table Panel. In this window, all linked pairs appear after the Accept-button is pressed, and it is possible to check the links as you go along. In HP1, I did not use this as I did not need it, but for the samples pre-aligned with I*Trix, it became indispensable as it diminished the problem with neighbouring links of almost the same colour.

One idea that occurred to me after aligning the first 350 sentence pairs of HP2 was that if one wanted to use the same guidelines that the heuristics of I*Link and I*Trix are based on, it might have been better to start with the strategy that was used for HP2, i.e. using I*Trix to pre-align HP1. The matches in the system’s output quite often differed from my own choices for matches, and this made me start to doubt my reasons for not using the exact same heuristics as the system. My theory is that using the same guidelines and starting with a pre-aligned text might have ensured the consistency of the chosen links, as it is much easier to just accept what the system suggests and simply correct the mistakes and link what the system has not been able to link. For the unexperienced annotator, this could be used as a way to simplify the process and ensure a greater consistency.
6.2.3 Aligning HP3

The dynamic resources created during the first two sessions were used as additional learning material for I*Trix in the third strategy. Using the created resources from HP1 and HP2 as input data to I*Trix, HP3 was automatically aligned and then revised in I*Link. The third strategy is thus basically the same as the second, apart from the fact that the dynamic user-specific resources were used. Aligning HP3 took 10.5 hours.

As was the case with HP1, some sentence pairs in HP3 are flawed. In pair 19, one sentence on the target side had somehow disappeared already in the scanning, which was unfortunately not discovered in the proofreading of the text. In pair 199, the sentence on the target side is repeated. In pairs 343 and 854 spelling mistakes that were present in the actual printed texts were marked with a [sic!] in the proofreading of the scanned texts, and they were unfortunately not removed before the texts were POS-tagged. Both [sic!]s were marked as NULL links. As with the flawed sentences in HP1, these sentences and tokens were not removed from the corpus, but left disregarded.

Although the dynamic resources were used in this session, I*Trix built-in heuristics still overrode the dynamic resources in some cases. For example, following guideline number two in marking as few words as possible in each link, I chose to link name and name, surname and surname, i.e. Harry was linked with Harry, and Potter with Potter. I*Trix aligns name and surname as one link, Harry Potter with Harry Potter, and this was retained in HP3.

6.2.4 Aligning HP4

The dynamic resources from the first three sessions were fed to I*Link, which was used to semi-manually align HP4 using the Match button in I*Link. In other words, this is the same strategy as for HP1, with the exception that the dynamic resources from the earlier sessions were used. The alignment of HP4 required 16 hours.

No sentences were flawed in HP4, and the used strategy did not pose any significant problems. However, in the process of aligning HP4, I got the impression that there were more differences between the source and target texts in this sample than in the previous ones.

6.3 Comments on the Alignment Process

The time it takes to align different parts of a sample is heavily dependent on the nature of the sentences included in that segment. If there are many short sentences, I*Link needs to do little work and few calculations and the aligning takes little time. If there are longer, more complex sentences and sentence pairs, there is a delay in the system and the aligning takes longer time, simply because there are more possible matches for every token. This is of course also dependent on the platform used, i.e. the capacity of the PC. In aligning the HP-corpus, two different PCs with different capacities were used. To balance
the effect of the PC-capacity, a simple time-test was done on the PCs, and the summary of the time required for aligning the different samples in table 6.5 has been modified to accommodate for that difference.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Tool(s) used</th>
<th>Resources</th>
<th>Required time</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>I*Link</td>
<td>Built-in</td>
<td>18 hours</td>
</tr>
<tr>
<td>HP2</td>
<td>I<em>Trix, I</em>Link</td>
<td>Built-in</td>
<td>16.5 hours</td>
</tr>
<tr>
<td>HP3</td>
<td>I<em>Trix, I</em>Link</td>
<td>Dynamic from HP1, 2</td>
<td>10.5 hours</td>
</tr>
<tr>
<td>HP4</td>
<td>I*Link</td>
<td>Dynamic from HP1, 2</td>
<td>16 hours</td>
</tr>
</tbody>
</table>

Table 6.5: The different strategies and the time required for aligning each sample.

6.3.1 Problems Common to the Samples

During the POS-tagging of the sentence aligned texts using Machinese Syntax, almost all of the quotation marks that surround dialogue on the source side disappeared, while those on the target side remained. This was unfortunate, but as there generally is a very high correspondence between the source and target use of quotation marks, it was deemed irrelevant. Furthermore, quotation marks are of no particular interest in this study, so all instances on both source and target sides were simply subtracted from the token-count.

Another problem was the frequent use of the three periods in sequence ("...") construction on the target side, in most cases corresponding to the use of a dash ("-") on the source side. This construction was also damaged in the POS-tagging, which proved to be a little more problematic as the use of periods is of interest to the focus of this study. The reason behind this problem is, in all likelihood, that when the files are saved in text format, these constructions are not converted properly.

In addition, there are as already mentioned some small problems with imperfections in the form of misspellings etc. that were overlooked in the proofreading and has since prevailed in the samples. All tokens with such minor imperfections have been left in the corpus and disregarded in the analysis, because of the very limited effect they might have on the results in proportion to how much time it would take to correct them.

6.4 Post-editing HP1

When all samples were aligned, it was possible to reflect on the consistency of the chosen links. It was discovered that in the beginning of HP1, the alignment was inconsistent regarding the aligning of names, i.e. how name and surname were treated. The chosen strategy for constructions like for example Harry Potter, was to link Harry in the ST with Harry in the TT, and Potter with Potter in the same manner. This, however, had not been done consistently in
the beginning of HP1, probably due to the fact that a heuristic used in I*Link always suggests that such constructions should be aligned as one link, not two. At the very start of the aligning I was apparently too preoccupied with handling the system to notice that I was not following my own guidelines. Because of this, HP1 was post-edited in order to make the links conform to the patterns used in HP2, 3 and 4. The post-editing session, in which all 1768 sentence pairs were checked and mistakes corrected, required 3 hours.
Chapter 7

Results

The first part of this chapter briefly states that the HP-corpus is a result in itself. The second part describes the translational results of the analysis of the corpus and some complications that occurred during the analysis. I have chosen to give a more detailed description of the analysis because there is no ready-to-use framework for studies like this, as mentioned in the introduction. The third part of the chapter presents the methodological results that concern the tools and strategies used in the project.

7.1 The HP-Corpus

The most tangible result of this study is the HP-corpus in itself. Fully aligned on a word level, this corpus of 189116 tokens in 5816 sentence pairs is ready to be used in other descriptive translation studies.

7.2 Translational Results

The translational results are divided in three main sections. The first section presents the results of the investigation into additions and deletions, and explains the consequences of the somewhat coarse FDG-tagging. Furthermore, one subsection is devoted to showing examples of additions and deletions in order to illustrate the effect these operations have on the actual texts. This is followed by the results of a closer investigation of the last 150 sentence pairs of HP4, giving a description of how additions, deletions and translation universals can be manifested if studied in more detail.

The second section contains the results concerning the translation universals. Manifestations of explicitation, normalisation and simplification are presented, both on a lexical level, and on a more general, pattern-oriented syntactical level.

The third section describes the effect of some translational choices made by the translator. Lexical choice is investigated, and one specific translation difficulty in the Harry Potter domain is used to exemplify a translational choice.
Finally, a way of organising and presenting the data obtained in this study using semantic mirroring is presented.

7.2.1 Additions and Deletions

The starting point for this analysis is the advice to translators from Peter Newmark, cited in section 4.5, to be careful not make unnecessary changes to the text, and “mind particularly your descriptive words: adjectives, adverbs, nouns and verbs of quality” (Newmark, p. 36). Consequently, addition and deletion were studied for verbs, adjectives, adverbs, nouns and pronouns. The last category was included because addition of pronouns can be a sign of explicitation and normalisation.

Analysing Verbs and Adjectives

As already mentioned, one of the most important advantages with I*Link and similar alignment tools is the possibility to distinguish additions and deletions. However, at the start of the analysis process, a problem was discovered concerning the analysis of verbs and adjectives. The basis of the problem is that not only the tokens tagged by Machinese Syntax as verbs, V, and adjectives, A, are verbs and adjectives. The parts of speech tagged EN and ING in the source samples, and AD and NDE in the target samples can have both adjective and verb function, and the functional dependency grammar (FDG) that was used in the POS-tagging does not distinguish between the two different uses.

In English, ING is either the present participle that has an adjectival function, or the gerund form of a verb, as in sentence pair 1503 of HP1: “They were going even deeper now and gathering speed”. EN is the past participle as used in sentence pair 19 of HP1: “Little tyke, chortled Mr Dursley as he left the house”. In Swedish, the problem is manifested in the words tagged AD and NDE. AD is a participle form, such as in brutit in sentence pair 290 of HP1: “Ett låg mullrande ljud hade brutit tystnaden runt dem”. An example of a NDE is mullrande in the same sentence. The corresponding source sentence reads: “A low rumbling sound had broken the silence around them”. Brutit in the equivalent of broken and mullrande corresponds to rumbling.

When analysing the corpus in terms of what has happened to verbs and adjectives, the words marked EN, ING, AD and NDE must in some way be accounted for. One way to do that is to manually go through and analyse every single token with such a mark to ascertain if it has verb or adjective function. However, I opted not to do so, since it would take an unproportional amount of time. Instead, a brief examination of approximately 100 sentence pairs was made, and the result of this was that the majority of these particles functioned as verbs.

Because of the somewhat uncertain classification of these particles, the deleted and added verbs are represented by two columns in tables 7.1 and 7.2. The first verb column, V, contains only the added or deleted tokens tagged V for verb. The second, V2, contains the same added or deleted tokens tagged V, plus the
EN, ING, AD and NDE marked tokens. In table 7.1, V2 thus contains nulled tokens in the TTs marked V, AD and NDE, because all additions are, naturally, made in Swedish and cannot be ING or EN. Consequently, V2 in table 7.2 thus includes nulled tokens from the STs tagged V, EN or ING.

**Results for Additions and Deletions**

The percentage of additions and deletions can be seen in table 7.1 and table 7.2. As is evident in the tables, there is a large and steady increase in the number of both additions and deletions made in the four different samples. The tendency is that HP1 has the lowest numbers of additions and deletions and HP4 the highest numbers, while HP2 and HP3 fall in between. In other words, the numbers grow sequentially and consistently, which is true for all the investigated word classes and both possibilities for analysing verbs.

<table>
<thead>
<tr>
<th>Sample</th>
<th>V1</th>
<th>V2</th>
<th>Adjectives</th>
<th>Adverbs</th>
<th>Nouns</th>
<th>Pronouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>6.2</td>
<td>5.3</td>
<td>6.3</td>
<td>15.7</td>
<td>3.7</td>
<td>10.5</td>
</tr>
<tr>
<td>HP2</td>
<td>9.4</td>
<td>7.8</td>
<td>9.3</td>
<td>23.2</td>
<td>4.8</td>
<td>15.2</td>
</tr>
<tr>
<td>HP3</td>
<td>12.4</td>
<td>10.5</td>
<td>9.6</td>
<td>22.2</td>
<td>6.3</td>
<td>16.2</td>
</tr>
<tr>
<td>HP4</td>
<td>16.0</td>
<td>13.4</td>
<td>16.9</td>
<td>25.3</td>
<td>7.1</td>
<td>20.5</td>
</tr>
</tbody>
</table>

Table 7.1: The percentage of additions in the samples. For each investigated word class, the number shown is in relation to the total amount of words of that word class in the source texts.

<table>
<thead>
<tr>
<th>Sample</th>
<th>V1</th>
<th>V2</th>
<th>Adjectives</th>
<th>Adverbs</th>
<th>Nouns</th>
<th>Pronouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>3.2</td>
<td>2.8</td>
<td>2.1</td>
<td>3.7</td>
<td>1.6</td>
<td>3.1</td>
</tr>
<tr>
<td>HP2</td>
<td>4.9</td>
<td>4.3</td>
<td>4.1</td>
<td>7.7</td>
<td>2.6</td>
<td>5.5</td>
</tr>
<tr>
<td>HP3</td>
<td>5.9</td>
<td>5.3</td>
<td>3.8</td>
<td>7.7</td>
<td>2.7</td>
<td>5.0</td>
</tr>
<tr>
<td>HP4</td>
<td>9.6</td>
<td>8.0</td>
<td>6.9</td>
<td>14.4</td>
<td>4.1</td>
<td>9.3</td>
</tr>
</tbody>
</table>

Table 7.2: The percentage of deletions in the samples. For each investigated word class, the number shown is in relation to the total amount of words of that word class in the source texts.

The results are not homogeneous, however, as there are differences in the distributions of additions and deletions. Figure 7.1 and figure 7.2 are included to show the results in a directly perceptible way. In studying these, it is obvious that there are few deletions for all the word classes in HP1, but for additions, the results have a much greater range for the different word classes.

---

1. The lines in the figures are the straight lines that minimise the sum of the squared distances between each line and its four corresponding dots. The purpose of the lines is to visualise trends in the data, mainly to illustrate the sequential differences between the samples. The underlying data is displayed in tables 7.1 and 7.2.
Despite this difference in distribution of additions and deletions in HP1, the tendency of steady and sequential increase in the percentages from HP1 to HP4 is the same for both additions and deletions. For both addition and deletion, the noun category has the lowest percentage of change, and the adverb category the highest.

The implications of the results evident in the tables and figures is that there have been significant changes made to the texts in terms of additions and deletions, and the tendency of Fries-Gedin to alter elements of text has increased over time.

Additions and Deletions in Context

Although the percentages of additions and deletions clearly show that there are differences between the source and target versions of the Harry Potter books, they do not say anything about the actual effect of the additions and deletions in the texts. As an illustration of what additions and deletions can look like, one example from each investigated word category is shown in table 7.3 for additions and table 7.4 for deletions. These are basic additions and deletions that have no corresponding element in the corresponding sentence at all. This
type of addition or deletion tend to be manifested in small pieces of information that have been added or omitted.

But oftentimes both addition and deletion are present in one sentence pair. In some such cases, there is no semantic correspondence at all between the added and deleted elements, as in pair 797 in HP4. The source sentence reads: “A bag of sweets had spilled out of Fred’s pocket and the contents were now rolling in every direction - big, fat toffees in brightly coloured wrappers”. In the Swedish sentence, *fat* has been omitted, but the Swedish *blanka*, equivalent of *shiny*, has been added to the description of the toffee wrappers: “En godisåse hade trillat ut ur Freds ficka, och innehållet rullade nu åt alla håll - stora kolor i blanka, färgglada papper”.

However, addition and deletion are also present in those sentence pairs where there is an element in the corresponding sentence that vaguely corresponds to the added or deleted element, but is not a direct equivalent. These are instances where a closer translation is possible than the option Fries-Gedin has chosen, meaning that a lexical shift has been made voluntarily. In such sentence pairs, the affected words are marked as deletions in the source sentence and additions in the target sentence. One example of this is sentence pair 1108 in HP4. The

Figure 7.2: Plot showing the percentage of deletions.
<table>
<thead>
<tr>
<th>Word class</th>
<th>Source sentence</th>
<th>Target sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>[Are] All these yours, Arthur?</td>
<td>Är alla de här dina, Arthur?</td>
</tr>
<tr>
<td>Adjective</td>
<td>Who did he know who sent letters by the [regular] postman?</td>
<td>Vem kände han som skickade brev med en vanlig brevbärare?</td>
</tr>
<tr>
<td>Adverb</td>
<td>Empty your pockets [immediately], go on, both of you!</td>
<td>Töm genast fickorna! Sätt igång, båda två!</td>
</tr>
<tr>
<td>Noun</td>
<td>War turned him funny [in the head], if you ask me, said the landlord.</td>
<td>Kriget gjorde honom konstig i huvudet, om ni vill veta vad jag tror, sa krogvärden.</td>
</tr>
<tr>
<td>Pronoun</td>
<td>[They were] Lying there with their eyes wide open!</td>
<td>De lag där med ögonen vidöppna!</td>
</tr>
</tbody>
</table>

Table 7.3: A representation of additions for the different word classes. The added information is shown in bold face in the target sentence column. A gloss translation of the information that has been added is given in the square brackets in the source sentence column.

<table>
<thead>
<tr>
<th>Word class</th>
<th>Source sentence</th>
<th>Target sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>Can you take this to Sirius for me? he said, picking up his letter.</td>
<td>Vill du ta med dig det här till Sirius åt mig? Han tog upp sitt brev.</td>
</tr>
<tr>
<td>Adjective</td>
<td>It took Harry several days to get used to his strange new freedom.</td>
<td>Det tog flera dagar för Harry att vänja sig vid sin nya frihet.</td>
</tr>
<tr>
<td>Adverb</td>
<td>The sky lightened very slowly as they made their way through the village, its inky blackness diluting to deepest blue.</td>
<td>Himlen ljusnade långsamt medan de gick genom byn, den bläcksvarta färgen späddes ut till djupblått.</td>
</tr>
<tr>
<td>Noun</td>
<td>There came the chink of a bottle being put down upon some hard surface, and then the dull scraping noise of a heavy chair being dragged across the floor.</td>
<td>Det kom ett klirr från en flaska som sattes ner på någon hård yta och sedan ett lågt skrapande från en tung stol som släpades över golvet.</td>
</tr>
<tr>
<td>Pronoun</td>
<td>Don’t you lie to me!</td>
<td>Ljug inte för mig!</td>
</tr>
</tbody>
</table>

Table 7.4: A representation of deletions for the different word classes. The deleted information is shown in bold face in the source sentence column.
source sentence reads: “He was wearing what appeared to be a golfing jumper and a very old pair of jeans, slightly too big for him and held up with a thick leather belt”. The target is: “Han var iförd något som såg ut som en golfröja och ett par slitna jeans, som var lite för stora för honom och därför hölls uppe av ett tjockt läderbälte”. Here, very old, which should be mycket gamla in Swedish, has instead become slitna, the equivalent of worn. In this example, it is evident that the Swedish and English constructions are not equivalent as closer translations are possible. However, there is still some degree of correspondence between the deleted and added elements, as the meaning of the Swedish word at least has some kind of semantic relationship with the meaning of the source words. Consequently, this combination of addition and deletion is in fact a lexical shift.

A Close Investigation of Additions and Deletions

In order to search for possible patterns of change that could not be detected with the search tools in I*Link, the last 150 sentence pairs of HP4 were subjected to a closer scrutiny. These sentence pairs were chosen because they are the most recently translated part of the HP-corpus, and were deemed most indicative of what the translations will be like in later parts of Harry Potter and the Goblet of Fire (Rowling 2001a), as well as in later Harry Potter-books not included in the corpus. The close investigation had two main purposes; to try to establish how common it is that additions and deletions in the same sentence pair are related, and to search for translation universals in more detail.

Bearing in mind that combinations of addition and deletion can indicate a lexical shift, the subsample was thoroughly investigated manually, and all additions and deletions, as well as their POS-tags, were recorded for further analysis. Table 7.5 shows the distribution of additions and deletions on a general surface level, i.e. how many sentences contained only an addition or a deletion, both addition and deletion, or neither. The numbers show that if this subsample of HP4 is representative of the translational style of Fries-Gedin, about 45 percent of the sentences in her current translating style have both additions and deletions in them, whereas 26 percent are translated without any addition or deletion being made to them. This cannot, naturally, be generalised to the rest of the corpus, but nevertheless gives an idea of the distribution of change in the translated material.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>18</td>
<td>68</td>
<td>39</td>
</tr>
</tbody>
</table>

Table 7.5: Results of the close investigation of the last 150 sentence pairs of HP4.

However, this surface relationship does not say anything about whether or not there is a relationship between the additions and deletions in the sentence pairs that contain both. It would be reasonable to expect that if there is a
relationship between an addition and a deletion, the words involved will often be of the same word class. A detailed presentation of the distribution of additions and deletions for each word class is given in table 7.6.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>45</td>
<td>32</td>
<td>24</td>
<td>97</td>
</tr>
<tr>
<td>Adjective</td>
<td>19</td>
<td>8</td>
<td>3</td>
<td>126</td>
</tr>
<tr>
<td>Adverb</td>
<td>37</td>
<td>24</td>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td>Noun</td>
<td>26</td>
<td>8</td>
<td>3</td>
<td>119</td>
</tr>
<tr>
<td>Pronoun</td>
<td>43</td>
<td>27</td>
<td>12</td>
<td>92</td>
</tr>
</tbody>
</table>

Table 7.6: Distribution of additions and deletions for the investigated word classes in the subsample.

On the surface, it appears that there is a higher degree of correlation between added and deleted verbs than between additions and deletions in the other word classes. Because of this, the verb category was further investigated. In the sentence pairs where both additions and deletions were made, a general pattern of relationship between the two was indeed observable. In almost all these 24 sentence pairs, there was an obvious relationship between one or more of the added and deleted verbs, but some type of unnecessary change had been made in the translation. In these cases, Fries-Gedin paraphrases the meaning of the source segment or does not use the closest possible translation. One example of this is sentence pair 1161, where the verb *said* in “We told you to destroy them!” said Mrs. Weasley [...]” is deleted as the target sentence instead uses the verb *skrek*, the equivalent of *shouted*. The target reads: “‘Vi sa ju åt er att förstöra dem!’ skrek Mrs. Weasley [...]”. This is a rather typical example of a word pair that are semantically related, but where a much closer translation could have been chosen, wherefore the pair is treated as an addition/deletion pair.

This close investigation also made it possible to investigate manifestations of the translation universals that were not feasible to investigate for the entire corpus, due to the extent of the corpus and the lack of automated possibilities for searches of this kind. Signs of explicitation, as well as patterns of simplification and normalisation were obtained in this way.

Explicitation was manifested in a way that could possibly be a pattern. In English, characters are sometimes affectionately called short forms of their names. Cedric Diggory, a boy Harry knows from Hogwarts, is repeatedly called Ced by his father, as in sentence pair 1241 that reads: “‘Ced’s talked about you, of course’,” said Amos Diggory”. This is translated as “’Cedric har förstås pratat om dig”, sa Amos Diggory”’. In cases like this where the short form of the name is used in the source text, the translation either reads Cedric, or the passage containing the name is deleted in the target text. I argue that this is a way of explicitating the text, but whether or not it is a pattern cannot, of course, be stated without a more thorough investigation of the entire corpus.
7.2. TRANSLATIONAL RESULTS

For simplification, one possible manifestation is that there are oftentimes more sentences in the target text than in the original, which proved to be true for the 150 sentence pair subsample of HP4. The source side consisted of 150 sentences, and the target of 176. This is an increase by 17 percent in the subsample. Although this cannot be generalised to the whole corpus, it is at least an indication that there are indeed more sentences in the target texts.

In translating between English and Swedish, some verb constructions will inevitably lead to differences between the source and target texts, which was apparent in a few cases. For example, when describing something a person does in Swedish, a construction based on stood and, sat and or similar combinations of verb plus and is often used as an equivalent to the gerund form in English in order to signal ongoing action. This is exemplified in the first words of sentence pair 1177, where the source sentence reads: “Harry, having been thinking about thousands of wizards [...].” The target sentence, however, is: “Harry, som gick och tänkte på alla de tusentals trollkarlar [...].” The literal translation of this back to English would be Harry, who was walking and thinking about all the thousands of wizards [...]. The target sentence contains other changes as well, but it still illustrates the point of how a gerund construction can be adapted to fit Swedish usage, which is an instance of normalisation.

7.2.2 Translation Universals

Manifestations of the translation universals were also investigated for the entire corpus.

Explicitation

For the core purpose of this study, explicitation was expected to be manifested primarily in two ways. The first was that if there were more additions than deletions in the samples, this could be seen as an indication of explicitation. Looking at the combined effect of table 7.1 and table 7.2, there is indeed more added than deleted information in the samples. However, addition of nouns is, as mentioned in section 3.4.1, considered to be a typical manifestation of explicitation. In the HP-corpus, the number of additions is lower for the noun category than for any of the other categories of words. This fact is an indication that explicitation is not so strongly manifested in the HP-corpus, at least not for the traditional type of lexical explicitation by the addition of nouns.

Notwithstanding this, the second expected manifestation of explicitation in the corpus follows logically from the first, in that the translated texts were likely to be longer than the original texts. As early as during the sentence alignment, it became clear that the samples seemed to conform to the tendency of translations to be longer than their originals. As can be seen in table 7.7, there are more tokens in all the TT-samples, compared to their STs. This increase, although consistent, struck me as being smaller than expected, as the impression during the sentence alignment was that the Swedish texts were noticeably longer. A
possible explanation for this will be given in section 8.1.3 in the discussion chapter.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample size ST</th>
<th>Sample size TT</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>26121</td>
<td>26642</td>
<td>+ 521</td>
</tr>
<tr>
<td>HP2</td>
<td>24780</td>
<td>25359</td>
<td>+ 579</td>
</tr>
<tr>
<td>HP3</td>
<td>20036</td>
<td>20671</td>
<td>+ 635</td>
</tr>
<tr>
<td>HP4</td>
<td>22546</td>
<td>22961</td>
<td>+ 415</td>
</tr>
</tbody>
</table>

Table 7.7: The respective sizes (in number of tokens) of the samples in the HP-corpus, and the difference in number of tokens.

Another way of explicitating translations is to add lexical elements that serve to explain events and notions in the source texts. Although this type of manifestation of explicitation is not in focus in this study, two examples from the corpus are brought up nevertheless. The point of doing so is partly to show that there are indeed examples of lexical explicitation in the corpus, and partly to give the reader a better grasp of what such an addition might look like.

The first example is in sentence pair 1206 in HP4, which reads: “Hermione came over the crest of the hill last, clutching a stitch in her side”. This has been translated as (added information in italics): “Hermione kom upp sist över branten. Hon höll sig i sidan, där hon hade fått håll av den ansträngande klättringen”. A gloss translation of this would be something like: Hermione came up last over the steep. She was holding her side, where she had gotten a stitch from the strenuous climb. This is a clear example of explicitation, as the fact that the stitch was caused by the strenuous climb has been added in the Swedish text.

Another example of explicitation is sentence pair 588 in HP4. The source sentence reads: “He had never seen anything that looked less like a pig”. The “anything” refers to Harry’s friend Ron’s owl, which Ron calls both Pig and Piggy. To explain the semantic meaning in English of the name Pig and its relation to the animal pig, Fries-Gedin has added information. The target sentence reads “Piggy, det var ju en liten gris, och han hade aldrig sett någonting som mindre liknade en gris”. The first part of this sentence, Piggy, det var ju en liten gris, och is an addition roughly corresponding to Piggy, that was a small pig, and. Apart from these two examples, there are many other instances of lexical explicitation in the corpus, but these sentence pairs suffice to illustrate that lexical explicitation is manifested, as well as some of its possible effects on a text.

Simplification

One of the possible manifestations of simplification investigated for the HP-corpus is the punctuation, and whether or not it has been strengthened. As can be seen in table 7.8, there is indeed evidence of strengthened punctuation in the
samples. Commas and semicolons have been changed to full stops, and significantly, there is a change over time concerning the strengthening of punctuation markers. The implication of this is that the texts have become more simplified from HP1 to HP4.

<table>
<thead>
<tr>
<th>Sample</th>
<th>, → .</th>
<th>; → .</th>
<th>; → ,</th>
<th>; → ;</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>HP2</td>
<td>44</td>
<td>17</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>HP3</td>
<td>95</td>
<td>13</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>HP4</td>
<td>88</td>
<td>31</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 7.8: The changes in punctuation in the samples.

In a few instances, other punctuation markers than commas have been strengthened, but in proportion to the total number of punctuation markers of each kind, commas dominated among the strengthened elements. The treatment of semicolons will be further discussed in the normalisation section below.

The complexity of the texts has also been decreased by the way Fries-Gedin has handled the regional dialects of the characters Rubeus Hagrid and Stan Shumpike (see table 7.9 for examples of the dialects). That the dialects were more prominent in the source texts became evident already during the word alignment. In order to investigate if this impression was correct, a small-scale case study was performed on subsamples of the corpus where the two characters make utterances. The case study had to be focused on different samples for the two characters, as there is only dialogue from Stan Shumpike in HP3, and Hagrid makes utterances in only HP1 and 2.

The case study revealed that Fries-Gedin has treated these two very different accents in the same way, to all intents and purposes. The dialogue of both Hagrid and Stan is translated using simple markers, i.e. spelling the words as they sound when spoken. Both characters, in Swedish, say å instead of och for the source use of an (which would be and in standard English spelling), mej and dej instead of mig and dig for me and you, and e instead of är for is. These expressions, that are used profusely for Hagrid and Stan, are not used for any of the other characters (except for å, which can also be an interjection corresponding to the English ah, or oh). The simple markers are the only elements that separate the utterances of Hagrid and Stan from the rest of the text in the TTs, and thus these two characters seem to speak in the same manner in Swedish. Furthermore, there is only very little contrast in dialect between Hagrid and Stan and the rest of the characters in Swedish, which also simplifies the dialogue.

Thus simplification is indeed manifested, but the treatment of dialects does not only mean that the texts have been simplified, but also that they have been normalised. Removing coarse dialects in dialogue adapts the text to a Swedish audience because in written Swedish, it is very uncommon for authors to use regional dialects. Further implications of the use of simple markers will be dealt
with in section 8.1.3.

In addition, there are indications of other manifestations of simplification in the HP-corpus. Long sentences have in some cases been divided into several shorter ones, as was revealed in the close investigation of the 150 last sentence pairs of HP4. In other instances, information that increases the complexity of a sentence has been removed. However, these manifestations have not been investigated for the full corpus in any structured way, due to the extent of the corpus and the lack of tools suitable for studying these aspects.

Normalisation

As mentioned above, normalisation is manifested in the way the translator has chosen to treat the dialects of Rubeus Hagrid and Stan Shupike. In addition to this, two other kinds of manifestations of normalisation have been found in the HP-corpus.

<table>
<thead>
<tr>
<th>Character</th>
<th>Source utterance</th>
<th>Target utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hagrid</td>
<td>'Best be off, Harry, lots ter do today, got ta get up ter London an buy all yer stuff fer school.'</td>
<td>“Bäst å ge oss iväg, Harry, massor å göra idag, måste fara opp till London å köpa alla dina saker till skolan.”</td>
</tr>
<tr>
<td>Hagrid</td>
<td>'Suppose the myst'ry is why You-Know-Who never tried to get 'em on his side before...probably knew they were too close ter Dumbledore ter want anythin' ter do with the Dark Side.'</td>
<td>“De konstiga e väl varför Du-Vet-Vem aldrig försökte få över dom på sin sida tidigare visste nog att dom stog för nära Dumbledore för å vilja ha nånting med Den Mörka Sidan att göra.”</td>
</tr>
<tr>
<td>Stan</td>
<td>'Very close to You-Know-Oo, they say anyway, when little Arry Potter put paid to You-Know-Oo' - Harry nervously flattened his fringe down again - 'all You-Know-Oo's supporters was tracked down, was n’t they, Ern?'</td>
<td>“Han stod vildigt nära Du-vet-vem, säjer dom. Men den där gången då lille Harry Potter gav Du-vet-vem på nöten så” Harry slätade nervöst till luggen över pannan igen “spåra dom opp alla Du-vet-vems anhängare, visst va de så, Ern?”</td>
</tr>
<tr>
<td>Stan</td>
<td>'Eleven Sickles, said Stan, but for firteen you get ot chocolate, and for fifteen you get an ot water bottle an a toothbrush in the colour of your choice.'</td>
<td>“Elva siklar”, sade Stan, “men för fjorton fåru varm choklad å för femton fåru en varmvattensflaska å en tandborste i vicken färg du vill ha.”</td>
</tr>
</tbody>
</table>

Table 7.9: The dialects in English and Swedish.
Firstly, there are sentences in the corpus that have been made more grammatically correct in the translations than they were in the source texts. One example of this is sentence pair 1225 in HP4. The original sentence reads: “‘Long walk, Arthur?’ Cedric’s father asked”. The Swedish sentence has been grammaticised by completing the sentence. The Swedish equivalent of *did you have a* has been added before *long walk*, as in “*Hade ni långt att gå, Arthur?*” frågade Cedrics far.” Another example of normalisation through grammaticising an ungrammatical utterance is in the top row of table 7.3, where the verb has been added, making the Swedish sentence complete.

Secondly, normalisation can also be manifested in the translation of punctuation markers. Translators tend to adapt the usage of punctuation markers to fit better with the target language usage, and evidence of this has been found in the HP-corpus. Particularly interesting is the treatment of semicolons in the translations. Semicolons are not very common in original Swedish texts, especially not in children’s literature, but as can be seen in table 7.8, many semicolons are kept in the target versions of the HP-samples. In comparing the numbers for the respective samples, it is evident also that there has been a change over time in the treatment of semicolons.

In HP1, 28 semicolons have been retained, and very few other changes have been made to this particular punctuation marker. In HP4, only 13 semicolons have been retained. Moreover, 31 semicolons have been changed into full stops and 12 into commas. For HP1, the corresponding figures are much lower, as no semicolon has been changed into a full stop, and only one semicolon has become a comma.

Also concerning commas there are indications of normalisation. As is evident in table 7.10, many commas have been omitted in the target texts, which is also an indication of normalisation because commas are used much more frequently in English, compared to Swedish. The conclusion I draw from this is that in the usage of syntactic markers, the texts are normalised through adapted punctuation, and the tendency for the translator to normalise the texts in this way has increased over time, at least regarding semicolons.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Nr. of commas in ST</th>
<th>Deleted commas</th>
<th>Added commas</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>1504</td>
<td>404</td>
<td>57</td>
</tr>
<tr>
<td>HP2</td>
<td>1639</td>
<td>519</td>
<td>83</td>
</tr>
<tr>
<td>HP3</td>
<td>1331</td>
<td>439</td>
<td>66</td>
</tr>
<tr>
<td>HP4</td>
<td>1461</td>
<td>528</td>
<td>170</td>
</tr>
</tbody>
</table>

Table 7.10: The changes in commas in the samples.

7.2.3 Investigating Translational Choices

In order to show that tools like I*Link have enormous potential in the investigation of lexical choices in the corpus, some examples of different translational
choices and situations are presented below. These examples are also meant to illustrate that the choice of translator does have a real effect on the produced translation, an effect that is manifested in the particular translational choices of that translator.

**Lexical Choice**

During the manual word alignment, I noticed that Fries-Gedin used two alternative translations for *wand*, namely *trollstav* and *trollspö*. Fries-Gedin has opted for using *trollstav* when the carrier of the wand is male, and *trollspö* when the carrier is female (see examples from HP2 in table 7.11).

<table>
<thead>
<tr>
<th>Source sentence</th>
<th>Target sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ron let go of the steering wheel completely and pulled his <em>wand</em> out of his back pocket.</td>
<td>Ron släppte ratten helt och hållet och drog fram sin <em>trollstav</em> ur bakfickan.</td>
</tr>
<tr>
<td>She was wearing a flowered apron with a <em>wand</em> sticking out of the pocket.</td>
<td>Hon var iförd ett blommigt förkläde och ur en av fickorna stack ett <em>trollspö</em> upp.</td>
</tr>
</tbody>
</table>

Table 7.11: A representation of the translator’s choices in translating *wand*.

There is no obvious explanation as to why Fries-Gedin has done this, as *wand* is gender-neutral in English, as is both *trollstav* and *trollspö* in Swedish. Moreover, *trollspö* is a compound in Swedish, and the second part, *spö* is an instrument used mainly by fairies, not witches and wizards. In Swedish a *spö* is quite different from a *stav*, and the difference is not based on the gender of the carrier, but rather on the context in which the instrument is used. Possibly, it could be argued that a *stav* somehow gives the impression of being longer and more sturdy than a *spö*, which in general corresponds to the properties that Rowling describes in the wands women and men use. However, since *spö* and *stav* are generally not used in the same context, the difference between them is emphasised. The consequence is, perhaps, that a Swedish reader interprets them as being different instruments, which they are not supposed to be.

Moreover, in the political climate of today, the act of gender-differentiating when it is not absolutely necessary is somewhat dubious, especially since children’s literature plays an important role in shaping children’s social and cultural identity (Puurtinen 1998). Interestingly, this gender-differentiating is present in the translation of the first two books, but not in the fourth. *Wand* is not mentioned in the third sample, so unfortunately no comparison can be made with the third book. Whether using *trollstav* consistently in HP4 is a conscious choice by Fries-Gedin or just inconsistency is, of course, impossible to say just by investigating the texts.

Apart from *wand*, there are other examples of expressions specific to the Harry Potter books where Fries-Gedin has changed her translation in the later
parts of the series. *The cupboard under the stairs* is a well-known concept to any Harry Potter reader, as it is what functions as Harry’s room in the Dursley house in the beginning of the series. Later on, it is where Harry’s Hogwarts things are kept when he is home for the holidays. *The cupboard under the stairs* is not translated consistently throughout the corpus. The construction can be found in all four samples, and in HP1 the full construction is translated as *skrymslet under trappan*. In other instances, where the source consists of only *cupboard*, the Swedish translation is *krypin*. Both *skrymsle* and *krypin* are in some cases modified with the adjective *trånga*, denoting narrow in English. In HP2, HP3 and HP4, the full construction is translated as *skrubben under trappan*, and shorter versions as *skrubben*. Why this change has been made is, naturally, impossible to say without asking the translator, but it is an indication that she is not averse to change, if it is called for. In this case, I argue that it is a change for the better, as *skrymsle* denotes a very small and narrow space, generally impossible to close off with a door, corresponding more closely to the English *nook* than to *cupboard*. *Skrubb*, on the other hand, is a more likely description, since it denotes a rather small, closed-off space, but still giving the impression of being large enough to hold an eleven year old boy.

**A Translation Difficulty in the Harry Potter World**

The fact that the very specific world of the Harry Potter books and its associated neologisms causes some problems to translators becomes obvious when the translation for the neologist use of *apparition* is studied. Rowling uses *Apparition* with a completely new meaning, to describe the way in which wizards can teleport themselves to another location instantly. Also the verbs used in connection with this activity are neologisms. To *Apparate* is to appear in the new location by Apparition and to *Disapparate* is to disappear by Apparition. The difficulties in translating this activity is illustrated in sentence pairs 1188 and 1189 of HP4 in table 7.12.

<table>
<thead>
<tr>
<th>Sentence pair</th>
<th>Source text</th>
<th>Target text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1188</td>
<td>Some <em>Apparate</em>, of course, but we have to set up safe points for them to appear, well away from Muggles.</td>
<td>Nägra använd sig förstås av <em>spöktransferens</em>, men vi måste välja ut säkra ankomstställen som de kan dyka upp på, väl dolda för mugglarna.</td>
</tr>
<tr>
<td>1189</td>
<td>I believe there’s a handy wood they’re using as the <em>Apparition point</em>.</td>
<td>Jag tror att de har hittat en lämplig skog för ändamålet.</td>
</tr>
</tbody>
</table>

Table 7.12: Examples of the translation of Apparate and related constructions.

Fries-Gedin has chosen the neologist compound noun *spöktransferens* as an
equivalent of both *Apparition* and the infinitives of the associated verbs, which has some interesting implications. The first half, *spök-* is derived from the Swedish word for *ghost*, *spöke*, which is very close in meaning to the original meaning of apparition. *Transferens* is a neologism that Fries-Gedin has probably built on the word *transferering*, denoting a transfer of some resource, usually money. *Spöktransferens* works as an equivalent of the infinitives of the verbs to *Apparate/Disapparate*, but it does not work in the active sense, when somebody Apparates, or Disapparates. The chosen translation for the active senses is *använda sig av spöktransferens*, the equivalent of *to use ghost transferal*, which is a cumbersome construction. In table 7.12, sentence pair 1188 shows the relationship between *Apparition* and *spöktransferens*. In 1189 the difficulties in translating constructions containing *Apparition* is illustrated. Fries-Gedin has in this case chosen to paraphrase and simplify by changing the *Apparition point* to the Swedish equivalent of *for the purpose*.

**Lexical Patterns**

From the resources built in the alignment, it is possible to create alphabetical lists of the words and their translations. By investigating these lists, lexical patterns of how words are translated can be discovered. In instances where words have many and diverse translations, this is generally a sign that they have been difficult to translate. Because the HP-books portray a complex, magical environment with quite detailed vocabulary, it would be reasonable to expect that words specific to this world might have been especially challenging for the translator. Therefore, a closer investigation was made into the translation of vocabulary typical to this domain.

Whether a person is wizard or Muggle is paramount in the Harry Potter world. *Muggle* is consistently translated with the neologism *mugglare*. In the cases where it is a part of a longer noun construction, such as *Muggle clothes*, this, equally consistently, becomes a compound noun in Swedish, *mugglarklädare*, with the stem *mugglar-*. Similarly, *wizard* is translated as *trollkarl* in the absolute majority of the cases. When it is a part of a longer noun construction, such as in *the wizarding bank*, this becomes a compound noun in Swedish with the stem *trollkarls-*, in this particular case *trollkarlsbanken*. One rare exception is *wizard gold*, translated as *trollmynt*, which changes the meaning of the word, since *troll* in Swedish means exactly what *troll* does in English. The second part of the word is also changed, as *gold* is translated into the equivalent of *coin*.

Patterns of consistency, as well as patterns of inconsistency, become apparent when investigating the resources in this way. For example, noun constructions with *magic* are oftentimes translated into a compound with *trollkarls-* as the stem, i.e. the same stem as used for translating compounds containing forms of *wizard*. The translation equivalent of *magic* is *magi*, but the word has eleven different translations (see table 7.13). *Magical*, however, has fewer translations, but they are built both around the *magi* and the *troll* stems. Judging from the amount of translation equivalents for words about wizards and magic, the
magical element of the Harry Potter world seems to have caused a problem in the translation process.

An interesting lexical pattern of inconsistency apparent in the HP-corpus is the translation of *the Dursleys*. Perhaps surprisingly, it has ten different translations ranging from similar constructions equivalent to *the Dursley couple* and *the Dursley spouses* to equivalents of *his uncle or aunt, them and the others* (see table 7.13). This illustrates a difference between English and Swedish; in Swedish a plural form of a family name is not used as consistently to describe the unit of that family as it is in English, which could account for the many different translation alternatives.

<table>
<thead>
<tr>
<th>Source word</th>
<th>Target translations</th>
</tr>
</thead>
<tbody>
<tr>
<td>magic</td>
<td>magi, magiska, trolla, trolldom, trolldomskraft, trolleri, trollkarlar, trollkarlsvärlden, trollkonst, trollkonster, trollkunskap</td>
</tr>
<tr>
<td>magical</td>
<td>förtrollande, magisk, magiska, magiskt</td>
</tr>
<tr>
<td>the Dursleys</td>
<td>de, de andra, Dursleys, familjen Dursley, familjen Dursleys, hans morbror eller moster, makarna Dursleys, mr och mrs Dursley, paret Dursley, paret Dursleys</td>
</tr>
</tbody>
</table>

Table 7.13: The patterns of translations for certain typical Harry Potter related words.

**Semantic Mirroring**

The resources built during the word alignment can also be used to create more powerful resources than the alphabetical lists, such as semantic mirrors. With semantic mirrors, it is possible to extract information about the translations that is not available in I*Link in itself.

Semantic mirroring of the resources built up in the HP-project was made by Helge Dyvik at the University of Bergen, Norway, and this resulted in two additional means for studying the material. One is a thesaurus-like file that shows all the words in the corpus that have many different translation alternatives, and does not contain words with few translation alternatives. As mentioned above, a large number of translation alternatives for a certain word indicates that it has been difficult to translate, as there is not only a few possibilities for equivalence. Thus the thesaurus shows words that have been translated inconsistently. The other is a search tool that makes it possible to search for specific words in the thesaurus. For a full description of semantic mirrors, see Dyvik (2003).
The semantic mirrors do not show direct correspondences, that is source word and its translation or translations in the corpus. Instead, they attempt to regroup the data so that semantic information such as hyperonyms and synonyms can be extracted from the corpus. These indirect correspondences can be very interesting to investigate closer if the object is to study language use in a particular semantic field, such as the fantasy genre of literature. One possible use of the thesaurus-listing is that it functions as a Harry Potter dictionary, and could be paired with other resources from other studies in order to create a genre dictionary for fantasy books.

The semantic mirrors are also practical tools for obtaining an overview of the words in the corpus, as they make the material more accessible and easier for a layman to read than the output from I*Link. A listing in the thesaurus can look like this, with the listed word in bold face, its translation or translations, synonym or synonyms, and where applicable, related words:

grinning
(Translation: log. )
Synonyms: grinned—1—
gripped
(Translation: tog. )
Synonyms: withdrew.
grow
(Translation: blev. )
Synonyms: became.
growled
(Translation: röt, brummade. )
Synonyms: roared—1—, snarled—1—
Related words: barked, bellowed.
grudgingly
(Translation: motvilligt. )
Synonyms: resentfully.

For the particular purpose of this study, the semantic thesaurus and search tool were used to some extent in the investigation of lexical choice. However, the main contribution of these two resources are perhaps to simplify searching the material and browsing the thesaurus for those with a particular interest, be it in semantic relationships of translations or in the Harry Potter books.

7.3 Methodological Results
The methodological analysis focuses on the different strategies used in the word alignment.
7.3. METHODOLOGICAL RESULTS

7.3.1 Evaluation of the Different Strategies

The four different strategies that were used for the four samples constitute the basis of a brief evaluation of I*Link and the strategies themselves. A summary of the strategies can be seen in table 7.14.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Tool(s) used</th>
<th>Resources</th>
<th>Required time</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>I*Link</td>
<td>Built-in</td>
<td>18 hours</td>
</tr>
<tr>
<td>HP2</td>
<td>I<em>Trix, I</em>Link</td>
<td>Built-in</td>
<td>16.5 hours</td>
</tr>
<tr>
<td>HP3</td>
<td>I<em>Trix, I</em>Link</td>
<td>Dynamic from HP1 and 2</td>
<td>10.5 hours</td>
</tr>
<tr>
<td>HP4</td>
<td>I*Link</td>
<td>Dynamic from HP1, 2 and 3</td>
<td>16 hours</td>
</tr>
</tbody>
</table>

Table 7.14: The different strategies used in the alignment of the samples.

The basis for the evaluation of the strategies presented above is their efficiency, measured in the time it took to align each sample. This is because manual aligning is, as already mentioned, despite its advantages, very time-consuming. Consequently, it is relevant to consider if any one strategy decreases the time required by the aligning more than the others. As is evident in table 7.14, HP3 required the least time to align, and HP1 the most. Running a sample through I*Trix takes only a few minutes, so this time can be disregarded in the comparison between the different strategies.

One of the two most obvious explanations to why HP1 required so much time is that it was the largest sample in terms of number of tokens. The other explanation is the fact that it was the first sample to be aligned, and a certain amount of the time was spent dealing with insecurities in using I*Link and trying to maintain consistency.

However, in investigating the efficiency of the strategies, it is of course pivotal to take the exact sizes of the samples, i.e. the token-count, into consideration. In table 7.15 below, such a comparison is made.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Seconds/sentence pair</th>
<th>Seconds/token</th>
<th>Words/sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>36.6</td>
<td>2.46</td>
<td>14.77</td>
</tr>
<tr>
<td>HP2</td>
<td>38.4</td>
<td>2.40</td>
<td>16.04</td>
</tr>
<tr>
<td>HP3</td>
<td>30.6</td>
<td>1.86</td>
<td>16.25</td>
</tr>
<tr>
<td>HP4</td>
<td>45.6</td>
<td>2.58</td>
<td>17.75</td>
</tr>
</tbody>
</table>

Table 7.15: The efficiency of the different strategies, in relation to sample size. The words/sentence count is the mean number of words per sentence on the source side.

The fact that there is such a great difference between the time required to align HP3 in comparison with the other samples is of course the most relevant finding of the strategy evaluation. The implication of the results is that it appears to be most efficient to align a subsample or a part of the corpus, use the
dynamic resources from that session to automatically align the next subsample in I*Trix, and then revise those results in I*Link.
Chapter 8

Discussion

First of all, I would like to point out that this thesis is in no way intended to be a value-judgement of the translations or the translator. As Newmark points out (1988), translations must be discussed as they are always made subjectively, and this thesis is a mere discussion of the translation of Harry Potter, also made from a subjective viewpoint.

That being said, the results unanimously indicate that there are indeed significant changes between the translations in relation to their respective originals, and that these changes increase sequentially. Additionally, the universals of translation are manifested in the corpus. Moreover, the different strategies used in the alignment process gave different results concerning their efficiency, and in summary, the third strategy seems to be the most efficient, at least in the case of this annotator. In other words, all the hypotheses stated in the introduction have been verified by the results, which is highly encouraging.

Following the format of the last chapter, the discussion will be divided into three main sections. The first section will discuss the translational results. The second will deal with the methodological results, and will also include some additional discussion about my experiences of I*Link and using this set of tools for projects of this kind. The third and final section contains suggestions for further research.

8.1 Discussion on the Translational Results

8.1.1 FDG Imperfections

Apart from the difficulties in obtaining exact word class labels for all lexical tokens in the samples, the functional dependency grammar standard that was used for this study has unfortunately flawed the material in other ways. As already described, the construction where three dots (...) indicate a pause in a character’s dialogue, disappeared in the tagging process. This caused delays in the aligning process, and made it impossible to fully study punctuation markers to the intended extent.
In the final hours of analysis carried out, another FDG-related problem was discovered. In sentence pair 1125 in HP4, the Swedish råkade, a verb equivalent to happened to, has not, as expected, been tagged V, but instead A, for adjective. Due to this discovery, all word class tags for all tokens of the last 150 sentence pairs of HP4 used in the close investigation of additions and deletions were checked manually. It was found that for both the source and the target sides, less than 2 percent of the tokens carried a faulty tag. If this is indicative for the whole HP-corpus, the FDG precision rate is approximately 98 percent. Based on this, I conclude that the few flaws concerning FDG-tags present in the HP-corpus are not likely to have affected the reliability of the results to any great extent.

8.1.2 The Relationship between Additions, Deletions and Lexical Shifts

Because the focus of this study is in part on the influence of the translator on the translated text, I have tried to distinguish between small and significant changes, which ties into the concepts of lexical shifts, additions and deletions.

In my definition of necessary lexical shifts, it is only natural that they have no closer or more accurate translations. In other words, these can be treated as regular translational equivalents, because to all intents and purposes, they are, as long as the meaning of the source words could not have been more preserved in any other target construction.

Concerning the nature of unnecessary, or voluntary, changes made to the texts, the close investigation revealed that at least regarding verbs, many additions and deletions are used in combination, as pairs. They are in fact lexical shifts or parts of paraphrases, not regular additions and deletions.

Because of the fact that significant lexical shifts seem to be rather common, I feel that it would be positive to be able to distinguish them from regular additions and deletions, at least if the focus is on the degree of change. In my opinion, this could be done by making it possible to mark lexical shifts in alignment programs such as I*Link. In the analysis, all types of lexical shifts could then be treated as indicators of change in the translation, and other lexical shifts could be analysed as having a similar effect to additions and deletions.

Implementing lexical shifts in I*Link would enrich the analysis of the corpus as an even more fine-grained analysis could be made into the nature of the changes the translator has made to the target text. Especially, more specific lexical shifts could be seen as clear indicators of explicitation. In addition, the changes that are simply more free in relation to the source text than necessary could be distinguished, and this could in turn be used to measure how free the translation is in relation to the original text. One possible disadvantage could be that it might make the aligning more time-consuming due to the time that would be spent distinguishing the different kinds of lexical shifts.
Lexical Shifts and Translation Universals

My strategy for aligning lexical shifts, as described in section 4.6, has some implications for the results on translation universals.

Regarding more specific lexical shifts, different types are aligned differently, some as regular links and some as addition and deletion pairs. This makes it more difficult to search for these manifestations in the HP-corpus. Explicitated references as manifestations of explicitation are to some extent hidden among the regular links and cannot easily be counted. Equally, less specific lexical shifts can indicate simplification, but they are also difficult to find, as they are treated as additions and deletions.

However, the intent in this study was never to be able to count the number of specific manifestations, as it is not feasible in the current version of I*Link. The goal concerning translation universals was to search for manifestations of the three different types, and such manifestations have indeed been found.

8.1.3 Translation Universals

Explicitation

The fact that there are much more additions than deletions in the target texts imply that they have been explicitated, and that they have been more explicitated over time. However, the reasons to why there are more additions and deletions after time can only be hypothesised about here. Perhaps the translator “knows” the characters better after translating many hundreds of thousands of words written by Rowling; she is more familiar with them and therefore interprets the texts in ways she did not in the beginning, reading between the lines and adding things that perhaps are implicit in the source texts, making them explicit in the target. In some lexical choices, it is possible to see her interpretations.

Another manifestation of explicitation was evident in that the target texts contained more tokens than the original texts. The difference in number of tokens was smaller than I expected, however, and one reason for this can possibly be the difference in how nouns are constructed in English and Swedish. In Swedish, one single compound noun is used as the equivalent of a string of nouns in English, which of course affects the token count as one of these tokens in Swedish can be the translational equivalent of two or more tokens in English. On the other hand, the reverse is true for Swedish equivalents of English verb constructions of the type was doing and was thinking etcetera, which become one token in Swedish, in this case gjorde and tänkte. This study does not include any further investigation into how these convergences and divergences due to differences between source and target languages have affected the token count, but it is reasonable to assume that they have indeed had an effect.

In addition to this, changes in the use of commas between the source and target texts can have contributed to the fact that the target texts were perceived as being longer in comparison to the source texts than the token count implies. Many commas have been removed from the ST’s and few added to the TT’s (as
is obvious in table 7.10). Each comma is counted as one token, just like each word is counted as one token. This is of course true for all punctuation markers.

In some instances, there might be a relationship between the added and deleted commas, like between added and deleted words. Some of the added commas could be replacements for deleted ones. If commas were moved within the same sentence or sentence pair, I tried to be consistent in linking them as each other’s equivalent, notwithstanding the fact that the comma was moved.

If the number of added commas is subtracted from the number of deleted commas, the difference equals the number of tokens that were commas in the source text, but are not commas in the target text. This way of analysing the use of commas using the data in table 7.10 indicates that even if there are only 521 more tokens in the TT of HP1, there might, in effect, be many more words, since 404 commas have been removed and only 57 added. In other words, the target texts might have been more explicitated by containing more words than a simple comparison of token counts between the samples reveals.

Normalisation and Simplification in the Treatment of Dialects

Though the case study on the dialects of Hagrid and Stan Shumpike reveals that they have been normalised and simplified, it can be argued that Fries-Gedin has tried to retain part of the effect of the dialects. In marking the speech of Hagrid and Stan with unconventional spelling, she still signals something of a dialect in these characters. However, the fact that they are two very different dialects has not been retained, which means that the texts have been simplified. In addition, keeping in mind that the Harry Potter series are children’s books, a substantial part of the intended audience will perhaps not actually read the text, but get it read to them, by adults. For this group, any effect of the simple markers will most likely be completely lost, as there will be no audible difference between the characters with dialect and those without, unless of course the person that does the reading picks up on the simple markers and in some imaginative way acts out the perceived difference.

8.1.4 The Development of the Translator

The fact that there are such large differences between the number of additions and deletions for the target texts of HP1 and HP4 appears to be a very clear sign of development in the translating style of the translator. In addition to this, the treatment of punctuation markers also indicates that the translating style has changed from HP1 to HP4. However, there is a serious threat to the validity of the results attained in this study, concerning the sequential difference between the samples.

8.1.5 Sources of Error for the Translational Results

One possible source of error is that the clear indication of change between HP1 and HP4 can have two different explanations. It could either be that there
is an actual difference between the samples, caused in the process of translation. However, the difference could also possibly depend not on the samples in themselves, but on the aligning and the annotator. Unintentionally, I may have changed my way of linking between the samples, thereby causing the sequential effects myself. However, that I did post-edit HP1 after finishing HP4, and only found it necessary to correct a few links, indicates that perhaps I was rather consistent and did not change my way of choosing links significantly during the project.

Trying to maximise the use of the built-in resources meant that the samples had to be aligned in sequence in the HP-project. Another reason for doing so was to, with I*Link, mimic the process of the translator. However, taking into account what is said above about possible sequence effects, it would have been better, perhaps, not to align the samples sequentially. One idea for future studies of this nature is to divide each sample into subsamples, and randomising the order in which they are aligned, to prevent sequence effects in the alignment process from influencing the results.

8.2 Discussion on Tools and Methodological Results

8.2.1 Using the Alignment Tools

Word alignment with software tools such as I*Link as a method for studying translations has many benefits compared to manual inspection and other low-tech methods. The fact that the texts are POS-tagged is absolutely pivotal, as it provides the researcher with so much readily available information which would otherwise have been very cumbersome to extract. The material becomes searchable, and most words and constructions can be closely examined from a linguistic perspective.

Notwithstanding this, I*Link, like alignment systems in general, needs to be developed further. One general issue traces back to Borin’s point I brought up in section 4.1, that the annotator is in the hands of the tools (2002). This is a valid point, both regarding the alignment and the subsequent analysis. Especially the analysis does become heavily reliant on what the tools allow for and simplify. In my case, the analysis of the results was very cumbersome, mainly perhaps because of the lack of a framework for this type of study. The results, however, are encouraging enough to motivate further research of this type.

8.2.2 Advantages and Disadvantages of Using I*Link

In relation to this study, the most significant advantage I*Link has in comparison with automatic alignment systems is that in I*Link, additions and deletions can be distinguished, whereas in the latter, such cases cannot be separated from instances where the system just does not find a suitable link. Naturally, if additions and deletions cannot be distinguished, they cannot be studied.
Above and beyond all, what a system such as I*Link provides to the field of translation studies is the ability to fully align and investigate a large corpus of texts in a structured way, using the tools integrated in the system. It makes it possible to search the material and get structured output in mere seconds, once the alignment is done. In my opinion, the challenge is to find a way to analyse the material and the outputs so that scientifically interesting results can be presented. A risk with all systems that provide a lot of results in the shape of numbers and statistics is that it is tempting to over-use the possibilities for making calculations. Consequently, the researcher must be very focused on the scope of the study and avoid presenting all figures that can be calculated on the aligned material.

In my personal experience, the built-in heuristics can also cause problems in some cases, because if the user chooses another strategy than the one pre-programmed in I*Link, the system does not respond to this as quickly as perhaps desirable, but continues to suggests links that are preferable according to the heuristics. This can be very frustrating to the user and the risk is that the system by continuously working against the user dominates the choice process and convinces the user to adapt to I*Link, which means that the links will be less consistent than necessary.

The specific situation in which this caused a problem to me was described in section 6.2.3, and relates to the linking of proper names. Had I instead of my own strategy chosen to make one link of a character’s name and surname, as I*Link is built to do, the automatic heuristic would have been very helpful, and could have aided me in keeping my links consistent.

8.2.3 Specifics of I*Link as Sources of Error

One of the biggest problems with I*Link is that it does not support the alignment of discontinuous phrases, which is a cause of frustration to the annotator, but more importantly, the quality of the alignment is affected. The implications of this for this particular study is that sometimes additions and deletions cannot be marked as NULL links, because they are, for example, surrounded by an auxiliary verb and its main verb, as in sentence pair 626 in HP1: “He dodged the Smeltings stick and went to get the post.” In Swedish, the information specifying that the stick is from Smeltings is deleted: “Han hoppade åt sidan för kappen och gick ut för att hämta posten.” Here, kappen is the equivalent of the stick, and the annotator is forced either to include Smeltings in the link, or mark the Smeltings as deleted and align just stick with kappen. This is a situation that the annotator often faces, and it is basically a matter of choice whether to delete the whole segment, or to align it and include the word or words that are in fact added or deleted.

In defence of I*Link can be said that the creators of the system are well aware of this problem (Merkel et al. 2003), and it is possible to mark links as discontinuous. However, in the current version of I*Link, this does not have any practical effect as the system lacks possibilities to treat the included segments in a link marked discontinuous any differently than if the link was simply accepted.
as one large continuous segment. In other words, a discontinuous link must be nullled or accepted in its entirety. In the alignment of the HP-corpus, I confess that I did not mark discontinuous segments as discontinuous, partly because I discovered the possibility to do so at a very late date in the alignment process, and partly because doing so would have no practical effect on the results of this study.

In the sentence pair above in which Smeltings is omitted, the problem could also have been solved if lexical shifts were implemented. The Smeltings stick and käppen could then be linked together as a less specific lexical shift.

Another issue with I*Link is that words cannot be divided and aligned in subsegments, which can be a problem with the tendency in Swedish to form compound nouns. Sometimes, a piece of information is added to the compound, but it cannot be marked as an addition. One example of this is sentence pair 385 in HP1: “The only thing Harry liked about his own appearance was a very thin scar on his forehead that was shaped like a bolt of lightning.” The target sentence reads: “Det enda Harry gillade i sitt eget utseende var ett mycket smalt ärr i pannan som hade formen av en sicksackblixt.” Sicksack is an addition that gives additional information about the shape of the scar, but it cannot be marked as an addition in I*Link. However, if lexical shifts were implemented in the program, sicksackblixt could be linked together with bolt of lightning as a more specific lexical shift, which would solve this problem.

I find it highly likely that the problems with discontinuous segments and not being able to divide words have affected the results of this study. How much and in precisely what way is, unfortunately, not possible for me to say.

8.2.4 Suggestions for Improvements of the Tools

Drawing on the experience of aligning and working with I*Link and I*Trix accumulated in this project, a few suggestions for augmentations that can be made to the tools does not seem out of place. The suggested changes could be of use both in projects like this one, and in working with the tools in general. No so-called usability evaluation of I*Link has been made during this project, and the suggested improvements are primarily motivated by either being able to shorten the time required by the alignment, or by making it possible to perform a more fine-grained and automated analysis of the material.

Although I*Link gives the researcher an opportunity to study translations in a structured and accessible manner, more could be wished for in terms of search options. For any study focusing on the differences between source and target texts, it would among other things have been very positive to be able to search for elements of a particular word class that are translated into a different word class, for example verbs that are translated as anything but verbs. This can be done to some extent, but since one actively has to search on every possible change separately (verbs translated as adjectives, nouns, etc.), it is either necessary to do a great deal of work, or to know exactly what to look for before the start of the analysis of the material. This can be a pity in explorative studies such as this one, where the results are sometimes not entirely predictable.
Another possible feature that would be very useful in I*Link would be to be able to select exactly what sentence pairs, or string of sentence pairs, a particular search covers. In this project, a feature like that would have come to much use for example in the case study made on the treatment of the dialects of Hagrid and Stan Shunpike. Was such a search option available, structured exact investigations of the characters’ specific ways of talking would have been very easy to conduct.

In addition, an algorithm of some sort that would make sure that links that are in close proximity of each other in the sentence would have more differentiated colours would be helpful.

It could possibly save time in the alignment process if I*Link included one button that, if pressed, performed some action like “Mark everything that is unmarked as a NULL link”. This could be useful in studies like this, because much time is devoted to nulling the links that have no match in the corresponding text. A button that marks all unmarked elements NULL might speed up the process, although using it would mean that the annotator must be very careful and always check that all unmarked elements are indeed without correspondences.

Lastly, as I have already hinted at, implementing lexical shifts would have a number of positive effects. The analysis could be made more fine-grained and more powerful searches of the material could be made automatically. This could simplify and strengthen the possibilities for investigating both the influence of the translator on the text and manifestations of translation universals.

8.3 Suggestions for Further Research

The results of this study seem promising, and gives much food for thought on further research in the field of descriptive translation studies. The first suggestion for future research is specific to the word aligned corpus used in this study. The other two suggestions are given in an order of increasing general applicability in studies of this kind.

Firstly, as I have reason to question if the seemingly clear sequential development of the translator stems from the translations or the word alignment, a study that deals with this question could be fruitful. On a more general level, I suggest research into possible sequence effects and how to avoid them in studies of this kind.

Secondly, one issue that needs to be paid attention to is the consistency of annotators, and ways of controlling and investigating consistency. I believe that this is necessary to ensure the reliability of the results of studies of this kind, as means to control consistency will hopefully make word alignment less subjective.

Lastly, fantasy, and even fiction, are more or less unexplored genres of literature when it comes to translation studies. I think that this is a pity, since much of the material that is translated and read is fiction. Moreover, the fantasy genre should prove to be particularly interesting to study as it contains
8.3. SUGGESTIONS FOR FURTHER RESEARCH

much domain specific language use and many neologisms. Consequently, more research into translation of fantasy and fiction is needed.


Att spåra översättningsuniversalier och översättarutveckling genom att ordlänka en Harry Potter-korpus

Tracing Translation Universals and Translator Development by Word Aligning a Harry Potter Corpus

Sofia Helgegren

For the purpose of this descriptive translation study, a translation corpus was built from roughly the first 20,000 words of each of the first four Harry Potter books by J.K. Rowling, and their respective translations into Swedish. I*Link, a new type of word alignment tool, was used to align the samples on a word level and to investigate and analyse the aligned corpus. The purpose of the study was threefold: to investigate manifestations of translation universals, to search for evidence of translator development and to study the efficiency of different strategies for using the alignment tools.

The results show that all three translation universals were manifested in the corpus, both on a general pattern level and on a more specific lexical level. Additionally, a clear pattern of translator development was discovered, showing that there are differences between the four different samples. The tendency is that the translations become further removed from the original texts, and this difference occurs homogeneously and sequentially. In the word alignment, four different ways of using the tools were tested, and one strategy was found to be more efficient than the others. This strategy uses dynamic resources from previous alignment sessions as input to I*Trix, an automatic alignment tool, and the output file is manually post-edited in 1*Link.

In conclusion, the study shows how new tools and methods can be used in descriptive translation studies to extract information that is not readily obtainable with traditional tools and methods.

word alignment, translation universals, translator development, corpus, additions, deletions