Conjunction in Translation from English to Arabic: A Corpus-Based Study

By

Ibrahim Fayz Alasmri

A thesis submitted in fulfilment of the requirements for the degree of

Master of Research in Linguistics (Translation Studies)

Department of Linguistics

Macquarie University

October 2016

Table of contents

List of tables

List of figures

Transliteration of Arabic conjunctions

Arabic Conjunction	English equivalent	Transliterated version
و	and	*wa*
أو	or	*aw*
ثم	then, subsequently	*thumma*
لكن	but, yet, however, nevertheless	*lakin*
حتى	until, as far as, till	*hatta*
عندما	when, at the time when	*in-da-maa*
كما	just as, similarly, likewise	*ka-maa*
إلا	but, except	*‘illa*
حيث	where	*hayth*
حين	when, at the time when	*hiin*
بل	rather, but actually	*bal*
أما	as for	*‘amma*
ربما	perhaps, maybe, possibly	*rub-ba-maa*
أم	or	*‘am*
بينما	while, whereas	*bay-na-maa*
إذ	since, inasmuch as	*idh*
بعد أن	after	*ba'ad ‘an*
إذن	therefore, so, thus, in that case	*idhan*
بالإضافة	in addition, additionally	*bel-ida-fah*
خصوصا	in particular, specially	*khoso-san*

Abstract

Against the background of existing corpus-based research on the recurrent features of translated language, this study set out to investigate two main research questions: (1) What are the differences between translated and non-translated Arabic regarding the use of conjunctive markers? (2) What are the differences in register-related preferences for conjunctive markers between translated and non-translated Arabic? This study was conducted utilising a comparable corpus of translated and non-translated Arabic across two registers, namely creative fiction and legal texts. Conjunctive markers, as a cohesive pattern, were used as an operationalisation to investigate five features of translated language in this language pair: explicitation, simplification, normalisation, levelling out, and interference. A quantitative and qualitative analysis was carried out to investigate whether the independent variables of corpus and register have any significant main effects on the frequency of a set of conjunctive markers, as well as whether there is a significant interaction in the effects of these independent variables on the frequency of conjunctive markers. The findings from the investigation confirm that there are some differences between conjunction use in translated and non-translated Arabic, providing some support for features such as interference, normalisation and explicitation. Register has a strong and consistent effect on conjunction use, and the study also demonstrates that there are interaction effects between translation status and register that are specific to particular conjunctions, such that explicitation effects are only evident in some registers and not others, and only for particular conjunctions.

Keywords: cohesion, conjunctive markers, corpus-based translation studies, features of translated language, English-Arabic translation.

Statement of candidate

I certify that the work in this thesis entitled "Conjunctive markers in translation from English to Arabic: A corpus-based study" has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Name: Ibrahim Alasmri

MQ ID:

Signature:

10 October 2016

Acknowledgements

Praise be to God for His endless favours and benefaction. Thanks are due Him for the desire and ability He granted me to accomplish this work.

I owe the greatest debt of gratitude to my supervisor, Dr Haidee Kruger, for her boundless support, guidance and inspiration. Without her unique direction, insightful comments and incessant motivation at every stage of this endeavour, this study would not have been completed.

I also would like to thank my adjunct supervisor, Dr Nesrine Basheer, from the University of Sydney, for her valuable comments and suggestions on the Arabic section in the literature review.

My warmest gratitude goes to my parents overseas. I have always been motivated by their faithful love, care and prayers.

I would like to express my heartfelt thanks to my wife and soulmate, Ayshah, who spares no effort to help, encourage and support me all the time; to my daughter, Banan; and to my son, Ghassan. Their unconditional love, passion and care have always been my source of hope and success.

Ibrahim

October 2016
Sydney-Australia

Chapter 1: Introduction

1.1 Introduction

The quest for universals of translation has found a place in the centre of discussions in descriptive translation studies. The notion of universals is based on the idea that translated language is qualitatively different from non-translated language, and that it has unique linguistic features that "typically occur in translated rather than original texts and are thought to be independent of the influence of the specific language pairs involved in the process of translation" (Laviosa, 1998, p. 288). The concept of the universal or recurrent features of translated language has been a point of departure for numerous corpus-based studies which have aimed to distinguish trends of translation behaviour, and then to make generalisations about the typical features of translations.

Against the background of existing corpus-based research on the recurrent features of translated language, this study sets out to investigate some potential features of translated Arabic texts by examining the use of conjunctive markers. Conjunctive markers are used as an operationalisation to investigate five proposed features of translated language in this language pair: explicitation, simplification, normalisation, levelling out, and interference. The use of conjunctive markers is analysed in a comparable corpus of translated and non-translated Arabic across two different text types, or registers,[1] namely creative and legal texts.

[1] The concept of register is discussed in more details in Chapter 2 (see Section 2.2.3 and 2.3.3).

1.2 Research significance

From the literature review presented in Chapter 2, it is evident that there is a lack of studies on the features of translated language in the English-Arabic language pair, which remains understudied compared with other language pairs. In particular, as far as conjunction and its cohesive role are concerned, there are limited comprehensive and systematic corpus-based investigations on contrastive differences in conjunction between English and Arabic, and the way in which these differences may affect translation from English to Arabic. To our knowledge, there is no existing study that uses conjunction as operationalisation to study a range of features of translated language in translation from English to Arabic, with an emphasis on contrastive and register-based differences as well.

To address this gap, this product-oriented study uses a quantitative corpus-based approach to examine the use of conjunctive markers in translated and non-translated Arabic texts across two different registers, namely creative fiction writing and legal texts, respectively. The ultimate goal of this study is to investigate some potential recurrent features of translated Arabic texts and to provide a coherent account of the way in which the use of conjunctive markers may reflect the recurrent features of translated Arabic. The study takes advantage of conjunction's vital but nevertheless differentiated role in English and Arabic, and across registers, to investigate its applicability as an operationalisation of various features of translated language. The investigation of the features of translated language in this language pair will further contribute to the generalisability of claims about the universality of these features in all translations.

1.3 Research questions

Adopting a quantitative corpus-based approach, and using the custom-designed and -compiled comparable corpus described in Chapter 3, this study aims to address the following three main research questions and subquestions:

1- What are the contrastive differences between non-translated English and non-translated Arabic texts regarding the use of conjunctive markers?

2- Are there differences between translated and non-translated Arabic regarding the use of conjunctive markers?

   a. If so, what are these differences?

   b. Can these differences be viewed as the result of translation-specific processes proposed in the literature on the recurrent features of translated language, including explicitation, simplification, normalisation/conventionalisation, and interference?

3- Are there differences in register-related preferences for conjunctive markers in translated and non-translated Arabic?

   a. If so, what are these differences?

   b. Can these differences be viewed as a result of translation-specific processes proposed in the literature on the recurrent features of translated language, particularly levelling out of register differences, interference of register-based preferences of the source language, or over-accommodation to the register-based preferences of the target language as a consequence of normalisation?

1.4 Overview of method

To answer research question 1, a literature study of existing research on contrastive differences between conjunctive preferences in English and Arabic is carried out. Research question 2 and 3 are answered by means of a quantitative corpus-based method, applied to a register-differentiated comparable corpus of translated and non-translated Arabic. Using a bottom-up method to determine the conjunctions selected for investigation, the 20 most frequent conjunctions in the corpus are analysed to provide an overall assessment, with more detailed analyses of the five most frequent conjunctions in the corpus.

For all the analyses, factorial Analysis of Variance (ANOVA) (see Section 2.4.3), is used to determine whether translation status (i.e. original vs translated text status) and register (i.e. creative vs legal texts) significantly affect the frequency of use of conjunctions. This analysis also determines whether there are any interactions between the effects for translation status and register – in other words, whether preferences for conjunction use across registers are significantly affected by translated or non-translated status.

1.5 Research hypotheses

In the light of the literature on conjunction use in English and Arabic, the recurrent features of translated language, and register variation, this study hypothesises that:

1. Translated and non-translated Arabic texts will demonstrate significant differences regarding the frequency and distribution of conjunctive markers (in other words, there will be a significant main effect for translation status), and these differences can be ascribed to translation-specific processes leading to features such as explicitation, normalisation, interference and simplification.

2. Creative and legal texts will demonstrate significant differences regarding the frequency and distribution of conjunctive markers (in other words, there will be a significant main effect for register).
3. Translated Arabic creative and legal texts will demonstrate a significantly different frequency and distribution of conjunctive markers, compared to original texts in the same registers (in other words, there will be a significant interaction between translation status and register), as a consequence of translation-specific processes.

1.6 Research overview

Following this introductory chapter, this study consists of four subsequent chapters. **Chapter 2** reviews the pertinent literature regarding two main concepts and areas of study relevant to this research: the recurrent features of translated language, and the concept of conjunction, particularly in relation to differences between English and Arabic. The discussion of the recurrent features of translated language includes an overview of the main theories and corpus-based empirical studies, as well as an overview of the main five features of translated language that have been proposed (explicitation, simplification, normalisation, levelling-out, and interference), and the ways in which these features have been operationalised. The second part of the literature review focuses on the concept of conjunction. It reviews the concept of conjunction and its cohesive role in English and Arabic, and outlines some contrastive differences between these two languages regarding the use of conjunctive markers. This section also reviews some corpus studies that have investigated conjunction or other cohesive patterns, with a particular focus on English-Arabic studies.

Chapter 3 provides a detailed description of the quantitative corpus-linguistic

method that was used to carry out the research. It briefly describes how the literature review in Chapter 2 answers the first research question, but the main focus of this chapter is the method used to answer research question 2 and 3, and to test the hypotheses outlined in Section 1.5. The corpus compiled for and used in this study is discussed, and the methods followed for the extraction of data from the corpus are outlined. The data analysis methods, and statistical tests used are also discussed in detail.

Chapter 4 consists of a detailed presentation of the main findings, and a discussion of the possible interpretations of these findings, in particular in relation to the hypotheses of this study. The main findings regarding the frequency and distribution of all conjunctions investigated in this study are presented, with particular attention to the effects of translation status and register on the frequency of conjunctions overall. In the second part of the chapter, the findings for the five most frequent conjunctions are presented. The effects of translation status and register on the frequency and distribution of these conjunctions are analysed and discussed in more detail, to provide a more nuanced picture of the potential effects of translation on conjunction use in different registers in Arabic written texts.

Chapter 5 concludes the study. It reviews the main findings of this study, and considers some implications, limitations, and possible avenues for further research in this area of study.

Chapter 2: Literature review

2.1 Introduction

This chapter provides the background to two sets of concepts. The first is the "universal" or "recurrent features" of translated language (see Section 2.2), while the second is the concept of conjunction, as a cohesive pattern (see Section 2.3). Furthermore, as the current study aims to examine the use of conjunctive markers in translated and non-translated Arabic texts across two different text types (or registers), and its ultimate goal is to investigate some potential features of translated Arabic texts using a corpus-based approach, this chapter also highlights the implications of a potential relationship between the two sets of concepts. This chapter therefore specifically discusses the theoretical and methodological implications of existing research on the recurrent features of translated language with a special focus on studies that have used conjunction and other patterns of cohesion as operationalisation, leading to the identification of a current research gap in product-oriented investigations of the features of translated language in Arabic translations.

The first part of this chapter (Section 2.2) provides an overview of the concept of the "universal" or "recurrent" features of translated language, particularly as investigated within corpus-based translation studies. Five features proposed as characteristic of translated texts, namely explicitation, simplification, normalisation, levelling-out, and interference, are discussed (see Section 2.2.2). The main focus in this section is on the definitions of these features, and operationalisations of the features that have been used in previous corpus-based translation studies with their corpus designs and key findings.

The second section of the literature review (Section 2.3) sheds light on the concept of cohesion, and particularly on conjunctive markers in English and Arabic. This section discusses the framework for the analysis of cohesion in English proposed by Halliday and Hasan (1976) (Section 2.3.1), and the categorisation of conjunctive markers (Halliday & Matthiessen, 2004) (Section 2.3.2.1). It also considers how the concept of conjunction has been viewed in both classical and modern standard Arabic (Section 2.3.2.2). This section also outlines some contrastive differences between English and Arabic regarding the use of conjunctive markers (Section 2.3.3). Furthermore, it briefly discusses register variation and its potential effect on the distribution of linguistic features, including conjunctions (Section 2.3.4). Finally, some previous studies that have used conjunctions or other cohesive patterns to investigate the features of translated language are reviewed (Section 2.3.5).

2.2 The features of translated language

2.2.1 Theoretical and methodological foundations

A considerable amount of literature has developed on the topic of the linguistic nature of translated language. The notion that translated language is "intrinsically, observably and, above all, meaningfully different from language that has not been translated" (Kruger & Van Rooy, 2012, p. 2) recurs throughout research on translation, and has been described in many different terms, such as "the third code" (Frawley, 1984), "the third language" (Duff, 1981), "translationese" (Gellerstam, 1986), "universal features of translation" (Baker, 1993), "laws of translation" (Toury, 2012), "translation universals" (Mauranen & Kujamäki, 2004), and "features of translation" (Olohan, 2004). All these different terms reflect the idea that translation demonstrates unique linguistic features, described by Laviosa

(1998, p. 288) as "linguistic features which typically occur in translated rather than original texts and thought to be independent of the influence of the specific language pairs involved in the process of translation".

The study of the universal or recurrent features of translated language received an influential theoretical impetus from the work of Gideon Toury (1980, 1995, 2012), and specifically his notion of "laws of translation". He proposed the formulation of general laws and regularities for translation, describing these as "theoretical formulations purporting to state the relations between all variables which have been found relevant to a particular domain" (Toury, 2012, p. 295). The goal behind formulating these laws is to distinguish trends of translation behaviour and the occurrence of certain translation phenomena, which Toury regards as a fundamental task for descriptive translation studies.

Toury (2012) proposes two main laws of translation: the law of growing standardisation, which he explains as the tendency that "in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of habitual options offered by a target repertoire" (Toury, 2012, p. 304). The second law is the law of interference, which refers to the observation that "phenomena pertaining to the make-up of the source text tend to force themselves on the translators and be transferred in the target text" (Toury, 2012, p. 310). These two laws of translation have been a point of departure for many studies and hypotheses that have addressed the features of translated language.

Baker (1993) is credited with combining the notion that translated language is intrinsically different from non-translated language with corpus linguistic

methodologies, and proposing the concept of "universals of translation".[2] She defines these universals as "features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems" (Baker, 1993, p. 243). Her aim in using corpus-linguistic methodologies is to systematically investigate these linguistic features, or "universals" as she calls them, that occur regularly in translated rather than non-translated language. Baker (1993) identifies corpus linguistics as a new paradigm able to supply the broad evidential base needed to research hypotheses about the recurrent features of translated language, since the corpus-based methodology offers a fruitful way to investigate such features by using large electronic corpora that allow for control over such variables as text type, translator, or historical period. She proposes four main features of translated language, termed explicitation, simplification, standardisation, and levelling-out.

In contrast with Toury's laws of translation, Baker's formulation of the "universals" of translated language excludes any features associated with source-text or -language transfer or interference effects, precisely because these universal features are supposed to be the consequence of factors other than source-text or source-language influence. Subsequently, however, many researchers have returned to the notion that source-text or source-language influence is one of the defining features of translation as a type of contact variety (see Ellis, 1996; Halverson, 2015; Mauranen, 2004; Thomason & Kaufman, 2001; Toury, 2012). Toury's law of interference clearly indicates the importance of transfer effects as a defining feature of translated language.

2. The term "universals" is contentious, and many researchers do not agree with it. Mauranen and Kujamäki (2004) describe the term as "too radical, too abrupt, too absolute" (p. 9). Other researchers and theorists suggest other terms instead, such as "regularities", "laws", "tendencies", "features", "regular features", or "recurrent features". Against the background of the contentious nature of the term "universals", this study uses the term "recurrent features", for preference.

In this regard, Chesterman (2004) proposes a distinction between T-Universals, which are "characteristics of the way translators use the target language", and S-Universals, which are "characteristics of the way in which translators process the source text" (p. 7). He provides some examples of features that may be classified as S-Universals or T-Universals. For example, S-Universals include "lengthening, when translations tend to be longer than their source texts, the law of interference, the law of standardisation, dialect normalisation, reduction of complex narrative voices, the explicitation hypothesis, sanitisation, the retranslation hypothesis, reduction of repetition", while T-Universals include "simplification, conventionalisation, untypical lexical patterning, under-representation of TL-specific items" (Chesterman, 2004, p. 8). Empirical research in this area includes work on interference, the unique-items hypothesis, and untypical collocations (see Eskola, 2004; Mauranen & Kujamäki, 2004; Tirkkonen-Condit, 2004).

Baker (1995) identifies three main types of corpora for translation research (see also Zanettin 2014). The first one is comparable corpora, which "consist of two separate collections of texts in the same language: one corpus consists of original texts in the language in question and the other consists of translations in that language from a given source language or languages" (p. 234). The second type of corpus is the parallel corpus, which consists of "original source language-texts in language A and their translated versions in language B" (p. 230). The third type of corpus is multilingual corpora, which are "sets of two or more monolingual corpora in different languages, built up either in the same or different institutions on the basis of similar design criteria" (p. 232).

The first investigations of the features of translated language relied on comparable corpora, in which translated texts are compared to non-translated texts, where both corpora must cover "a similar domain, variety of language and time span, and be of

comparable length" (Baker, 1995, p. 234). According to Baker (1995), the most important contribution that comparable corpora can make to the discipline is "to identify patterning which is specific to translated texts, irrespective of the source or target languages involved" (p. 234). The preference for this methodology is closely related to Baker's target-language orientation in her concept of the "universals" of translated language, which largely discounts source-language transfer effects.

This target-oriented methodology, utilising comparable corpora, has yielded some important findings (see De Sutter, Delaere, & Plevoets, 2012; Kruger & Van Rooy, 2012; Laviosa, 1998; Olohan & Baker, 2000; Pastor, 2008; Xiao, 2010); however, it has been criticised for not taking sufficient account of the potential effects of source-text and source-language influence. Studying the effects of transfer (and other S-universals), however, typically requires a comparison between translations and their source texts, and for this purpose, the use of parallel corpora is required. Parallel corpus analysis is particularly concerned with features associated with interference, the translation of unique items, and untypical collocations (see Zanettin, 2014). However, parallel corpora have the drawback that they do not allow for the investigation of translations in relation to the recipient system, as comparable corpora do. In more recent work in corpus-based translation studies, a new type of corpus has been developed to overcome the limitations of using either a comparable or a parallel corpus design, and to allow for all types of comparisons. This type of corpus is called a "bidirectional parallel and comparable corpus", or a "reciprocal corpus", and "contains a combination of two types of corpora, bilingual parallel and monolingual comparable, in two directions of translation" (Zanettin, 2013, p. 28).

To conclude this brief overview of the theoretical underpinnings and methodological orientations of corpus-based studies of the recurrent features of

translated language, it should be emphasised that these features are not uncontroversial. A number of scholars, including Baker herself, have questioned these features, as well as the explanatory hypotheses for why translated language would demonstrate a greater incidence of these features (see Baker, 2004, 2007; Laviosa, 1998; Mauranen, 2000; Mutesayire, 2004; Olohan & Baker, 2000; Pápai, 2004; Pym, 2005). There is, even among those who accept the notion of recurrent features of translation, considerable disagreement about the causes of these features. For instance, Baker (1995) explains these features as a consequence of translators' attempt to make their translation more accessible to readers, while Pym (2005) and Becher (2010) explain them as a result of translators' "risk-avoidance strategies". On the other hand, Halverson (2003) explains these features with reference to human cognition, as a consequence of the bilingual language processing taking place in translators' minds.

The following section provides a brief background to five features of translated language, namely explicitation, simplification, standardisation, levelling-out, and interference. These are the features which will be investigated in this study, particularly in relation to conjunction markers in Arabic.

2.2.2 Overview of the features of translated language

2.2.2.1 Explicitation

The first and most widely investigated feature of translated language is explicitation. The concept was introduced to translation studies as a stylistic technique by Vinay and Darbelnet (1958), who define explicitation as "a stylistic translation technique which consists of making explicit in the target language what remains implicit in the source language because it is apparent from either the context or the situation" (p. 342). Explicitation as a feature of translated language

is explained more specifically by Baker (1996) as "the tendency to spell things out in translation, including, in its simplest form, the practice of adding background information" (p. 197). It is therefore the conscious and/or unconscious attempt of translators to increase the explicitness of their translations, in comparison to either the source text, or in comparison to similar non-translated texts in the target language. Explicitation may therefore be viewed as an S-universal as well as a T-universal.

Explicitation can be reflected in a number of ways, such as the use of supplementary explanatory phrases in translation; the resolution of source text ambiguities in the translation; repetition; rationalising tenses; the addition of conjunctions, cohesive devices and discourse-organising items to the target text; the use of optional or redundant syntactic elements; and the avoidance of elliptical structures (Klaudy, 2008; Olohan & Baker, 2000; Vanderauwera, 1985).

Explicitation has been investigated using both monolingual comparable and bilingual and multilingual parallel corpora at different levels. According to Zanettin (2013), explicitation has been investigated primarily at the levels of syntax and discourse. For example, a number of studies have been carried out at the syntactical level analysing the distribution of explicitating devices involving optional choices in translated and non-translated texts. For instance, Olohan and Baker (2000), Williams (2005), Kruger and Van Rooy (2012) and Redelinghuys and Kruger (2015) all investigate explicitation utilising the optional complementiser *that* in English as an indicator of explicitness. Based on comparable designs, all these studies find support for explicitation at the syntactic level.

Another syntactic operationalisation of explicitation is used in Jiménez-Crespo (2011), in which a comparable corpus of original Spanish texts and texts translated

from "localised American English" to Spanish is analysed to examine the frequency of optional Spanish personal pronouns in the translated and non-translated subcorpora of the comparable corpus. This study finds a more frequent use of articles and personal pronouns in the translated texts, which is taken as evidence of explicitation.

Discourse-level linguistic indicators have also been used as operationalisation of explicitation in some studies. For example, explicitating shifts in cohesion have been analysed by Hansen-Schirra, Neumann, and Steiner (2007) in their bidirectional parallel corpus study, where they compare English and German across eight registers. They find significant support for explicitation in the use of some cohesive devices such as relative pronouns, ellipsis, and conjunctive markers. Øverås (1998), in a bidirectional parallel corpus study, compares Norwegian original novels with their translations into English, and English original novels with their translations into Norwegian. She investigates the addition of cohesive devices to explore the degree of explicitation, and the omission of cohesive devices to explore the degree of implicitation. The study finds that explicitation overall is more dominant than implicitation in both translation directions. Pápai (2004) uses a number of discourse-related features as operationalisation. This study is based on both a parallel corpus consisting of English texts and their Hungarian translations, and a comparable corpus containing Hungarian original texts that are compared with Hungarian translations. The addition of punctuation marks, derivatives, conjunctions, cataphoric reference, and discourse particles to the target texts is investigated, and it is shown that both corpora display a clear-cut tendency of translation-related explicitation. Mutesayire (2004) uses apposition markers in English, such as *in other words*, *namely* and *that is to say*, as operationalisation of explicitness. This comparable corpus study shows a significantly higher frequency

of reformulation markers in translated texts compared to non-translated texts, which supports the notion that there is a higher degree of explicitness in translated language compared to non-translated language.

2.2.2.2 Simplification

The second proposed recurrent feature of translation is simplification. This feature refers to "translators' tendency to simplify the language or message or both" (Baker, 1996, p. 183). It thus reflects the idea that the language of translated texts is simpler than that of non-translated texts, both lexically and syntactically. Some researchers (see Baker, 1996; Laviosa, 1998; Williams, 2005) argue that simplification can be evident in the use of shorter sentences, a more restricted vocabulary, and/or a smaller number of content words (compared to function words) in translated texts. At the lexical level, lexical variety (which refers to the range of vocabulary) and lexical density (which refers to content words that represent the information load), have been used in a number of studies as operationalisation of simplification (Kruger & Van Rooy, 2012; Laviosa, 1998; Pastor, 2008; Xiao, He, & Yue, 2010). Laviosa (1998), for example, compares lexical variety and lexical density in translated and non-translated English literary and newspaper texts. She finds that translated language tends to be simpler than non-translated language, demonstrating less lexical variety and a lower proportion of content words. Laviosa's findings have been replicated for a number of languages, such as Pastor (2008) for Spanish, and Xiao et al. (2010) for Chinese.

At the syntactic level, simplification has been operationalised using measures such as sentence length. Laviosa (1998) hypothesises that translated texts have shorter and thus less complex sentences than non-translated texts, and her findings partially confirm this hypothesis. Other studies, such as Williams (2005) and Redelinghuys

(2013), combine lexical and syntactic measures of complexity in the single measure of readability, to determine whether translated texts are simpler than non-translated texts. Williams (2005), in a bidirectional corpus of translated and non-translated English and French, finds that her translated subcorpus has shorter and less complex sentences, but has larger information loads and vocabularies. Redelinghuys (2013) similarly finds only partial support for the simplification hypothesis.

2.2.2.3 Normalisation

The third feature of translated language is normalisation, also sometimes referred to as "standardisation", "conventionalization", "conservatism" and "sanitisation". It is defined as "the tendency to conform to patterns and practices that are typical of the target language, even to the point of exaggerating them" (Baker, 1996, p. 177). In other words, normalisation can be viewed as translators' choices in favour of conventional expressions rather than creative ones (Zanettin, 2014). Normalisation in translated texts can be measured by identifying the degree of conformity to convention compared to non-translated texts in the same language, or source texts in a different language.

Normalisation has been operationalised in various ways. Many studies have focused on lexis and collocation, assuming that the tendency towards standardisation predisposes translators to avoid creative lexis and collocation to a greater degree than original writers would. For instance, Kenny (2001), comparing a parallel corpus of German literary source texts and their English translations, concludes that about 44% of creative lexical items and 16% of creative collocations in the source texts were normalised in the target texts.

Comparable corpus designs have also been used to investigate normalisation at the lexical and collocational level. As far as lexis is concerned, the frequency of

coinages and loanwords has been used as an operationalisation of lexical creativity, with the hypothesis that translated language would demonstrate a lower frequency of such word-forms because they are less conventional (Bernardini & Ferraresi, 2011; Kruger & Van Rooy, 2012; Redelinghuys & Kruger, 2015). Bernardini and Ferraresi's (2011) combined comparable and parallel study, using non-translated Italian texts, non-translated English texts, and their Italian translated counterparts, confirms this hypothesis by demonstrating that translators avoid using loanwords more frequently than the source-text writers do, following the trend of normalisation. Redelinghuys and Kruger (2015) arrive at a similar conclusion using their corpus of English translated from Afrikaans, across professional and non-professional translation.

Other studies have examined lexical patterning, collocations, or lexical bundles/clusters instead of specific lexical items (Baker, 2004; Bernardini, 2007; Bernardini & Ferraresi, 2011; Dayrell, 2007; Kruger & Van Rooy, 2012; Kruger, 2012). For example, in Kruger's (2012) comparable corpus study, she finds a statistically significant higher frequency of trigrams in her translated subcorpus, which represents more normalised language.

2.2.2.4 Levelling-out

Levelling-out refers to the "tendency of translated text to gravitate around the centre of any continuum rather than move toward the fringes" (Baker, 1996, p. 177). This feature has not been widely investigated. One of the studies that have attempted to operationalise levelling-out is Williams (2005), which uses readability scores as a measure to investigate to what extent translated texts are more homogeneous and gravitate to a "neutral middle" in terms of readability (rather than demonstrating

more variability), and may accordingly be regarded as more "levelled-out" (p. 151). However, her findings do not support this hypothesis.

Two other studies have investigated levelling-out as a register feature. Kruger and Van Rooy (2012) included register as an independent variable in their study comparing texts translated from Afrikaans and European languages into English and non-translated English texts. Redelinghuys (2013) similarly investigates register variability in her comparable corpus study which uses three English subcorpora: translations produced by experienced translators, translations by inexperienced translators, and non-translated texts. Both studies operationalised levelling out as the degree of register variability demonstrated by a set of linguistic features in translated texts compared with non-translated texts. Both studies find limited support for the feature of levelling-out across registers.

2.2.2.5 Interference

Interference is considered as a recurrent feature of translation by a number of studies that advocate the recognition of source-text and source-language transfer phenomena (contra the work of, for example, Baker, 1993). The important role of interference is outlined in Toury's law of interference (Toury, 1995, 2012), which defines interference as follows: "Phenomena pertaining to the make-up of the source text tend to be transferred to the target text" (1995, p. 275). The concept of interference was first used in second-language learning studies to mean a "deviation from the norms of either language which occurs in the speech of bilinguals as a result of their familiarity with more than one language" (Weinreich, 1953, p. 1). In translation studies, the term "interference" is used to refer to the influence of a source language or text on a target text, where the translation process itself represents a language contact situation between two languages in which the cross-

linguistic influence is expected (see Ellis, 1996; Grosjean & Soares, 1986; Thomason & Kaufman, 2001; Toury, 2012). Interference has been referred to using different terms and concepts, such as "transfer" and "SL interference" (Mauranen, 2004; Toury, 2012), the "unique items hypothesis" (Eskola, 2004; Tirkkonen-Condit, 2004), "shining through" (Hansen-Schirra, 2011), and "untypical collocations" or "usage patterns" (Jantunen, 2004; Mauranen, 2000b).

This cross-linguistic influence between the source-text or -language and target text can be found at different levels, including lexical, syntactic, pragmatic and/or textual levels, with the syntactic level the most commonly influenced by interference phenomena (Mauranen, 2004, p. 67). Transfer effects not only include more obvious or overt transfer of source-language items or structures, but may also take place in more subtle ways. For example, it may also manifest when target-language-preferred constructions are replaced with less preferred or conventional constructions under the influence of the source language, which may lead translators to produce usage patterns in the target language that are distinct from those in naturally occurring language.

In terms of lexis, for example, Mauranen (2004) investigates interference as a feature of translated language by comparing translated Finnish texts from ten different source languages with original Finnish fiction texts in a comparable corpus study. This study demonstrates evidence of lexical interference through examining the distribution of the most frequent words, where the translated texts deviate significantly from non-translated texts in this respect. It also confirms the phenomenon of interference as a universal or feature of translated language in the finding that all three translated subcorpora "bore a closer affinity to each other than to untranslated texts" (Mauranen, 2004, p. 79).

Tirkkonen-Condit (2004), who compares translated Finnish texts with non-translated Finnish texts, introduces an indirect type of interference in a new hypothesis called the "unique items hypothesis". This refers to the notion that target-language items that do not have a straightforward source-language equivalent tend to be underrepresented in translated language as compared to non-translated language. The activation of the source language suppresses the selection of features unique to the target language. The findings of Tirkkonen-Condit (2004) confirm this hypothesis, with frequencies of such unique items significantly lower in translated Finnish than in non-translated Finnish texts.

As is evident from the above brief overview, a number of operationalisations have been used to investigate the features of translated language. Linguistic features associated with cohesion, and in particular conjunctions, have been investigated in a number of studies, since they may be seen as operationalisations of a number of the features of translated language. The following section provides a brief overview of the concept of conjunction as a cohesive device, and the investigation of cohesion in the context of studies of the features of translated language.

2.2.3 Features of translated language across registers

Register has become an important factor in many contrastive linguistic and translation studies. Register, as functional language variation, can be defined as a "contextual category correlating groupings of linguistic features with recurrent situational features" (Gregory & Carroll, 1978, p. 4). Register variation has been widely investigated (see Biber, 1988, 1993, 1995; Biber & Finegan, 1994; Ghadessy, 1988, 1993). A distinction between register, genre, and text type can be made by defining genre and text type as "a category assigned on the basis of external criteria such as intended audience, purpose, and activity type, that is, it

refers to a conventional, culturally recognised grouping of texts based on properties other than lexical or grammatical (co-)occurrence features, which are, instead, the internal (linguistic) criteria forming the basis of text type categories" (Lee, 2002, p. 38). For example, while the "short story" is a genre, its text type is narrative, and its register is literary writing.

The relationship between register, genre or text type on the one hand, and the features of translated language on the other, has not received much attention, and the predominant register in many (especially early) corpus-based studies of the features of translated language was the literary register. This, of course, limits the generalisability of claims about the potential universality of the features of translated language. It is only in recent years that researchers have begun to explicitly consider the role that register plays in conditioning the features of translated language (Delaere, De Sutter, & Plevoets, 2012; Kruger & Van Rooy, 2012; Kruger & Van Rooy, 2016; Neumann, 2014; Redelinghuys & Kruger, 2015). This therefore remains an area where more research is required, a gap that this study aims to fill.

2.3 Conjunction and its cohesive role in translation

2.3.1 Theoretical overview

The concept of cohesion has been proposed and investigated by a number of linguists, such as Callow (1974), Gutwinski (1976), De Beaugrande (1981), and Hoey (1991), but the best known and the most detailed model has been developed by Halliday and Hasan (1976), in *Cohesion in English*. According to Halliday and Hasan (1976), cohesion occurs "where the interpretation of some element in the text is dependent on that of another. The one presupposes the other, in the sense that it cannot be effectively decoded except by recourse to it" (p. 4). Baker (1992)

defines cohesion as "a network of lexical, grammatical, and other relations which provide links between various parts of a text. These relations… organise and, to some extent create a text, for instance by requiring the reader to interpret words and expressions by reference to other words and expressions in the surrounding sentences or paragraphs" (p. 180). These relations or links are known as patterns of cohesion. Halliday and Hasan (1976) identify five main devices by means of which cohesion may be created: reference, substitution, ellipsis, conjunction, and lexical cohesion.

The first cohesive device is **reference,** which is defined as "a network of semantic relationships that allows the reader or the hearer to follow the participants, entities, events and so on in a text" (Baker, 2011, p. 191). In many languages, reference is achieved by various functional items such as pronouns, definite articles, demonstratives, and comparatives.

The second and the third categories of cohesive device are **substitution** and **ellipsis**, which involve networks of grammatical relations that presuppose the existence of certain elements in the text. Substitution works as a "replacement of one item by another" while ellipsis works as "the omission of an item. Essentially the two are the same process; ellipsis can be interpreted as that form of substitution in which the item is replaced by nothing" (Halliday & Hasan, 1976, p. 88).

The fourth cohesive device is **conjunction**, which is represented in the conjunctive markers that relate sentences, clauses, and paragraphs to one other. This cohesive device forms the main focus of this study, and will be discussed in detail in Section 2.3.2.

The last cohesive device is **lexical cohesion**. Unlike the four previous patterns, which are grammatical, lexical cohesion is "the cohesive effect achieved by the

selection of vocabulary" (Halliday & Hasan, 1976, p. 274). Lexical cohesion has two main types, namely reiteration and collocation.

2.3.2 Conjunction in English and Arabic

2.3.2.1 Conjunction in English

The fourth cohesive pattern Halliday and Hasan (1976) propose is conjunction. Unlike all the other cohesive devices, "conjunctive elements are cohesive not in themselves but indirectly, by virtue of their specific meanings; they are not primary devices for reaching out into the preceding (or following) text, but they express certain meanings which presuppose the presence of other components in the discourse" (p. 226). Conjunctive markers consist of lexical items that mark the relationship between the propositions in a text, such as additive, adversative or continuative relations, thereby connecting sentences, clauses, and paragraphs to one another (Baker, 2011, p. 200). In their initial work, Halliday and Hasan (1976) classified conjunctive markers into four main categories in terms of their cohesive relations in discourse: additive, adversative, causal and temporal. However, in later work, Halliday and Matthiessen (2004) systematically categorise conjunctions in a more detailed model in which they propose all its possible subclasses. These subclasses include elaboration, extension and enhancement, which Halliday and Matthiessen (2004) divide into further subtypes.

According to Halliday and Matthiessen (2004), the first type of conjunction is **elaboration**, which occurs when "one clause elaborates on the meaning of another by further specifying or describing it" (p. 396). Elaboration has two subtypes, **apposition**, which involves restating or representing the elaborated items by exposition or example, and **clarification**, which involves restating, summarising or clarifying the elaborated items (Halliday and Matthiessen, 2004). Table 2.1

summarises the elaboration subtypes and markers identified by Halliday and Matthiessen (2004, p. 541-543).

Type of conjunction	Subtypes		Items
Elaboration	Appositive	Expository	*in other words, that is, I mean, to put it another way.*
		Exemplifying	*for example, for instance, thus, to illustrate.*
	Clarification	Corrective	*or rather, at least, to be more precise*
		Distractive	*by the way, incidentally*
		Dismissive	*in any case, anyway, leaving that aside*
		Particularising	*in particular, more especially*
		Resumptive	*as I was saying, to resume, to get back to the point*
		Summative	*in short, to sum up, in conclusion, briefly*
		Verifactive	*actually, as a matter of fact, in fact*

Table 2.1: List of elaboration subtypes and markers

The second type of conjunction is **extension**, which occurs when "one clause extends the meaning of another by adding something new to it" (Halliday & Matthiessen, 2004, p. 405). Extension has two subtypes, **addition** and **variation**. Table 2.2 summarises the subtypes and items involved in this type of conjunction.

Types of conjunction	Subtypes		Items
Extension	Addition	Positive	*and, also, moreover, in addition*
		Negative	*nor*
	Adversative		*but, yet, on the other hand, however*
	Variation	Replacive	*on the contrary, instead*
		Subtractive	*apart from that, except for that*
		Alternative	*or (else), alternatively*

Table 2.2: List of extension subtypes and markers

The third type of conjunction is **enhancement**, which refers to when "one clause enhances the meaning of another by qualifying it in one of a number of possible ways: by reference to time, place, manner, cause or condition" (Halliday & Matthiessen, 2004, p. 410). The enhancement category of conjunction has four subtypes: matter, manner, spatio-temporal and causal-conditional. Table 2.3 shows these subtypes and some example items.

Types of conjunction	Subtypes			Items
Enhancement	Matter	Positive		*here, there, as to that, in that respect*
		Negative		*in other respects, elsewhere*
	Manner	Comparative		*likewise, similarly; in a different way*
		Means		*in the same manner*
	Spatio-temporal	Simple	following	*then, next, afterwards [including correlatives first...then]*
			simultaneous	*just then, at the same time*
			preceding	*before that, hitherto, previously*
			conclusive	*in the end, finally*
		Complex	immediate	*at once, thereupon, straightaway*
			interrupted	*soon, after a while*
			repetitive	*next time, on another occasion*
			specific	*next day, an hour later, that morning*
			durative	*meanwhile, all that time*
			terminal	*until then, up to that point*
			punctiliar	*at this moment*
	Causal - conditional	Casual	general	*so, then, therefor, hence*
			specific	*Result: as a result*
				Reason: on account of this
				Purpose: for that purpose
		Conditional	positive	*then, in that case, in that event, under the circumstances*
			negative	*otherwise, if not*
			concessive	*yet, still, though, despite this, however, even so, all the same, nevertheless*

Table 2.3: List of enhancement subtypes and markers

2.3.2.2 Conjunction in Arabic

The concept of conjunction in Arabic has been viewed either as a grammatical linking device that mainly coordinates units, such as words, phrases, clauses and sentences; or as a cohesive pattern that semantically and syntactically links different parts of the text. The former view, which the majority of Arab grammarians and rhetoricians espouse, introduces conjunctions as "connectives" or "connective particles" (*ḥurūf al-ʿatf*, in Arabic) that connect different parts at the sentence level. Nine main connective particles are identified, namely *wa* 'and', *fa* 'and/then', *thumma* 'then', *hatta* 'until', *aw* 'or' *am* 'or', *lakin* 'but', *bal* 'rather/but actually', and *laa* 'not' (Eid, 1975). This prominent view of Arabic grammar mainly depends on the "grammar of the sentence" which focuses on case or mood inflection, and neglects the textual function (Al-Amri, 2004).

On the other hand, the cohesive role of conjunctive particles has more recently attracted the attention of numerous Arab discourse analysts and modern standard Arabic (MSA) researchers (Al-Batal, 1990; Al-Jubouri, 1984; Alsaif, 2012; Beeston, 1983; Cantarino, 1975; Hassan, 1979; Holes, 2004; Ryding, 2005; Williams, 1989; Wright, 1967). These studies have attempted to expand the concept of conjunction in Arabic by describing and analysing different syntactic and semantic functions of these particles, and/or by introducing some new conjunctive markers. For example, Ryding's work (2005) is one of the recent works in MSA, in which she categorised Arabic conjunctive markers according their grammatical functions and effect (see Table 2.4).

Type of conjunction	Item in Arabic		English equivalent/Function
wa	و		*and*
fa	فـ		*and so, and then, yet, and thus*
Contrastive conjunctions	لكن		*lakin* (‘but, yet, however, nevertheless’)
	بل		*bal* (‘rather, but actually’)
	إنما		*inna-maa* (‘but, but moreover, rather’)
Explanatory conjunctions	أي		*ay* (‘that is, i.e.’)
Resultative conjunctions	إذ		*idh* (‘since, inasmuch as’)
	إذن		*idhan* (‘therefore, so, thus, in that case’)
	حتى		*hatta* (‘until, as far as, till’)
Adverbial conjunctions	Place	حيث	*hayth* (‘where’)
	Time	بينما	*bay-na-maa* (‘while, whereas’)
		بعدما	*ba’-da-ma* (‘after’)
		بعد أن	*ba’-da ‘an* (‘after’)
		بعدئذ	*ba’-da idhin* (‘after’)
		حين/حينما	*hiin/hiina-ma* (‘when, at the time when’)
		عندما	*in-da-maa* (‘when, at the time when’)
		عندئذ	*in-da-idhin* (‘then, at that time’)
		قبل أن	*qablq ‘an* (‘before’)
		ثم	*thumma* (‘then, and then, subsequently’)
	Similarity	كما	*ka-ma* (‘just as, similarly, likewise’)
		مثلما	*mithil-a-ma* (‘like, just as, as’)
		قدر ما	*qadr-a-ma* (‘as much as, just as, as … as’)
		حسبما	*hasb-a-ma* (‘according to, in accordance with, depending on’)
	Possibility	ربما	*rubb-a-ma* (‘perhaps, maybe, possibly’)
Disjunctives	أو		*aw* (‘or’)
	أم		*am* (‘or’)
Sentence-starting connectives	topic shift	أما	*amma* (‘as for’)
	Addition	إلى ذلك	*ila dhalik* (‘in addition to that, moreover, furthermore’)

Table 2.4: Ryding’s categorisation of Arabic conjunction (Ryding, 2005, pp. 409-421)

Another comprehensive categorisation model for conjunctions in Arabic is found in Alsaif (2012). It is one of the recent works in MSA, which identified and categorised Arabic discourse conjunctives according to their discourse relations, and by an automated process. Alsaif (2012, pp. 75-87) categorises Arabic conjunctions in eight groups, as follows:

1. The coordinating conjunction connectives

- *wa* (و) 'and'
- *fa* (و) 'and so/and then/yet/thus'
- *idh* (إذ) 'since, inasmuch as, as'
- *aw* (أو) 'or'
- *lakin* (لكن) 'but, yet'

2. The subordinating conjunction connectives

- *idha* (إذا) 'if'
- *illa* (إلا) 'except'
- *illa idha* (إلا إذا) 'except if'
- *illa an* (إلا أن) 'but'
- *illa baa'd* (إلا بعد) 'except after'
- *amma* (أما) 'as for'
- *inna-maa* (إنما) 'but, but moreover, rather'
- *hayth* (حيث) 'where'
- *be-sabab* (بسبب) 'because of'
- *ba'-da-ma* (بعدما) 'after'
- *in-da-ma* (عندما) 'when, at the time when'
- *gyr anna* (غير أن) 'however'
- *bay-na-maa* (بينما) 'while, whereas'
- *kulla-ma* (كلما) 'whenever'
- *be-ma'na akhar* (بمعنى آخر) 'in other words'
- *ka-anna* (كأن) 'as'
- *ka-ma* (كما) 'just as, similarly, likewise'
- *bal* (بل) 'rather, but actually'
- *ky* (كي) 'to'
- *li-dha* (لذا) 'for this'
- *la-siya-ma* (لا سيما) 'particulary'
- *li-anna* (لأن) 'because'
- *li-ky* (لكي) 'for, in order to'
- *lw* (لو) 'if' (in past)
- *lw-la* (لولا) 'if not'
- *ta-la-ma* (طالما) 'as long as'
- *wa-qabl* (وقبل) 'and before'
- *idha-fatan ila* (إضافة إلى) 'in addition to'
- *bog-ya-ta* (بغية) 'desire, to'
- *byd* (بيد) 'but'
- *byd an* (بيد أن) 'but'
- *fadh-lan an* (فضلا عن) 'as wll as'
- *hin-na-ha* (حينها) 'when that'
- *na-tija-ta* (نتيجة) 'result of'
- *qubail* (قبيل) 'shortly before'
- *ragma* (رغم) 'though'
- *ragma* anna (رغم أن) 'although'
- *khilafan le* (خلافا ل) 'unlike'
- *nadhran le* (نظرا ل) 'because of'

3. The noun connectives

- *idha-fatan ila* (إضافة إلى) 'in addition to'
- *byd* (بيد) 'but'
- *fadh-lan an* (فضلا عن) 'as well as'
- *na-tija-ta* (نتيجة) 'result of'
- *ragma* (رغم) 'though'
- *khilafan le* (خلافا ل) 'unlike'
- *bog-ya-ta* (بغية) 'desire, to'
- *byd an* (بيد أن) 'but'
- *hin-na-ha* (حينها) 'when that'
- *qubail* (قبيل) 'shortly before'
- *ragma* anna (رغم أن) 'although'
- *nadhran le* (نظرا ل) 'because of'

4. The adverbial connectives

- *aidhan* (أيضا) 'also'
- *hatta* (حتى) 'until, till, as far as'
- *hiin* (حين) 'when, at the time when'
- *li-dali-ka* (لذلك) 'therefore'
- *thumma* (ثم) 'then, subsequently'
- *hal* (حال) 'when'
- *hatta lw* (حتى لو) 'even if'
- *kadha-lika* (كذلك) 'and that'
- *min thamma* (من ثم) 'then'
- *koso-san* (خصوصا) 'in particular, specially'

5. The (prepositions + relative pronoun) connectives

- *fi-ma* (فيما) 'while'
- *mimma* (مما) 'which (+past verb)'

6. The preposition connectives

- *ithra* (إثر) 'after'
- *be* (ب) 'by'
- *aqiba* (عقب) 'shortly after'
- *jarra* (جراء) 'because'
- *li* (ل) 'for'
- *mondho* (منذ) 'since'
- *qabl* (قبل) 'before'
- *qabla an* (قبل أن) 'before that'
- *ba'ad* (بعد) 'after'
- *khilal* (خلال) 'during'

7. The prepositional phrase connectives

- *bel-moqabil* (بالمقابل) 'in contrast'
- *be-fadhl* (بفضل) 'thanks to'
- *be-hadaf* (بهدف) 'in order to'
- *be-ragm* (برغم) 'although'
- *be-ttali* (بالتالي) 'consequently'
- *ala ragm* (على الرغم من) 'although'
- *fil moqabil* (في المقابل) 'in contrast'
- *fi hal* (في حال) 'in case'

- *bel-ida-fah* (بالإضافة) ‘in addition’
- *be-rragm* (بالرغم من) ‘although’
- *fi hiin* (في حين) ‘while’
- *fi dil* (في ظل) ‘under’

8. Other discourse connectives

- *a’la alo’mom* (على العموم) ‘in general’
- *mathalan* (مثلا) ‘for example’'
- *bekh-ti-sar* (باختصار) ‘briefly, in sum’
- *bel-asas* (بالأساس) ‘basically’
- *bel-edafa* (بالإضافة) ‘in addition, additionally’
- *bel-fee’l* (بالفعل) ‘indeed’
- *be-hujati an* (بحجة أن) ‘because of’
- *ba’ad dhalika* (بعد ذلك) ‘after that’
- *jadyir bi-dhiker* (جدير بالذكر) ‘it should be noted’
- *dalilan a’la* (دليلا على) ‘evidence for’
- *ghalik anna* (ذلك أن) ‘that because’
- *a’la sabil al-mithal* (على سبيل المثال) ‘for example’
- *fi hadhih elathna’* (في هذه الأثناء) ‘in the meantime’
- *lihadha assabab* (لهذا السبب) ‘for this reason’
- *a’lawh a’la* (علاوة على) ‘in addition to’
- *a’la lax* (على العكس) ‘opposite’
- *a’la naqid* (على النقيض) ‘in contrast’
- *natijatan li* (نتيجة ل) ‘resulted by’
- *o’moman* (عموما) ‘generally’
- *fi’lan* (فعلا) ‘indeed’
- *fil-waqi* (في الواقع) ‘in fact, of course’
- *fi a’qab* (في أعقاب) ‘after all’
- *khitaman* (ختاما) ‘finally’
- *ka-daleel* (كدليل) ‘as an evidence’
- *li-a’jel* (لأجل) ‘for’
- *kholasah* (خلاصة) ‘to sum up’
- *li-a’lla* (لئلا) ‘for not’
- *wa fi alkhitam* (وفي الختام) ‘finally’

These two categorisation models of Arabic conjunctives formed the framework of this study, used in supplementary ways. All these conjunctions were considered in the bottom-up inductive frequency-based identification of conjunctives to investigate for the purposes of this study (see Section 3.4.2 for more details). The following paragraphs briefly discuss the five most frequent conjunctions analysed in more detail in this study (see Chapter 3 for details on how frequency was determined using the corpus designed in this study and Chapter 4 for exemplification of some of these uses in Section 4.4).

The most used conjunctive marker in Arabic is *wa* ‘and’, which has various grammatical and rhetorical uses in different contexts. According to Holes (1995), *wa* ‘and’ can act as a textual device or sentence connector and can express the following relations:

- It signifies the beginning of information chunks. It usually occurs at the beginning of a paragraph and mostly in narrative texts, where it is "performing the same function of marking the end of one sentence and the beginning of another" (Holes, 1995, p. 217).
- It expresses additive relations, in which two or more equal clauses are added to each other. In this additive relation, *wa* 'and' acts like *and* in English.
- It expresses temporal relations such as *then* in English, where it signals the successive relationship between clauses.
- It expresses simultaneous action, where two actions in different clauses are connected with *wa* 'and', which gives the meaning of *at the same time* and without giving "particular topical prominence" by showing which action happens first (Holes, 1995, p. 219).
- It expresses circumstantial relations, usually giving the meaning of *when/while* in English. In this relation, *wa* 'and' introduces and connects the circumstantial clause to the main clause.
- It expresses adversative relations, in which two or more adversative clauses are connected and *wa* 'and' acts as *but/yet*.

The second most frequent conjunctive marker analysed in this study is *aw* 'or'. In Arabic, *aw* 'or' is "a disjunctive indicates an option between two or more elements, but that option is inclusive, that is, it may include one, both, or all the elements" (Ryding, 2005, p. 418).

The third most frequent conjunctive marker analysed is *thumma* 'then', which usually connects sequential actions, and expresses the temporal relation between actions: "As a consequence of its temporal meaning, *thumma* usually implies that the action of the preceding sentence has been completed, thus introducing a new event or situation" (Cantarino, 1975, p. 36).

The fourth most frequent conjunctive marker analysed in this study is *lakin* 'but/yet'. It usually expresses an adversative relationship to the previous clause, sentence or situation (Cantarino, 1975).

Lastly, the fifth most frequent conjunctive marker analysed in this study is *hattaa* 'until'. It can be used for a variety of functions, but when it used as a conjunctive marker, it usually introduces a clause that "shows the consequences or result of the previous clause. Used in this way, it refers to an event or action that has taken place in the past" (Ryding, 2005, p. 413).

2.3.3 Contrastive differences in preferences for conjunctive markers between English and Arabic

While cohesive devices are common in most languages, languages do demonstrate contrastive differences in preferences for achieving cohesion, by either using different cohesive devices more frequently than others, or in specific combinations that may not match English patterns of cohesion. This section briefly discusses some contrastive differences between English and Arabic in terms of preferences in the use of conjunctive markers.

A number of studies in Arabic have been carried out discussing the concept of cohesion, including conjunction, from different angles (Al-Jabr, 1987; Hassan, 1979; Mehamsadji, 1988; Williams, 1989). These studies include some discussions of contrastive differences between English and Arabic in cohesive devices and other linguistic features. Other studies (Al-Amri, 2004; Al-Kashef, 2011; Al-Shabab, 1996; Djamila, 2010; Fattah, 2010; Lulu, 2013; Najjar, 2015; Shamaa, 1978), have examined some cohesive devices in English and Arabic in relation to translation studies, which will be discussed in more detail in Section 2.3.4 and Section 2.3.5.

As regards contrastive differences relating to the pattern of conjunction, there are some structural and discourse-related factors that may account for these differences. For instance, some differences exist as a result of different ways of constructing and chunking information, and different ways of linking such chunks. It is common in English to present information in relatively small chunks. Conversely, Arabic tends to use large chunks of information, where it is common to have an entire paragraph composed of only one sentence (Baker, 2011, p. 201). Therefore, the high frequency of conjunctive markers in Arabic may be regarded as a consequence of the need to connect different parts in the Arabic sentence.

Moreover, English depends on a very well developed punctuation system that further marks the relations between chunks of information. In contrast, a punctuation and paragraphing system is a recent development in Arabic (Holes, 2004). The more limited use of punctuation to mark the relations between propositions may increase the number of conjunctive markers that is used in the Arabic sentence instead of using punctuation marks (Holes, 2004).

In addition, the preference for presenting information in smaller chunks in English is obviously associated with the wide variety of conjunctive markers with many semantic relations which are used to link these chunks of information in English. In contrast, Arabic uses fewer conjunctions, but with various functions and meanings which can be understood from the context of the discourse (Baker, 2011) (see Section 2.3.2.2 for examples). This difference could explain the high frequency of specific conjunctive markers in Arabic texts.

Any discussion of the differences between languages regarding the use of conjunctive markers also needs to take account of register-based differences. Register may be regarded as conventionalised “patterned linguistic reflections of

contextual variation" (Neumann 2014:36). Register studies emphasise that language use varies systematically with different contexts of use. Such linguistic variation is usually analysed in terms of multi-dimensional constellations of linguistic features, characterised by probabilistic co-occurrence relationships (see Biber, 1988, 1993, 1995; Biber & Finegan, 1994; Ghadessy, 1988, 1993). Every register has its own linguistic characteristics or co-occurrence patterns of features that distinguish it from other registers.

The relation between conjunction and register variation is manifested when some registers tend to be more conjunctive than others, and each genre has its own preference for different types of conjunctions. Religious and fiction registers, for example, use more conjunctions than others (Smith & Frawley, 1983). While the main focus of this study is on the use of conjunctive markers in translated and non-translated Arabic, the study proceeds with an awareness that conjunctive markers may demonstrate differential uses not only across languages, but also across registers. For this reason, it was deemed necessary to include two distinct registers in the study (see Section 3.3.1).

2.3.4 Conjunction and the features of translated language

Section 2.2.2 has reviewed some of the main studies that have examined the recurrent features of translated language, using various linguistic indicators as operationalisations. This section sheds light specifically on those studies that have used conjunctive markers as operationalisations to investigate the features of translated language, and therefore demonstrates the relationship between the two main sets of concepts used in the study: conjunction, and the features of translated language.

Adding conjunctive markers to the translated text makes the relations between text

ideas clearer and more explicit to the reader, which is evidently related to the feature of **increased explicitness**. For example, Vanderauwera (1985) compares Dutch novels translated to English with non-translated English novels, and observes significant addition of conjunctions in the translated texts as a way to greater transparency. This finding is interpreted as evidence of explicitation (Laviosa, 2002). Other studies (Becher, 2011; Hansen-Schirra et al., 2007; Pápai, 2004; Puurtinen, 2004; Redelinghuys & Kruger, 2015) have also analysed the feature of explicitation in terms of the frequency of conjunctive markers. Pápai (2004), for example, in a parallel and comparable corpus study of Hungarian, finds significant addition of conjunctions to the target texts, demonstrating a tendency of translation-related explicitation. Mutesayire (2004) analyses the use of apposition markers in a comparable corpus study, which shows a significantly higher frequency of apposition markers in translated texts. This finding, too, provides support for a higher degree of explicitness in translated language.

As far as the feature of **interference** is concerned, translated texts may demonstrate evidence that conjunctive markers are used in ways that are more typical of the source language than of the target language. For instance, Bystrova-McIntyre (2012)'s study of a comparable corpus of English translated from Russian, and non-translated English, demonstrates the effects of the "law of interference" many times in cases where the English equivalents of cohesive particles typical of Russian are overused in the translated texts compared with non-translated texts written in English. For example, she ascribes the highest frequency of additive and adversative conjunction devices in translated texts to interference from Russian preferences for conjunctive devices.

The use of conjunctive markers can also be related to the features of **normalisation,** and **simplification**. If translated texts overuse the typical conjunctive markers in

the target language (compared to non-translated texts in the same language), there is evidence of normalisation or conventionalisation, (see Al-Kashef, 2011). As regards the feature of simplification, it is uncommon to investigate this feature in relation to the use of conjunctive markers. However, Al-Kashef (2011), in her comparable study, explained the use of conjunctions to break up long sentences in translated texts as a simplification procedure.

Lastly, if translated texts do not demonstrate the same degree of register-based variation in the use of some cohesive devices, such as conjunctive markers, as original texts in the same language do, there is evidence for the feature of **levelling-out** of register-based differences in favour of a neutral "middle" register. Kruger and Van Rooy (2012) and Redelinghuys (2013) in their comparable corpus studies of English translated from Afrikaans, and non-translated English, investigate conjunctive particles across different registers. Neither study finds convincing evidence for levelling-out; however, Redelinghuys (2013) does demonstrate that expertise plays a role, since translation-related levelling-out occurs when "inexperienced translators misjudge the appropriateness of the addition of conjunctive particles particularly in more informal registers" (p. 122).

2.3.5 Corpus-based studies of cohesion in general, and the features of translated language in Arabic translation

The number of studies focusing on the recurrent features of translated language in general has increased over time, and an increasing range of language pairs has been investigated, which has contributed to the generalisability of the theory of "universals" of translated language. However, there have been comparably few corpus-based studies on the recurrent features of translated language in relation to translated Arabic. There are numerous studies on translation between Arabic and

English, addressing different topics with different approaches; however, the number of corpus-based investigations that have analysed the features of translated language in this language pair is very limited.

Shamaa (1978) examines the frequency of some lexical items in a small monolingual comparable corpus of English translations of Arabic novels. Al-Shabab (1996) analyses the feature of simplification through examining lexical density in an English monolingual comparable corpus. Fattah (2010) focuses on explicitation, by investigating clause complexing and conjunctive explicitation in an Arabic parallel and comparable corpus. Lastly, Al-Kashef (2011) analyses the features of explicitation, simplification, and normalisation by examining cohesion in an English monolingual comparable corpus. With the exception of Fattah (2010), these studies all focus on translation from Arabic into English, and corpus-based research focusing on translated Arabic is scarce. While these studies do use cohesive devices, including conjunctive markers, as operationalisation, existing studies also do not generally take account of register variation and its implications.

These corpus-based studies present some significant findings in respect of the recurrent features of translated language between Arabic and English. According to Kruger (2004, p. 85), the study of Shamaa (1978), which finds a higher frequency of some common words such as *day* and *say* in English translations (from Arabic) compared to original English texts, supports a conventionalising trend in translation from Arabic to English. Al-Shabab (1996), who compares a translated corpus (translated from Arabic to English) with a corpus of original English texts, finds a lower type-token ratio with a greater level of repetition in the translated texts. He argues that this lexical repetition is the consequence of a simplification process. Fattah (2010), in his parallel and comparable study of conjunctives, uncovers some patterns of explicitating shifts and finds evidence of explicitation in the high

frequency of conjunctive particles in the translated texts compared with the original and non-translated texts. Finally, Al-Kashef (2011), in her study based on comparable corpora, finds evidence of explicitation in the high degree of repetition; and normalisation in the more frequent use of preferred conjunctive particles in the translated texts.

Despite the research on cohesion in Arabic, and some research on the features of translated language in translation between English and Arabic, conjunctive markers and their contrastive and register-based differences have hardly been used in corpus-based studies of the features of translated language in translation from English to Arabic. This study aims to fill this gap in existing research.

2.4 Conclusion

This chapter provided background to the two main sets of concepts that this study investigates: the recurrent features of translated language, and the concept of conjunction. A brief background discussion of the main recurrent features identified in the literature and the main studies in this area was provided, together with an overview of the concept of conjunction in English and Arabic. The relation between these two concepts was set out. The aim of this discussion was to investigate the possible implications of the relationship between these two sets of concepts, to gain a sound understanding of the use of conjunction as operationalisation in studies of the recurrent features of translated language (specifically in translation between English and Arabic), and finally to delineate the existing research gap in this research area.

The discussion in this chapter has answered the first research question of this study. In the next chapter, the corpus methodology used to investigate research question 2 and 3, and test the hypotheses formulated for this study, is set out in more detail.

Chapter 3: Methodology

3.1 Introduction

This chapter provides a detailed description of how this study was conducted. The chapter consists of three main sections. Section 3.2 identifies the general methodological approach that is employed in this study. Section 3.3 discusses the methodological role of the literature review and its contribution in answering the research questions and identifying the data that are used for analysis. Section 3.4 discusses the corpus analysis, and is divided into three parts. Section 3.4.1 outlines the corpus composition and compilation; Section 3.4.2 explains how the data were extracted from the corpus; and Section 3.4.3 explains how the data were analysed.

The ultimate goal of this corpus-based study is to investigate some potential recurrent features of translation from English to Arabic by examining the use of conjunctive markers. To attain this goal, it set three objectives. The first objective was to identify contrastive differences between English and Arabic regarding the use of conjunctions. The second objective was to identify differences between translated and non-translated Arabic by analysing the frequency of occurrence of conjunctive markers, and exploring the potential processes of simplification, interference, explicitation, and normalisation as explanations for any differences in the frequency of conjunctive markers. The third objective was to shed light on register variability in the use of conjunction in translated Arabic as compared to non-translated Arabic, to determine whether there is evidence of register differences being levelling out in translation, interference of source-language register preferences, or normalisation to target-language register preferences for the use of conjunctions. This chapter sets out to explain the method used to meet these objectives.

3.2 General methodological approach

In order to achieve the study objectives set out above, and to answer the research questions set out in Chapter 1, this study employed a primarily quantitative corpus-based approach, with some elements of qualitative analysis. The corpus-based approach relies on the use of large collections of authentic texts, known as corpora, which are collected in an electronic format according to certain pre-set criteria (Zanettin, 2014). Corpus-based approaches use these collections of texts as a source from which linguistic data can be extracted, processed and then analysed.

The corpus-based approach in descriptive translation studies has primarily been used for "research into regular patterns of behaviour which characterize the language of translated text, often referred to as 'translation universals'" (Zanettin, 2014, p. 11). As discussed in Chapter 2, many corpus-based studies have been carried out aiming to discover general norms, laws or universals of translation (Baker, 1993, 1995, 1996; Blum-Kulka, 1986; Chesterman, 1998, 2000; Frawley, 1984; Mauranen & Kujamäki, 2004; Toury, 1995, 2012; Vanderauwera, 1985). Therefore, it is an appropriate and valid approach for the current study, which is framed by the notion of "translation universals", or the recurrent features of translated language.

Quantitative corpus-linguistic methods (see Oakes & Ji, 2012), are used to answer research questions 2 and 3 of this study, although some qualitative discussion is used to supplement the quantitative analysis. According to Williams and Chesterman (2011), the purpose of qualitative research is to "describe a quality of something in some enlightening way" to "lead to conclusions about what is possible, what can happen, or what can happen at least sometimes; it does not allow conclusions about what is probable, general, or universal" (p. 64). On the other

hand, the purpose of quantitative research is to draw conclusions about "the generality of a given phenomenon or feature, about how typical or widespread it is … about regularities, tendencies, frequencies, and distributions" (p. 64).

Since the main goals of this study are to investigate the regularities and tendencies of translated Arabic texts across registers, by analysing the frequencies and distributions of conjunctive markers, quantitative analysis is an appropriate method for answering the research questions of this study.

3.3 Literature review

The literature review presented in Chapter 2 comprises two main sections. The first section (Section 2.2) provided an overview of the concept of the recurrent features of translated language, particularly as investigated within corpus-based translation studies. The second section (Section 2.3) of the literature review mostly discussed the concept of conjunction and the contrastive differences between English and Arabic regarding the use of conjunctive markers. This part of the discussion directly answers the first research question posed for this study by summarising the main concepts of conjunction in English and Arabic, and giving a brief theoretical background about the functions, uses, and markers of conjunction in both languages.

Moreover, by providing the necessary background to the concept of conjunction and its markers in both languages, a framework informing the extraction and analysis of data on conjunction use in the corpus analysis could be determined.

3.4 Corpus analysis

3.4.1 Corpus composition and compilation

Following one influential approach in corpus-based translation studies for investigating the features of translated language, this study used a monolingual comparable corpus of Arabic translated and non-translated texts. In this approach, translated texts are compared to non-translated texts in the same language. This target-oriented corpus method, which characterises much of the study of translation universals (Section 2.2.1), has been used in a number of studies (De Sutter et al., 2012; Kruger & Van Rooy, 2012; Laviosa, 1998; Olohan & Baker, 2000; Pastor, 2008; Xiao, 2010).

The comparable corpus used in this study consists of two subcorpora. The first subcorpus consists of texts translated from English to Arabic. This subcorpus included two registers: translated creative texts,[3] and translated legal texts. The second subcorpus was made up of non-translated Arabic texts, across the same two registers. This design allowed for a comparison of the use of conjunctive markers in translated and original writing across two registers. It also allowed for the investigation of whether there is any interaction between "translation status" (i.e. original or translated) and "register" (i.e. creative or legal) as factors that affect the use of conjunctive markers in the corpus – in other words, whether translations and non-translations are significantly different in their use of conjunctive markers across different registers. This analysis essentially answers the second and the third research questions formulated for this study.

[3] "Creative texts" in this study refers to fictional narratives.

As suggested in the discussion in Section 3.3.1, there are significant differences in conjunction use across registers. Different registers, such as creative and legal registers, may have distinct preferences for conjunction use, since texts in these registers have very different functions, which affect the use of linguistic resources, such as conjunctions and other cohesive devices. According to Smith and Frawley (1983), some registers tend to be more conjunctive than others, and each genre has its own preference for different types of conjunctions. For example, religious and fiction registers use more conjunctions than others. Preferences for particular conjunctions can be linked to the function of a particular text type or register. Creative fiction has the function of narration, which usually requires some indication of temporal or sequence relationships, necessitating the frequent use of conjunctions such as *wa* 'and' or *thumma* 'then'. Indicating temporal relationships, in contrast, would be less important in legal texts, but in the latter register other conjunctions (such as conjunctions indicating concessive relationships) may be more prominent, as a result of the function of legal texts. Another point in this regard is that the patterns of conjunctive use in English and Arabic across these two registers may potentially be very different, creating the possibility that interference effects might arise in specific registers.

The creative register in both translated and non-translated subcorpus is composed of a selection of stories and extracts from novels, collected from a number of published books and open-source corpora websites, where the source language of all the translations is English. The open-source corpora websites include:

- Arabic corpora, (Al-Sulaiti, 2009)

(http://www.comp.leeds.ac.uk/latifa/arabic_corpora.htm).

- Web-based (Searchable) corpora (Eddakrouri, 2016)

(https://sites.google.com/a/aucegypt.edu/infoguistics/directory/Corpus-Linguistics/arabic-corpora)

- Arabic Corpora resource

(http://aracorpus.e3rab.com/index.php?content=english)

All published books used in this creative register subcorpus are listed in details in the appendix.

The legal register is composed of government legal documents. In the translated subcorpus, documents translated from English into Arabic were sourced from the websites of governments and international organisations, including:

- Legal Aid NSW, Australia (http://www.legalaid.nsw.gov.au/publications/order-a-publication);
- Fair Work Ombudsman NSW, Australia (https://www.fairwork.gov.au/contact-us/language-help/arabic);
- The Department of Human Services, Australia (https://www.humanservices.gov.au/customer/information-in-your-language/arabic);
- Fair Trading NSW, Australia (http://www.fairtrading.nsw.gov.au/ftw/language_arabic.page);
- Ministry of Justice, United Kingdom (https://www.gov.uk/government/publications/screening-tests-for-you-and-your-baby-description-in-brief);
- The United Nations (https://documents.un.org/prod/ods.nsf/home.xsp).

For the subcorpus of original legal writing, Arabic texts were collected from ministries' and governmental organisations' websites in Arabic-speaking countries, including:

- Ministry of Finance, Saudi Arabia (https://www.mof.gov.sa/Pages/default.aspx);
- Ministry of Justice, Saudi Arabia (https://www.moj.gov.sa/ar/Pages/default.aspx);
- Ministry of Municipal and Rural Affairs, Saudi Arabia (https://www.momra.gov.sa);
- Bureau of Experts at the Council of Ministers, Saudi Arabia (https://www.boe.gov.sa/MainDefault.aspx?lang=ar);
- Ministry of Human Resources and Emiratisation, The United Arab Emirates; (http://www.mohre.gov.ae/MOLWebsite/en/home.aspx);
- Ministry of Justice and Freedom, Morocco (http://adala.justice.gov.ma/AR/Legislation/TextesJuridiques.aspx).

These corpus components were designed to be as comparable as possible, in terms of word count and publication time span. Regarding the word count, all texts ranged from 1000 to 4000 words, and each subcorpus contained about 100,000 tokens. In respect of the time span of publication, it was restricted to texts published from the 1980s to the present, which constitutes a synchronic corpus. The composition of the corpus is set out in Table 3.1.

	Creative	**Legal**	**Total**
Translated Arabic subcorpus	95,411	100,421	195,832
Non-translated Arabic subcorpus	91,303	98,028	189,331
Total	186,714	198,449	385,163

Table 3.1. Corpus composition

The first step in processing the corpus was capturing different types and formats of texts from the Internet. These formats included PDF documents, Word documents, HTML texts, and TXT files. This was followed by the second step, which involved converting these different formats to TXT files. The texts were copied, and then pasted in an editor file, such as a Word file. In the third step, the texts were edited by correcting all errors that resulted from the conversion process. This step also made a random selection of longer texts, to ensure that all texts would be between 1000 and 4000 words. The final step was converting the document to a plain TXT file. After this, the texts were ready for data collection.

3.4.2 Data collection

Data about the frequency of conjunctive markers were extracted by using the corpus software WordSmith Tools 7 (Scott, 2016), using the concordancing function. Corpus-linguistic researchers in Arabic face a number of challenges, in particular the fact that some corpus-processing software does not meet important criteria for searching and analysing Arabic corpora, such as "displaying Arabic text in its right-to-left direction, normalising diacritics and Hamza, providing an Arabic user interface" (Alfaifi & Atwell, 2016, p. 347). WordSmith Tools meets a number of these criteria and has the necessary functions for analysing (custom-built) Arabic corpora, such as displaying Arabic text in its right-to-left direction, and enabling the use of a personal corpus. For this reason, it was selected for the analysis in this study.

Extracting instances of the use of conjunctions was done by creating concordances, using conjunctive markers as search terms. The concordance function "retrieves all the occurrences of a particular search pattern in its immediate contexts and displays these in an easy-to-read format" (Bowker, 2002, p. 53).

A study investigating all Arabic conjunctions was not feasible within the constraints of this thesis project. A decision therefore had to be made about which conjunctions would be analysed. An inductive, bottom-up method of selection was used. A word list was created in WordSmith Tools, which lists all words occurring in the corpus, in order of frequency. The word list was compared with the typologies of conjunctives discussed in Chapter 2, and from this word list, the 20 most frequent conjunctive markers in the corpus were identified. Table 3.2 shows this list of 20 most frequent conjunctive markers selected for investigation in this study. [4]

No.	Arabic Conjunction	English equivalent	Transliterated version	Frequency
1	و	and	*wa*	31,192
2	أو	or	*aw*	4,374
3	ثم	then, subsequently	*thumma*	762
4	لكن	but, yet, however, nevertheless	*lakin*	720
5	حتى	until	*hatta*	646
6	عندما	when, at the time when	*in-da-maa*	482
7	كما	just as, similarly, likewise	*ka-maa*	472
8	إلا	but, except	*'illa*	343
9	حيث	where	*hayth*	299
10	حين	when, at the time when	*hiin*	225
11	بل	rather, but actually	*bal*	198
12	أما	as for	*'amma*	154
13	ربما	perhaps, maybe, possibly	*rub-ba-maa*	144
14	أم	or	*'am*	131
15	بينما	while, whereas	*bay-na-maa*	120
16	إذ	since, inasmuch as	*idh*	105
17	بعد أن	after	*ba'ad 'an*	96
18	إذن	therefore, so, thus, in that case	*idhan*	66
19	بالإضافة	in addition, additionally	*bel-ida-fah*	56
20	خصوصا	in particular, specially	*khoso-san*	53

Table 3.2: List of 20 most frequent conjunctive markers

[4] It should be noted that the conjunction *fa*, 'and, then' was the second most frequent conjunction, with 17,517 occurrences. *Fa* 'and, then' has a very similar function as *wa* 'and'. It was, however, very difficult to analyse, since the letter sequence *fa* occurs frequently at the beginning of many words, necessitating time-consuming manual analysis. For these reasons, this conjunction was excluded from the study.

For the 20 conjunctions with the exception of *wa* 'and', a concordance was created in WordSmith, and all cases were analysed. In other words, for these conjunctions, no sampling was done. The concordance was cleaned manually, by deleting all irrelevant (non-conjunctive uses) of the lexical items in question.

The cleaned concordances were exported to an Excel spreadsheet, and used to calculate the frequencies of conjunctive markers, per text. Since the texts in the corpus vary in length, frequencies were normalised to 1000 words. The frequency of individual markers (the dependent variable) was captured for each text, together with information about the independent variables of corpus and register.

The conjunction *wa* 'and' necessitated somewhat different treatment from the other conjunctive markers. This conjunction is the most frequently used in the corpus, as well as the most common conjunction in Arabic (although its frequency is somewhat inflated by the fact that it can also be used as a phrasal rather than clausal coordinator). As a consequence of the limited scope and time constraints of this thesis, all *wa* 'and' cases could not be analysed.[5] A sampling procedure was therefore followed for *wa* 'and', after which frequencies were extrapolated to the full corpus. This process involved a number of steps. First, a full concordance of *wa* 'and' cases was created, for the full corpus, the creative register, and the legal register. The next step was to reduce the number of concordance lines for *wa* 'and' by reducing the number of entries in the concordance to 500 in both registers, using the option to create a random downsample in Wordsmith Tools. After this, the concordance was cleaned by deleting all irrelevant cases, such as cases where *wa*

[5] Furthermore, *wa*, 'and' is one letter in Arabic 'و', which can be used as a conjunctive/connective marker, and can be as a letter in beginning of some words, like *wardah* (flower) 'وردة'. In the data collection process, this necessitates time-consuming manual analysis to distinguish conjunctive or connective markers from other uses. This is further motivation for the sampling process. Moreover, it is important to add that the search terms for the concordance were designed to also find prefixed uses in WordSmith Tools.

'and' was used as a phrasal rather than clausal coordinator. All non-conjunctive uses were therefore removed before further analysis.

For *wa* 'and', sample frequencies therefore had to be extrapolated to estimated frequencies for the whole corpus, before the data for *wa* 'and' could be added to the data for the other conjunctions. This calculation was done as follows. First a calculation was done to determine per register and per translated status (original versus translated) the average proportion of clausal uses based on the samples. Then the proportion of clausal uses per file was calculated, based on the samples. The average of the per file and per register and translated status for each file was then computed. Subsequently, the number of clausal uses of *wa* 'and' per file was estimated, based on the average proportion calculated in the previous step.

For the estimation, a smoothed estimate was computed, to account for differences in sample frequency. For files with samples of 20 or more, no smoothing was applied, and the estimate was directly based on the sample distribution (i.e. P(clausal *wa*) sample * overall count of *wa* in file). For files with fewer than 20 samples (1-19) a weight was assigned to the probability of the estimate derived from the sample, a linear weighting of sample size 1-19 itself. A weight of 20-sample size was assigned to the probability derived from the entire random sample for the subcorpus, and the estimate was calculated based on the trade-off of weights between the file estimate and the subcorpus estimate. This smoothed probability was multiplied with the number of observations in the file.[6]

[6] Smoothing is derived from computational linguistics, usually applied to n-gram data when that is used for language models (see http://arxiv.org/pdf/cmp-lg/9606011.pdf). We are indebted to Professor Bertus van Rooy for his assistance with these calculations.

Subsequent to this, the extrapolated or estimated frequencies for *wa* 'and' were exported to the same spreadsheet as the other conjunctions, and the same data about independent variables captured for the normalised estimated frequency of *wa* 'and' for each text. This spreadsheet was used for the data analysis.

3.4.3 Data analysis

To answer research question 2 and 3, the data were analysed using the statistical software R (R Core Team, 2016). The statistical method selected for the analysis is factorial Analysis of Variance (ANOVA). Factorial ANOVA is an analysis of variance which tests the hypothesis that the means of two or more populations are equal. It is used to investigate the effects of two main independent variables (in this study, translation status and register) on a dependent variable (in this study, the normalised frequency of conjunctions). It also allows for the investigation of an interaction between the two independent variables. In this study, its use therefore allows for an investigation of whether conjunctions demonstrate different frequency profiles across registers, depending on whether the texts in those registers are translated or not. The use of this statistical method therefore allows for the testing of the three hypotheses outlined in Chapter 1.

ANOVA has three important assumptions: independence of observations, normal distribution of data, and homogeneity of variance (Hill & Lewicki, 2006). The first assumption, which is independence of observations, was met in all cases. To assess the second and the third assumptions, additional tests were conducted. In the first step, a quantile-quantile plot was produced and visually inspected to assess whether the data were sufficiently normally distributed to use factorial ANOVA. In the second step, Levene's Test was carried out to test the assumption of homogeneity

of variance. If the test statistic (p-value) was larger than 0.05, then the equal variances assumption could not be rejected.

Where the data were sufficiently normally distributed and the assumption of homogeneity of variance was met, subsequent descriptive statistics are presented using the means as measure of central tendency, and factorial ANOVA was used for the analysis of the main effects and interaction of the independent variables of translation status and register. As post-hoc tests for the factorial ANOVA (where relevant), the t-test was used.

However, if data were not sufficiently normally distributed for the use of factorial ANOVA, medians are reported as measure of central tendency, and interaction plots are used to assess the main effects and interaction of the independent variables of translation status and register. Finally, in some cases when the data was non-normally distributed, the non-parametric Mann-Whitney *U*-test (a rank-order test which does not depend on the assumption of normal distribution of data) was used for pairwise comparisons.

Chapter 4: Findings and discussion

4.1 Introduction

This chapter reports the findings of the study by presenting a quantitative and qualitative analysis of the frequency and use of Arabic conjunctive markers in the comparable Arabic corpus designed for this study. The findings presented in this chapter are discussed in relation to research question 2 and 3 formulated for this study. The first question addresses the differences between translated and non-translated Arabic regarding the use of conjunctive markers, whereas the second question focuses on the differences in register-related preferences for conjunctive markers in translated and non-translated Arabic. The first part of this chapter presents the findings for all 20 conjunctive markers investigated in this study (see Section 4.2), taken together, by considering the overall frequency of these markers in the corpus. The second part investigates the effects of translation status (i.e. translated versus non-translated) and register (i.e. creative and legal) on the total normalised frequency of conjunctive markers (see Section 4.3). The third part focuses more specifically on the five most frequent conjunctive markers in this study, analysing the effects of translation status and register on these markers (see Section 4.4). Lastly, Section 4.5 summarises the findings and key conclusions.

4.2 Overall normalised frequency of conjunctions

As outlined in Section 3.4.2, the 20 most frequent conjunctive markers in the corpus were selected for investigation in this study. Figure 4.1 displays the overall normalised frequency (per 1000 words) of these 20 conjunctive markers in the corpus as a whole.

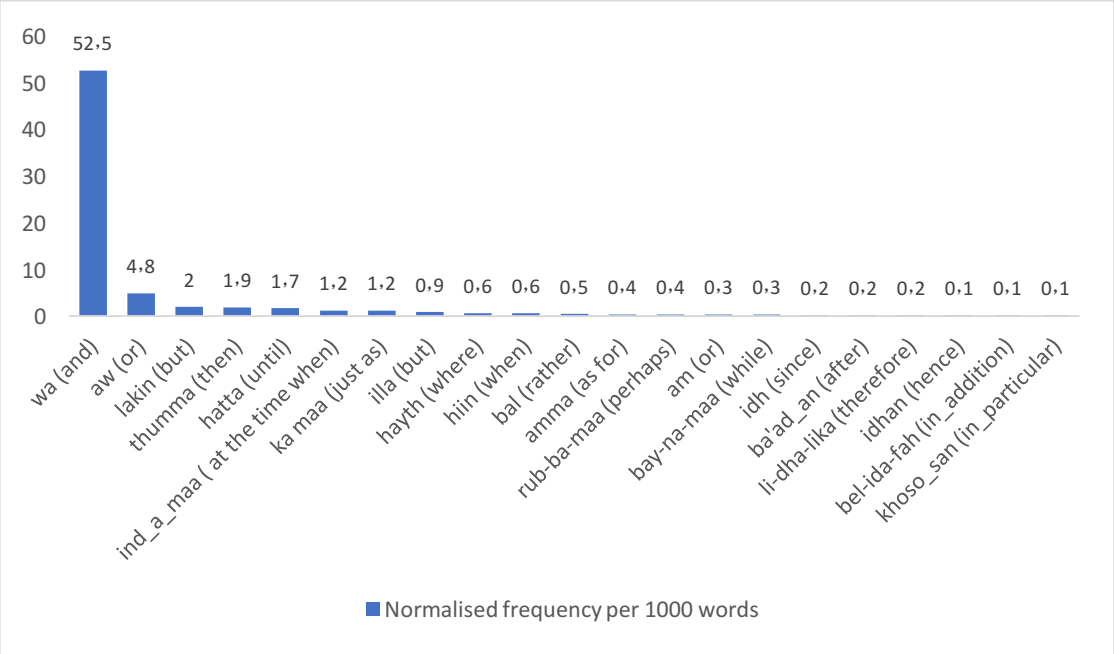

Figure 4.1: Normalised frequency of conjunctions (frequency of occurrence per 1000 words)

Figure 4.1 supports the literature on the most common or preferred conjunctions in Arabic, starting with the eight Arabic conjunctions (or connective particles) which are identified by the majority of Arabic grammarians and rhetoricians (see Section 2.3.2.2), and which have also been a point of departure for many later discourse analysts and MSA researchers who expanded on this set and categorised more conjunctions and functions.

In addition, as can be seen in Figure 4.1, the conjunction *wa* 'and' demonstrates, by far, the highest normalised frequency (52.5 per 1000 words) among all conjunctions. Consistent with the literature, this confirms that *wa* 'and' is the most widely used conjunction in Arabic, which has numerous functions, and can express various relations. The examples in Table 4.1 are taken from the corpus to illustrate different relations that can be expressed by the conjunction *wa* 'and'.

Wa 'and' function	Example	Translation
Expresses additive relations.	اتصل هاتفيا مع المحكمة المحلية التي سوف يجري فيها استماع لقضيتك واطلب منهم ان يحجزوا لك مترجما شفهيا بلغتك. Text: TR-Legal-35	Ring the Local Court where your case is to be heard and ask them to book an interpreter in your language.
Expresses temporal relations such as *then.*	وصعد مباشرة لبيت العمة ونام نوما عميقا Text: OR-Creative-35	And he went up to the aunt's house, and he fell in a deep sleep.
Signifies the beginning of information chunks.	وراقبت ذلك الشاطئ المألوف يدنو رويدا رويدا Text: TR-Creative-10	And I watched that common beach approaching slowly.

Table 4.1: Some examples of the varied functions of *wa* 'and'

Due to constraints of space, the functional differentiation of *wa* 'and' is not discussed in further detail in this study, but is foreseen as an important area of future investigation.

The following section focuses on the effects of translated versus non-translated status, as well as register, on the normalised frequency of the 20 conjunctions investigated in this study, viewed together.

4.3 The effects of translation status and register on the overall frequency of conjunctions

In order to measure the effects of translation status and register on the normalised frequency of conjunctions, factorial ANOVA was performed. While the first assumption of factorial ANOVA, the independence of observations, is met in all cases, two further analyses were carried out to test the two further assumptions of ANOVA: the normal distribution of data and homogeneity of variance (see Section 3.4.3). First, a quantile-quantile plot was produced and visually inspected to assess

whether the data are sufficiently normally distributed. Second, Levene's Test was carried out to test the assumption of homogeneity of variance.

The quantile-quantile plot in Figure 4.2 shows that the data are sufficiently normally distributed – in the plot, dots following roughly a straight line are indicative of normality. For a large sample size (such as this), this plot shows that the data are sufficiently normally distributed to use parametric statistics.

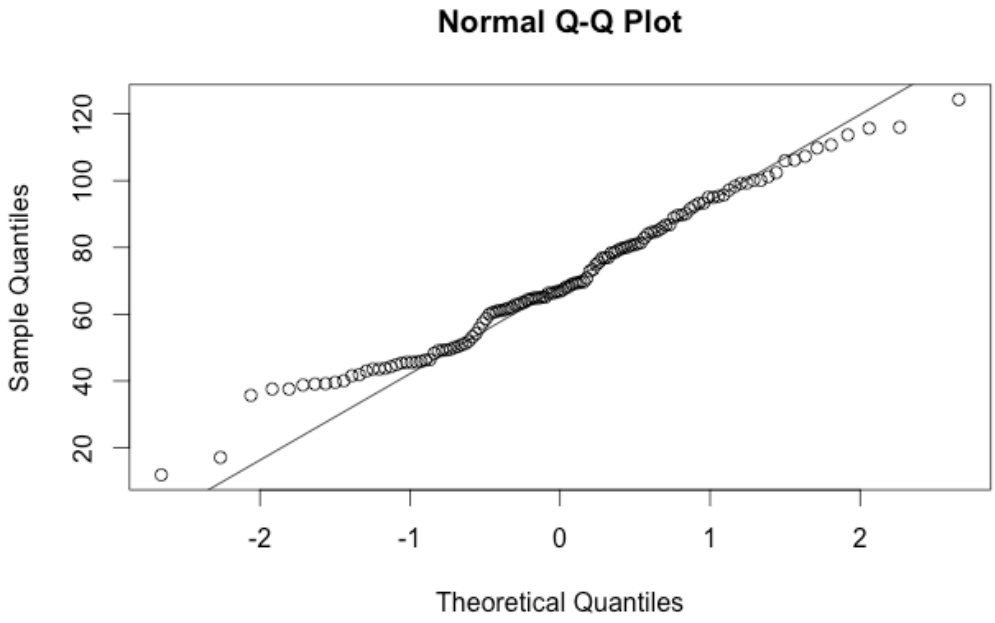

Figure 4.2: Quantile-quantile plot to assess normality: All conjunctions

Furthermore, the results of Levene's Test show that the assumption of homogeneity of variance is met ($F = 0.94$, $p = 0.45$). Since $p > 0.05$, the equal variances assumption cannot be rejected.

Table 4.2 summarises the mean normalised frequency of all conjunctions, by corpus and register. These data are visually represented by the interaction plot in Figure 4.3.

	Creative	Legal
Original	86.6	54.6
Translated	80.3	56.3

Table 4.2: Mean normalised frequency of all conjunctions (per 1000 words), by corpus and register

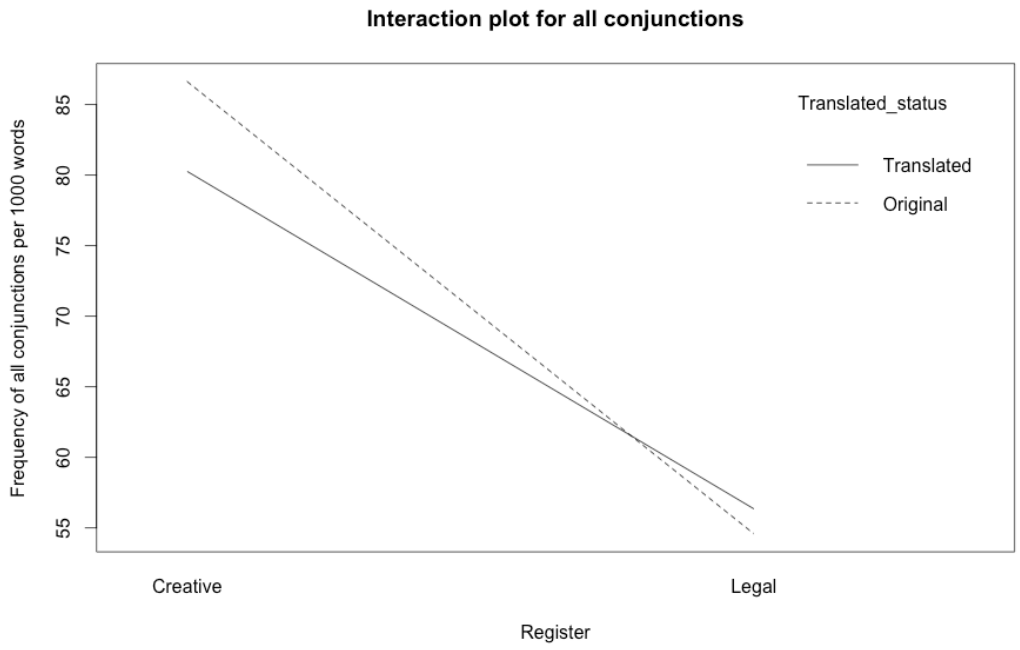

Figure 4.3: Interaction plot (corpus x register) for mean normalised frequency of all conjunctions

Figure 4.3 illustrates a clear main effect for register, with creative texts using conjunctions much more frequently (more than 80 times per 1000 words) than legal texts (around 54-56 times per 1000 words). Overall, therefore, conjunctions are more common in creative than in legal registers in Arabic, irrespective of translated or non-translated status. This is a clear register effect, which confirms the strong register conditioning of conjunction use and is consistent with the literature which states that some registers and genres tend to use more conjunctions than others, such as religious and fiction registers (see Section 2.3.3). The plot in Figure 4.3 also

suggests that there is no levelling-out effect: the register conditioning for conjunction use is equally visible in translated texts as in original texts.

The results of the factorial ANOVA (see Table 4.3) show a strong significant main effect for register (evident in the plot in Figure 4.3), but also a (weaker) but still significant main effect for translated status. There is no significant interaction effect between translation status and register.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Translation status	1	1704	1704	5.80	0.02 *
Register	1	24242	24242	82.46	2.22e-15 ***
Translation status x Register	1	506	506	1.72	0.19

* Signif. codes: 0 '***' / 0.001 '**' / 0.01 '*' / 0.05 '.' / 0.1 ' ' 1.

Table 4.3: Results of factorial ANOVA: All conjunctions

The main effect for register is highly significant (see Table 4.3). A post-hoc t-test confirms that the frequency of conjunctions is significantly different in the legal and creative registers ($t = 9.33$, $p < 0.001$), also clearly evident in the interaction plot in Figure 4.3.

The direction of the main effect for translated status, however, runs counter to the expectations of explicitation: conjunctions are more frequent in original Arabic texts than in translated Arabic texts (as shown in Figure 4.4). In original Arabic texts, conjunctions occur more frequently (73 per 1000 words), compared to translated Arabic texts (66 per 1000 words).

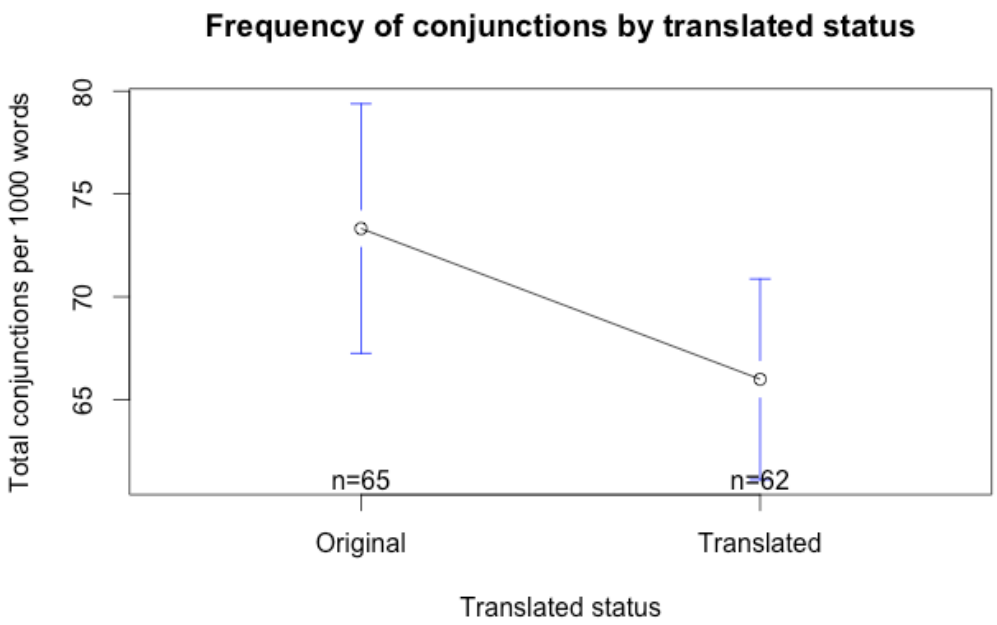

Figure 4.4: Mean normalised frequency of total conjunctions, by translated status

A post-hoc two-sample t-test shows that the difference between the total number of conjunctions, by translated status, approaches statistical significance ($t = 1.88$, $p = 0.06$).

This finding runs counter to previous studies which have suggested that translated language tends to be more explicit than non-translated language, as reflected in the higher frequency of conjunctions in translated language (see Section 2.3.4). Overall, therefore, this study does not find any support for the feature of increased explicitness in translated Arabic texts, as would be reflected in the higher frequency of conjunctions in translated Arabic. A possible explanation for this result may be sought in a type of indirect interference, in that the source-language preferences of English (for the use of less conjunctive language) may be transferred to the Arabic translations, leading to less frequent use of conjunctions in Arabic, compared to non-translated Arabic. However, in order to investigate possible explanations for the overall findings in more detail, it is necessary to investigate individual conjunctions in more detail. This analysis is presented in the following section.

4.4 The effects of translation status and register on the five most frequent conjunctions

This section analyses the five most frequent conjunctions in the corpus: *wa* 'and', *aw* 'or', *thumma* 'then', *lakin* 'but', and *hatta* 'until'. This section aims to shed light on the effects of translation status and register for each of these five conjunctions, and seeks to explain some potential features of translated texts in relation to these five conjunctions.

4.4.1 *wa* 'and'

The conjunction *wa* 'and' is by far the most frequent conjunction in this study, occurring 52.5 times per 1000 words. It has a special status in Arabic (see Section 4.2). Visual inspection of the quantile-quantile plot in Figure 4.5 shows that the data are sufficiently normally distributed for parametric statistics. The results of Levene's Test ($F = 1.24$, $p = 0.3$) show that the assumption of homogeneity of variance is met ($p > 0.05$), where the equal variances assumption cannot be rejected.

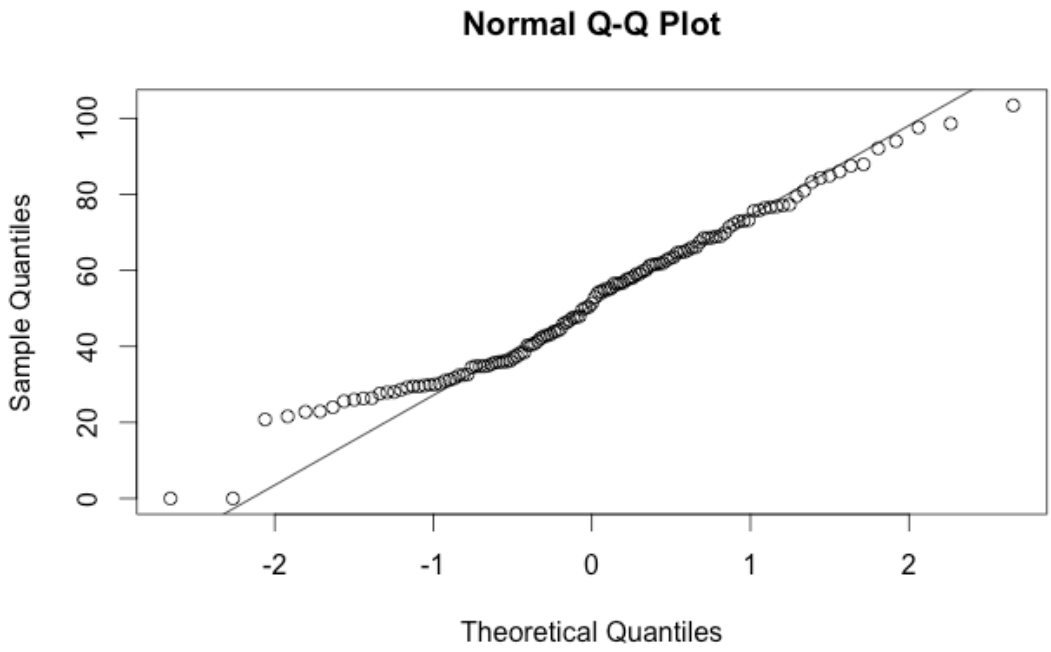

Figure 4.5: Quantile-quantile plot to assess normality: *wa* 'and'

Table 4.4 summarises the mean frequency of *wa* 'and' in the two subcorpora and two registers. These findings are visually represented in the interaction plot in Figure 4.6.

	Creative	**Legal**
Original	66.3	38.8
Translated	58.6	42.6

Table 4.4: Mean normalised frequency of *wa* 'and' (per 1000 words), by corpus and register

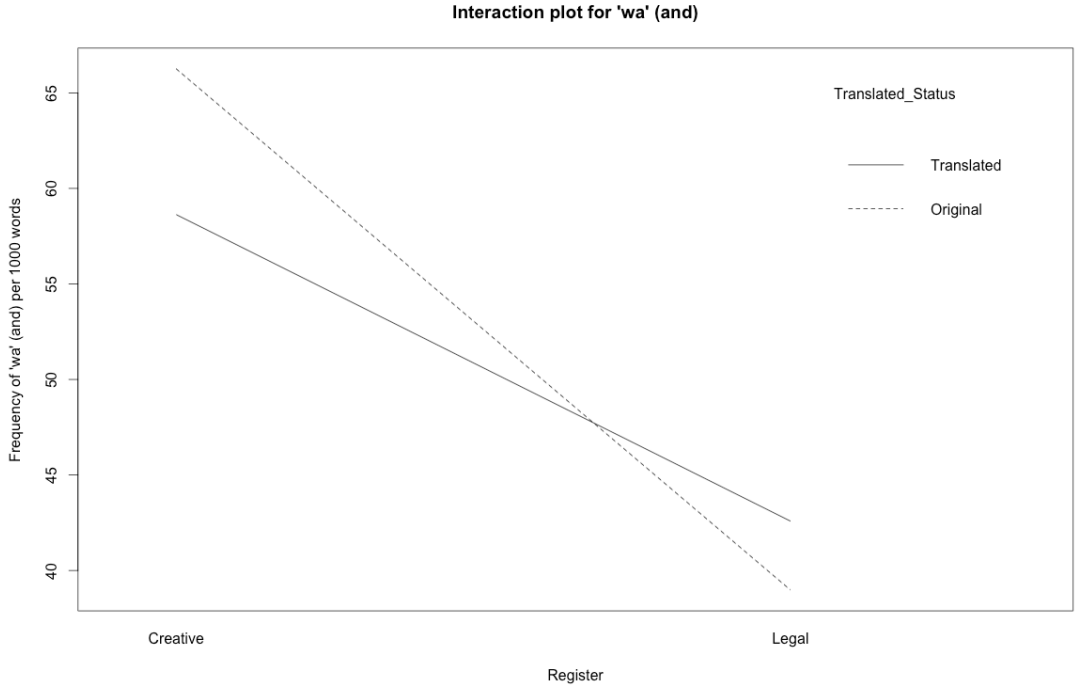

Figure 4.6: Interaction plot (corpus x register) for mean normalised frequency of *wa* 'and'

Once again, there is a strong register effect: *wa* 'and' occurs much more frequently in the creative than in the legal register. *Wa* 'and' is therefore generally more common in (translated and non-translated) creative Arabic texts, as compared to (translated and non-translated) legal Arabic texts. Since *wa* 'and' is considered as the most used conjunctive marker in Arabic, and the fiction texts tend to be more conjunctive than the legal texts, these two factors combine to explain the high frequency of *wa* 'and' in the creative register.

Figure 4.6 shows that there is some differentiation by translated and non-translated status. In the creative register, *wa* 'and' is used more frequently in original (66 per 1000 words) than in translated texts (59 per 1000 words); while in the legal register, translated texts use it more frequently (43 per 1000 words) than original texts (39 per 1000 words). This differentiation may be the result of the way in which translation purpose (or skopos) is affected by register. Entertainment is usually the purpose of the creative register, while the legal register aims to give instructions or

laws. In the translation of legal texts, both the specificity and the accessibility of the translation are of key importance to readers. The risk-avoidance strategy (see Section 2.2.1) may also be more strongly felt in the legal register, leading translators to use *wa* 'and' more frequently in the legal register only. There is therefore evidence of an explicitation effect, only in the translated Arabic legal register.

The results of the factorial ANOVA (see Table 4.5) show a strong main effect for register. However, the main effect for translated status is marginally significant, and there is also an interaction effect approaching significance.

	Df	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
Translation status	1	1096	1096	3.52	0.06 .
Register	1	14623	14623	46.91	3.15e-10 ***
Translation status x Register	1	966	966	3.10	0.08 .

* Signif. codes: 0 '***' / 0.001 '**' / 0.01 '*' / 0.05 '.' / 0.1 ' ' 1.

Table 4.5: Results of factorial ANOVA: *wa* 'and'

The main effect for translated status is, again, in the opposite direction as would be predicted by the explicitation hypothesis – with *wa* 'and' more frequent in original (55 per 1000 words) than in translated (49 per 1000 words) Arabic texts (see Figure 4.7). Again, this may be the result of an indirect interference effect. It is possible that translators underuse *wa* 'and' in comparison to original texts, under the influence of English patterns of conjunction use, which uses *and* as conjunction comparatively less frequently and favours a wider range of conjunctions.

The marginally significant interaction between translation status and register reflects the pattern discussed above in relation to Figure 4.6, and shows that there is a salient difference in how conjunctions are used across the two registers. In the creative register, conjunctions are underused in translated Arabic creative texts

compared to original Arabic creative texts; whereas in the legal register, the opposite is the case. In the latter register, therefore, an explicitation hypothesis is partially confirmed.

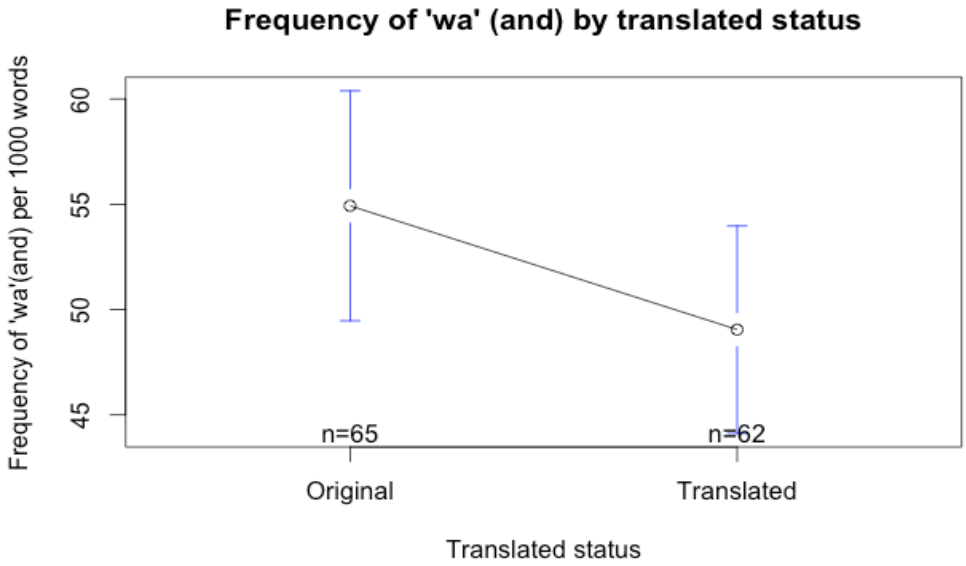

Figure 4.7: Mean normalised frequency of *wa* 'and' by translated status

4.4.2 *aw* 'or'

The conjunction *aw* 'or' occurs as the second most frequent conjunction in this study, at a rate of 4.8 occurrences per 1000 words. As before, a quantile-quantile plot and Levene's Test were used to assess the normality of the data distribution and the homogeneity of variance.

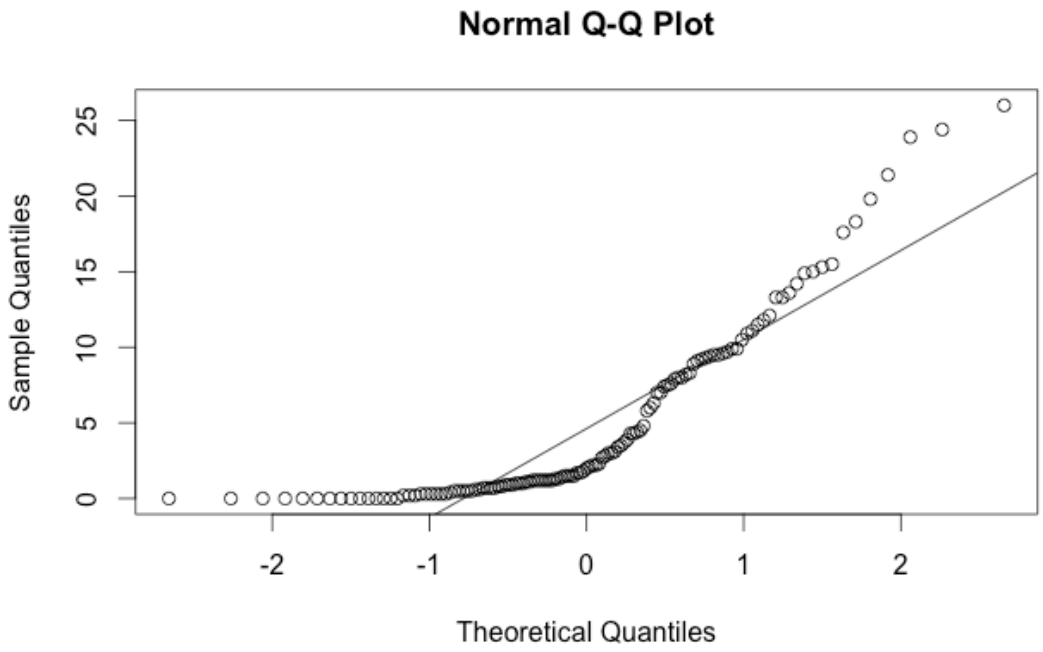

Figure 4.8: Quantile-quantile plot to assess normality: *aw* 'or'

The plot in Figure 4.8 shows considerable deviation from normality. In addition, the results of the Levene's Test ($F = 20.00$, $p < 0.001$) show that the assumption of homogeneity of variance is not met ($p < 0.05$). In this analysis, therefore, medians are reported, and the interaction plot is used to assess any interaction effects. Main effects for translated status and register were tested using the Mann-Whitney *U*-test.

Table 4.6 summarises the median normalised frequency for *aw* 'or' per 1000 words, while Figure 4.9 represents these data visually in the form of an interaction plot.

	Creative	**Legal**
Original	1.0	9.9
Translated	0.5	7.0

Table 4.6: Median normalised frequency of *aw* 'or' (per 1000 words), by corpus and register

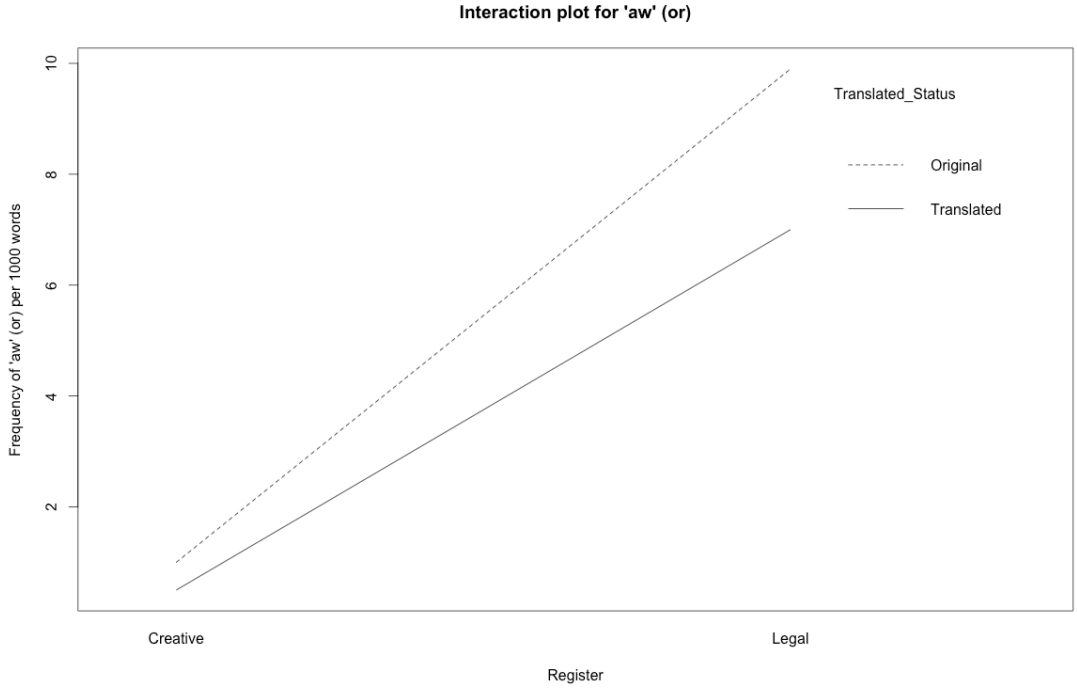

Figure 4.9: Interaction plot (corpus x register) for median normalised frequency of *aw* 'or' per 1000 words

There is a clear register effect. In contrast with *wa* 'and', *aw* 'or' is used much more frequently in the legal register than in the creative register. The boxplot in Figure 4.10 shows that *aw* 'or' is used with a median frequency of 0.8 per 1000 words in creative texts, and 8.6 per 1000 words in legal texts. Certain alternative conjunctions, such as *aw* 'or/alternatively', are clearly much more strongly associated with legal texts, than the more general conjunction *wa* 'and'. This usage is illustrated in the example below.

يقصد بمصطلح " برنامج الحاسوب"، كل مجموعة من التعليمات المعبر عنها بكلمات أو برموز أو برسوم أو بأي طريقة أخرى قابلة لفك رموزها – ان تنجز أو تحقق مهمة محددة، أو تحصل على نتيجة بواسطة حاسوب أو بأي طريقة إلكترونية قادرة على معالجة المعلومات.

(Text: OR-Legal-26)

Translation: *The term "computer program" means that each set of instructions expressed in words, codes, drawings, or by any other way that can be decoded - which is accomplished, achieves a specific task, or gets a result by the computer or by any other electronic way that is able to process the information.*

In this example, there are three conjunctive *aw* 'or' in the sentence, while there are about 104 cases of *aw* 'or' in the entire text.

In the creative register, original and translated texts are relatively similar in frequency (median 1.0 and 0.5 respectively), but in the legal register, original and translated texts diverge: *aw* 'or' is used more frequently in original (median 10 per 1000 words) than in translated texts (median 7 per 1000 words). There appears to be a translation-specific effect evident in the legal register only, such that this conjunction is used less frequently in translated legal texts than in original legal texts.

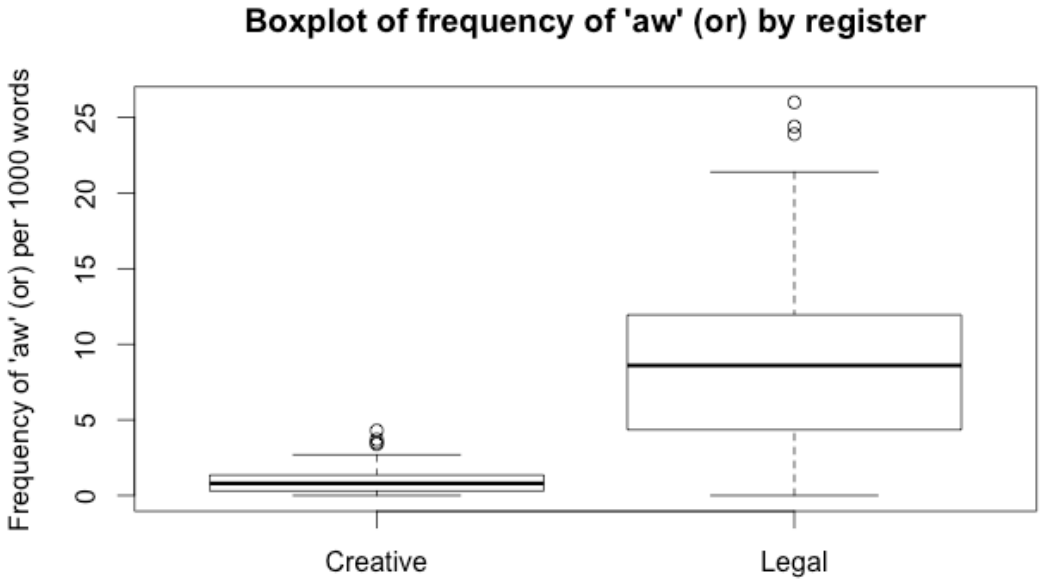

Figure 4.10: Boxplot for frequency of *aw* 'or' by register

A Mann-Whitney U-test shows that the difference between the rank-ordered frequency of *aw* 'or' is significantly affected by register ($W = 362$, $p < 0.001$). However, translation does not have an effect on the rank-ordered frequency of *aw* 'or', overall ($W = 2142$, $p = 0.54$).

4.4.3 *lakin* 'but'

The third most frequent conjunction in this study is *lakin* 'but', which occurs at a rate of 2 per 1000 words.

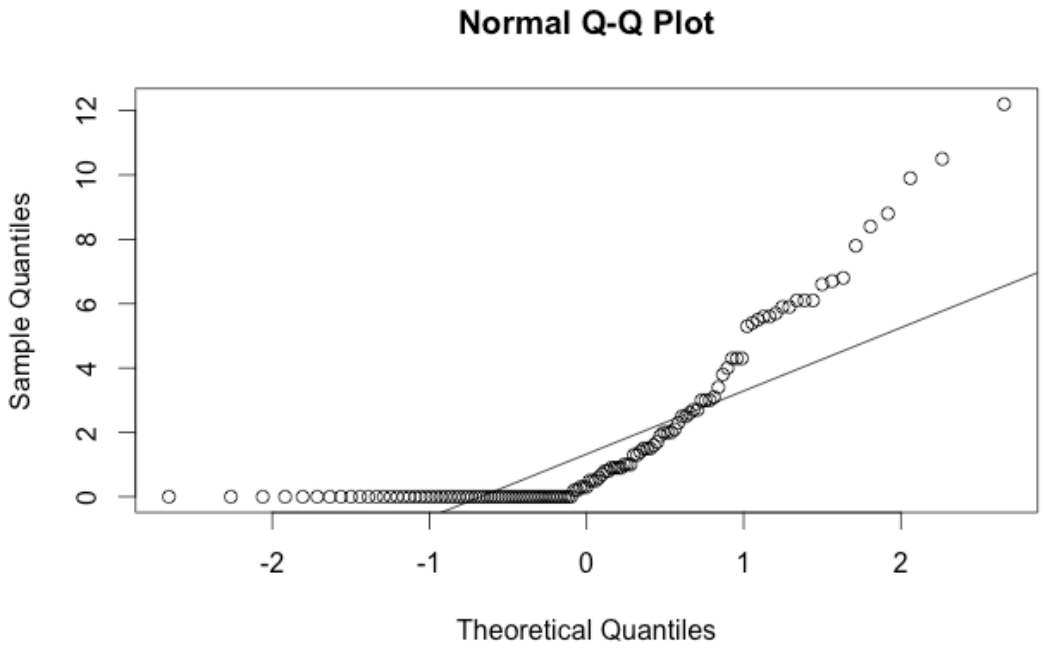

Figure 4.11: Quantile-quantile plot to assess normality: *lakin* 'but'

The quantile-quantile plot in Figure 4.11 shows clear non-normality (also as a result of the large number of 0 cases), and the results of Levene's Test ($F = 20.76$, $p < 0.001$) show that the assumption of homogeneity of variance is not met ($p < 0.05$). In the following discussion, medians are reported, and the interaction plot is used to assess any interaction effects. The Mann-Whitney *U*-test was used to test the main effects for translated status and register.

Table 4.7 summarises the median frequency of *lakin* 'but' per 1000 words. These data are visually represented in the interaction plot in Figure 4.12.

	Creative	Legal
Original	1.8	0
Translated	5.6	0

Table 4.7: Median normalised frequency of *lakin* 'but' (per 1000 words), by corpus and register

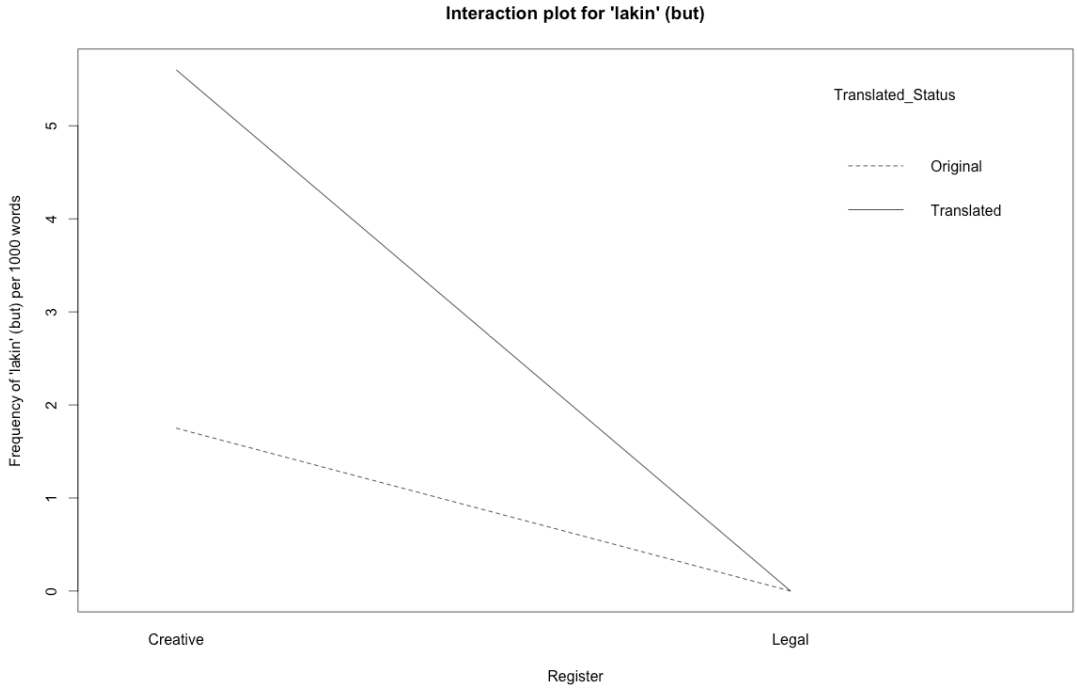

Figure 4.12: Interaction plot (corpus x register) for median normalised frequency of *lakin* 'but'

Once again there is evidence of a strong register effect: *lakin* 'but' is hardly used in the legal register, but more frequently used in the creative register (see also the boxplot in Figure 4.13). This finding can be linked to the adversative function of this conjunction, which includes expressing unexpected contrast, negating a former clause, or connecting ideas with the meaning of the exception, which is more relevant to the creative register than to the legal. This use is illustrated in the example below, which contains three conjunctive *lakin* 'but':

ولكن سأقول لك ما هو أهم؛ عمك مريض لكنه لا يعلم ذلك، ولو ترك الخيار لي لما وافقت على مجيئك الى هنا فانا اعرف كل شيء عنك، وما أعرفه لا يعجبني ولكن (اليستر) زوجي ...

(Text: TR-Creative-13).

Translation: *However, I will tell what is more important; your uncle is sick, but he doesn't know that. If I could, I wouldn't let you come here. I know everything about you, and I don't like what I know. Nevertheless, Allister is my husband…*

In the legal register, there is no differentiation between translated and non-translated texts. In both subcorpora, *lakin* 'but' occurs with a median frequency of 0. However, in the creative register, *lakin* 'but' occurs much more frequently in translated texts (median 5.6 per 1000 words) than in original texts (median 1.8 per 1000 words). There appears to be a clear translation-specific effect evident in the creative register only. The higher frequency of this conjunction in the translated texts in this register supports the hypothesis of explicitation. However, these results may also be the consequence of a transfer effect, if higher frequencies of *but* or other similar conjunctions are transferred from the English source texts to the Arabic translations. Furthermore, there may also potentially be a normalisation effect at play. These interpretations are discussed in more detail, below.

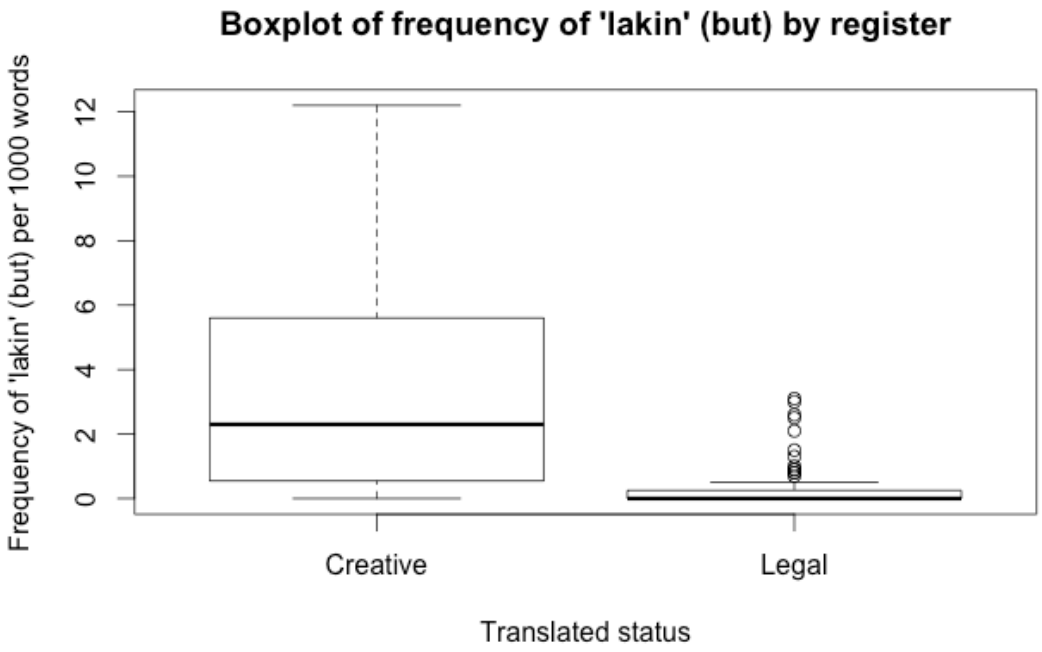

Figure 4.13: Boxplot of frequency of *lakin* 'but' by register

The Mann-Whitney *U*-test shows that the register effect is highly significant (W = 3370.5, $p < 0.001$), while the effect for translated status approaches statistical significance (W = 1659, $p = 0.07$), with the translated texts using *lakin* 'but' more frequently, overall. The boxplot in Figure 4.14 shows the register effect clearly: in creative texts, *lakin* 'but' is used with a median frequency of 2.3 per 1000 words, whereas in the legal register, its median frequency is 0 per 1000 words.

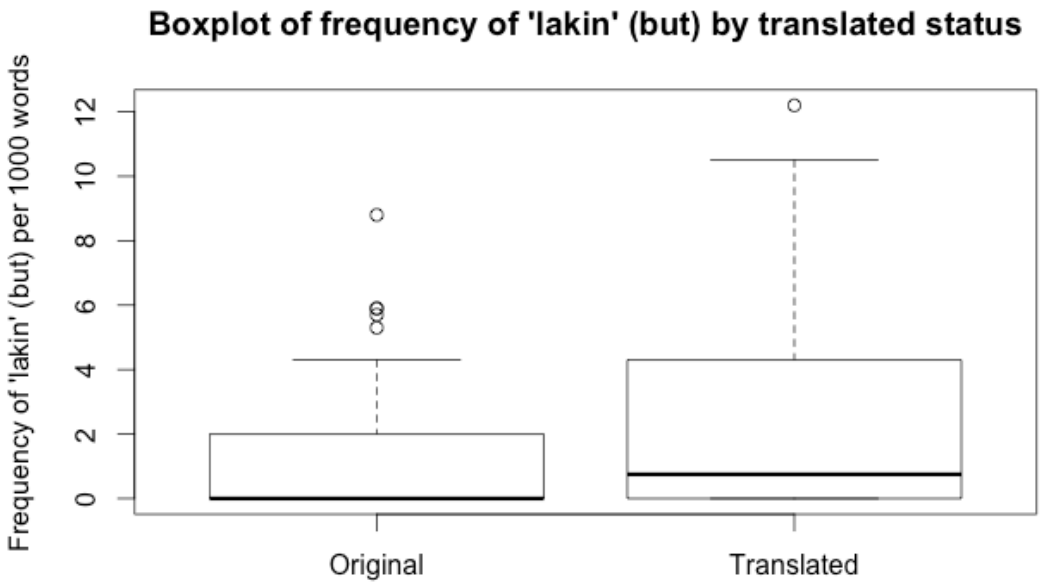

Figure 4.14: Boxplot of frequency of *lakin* 'but' by translated status

The boxplot in Figure 4.14 shows the translation-specific effect. Translated texts use *lakin* 'but' at a median frequency of 0.75 per 1000 words, whereas in original texts, the median frequency is 0 per 1000 words. This finding may be seen to provide support for the explicitation hypothesis (though the interaction plot in Figure 4.12 suggests that this effect may be restricted by register). Nevertheless, the findings for *lakin* 'but' suggest that in some cases, translators do add conjunctions that increase the explicitness with which the relationship between propositions in the text is set out in the translation.

In addition, as mentioned in the literature review (see Section 2.3.3), Arabic tends to use a less varied number of conjunctions with various functions and meanings. For example, the conjunctive marker *lakin* 'but' can have many meanings, reflected in English by different adversative markers such as *but, yet, however, still, nevertheless,* and *on the other hand*. This means that the high frequency of *lakin* 'but' in the translated texts can be explained as the consequence of a normalisation

tendency, in which translators tend to use more conventional or preferred conjunctions in the target language. Consider the example already cited above:

ولكن سأقول لك ما هو أهم؛ عمك مريض لكنه لا يعلم ذلك، ولو ترك الخيار لي لما وافقت على مجيئك الى هنا فانا اعرف كل شيء عنك، وما أعرفه لا يعجبني ولكن (اليستر) زوجي وأحبه كثيرا.

(Text: TR-Creative-13).

Translation: *However, I will tell what is more important; your uncle is sick, but he doesn't know that. If I could, I wouldn't let you come here. I know everything about you, and I don't like what I know. Nevertheless, Allister is my husband…*

There are three *lakin* 'but' in this sentence, where the first one can be translated as *however*, the second as *but*, and the third one as *nevertheless*.

Lastly, there may be an interference effect, if it is the case that preferences for the English use of *but* or related adversative markers are transferred to the Arabic translations, leading to patterns for the use of *lakin* 'but' that are distinct from those in original Arabic texts.

These three competing explanations can, however, not be tested using a comparable corpus design, but requires the use of a parallel corpus, and is therefore foreseen as a future research possibility.

4.4.4 *thumma* 'then'

The conjunction *thumma* 'then' occurs as the fourth most frequent conjunction in this study, at a frequency of just below 2 per 1000 words.

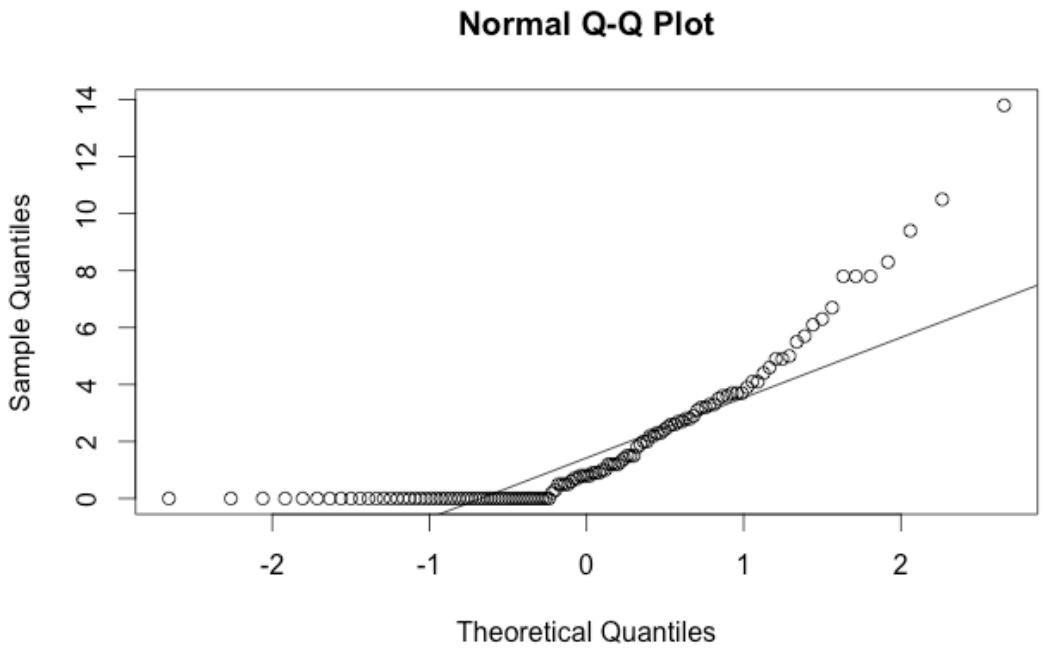

Figure 4.15: Quantile-quantile plot to assess normality: *thumma* 'then'

The quantile-quantile plot in Figure 4.15 shows a clear non-normal distribution (also as a result of many 0 cases). In addition, the results of Levene's Test ($F = 15.6$, $p < 0.001$) show that the assumption of homogeneity of variance is not met ($p < 0.05$). As before, therefore, medians are reported, and an interaction plot is used to assess any interaction effects. The Mann-Whitney *U*-test was used to test for main effects for translated status and register.

Table 4.8 summarises the median frequency of *thumma* 'then' per 1000 words in the two corpora and two registers, while Figure 4.12 visualises these data as an interaction plot.

	Creative	Legal
Original	2.7	0
Translated	3.3	0

Table 4.8: Median normalised frequency of *thumma* 'then' (per 1000 words), by corpus and register

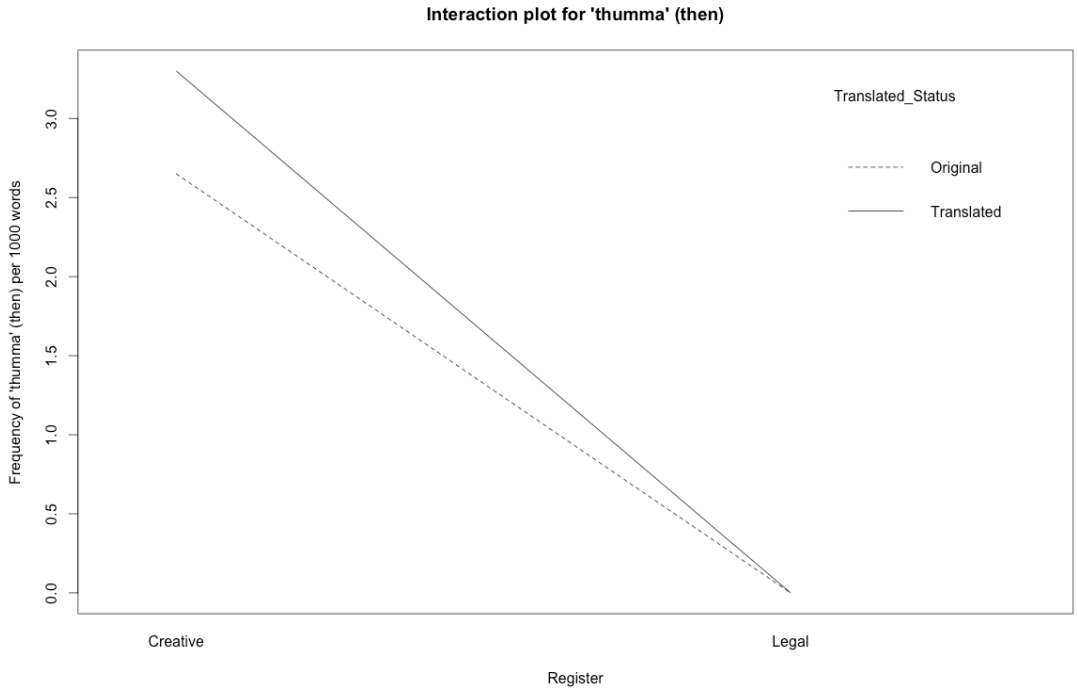

Figure 4.16: Interaction plot (corpus x register) for median normalised frequency of *thumma* 'then'

Once again there is a strong register effect: *thumma* 'then' is hardly used in the legal register, but more frequently used in the creative register (see also the boxplot in Figure 4.13). This register effect for *thumma* 'then' may be linked to the function of this conjunction, which expresses temporal relations between actions or events, and which is more relevant to the creative register, which narrates events. On the

other hand, this function is not as relevant in the legal register, where *thumma* 'then' is hardly used. This function is illustrated in the example below.

ولكنها لم تعرها اهتماما ثم كررت ذلك مرة أخرى بصوت أعلى " جيسيكا هيا بنا " ، فارتبكت جيسيكا ثم قالت : " حسنا هيا بنا " ، ثم مالت جين الى الخارج و هي تقول : " علام تنظرين "؟

(Text: OR-Creative-32)

Translation: *However, she did not pay attention to her, then she asked again, louder, "Jessica, let's go". Jessica got confused, then she said, "Well, let's go".*

Then, Jane leant outside saying, "What are you looking at?"

In this example, there are three (back-translated) conjunctive uses of *thumma* 'then', while there are about 45 *thumma* 'then' in the entire text.

In the legal register, there is no differentiation between translated and non-translated texts (with both subcorpora having a median value of 0). However, in the creative register, *thumma* 'then' occurs more frequently in translated texts (median 3.3) than in original texts (median 2.7). Explicitation is therefore evident in the creative register only, for this particular conjunction. However, this finding may also possibly be attributed to normalisation, with translators overusing the conjunctive preferences of Arabic fiction, in an attempt to adhere to target-language norms.

The Mann-Whitney *U*-test shows that the register effect is significant ($W = 3756$, $p < 0.001$), but not the overall effect for translated status ($W = 2095.5$, $p = 0.69$). Figure 4.17 shows that overall, *thumma* 'then' is used at a median frequency of around 2.9 per 1000 words in creative texts, and 0 per 1000 words in legal texts. The fact that there is no overall effect for translated status cautions that effects such as explicitation or normalisation (evident from the interaction plot above) may only

be visible in some registers, and not others. This is a reminder of the importance of factoring in register in studies of the features of translated language.

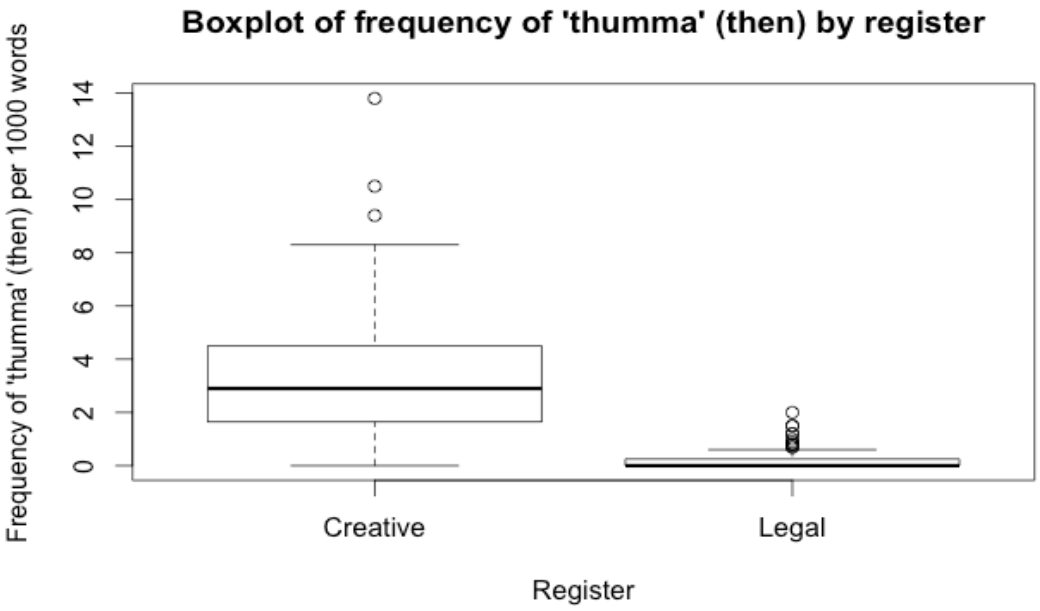

Figure 4.17: Boxplot for frequency of *thumma* 'then' by register

4.4.5 *hatta* 'until'

The fifth most frequent conjunction in this study is *hatta* 'until', occurring 1.7 times per 1000 words. The quantile-quantile plot in Figure 4.18 shows clear non-normal distribution of the data for this conjunction (as a result of many 0 cases), and the results of Levene's Test ($F = 15.8$, $p < 0.001$) show that the assumption of homogeneity of variance is not met ($p < 0.05$).

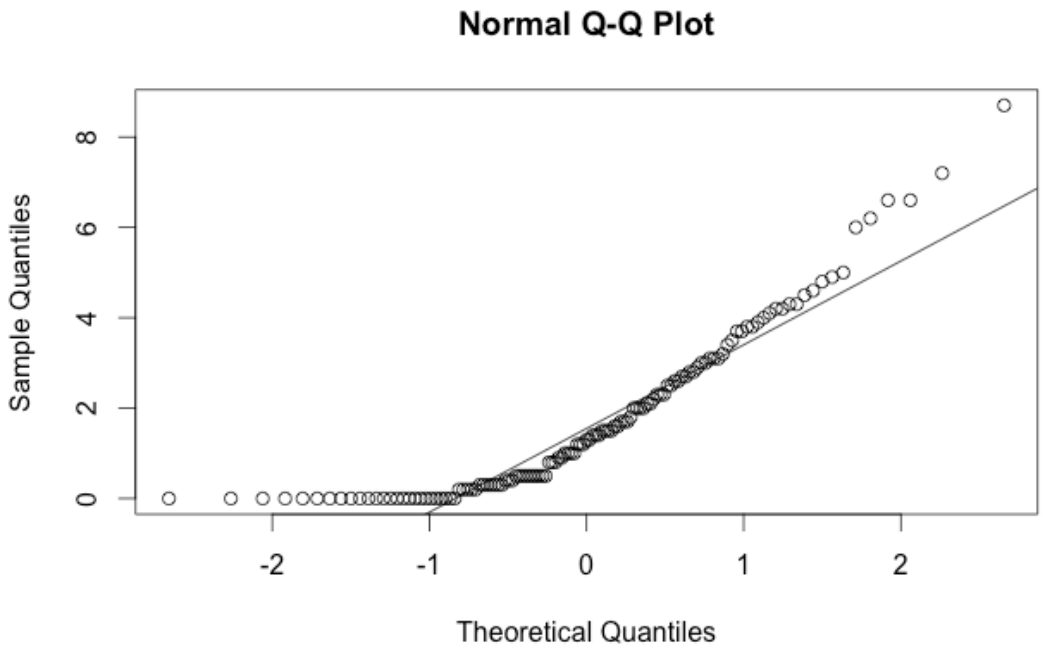

Figure 4.18: Quantile-quantile plot to assess normality: *hatta* 'until'

Table 14.9 summarises the median frequency of *hatta* 'until' per 1000 words, in the two subcorpora and two registers. Figure 4.19 present these data as an interaction plot.

	Creative	**Legal**
Original	3.0	0.2
Translated	2.6	0.5

Table 4.9: Median normalised frequency of *hatta* 'until' (per 1000 words), by corpus and register

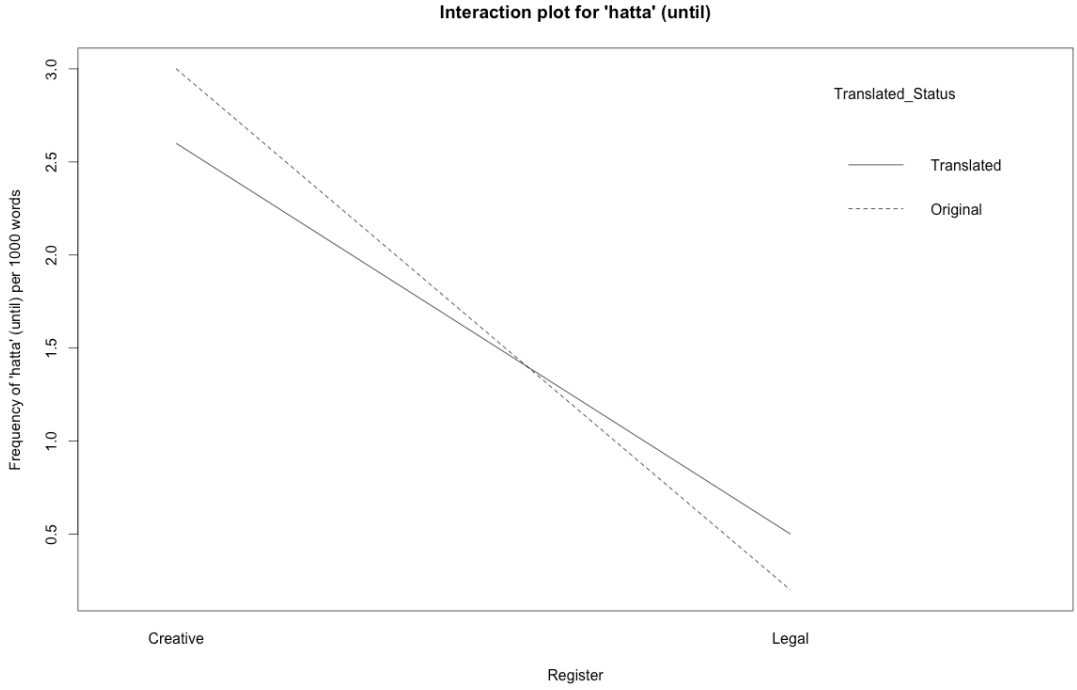

Figure 4.19: Interaction plot (corpus x register) for median normalised frequency of *hatta* 'until'

Once again, there is a clear main effect of register: *hatta* 'until' is used much more frequently in the creative register than in the legal register (see also Figure 4.20). This register effect for *hatta* 'until' can be linked to the resultative function of this conjunction. This function shows the consequences or result of the previous clause, and connects that to the next clause. That is more relevant to the creative register, which usually connects events in such a manner, than to the legal register. This function is illustrated in the example below.

وما إن وصلا ساحة البحيرات حتي اتجها صوب شارع البرلمان المزدحم بالبشر والمحال التجارية المتنوعة، فانقلبت الصورة وأصبحا يستعجلان المشي حتي اقتربا من الكاندلز.

(Text: OR-Creative-25)

Translation: As soon as they arrived at Lakes Square, they headed to the Parliament Street, which is crowded with people and shops. Thus, the image in their minds was turned over, and they started walking faster until they approached The Candles.

In the creative and legal registers, opposite tendencies are evident for translated and non-translated texts. In the creative register, original texts use *hatta* 'until' more frequently (median 3.0 per 1000 words) than translated texts (median 2.6 per 1000 words). In contrast, in the legal register, original texts use *hatta* 'until' less frequently (median 0.2 per 1000 words) than translated texts (median 0.5 per 1000 words).

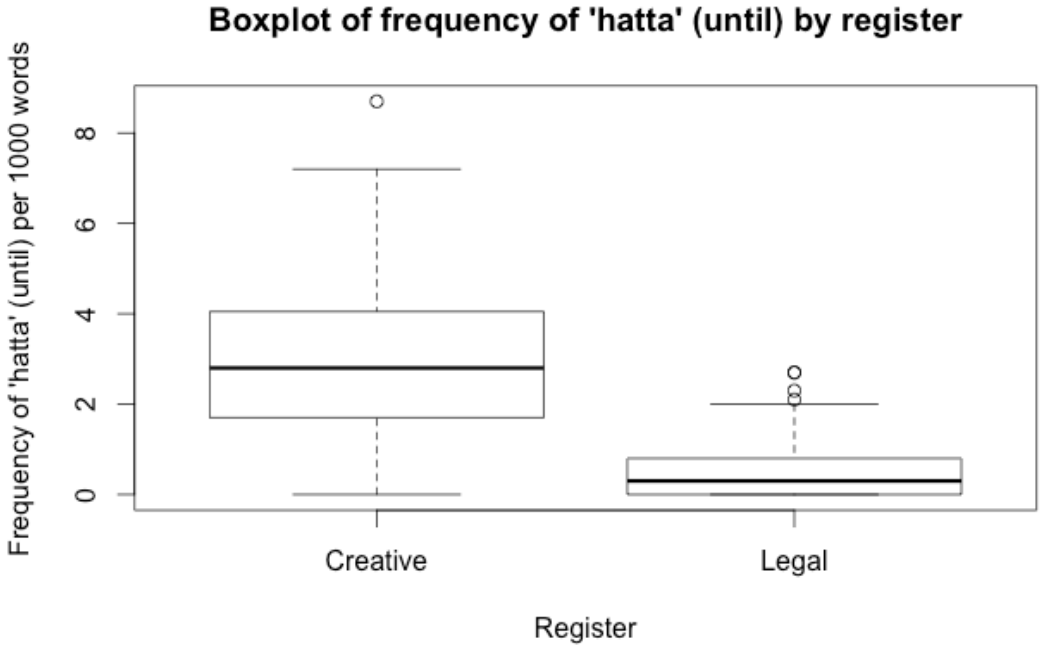

Figure 4.20: Boxplot for frequency of *hatta* 'until', by register

The Mann-Whitney *U*-test shows that the effect for register is significant (W = 3701, $p < 0.001$). As shown in Figure 4.20, the median frequency of *hatta* 'until' in creative texts is 2.8 per 1000 words, dropping to 0.3 in legal texts. However, the effect for translated status is not significant (W = 2156.5, $p = 0.5$).

4.5 Summary and interpretation of findings

The first research question of this study, which concerned the contrastive differences between English and Arabic in respect of conjunction use, was

addressed by the literature review. The findings presented in this chapter answer the second and third research questions, testing the main hypotheses of the study.

A first, important, point to make in respect of the findings of this study is that the importance of register in studies of the features of translated language cannot be underestimated. Across all the analyses presented in this chapter, register effects are by far the strongest, and there is no convincing evidence that translation levels out register differences in preferences for conjunction use. These effects are clearly linked to the functions that conjunctions generally, and individual conjunctions specifically, fulfil in different registers. Moreover, in some cases, translation-specific effects are only evident in particular registers.

Research question 2 set out to identify the difference between translated and non-translated Arabic regarding the use of conjunctive markers, as well as to investigate the potential features of translation that may account for this difference. The overall statistical analysis for all conjunctions showed some unexpected findings for the translation effect, such that the conjunctions investigated were, overall, more frequent in original Arabic texts than in translated Arabic. This finding therefore does not provide support for the notion that translated texts are more explicit than non-translated texts. A potential explanation for this result is the possibility of indirect interference as a consequence of source-language preferences for using fewer conjunctions.

The analyses of the individual conjunctions allow for a more nuanced interpretation of the overall findings. Much of the overall effects discussed in the previous paragraph can be ascribed to the influence of *wa* ‘and’, by far the most frequent conjunction in the corpus. There is an effect for translation status approaching significance – but the effect runs in the opposite direction than would be predicted by the explicitation hypothesis. The potential explanation for this finding is the

same kind of interference effect postulated for the overall findings.

In contrast, translation status has a marginally significant effect on the frequency of *lakin* 'but', which occurs more frequently in translated Arabic texts than in original Arabic texts. This result supports the assumption of explicitation in the translated texts, although explanations involving normalisation and interference can also not be ruled out. For the rest of the individual conjunctions analysed, *aw* 'or', *thumma* 'then', and *hatta* 'until', there was no significant main effect for translation status. These findings therefore confirm that there are some differences between translated and non-translated Arabic in relation to the frequency of conjunctions, and provide some (though not unqualified) support for the first hypothesis proposed for this study. As proposed, these differences may be the consequence of some of the proposed features of translated language, such as interference, explicitation, and normalisation.

Research question 3 aimed to identify the difference in register-related preferences for conjunctive markers, to measure the interaction between translation status and register, and finally to examine some potential features of translation that may be associated with these differences and interactions. In all the analyses, there is a significant main effect for register, providing strong support for the second hypothesis proposed for this study. Except for the case of *aw* 'or', conjunctions occur more frequently in creative texts than in legal texts, and these preferences can be linked to the particular functions of creative narrative texts. On the other hand, the explanation for the frequent use of *aw* 'or' in the legal register may be related to the fact that many legal texts enumerate definitions, types of offences, circumstances of application, and so on, and therefore require the use of *aw* 'or' in Arabic (whose function is realised with punctuation in English).

In respect of the interaction between translation status and register, the analysis of the overall frequency of conjunctions showed no statistically significant interaction effect. This suggests that overall, translated and non-translated texts use conjunctions in similar ways in the creative and legal registers. However, in the analysis of individual conjunctions, there are some potentially interesting findings. For *wa* 'and' there is a marginally significant interaction effect, such that in the legal register, translated Arabic texts use conjunctions more frequently, whereas the opposite is the case in the creative register. An explicitation effect is therefore evident, but only in the legal register. For *aw* 'or', translated legal texts use the conjunction less frequently than non-translated legal texts in Arabic, whereas there is hardly in difference in the creative register. For *thumma* 'then' an opposite pattern is evident; in the creative register, translations use the conjunction more frequently than the non-translated texts, with hardly any difference in the legal register. Here, therefore, the explicitation effect is evident only in translated creative texts. For *lakin* 'but' the same pattern is evident, even more strongly, whereas for *hatta* 'until' the findings are less conclusive, but there is evidence that in legal texts, translations use this conjunction more frequently than non-translations, but in creative texts, the opposite is the case. These findings therefore provide some support for the third hypothesis proposed for this study.

These findings emphasise the importance of accounting for register in studies of the features of translated language. Not only is register an important conditioning variable for conjunction use; some of the features of translated language, such as explicitation, only become visible in particular registers, and for the use of particular conjunctions.

4.6 Conclusion

This chapter has reported the findings of the corpus analysis, which aimed to investigate conjunctive markers in translated and non-translated Arabic texts across two registers. After presenting the findings regarding the normalised frequency of the 20 most frequent conjunctive markers in the first part of this chapter, the effects of translation status and register on the frequency of conjunctions, and individually for the five most frequent conjunctions, were investigated in the third and fourth part of this chapter. Finally, a summary of the findings and interpretation was presented.

Chapter 5: Conclusion

5.1 Introduction

The research questions informing this study were threefold. Firstly, it aimed to explore some contrastive differences in the ways in which conjunctive markers are used in English and in Arabic. Secondly, it intended to investigate whether there are differences in the way in which conjunctive markers are utilised in translated and non-translated Arabic, and whether these differences may be ascribed to the postulated features of translated language, including explicitation, normalisation, simplification, and interference. Third, it aimed to determine register-related preferences for using conjunctive markers in creative and legal texts, and to examine whether register effects are similar and different in translated and non-translated Arabic texts.

Three hypotheses were proposed for the study:

1. Translated and non-translated Arabic texts will demonstrate significant differences regarding the frequency and distribution of conjunctive markers (in other words, there will be a significant main effect for translation status), and these differences can be ascribed to translation-specific processes leading to features such as explicitation, normalisation, interference and simplification.
2. Creative and legal texts will demonstrate significant differences regarding the frequency and distribution of conjunctive markers (in other words, there will be a significant main effect for register).
3. Translated Arabic creative and legal texts will demonstrate a significantly different frequency and distribution of conjunctive markers, compared to original texts in the same registers (in other words, there will be a significant

interaction between translation status and register), as a consequence of translation-specific processes.

This chapter reviews the methodology that was applied in conducting this study (Section 5.2), summarises the findings of the study (Section 5.3), and concludes by discussing some implications of these findings, the limitations of the study, and further avenues of research.

5.2 Review of the methodology

The first research objective of this study set out to explore the contrastive differences between English and Arabic in relation to the concept of conjunction. The literature review presented in Chapter 2 answered this question by providing the necessary background regarding conjunction in both languages, outlining some differences in conjunction use between Arabic and English, and providing the theoretical background to the conjunctive markers to be used in the data analysis presented in Chapter 4.

The second and the third research objectives aimed to explore the differences between translated and non-translated Arabic, as well as register-related differences, in respect of the use and distribution of conjunctive markers. To answer these questions and their sub-questions, which focused on how the proposed features of translated language may account for differences in conjunction use across registers in translated and non-translated Arabic, this study adopted a corpus-based approach using a comparable, register-controlled corpus of translated and non-translated Arabic.

This comparable corpus consisted of translated and non-translated subcorpora across two registers, namely creative and legal texts. This comparable design with

the variance of registers allowed for the quantitative measurement of the effects of translation status and register on the frequency of conjunctions, as well as for an analysis of the interaction of these two factors in their effects on the frequency of conjunctions.

The first subcorpus, of translated and non-translated creative texts, was composed of a selection of stories and extracts of novels, sourced from open-source literary websites. The second subcorpus was composed of translated and non-translated legal documents, sourced from the websites of government departments and international organisations. The subcorpora were comparable in terms of word count and publication time span, which was restricted to texts published from the 1980s to the present.

These texts were electronically collected and analysed using the concordancing software WordSmith Tools 7 (Scott, 2016), in order to extract the data from the corpus. By using the concordance function, the 20 most frequent conjunctive markers (see Table 3.2) were selected as search terms to extract all incidences of these conjunctions. These data were analysed using the statistical software R (Team, 2016) to investigate whether translation status (i.e. translated or non-translated) and register (i.e. creative or legal) had significant effects on the frequency of conjunctions, and whether these two independent variables interacted in any significant way to affect the frequency of conjunctions. This quantitative analysis was followed by a qualitative analysis of the nature of these differences between the subcorpora, to determine whether the findings could be accounted for by means of the recurrent features of translated language, including explicitation, normalisation, interference, levelling-out and simplification.

5.3 Summary of findings

In relation to the first hypothesis, the study found some (though not unqualified) differences in conjunction use between the translated and non-translated subcorpora, with some unexpected findings for the translation effect. One unexpected finding was that conjunctions were, overall, more frequent in original Arabic texts than in translated Arabic. This finding does not support the explicitation hypothesis, but an explanation may be sought in the possibility of indirect interference as a consequence of source-language (English) preferences for using fewer conjunctions.

In relation to the second hypothesis, the study found a very significant and very consistent difference in register-related preferences for conjunctive markers, where conjunctions (with the exception of *aw* 'or') occur more frequently in creative texts than in legal texts. These preferences can be linked to the particular functions of the registers in question.

In relation to the third hypothesis, the findings, overall, demonstrated no statistically significant interaction between translation status and register, which means that in general, translated and non-translated texts use conjunctions in similar ways in the creative and legal registers. This study therefore also finds little support for the feature of levelling-out of register differences in translation. However, the individual analyses of the most frequent five conjunctions in this study found some statistically significant interactions between translation status and register.

Overall, the objectives of this study have been met, by investigating some potential features of translated Arabic using conjunctions as an operationalisation. This study therefore provides further support for the notion of recurrent features of translation, with a specific application of the concept of conjunction in Arabic. Furthermore,

this study also confirms the vital role of register effects in studies of the features of translated language.

5.4 Implications and further suggested research

Following an influential approach in corpus-based translation studies for investigating the features of translated language, this study used a monolingual comparable corpus to examine translated Arabic in relation to non-translated Arabic. This investigation of translated Arabic could conceptually contribute to the generalisability of claims about the universality of translation features, since the Arabic language has substantial structural, stylistic and translational differences compared to other languages previously investigated in studies of the features of translated language. Furthermore, this study is one of very few studies in this language pair that empirically investigates the potential effects of register. While the importance of register in studies of the features of translated languages has been proposed by a number of scholars (Kruger & Van Rooy, 2012; Neumann, 2014), studies are still limited.

The findings of this study in particular, and findings of corpus-based translation studies in general, may also have important implications for the area of translation pedagogy and training. These findings could be used to establish a theoretical base which would be a valuable source of translation strategies for students and trainees. For instance, the findings of this study provide useful background about the register preferences of Arabic conjunctions, and how translated language can be more explicit, simpler, or conventional in its use of conjunctive markers.

However, the findings of this study also suggest that more extensive investigations are required. The MRes thesis imposes both time and space constraints, which has

limited the scope of the study. Furthermore, the type of corpus design places limitations on the analysis, and moreover, this corpus is somewhat limited in size. Lastly, this study investigated conjunction (as one feature of cohesion) only. All these limitations suggest possibilities for expanding the scope of future corpus studies on translated Arabic language.

More sophisticated corpus designs are necessary, which can combine a comparable corpus approach with a parallel corpus methodology. This design would allow for the more definite disentanglement of particularly the role of interference, if translations could also be compared to their source texts. Secondly, a larger range of features could be investigated, by using different linguistic indicators at all language levels as operationalisations, including the lexical, syntactic, semantic, and discourse levels. Features for analysis could include other cohesive devices, lexical variety and density, readability and speakability, and the distribution of typical and atypical register features. Thirdly, examining more registers and investigating texts from different eras of time would produce results that would contribute to the aim of generalisability, which is the ultimate goal of descriptive corpus-based translation studies of the recurrent features of translated language.

References

Al-Amri, K. (2004). *Arabic/English/Arabic translation: Shifts of cohesive markers in the translation of argumentative texts: A contrastive Arabic-English text-linguistic study.* (PhD), University of Durham.

Al-Batal, M. (1990). Connectives as cohesive elements in a modern expository Arabic text. In M. Eid & J. McCarthy (Eds.), *Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics* (Vol. 72, pp. 234-268). Amsterdam and Philadelphia: John Benjamins Publishing.

Al-Jabr, A. (1987). *Cohesion in text differentiation: A study of English and Arabic.* (PhD), Aston University.

Al-Jubouri, A. (1984). The role of repetition in Arabic argumentative discourse. In J. Swales & H. Mustafa (Eds.), *English for specific purposes in the Arab world* (pp. 100-125). Birmingham: University of Aston.

Al-Kashef, Y. (2011). *Cohesion in Arabic into English Translation: A Corpus-Based Investigation of the Translation Universals.* (MA), Ain Shams University, Cairo.

Al-Shabab, O. (1996). *Interpretation and the language of translation: Creativity and conventions in translation*. Edinburgh: Janus Publishing Company.

Al-Sulaiti, L. (2009). Arabic Corpora. Retrieved from http://www.comp.leeds.ac.uk/latifa/arabic_corpora.htm

Alfaifi, A., & Atwell, E. (2016). Comparative evaluation of tools for Arabic corpora search and analysis. *International Journal of Speech Technology, 19*(2), 347-357.

Alsaif, A. (2012). *Human and automatic annotation of discourse relations for Arabic.* PhD thesis, University of Leeds.

Baker, M. (1992). *In other words: a coursebook on translation* (1 ed.). London and New York: Routledge.

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology: in Honour of John Sinclair* (pp. 17-45). Amsterdam: John Benjamins.

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target, 7*(2), 223-243.

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sager* (pp. 175-186). Amsterdam: John Benjamins.

Baker, M. (2004). A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics, 9*(2), 167-193.

Baker, M. (2007). Patterns of idiomaticity in translated vs. non-translated text. *Belgian Journal of Linguistics, 21*(1), 11-21.

Baker, M. (2011). *In other words: A coursebook on translation* (2 ed.). London and New York: Routledge.

Becher, V. (2010). Abandoning the notion of "translation-inherent" explicitation: Against a dogma of translation studies. *Across Languages and Cultures, 11*(1), 1-28.

Becher, V. (2011). When and why do translators add connectives? A corpus-based study. *Target, 23*(1), 26-47.

Beeston, A. (1983). The role of parallelism in Arabic prose. In A. Beeston, T. Johnstone, R. Serjeant, & S. G (Eds.), *Arabic Literature to the End of the Umayyad Period* (pp. 180-185). London: Cambridge University Press.

Bernardini, S. (2007). Collocations in translated language: combining parallel, comparable and reference corpora. *Proceedings of the fourth Corpus Linguistics conference in the University of Birmingham* (pp. 27-30). Birmingham: University of Birmingham.

Bernardini, S., & Ferraresi, A. (2011). Practice, description and theory come together: Normalization or interference in Italian technical translation? *Meta: Journal des traducteursMeta:/Translators' Journal, 56*(2), 226-246.

Biber, D. (1988). *Variation across speech and writing*. New York: Cambridge University Press.

Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics, 19*(2), 219-241.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*: Cambridge: Cambridge University Press.

Biber, D., & Finegan, E. (1994). *Sociolinguistic perspectives on register*. Oxford: Oxford University Press.

Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-Kulka (Eds.), *Interlingual and Intercultural Communication* (pp. 17-35). Tübingen: Narr.

Bowker, L. (2002). *Computer-aided translation technology: A practical introduction*. Ottawa: University of Ottawa Press.

Bystrova-McIntyre, T. (2012). *Cohesion in translation: A corpus study of human-translated, machine-translated, and non-translated texts (Russian into English)*. (PhD), Kent State University.

Callow, K. (1974). *Discourse considerations in translating the Word of God*. Grand Rapids: Zondervan.

Cantarino, V. (1975). *Syntax of modern Arabic prose*. Bloomington: Indiana University Press for the International Affairs Center.

Chesterman, A. (1998). Causes, translations, effects. *Target, 10*(2), 201-230.

Chesterman, A. (2000). A causal model for translation studies. In M. Olohan (Ed.), *Intercultural Faultlines. Research Models in Translation Studies I: Textual and Cognitive Aspects* (pp. 15-27). Manchester: St. Jerome Publishing.

Chesterman, A. (2004). Beyond the particular. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals: Do They Exist?* (pp. 33-49). Amsterdam: John Benjamins Publishing.

Dayrell, C. (2007). A quantitative approach to compare collocational patterns in translated and non-translated texts. *International Journal of Corpus Linguistics, 12*(3), 375-414.

De Beaugrande, R., & Dressler, W. (1981). *Introduction to Text Linguistics*. London & New York: Longman.

De Sutter, G., Delaere, I., & Plevoets, K. (2012). Lexical lectometry in corpus-based translation studies. In M. Oakes & M. Ji (Eds.), *Quantitative methods in corpus-based translation studies: a practical guide to descriptive translation research* (pp. 325-345). Amsterdam/Philadelphia: John Benjamins.

Delaere, I., De Sutter, G., & Plevoets, K. (2012). Is translated language more standardized than non-translated language?: Using profile-based correspondence analysis for measuring linguistic distances between language varieties. *Target, 24*(2), 203-224.

Djamila, L. (2010). *Shifts in translating lexical cohesion from Arabic into English.* Master in Applied Language Studies thesis, University of Mentouri Constantine.

Duff, A. (1981). *The third language: Recurrent problems of translation into English: It ain't what you do, it's the way you do it*. Oxford/New Yourk: Pergamon Press.

Eddakrouri, A. (2016). Web-based (Searchable) corpora. Retrieved from https://sites.google.com/a/aucegypt.edu/infoguistics/directory/Corpus-Linguistics/arabic-corpora

Eid, M. (1975). *Al Nahw Al Mussafah [The Refined Grammar]*. Cairo: Maktabat al-shabab.

Ellis, N. (1996). Sequencing in SLA. *Studies in second language acquisition, 18*(01), 91-126.

Eskola, S. (2004). Untypical frequencies in translated language: A corpus-based study on a literary corpus of translated and non-translated Finnish. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals: Do They Exist?* (pp. 83-100). Amsterdam: John Benjamins Publishing.

Fattah, A. (2010). *A corpus-based study of conjunctive explicitation in Arabic translated and non-translated texts written by the same translators/authors.* (PhD), University of Manchester, Manchester, UK.

Frawley, W. (1984). Prolegomenon to a Theory of Translation. In W. Frawley (Ed.), *Translation: Literary, Linguistic, and Philosophical Perspectives* (pp. 159-175). London: Associated University Presses.

Gellerstam, M. (1986). Translationese in Swedish Novels Translated from English. In L. Wollin & H. Lindquist (Eds.), *Translation studies in Scandinavia* (pp. 88-95). Lund: CWK Gleerup.

Ghadessy, M. (1988). *Registers of Written English: Situational Factors and Linguistic Factors*. London: Pinter Publishers.

Ghadessy, M. (1993). *Register analysis: Theory and practice*. London: Pinter Publishers.

Gregory, M., & Carroll, S. (1978). *Language and situation: Language varieties and their social contexts*: London: Routledge.

Grosjean, F., & Soares, C. (1986). Processing mixed language: Some preliminary findings. In J. Vaid (Ed.), *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives* (pp. 145-179). Hillsdale, New Jersey: Lawrence Erlbaum.

Gutwinski, W. (1976). *Cohesion in literary texts: A study of some grammatical and lexical features of English discourse*. Berlin: Walter de Gruyter.

Halliday, M., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M., & Matthiessen, C. (2004). *An introduction to functional grammar*. London: Arnold.

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

Halverson, S. (2015). Cognitive Translation Studies and the merging of empirical paradigms: The case of 'literal translation'. *Translation Spaces, 4*(2), 310-340.

Hansen-Schirra, S. (2011). Between normalization and shining-through. Specific properties of English-German translations and their influence on the target language. In S. Kranich, V. Becher, H. d. S, & J. House (Eds.), *Multilingual Discourse Production: Diachronic and Synchronic Perspectives* (pp. 133-162). Amsterdam: John Benjamins Publishing.

Hansen-Schirra, S., Neumann, S., & Steiner, E. (2007). Cohesive explicitness and explicitation in an English-German translation corpus. *Languages in Contrast, 7*(2), 241-266.

Hassan, T. (1979). *Allughah Alarabiyyah: Ma'naahaa wa Mabnaahaa [Arabic Language: Its Structure and Meaning]*. Cairo: General Egyptian Book Organization.

Hill, T., & Lewicki, P. (2006). *Statistics: methods and applications*. Cape Town: StatSoft.

Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.

Holes, C. (1995). *Modern Standard Arabic: Structure, functions and varieties* (1 ed.). London/New Yourk: Longman.

Holes, C. (2004). *Modern Standard Arabic: Structure, functions and varieties*(2 ed.). Washington: Georgetown UP.

Jantunen, J. (2004). Untypical patterns in translations: Issues on corpus methodology and synonymity. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals: Do They Exist?* (pp. 101-128). Amsterdam / Philadelphia: John Benjamins Publishing.

Jiménez-Crespo, M. (2011). The future of general tendencies in translation: Explicitation in web localization. *Target, 23*(1), 3-25.

Kenny, D. (2001). *Lexis and creativity in translation: A corpus-based study.* Manchester: St. Jerome Publishing.

Klaudy, K. (2008). Explicitation. In M. Baker, & Saldanha, G. (Ed.), *Routledge Encyclopaedia of Translation Studies* (pp. 104-108). London: Routledge.

Kruger, & Van Rooy, B. (2012). Register and the features of translated language. *Across Languages and Cultures, 13*(1), 33-65.

Kruger, A. (2004). Corpus-based translation research: Its development and implications for general, literary and Bible translation. *Acta Theologica, 22*(1), 70-106.

Kruger, H. (2012). A corpus-based study of the mediation effect in translated and edited language. *Target, 24*(2), 355-388.

Kruger, H., & Van Rooy, B. (2016). Syntactic and pragmatic transfer effects in reported-speech constructions in three contact varieties of English influenced by Afrikaans. *Language Sciences, 56*, 118-131.

Laviosa, S. (1998). The corpus-based approach: A new paradigm in translation studies. *Meta: Journal des traducteurs / Meta: Translators' Journal, 43*(4), 474-479.

Laviosa, S. (2002). *Corpus-based translation studies: Theory, findings, applications*. Amsterdam: Rodopi.

Lee, D. (2002). Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language and Computers, 42*(1), 247-292.

Lulu, R. (2013). *Grammatical cohesion in the English to Arabic translation of political texts.* Master of Linguistics thesis, University of Malaya.

Mauranen, A. (2000). Strange Strings in Translated Language: A Study on Corpora. In M. Olohan (Ed.), *Intercultural Faultlines: Research Models in Translation Studies 1: Textual and Cognitive Aspects* (pp. 119–141). Manchester: St Jerome Publishing.

Mauranen, A. (2004). Corpora, Universals and Interference. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals Do they exist?* (pp. 65-82). Amsterdam / Philadelphia: John Benjamins Publishing.

Mauranen, A., & Kujamäki, P. (Eds.). (2004). *Translation universals: Do they exist?* Amsterdam / Philadelphia: John Benjamins Publishing.

Mehamsadji, M. (1988). *Cohesion and text development in written Arabic.* (PhD), University of Salford, Manchester, UK.

Mutesayire, M. (2004). Apposition markers and explicitation: A corpus-based study. *Language Matters, 35*(1), 54-69.

Najjar, I. (2015). 'Repetition' in Arabic-English Translation: The case of Adrift on the Nile. *International Journal of Foreign Language Teaching and Research, 3*(10), 24-34.

Neumann, S. (2014). Cross-linguistic register studies: Theoretical and methodological considerations. *Languages in Contrast, 14*(1), 35-57.

Oakes, M., & Ji, M. (2012). *Quantitative methods in corpus-based translation studies: A practical guide to descriptive translation research*. Amsterdam: John Benjamins Publishing.

Olohan, M. (2004). *Introducing corpora in translation studies*. London: Routledge.

Olohan, M., & Baker, M. (2000). Reporting THAT in translated English. Evidence for subconscious processes of explicitation? *Across Languages and Cultures, 1*(2), 141-158.

Øverås, L. (1998). In search of the third code: An investigation of norms in literary translation. *Meta: Journal des traducteurs / Meta: Translators' Journal, 43*(4), 557-570.

Pápai, V. (2004). Explicitation: A universal of translated text? In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals: Do They Exist?* (pp. 143-164). Amsterdam / Philadelphia: John Benjamins Publishing.

Pastor, G. (2008). *Investigar con corpus en traducción: los retos de un nuevo paradigma* (Vol. 49). Bern: Peter Lang.

Puurtinen, T. (2004). Explicitation of clausal relations: A corpus-based analysis of clause connectives in translated and non-translated Finnish children's literature. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals: Do They Exist?* (pp. 165-176). Amsterdam / Philadelphia: John Benjamins Publishing.

Pym, A. (2005). Explaining Explicitation. In K. Karoly & A. Fóris (Eds.), *New Trends in Translation Studies. In Honour of Kinga Klaudy* (pp. 29-34). Budapest: Akadémia Kiadó.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/.

Redelinghuys, K. (2013). *Using the features of translated language to investigate translation expertise: A corpus-based study.* MA thesis, North-West University.

Redelinghuys, K., & Kruger, H. (2015). Using the features of translated language to investigate translation expertise: A corpus-based study. *International Journal of Corpus Linguistics, 20*(3), 293-325.

Ryding, K. (2005). *A reference grammar of modern standard Arabic*. Cambridge: Cambridge University Press.

Scott, M. (2016). WordSmith tools (Version 7). Stroud: Lexical Analysis Software.

Shamaa, N. (1978). *A linguistic analysis of some problems of Arabic to English translation.* (PhD), University of Oxford, Oxford, UK.

Smith, R., & Frawley, W. (1983). Conjunctive cohesion in four English genres. *Text-Interdisciplinary Journal for the Study of Discourse, 3*(4), 347-374.

Thomason, S., & Kaufman, T. (2001). *Language contact*. Edinburgh: Edinburgh University Press.

Tirkkonen-Condit, S. (2004). Unique items – over- or under-represented in translated language? In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals: Do They Exist?* (pp. 177-186). Amsterdam / Philadelphia: John Benjamins Publishing.

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

Toury, G. (1995). *Descriptive translation studies and beyond*. Amsterdam: John Benjamins Publishing.

Toury, G. (2012). *Descriptive translation studies and beyond*. Amsterdam: John Benjamins Publishing.

Vanderauwera, R. (1985). *Dutch novels translated into English: The transformation of a "minority" literature*. Amsterdam: Rodopi.

Vinay, J., & Darbelnet, J. (1958). *Comparative Stylistics of French and English: A Methodology for Translation, translated and edited by Sager, J. and Hamel, M.* Amsterdam/Philadelphia: John Benjamins Publishing.

Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.

Williams, D. (2005). *Recurrent features of translation in Canada: a corpus-based study.* (PhD), University of Ottawa, Ottawa.

Williams, J., & Chesterman, A. (2011). *The map: A beginner's guide to doing research in translation studies*. London: Routledge.

Williams, M. (1989). *A comparison of the textual structures of Arabic and English written texts: A Study in the comparative orality of Arabic.* PhD thesis, University of Leeds.

Wright, W. (1967). *A grammar of the Arabic language*. Cambridge: Cambridge University Press.

Xiao, R. (2010). How different is translated Chinese from native Chinese? A corpus-based study of translation universals. *International Journal of Corpus Linguistics, 15*(1), 5-35.

Xiao, R., He, L., & Yue, M. (2010). In pursuit of the third code: Using the ZJU corpus of translational Chinese in translation studies. In R. Xiao (Ed.), *Using corpora in contrastive and translation studies* (pp. 182-214). Newcastle, UK: Cambridge Scholars.

Zanettin, F. (2013). Corpus methods for descriptive translation studies. *Procedia: Social and Behavioral Sciences, 95*, 20-32.

Zanettin, F. (2014). *Translation-driven corpora: Corpus resources for descriptive and applied translation studies*. London: Routledge.

Appendix

Creative register sources

1. Translated subcorpus

Literary work	Author	Publication year	Translator	Translation year
A Perfect Stranger	Danielle Steel	1994	Thyab Rajab	2015
chill factor	Sandra Brown	2005	Thyab Rajab	2015
Pride and Prejudice	Jane Austen	1813	Ameen Salamah	1996
Twilight	Stephenie Meyer	2005	Alhareth Nabhan	2009
The Black magic	Christina Vitoyo	-	Edouard Abu Nasr	2003
unexpected love	Mary Rock	2003	Kortam	2013
His pretend mistress	Jessica Steele	2002	Alfarasha Publishing House	2003
The Melting Heart	Claudia Jameson	1983	Alfarasha Publishing House	1999
The Black Tulip	Alexandre Dumas	-	Shoa'a for publishing & Science	2008
The Big Four	Agatha Christie	1927	Jareer Bookstore	2011
Unwanted Wedding	Penny Jordan	1996	Alfarasha Publishing House	2001
Chateau of Flowers	Margaret Rome	1971	Alnahas Publishing House	2009
Lord of the Island	Mary Wibberley	1978	Alnahas Publishing House	2008
By Love Alone	Kathryn Ross	1992	Alfarasha Publishing House	2012
So Much to Tell You	John Marsden	1987	Yousf Sahari	2014

Literary work	Author	Publication year	Translator	Translation year
The Marrying Game	Lindsey Armstrong	1989	Alfarasha Publishing House	2007
Turbulent Covenant	Jessica Steele	1980	Alfarasha Publishing House	2000
Engaged to Jarrod Stone	Carole Mortimer	1980	Alfarasha Publishing House	2003
Tidewater Lover	Janet Dailey	2001	Alfarasha Publishing House	2008
Terms of Possession	Elizabeth Power	1995	Alnahas Publishing House	2011
Summer Mahogany	Janet Dailey	1979	Alnahas Publishing House	2010
Return to Silbersee	Jane Arbor	1979	Alnahas Publishing House	2009
Rules of the Game	Penny Jordan	1985	Alnahas Publishing House	2014
Don't Play Games	Emma Darcy	1985	Alnahas Publishing House	2010
A Taste for Rich Things	Joan Hohl	1985	Alnahas Publishing House	2010

2. Non-translated subcorpus

Author	Publication year	URL
Mohammad Abdulmalik	2001	http://www.arabicstory.net/?p=text&tid=85
Suad Khalifah	2001	http://www.arabicstory.net/?p=text&tid=87
Abdoh Khal	2001	http://www.arabicstory.net/?p=text&tid=228
Jamal Alkhayat	2003	http://www.arabicstory.net/?p=text&tid=88
Ghalia Gabbani	2003	http://www.ArabicStory.net
Khalaf Ahmad	2003	http://www.arabicstory.net/?p=text&tid=151
Aysha Ghaloom	2003	http://www.arabicstory.net/?p=text&tid=3134
Aysha Ghaloom	2003	http://www.arabicstory.net/?p=text&tid=3134
Sara Alnawaf	2003	http://www.ArabicStory.net
Suad Khalifah	2001	http://www.arabicstory.net/?p=text&tid=2779
Khalaf Ahmad	2003	http://www.arabicstory.net/?p=text&tid=151
Sara Alnawaf	2003	http://www.ArabicStory.net
Ghalia Gabbani	2002	http://www.ArabicStory.net
Naima Sammak	2006	http://www.arabicstory.net/?p=text&tid=8410
Anisah Azayani	2004	http://www.arabicstory.net/?p=text&tid=4258
Abdoh Khal	2001	http://www.arabicstory.net/?p=text&tid=743
Suad Khalifah	2003	http://www.arabicstory.net/?p=text&tid=2863
Abdoh Khal	2001	http://www.arabicstory.net/?p=text&tid=227
Suad Khalifah	2003	http://www.arabicstory.net/?p=text&tid=2971
Jamal Alkhayat	2003	http://www.arabicstory.net/?p=text&tid=2925

Author	Publication year	URL
Saad Alajmi	2011	http://www.rewity.com/forum/t338162.html
Jamal Alkhayat	2003	http://www.arabicstory.net/?p=text&tid=2923
Mona Almarshood	2007	http://www.rewity.com/forum/t304038.html
Yaseen Refaiah	1985	http://www.rewity.com/forum/t150491.html
Jamal Alkhayat	2003	http://www.arabicstory.net/?p=text&tid=2924
Samya Ahmad	2009	http://www.rewity.com/forum/t282107.html
Shareef Mustafa	2014	http://www.rewity.com/forum/t298236.html
Yahya Saleem	2009	http://www.rewity.com/forum/t296424.html
Noha Talaba	2015	http://www.rewity.com/forum/t332076.html
Khalid Abu Ashamat	2006	http://www.rewity.com/forum/t321821.html
Hawra Alomairi	2015	http://www.rewity.com/forum/t326196.html
Mahmoud Ghassan	2011	http://www.rewity.com/forum/t335320.html
Mohammad Dawood	2009	http://www.rewity.com/forum/t300461.html
Elias Khori	2007	http://www.rewity.com/forum/t327388.html
Zaid Dammaj	1984	http://www.rewity.com/forum/t292463.html
Abdulaziz Saken	2012	http://www.rewity.com/forum/t313332.html
Malek Alghamedi	2014	http://www.rewity.com/forum/t295048.html
Ibrahim Nasrallah	2009	http://www.rewity.com/forum/t153505.html