Routledge Studies in Empirical Translation

CORPUS TRIANGULATION

COMBINING DATA AND METHODS IN CORPUS-BASED TRANSLATION STUDIES

Sofia Malamatidou

Corpus Triangulation

Despite the recognition that corpus-based translation research would benefit from the triangulation of corpora, little has been done in the direction of actually employing combined corpus data and methods in the field. This book aims to address this gap by providing a much-needed detailed account of corpus triangulation, where different corpora (e.g. parallel, comparable, synchronic, diachronic) and/or different methods of analysis (e.g. qualitative, quantitative) can be used to increase our understanding of the phenomena where translation plays a key role. The book also demonstrates clearly how the proposed methodology can be fruitfully employed to investigate different linguistic features, through its systematic application to empirical data. The first part of the book introduces the innovative framework for corpus triangulation, which is based on a new and comprehensive corpus typology, while the second part applies the methodological framework to two case studies examining the language of translation and the relationship between translation and language change. The book advances current translation studies in terms of methodology innovation and offers a model on which future studies investigating the network of relationships surrounding translated texts can be based.

Sofia Malamatidou is a Lecturer in Translation Studies at the University of Birmingham, UK. Her main research interests are in the fields of corpus linguistics, translation studies and contact linguistics. She has published papers in international journals (*The Translator*, *Target*, *Meta*) and edited volumes.

Routledge Studies in Empirical Translation

Series Editor: Meng Ji, *The University of Sydney, Australia*

For a full list of titles in this series, visit www.routledge.com/Routledge-Studies-in-Empirical-Translation/book-series/RSET

Empirical Translation Studies (ETS) represents a rapidly growing field of research which came to the fore in the 1990s. From the early tentative use of computerised translation to the systematic investigation of large-scale translation corpora by using quantitative/statistical methods, ETS has made substantial progress in the development of solid empirical research methodologies which lie at the heart of the further development of the field. There is a growing volume of research pursued in ETS as corpus translation studies has become a core component of translation studies at the postgraduate and research levels. To offer an appropriate and much-needed outlet for high-quality research in ETS, this proposed book series intends to select and publish latest translation research around the world, in which the innovative use of corpus materials and related methodologies is essential. An important shared feature of the manuscripts accepted is their original contribution made to the advancement of empirical methodologies in translation studies which includes but is not limited to the quantitative/statistical processing, modelling and interpretation of translation corpora.

1 Corpus Triangulation
Combining Data and Methods in Corpus-Based Translation Studies
Sofia Malamatidou

Corpus Triangulation

Combining Data and Methods in Corpus-Based Translation Studies

Sofia Malamatidou

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2018
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book has been requested

ISBN: 978-1-138-94850-1 (hbk)
ISBN: 978-1-315-66959-5 (ebk)

Typeset in Galliard
by Apex CoVantage, LLC

To
E
&
N
the two most important letters in my life

Contents

Figures

Tables

Acknowledgements

I would like to express my profound gratitude to the many people who saw me through this book. To my family, friends and colleagues who talked things over, supported and encouraged me. To Routledge and Prof Meng Ji, the series editor, for enabling me to publish this book and for their full support during the process of writing, editing and design. I would like to thank authors and publishers for permissions to reproduce illustrations in this book. While I have made every effort to contact copyright holders of material used in this volume, I would be grateful to hear from any I was unable to contact. I am also grateful to Federico Zanettin for his careful reading of the manuscript, extremely helpful critical comments and valuable suggestions. Last but not least, a special thanks goes to Mona Baker for her belief in my potential and continuous support—without her, I would probably not even be in academia today.

Part I

Theoretical considerations

Introduction

Aim

When corpora were first introduced some 25 years ago in the field of translation studies (Baker 1993), few could have anticipated the impact they would have had on a relatively young discipline. Since then, corpus-based translation research has seen some remarkable developments. Large corpora focusing specifically on translation, and even interpreting, have been created, such as the Translational English Corpus (TEC), the Oslo Multilingual Corpus, and the European Parliament Proceedings Parallel Corpus, and a broad range of tools are now available for the investigation of a wide variety of linguistic features in monolingual, bilingual, and even multilingual corpora. Corpus-based translation studies has by now overcome its "teenage angst" (Olohan 2004, p. 1) and has entered its mature years regarding developing its own identity and establishing itself as a fast-growing area of research. As a result, there is now a growing need for research that will report on the most recent corpus approaches and methodologies to help move corpus-based translation studies forward.

This book aims to contribute to this coming of age, by proposing and discussing an innovative methodological framework where corpus data and/or methods are combined in a principled way for the study of phenomena related to translation (e.g. the relationship between translated and non-translated texts). In particular, this book aims at introducing triangulation techniques and demonstrating how these can be fruitfully employed to investigate different translation phenomena and advance corpus-based translation studies by providing answers to both existing questions and new questions in the field. By the end of the book, readers should have a good understanding of the main principles of triangulation, and be able to appreciate its advantages, while being aware of its limitations. They should also be in a position to apply corpus triangulation techniques to their own corpus projects.

Ultimately, this book fills a conspicuous gap in existing corpus-based translation research, related to the complementarity of corpus data and methods. Despite the positive progress, corpus-based translation studies has remained rather static in the last years, refining existing methodologies, for example, parallel and comparable corpora, which tend to be understood as separate, rather than

complementary approaches. While different scholars (Johansson 2003; Kenny 2004; McEnery and Xiao 2008) rightly observe that parallel and comparable corpora can complement each other, the applications of such combinations have been made ad hoc, without a clear rationale and generally without the possibility of being replicated for the benefit of future research. Similarly, although it is considered standard practice to combine quantitative and qualitative analyses in corpus-based translation research, such a combination lacks a clear rationale, and its purpose and benefits are rarely mentioned, while different quantitative measures are seldom combined. This book focuses specifically on the considerable advantages of combining corpus data (e.g. parallel, comparable, synchronic and diachronic), as well as corpus methods (e.g. different quantitative measures or quantitative and qualitative methods) in a principled way, and presents the results from two specific case studies, demonstrating how the use of corpus triangulation can help answer pressing questions in translation studies.

An important aspect of the corpus triangulation framework presented in this book is that it heavily relies on a novel corpus typology, which is presented in detail, and which is based on the idea of variables, values and attributes (hence named the VVA typology). This new corpus typology is comprehensive and flexible at the same time since it can cater for all existing corpora available today, while it is sufficiently open-ended to allow for more corpora to be added in the future. The VVA typology has been purposefully created to operationalise the idea of corpus data triangulation and, as a result, needs to be adopted by any study employing this type of triangulation. However, it can also be used as a descriptive framework in any corpus-based translation project, even in those that do not employ triangulation techniques. Adopting the VVA typology more widely can considerably facilitate the selection of corpus components from different sources and allow for a common and clear terminology to be used across corpus-based translation studies.

A unique aspect of the book is that it consists of both a theoretical and an empirical part, which are interactively linked. The theoretical part is centred on the model for corpus triangulation, which has been developed through the analysis of empirical data and can be re-used for analysing similar data from different genres, contexts and/or languages. The model is then applied to empirical data, demonstrating its effectiveness and producing valuable insights about translation-related phenomena. Even though the case studies discussed inevitably reflect the author's own research interests and access to material, the corpus triangulation approach presented in this book is envisaged to be applicable to a wide range of contexts and languages, and an attempt has been made to include in the case studies different languages, genres and linguistic features.

It must be noted that the present book aims to move away from the pedagogical perspective of existing books in the field, and is, thus, not a textbook of corpus-based methods in translation studies. Readers who are unfamiliar with corpus techniques can choose from a growing number of books in the field (e.g. Olohan 2004; Zanettin 2012). Such books offer very comprehensive introductions to corpus-based translation studies and guide readers through the exciting

process of getting started with corpora in translation studies. Conversely, this book will prove of interest to translation and linguistics researchers, who have a good understanding of corpus methods and techniques, and would like to examine advanced methodologies in a fast-growing area of research. At the same time, the book will be of value to students who wish to study corpus-based translation studies in more depth and who are after an advanced resource.

Outline

The book is organised in such a way that it reports on all essential aspects of corpus triangulation and is divided into two parts, each consisting of four chapters. Part I is the theoretical part, and it is here that the theoretical and methodological underpinnings of corpus triangulation are discussed. This part begins by providing an overview of previous attempts at corpus triangulation in corpus-based translation studies and examines their limitations (Chapter 1). It then offers a general introduction to triangulation techniques and how these can be adapted to corpus-based translation studies, providing a detailed definition of corpus triangulation (Chapter 2). Further, it offers a detailed presentation of corpus data (Chapter 3) and corpus methods (Chapter 4) triangulation. Part II reports on the application of corpus triangulation to two case studies. Each case study is divided into two chapters and applies the corpus triangulation framework described in Part I to the study of different linguistic phenomena pertinent to translation and for which a combination of corpus data has been attempted in the past. The first case study (Chapter 5 and Chapter 6) examines the language of translation, while the second case study (Chapter 7 and Chapter 8) focuses on the relationship between translation and language change.

Chapter 1 begins with an overview of the way in which triangulation has been understood in translation studies. It then focuses on corpus-based translation studies and offers an in-depth examination of previous attempts at combining different corpora, as well as quantitative and qualitative methods. Models of corpus combination that have been attempted in the past are presented, especially those employed in the investigation of the (ir)regularities of the language of translation and the relationship between translation and language change. These models are examined closely regarding their advantages and limitations to establish whether they constitute examples of 'real' triangulation or an ad hoc combination of data or methods.

Chapter 2 introduces readers to the concept of triangulation, tracing its evolution and development in the social sciences and how it has come to be implemented in a growing number of disciplines. Particular attention is paid to the idea of integration, as a distinguishing characteristic of triangulation, as well as to the purpose of triangulation. These concepts are then applied to corpus-based translation studies to arrive at a definition of corpus triangulation. Two distinct types of corpus triangulation are identified: corpus data and corpus method. Finally, this chapter examines the potential advantages that triangulation techniques demonstrate, as well as their limitations and how they might be addressed.

Chapter 3 begins by providing a comprehensive definition of *corpus* as understood in corpus-based translation studies and briefly examines existing typologies of corpora in the field regarding their suitability for informing a corpus triangulation model. A novel corpus typology is proposed, namely the Variables–Values–Attributes (VVA) typology, which forms the basis of the corpus triangulation model. The remainder of this chapter explains how corpus data triangulation can be achieved based on the new corpus typology and discusses issues of corpus design that might be affected by triangulation considerations, namely corpus size, representativeness and balance.

Chapter 4 focuses on the different aspects of corpus methods that can be combined to achieve corpus method triangulation and introduces its two subtypes: within-method and between-method. The former refers to the combination of methods belonging to the same paradigm (i.e. quantitative or qualitative), while the latter relates to the combination of methods belonging to different paradigms (e.g. qualitative and quantitative). Since quantitative methods are seldom combined in corpus-based translation studies, more emphasis is placed in this chapter on how quantitative methods can be used complementarily, with a special reference to inferential statistics, which are considered crucial for the cross-comparison of different corpora.

Chapters 5 and 6 see the application of corpus triangulation to the investigation of the language of translation and in particular the examination of the factors that affect the distribution of adversative connectives in a 9-million-word corpus of non-translated and translated Russian texts and their English source texts. The factors examined are the genre of the text (fiction, children's fiction and non-fiction), the audience it addresses (adults vs. children), the influence from target linguistic conventions, and the influence of the source texts. Chapter 5 establishes the research background for the case study, examining how adversative connectives are used in English and Russian, while Chapter 6 describes the corpus design and methodology, focusing on how triangulation is achieved, and reports on the corpus findings. Results suggest that a complex interplay of factors affects the use of connectives, which can be related to influence from target linguistic conventions, source language interference and other genre-specific considerations (e.g. audience).

Finally, in Chapters 7 and 8, corpus triangulation techniques are applied to the investigation of the possible relationship between translation and language change. More specifically, an 800,000-word corpus of non-translated Greek and English popular science articles, together with translated Greek popular science articles and their English source texts, is examined to establish whether the frequency and patterning of cleft and pseudo-cleft constructions have changed in Greek texts as a result of contact with English through translation. As with the previous case study, Chapter 7 sets the background, discussing the main properties of these constructions in English and Greek, while Chapter 8 focuses on the corpus design, methodology and findings. Results suggest that, while there is no quantitative evidence to suggest that the frequency of these constructions has changed in Greek, there is some indication based on the qualitative analysis which

allows us to hypothesise that changes might be observed in their patterning due to contact with English through translation.

References

Baker, M., 1993. Corpus Linguistics and Translation Studies: Implications and Applications. In M. Baker, G. Francis, and E. Tognini-Bonelli, eds. *Text and Technology: In Honour of John Sinclair*. Amsterdam and Philadelphia: John Benjamins, pp. 231–250.

Johansson, S., 2003. Contrastive Linguistics and Corpora. In S. Granger, J. Lerot, and S. Petch-Tyson, eds. *Corpus-Based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam and New York: Rodopi, pp. 31–44.

Kenny, D., 2004. Parallel Corpora and Translation Studies: Old Questions, New Perspectives? Reporting "that" in Gepcolt: A Case Study. In G. Barnbrook, P. Danielsson, and M. Mahlberg, eds. *Meaningful Texts: The Extraction of Semantic Information From Monolingual and Multilingual Corpora*. London and New York: Continuum, pp. 154–165.

McEnery, T. and Xiao, R., 2008. Parallel and Comparable Corpora: What Is Happening? In G. Anderman and M. Rogers, eds. *Incorporating Corpora: The Linguist and the Translator*. Clevedon: Multilingual Matters, pp. 18–31.

Olohan, M., 2004. *Introducing Corpora in Translation Studies*. London and New York: Routledge.

Zanettin, F., 2012. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome.

1 Triangulation in corpus-based translation studies

1.0 Introduction

This chapter offers an overview of existing attempts towards triangulation in corpus-based translation studies. These attempts are not examined in isolation, but rather as part of a more general trend in translation research, and the chapter begins by examining how triangulation is understood in translation studies research more generally. In corpus-based translation studies, we can identify two main types of triangulation: one involving the combination of different types of corpora and one consisting of the combination of quantitative and qualitative methods of analysis. More emphasis is placed here on triangulation where different types of corpora are combined since it presents more difficulties compared to the other type, which is considered somewhat more established in corpus-based translation research. Since attempts at combining corpus data and methods are rarely labelled as triangulation, the terms *combined corpora* and *combined methods* are used in this chapter to clearly differentiate such approaches from the corpus triangulation model presented in this book. The aim of this chapter is to reveal the gaps in existing research and stress the need for a systematic and comprehensive account of corpus triangulation.

1.1 Triangulation in translation research

The most informative theoretical account of triangulation in translation studies is provided by Hansen, who defines triangulation as "a mix of procedures to grasp complex phenomena" (2010, p. 207). She also identifies the different elements that can be combined: subjects, material, strategies, methods, purposes, perspectives and investigators. What is important in Hansen's account of triangulation is that she clearly distinguishes it from *combination* and stresses the importance of integration as a distinguishing characteristic of triangulation. Similarly, Saldanha and O'Brien consider triangulation "the backbone of solid, high-quality research" (2013, p. 5). Still, they define triangulation rather simply as "cross-checking the results of one set of data provides with results from another set of data" (2013, p. 39), which have been obtained using different methods. According to them, triangulation in translation studies consists of a combination of different data and

methods. Like Hansen, Saldanha and O'Brien make an attempt to systematise the essential characteristics of triangulation, notably integration. However, no detailed examples of how triangulation can be implemented in a research project are provided. Hansen offers only one example where different types of data (questionnaires, interviews, the evaluation of target texts, and the log files of the revision process) are used to reveal the relationship between translation quality and time management.

Empirically, triangulation was initially used in process-oriented translation research, where it is considered as "a desirable best practice" (Shreve and Angelone 2010, p. 6), which can be used to "throw light on the nature of the process of translation" (Alves and Gonçalves 2003, p. vii). Since triangulation has mostly been used in the social sciences, it is not surprising that it found its home in process-oriented translation research, where the focus is on how translators translate. Within this context, triangulation is defined as "the use of two or more data acquisition methodologies within a single study to improve the quality, validity, and reliability of research findings" (Shreve and Angelone 2010, p. 6). From this definition, it is clear that translation research, or at least process-oriented research, has a rather limited view of triangulation, which ignores the potential advantages of triangulation for understanding complex phenomena better, as has been suggested by Hansen (2010). We can find some examples of process-oriented translation research making explicit references to the use of triangulation techniques in the research of the different cognitive aspects of professional translation. For example, Muñoz Martín (2009) combines an impressive range of different research methods, including those involved in Translog, WordSmith corpus tools, questionnaires, post-evaluation and statistics, while Jakobsen (2003) and Alves and Gonçalves (2003) combine Translog and Thinking Aloud Protocols. Similarly, Zheng and Xiang (2013) employ four different techniques, namely processing times, translation quality assessments, questionnaires, and interviews, to investigate the challenges that metaphors present for sight translators. Triangulation is also used in product-oriented translation research, albeit to a more limited extent (Baumgarten 2009). It has also been used as a means to overcome the methodological weaknesses of product and process-oriented research, by combining process and product data, for example, qualitative experimental psycholinguistics data and quantitative corpus data (Alves et al. 2010). However, the focus is still on the cognitive processes associated with translation, which suggests that the ultimate goal is process oriented. Typically, qualitative and quantitative methods are combined to overcome the deficiencies of any one method. This combination of methods has also been used in interpreting research (Davitti and Pasquandrea 2014; Gile 2003; Hild 2004)

Although triangulation is employed to some extent in translation research, studies in which some sort of mixing has taken place do not foreground the triangulation approach, and hardly any treat triangulation in a principled way. Most research in the field, simply mentions triangulation in passing, without providing information about how exactly triangulation has been interpreted and applied. Similarly, it is hard to find accounts of the specific advantages that triangulation

brought to a study. This state of affairs might be evidence of the fact that translation studies, although employing triangulation techniques, is not recognising their full potential. It also raises the important question of whether these studies have employed 'real' triangulation techniques, and not a collation of different data or methods, and, ultimately, of how exactly triangulation is understood within translation studies. Similar considerations (regarding whether or not 'real' triangulation has been used) become highly relevant when we examine how triangulation has been employed in corpus-based translation studies.

1.2 Why combine corpora

While corpora present numerous advantages, for example, they capture a significant amount of naturally occurring linguistic data, each type of corpus has its limitations, which have become the focus of criticism. On the one hand, parallel bilingual corpora are particularly useful when it comes to revealing cross-linguistic equivalences, since they have "the advantage of keeping meaning and function constant across the compared languages"(Altenberg and Granger 2002, p. 9), thus allowing for direct comparability. However, they do not provide any information about the features of translation-specific language (often referred to as tranlsationese) or other general characteristics of translated texts. For the examination of these features, comparable monolingual corpora of translated and non-translated texts are more appropriate than parallel bilingual corpora. Studies employing such comparable corpora recognise translation as a distinct communicative event with its own context, goals and pressures (Baker 1996). They allow researchers to identify linguistic patterns in translated texts, which are a result of the complex nature of the translation activity (Olohan 2001), and compare them to those found in non-translated texts, something that is not possible with the use of parallel corpora. On the other hand, contrary to parallel corpora, comparable corpora are not suitable for revealing cross-linguistic variation, and they focus primarily on the product of translation, that is, the translated text, rather than the process of translation. Since comparable monolingual corpora focus on the examination of target texts without any reference to their source texts, they have been criticised for ignoring the important role played by the source text in understanding the nature of translation (Stewart 2000).

Additionally, parallel bilingual corpora tend to be rather small and imbalanced (Altenberg and Granger 2002), thus not representative of the languages compared, since they mostly focus on a very limited number of texts—in many cases a single source text and its translation. This is of course not the case with the larger parallel corpora available, such as the Intersect corpus (English–French), which consists of approximately 1.5 million words. However, the majority of parallel corpora used are much smaller, thus affecting the perspective of the observer to an undesirable degree (Malmkjaer 1998). To address this problem, Malmkjaer advocates the use of several translations of a single source text, which is, however, not always possible. Contrary to parallel bilingual corpora, comparable monolingual corpora can be much more representative and balanced, as they also tend to

be larger. Still, when using comparable monolingual corpora, it is hard to achieve real comparability of data, particularly regarding style and function (Altenberg and Granger 2002). Comparability presupposes that languages are in some way symmetrical, which, in reality, is never the case (Zanettin 2013), and seems to clash with representativeness: the greater comparability one achieves, the more the corpus will be distorted regarding representativeness and vice versa (Leech 2007). The problem of comparability is further complicated by the fact that in most cases it is difficult to know what to compare (Johansson 1998). These issues of comparability have resulted in monolingual comparable corpora consisting of specialised texts limited to a very specific genre or topic, such as the Translational English Corpus (TEC), which consists predominantly of translated fiction. Alternatively, they can be very large balanced corpora, where factors such as topic, register and function can be controlled (Altenberg and Granger 2002).

The discussion of the limitations of parallel and comparable corpora reveals that they are complementary in many respects and that an approach where the two can be combined presents considerable advantages. As a reaction to the criticism directed towards studies that employ only one type of corpora, some voices have been raised to claim that different types of corpora can, and should, complement each other. The first explicit reference to the benefits of combining corpora is made by Teubert (1996, p. 252), who argues that "ideally, parallel corpora should be viewed as complementary to comparable corpora". Although he views this combination from the perspective of lexicography, it is not difficult to extend his suggestion to encompass translation. Similarly, Johansson (1998) identifies the possibility of combining comparable and parallel corpora, when he argues that we need to compare results from parallel bilingual corpora to those obtained from a corpus consisting of non-translated and translated texts in the same language, which he calls a control corpus. All in all, translation scholars have started treating parallel and comparable corpora as "complementary sources of cross-linguistic data" (Altenberg and Granger 2002, p. 9) and realised that a combination of different types of corpora is crucial if research in corpus-based translation studies is to move forward. For instance, Zanettin (2000) argues that the examination of a monolingual comparable corpus of translated texts needs to be complemented with the comparison of these texts and their source texts, as well as a reference corpus in the languages involved in the comparison. He also urges translation studies to make use of more diversified types of analyses, where different corpus components can be compared and contrasted. His ideal corpus "is not . . . a pre-formed set of texts but an open-ended corpus comprising different components" (Zanettin 2000, p. 109), which as will be explained in Chapter 3 is the fundamental principle of corpus triangulation. Later, Zanettin (2012, p. 12) refines even further his ideas regarding the combined use of corpora and explicitly refers to corpus triangulation, in the form of a combination of data:

> [C]orpus-based translation studies seem to profit mostly not only from the comparison of different corpus components but also from the triangulation of data, and the combination of different components of multilingual corpora

as well as of reference corpora not originally created for translation-oriented purposes.

Despite the recognition that corpus-based translation research would benefit from the triangulation of corpora (see also Kenny 2004; McEnery and Xiao 2008; Olohan 2004), little has been done in the direction of actually employing combined corpora in translation studies. For example, relevant studies in the field, which employ comparable corpora, do not make use of Zanettin's ideal corpus design, and they typically omit the analysis of a parallel corpus (for example, Olohan and Baker 2000). This might be explained by the fact that, while in contrastive linguistics parallel multilingual corpora are quite often used together with comparable multilingual corpora, in translation studies parallel and comparable corpora are still often treated as mutually exclusive.

1.3 Research with combined corpora

There are two specific areas of translation research where attempts have been made to combine corpora. The first area is the examination of the language of translation, not just to reveal potentially common characteristics across languages, but also to understand the way in which specific linguistic features are used in translated texts or by specific translators. The second area is the investigation of the relationship between translation and language change and, more specifically, the possible influence that translation might have on linguistic developments in the target language.

1.3.1 The language of translation

The first area of study, where more than one corpora tend to be used is the examination of the (ir)regularities of translated texts, compared to non-translated texts in the same language, and/or their respective source texts, and/or other translations into the same or other languages. This area of study is often associated with the investigation of recurrent features of translation (typically labelled as *translation universals*), although the combined use of corpora does not need to be exclusive to the study of such features. Instead, it can be used to explore the linguistic properties of translated texts more generally, irrespective of whether or not similar properties appear in translated texts from/into other languages (for an empirical study of such properties see Chapters 5 and 6).

A combined corpus methodology for the investigation of the language of translation that gained some popularity in corpus-based translation studies is that proposed by Johansson (Johansson 1998; Johansson 2002; Johansson 2003a; Johansson 2003b), and which has been used for the creation of the English–Norwegian Parallel Corpus (ENPC) and the Oslo Multilingual Corpus (OMC). Johansson proposes two corpus models: the diamond and the star. The diamond model (Figure 1.1) consists of non-translated and translated texts, as well as the source texts of the translations in three different languages. Thus, four different

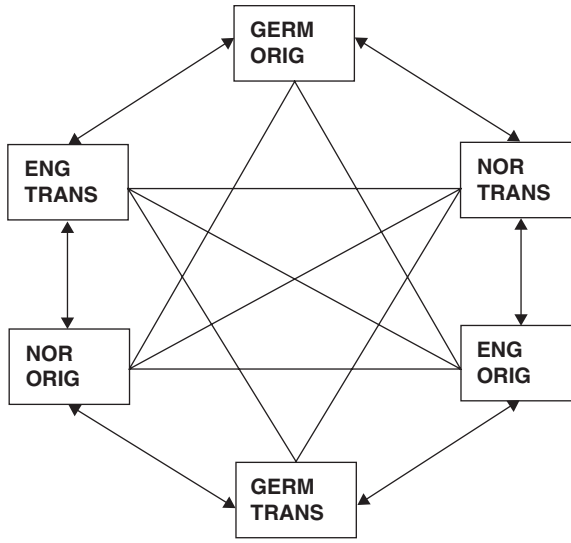

Figure 1.1 Diamond corpus model (Johansson 2002, p. 49)

corpora are available for contrastive analysis: (a) a comparable multilingual corpus of non-translated texts, (b) a comparable multilingual corpus of translated texts, (c) three comparable monolingual corpora of translated and non-translated texts in the same language and (d) a parallel multilingual bi-directional corpus. Although three languages are employed in OMC, namely English, Norwegian and German, in principle, an unlimited number of languages can be used, although the more complex the model, the bigger the problems of availability and comparability.

The advantages of this corpus structure are the combination of comparable and parallel elements, as well as the emphasis placed on incorporating a broad range of languages. The fact that it is also bi-directional allows for different comparisons among texts, as the various lines in Figure 1.1 suggest, resulting in the diamond shape. A diamond corpus design can be used in the examination of how a specific linguistic feature of a source language A has been translated into different target languages (B, C, D). The patterns found in the translated texts can subsequently be compared to a corpus of non-translated texts in these languages. Then, a parallel corpus in the reverse direction of translation can be used to examine how the patterns found in the translated texts (B, C, D) are translated into language A. According to Zanettin (2014, p. 185), it is possible to repeat this process "in a cyclical fashion". Thus, regarding visual semiotics, instead of a diamond shape, we can also understand this design as a circle.

An alternative to the diamond model is the star model (Figure 1.2). This model consists of a single source text and its translations into several target languages,

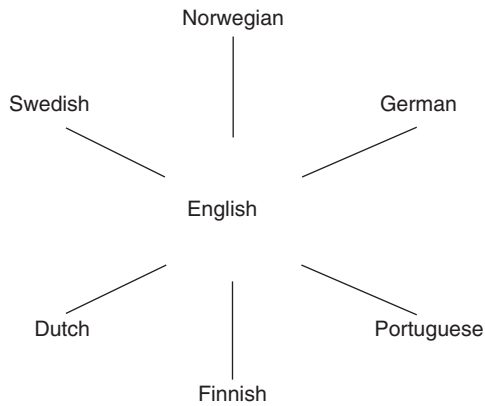

Figure 1.2 Star corpus model (Johansson 2002, p. 48)

allowing for comparisons to be made across the different translations. In principle, two different types of corpora are available: (a) a comparable multilingual corpus of translated texts, and (b) a parallel multilingual corpus. The aim of this design is to reveal translation-specific features, as well as language-specific translation patterns. A much larger number of languages can be included, without the problem of comparability that affects the diamond model. The star model can also be used to examine multiple translations in the same language of a single source text, for example, in the investigation of stylistic features of different translators. The main difference between the diamond and the star model is that the latter does not cater for reciprocity, but, because of its rather simpler corpus design, allows for more languages/texts to be compared and contrasted. For these reasons, the star model tends to be more frequently employed (for example, Bowker and Bennison 2003; Malmkjaer 2003; Mouka et al. 2015). However, the research potential of a diamond corpus is considerably higher, not least because of the greater number of possible corpus combinations.

Similarly to Johansson, Bernardini (2011) also identifies the false dilemma between monolingual comparable and bilingual parallel corpora and rightly observes that neither type of corpus would be sufficient in isolation. She takes forward the ideal corpus design proposed by Zanettin (2000) and advocates for the use of a tripartite corpus structure, which consists of source texts in language A, their translations in language B, and a comparable/reference corpus of non-translated texts in language B. A small illustration of this model is given in Figure 1.3. Bernardini considers this corpus structure as the minimum requirement for the study of the particularities of translated texts. The main advantage of this corpus structure is that it can be more easily employed compared to Johansson's models, as it presents fewer difficulties related to the availability of data and is simpler regarding design. As a result, it is more economical and could lend itself

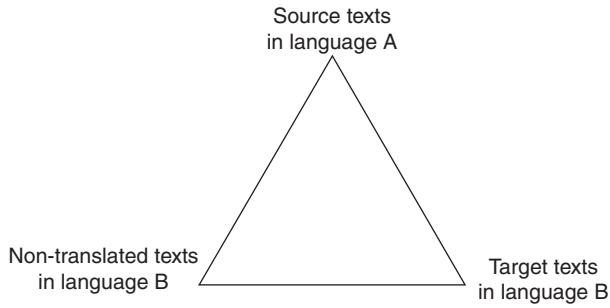

Figure 1.3 Tripartite corpus structure

more easily to different research projects; for example, Bernardini employs this corpus design to investigate borrowing in technical translation and phraseological regularities in fiction translation. However, this corpus structure is still limited regarding what types of corpora might be combined, and it does not cater for the possibility of adding more corpora in other languages.

Despite the individual differences, these three corpus models are important for two reasons. Firstly, they constitute the first attempt at developing a methodological framework for corpus-based translation studies, which places the combination of corpora at its centre. They recognise the insufficiency in previous studies of focusing only on one type of corpus and stress the importance of data from different languages and both translated and non-translated texts. Secondly, they succeed in presenting for the first time a clear and comprehensive combined corpus design, which can be replicated. However, despite their advantages, both models are limited to a specific type of research, that is the investigation of certain linguistic features of translation, and cannot be easily used outside this research context.

1.3.2 Translation and language change

The second area of study, where a combination of corpora tends to be used, is the investigation of translation as a language contact phenomenon and a possible facilitator of linguistic developments in the target language, which has seen a growing interest in recent years. Numerous studies have been conducted in the field (Amouzadeh and House 2010; Bisiada 2013; Gellerstam 1986; McLaughlin 2011), all of which have employed a combination of diachronic parallel (either bilingual or multilingual) and comparable (both monolingual and bilingual/multilingual) corpora. Although some commonalities can be observed, all studies seem to employ a very different corpus design. This is an indication that a clear corpus methodology, which other researchers interested in the investigation of the relationship between translation and language change could easily adopt, is still missing.

There is one corpus design that stands out from the rest because it makes an attempt at systematising the way in which the different corpora can be combined, and at identifying the potential advantages of employing each corpus type. This is the corpus design developed for the 'Covert Translation' project (Baumgarten and Özçetin 2008; Becher 2011; House 2011; Kranich et al. 2012). This project investigates the translational language contact between English and German in the genres of business communication, popular science and computer science. The corpus (Figure 1.4), which is a "dynamic, implicitly diachronic translation and parallel text corpus" (House 2011, p. 190) combines synchronic and diachronic elements and consists of both translated and non-translated texts. The corpus is divided into three components: (a) the *Primary Corpus* is a parallel corpus of non-translated English texts and their translations into German, (b) the *Parallel Corpus* is a comparable corpus of English and German non-translated texts, as well as non-translated texts from French and Spanish, and (c) the *Validation Corpus* is a parallel corpus of non-translated German texts and their translations into English, as well as translated texts from English into French and Spanish. The texts, at least the English and German ones, capture two points in time: 1978–1982 and 1999–2002.[1]

This model makes a significant step towards the combined use of corpora, as it recognises the importance of increasing the understanding of a phenomenon

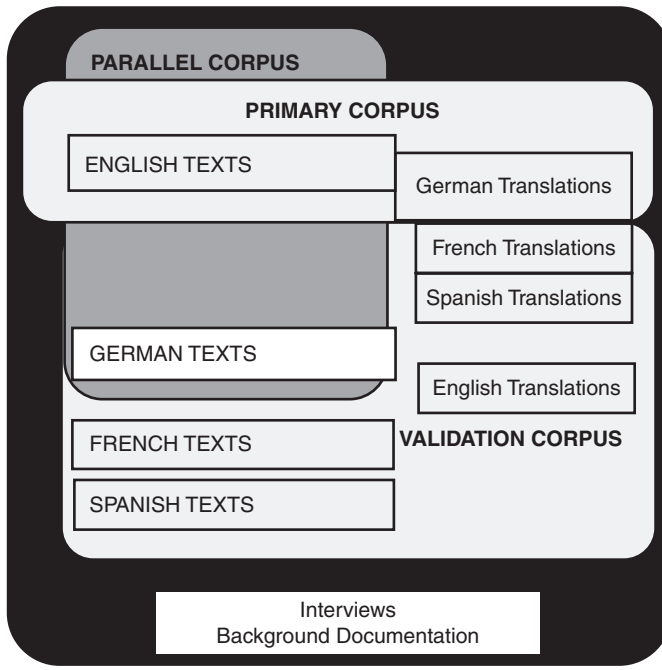

Figure 1.4 The corpus of the 'Covert Translation' project (House 2011, p. 191)

by collecting data from a wide range of language pairs, and also from the reverse translation direction. Similar considerations are present in Johansson's diamond and star models, albeit they are not as explicitly stated. Another advantage of this model is that it presents a stepped approach to corpus analysis, where each stage serves a different purpose, and the results from all analyses are combined to answer a single research question. In this sense, the corpus model developed as part of the 'Covert Translation' project has come closer to corpus triangulation than any other corpus design, where different corpora are combined. Nevertheless, this corpus design does not present the same methodological potential for corpus-based translation studies, as Johansson's and Bernardini's models do. Even regarding visual representation, Figure 1.4 appears convoluted, and a visual depiction of how corpus components can be compared is missing. The terms used to refer to the different corpus components also minimise its potential to develop into a strong methodological framework. The corpus components are presented in terms of their purpose (e.g. validation) or their significance (e.g. primary), rather than in terms of what they consist of.

When examining all models where a combination of corpora is encouraged, it is clear that the combination of different corpora is not a new phenomenon in translation studies. All models recognise that the analysis of one type of corpus can provide us with interesting results, but it is only through a combination of different corpora that we might answer more complicated and pressing questions, such as the examination of the language of translated texts, and the role that translation might play in language contact situations. However, while it is possible to infer some of the benefits that this combination of corpora can bring to a study, their purpose and advantages are often considered obvious and are rarely explained. This is a good indication that triangulation has not been sufficiently recognised and developed in corpus-based translation studies. This state of affairs suggests that existing combined corpus methodologies have been created ad hoc, even though they might demonstrate some characteristics of triangulation. While the models discussed here can be considered the first notable attempts towards corpus triangulation, corpus triangulation still needs to be operationalised to reach its full potential and become an integral part of translation studies. What seems to be missing is a flexible corpus framework, which is not restricted to one research topic, but caters for a wide range of research areas and allows for corpora to be combined in novel ways.

1.4 Combining methods in corpus-based translation research

So far, we have examined previous attempts at a combination of different types of corpora, where different sources of data are employed. There is another type of triangulation often used in translation studies, which consists of the combination of quantitative and qualitative methods. This combination is the outcome of the criticism addressed towards corpus-based approaches in the 1990s when the focus was mostly on quantitative methods.[2] As a result, there is now a growing number

of corpus-based studies, which combine quantitative and qualitative methods (for example, Dayrell 2004; Kenny 2001; Liao 2010; Marco 2013; Nilsson 2002; Williams 2009).

To demonstrate how quantitative and qualitative methods have been combined in corpus-based translation studies, this section focuses on three specific studies, which have been selected because they also use a combination of corpora. The first example of a combination of quantitative and qualitative methods can be found in a study conducted by Shuttleworth (2014), who uses Johansson's star model and creates a multilingual corpus of popular science texts from *Scientific American* to examine how metaphorical expressions in English are translated into different languages. He makes use of quantitative methods and provides descriptive statistics of the proportions of techniques that different translators employ, followed by an in-depth discussion of selected examples. Similarly, Winters (2009) also makes use of the star model to create a corpus consisting of an English source text, namely the novel *The Beautiful and Damned*, and its two translations into German. She conducts a quantitative analysis to reveal significant patterns of use of modal particles, and a qualitative analysis of data from a specific modal particle to reveal its pragmatic properties. A final example can be found in the work of Bernardini and Ferraresi (2011), who use the tripartite corpus structure to examine the use of Anglicisms in translated and non-translated texts. They complement the results from the quantitative analysis, with qualitative observations and discussion of selected examples. The common ground in all these studies is that they use the different methods complementarily to highlight different dimensions of the phenomenon under investigation.

Despite the fact that these studies demonstrate characteristics of triangulation, this is a result of a rather ad hoc combination of methods, since they neither explain the clear benefits that the combination of methods provides nor identify the aim of combining different methods from the outset of the research project. Rather, the qualitative analysis often follows the quantitative one, and it appears that this methodological approach is considered mainstream in corpus-based translation studies. As a result, integration, which has been identified as an important characteristic of triangulation in translation studies, is missing. Thus, the same problems, which were identified for the combined use of corpora, are pertinent to the combined use of methods in corpus-based translation studies. However, due to the considerably less complex nature of the combination of methods compared to the combination of corpus data, the former can be successfully achieved more easily than the latter, even if it is ad hoc. This is the most probable reason why more studies combine methods rather than corpora.

A further characteristic that supports the idea that methodological triangulation has not been so far adequately developed in corpus-based translation studies is that, by far, the most popular combination attempted is that of quantitative and qualitative methods. However, when it comes to corpus-based translation studies, a combination of methods from the same paradigm, especially quantitative, can be particularly useful. Even though it might be challenging to combine quantitative measures, the fact that some studies have started using different types of quantitative analysis is a positive sign towards the development of this

type of triangulation. Any corpus-based study that makes use of descriptive statistics, for example, mean, range, standard deviation, as well as inferential statistics, for example, statistical significance tests, tests of difference, tests of relationship, might be said to have used a combination of quantitative methods (see also Chapter 4). Examples of such studies are Dong and Lan (2010), Ji (2012), Bisiada (2013) and Malamatidou (2016). However, there is still a need for a more rigorous corpus-based methodology in translation studies, which will recognise and employ a wider range of inferential statistics (de Sutter et al. 2012), and quantitative methods more generally. This is why recent books on the topic of quantitative methods in corpus-based translation studies (Ji and Oakes 2012), and translation studies more generally (Mellinger and Hanson 2016) are not only welcomed but necessary.

1.5 Conclusion

Because triangulation demonstrates such flexibility and has wide applicability, it has led researchers in different disciplines, including translation studies, to misuse the term or apply it very loosely. As a result, not all attempts at triangulation can be considered successful. Such is the case of previous studies in the field of corpus-based translation studies, which have not been conducted in a genuine spirit of triangulation, but rather constitute an ad hoc combination of different corpus types and quantitative and qualitative methods. This is strong indication that principles of triangulation are not adequately understood in corpus-based translation studies and that it has not yet been possible to benefit fully from triangulation. What is missing is a guided and principled account of how the principles of triangulation can be applied to corpus-based research, which can act as a general model explaining all the different possible combinations. This model can then serve as a guide for corpus-based translation research and encourage triangulation to be used more widely, but also in a more integrated manner.

Notes

1 Although the aim here is not to criticise the way in which the corpus is described, it is worth noting that the terms used to refer to the different corpus components are not very successful. For example, the *Parallel Corpus* does not include a comparison of translated and non-translated texts, which is normally how a parallel corpus is understood in corpus-based translation studies (see also Chapter 3).
2 For example, Malmkjaer (1998) warned that the focus on quantitative methods might result in scholars treating as marginal, or even ignoring, problematic cases, and suggested that more context is necessary compared to what the computers are capable of displaying, which implies complementing corpus-based analyses with a manual examination of data.

References

Altenberg, B. and Granger, S., 2002. Recent Trends in Cross-Linguistic Lexical Studies. In B. Altenberg and S. Granger, eds. *Lexis in Contrast: Corpus-Based Approaches.* Amsterdam: John Benjamins, pp. 3–48.

Alves, F. and Gonçalves, J.L., 2003. A Relevance Theory Approach to the Investigation of Inferential Processes in Translations. In F. Alves, ed. *Triangulating Translation*. Amsterdam and Philadelphia: John Benjamins, pp. 3–24.

Alves, F. et al., 2010. Translation Units and Grammatical Shifts: Towards an Integration of Product- and Process-Based Translation Research. In G.M. Shreve and E. Angelone, eds. *Translation and Cognition*. Amsterdam and Philadelphia: John Benjamins, pp. 109–142.

Amouzadeh, M. and House, J., 2010. Translation as a Language Contact Phenomenon: The Case of English and Persian Passives. *Languages in Contrast*, 10(1), pp. 54–75.

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*. Amsterdam and Philadelphia: John Benjamins, pp. 175–185.

Baumgarten, N. and Özçetin, D., 2008. Linguistic Variation Through Language Contact in Translation. In P. Siemund and N. Kintana, eds. *Language Contact and Contact Language*. Amsterdam and Philadelphia: John Benjamins, pp. 293–316.

Baumgarten, S., 2009. *Translating Hitler's "Mein Kampf": A Corpus-Aided Discourse Analytical Study*. Saarbrucken: VDM.

Becher, V., 2011. When and Why Do Translators Add Connectives? A Corpus-Based Study. *Target*, 23(1), pp. 26–47.

Bernardini, S., 2011. Monolingual Comparable Corpora and Parallel Corpora in the Search for Features of Translated Language. *SYNAPS: A Journal of Professional Communication*, 26, pp. 2–13.

Bernardini, S. and Ferraresi, A., 2011. Practice, Description and Theory Come Together – Normalization or Interference in Italian Technical Translation? *Meta: Journal des traducteurs*, 56(2), p. 226.

Bisiada, M., 2013. Changing Conventions in German Causal Clause Complexes: A Diachronic Corpus Study of Translated and Non-Translated Business Articles. *Languages in Contrast*, 13(1), pp. 1–27.

Bowker, L. and Bennison, P., 2003. Student Translation Archive and Student Translation Tracking System: Design, Development and Application. In F. Zanettin, S. Bernardini, and D. Stewart, eds. *Corpora in Translator Training*. Manchester: St. Jerome, pp. 103–118.

Davitti, E. and Pasquandrea, S., 2014. Enhancing Research-Led Interpreter Education: An Exploratory Study in Applied Conversation Analysis. *The Interpreter and Translator Trainer*, 8(3), pp. 374–398.

Dayrell, C., 2004. Towards a Corpus-Based Research Methodology for Investigating Lexical Patterning in Translated Texts. *Language Matters*, 35(1), pp. 70–101.

de Sutter, G., Delaere, I. and 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), pp. 203–224.

Dong, D.-H. and Lan, Y.-S., 2010. Textual Competence and the Use of Cohesion Devices in Translating Into a Second Language. *The Interpreter and Translator Trainer*, 4(1), pp. 47–88.

Gellerstam, M., 1986. Translationese in Swedish Novels Translated From English. In L. Wollin and H. Lindquist, eds. *Translation Studies in Scandinavia: Proceedings From the Scandinavian Symposium on Translation Theory (SSOTT) II, Lund 14–15 June, 1985*. Lund: CWK Gleerup, pp. 88–95.

Gile, D., 2003. Empirical Research Into the Role of Knowledge in Interpreting: Methodological Aspects. In H.V. Dam, J. Engberg, and H. Gerzymisch-Arbogast, eds. *Knowledge Systems and Translation*. Berlin: Mouton de Gruyter, pp. 149–172.

Hansen, G., 2010. Integrative Description of Translation Processes. In G.M. Shreve and E. Angelone, eds. *Translation and Cognition*. Amsterdam and Philadelphia: John Benjamins, pp. 189–211.

Hild, A., 2004. Establishing Rigour in a Between-Methods Investigation of SI Expertise. In Y. Gambier, M. Shlesinger, and R. Stolze, eds. *Doubts and Directions in Translation Studies*. Amsterdam and Philadelphia: John Benjamins, pp. 99–112.

House, J., 2011. Using Translation and Parallel Text Corpora to Investigate the Influence of Global English on Textual Norms in Other Languages. In A. Kruger, W. Kim, and J. Munday, eds. *Corpus-Based Translation Studies: Research and Applications*. London and New York: Bloomsbury, pp. 187–208.

Jakobsen, A.L., 2003. Effects of Think Aloud on Translation Speed, Revision, and Segmentation. In F. Alves, ed. *Triangulating Translation*. Amsterdam and Philadelphia: John Benjamins, pp. 69–96.

Ji, M., 2012. Translation and Scientific Terminology. In M. Oakes and M. Ji, eds. *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam and Philadelphia: John Benjamins, pp. 249–274.

Ji, M. and Oakes, M., 2012. A Corpus Study of Early English Translations of Cao Xueqin's Hongloumeng. In M. Oakes and M. Ji, eds. *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam and Philadelphia: John Benjamins, pp. 177–208.

Johansson, S., 1998. On the Role of Corpora in Cross-Linguistic Research. In S. Johansson and S. Oksefjell, eds. *Corpora and Cross-Linguistic Research: Theory, Method and Case Studies*. Amsterdam: Rodopi, pp. 3–24.

Johansson, S., 2002. Towards a Multilingual Corpus for Contrastive Analysis and Translation Studies. In L. Borin, ed. *Parallel Corpora, Parallel Worlds. Selected Papers From a Symposium on Parallel and Comparable Corpora at Uppsala University, Sweden, 22–23 April, 1999*. Amsterdam and New York: John Benjamins, pp. 47–59.

Johansson, S., 2003a. Contrastive Linguistics and Corpora. In S. Granger, J. Lerot, and S. Petch-Tyson, eds. *Corpus-Based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam and New York: Rodopi, pp. 31–45.

Johansson, S., 2003b. Reflections on Corpora and Their Uses in Cross-Linguistic Research. In F. Zanettin, S. Bernardini, and D. Stewart, eds. *Corpora in Translator Training*. Manchester: St. Jerome, pp. 135–144.

Kenny, D., 2001. *Lexis and Creativity in Translation*. London and New York: Routledge.

Kenny, D., 2004. Parallel Corpora and Tranlsation Studies: Old Questions, New Perspectives? Reporting "that" in Gepcolt: A Case Study. In G. Barnbrook, P. Danielsson, and M. Mahlberg, eds. *Meaningful Texts: The Extraction of Semantic Information From Monolingual and Multilingual Corpora*. London and New York: Continuum, pp. 154–165.

Kranich, S., House, J. and Becher, V., 2012. Changing Conventions in English-German Translation of Popular Scientific Texts. In K. Braunmüller, C. Gabriel, and B. Hänel-Faulhaber, eds. *Multilingual Individuals and Multilingual Societies*. Amsterdam and Philadelphia: John Benjamins, pp. 315–334.

Leech, G., 2007. New Resources, or Just Better Old Ones? The Holy Grail of Representativeness. In C. Biewer, N. Nesselhauf, and M. Hundt, eds. *Language and*

Computers: Studies in Practical Linguistics – Volume 59: Corpus Linguistics and the Web. Amsterdam: Rodopi, pp. 135–149.

Liao, M.-H., 2010. Influence of Translations on Non-Translations: Popular Science as a New Genre in Taiwan. *Perspectives: Studies in Translatology*, 18(2), pp. 127–140.

Malamatidou, S., 2016. Understanding Translation as a Site of Language Contact: The Potential of the Code-Copying Framework as a Descriptive Mechanism in Translation Studies. *Target*, 28(3), pp. 399–423.

Malmkjaer, K., 1998. Love Thy Neighbour: Will Parallel Corpora Endear Linguists to Translators? *Meta: Journal des traducteurs*, 43(4), p. 534.

Malmkjaer, K., 2003. On a Pseudo-Subversive Use of Corpora in Translator Training. In F. Zanettin, S. Bernardini, and D. Stewart, eds. *Corpora in Translator Training*. Manchester: St. Jerome, pp. 119–134.

Marco, J., 2013. Tracing Marked Collocation in Translated and Non-Translated Literary Language. In C. Way et al., eds. *Tracks and Treks in Translation Studies: Selected Papers From the EST Congress, Leuven 2010*. Amsterdam and Philadelphia: John Benjamins, pp. 167–188.

McEnery, T. and Xiao, R., 2008. Parallel and Comparable Corpora: What Is Happening? In G. Anderman and M. Rogers, eds. *Incorporating Corpora: The Linguist and the Translator*. Clevedon: Multilingual Matters, pp. 18–31.

McLaughlin, M., 2011. *Syntactic Borrowing in Contemporary French: A Linguistic Analysis of News Translation*. Oxford: Legenda.

Mellinger, C.D. and Hanson, T.A., 2016. *Quantitative Research Methods in Translation and Interpreting Studies*. London and New York: Routledge.

Mouka, E., Saridakis, I. and Fotopoulou, A., 2015. Racism Goes to the Movies: A Corpus-Driven Study of Cross-Linguistic Racist Discourse Annotation and Translation Analysis. In F. Zanettin and C. Fantinuoli, eds. *New Directions in Corpus-Based Translation Studies*. Berlin: Language Science Press, pp. 35–70.

Muñoz Martín, R., 2009. Expertise and Environment in Translation. *Mutatis Mutandis*, 2(1), pp. 24–37.

Nilsson, P.-O., 2002. An Assessment of the Translation-Specificity of Over-Represented Multi-Word Patterns in Swedish Fiction Texts Translated From English. In L. Van Vaerenbergh, ed. *Linguistics and Translation Studies*. Antwerp: HIVT, pp. 407–417.

Olohan, M., 2001. Spelling Out the Optionals in Translation: A Corpus Study. *UCREL Technical Papers*, 13, pp. 423–432.

Olohan, M., 2004. *Introducing Corpora in Translation Studies*. London and New York: Routledge.

Olohan, M. and Baker, M., 2000. Reporting That in Translated English: Evidence for Subconscious Processes of Explicitation? *Across Languages and Cultures*, 1(2), pp. 141–158.

Saldanha, G. and O'Brien, S., 2013. *Research Methodologies in Translation Studies*. London and New York: Routledge.

Shreve, G.M. and Angelone, E., 2010. Translation and Cognition: Recent Developments. In G.M. Shreve and E. Angelone, eds. *Translation and Cognition*. Amsterdam and Philadelphia: John Benjamins, pp. 1–13.

Shuttleworth, M., 2014. Scientific Rich Images in Translation: A Multilingual Study. *The Journal of Specialised Translation*, 21, pp. 35–51.

Stewart, D., 2000. Poor Relations and Black Sheep in Translation Studies. *Target*, 12(2), pp. 205–228.

Teubert, W., 1996. Comparable or Parallel Corpora? *International Journal of Lexicography*, 9(3), pp. 238–264.

Williams, I.A., 2009. A Corpus-Based Study of Spanish Translations of the Verb "report" in Biomedical Research Articles. *Meta: Journal des traducteurs*, 54(1), p. 146.

Winters, M., 2009. Modal Particles Explained: How Modal Particles Creep Into Translations and Reveal Translators' Styles. *Target*, 21(1), pp. 74–97.

Zanettin, F., 2000. Parallel Corpora in Translation Studies: Issues in Corpus Design and Analysis. In M. Olohan, ed. *Intercultural Faultlines: Research Models in Translation Studies I: Textual and Cognitive Aspects*. Manchester: St. Jerome, pp. 105–118.

Zanettin, F., 2012. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome.

Zanettin, F., 2013. Corpus Methods for Descriptive Translation Studies. *Procedia–Social and Behavioral Sciences*, 95, pp. 20–32.

Zanettin, F., 2014. Corpora in Translation. In J. House, ed. *Translation: A Multidisciplinary Approach*. Basingstoke: Palgrave Macmillan, pp. 178–199.

Zheng, B. and Xiang, X., 2013. Processing Metaphorical Expressions in Sight Translation: An Empirical-Experimental Research. *Babel*, 59(2), pp. 160–183.

2 Introduction to corpus triangulation

2.0 Introduction

This chapter introduces the main principles and types of triangulation as understood in the social sciences, where it has been mostly used. Because triangulation is a time-consuming and expensive technique, situations of triangulation for triangulation's sake need to be avoided, and there needs to be a clear understanding of what we can achieve with the help of triangulation. For this reason, a separate section is dedicated to the purpose of triangulation. These principles serve as a guide for the development of corpus triangulation techniques introduced in this chapter. Two distinct types of corpus triangulation are identified here: corpus data and corpus method. This categorisation also reflects previous attempts at combining corpus data and methods in corpus-based translation studies presented in Chapter 1. Readers can also find here a discussion of the multiple advantages of triangulation in general, and corpus triangulation in particular, as well as an examination of their limitations, together with ways in which they can be minimised.

2.1 Basic principles

Triangulation as a research methodology has been employed in a range of disciplines, including the social sciences (Brannen 1992; Denzin 1989; Denzin 2006; Silverman 1985; Webb et al. 1981), nursing and health services research (Begley 1996; Fotheringham 2010; Mitchell 1986; Shih 1998; Sohier 1988), education (Altrichter et al. 2008; Cohen and Manion 2011; Oliver-Hoyo and Allen 2006), management (Jack and Raturi 2006), applied linguistics (Dörnyei 2007; Magnan 2006) and translation studies (see Chapter 1). The wide applicability of triangulation techniques has naturally led to different definitions of it, which are continuously being refined. The diversity of definitions, as well as their broadness, risk categorising any instance where two or more research methods are used as triangulation, ending up with "an 'anything goes as long as you mix them' mentality" (Dörnyei 2007, p. 46) and resulting in triangulation losing some of its research potential. In this chapter, and throughout the book, we will be using Denzin's (1970; 1989; 2012) account for two interrelated reasons. Firstly, Denzin was

the first to try and systematise triangulation; he has written extensively on the topic and has continuously refined it through the years by addressing the criticism directed to his work. This makes his account well-developed and organic, evolving through time, instead of resisting change, and thus becoming obsolete. Secondly, thanks to its flexibility, Denzin's account has been adopted by many other researchers in the social sciences and other disciplines, providing evidence for its significance.

Triangulation originates in mathematics, and in particular, positioning geometry, where triangulation literally means the act of making a triangle (Denzin 1989). Based on the concept of the triangle, it is possible to determine the position of an unknown object with the help of two known reference points. Let us consider the following example: a vessel at sea makes a distress call. One way of locating the ship is by triangulation. One rescue ship, picking up the distress call, may determine that the vessel is to the southeast, while another may determine that the vessel is to the northwest. The vessel will be at the intersection of the two lines of sight (Figure 2.1); the third point in a triangle. If we consider this example, it is not difficult to understand why triangulation was first used in land surveying and navigation, particularly in the military, as a technique for determining a ship's or aircraft's position (Ammenwerth et al. 2003). As such, it is still used today in modern GPS technology.

Triangulation acquired a less literal sense when it was first introduced in social sciences research in the discussion of nonreactive measurement (Webb et al. 1966), as a technique for discovering whether a hypothesis can be corroborated using different complementary testing methods. Denzin (1989) defines triangulation as the combination of multiple (two or more) theories, data sources, methods, or investigators in one study of a single phenomenon.[1] This definition highlights the central role of the research question since triangulation can be

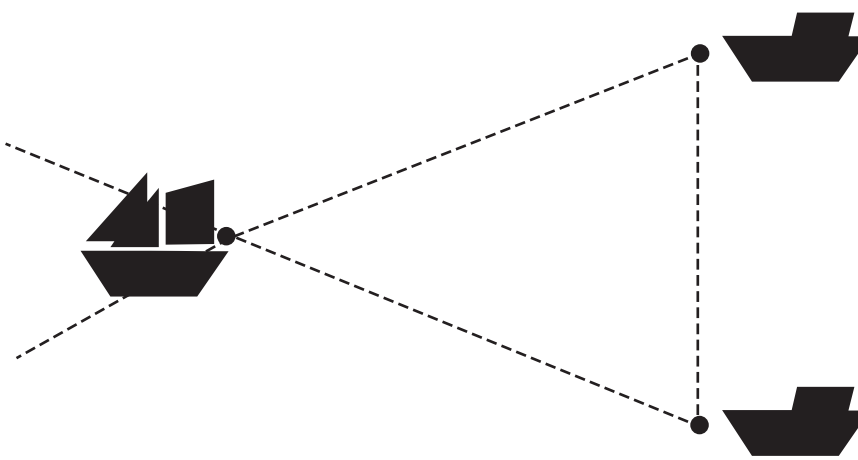

Figure 2.1 Using triangulation to locate a distressed vessel

considered as having taken place only when it is used to answer one and the same research question (Ammenwerth et al. 2003). The research question is important for an additional reason: the decision whether triangulation is to be used, and if so, which type of triangulation is the most appropriate depends on the research question at hand (Begley 1996; Cohen and Manion 2011; Flick 1992).

According to Dörnyei (2007, p. 46), it is important that triangulation is not equated to a mere collation of different methods and perspectives, and for this to be achieved research using triangulation needs to rely on "a principled approach". One way of getting closer to the principled approach advocated by Dörnyei is by examining more closely the importance of integration for triangulation. Moran-Ellis et al. (2006, p. 50) aptly summarise the concept of integration:

> Integrated transport provides a model of integration which we find useful as a metaphor for understanding methodological integration in cross-paradigm research. In transport, integration is 'the principle of ensuring transport modes operate in conjunction with one another' (Commission for Integrated Transport, UK 2005). Ideally, an integrated transport system allows a passenger to purchase one ticket for the whole journey despite changing between modes of travel (bus, train, plane), and alight from one vehicle and board the next at the same location with minimum waiting. In other words, to change vehicles effortlessly in pursuit of the goal of reaching a destination. In this system, the process is smooth, efficient and relatively trouble-free for the passenger because of the ordered, integrated relationships between the different modes of transport.

Triangulation can be understood in the same way: different approaches are combined in such a way that, while they do not lose their individual characteristics, through their combination researchers address a single research question. According to Moran-Ellis et al. (2006), it is important that when using triangulation techniques equal weight is given to the different methods, data and theories, with respect to operationalization (the bus is not more important than the train), and that they are understood as interdependent (it would have been impossible to reach our destination by using only one means of transport). What often happens in studies using triangulation techniques is that different methods or data are combined, but each of these addresses different research questions, and there is little or no interaction during analysis, when contradictions, divergences and convergences would be revealed. This is to an extent evident in previous studies employing combined corpora discussed in Chapter 1. However, if triangulation is to be successful, then integration needs to occur "from the point of conceptualisation and across all phases of the research" (Moran-Ellis et al. 2006, p. 54). Otherwise, there is a higher probability of error (Fielding and Fielding 1986), and the possible advantages of triangulation are undermined. An example of integration from the triangulation of different methods (e.g. qualitative and quantitative), is when the results obtained from one method point towards questions that can be addressed by the other (Risjord et al. 2002).

2.2 Types of triangulation

A large proportion of the available literature on triangulation focuses on the combination of qualitative and quantitative methods. Consequently, it is not surprising that triangulation has come to be equated to mixed methods research (Hurmerinta-Peltomäki and Nummela 2004), with other types of triangulation often being ignored. However, Denzin (1989) identifies many more types of triangulation, namely data, investigator, theory and methodological. There is also the possibility of multiple triangulation, where two (or more) of these types are combined, for example, data and theory triangulation. The types of triangulation are described below with suitable examples from a hypothetical study aiming at examining how undergraduate students experience stress. Not all of these types are expected to be relevant to corpus-based translation studies, but their discussion will help illustrate the considerable potential of triangulation and the fact that it is more elaborate than a simple mixing of different methods.

2.2.1 Data triangulation

Data triangulation occurs when multiple sources of data, with a similar focus, are combined (Denzin 1989), aiming at obtaining diverse views on the same phenomenon (Kimchi et al. 1991). It allows researchers to use the same methods on a range of data to achieve maximum theoretical advantage and increase the possible range of data that might contribute to a better understanding of the phenomenon under investigation (Knafl and Breitmayer 1989). Data triangulation can be refined even further, by focusing on one of the three subtypes: time, space and person. These subtypes result from the idea that research findings can be strengthened based on the time data were collected, the setting from which they were collected, and the people involved in their collection (Begley 1996). In sociological research, these three subtypes of data triangulation are interrelated, and the study of one might require the study of others.

Time triangulation occurs when data on the same phenomenon are collected at different intervals, for example, times, days, weeks, months or years. It is not simply about gathering data from various points in time, but rather about gathering data from various points in time, which are relevant to the study. It has been argued that longitudinal studies, for example, diachronic analyses, do not make use of time triangulation, since they are interested in examining how the phenomenon changes over time, rather than focusing on commonalities across time (Kimchi et al. 1991). Nevertheless, as will be explained later, divergence can also be important in triangulation. An example of time triangulation is the collection of data from undergraduate students on each of the three or four years of a degree to discover common experiences of stress across years. *Space triangulation* occurs when data on the same phenomenon are collected from different sites. The main aim of space triangulation is to test multi-site consistency (Shih 1998) and eliminate cross-site variation (Kimchi et al. 1991). As a result, it is important that space is the central variable of the study. An example of space triangulation is the collection of

data from undergraduate students from different universities in the same country (or even different countries) on their stress experiences. If congruence is demonstrated, the findings from one university will support the findings from all others. Another example might involve gathering data from different populations, that is cultures and ethnic groups, which are distributed geographically. *Person triangulation* occurs when data on the same phenomenon are collected from different levels of persons: individuals or groups. An example of person triangulation is the collection of data from different groups of people, for example, undergraduate students, academic staff, support staff, using focus groups (group level) and in-depth interviews with selected participants (individual level) to gain a better understanding of how students experience stress during their studies.

The main criticism against data triangulation is that each type of data will capture a different aspect of the phenomenon, and it is impossible for any two datasets to measure the same aspects. Any attempt to this will lead to false interpretations of the phenomenon under study (Porter 1989). Denzin (1989) addresses this by arguing that the aim of data triangulation is not to ignore the different picture of the world that each set of data captures, but rather to try and understand their differences and offer an interpretation of these. By acknowledging this, Denzin recognises that data triangulation (and triangulation in general) is interested not only in congruency and consistency, but also in divergence in the results.

2.2.2 Investigator triangulation

Investigator triangulation refers to the combination of multiple researchers, for instance, in the collection and analysis of data, to study a single phenomenon (Denzin 1989). According to Kimchi et al. (1991), three conditions are necessary for investigator triangulation to take place. Firstly, all investigators need to have a prominent role in the study. In many research projects, a principle investigator is working together with a co-investigator, as well as some other researchers, all interested in studying the same phenomenon. However, not all of these stakeholders will have an equally prominent role and the same responsibilities. Thus, these are not examples of successful investigator triangulation, even though we must recognise that some care has been taken to minimise subjectivity and bias. Secondly, each investigator needs to possess different expertise, which is relevant to the study. Having multiple investigators with the same expertise, although reduces subjectivity, does not increase the understanding of the phenomenon under study. Thus, co-authored studies between researchers belonging to the same discipline and following the same paradigms are not considered examples of investigator triangulation. Thirdly, the expertise of each investigator needs to be evident in the study; there needs to be a clear benefit from employing multiple investigators. Investigator triangulation occurs when, in the example of measuring stress, a researcher from the field of psychology and a researcher from education analyse the findings of the undergraduate students' experience of stress. If, for example, similar observations are made by the two researchers, then the reliability of conclusions is increased.

The aim of this type of triangulation is to neutralise or minimise the possible bias and subjectivity that occurs from employing a single investigator. It can also add reliability to the observations made by bringing together different perspectives and epistemological assumptions, which can inform the results (Rothbauer 2008). Investigator triangulation, although not difficult to achieve, is considered rather difficult to demonstrate (Kimchi et al. 1991), and has been criticised for assuming that it is possible for two investigators to make the same observation about a single phenomenon, and ignoring the essential element of subjectivity. However, this element of subjectivity is much larger in studies that employ a single investigator.

2.2.3 Theory triangulation

Theory triangulation occurs when the same empirical material is approached from different theoretical angles, "with multiple perspectives and hypotheses in mind" (Denzin 1978, p. 297). The aim of this type of triangulation is to support or refute different findings since the same data are tested against alternative theories. In this way, the shortcomings of each theory are addressed, and the strength of the conclusions is increased, providing a more in-depth understanding of the phenomenon under investigation (Banik 1993). Theory triangulation can most typically be observed in studies that focus on either theory testing, where it occurs from the outset (Shih 1998), or theory generation, where it occurs at the conclusion (Knafl and Breitmayer 1989). An example of theory triangulation would be the analysis of data from the experience of stress by undergraduate students using different theoretical models from psychology and learning theories. Theory triangulation is perhaps the type of triangulation that is most difficult to achieve, but it has the considerable advantage of addressing the criticisms expressed about any single theoretical framework, and also assessing the power and value of competing theories, hypotheses and interpretations (Denzin 1989). It also allows researchers not to ignore contradictory presuppositions and to consider all possible interpretations of data and enhances the significance of the findings, by allowing for the "widest possible theoretical use of any set of observations" (Denzin 1989, p. 242). The most significant advantage of theory triangulation, however, is that it allows us to develop coherent theoretical models for the study of the same phenomenon that leave fewer questions unanswered and fewer answers unquestioned (Risjord 2000).

Some criticism has been expressed against theory triangulation, mainly by Lincoln and Guba, who dismiss even its possibility by referring to it as "epistemologically unsound and empirically empty" (1985, p. 307). They base their criticism on the idea that it is not possible for a phenomenon to be consistent with different theoretical frameworks and that there will inevitably be contradictions. In other words, it is not possible to arrive at the same conclusion using different theories. Such criticism has misinterpreted the true aim of theory triangulation, which is to approach the phenomenon from various angles, without necessarily aiming for convergence. Finally, care must be taken when interpreting the results.

If a fact seems to be confirmed by two theories, this might be more a sign of the similarities of the theories rather than "of the empirical meaningfulness of the fact" (Lincoln and Guba 1985, p. 307).

2.2.4 Methodological triangulation

Perhaps the most distinctive type of triangulation is methodological triangulation (Duffy 1987; Fontana and Frey 2000; Mitchell 1986), which refers to the combination of different methods for the analysis of the same dataset. Methodological triangulation is chosen in a study as it highlights different aspects or dimensions of the phenomenon under investigation (Kimchi et al. 1991; Knafl and Breitmayer 1989). It can occur both at the level of research design (different research approaches) and data collection (different tools) (Burns and Grove 2005; Kimchi et al. 1991), and it can be simultaneous or sequential (Morse 1991). Simultaneous methodological triangulation occurs when different methods, usually qualitative and quantitative, are used at the same time, but the interaction between them is minimum, and findings are combined only at the end of the study. Sequential triangulation occurs when one method is treated as an essential step for planning the next method. For example, a researcher may use questionnaires to highlight areas that are important and then conduct interviews to examine these in more depth; it is the questionnaire that informs the areas on which the interview will focus. This is clearly related to issues of integration discussed earlier and suggests that greater integration can be achieved by using sequential methodological triangulation. Denzin (1989) distinguishes two subtypes of methodological triangulation: within-method and between-method, also known as across-method, triangulation.

Within-method triangulation occurs when "multiple complementary methods within a single given paradigm" (Hussein 2009, p. 4) are used in the same study to measure the same variable. This type of triangulation is best suited for multidimensional data. The way in which similarity between methods is interpreted is found to be quite subjective and even the use of different qualitative methods, for example, interviews and observation, has been argued to be an example of within-method triangulation by some (Begley 1996). However, this type of triangulation can also be understood in a much narrower sense, for instance using different types of questionnaires with different scales to measure stress experience in undergraduate students. The main limitation of this approach is that, even though multiple variations of the same method are used, in reality, only one method is employed. Thus, all the limitations of employing one method hold true, although this research design is still better than using a single method without any variation.

A preferred form of methodological triangulation according to Denzin (1989) is between-method triangulation, which occurs when two or more research strategies, techniques, or methods, are combined in the study of the same set of data. Between-method triangulation is the type that is most strongly associated with

the keyword triangulation, and it often means combining qualitative and quantitative methods. The methods employed in between-method triangulation can be dissimilar, yet complementary, and the main advantage of this form of triangulation is that investigators can benefit from the advantages of different methods, while at the same time minimise their limitations. For instance, in the study of stress experienced by undergraduate students, interviews with students might be combined with quantitative questionnaires, inferring statistical data. One of the main criticisms addressed towards between-method triangulation is that the inaccuracies of the data obtained using one approach will not necessarily be addressed by those obtained using another (Fielding and Fielding 1986). Morse (1991) responds to this criticism by advocating the importance of a primary method, which must be rigorous enough to sustain the study, while any additional method used further strengthens the study.

2.2.5 Multiple triangulation

There is one last type of triangulation, which we need to consider, and that is multiple triangulation. Multiple triangulation can be defined as a means of 'triangulating triangulation' since it involves the combination of "multiple methods, data types, observers and theories in the same investigation" (Denzin 1970, p. 472). For example, a study might combine the use of investigator triangulation and within-method triangulation, where each investigator will be familiar with a different research methodology. This type of triangulation offers additional advantages since the benefits of different types of triangulation are combined. Although not mentioned by Denzin, multiple triangulation can also occur by combining different subtypes of triangulation, for example, time and space data triangulation, or within-method and between-method triangulation. The latter combination will be particularly relevant when we apply triangulation to corpus-based translation studies in Section 2.4. Figure 2.2 offers a visual representation of how triangulation (in this example methodological triangulation) and multiple

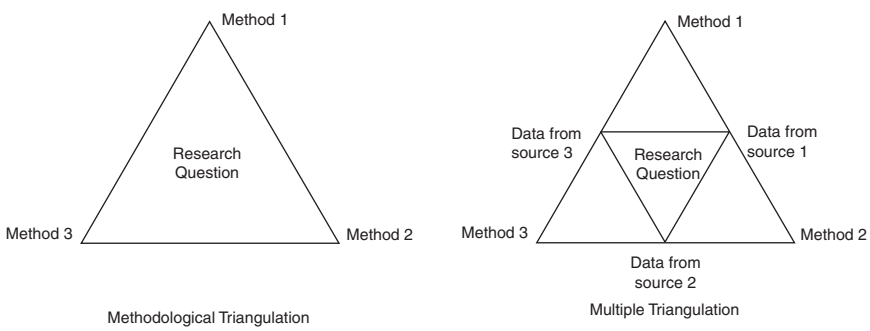

Figure 2.2 Graphic representation of triangulation

triangulation (in this example methodological and data triangulation) work. Here, the shape of the triangle is used, which is also linked to the etymology of the term *triangulation*. This is most appropriate when three different theories, sources of data, investigators and methods are employed. There are, of course, cases, where only two elements are used, or more than three. In the latter case, a more appropriate shape will be that of the crystal (for a detailed discussion see Richardson and St. Pierre 2005)

It must be noted that the advantages of multiple triangulation come at a cost. According to Dootson (1995), the main problem with multiple triangulation is its complexity. Each method and theory needs to be fully understood to achieve the full potential of multiple triangulation. Otherwise, bias might be increased (Duffy 1987; Morse 1991). As a result, multiple triangulation tends to be time-consuming (Mitchell 1986) and preferred by more experienced investigators (Begley 1996). In an attempt to address this particular problem, Mitchell (1986) stresses the importance of the research question, which needs to be as clearly focused as possible, to help guide the investigator select the right methods, theories and data. The central role that the research question plays in triangulation is, of course, not limited to multiple triangulation, but is relevant to any type of triangulation.

2.3 The purpose of triangulation

Successful triangulation can be achieved only if we have a clear understanding of what we hope to achieve from using this technique in the study of a phenomenon. In other words, we should be able to answer the question: What is the purpose of triangulation? The answer to this question becomes ever more important if we consider that there are contrasting views in the literature as to the ultimate purpose of triangulation.

Traditionally, triangulation has been considered as a means of either confirmation (Denzin 1970; Jick 1979; Webb et al. 1981) or validation (Denzin 1978). Denzin refers to methodological triangulation, arguing that it involves "a complex process of playing each method off against the other so as to maximise the validity of field efforts" (Denzin 1978, p. 304). According to this view, each new piece of data is expected to support or confirm existing data, the underlying assumption being that there exists some unified reality that can be captured (Modell 2009). As a result, triangulation is also often viewed as a means for increasing the reliability and accuracy of a study (Duffy 1987; Knafl and Breitmayer 1989; Mitchell 1986), increasing confidence (Jick 1979) and overcoming problems of bias (Blaikie 1991; Risjord et al. 2002). For confirmation to be obtained, investigators need to have a clear understanding of the strengths and weaknesses of different theories and methods and counterbalance these to increase validity. However, the possibility of using triangulation for validation purposes has been heavily criticised, mainly by Silverman, who argues against Denzin's idea of a "master reality" and a "total picture of some phenomenon"

(1985, p. 21). Although validation is possible when results converge, the real question that needs to be addressed is how to interpret divergence. Hammersley and Atkinson rightly observe that it is not the case that "the aggregation of data from different sources will unproblematically add up to produce a more complete picture" (1983, p. 199). In many cases, the differences between various types of data are more revealing than their similarities, as long as the triangulation of these is conducted in an integrated manner. In such cases, confirmation or validation is problematic, which means that we might need to talk about a different purpose of triangulation.

As a response to the criticism addressed against the increased validity view, some scholars advocate that the purpose of triangulation is comprehensiveness or completeness (Redfern and Norman 1994; Fielding and Fielding 1986), that is, the potential for increasing knowledge about a phenomenon. Triangulation, according to this view, is used to acquire a complete picture of the phenomenon under study, "to get additional pieces to the overall 'puzzle'" (Ammenwerth et al. 2003, p. 239). This view acknowledges that there are multiple realities (Tobin and Begley 2004), instead of a single 'true' version; a kaleidoscopic picture, instead of a uniform one (Denzin 1989). It recognises the complementarity of different methods, data, theories and investigators, and supports that they can be triangulated to increase our understanding of the complex nature of various phenomena. Completeness is achieved when the results obtained from one part of the study present results that have not been found in other parts (Ammenwerth et al. 2003); it is, in a way, the exact opposite of validation. Compared to validity, when triangulation is used for completeness purposes, the different results do not suggest flawed measurements, but rather reflect alternative aspects of the phenomenon (Moran-Ellis et al. 2006). Denzin later revised his view on the outcomes of triangulation, acknowledging the limitations of the increased validity view. He focused more on the interpretive potential of triangulation and admitted that "objective reality can never be captured", and proposed that triangulation should be regarded "not as a tool or a strategy for validation but an alternative to validation" (Denzin 2012, p. 82).

2.4 Towards corpus triangulation

Using Denzin's (1989) account of triangulation as a starting point, two main types of corpus triangulation are identified in this book: corpus data and corpus method triangulation. The other types of triangulation, namely investigator and theory triangulation are not discussed since previous studies in corpus-based translation studies suggest a combination of investigators or theories does not occur very frequently. In particular, corpus-based translation studies focus on the use of a specific methodology, i.e. corpora, for the study of translation, and as a result are more concerned with data and methods. Since theory and investigator triangulation cannot be easily associated with corpora, they are excluded from the discussion of corpus triangulation in this book. However, they are possible within

the wider context of translation research. Taking these into consideration, the definition of corpus triangulation proposed in this book is the following:

> Corpus triangulation is the combination, in an integrated manner, of multiple (two or more) corpus values and/or attributes from one or more corpus variables and/or the use of (two or more) corpus analysis techniques in one study of a single phenomenon.

We might be tempted to consider that corpus data triangulation can be simply defined as the combination of different corpora in a single study, for example, the combination of a comparable monolingual corpus and a parallel bilingual corpus. No matter how attractive such a simple definition of corpus data triangulation might be, it presents a problem of interpretation, given the availability of the different corpus typologies in the literature. To avoid the risk of loose interpretation, corpus triangulation needs to go hand in hand with a corpus typology that is well-defined and flexible at the same time. A typology that is based on the idea of variables, values and attributes can offer a solution to this. This is necessary to ensure that there is a common understanding of which corpora can be combined and how, and avoid further ad hoc attempts towards corpus triangulation. Such a typology is provided in Chapter 3. Thus, a corpus data triangulation definition that relies on the idea of corpus variables, values and attributes offers some guidance regarding what exactly can be combined when we refer to corpus triangulation and at the same time allows flexibility regarding which specific values/attributes need to be combined for each study.

Corpus method triangulation refers to the combination of different methods for the analysis of the same corpus. It can be further divided into within-method and between-method. Within-method corpus triangulation occurs when different quantitative techniques are employed in the analysis of the same corpus, while between-method corpus triangulation occurs when both qualitative and quantitative techniques are employed. Since corpora provide by definition quantitative data (they are, after all, large electronic collections of texts), within-method corpus triangulation cannot occur with qualitative techniques. This does not mean that qualitative techniques cannot be combined in corpus-based studies. Their combination might involve a close examination of corpus data from different perspectives. However, even in such a case, some quantitative analysis, no matter how basic, will also need to be conducted. In other words, quantitative methods are always considered primary in corpus-based studies. The combination of corpus methods can also occur at different stages of the study, either at the end of the analysis stage (simultaneous triangulation) or during each step of the analysis (sequential), where one type of analysis is a prerequisite for the next. For example, it is typical for quantitative analyses to highlight areas that would benefit from

a qualitative analysis, in the form of a close examination of some aspect of the corpus data.

Finally, multiple corpus triangulation occurs when corpus data triangulation is combined with corpus method triangulation, for example, in a study that employs both comparable and parallel corpora and analyses these using both quantitative and qualitative techniques. In this case, corpus data triangulation will normally take place before corpus method triangulation. Multiple corpus triangulation can also occur when both within-method and between-method triangulation are employed. For example, when results obtained from two or more quantitative analyses are combined with those obtained from a qualitative analysis.

The principle that unites the different corpora employed and the various types of corpus analysis performed is the research question, which plays a central role and guides the triangulation process. Thus, the way in which the research question is formulated has significant implications on the type of corpora, which need to be designed, and the way in which these can be analysed, and ultimately on how triangulation is achieved. It is important to remember that triangulation occurs only when all the different corpora and/or corpus analyses used provide an answer to the same research question(s). Cases where different corpora address different, yet related, research questions cannot be considered instances of corpus triangulation. For example, it is often the case that some elements of corpus design are difficult to link with a specific aspect of the research question, even though their contribution to the study might be obvious. This is particularly the case when reference corpora are used for comparison purposes (see for example, Liao 2010; Scarpa 2006).[2] Additionally, some corpus studies make use of fewer corpora than it might be suggested by their research aims, which, in turn, raises questions as to the extent to which these studies have managed to achieve a holistic view of the phenomenon under investigation. A typical example can be found in studies interested in the examination of translated language. Such studies are often limited to analyses of a comparable monolingual corpus, sometimes accompanied by an analysis of a monolingual reference corpus (see for example, Olohan and Baker 2000). However, they are rarely accompanied by an analysis of the source texts, and/or the analysis of multilingual corpora, a problem which has also been identified by Zanettin (2000). Such studies have missed the possibility of corpus data triangulation, either because they have not taken advantage of the potential of combining different corpora, or because they have not used an integrated approach towards the combination of corpora.

When it comes to corpus triangulation, it is also important to have a clear working definition of the linguistic feature under investigation. If the linguistic feature is not clearly defined, the inconsistencies in the data obtained from the corpus might be too high, and, as a result, they might be difficult to detect and correct. This is particularly important to bear in mind when analysing more than one corpora (e.g. belonging to different languages or genres), as the same linguistic feature might be realised differently in different languages, or have a different function depending on the genre. A smaller pilot corpus can be used

in cases where a linguistic feature proves difficult to define and capture, which should help fine tune the research parameters.

Given the considerable advantages of completeness compared to validation, the purpose of corpus triangulation should be to increase the knowledge about the linguistic and/or translation phenomenon under investigation, by approaching it from different perspectives and examining various relevant parameters. According to Hussein (2009), completeness is also more appropriate when studying complex phenomena, less explored or unexplored research questions, as is often the case with translation. Additionally, for those new to triangulation techniques, the purpose of completeness is easier to master, compared to that of validity, and thus it should be their primary aim. However, it must be noted that there are still many researchers who treat triangulation as a validation technique, not only in the social sciences (Blaikie 1991; Dootson 1995; Oliver-Hoyo and Allen 2006), but also applied linguistics (Dörnyei 2007), and translation studies (Saldanha and O'Brien 2013; Shreve and Angelone 2010).

2.5 Advantages and limitations

The main advantage of corpus triangulation is that it can offer detailed data, which would not be easily obtained using just one corpus and/or a single method. As in social sciences research, corpus triangulation reduces bias and increases the confidence in the results (Fielding and Fielding 1986; Redfern and Norman 1994; Seale 1999). Although confidence is typically associated with triangulation, caution should be exerted, as increased confidence does not suggest that the data are unquestionable, but rather that triangulation results in rich data. Triangulation also helps bring to the fore possible contradictions in the data, which allows for a more profound interpretation of the results (Altrichter et al. 2008). In turn, where contradictions are not observed, and the different perspectives converge, credibility can be added to the interpretation (Sands and Roer-Stier 2006). Additionally, corpus triangulation can help deepen our understanding of a phenomenon, by adding more knowledge (Seale 1999), or by "turning the prism" (Wilson and Hutchinson 1991, p. 274) to view a phenomenon from a different perspective, adding breadth, depth, complexity, richness and rigor to any inquiry (Denzin 2012). This particular strength of triangulation is very successfully captured by Wilkinson (2007, p. 631) in his photography metaphor, when he argues that "multiple snapshots, even if some are not totally in focus, give a better picture than one poorly aimed photograph". This suggests that by using triangulation more coherent results can be obtained, offering a more detailed and balanced picture. Finally, by analysing different types of corpora, the limitations associated with each corpus type are addressed, while their advantages are combined. Triangulation is also recommended when a more holistic view of a phenomenon is required (Cohen and Manion 2011), for example when different aspects need to be studied. This is even more so the case if the phenomenon under investigation is particularly complex (Cohen and Manion 2011). Translation, by its nature, is a complex phenomenon, which is situated between languages and cultures and can

be studied from a range of different perspectives, including comparisons between translated and/or non-translated material. For this reason, corpus triangulation can significantly increase our understanding of translation phenomena.

Regarding limitations, conflicting results might prove difficult to interpret, for example, if evidence from one type of corpus analysis contradicts that of another. In such occasions, it is important to be able to control all parameters that affect the corpus design to be in a position to interpret this contradiction in the data. Another potential problem for corpus triangulation is the unit of analysis, which needs to remain consistent (Redfern and Norman 1994). Especially in corpus method triangulation, where qualitative and quantitative methods are combined, the unit of analysis that is appropriate for one method might not be appropriate for another. For example, a study that aims at examining the use of conjunctions in translation might triangulate methods by examining the frequency of individual conjunctions, and by closely examining how they contribute to the function of texts. For the former, the unit of analysis will be the individual conjunction, while for the latter it will be the text. This creates problems of comparability and raises the question of the potential incompatibility of different methods and/or data. Within-method corpus triangulation presents the additional limitation that it only uses a single method, which refers to quantitative analyses. However, if care is taken during the design of the methodology, this limitation can be addressed by employing a wide range of possible quantitative approaches to reduce the bias of a single approach. Similarly, the fact that corpus-based research typically involves a single investigator creates a researcher bias. Corpus triangulation is no exception to this, but the fact that different corpora and methods are employed increases objectivity when compared to single corpus/methods approaches. Finally, corpus triangulation, as any type of triangulation, is time-consuming and challenging to design and undertake, as well as expensive to implement (Denzin 1989; Redfern and Norman 1994). This limitation is further aggravated by the lack of a clear methodological framework, which can be easily adopted in corpus-based translation research. As a result, many researchers, even though they recognise the need for triangulation, are not in a position to benefit from it. The present book aims at addressing this limitation by providing for the first time a comprehensive account of corpus triangulation, which gives equal weight to theoretical and methodological considerations of corpus triangulation, as well as to the applications of the methodology to real data with examples from a range of different contexts.

2.6 Conclusion

Triangulation is a flexible technique that presents many advantages, which can be easily extended to corpus based-translation studies. It is not simply an alternative way of obtaining results. Rather, it is a way of getting results that could not have been obtained otherwise. Given the multiple advantages of triangulation techniques, one cannot help but wonder: Why do not researchers in translation studies use them more often? We can speculate many possible reasons for this, similar to what Dörnyei (2007) suggests for applied linguistic research.

One possible explanation is the lack of sufficient knowledge on triangulation techniques and a lack of expertise in implementing these. This might be related to the lack of available literature, which would provide a guided and principled account of main triangulation techniques and how they might be used in translation studies. As already mentioned, existing literature in translation studies touches lightly upon triangulation, which does not make it easy for other researchers to familiarise themselves with such techniques and use them in their own research, while previous attempts at corpus triangulation have been rather poorly executed. More awareness needs not be created around triangulation techniques more generally, and corpus triangulation techniques, in particular, for researchers to be able to use them with confidence. Although this book does not have the ambition to present a systematic account of triangulation in translation studies research more generally, it does, however, aim to provide a rigorous and comprehensive account of how triangulation can be achieved within corpus-based translation studies.

Notes

1 No guidance as how many and which theories, sources, data and investigators can be combined is found in the literature, and it is not infrequent for studies to only employ two of these (e.g. a qualitative and a quantitative method). Oppermann (2000) reminds us of the importance of the third measurement, arguing that true triangulation should consist of at least three sets of theories, methods, etc.
2 Such comparisons are often seen as additional analyses, and do not form part of the corpus design (for example, reference corpora, and their design, are not part of the methods and data sections).

References

Altrichter, H., Posch, P. and Somekh, B., 2008. *Teachers Investigate Their Work: An Introduction to the Methods of Action Research*. London and New York: Routledge.

Ammenwerth, E., Iller, C. and Mansmann, U., 2003. Can Evaluation Studies Benefit From Triangulation? A Case Study. *International Journal of Medical Informatics*, 70(2–3), pp. 237–248.

Banik, B.J., 1993. Applying Triangulation in Nursing Research. *Applied Nursing Research*, 6(1), pp. 47–52.

Begley, C.M., 1996. Using Triangulation in Nursing Research. *Journal of Advanced Nursing*, 24(1), pp. 122–128.

Blaikie, N.W.H., 1991. A Critique of the Use of Triangulation in Social Research. *Quality and Quantity*, 25(2), pp. 115–136.

Brannen, J., 1992. Combining Qualitative and Quantitative Approaches: An Overview. In J. Brannen, ed. *Mixing Methods: Qualitative and Quantitative Research*. Aldershot: Avebury, pp. 3–37.

Burns, N. and Grove, S.K., 2005. *The Practice of Nursing Research: Conduct, Critique, and Utilization*. St. Louis, MI: Elsevier/Saunders.

Cohen, L. and Manion, L., 2011. *Research Methods in Education* 7th ed. London and New York: Routledge.

Denzin, N.K., 1970. *The Research Act: A Theoretical Introduction to Sociological Methods* 1st ed. Chicago: Aldine Pub. Co.

Denzin, N.K., 1978. *The Research Act: A Theoretical Introduction to Sociological Methods* 2nd ed. New York: McGraw-Hill.

Denzin, N.K., 1989. *The Research Act: A Theoretical Introduction to Sociological Methods* 3rd ed. Englewood Cliffs: Prentice Hall.

Denzin, N.K., 2006. *Sociological Methods: A Sourcebook*. Chicago: Aldine Pub. Co.

Denzin, N.K., 2012. Triangulation 2.0. *Journal of Mixed Methods Research*, 6(2), pp. 80–88.

Dootson, S., 1995. An In-Depth Study of Triangulation. *Journal of Advanced Nursing*, 22, pp. 183–187.

Dörnyei, Z., 2007. *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.

Duffy, M.E., 1987. Methodological Triangulation: A Vehicle for Merging Quantitative and Qualitative Research Methods. *Image: The Journal of Nursing Scholarship*, 19(3), pp. 130–133.

Fielding, N.G. and Fielding, J.L., 1986. *Linking Data*. Thousand Oaks, CA: Sage Publications.

Flick, U., 1992. Triangulation Revisited: Strategy of Validation or Alternative? *Journal for the Theory of Social Behavior*, 22(2), pp. 175–197.

Fontana, A. and Frey, J.H., 2000. The Interview: From Structured Questions to Negotiated Text. In N.K. Denzin and Y.S. Lincoln, eds. *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage Publications, pp. 645–672.

Fotheringham, D., 2010. Triangulation for the Assessment of Clinical Nursing Skills: A Review of Theory, Use and Methodology. *International Journal of Nursing Studies*, 47(3), pp. 386–391.

Hammersley, M. and Atkinson, P., 1983. *Ethnography: Principles in Practice*. London: Routledge.

Hurmerinta-Peltomäki, L. and Nummela, N., 2004. First the Sugar, Then the Eggs . . . Or the Other Way Round? In R. Piekkari and C. Welch, eds. *Mixing Methods in International Business Research: Handbook of Qualitative Research Methods for International Business*. Cheltenham: Edward Elgar, pp. 162–180.

Hussein, A., 2009. The Use of Triangulation in Social Sciences Research: Can Qualitative and Quantitative Methods Be Combined? *Journal of Comparative Social Work*, 1, pp. 1–12.

Jack, E.P. and Raturi, A.S., 2006. Lessons Learned From Methodological Triangulation in Management Research. *Management Research News*, 29(6), pp. 345–357.

Jick, T.D., 1979. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4), pp. 602–610.

Kimchi, J., Polivka, B. and Stevenson, J.S., 1991. Triangulation: Operational Definitions. *Nursing Research*, 40(6), pp. 364–366.

Knafl, K.A. and Breitmayer, B.J., 1989. Triangulation in Qualitative Research: Issues of Conceptual Clarity and Purpose. In J. Morse, ed. *Qualitative Nursing Research: A Contemporary Dialogue*. Aspen, Rockville, MD: Sage Publications, pp. 209–220.

Liao, M.-H., 2010. Influence of Translations on Non-Translations: Popular Science as a New Genre in Taiwan. *Perspectives: Studies in Translatology*, 18(2), pp. 127–140.

Lincoln, Y.S. and Guba, E.G., 1985. *Naturalistic Inquiry*, Thousand Oaks, CA: Sage Publications.

Magnan, S.S., 2006. From the Editor: The MLJ Turns 90 in a Digital Age. *The Modern Language Journal*, 90(1), pp. 1–5.

Mitchell, E.S., 1986. Multiple Triangulation: A Methodology for Nursing Science. *ANS: Advances in Nursing Science*, 8(3), pp. 18–26.

Modell, S., 2009. In Defence of Triangulation: A Critical Realist Approach to Mixed Methods Research in Management Accounting. *Management Accounting Research*, 20(3), pp. 208–221.

Moran-Ellis, J. et al., 2006. Triangulation and Integration: Processes, Claims and Implications. *Qualitative Research*, 6(1), pp. 45–59.

Morse, J., 1991. Approaches to Qualitative-Quantitative Methodological Triangulation. *Nursing Research*, 40(1), pp. 120–123.

Oliver-Hoyo, M. and Allen, D., 2006. The Use of Triangulation Methods in Qualitative Educational Research. *Journal of College Science Teaching*, 35(4), pp. 42–47.

Olohan, M. and Baker, M., 2000. Reporting That in Translated English: Evidence for Subconscious Processes of Explicitation? *Across Languages and Cultures*, 1(2), pp. 141–158.

Oppermann, M., 2000. Triangulation: A Methodological Discussion. *International Journal of Tourism Research*, 146(2), pp. 141–146.

Porter, E.J., 1989. The Qualitative-Quantitative Dualism. *Image: The Journal of Nursing Scholarship*, 21(2), pp. 98–102.

Redfern, S.J. and Norman, I.J., 1994. Validity Through Triangulation. *Nurse Researcher*, 2, pp. 41–56.

Richardson, L. and St. Pierre, E.A., 2005. Writing: A Method of Inquiry. In K. Denzin, Norman and Y.S. Lincoln, eds. *The SAGE Handbook of Qualitative Research*. Thousand Oaks, CA: Sage Publications, pp. 959–978.

Risjord, M.W., 2000. *Woodcutters and Witchcraft: Rationality and Interpretive Change in the Social Sciences*. Albany: SUNY Press.

Risjord, M.W., Dunbar, S.B. and Moloney, M.F., 2002. A New Foundation for Methodological Triangulation. *Journal of Nursing Scholarship: An Official Publication of Sigma Theta Tau International Honor Society of Nursing/Sigma Theta Tau*, 34(3), pp. 269–275.

Rothbauer, P.M., 2008. The SAGE Encyclopedia of Qualitative Research Methods. In L.M. Given, ed. *The SAGE Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: Sage Publications, pp. 892–894.

Saldanha, G. and O'Brien, S., 2013. *Research Methodologies in Translation Studies*. London and New York: Routledge.

Sands, R.S. and Roer-Stier, D., 2006. Using Data Triangulation of Mother and Daughter Interviews to Enhance Research About Families. *Qualitative Social Work*, 5(2), pp. 237–260.

Scarpa, F., 2006. Corpus-Based Quality-Assessment of Specialist Translation: A Study Using Parallel and Comparable Corpora in English and Italian. In M. Gotti and S. Sarcevic, eds. *Insights Into Specialized Translation*. Bern: Peter Lang, pp. 155–172.

Seale, C., 1999. Quality in Qualitative Research. *Qualitative Inquiry*, 5(4), pp. 465–478.

Shih, F.J., 1998. Triangulation in Nursing Research: Issues of Conceptual Clarity and Purpose. *Journal of Advanced Nursing*, 28(3), pp. 631–641.

Shreve, G.M. and Angelone, E., 2010. *Translation and Cognition* G.M. Shreve and E. Angelone, eds. Amsterdam and Philadelphia: John Benjamins.

Silverman, D., 1985. *Qualitative Methodology and Sociology: Describing the Social World*. Brookfield, VT: Gower Publishing.

Sohier, R., 1988. Multiple Triangulation and Contemporary Nursing Research. *Western Journal of Nursing Research*, 10(6), pp. 732–742.

Tobin, G.A. and Begley, C.M., 2004. Methodological Rigour Within a Qualitative Framework. *Journal of Advanced Nursing*, 48(4), pp. 388–396.

Webb, E.J. et al., 1966. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.

Webb, E.J. et al., 1981. *Nonreactive Measures in the Social Sciences*. Boston: Houghton Mifflin.

Wilkinson, T.J., 2007. Assessment of Clinical Performance: Gathering Evidence. *Internal Medicine Journal*, 37(9), pp. 631–636.

Wilson, H.S. and Hutchinson, S.A., 1991. Triangulation of Qualitative Methods: Heideggerian Hermeneutics and Grounded Theory. *Qualitative Health Research*, 1(2), pp. 263–276.

Zanettin, F., 2000. Parallel Corpora in Translation Studies: Issues in Corpus Design and Analysis. In M. Olohan, ed. *Intercultural Faultlines: Research Models in Translation Studies I: Textual and Cognitive Aspects*. Manchester: St. Jerome, pp. 105–118.

3 Corpus data triangulation

3.0 Introduction

The popularity of corpora in translation studies and their wide applicability has inevitably resulted in ambiguities in the way some of the key terms have been used. However, if we want to develop a framework that relies on the combination of different corpora, it is important to understand what constitutes a corpus and draw clear distinctions among the different corpora. This chapter begins by providing a comprehensive definition of what constitutes a corpus in corpus-based translation studies and a brief overview of existing typologies of corpora in the field. The discussion of previous corpus typologies demonstrates the need for a new typology based on the idea of corpus variables, values and attributes, which forms the basis of corpus data triangulation. This new VVA typology is introduced here together with a detailed examination of its parameters and how these can be combined to achieve corpus data triangulation. Lastly, we shall examine in this chapter how corpus size, representativeness and balance are affected when corpus triangulation is employed in a project.

3.1 Corpora in translation studies

Many attempts to define *corpus* and its key features have been made in translation studies throughout the years, which are naturally very close to the way in which corpora are understood in corpus linguistics. One of the first definitions is that proposed by Baker, according to which a corpus is "a collection of texts held in machine-readable form and capable of being analysed automatically or semi-automatically in a variety of ways" (Baker 1995, p. 225). Machine-readability is an important characteristic of corpora as it allows for much more systematic and reliable analyses (Mason 2001). Thus, collections of texts which are not machine-readable cannot be considered as corpora in the strict sense of the term. In more recent definitions, such as the one provided by Olohan (2004, p. 1), the electronic aspect of corpora is further highlighted: "[t]he texts are held in electronic format, i.e. as computer files, so that various kinds of corpus tools, i.e. software, can be used to carry out analysis on them". Similarly, Zanettin (2012, p. 7) sees a corpus as "a collection of texts in electronic format which are processed and analysed using software specifically created for linguistic research". Such definitions clearly

distinguish corpora from simple electronic and non-electronic collections of texts, which can be analysed manually, even more so in recent years with the plethora of specifically designed electronic tools. However, although most of the corpus analysis is conducted automatically, for example, by generating concordance lines, part of it is carried out manually. Thus, in reality, most corpora are analysed semi-automatically, as suggested by Baker. Other characteristics of corpora are specific design criteria and a clear purpose (Baker 1995; Johansson 1998; Olohan 2004; Zanettin 2012), which is also what distinguishes corpora from other electronic collections of texts. Based on the above, the corpus definition employed in this book is the following:

> A corpus is an electronic collection of texts, which is compiled according to specific design criteria, and which can be analysed automatically or semi-automatically using different types of software specifically created for linguistic research.

Ever since corpora were introduced in translation research, scholars have tried to develop a corpus typology, which would incorporate the different types of corpora typically used in corpus-based translation studies (Altenberg and Granger 2002; Bowker and Pearson 2002; Johansson 1998; Johansson 2003; Laviosa 2002; Tognini-Bonelli 2001; Zanettin 2012). A comprehensive evaluation of their descriptive potential reveals that these typologies present significant limitations, particularly if they are to inform corpus triangulation practices. Therefore, we need a new corpus typology, which clearly presents the various corpus categories and how these might be combined, and informs the definition of corpus triangulation. At the same time, this corpus typology needs to be flexible, not only describing which corpora have been used until now, but also capturing possible corpora that might be developed in the future. Before we introduce this new corpus typology, we will revisit two of the available corpus typologies to offer some examples of their limitations. Laviosa's (2002) typology has been chosen as it is the most comprehensive in the literature, while Zanettin's (2012) typology is the most recent and economical.

Laviosa (2002) was the first to introduce the idea of a corpus typology to corpus-based translation studies. She bases her typology on a number of parameters, which can be general or specific, and identifies four levels (I–IV) of corpus description. Figure 3.1 provides a visual depiction of the typology. Although the different levels are described in detail, they are not given descriptive names or labels, which is problematic when we try to refer to them. This might be a reason why other scholars in the field have not subsequently successfully adopted this typology.

The typology becomes more complex if we consider that Laviosa identifies additional categories for Level I (Table 3.1), apart from the number of languages

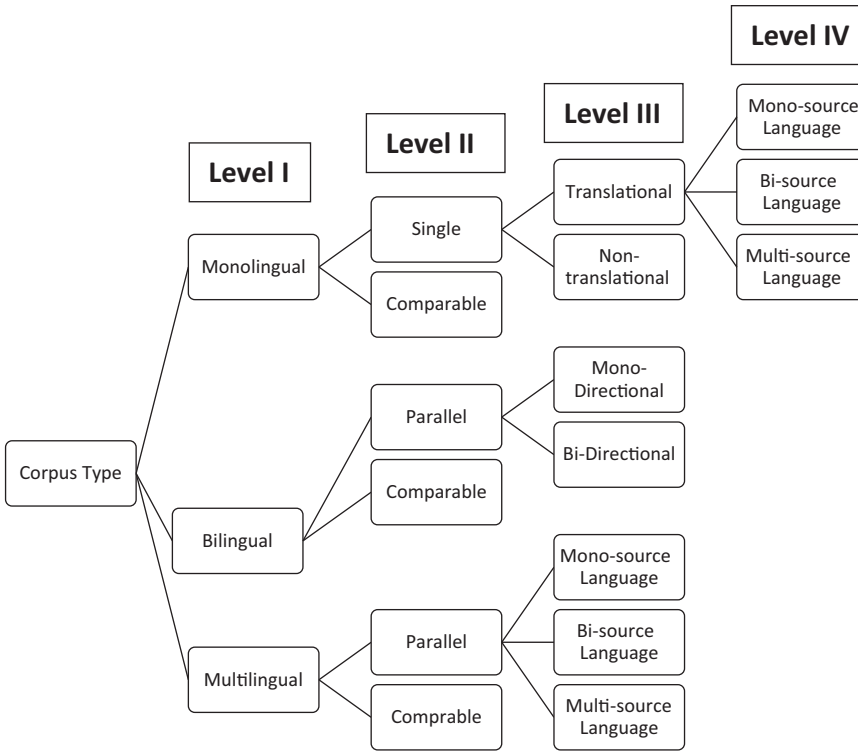

Figure 3.1 Laviosa's corpus typology

Table 3.1 Additional Level I categories of Laviosa's corpus typology

Full-text	Synchronic	General	Language(s)	Written
Sample	Diachronic	Terminological		Spoken
Mixed				Mixed
Monitor				

included in the corpus (Figure 3.1). These additional categories for Level I are not further specified, as is the case with the number of languages (Levels II–IV).

Laviosa (2002) acknowledges that this typology is not exhaustive, but rather serves as a point of departure, where more parameters can be added according to individual research projects. Its main advantage is that it aims at comprehensiveness. However, its main limitations are the complexity and lack of economy. The corpus typology is repetitive, both within levels and across levels. For instance, the comparable category is repeated three times in Level II, while the number of source languages included in the corpus is mentioned in both Level III and Level IV. Descriptive labels, which would facilitate reference to

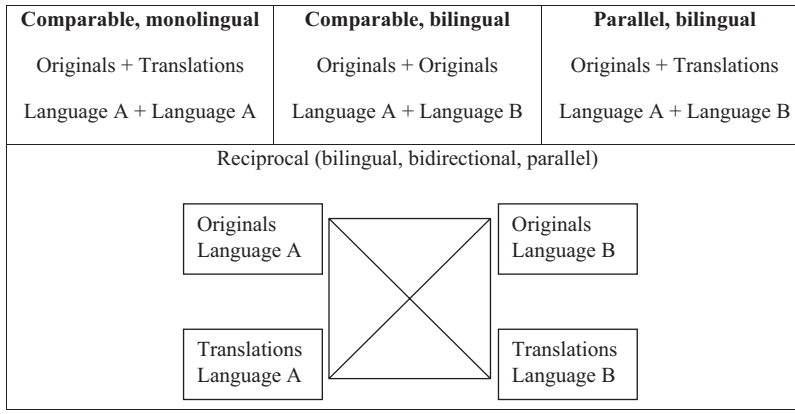

Figure 3.2 Zanettin's corpus typology (taken from Zanettin 2012, p. 11)

different categories, are also missing. For example, a label such as *Medium* could have been used for the written/spoken/mixed category. As can be seen from this brief description of Laviosa's typology, although it tries to account for as many different types of corpora as possible, it is a rather complex typology.

Zanettin's (2012) attempt at a typology of corpora in translation studies resulted in a concise and economical model (Figure 3.2). The main strength of this typology is that it describes existing types of corpora, but also identifies the elements that constitute them, pointing towards the possibility of combining existing corpora, especially as far as the model of the reciprocal corpus is concerned. However, despite its economy, this typology is perhaps too restrictive, and it might not be easy to add other types of corpora, for example, synchronic or diachronic corpora, without resulting in a complex corpus typology such as Laviosa's. Thus, although Zanettin's corpus typology succeeds in capturing the combined use of corpora, if a corpus triangulation model was to be based on such a typology it would have been necessarily limited to triangulation based on only reciprocal corpora, that is, the combination of comparable and parallel corpora in two languages.

3.2 A new corpus typology

The discussion of limitations of existing corpus typologies demonstrates the strong need for a new corpus typology, which is the golden mean between Laviosa's detailed and flexible, yet intricate, typology and Zanettin's concise, yet restricted model. This new typology will need to clearly present corpora based on different criteria, and allow for new corpora to be easily added, as corpus-based translation research advances. The idea of parameters introduced by Laviosa is worthwhile and can be extended to capture variables, values and attributes, which have proven useful in other disciplines, such as information science. A corpus

typology that relies on these can offer corpus-based translation studies significant descriptive potential, considerably facilitate the selection of corpus components from different sources and allow for a common and clear terminology to be used across corpus-based translation studies.

The first step in designing this corpus typology is to identify the most significant aspects of corpora, which will constitute the variables, for example, the type of corpus and languages. The variables are then assigned specific values, which form close-ended categories, for example, parallel, comparable and reference for the type of corpus variable. Values can be further defined, if needed, regarding specific attributes, for example, the specific languages for the languages variable.[1] Although it might appear that attributes replace their respective values, it is generally good practice to specify both when describing a corpus. For instance, if a corpus is described as *a comparable German–Russian corpus*, it is implied that the corpus is bilingual. For reasons of completeness and clarity, and to facilitate comparisons across available corpora, it is preferable to describe the corpus as *a comparable bilingual (German–Russian) corpus.* It is also recommended that the order of variables as presented in Table 3.2 is followed and that attributes appear in a parenthesis since they often provide additional information, for example, *a parallel bilingual (Italian–English) synchronic (2002–2003) corpus (news articles)*. This Variables–Values–Attributes (VVA) typology is flexible, in the sense that, while it allows for a clear understanding of the most central aspects of corpora, variables relevant to individual projects (e.g. annotation and alignment, translator/author profile) can be added if necessary.

As discussed in Chapter 2, corpus data triangulation is the combination, in an integrated manner, of multiple (two or more) corpus values and/or attributes from one or more corpus variables. Thus, corpus data triangulation is characterised by considerable flexibility and involves different possible combinations, based on the research aim of the project. These are listed below:

- the combination of values from the same variable, for example synchronic and diachronic corpora (time variable), or translated and non-translated texts (text variable)

Table 3.2 VVA (variables, values, and attributes) typology of corpora

	Corpus Variables			
	Type	*Languages*	*Time*	*Texts*
Corpus Values	Parallel Comparable Reference	Monolingual Bilingual Multilingual	Synchronic Diachronic	Translated Non-translated
Corpus Attributes	–	Specific languages	Specific time spans	Specific genres

- the combination of attributes from the same value, for example using two comparable monolingual corpora in different languages (specific languages attribute), or comparing the translation of different genres (genres attribute).
- the combination of values from different variables, for example employing a comparable monolingual corpus and a parallel bilingual corpus (type and languages variable)
- the combination of attributes from different values, for example comparing the translation of different genres in different languages (genres and specific languages attribute)

For instance, a study focusing on the examination of the use of creative language in the translation of fiction from English into German (similar to that conducted by (Kenny 2001b)) can combine the following corpora:

i a parallel, bilingual (English-German), synchronic (1980–1999) corpus (fiction)
ii a comparable, monolingual (German), synchronic (1980–1999) corpus of translated and non-translated texts (fiction)

Corpus triangulation, in this case, can be achieved by combining corpus values from three variables—corpora i and ii differ regarding the type, languages and texts variable. The exact way in which corpora will be combined will be dictated by the focus of the research study and the specific research question(s). The aim of the corpus project plays a central role in any corpus project (Hunston 2002; Kennedy 1998; Olohan 2004), but it is even more important in corpus data triangulation, since it can only occur when all the different elements of the corpus project provide an answer to the same research question or test the same hypothesis. In other words, the more concrete the research aim of the project, the easier it will be to draw a successful corpus outline and apply corpus data triangulation.

By focusing on specific corpus variables, values and attributes, instead of corpora in general, researchers are encouraged to incorporate triangulation considerations in their study from the formulation of the research questions to the corpus design and to reflect on which specific aspects of their corpora they are triangulating and why. This allows corpus triangulation to take place in an integrated manner. Corpus data triangulation also means deciding on how many and which values and attributes will be combined. Because their number can be quite large (especially if more variables are added to the original corpus typology presented in Table 3.2), it is important to have clearly articulated research questions and hypotheses, which can guide the process, but also to understand the key aspects of each variable. These are presented in detail below. It must be noted that, although a separate section is dedicated to each variable, these are often discussed in relation to each other, which, in turn, demonstrates how variables and values are interconnected in an integrated manner.

3.2.1 Type

The first type of corpora that most people associate with translation is the parallel (typically bilingual) corpus, which consists of a set of texts in a source language and their translations in the target language.[2] Parallel corpora focus specifically on the relationship between the source and the target text (Stewart 2000), and this is the reason why they have also been defined as *translation corpora* (Altenberg and Granger 2002; Johansson 1998; Tognini-Bonelli 2001). In this book, the term *parallel corpus* is preferred as it is frequently used by many translation scholars (House 2009; Kenny 2001a; Olohan 2004; Xiao 2010; Zanettin 2012), especially in recent years.

Parallel (both bilingual and multilingual) corpora have been widely used in translation studies to conduct contrastive analyses for a variety of purposes (Bosseaux 2001; Falkum 2007; Fernandes 2009; Kenny 2004; Mason and Şerban 2004; Munday 2002; Wang and Quin 2010). Typically, parallel corpora are used to examine to what extent the translated text is similar or different from its source text, as well as what decisions translators have made in the process of translation. Other applications of parallel corpora include translator training and machine translation. Most studies employing parallel corpora tend to extract data by aligning texts (Véronis 2000) to compare the linguistic features of source texts and their translations. The need for alignment considerably affects the ease with which parallel corpora can be compiled and, for some translation scholars, alignment is a necessary criterion for a corpus to be called parallel (Tognini-Bonelli 2001). Some examples of parallel corpora are the Oslo Multilingual Corpus (OMC) and the Portuguese/English Parallel Corpora (COMPARA).

The second type of corpus used in translation research is the comparable corpus, which can be defined as "a corpus containing components that are collected using the same sampling methods" (McEnery and Hardie 2012, p. 20). Comparable corpora are typically of two kinds: (a) monolingual, consisting of translated and non-translated texts in the same languages and (b) bilingual or multilingual, consisting of texts, either translated or non-translated, in different languages. Comparable monolingual corpora are used to examine features—linguistic, discursive or other—of translated texts and compare them to those found in non-translated texts produced in the same language. Most studies employing comparable monolingual corpora focus on language-independent features of translated texts often referred to as *universals* (Dayrell 2004; Dayrell 2008; Konšalová 2007; Laviosa-Braithwaite 1997; Olohan and Baker 2000). What is unique about such corpora, which are available online, is that their components (i.e. a corpus component of translated texts and a corpus component of non-translated texts) are usually not jointly available on the online platform. Examples include the Translational English Corpus (TEC) and the Corpus of Translated Finish (CTF), which need to be used in conjunction with monolingual corpora such as the British National Corpus (BNC) or the Finnish Literature corpus. Similarly, the ZJU Corpus of Translational Chinese (ZCTC) is the translational counterpart of the Lancaster Corpus of Mandarin Chinese (LCMC).

Thus, TEC, CTF and ZCTC are not comparable corpora in the strict sense of the term, but rather reference corpora (see below) or potential components of comparable corpora.

Comparable bilingual or multilingual corpora can be further divided into two categories: (a) non-translated, consisting of non-translated texts in two or more languages and (b) translated, consisting of translated texts of the same or different source texts in two or more languages. The first category is mainly used for comparative purposes, for example, to compare linguistic or other features in various languages. As a result, it is often used in applied research, for example, in translation training or terminological research (Zanettin 2000). However, it is also possible to use comparable corpora of this type in translation research. For instance, the reciprocal corpus in Zanettin's (2012) typology includes the possibility of a comparable corpus of this type where source texts in Language A and Language B are compared. It is worth noting, however, that such a corpus has a complementary role in translation research and can only be used in conjunction with other types of corpora which include translated texts. The second category has often been employed for the examination of regularities and irregularities of translated texts across languages, especially when the source language is the same (Cartoni et al. 2013; Johansson 2002). To find examples of such corpora we often need to look into parallel multilingual corpora that offer the possibility of comparing target texts, without necessarily consulting the source texts. Such examples are the Dutch Parallel Corpus (DPC), the Online Parallel Bible, the Multitext-East "1984" corpus, and the Oslo Multilingual Corpus (OMC).

The close link of comparable corpora consisting of translated texts to multilingual parallel corpora reflects a significant terminological problem in translation studies, which has been identified and elaborated by Zanettin (2012). As far as the terms are concerned, there is considerable overlap between parallel and comparable corpora and the distinction between them is not always clear-cut. Parallel corpora might not always contain translation in the strict sense of the term, while comparable corpora, as has been demonstrated above, might be derived from parallel corpora. The solution proposed by Fantinuoli and Zanettin (2015) is to treat *parallel* and *comparable* as characteristics of the corpus architecture, and not of the status of the texts in the corpus. Such terminological considerations indicate that parallel and comparable corpora are more closely linked than what literature on corpus-based translation studies seems to suggest and that corpus triangulation is not only desirable but might be inevitable in corpus-based translation research.

Finally, we need to consider reference corpora, which are sometimes used in translation studies. A reference corpus is monolingual, that is captures only one language, and is designed to offer detailed information about that language (Sinclair 1996). Typically, it is large enough to capture different varieties of language, and it might be used for the creation of grammar books and dictionaries. Examples of reference corpora are the British National Corpus (BNC) and Lancaster Corpus of Mandarin Chinese (LCMC) mentioned earlier. Regarding translation, these can be examined on their own, for example, as part of a diachronic study, or constitute a component of a comparable corpus as already explained. The fact that

such corpora can be used independently (i.e. not as corpus components) creates the need for a separate category for these. For example, a study might be interested in investigating the relationship between translation and language change (see Chapters 7 and 8), and it might begin by examining a reference monolingual corpus in the target language to identify areas in which change might be observed. Similarly, a study might focus on the investigation a specific genre. It might begin by examining a reference corpus of translated texts in the target language to establish features that might be worth examining in more detail using a parallel corpus. Depending on the aim of the research project, reference corpora can be large (such as BNC) or much smaller, focusing on a specific time period, genre, author, etc. It is worth stressing again the close link between different corpus types, not only comparable and parallel, but also comparable and reference.

3.2.2 Languages

A corpus can be monolingual, bilingual or multilingual. For reasons of economy, bilingual corpora are sometimes included under multilingual corpora (Altenberg and Granger 2002); however, given the special difficulties involved in creating multilingual corpora, it is worth assigning them a separate value. In general, the creation of bilingual or multilingual corpora is "more time consuming" and "technically complex"(Fantinuoli and Zanettin 2015, p. 3) than that of monolingual corpora. Creating bilingual corpora can be challenging, regarding the availability of appropriate material as these might not be available in a specific language combination or might be difficult to acquire. Once more languages are added to the corpus, these difficulties are multiplied. This is the reason why very few multilingual corpora have been used in descriptive translation studies (Fantinuoli and Zanettin 2015). However, since multilingual corpora allow for data from multiple languages to be captured, the confidence in findings can be increased, especially when these corpora are used to investigate claims about translation behaviour or translated language more generally and not limited to a specific language pair.

As has already been mentioned in the previous section, when we examine corpus type in conjunction with the number of languages, some restrictions apply. In particular, a parallel corpus cannot be monolingual, but we might have both bilingual and multilingual parallel corpora. Regarding comparable corpora, these can take all three values: monolingual, bilingual, or multilingual. Finally, reference corpora are always monolingual. The languages variable can also be assigned an attribute referring to the specific language or languages that make up the corpus, for example, French, Arabic, Chinese. Further restrictions apply here since a monolingual corpus can only take one language attribute, a bilingual corpus two, while a multilingual corpus three or more. It is worth noting that in the case of parallel corpora, the specific languages attribute can imply the directionality of the translation. Thus, if the corpus is described as English–French, the direction is from English into French, while if the corpus is described as French–English, the direction is reversed.[3] Although for most projects the specific language(s) attribute list will include different languages, in some cases, it might involve dialects or regional variations, for example, British English and American English.

3.2.3 Time

The third variable refers to whether the corpus consists of synchronic material, that is, texts produced during the same point in time, which is usually the present, or whether it consists of diachronic material, produced at different points in time. All types of corpora (i.e. parallel, comparable and reference) can be either synchronic or diachronic. The material included in the corpus can be further specified in terms of the specific time span they cover, which is normally the years, or in some cases the centuries, in which the texts have been produced. For a synchronic corpus, only one time span needs to be selected, which can be either an individual year, for example, 2005, or a more extended period, for example, 2005–2010.[4] Although two points in time are the minimum number necessary for a diachronic study to be conducted, this does not mean that more points in time cannot be selected. Depending on the time span of the corpus and its design criteria, some corpora may include three points in time or more. For instance, the Helsinki Corpus includes texts from Old English, Middle English and Modern English, while A Representative Corpus of Historical English Registers (ARCHER) includes, among other genres, science texts from 1800–1849, 1900–1949 and 1950–1999. Although these examples do not involve corpora of translated texts, their design and compilation should easily be extended to translation corpora. For example, a third point in time might be selected in a corpus-based translation project capturing the year(s) when translations started to circulate more widely, or when an event occurred, which might have had an impact on the texts under investigation. It is important to note here that the additional points in time are not selected randomly, but by consulting the research context. Such considerations allow corpus data triangulation to occur in a principled manner.

Apart from a limited number of studies (Amouzadeh and House 2010; Bisiada 2013; Ho-yan 2009; House 2003; House 2006; Malamatidou 2016; McLaughlin 2011), diachronic corpus methods have generally been disregarded in translation studies. Thus, corpus-based translation studies are typically restricted to synchronic analyses of languages, and there seems to be a lack of diachronic corpus-based studies in the field. This conspicuous gap in corpus-based translation studies cannot be easily explained, since the compilation of diachronic corpora does not involve any significant problems, in the same way, that compiling bidirectional or multilingual corpora does. There is a need for more diachronic corpora that would add a temporal dimension to corpus-based translation studies and would address questions such as the relationship between evolving translation styles and norms and evolving language norms and on the relationship between translation and language change (Zanettin 2013).

3.2.4 Texts

The variable of texts, although important, is seldom specifically discussed in corpus-based translation studies, and no clear reference framework is followed, although some tendencies are observed. When referring to translated texts, most studies use terms such as *translations, target texts* or *translated texts*, depending

on the context of the study. Regarding non-translated texts, the terms *source texts* and *original texts* are typically used in parallel studies, although most translation scholars these days prefer the former, since *original texts* might imply that translations are in some way unoriginal textual productions. For comparable studies, the term *texts written originally in Language A* is also frequently used. It is proposed here that the term *non-translated texts* is used for all instances where texts that are not translations are used and that the term *translated texts* is reserved for all cases where translations are used. The context of the research project and the language attribute will help distinguish between a text that is the source text of translation and a non-translated text written in the target language. The only possible exception to this is parallel corpora where the terms *source/target texts* might be employed to signal the relationship between the two sets of texts, and the direction of the translation.

As with other variables, when the texts variable is examined in conjunction with the type of corpus, some restrictions apply. Thus, a parallel corpus needs to consist of both translated and non-translated texts, and for that reason, specifying the text value for parallel corpora is not necessary (although a text attribute is), as it is assumed that they will include by definition both types of texts.[5] On the contrary, the text type value needs to be specified for comparable corpora, as they can consist of either translated or non-translated texts, depending on whether it is a monolingual or bilingual/multilingual comparable corpus (see Section 3.2.1). For instance, a comparable monolingual corpus will normally consist of both translated and non-translated texts. Conversely, bilingual or multilingual comparable corpora will normally consist of either translated texts or non-translated texts in different languages. Similar considerations apply to reference corpora, which can consist of either translated or non-translated texts, although the latter is more common. It is worth mentioning that the variable of text type is the only variable whose values are not mutually exclusive, and thus both translated and non-translated texts can be included in the same corpus. The only exception to this is reference corpora.

The texts variable can also be assigned a genre attribute, which refers to the specific discourse type to which texts in the corpus belong, for example, advertising, fiction, academic texts. Although traditionally translation scholars have been interested in the examination and in-depth analysis of literary texts, in recent years, they have begun to take into consideration more specialised material, such as legal, scientific and technical texts (Biel 2010; Ji 2012; Kranich 2009; Krein-Kühle 2011).

3.3 Issues of corpus design

3.3.1 Size

An important question that needs to be addressed in any corpus project is how big the corpus should be since size is a decisive factor regarding the reliability of the results (Zanettin 2012). The answer to this question, however, is not

straightforward, even for corpus studies that do not employ data triangulation, as it depends on a number of factors (Hunston 2002; McEnery et al. 2006). These factors can be divided into two main categories: the research aim of the study, which includes the types of linguistic analyses to be conducted, and practical considerations, such as availability of data, and time restrictions.

Corpus triangulation projects are affected more by practical considerations than projects where corpora are not combined and are, thus, likely to favour smaller corpora of a few hundred thousand words. Triangulation is a time-consuming process, both regarding corpus building and corpus analysis, and the more corpora that need to be built, the more time is required. Similarly, more time is required for analysing different corpora, compared to a single one. Availability and accessibility of data are also an acute problem when it comes to corpus data triangulation, as the same amount of data might not be available in different languages or for different text types, and if it is, it might not be easily accessible. Finally, comparability is a central concern of corpus data triangulation, and it might significantly affect not only corpus size but even the feasibility of the whole corpus project. As Kenny (1998, p. 53) explains:

> It is in the very nature of translation that new genres are introduced from one literature to another, and there may be nothing 'comparable' in the host literature to a text introduced to it through translation from another textual tradition.

Some readers might consider the smaller corpus size associated with triangulation as a possible limitation. For example, Zanettin argues that small corpora set "a further limit to the claims and generalisations that can be made based on the findings" (2012, p. 44). However, this is not necessarily the case, and smaller corpora present some advantages: they allow for careful sampling and accurate tagging, as well as exhaustive scrutinising (Hundt and Leech 2012). In cases where much noise might be present in the data, a smaller corpus might even be necessary (Hunston 2002). Additionally, the possible limitations of small corpora are counterbalanced by the depth and breadth of understanding that corpus data triangulation offers. In other words, it is preferable to use a number of smaller corpora, which capture different aspects of a phenomenon, than a single one, which, albeit large, captures only one aspect.

3.3.2 Representativeness

While issues of corpus size are associated with the quantity of text, representativeness and balance deal with the quality of the texts in the corpus. Representativeness is closely related to sampling and is notoriously difficult to achieve in a corpus project (Hunston 2002; McEnery and Hardie 2012; Olohan 2004; Renouf 1987; Tognini-Bonelli 2001).[6] In order to address some of the problems, it has been argued that representativeness needs to be understood as an approximation (Kennedy 1998) or a matter of degree (McEnery and Hardie 2012),

"something to strive for rather than something which can reasonably be attained" (Zanettin 2012, p. 45). Thus, even if we cannot achieve full representativeness, it is better to have a corpus that is somewhat representative, than a completely random selection of texts. As a solution for broader linguistic representation, Biber (1993) suggests using smaller samples of a wider range of texts, rather than longer samples of fewer texts. This can be interpreted as an argument favouring triangulation using smaller corpora. Representativeness is more easily achieved with smaller specialised corpora, which typically characterise corpus data triangulation projects, than with large general corpora.

Representativeness is often impeded by considerations of comparability, and creating a representative corpus might often mean that "certain comparisons between the sub-corpora are likely to prove difficult or unreliable" (Kennedy 1998, p. 65). Zanettin (2012) explains that when describing translated and non-translated fiction in one language, the criteria used might be different and they might also differ from the criteria used to describe translated or non-translated fiction in other languages. For instance, Western fiction (defined as fiction set in nineteenth century American West) might only be available as translated fiction outside North America. Although these considerations are relevant to any corpus-based study, it is in studies employing corpus data triangulation that they become most significant. To be able to compare corpora, each corpus needs to be representative of the respective population that it aims to capture, but at the same time allow for comparisons with other corpora to be made. The more corpora are added, the more difficult it becomes to address both these parameters. Thus, in corpus data triangulation projects, a decision often needs to be made between representativeness and comparability and which of the two will be prioritised. Because the aim of most corpus triangulation projects will be the comparison of data from different corpora, in most cases, representativeness will need to give its place to comparability. This does not mean that representativeness is not important in corpus triangulation and that we should not make an attempt to achieve it. Considerations of representativeness are central for corpus data triangulation projects, and it is important to make sure that the different corpora adequately represent a different aspect (e.g. linguistic or stylistic feature) of the same population (e.g. language or genre) or the same aspect of a different population. If we understand representativeness and comparability as matters of degree, as suggested by Leech (2007), we should aim at achieving the highest degree of representativeness, while maintaining the desired degree of comparability, in any project employing corpus data triangulation techniques.

3.3.3 Balance

Balance is difficult to define and is often equated with including into the corpus texts from a wide range of different sources.[7] As Kennedy (1998) rightly points out balance is not achieved by having equal amounts of texts from various sources, but by ensuring diversity, while taking into consideration issues of currency and influentialness when selecting from a wide range of sources. Diversity is

an important parameter, not only because it allows for a range of comparisons to be made across corpus components, but also because it allows for a better representation of the whole population. However, on the one hand, statistics on the circulation of textual productions are not always available, and even if they are, they might not be entirely reliable. On the other hand, a not so widely circulating text might be influential and attract significant cultural capital. For this reason, very few corpus studies have been based on similar statistics. Balance is equally complex and challenging to achieve as representativeness, and, thus, it is better to understand it as a matter of degree (McEnery and Hardie 2012). Although balance is typically more intractable with large general corpora, Kennedy (1998) argues that it is also problematic with smaller specialised corpora, as it is rarely possible to have a corpus consisting of everything published in a specific period.

Considerations of balance present some additional problems for projects relying on corpus data triangulation. As already discussed, statistics on the circulation of texts are not always available, and it is often the case that they might be available for one aspect of the corpus design, for example, language or genre, and not for another. Thus, consistency regarding design is compromised. In the absence of such figures, researchers cannot do much but rely on personal intuition, or, if everything else fails, aim at including equal amounts of texts for each language or genre. Both of these solutions have their limitations, but it is better to strive for some degree of balance than to completely abandon the concept. Corpus data triangulation is also related to issues of balance across corpora. The aim of any corpus triangulation project should be to gain a better understanding of a phenomenon, by analysing it from a range of different perspectives. Diversity, in this case, is important; the more corpora can be built, each focusing on a different aspect of the phenomenon under investigation, the more complete the picture of the phenomenon. Still, time restrictions and other practical considerations will allow for only a limited number of corpora to be built, which means that issues of influentialness are also important to keep in mind.

3.4 Conclusion

By providing corpus-based translation studies with a clear framework for corpus data triangulation, which is based on a comprehensive and flexible corpus typology such as the VVA, the quality and rigour of research conducted in the field can be improved, as clarity and objectivity will increase. This means that not only can we answer existing question in translation studies with more confidence, and increase the reliability of research, but that we can also attempt to address new questions, which have been left unanswered until now. Despite the limitations associated with small corpus size and the problems of representativeness and balance, corpus data triangulation is still a valid endeavour, which allows researchers to compare and contrast different aspects of the same phenomenon. However, we must remember that the results that are reached using a particular corpus are valid only for that corpus, and any generalisations still need to be made with caution. Moreover, since corpus triangulation creates additional difficulties for

corpus configuration, the interpretation of corpus data needs to be made in the light of the potential limitations as a result of corpus design.

Notes

1 An attributes list is much longer and varied compared to the values list. What exactly will be included in the attributes list will depend on the research aims of each project.
2 It must be noted, however, that some parallel corpora do not officially contain translations. For example, corpora consisting of EU texts.
3 For bi-directional corpora, two separate corpora are needed, which will need to be described separately.
4 The time variable might be more complex in studies using material that have been translated many years after their original publication or have numerous reprints.
5 This is the reason that corpora such as TEC need to be considered as corpus components since they do not include non-translated texts.
6 According to Leech, a corpus is representative when its study "can stand proxy for the study of some entire language or variety of a language" (Leech 2007). In other words, when a population is very large, as is often the case with linguistic studies, it is necessary to select a sample that is representative of that population.
7 The factors that will inform diversity will depend on the research aim of the project, and might include the gender, age, nationality, etc. of the writer, the genre, the time and place of publication, the status of texts, their reputation and many more. The larger the population that needs to be presented, the more variables will need to be considered.

References

Altenberg, B. and Granger, S., 2002. Recent Trends in Cross-Linguistic Lexical Studies. In B. Altenberg and S. Granger, eds. *Lexis in Contrast: Corpus-Based Approaches*. Amsterdam: John Benjamins, pp. 3–48.

Amouzadeh, M. and House, J., 2010. Translation as a Language Contact Phenomenon: The Case of English and Persian Passives. *Languages in Contrast*, 10(1), pp. 54–75.

Baker, M., 1995. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target*, 7(2), pp. 223–243.

Biber, D., 1993. Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4), pp. 243–257.

Biel, Ł., 2010. Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential. In *Reconceptualizing LSP: Online Proceedings of the XVII European LSP Symposium 2009*.

Bisiada, M., 2013. Changing Conventions in German Causal Clause Complexes: A Diachronic Corpus Study of Translated and Non-Translated Business Articles. *Languages in Contrast*, 13(1), pp. 1–27.

Bosseaux, C., 2001. A Study of the Translator's Voice and Style in the French Translations of Virginia Woolf's The Waves. *CTIS Occasional Papers*, 1, pp. 55–75.

Bowker, L. and Pearson, J., 2002. *Working With Specialized Language: A Practical Guide to Using Corpora*. London and New York: Routledge.

Cartoni, B., Zufferey, S. and Meyer, T., 2013. Using the Europarl Corpus for Cross-Linguistic Research. *Belgian Journal of Linguistics*, 27, pp. 23–42.

Dayrell, C., 2004. Towards a Corpus-Based Research Methodology for Investigating Lexical Patterning in Translated Texts. *Language Matters*, 35(1), pp. 70–101.

Dayrell, C., 2008. Investigating the Preference of Translators for Recurrent Lexical Patterns: A Corpus-Based Study. *Trans-Kom*, 1(1), pp. 36–57.

Falkum, I.L., 2007. A Relevance-Theoretic Analysis of Concept Narrowing and Broadening in English and Norwegian Original Texts and Translations. *Languages in Contrast*, 7(2), pp. 119–141.

Fantinuoli, C. and Zanettin, F., 2015. Creating and Using Multilingual Corpora in Tranlsation Studies. In C. Fantinuoli and F. Zanettin, eds. *New Directions in Corpus-Based Translation Studies*. Berlin: Language Science Press, pp. 1–11.

Fernandes, L., 2009. A Portal Into the Unknown: Designing, Building, and Processing a Parallel Corpus. *CTIS Occasional Papers*, 4, pp. 16–36.

Ho-yan, C.C., 2009. Third Person Pronouns in Indigenous Chinese Texts and Translated Chinese Texts: The Westernization of Modern Written Chinese. *New Voices in Translation Studies*, 5, pp. 1–15.

House, J., 2003. English as a Lingua Franca: A Threat to Multilingualism? *Journal of Sociolinguistics*, 7(4), pp. 556–578.

House, J., 2006. Covert Translation, Language Contact, Variation and Change. *SYNAPS: A Journal of Professional Communication*, 19, pp. 25–47.

House, J., 2009. *Translation*. Oxford: Oxford University Press.

Hundt, M. and Leech, G., 2012. "Small is Beautiful": On the Value of Standard Reference Corpora for Observing Recent Grammatical Change. In T. Nevalainen and E. Closs Taugott, eds. *The Oxford Handbook of the History of English*. Oxford: Oxford University Press, pp. 175–188.

Hunston, S., 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Ji, M., 2012. Translation and Scientific Terminology. In M. Oakes and M. Ji, eds. *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam and Philadelphia: John Benjamins, pp. 249–274.

Johansson, S., 1998. On the Role of Corpora in Cross-Linguistic Research. In S. Johansson and S. Oksefjell, eds. *Corpora and Cross-Linguistic Research: Theory, Method and Case Studies*. Amsterdam: Rodopi, pp. 3–24.

Johansson, S., 2002. Towards a Multilingual Corpus for Contrastive Analysis and Translation Studies. In L. Borin, ed. *Parallel Corpora, Parallel Worlds: Selected Papers From a Symposium on Parallel and Comparable Corpora at Uppsala University, Sweden, 22–23 April, 1999*. Amsterdam and New York: John Benjamins, pp. 47–59.

Johansson, S., 2003. Contrastive Linguistics and Corpora. In S. Granger, J. Lerot, and S. Petch-Tyson, eds. *Corpus-Based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam and New York: Rodopi, pp. 31–44.

Kennedy, G., 1998. *An Introduction to Corpus Linguistics*. London and New York: Longman.

Kenny, D., 1998. Creatures of Habit? What Translators Usually Do With Words. *Meta: Journal des traducteurs*, 43(4), pp. 515–523.

Kenny, D., 2001a. Corpora. In M. Baker, ed. *Routledge Encyclopedia of Translation Studies*. London and New York: Routledge, pp. 63–68.

Kenny, D., 2001b. *Lexis and Creativity in Translation*. London and New York: Routledge.

Kenny, D., 2004. Parallel Corpora and Translation Studies: Old Questions, New Perspectives? Reporting "that" in Gepcolt: A Case Study. In G. Barnbrook,

P. Danielsson, and M. Mahlberg, eds. *Meaningful Texts: The Extraction of Semantic Information From Monolingual and Multilingual Corpora*. London and New York: Continuum, pp. 154–165.

Konšalová, P., 2007. Explicitation as a Universal in Syntactic De/Condensation. *Across Languages and Cultures*, 8(1), pp. 17–32.

Kranich, S., 2009. Epistemic Modality in English Popular Scientific Texts and Their German Translations. *Trans-Kom*, 2(1), pp. 26–41.

Krein-Kühle, M., 2011. Register Shifts in Scientific and Technical Translation: A Corpus-in-Context Study. *The Translator*, 17(2), pp. 391–413.

Laviosa-Braithwaite, S., 1997. Investigating Simplification in an English Comparable Corpus of News Articles. In K. Klaudy and J. Kohn, eds. *Transference Necesse Est: Proceedings of the Second International Conference on Current Trends in Studies of Translation and Interpreting, 5–7 September 1996, Budapest, Hungary*. Budapest: Scholastica, pp. 531–540.

Laviosa, S., 2002. *Corpus-Based Translation Studies: Theory, Findings, Applications*, Amsterdam and New York: Rodopi.

Leech, G., 2007. New Resources, or Just Better Old Ones? The Holy Grail of Representativeness. In C. Biewer, N. Nesselhauf, and M. Hundt, eds. *Language and Computers: Studies in Practical Linguistics—Volume 59: Corpus Linguistics and the Web*. Amsterdam: Rodopi, pp. 135–149.

Malamatidou, S., 2016. Understanding Translation as a Site of Language Contact: The Potential of the Code-Copying Framework as a Descriptive Mechanism in Translation Studies. *Target*, 28(3), pp. 399–423.

Mason, I., 2001. Translator Behaviour and Language Usage: Some Constraints on Contrastive Studies. *Hermes*, 26, pp. 65–80.

Mason, I. and Şerban, A., 2004. Deixis as an Interactive Feature in Literary Translations From Romanian Into English. *Target*, 15(2), pp. 269–294.

McEnery, T. and Hardie, A., 2012. *Corpus Linguistics*. Cambridge: Cambridge University Press.

McEnery, T., Xiao, R. and Tono, Y., 2006. *Corpus-Based Language Studies*. London and New York: Routledge.

McLaughlin, M., 2011. *Syntactic Borrowing in Contemporary French: A Linguistic Analysis of News Translation*. Oxford: Legenda.

Munday, J., 2002. Translation Studies and Corpus Linguistics: An Interface for Interdisciplinary Co-Operation. *Logos and Language*, 3(2), pp. 11–20.

Olohan, M., 2004. *Introducing Corpora in Translation Studies*. London and New York: Routledge.

Olohan, M. and Baker, M., 2000. Reporting That in Translated English: Evidence for Subconscious Processes of Explicitation? *Across Languages and Cultures*, 1(2), pp. 141–158.

Renouf, A., 1987. Corpus Development. In J. Sinclair, ed. *Looking Up: An Account of the COBUILT Project in Lexical Computing*. Birmingham: Harper Collins, pp. 1–41.

Sinclair, J., 1996. *Preliminary Recommendations on Corpus Typology*. EAGLES Document EAG-TCWG-CTYP/P, May 1996.

Stewart, D., 2000. Poor Relations and Black Sheep in Translation Studies. *Target*, 12(2), pp. 205–228.

Tognini-Bonelli, E., 2001. *Corpus Linguistics at Work*. Amsterdam and Philadelphia: John Benjamins.

Véronis, J., 2000. *Parallel Text Processing: Alignment and Use of Translation Corpora*. Dordrecht: Kluwer Academic Publishers.

Wang, K. and Quin, H., 2010. A Parallel Corpus-Based Study of Translational Chinese. In R. Xiao, ed. *Using Corpora in Contrastive and Translation Studies*. Newcastle: Cambridge Scholars Publishing, pp. 164–182.

Xiao, R. ed., 2010. *Using Corpora in Contrastive and Translation Studies*. Newcastle: Cambridge Scholars Publishing.

Zanettin, F., 2000. Parallel Corpora in Translation Studies: Issues in Corpus Design and Analysis. In M. Olohan, ed. *Intercultural Faultlines: Research Models in Translation Studies I: Textual and Cognitive Aspects*. Manchester: St. Jerome, pp. 105–118.

Zanettin, F., 2012. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome.

Zanettin, F., 2013. Corpus Methods for Descriptive Translation Studies. *Procedia – Social and Behavioral Sciences*, 95, pp. 20–32.

4 Corpus method triangulation

4.0 Introduction

This chapter explains in detail the different types of corpus method triangulation, which can be defined as the use of (two or more) corpus analysis techniques in one study of a single phenomenon. As explained in Chapter 1, a combination of quantitative and qualitative analyses is typical in corpus-based studies, while it is not possible to employ only qualitative methods in a corpus-based study. Thus, more emphasis is placed in this chapter on the combination of different quantitative methods, than on the combination of quantitative and qualitative methods or the combination of qualitative methods. In particular, attention is paid to the use of statistics to scrutinise corpus data, as these have not been extensively used in corpus-based translation studies. The aim here is not to offer an introduction to corpus analysis techniques, for example, explain how statistical analyses can be performed with linguistic data. It is assumed that the reader will be familiar with these concepts, and if not, there are a plethora of books offering excellent introductions to these topics (see for example McEnery et al. 2006; Mellinger and Hanson 2016; Meyer 2002; Oakes 1998). Rather, the focus is on how corpus method triangulation can be achieved using such techniques. Before we proceed, it must be noted that corpus method triangulation is intrinsically linked to the way in which corpus data can be displayed allowing researchers to identify patterns more easily. Different corpus software (e.g. the Sketch Engine, WordSmith Tools, AntConc, AntPConc, ParaConc) offer different tools for accessing and displaying data. The most frequently mentioned in the literature are frequency lists, keyword lists and concordances. The potential of these can be further enhanced by using a lemmatiser or annotating a corpus for parts of speech (POS tagging) or clauses (syntactic parsing). Readers not familiar with these tools are encouraged to consult introductory books on corpus methods (e.g. Hunston 2002; McEnery et al. 2006; Olohan 2004; Zanettin 2012).

4.1 Within-method corpus triangulation

Within-method corpus triangulation involves the combination of methods, which belong to the same or similar paradigm, in the examination of the same

corpus, or set of corpora. Although it is possible to have a corpus-based study that relies solely on quantitative methods, it is not possible to have a corpus-based study that relies only on qualitative approaches. As a result, within-method corpus triangulation is only available with quantitative methods. The examination of corpus data can be performed with a range of statistical analyses, which can be broadly divided into two categories: descriptive, and inferential. Within-method corpus triangulation occurs when one (or more) descriptive statistical analyses are combined with one (or more) inferential statistical analyses. While it is good practice to employ a broad range of different types of quantitative methods since these will help reveal patterns, care must be taken to ensure that statistics are used with a clear rationale, and have an identifiable aim. This is to avoid a situation where measurements are obtained for the sake of measurements and do not add anything to our understanding of the phenomenon under investigation. It is advisable that whenever different quantitative methods are employed, this is accompanied by a short discussion of how each of these informs the analysis, that is what new and crucial data they offer. Equally, the decision of which types of methods will be combined needs to be taken at the outset of the research project, and directly related to the research question or hypothesis.

It has been argued that within-method triangulation is particularly useful in the study of phenomena that involve multidimensional data (Denzin 1989), and translation can be considered as such a phenomenon as it typically involves the analysis of different types of texts in different languages (e.g. translated and non-translated, source and target texts). However, translation studies have not yet fully benefited from this type of triangulation, and accounts of what could be considered as within-method triangulation are very recent and still rare (see also Chapter 1). Although there are several volumes and articles, where different types of statistical tests are presented (Balling and Hvelplund 2015; Oakes 2012), no explicit reference is made as to the possible combination of different quantitative techniques. The only exception to this is Mellinger and Hanson, who argue that to advance translation and interpreting studies "[r]eplication of research studies and triangulation with additional quantitative and qualitative data points are of utmost importance" (2016, p. 241). The present book takes this idea forward and formulates a solid methodological framework, which constitutes a detailed account of how methods belonging to the same paradigm can be combined in corpus-based translation studies.

4.1.1 Descriptive statistics

Descriptive statistics are part and parcel of corpus analysis, and they are central to many built-in features of corpus tools, such as frequency lists and type-token ratios to take the most common examples. According to Mellinger and Hanson (2016), descriptive statistics serve three main purposes. Firstly, they are an effective way of capturing and summarising important characteristics, identifying patterns, and making comparisons across different populations. Secondly, they can help test the assumptions made by statistical tests, which suggests that

a combination of descriptive and inferential statistics is necessary if we want to increase our understanding of the phenomenon under investigation. Finally, they are essential to make generalisations about the broader population based on a sample, although such generalisations should always be made with caution.

The most basic descriptive measure is the frequency count, where, the occurrences of a specific linguistic item, for example, a word, are calculated against the total number of words in a corpus. The resulting number corresponds to the *raw frequency* of the linguistic item (also known as *absolute frequency*). While a simple frequency count can be quite useful, it is often more interesting to examine the relative size or the relationship to the whole, especially when two or more corpora need to be compared. In that case, *normalised frequency* (also known as *relative frequency*) is used, which calculates the occurrence of the linguistic item out of a specific number of words in a corpus, for example, 100, 1,000, or 1,000,000. The basis of the comparison depends on "the typical text length in a corpus" (Biber et al. 1998, p. 264), or other parameters that are pertinent to the corpus under investigation.[1] Raw frequency is generally preferred with smaller samples since normalised frequency might be misleading in this case (Mellinger and Hanson 2016).[2] Two other useful concepts for descriptive statistics are central tendency, which is the typical value that best represents a set of data, and dispersion, which is a measure of how spread out or clustered together data are. There are three common measures of central tendency, namely mean, mode and median, while dispersion can be measured in terms of range, (co)variance, quartiles and standard deviation. See Oakes (1998), McEnery et al. (2006) and Mellinger and Hanson (2016) for more information on how these can be calculated, as well as Gómez (2013) for their suitability to linguistic analyses.

Descriptive measures can be combined, and McEnery and Hardie advise to "report *both* raw and normalised frequencies"(McEnery and Hardie 2012, p. 51, emphasis in the original), since normalised frequencies tend to "abstract from, and simplify, the reality of 'what's there' in the corpus". For instance, Kenny (2001) compares the use of hapax legomena in GEPCOLT by reporting both the raw and the normalised frequencies, though without employing any other quantitative technique. It is also possible to combine frequencies with central tendency or dispersion, but this combination is rarely used in corpus-based translation studies. However, these additional measures can offer new and valuable insight into the corpus data. It is advised here that whenever corpus method triangulation is employed an attempt is made to go beyond frequencies and explore the usefulness of other descriptive measures.

4.1.2 Inferential statistics

No matter how useful descriptive measures might be, they are, as their name suggests, descriptive, and, thus, provide only a limited understanding of the phenomenon under investigation. Inferential statistics go a step further and allow corpus analysts to study variation and test the significance of the differences observed either within or across corpora. Additionally, they can add insight and rigour,

allow for more focused corpus-based analyses of language, and can serve the purpose of triangulation. For example, by using statistical significance tests, "corpus linguists can not only be more confident about the results they obtain but may even gain new insights into the linguistic issues under investigation" (Meyer 2002, pp. 120–121). For this reason, it is highly recommended that inferential statistics be employed in every corpus-based study aiming towards triangulation.

Despite the benefits that inferential statistics offer, some translation scholars seem to believe that statistical tests are not always necessary (see for example Olohan 2004). The reason for their reluctance to employ inferential statistics is not always grounded on arguments about their usefulness. While the science of statistics offers researchers a range of tools to analyse data and identify patterns, many researchers in the humanities, including translation studies, often lack the necessary foundation to employ statistics (Mellinger and Hanson 2016). Corpus-based translation studies, in particular, seem to be lagging behind regarding the use of quantitative analytical methods (Ji and Oakes 2012). As a result, many translation scholars still work mainly with simple descriptive statistics, and translation studies has not yet fully benefited from statistical analyses.

Since it is rarely possible to be sure that an observed difference between corpora, as expressed in terms of normalised frequencies, has not arisen by chance, statistical tests are used to ascertain whether the differences observed in the frequencies are not the result of coincidence. All statistical significance tests are based on the same principle and require a *null hypothesis* (H_0) and an *alternate hypothesis* (H_1). The null hypothesis suggests that there is no difference or relationship between the variables and that any differences that might be observed are due to chance. The alternate hypothesis suggests that there is some difference or relationship between the variables, which can be attributed to some factor other than chance.[3] Statistical testing does not aim to prove the alternate hypothesis, but rather to reject the null hypothesis (Mellinger and Hanson 2016), for which the probability value p needs to be calculated. The closer the probability value p is to 0, the more statistically significant the results. It is generally agreed that a 95 per cent certainty is suitable for linguistic analyses and "a[n] [alternate] hypothesis can be accepted only when the level of significance is less than 0.05 (i.e. $p < 0.05$)" (McEnery et al. 2006, p. 55). If p is smaller than 0.05, the observed difference has most probably not arisen by chance and, thus, the null hypothesis can be refuted. We will not go into further detail about how the p value is calculated, but readers interested in finding out more about statistical significance tests can consult Dunning (1993), Oakes (1998), Baayen (2008), Gries (Gries 2009; 2010) and Gómez (2013) to understand their workings, as well as their limitations.

There are two main categories of statistical significance tests: parametric tests, which assume a normal distribution of data, and non-parametric, which assume that data are distribution-free. Data are normally distributed "if most of the values cluster relatively tight around a mean (average) value" (McEnery and Hardie 2012, p. 51), which is typically not the case with linguistic data that are likely to be distributed more randomly. As a result, linguistic research tends to prefer

non-parametric tests. There are also different types of statistical tests based on the focus of the study. There are tests that examine differences between groups, and tests that examine the relationship between two variables. In corpus-based translation studies, the focus is typically on variables. Examples of non-parametric tests of this type include the Spearman rank correlation coefficient, the chi-square test, and the log-likelihood test (also referred to as the G-test). The former is typically used with continuous data, that is data that can be measured in terms of degree (e.g. calculating the relationship between the quality of a translation and the hours spent working on it). In linguistic research, however, the focus is often on categorical or nominal variables, which are not characterised by degrees, but by a yes/no relationship (e.g. how many nouns in a text) (Gómez 2013), for which the chi-square test and the log-likelihood test are more appropriate than the Spearman's test. The main difference between these two tests is that, while the log-likelihood test can be used with any number of observed examples, the chi-square test tends to be unreliable with small samples.[4] With large samples, the two tests perform almost the same (Oakes 1998), although there seems to be a preference in corpus linguistics for the log-likelihood test, which is also preferred in this book.

Wilson (2013) identifies a problem with the interpretation of statistical significance tests and rightly observes that corpus linguists often treat the *p* value as the actual probability that a difference between corpora is due to chance. For example, if the *p* value is 0.03, they would assume that there is only 0.03 per cent chance that the difference is due to chance. As a result, they will often compare *p* values, assuming that the higher it is, the more likely it is that the difference is accidental. Wilson considers this "an extremely widespread, but normally false belief" (2013, p. 4). To measure the probability that a difference in frequency is due to chance, he advocates the use of the Bayes Factor (BIC), based on Kass and Raftery (1995), who calculated BIC approximations for the log-likelihood test. Positive figures offer evidence against the H_0, while negative figures offer evidence in favour of the H_0. According to Wilson, a figure between 0 and 2 is not worth more than a bare mention, a figure between 2 and 6 is positive evidence, a figure between 6 and 10 is strong evidence, while any figure larger than 10 is very strong evidence against/in favour the H_0. Approximations for the chi-square test can be found in Johnson (2005) and Yuan and Johnson (2008). Thus, it is useful to complement statistical significance tests with the Bayes Factor, as the latter allows corpus researchers to quantify the degree of evidence more rigorously compared to significance levels.

Statistical tests also need to be complemented by effect sizes, which measure the strength of a difference or relationship between variables. This is because it is not sufficient to argue, for instance, that a particular linguistic feature is used differently (i.e. based on normalised frequencies) in translated texts compared to non-translated texts. It is also important to quantify this difference, as a statistically significant result does not mean that the difference is important (i.e. large enough to be meaningful). Lakens argues that an effect size needs to be included for every statistical test since it allows researchers to "communicate the

practical significance of their results (what are the practical consequences of the findings for daily life), instead of only reporting the statistical significance (how likely is the pattern of results observed in an experiment)" (2013, p. 1). Similarly, Gries (2010) argues that "one should always provide an effect size so that one's significant result can be better evaluated", especially if large corpora are employed. This is because even small numerical differences tend to be statistically significant in large corpora. However, the reliability of the effect size is independent of corpus size (Rosenfeld and Penrod 2011).

A range of different metrics of effect sizes is available depending on which statistical test is being used. The most straightforward effect size metric is percentage difference (%DIFF), which indicates the proportion of the difference between the normalised frequencies of specific linguistic features (e.g. word) in two corpora[5] and can be used with any statistical test. The threshold for considering a difference large is relative and depends on the corpora compared. For instance, if normalised frequencies are relatively large, then a 50 per cent difference is considered small, while the same %DIFF is large if normalised frequencies are relatively low. However, %DIFF can only be used when two corpora are examined. When more than two corpora need to be compared, different effect size metrics need to be used based on the statistical significance test employed. For statistical tests on categorical data, as is often the case with linguistic data, effect size can be measured with Cramer's V, if the chi-square test is being used, or with Effect Size for Log Likelihood (ELL), if the log-likelihood test has been used. The effect size will be a figure between 0 and 1, with 0 suggesting no difference or association between the variables, and 1 suggesting strong difference or association between the variables. According to Ellis (2010), a general rule is that an effect size of 0.1 is considered small, 0.3 medium, while 0.5 large.

An additional inferential measure used in corpus-based studies is Mutual Information (MI), which measures how strongly two words are associated with each other and is typically employed in the examination of collocations or other lexical and grammatical associations. Another useful measure when examining corpora is the t-score, which measures the confidence with which we can make claims about associations. According to McEnery et al. (2006), MI scores tend to reveal low-frequency word associations, whereas t-scores tend to reveal high-frequency pairs. Since the two tests capture different aspects of collocations, it is useful to combine them (Church et al. 1994). This type of methodological triangulation will reveal patterns that score highly in both, increasing the reliability of findings. Many corpus analysis applications, such as the Sketch Engine and WordSmith Tools, provide such statistical information when they display collocations.

Within-method corpus triangulation involving inferential statistics can be achieved in a corpus-based study when one (or more) descriptive statistical measures are combined with one (or more) inferential statistical measures. For example, it is possible to use raw frequencies and combine these with the results from a log-likelihood test (one descriptive and one inferential measure). Alternatively, it is possible to have both raw and normalised frequencies and combine these with the results obtained from a chi-square test and Cramer's V (two descriptive and

two inferential). This type of methodological triangulation is necessarily sequential, as descriptive statistics need to be produced before statistical significance tests can be performed. In recent years, translation scholars have started making use of inferential statistics in corpus-based studies (Bernardini 2011; Bisiada 2013; Kruger 2012; Malamatidou 2016; Oakes 2012). In these studies, descriptive statistics, usually raw and normalised frequencies, are used in combination with one statistical significance tests. Very few studies make use of two inferential statistical methods in the same study (e.g. de Sutter et al. 2012; Ji and Oakes 2012). There is a strong need for the use of more inferential statistics in the study of translation, which will be combined both with descriptive and other inferential measures, as well as with qualitative methods.

4.2 Between-method corpus triangulation

Between-method corpus triangulation refers to the combination of quantitative and qualitative corpus analysis techniques in the study of the same corpus or set of corpora. A typical example is the use of descriptive statistics for the examination of the frequency of a specific linguistic feature in translated and non-translated texts followed by a meticulous exploration of the specific—linguistic or other—context in which this feature occurs. Analysing a corpus using quantitative methods helps researchers reach more objective conclusions, and make controlled observations, based on a large number of authentic easily searchable data. Nevertheless, quantitative methods do not take into consideration the context and co-text of the situation where the phenomenon occurs, and they cannot be used to analyse linguistic features in depth or contextualise texts (Kenny 2006). For this, qualitative methods are needed, which offer "descriptive information about the results that cannot be presented quantitatively" (Meyer 2002, p. 124). Qualitative methods rely to a large extent on the subjective interpretation of the corpus analyst, a limitation which is addressed by quantitative analyses. It must also be noted that the size and type of corpus affect the extent to which qualitative analyses can be conducted with it (Evison 2010). Quantitative analyses, while favouring large corpora, can be performed on a corpus of any size, whereas qualitative analyses work better with smaller corpora. Similarly, qualitative analyses are more easily conducted with domain-specific corpora, where the domain is clearly defined, and more information about the context is available. Thus, when corpora of different genres or different languages are examined, between-method corpus triangulation might be harder to achieve, because corpora in these cases tend to belong to different domains, or the domain is conceptualised differently in different languages.

This combination of research methodologies has already been used in corpus linguistics (McCarthy 1998; O'Keeffe 2006), but also in the investigation of a wide range of research questions in corpus-based translation studies, including translation style (Huang 2015; Winters 2013), source language interference (Becher 2009; Liao 2010), and patterning in translated texts (Dayrell 2004; Marco 2013; Nilsson 2002; Williams 2009). These studies have relied on the

examination of a corpus from both a quantitative perspective, relying mostly on frequencies and a qualitative standpoint, relying on close readings of corpus transcripts. However, it must be repeated here that existing corpus-based studies employing a combination of quantitative and qualitative corpus methods do not make any explicit reference to triangulation. Quite often, the combination of methods is done ad hoc, and not in an integrated manner. If between-method triangulation is to be successfully implemented, it needs to be clearly identified from the outset of the project, and its motivation fully explained. In other words, it is necessary that the benefits of each type of analysis be showcased, together with a discussion of how their combination contributes towards a better understanding of the phenomenon under investigation.

A distinguishing characteristic of corpus method triangulation is that one type of methodological triangulation does not exclude the other. On the contrary, corpus method triangulation can combine both within-method and between-method triangulation. Although complex in nature, when corpus method triangulation is achieved on both levels, corpus data can be examined in more depth. The combination of between and within-method triangulation occurs when a qualitative analysis is added to the methods employed to achieve within-method triangulation. For instance, a study might employ descriptive statistics (e.g. normalised frequencies) and inferential statistics (e.g. log-likelihood test) and also analyse data qualitatively. The more methods combined, the better, and corpus method triangulation is most successful when different types of triangulation are used complementarily. However, the suitability for each of the combinations will depend on the research aim of the project and the practical considerations surrounding it. As a rule of thumb, two descriptive measures, together with a test for statistical significance, should be considered the minimum when reporting corpus results.

4.3 A note on display

Before we conclude our discussion of corpus method triangulation, it is important to comment on the different means available for displaying corpus results. Even though these cannot be considered as a method of analysis, nevertheless, tables and graphs are valuable means which help researchers communicate information to the reader. Frequency tables normally include both raw and normalised frequencies, as well as some inferential statistics, if necessary. These tables are an easy way of summarising corpus findings and allow for simple comparisons across categories to be made. Graphs do not provide the same level of detail as tables, but they are more powerful in visualising patterns, trends, and exceptions (Few 2012), as well as relationships between different variables (Mellinger and Hanson 2016).They should not replace the verbal description of the results, but rather complement and support it. The most widely used graphs are bar charts and histograms, while line scatter plots and box plots (or box-and-whisker plots) are also used in corpus-based studies. For a more detailed account of a range of different visualisation tools techniques, see Yau (2011).

The aim of this brief note is to stress the integral part that display methods play in corpus analysis. Their considerable descriptive potential should be harnessed to drive the corpus-based study, irrespective of whether or not triangulation is employed. When corpus data triangulation is employed, tables and graphs can be particularly useful to demonstrate the results of the analysis of different corpora, and identify convergence or divergence in the corpus data. Since corpus data triangulation involves the analysis of various corpora, any way in which results can be summarised and presented helps triangulation achieve its aim of gaining a better understanding of the phenomenon under investigation.

4.4 Conclusion

Although the advantages of employing multiple methods are often recognised in corpus-based translation studies, the emphasis is rarely on the combination of quantitative methods. The model of corpus method triangulation developed in this book stresses the importance of employing not only both qualitative and quantitative methods of corpus analysis, but also combining different quantitative methods, not least inferential statistics. It is rather surprising how a field of study that focuses on examining large quantities of data has managed to disregard inferential statistics for so long. Once achieved, corpus method triangulation, irrespective of whether it is between-method or within-method, can significantly increase the reliability of findings since corpus data have been scrutinised from different perspectives. However, even if reliability is increased, researchers need to be aware of the fact that quantitative methods can be unreliable with too small or too large samples, while qualitative methods tend to rely on subjectivity. While these limitations can be addressed by combining different methods, the results need to be interpreted with caution, and, in particular, care must be taken when attempting to generalise for a larger population. This chapter concludes the theoretical part of the book presenting the corpus triangulation framework. The following chapters constitute the empirical part, where the framework is applied to two case studies, demonstrating its broad applicability.

Notes

1 For example, a study might be examining the use of a specific verb category (e.g. motion verbs), and it might, thus, be more meaningful to compare its frequency to that of all verbs in a corpus, rather than to the total number of words. Ensuring that the right unit is employed allows for meaningful comparisons and facilitates argumentation.
2 For instance, an absolute frequency of 1 out of 2 corresponds to a relative frequency of 50 per cent which also corresponds to an absolute frequency of 100 out of 200. The former is more likely to have arisen by chance than the latter; a fact that is better captured by reporting the absolute frequency.
3 However, even if a result is not statistically significant and, thus, needs to be ignored, "it may be useful as an indication of where to start doing further research" (McEnery and Hardie 2012, p. 51).

4 In cases where low frequencies are reported, it is also possible to use Fisher's exact test, which is very similar to the chi-square test.
5 For diachronic corpora, another useful effect size metric is percentage change, which indicates the proportion of change between the normalised frequencies of specific linguistic features (i.e. word) in two corpora, for example how much more or less frequent a feature has become over the years.

References

Baayen, H., 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Balling, L.W. and Hvelplund, K.T., 2015. Design and Statistics in Quantitative Translation (Process) Research. *Translation Spaces*, 4(1), pp. 170–187.

Becher, V., 2009. The Decline of Damit in English-German Translations: A Diachronic Perspective on Source Language Interference. *Skase Journal of Translation and Interpretation*, 4(1), pp. 2–24.

Bernardini, S., 2011. Monolingual Comparable Corpora and Parallel Corpora in the Search for Features of Translated Language. *SYNAPS: A Journal of Professional Communication*, 26, pp. 2–13.

Biber, D., Conrad, S. and Reppen, R., 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Bisiada, M., 2013. Changing Conventions in German Causal Clause Complexes: A Diachronic Corpus Study of Translated and Non-Translated Business Articles. *Languages in Contrast*, 13(1), pp. 1–27.

Church, K., Hanks, P. and Moon, R., 1994. Lexical Substitutability. In B. Atkins and A. Zampolli, eds. *Computational Approaches to Lexicon*. Oxford: Oxford University Press, pp. 153–177.

Dayrell, C., 2004. Towards a Corpus-Based Research Methodology for Investigating Lexical Patterning in Translated Texts. *Language Matters*, 35(1), pp. 70–101.

Denzin, N.K., 1989. *The Research Act: A Theoretical Introduction to Sociological Methods* 3rd ed. Englewood Cliffs: Prentice Hall.

de Sutter, G., Delaere, I. and 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), pp. 203–224.

Dunning, T., 1993. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19, pp. 61–74.

Ellis, P.D., 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press.

Evison, J., 2010. What Are the Basics of Analysing a Corpus? In A. O'Keeffe and M. McCarthy, eds. *The Routledge Handbook of Corpus Linguistics*. Abingdon and New York: Routledge, pp. 122–135.

Few, S., 2012. *Show Me the Numbers: Designing Tables and Graphs to Enlighten* 2nd ed. Burlingame, CA: Analytics Press.

Gómez, P.C., 2013. *Statistical Methods in Language and Linguistic Research*. Sheffield: Equinox Pub. Limited.

Gries, S., 2009. *Quantitative Corpus Linguistics With R: A Practical Introduction*. Abingdon and New York: Routledge.

Gries, S., 2010. Useful Statistics for Corpus Linguistics. In A. Sánchez and M. Almela, eds. *A Mosaic of Corpus Linguistics: Selected Approaches*. Frankfurt am Main: Peter Lang, pp. 269–291.

Huang, L., 2015. *Style in Translation: A Corpus-Based Perspective*. New York: Springer.

Hunston, S., 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Ji, M. and Oakes, M., 2012. A Corpus Study of Early English Translations of Cao Xueqin's Hongloumeng. In M. Oakes and M. Ji, eds. *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam and Philadelphia: John Benjamins, pp. 177–208.

Johnson, V., 2005. Bayes Factors Based on Test Statistics. *Journal of the Royal Statistical Society*, 67(5), pp. 689–701.

Kass, R. and Raftery, A., 1995. Bayes Factors. *Journal of the American Statistical Association*, 90(430), pp. 773–795.

Kenny, D., 2001. *Lexis and Creativity in Translation*. London and New York: Routledge.

Kenny, D., 2006. Corpus-Based Translation Studies: A Quantitative or Qualitative Development? *Journal of Translation Studies*, 9(1), pp. 43–58.

Kruger, H., 2012. A Corpus-Based Study of the Mediation Effect in Translated and Edited Language. *Target*, 24(2), pp. 355–388.

Lakens, D., 2013. Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, p. 863.

Liao, M.-H., 2010. Influence of Translations on Non-Translations: Popular Science as a New Genre in Taiwan. *Perspectives: Studies in Translatology*, 18(2), pp. 127–140.

Malamatidou, S., 2016. Understanding Translation as a Site of Language Contact: The Potential of the Code-Copying Framework as a Descriptive Mechanism in Translation Studies. *Target*, 28(3), pp. 399–423.

Marco, J., 2013. Tracing Marked Collocation in Translated and Non-Translated Literary Language. In C. Way et al., eds. *Tracks and Treks in Translation Studies: Selected Papers From the EST Congress, Leuven 2010*. Amsterdam and Philadelphia: John Benjamins, pp. 167–188.

McCarthy, M., 1998. *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.

McEnery, T. and Hardie, A., 2012. *Corpus Linguistics*. Cambridge: Cambridge University Press.

McEnery, T., Xiao, R. and Tono, Y., 2006. *Corpus-Based Language Studies*. London and New York: Routledge.

Mellinger, C.D. and Hanson, T.A., 2016. *Quantitative Research Methods in Translation and Interpreting Studies*. London and New York: Routledge.

Meyer, C., 2002. *English Corpus Linguistics*. Cambridge: Cambridge University Press.

Nilsson, P.-O., 2002. An Assessment of the Translation-Specificity of Over-Represented Multi-Word Patterns in Swedish Fiction Texts Translated From English. *Linguistica Antverpiensia*, 1, pp. 407–417.

Oakes, M., 1998. *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Oakes, M., 2012. Describing a Translational Corpus. In M. Oakes and M. Ji, eds. *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam and Philadelphia: John Benjamins, pp. 115–148.

O'Keeffe, A., 2006. *Investigating Media Discourse*. London: Routledge.
Olohan, M., 2004. *Introducing Corpora in Translation Studies*. London and New York: Routledge.
Rosenfeld, B. and Penrod, S., 2011. *Research Methods in Forensic Psychology*. Hoboken: Wiley.
Williams, I.A., 2009. A Corpus-Based Study of Spanish Translations of the Verb "report" in Biomedical Research Articles. *Meta: Journal des traducteurs*, 54(1), p. 146.
Wilson, A., 2013. Embracing Bayes Factors for Key Item Analysis in Corpus Linguistics. In M. Bieswanger and Ko.-S. Amei, eds. *New Approaches to the Study of Linguistic Variability: Language Competence and Language Awareness in Europe*. Frankfurt: Peter Lang, pp. 3–11.
Winters, M., 2013. German Modal Particles: From Lice in the Fur of Our Language to Manifestations of Translators' Styles. *Perspectives*, 21(3), pp. 427–445.
Yau, N., 2011. *Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics*. Indianapolis: Wiley Publishing.
Yuan, Y. and Johnson, V., 2008. Bayesian Hypothesis Tests Using Nonparametric Statistics. *Statistica Sinice*, 18, pp. 1185–1200.
Zanettin, F., 2012. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome.

Part II
Empirical applications

5 The language of English–Russian translation

Connectives

5.0 Introduction

This chapter will, together with Chapter 6, apply the triangulation framework described earlier in the book to the study of the language of translation. More specifically, the first case study presented in this book focuses on the pragmatic factors affecting the use of connectives in translated texts. A corpus design which facilitates data triangulation based on the VVA typology allows for the examination of a number of different factors that might affect the distribution of specific linguistic features. At the same time, relying on methodological triangulation to carry out the analysis allows to establish which of these factors are most strongly affecting the use of these features. The chapter begins by outlining the rationale for the case study and explaining why this specific linguistic feature has been chosen. It then presents some general properties of connectives and focuses particularly on their taxonomy. To provide a focused discussion and to facilitate analysis, this case study explores only one type of connectives, namely adversative connectives. The chapter continues by discussing properties (i.e. categories, problems of identification and frequency) of adversative connectives in English and Russian, examining previous studies that have investigated the use of connectives in these languages. The aim is to reveal some of the most significant factors that affect the use of connectives in both languages, which inform the research questions that the case study will address. The factors studied are the genre to which a text belongs (i.e. fiction, children's fiction and non-fiction), the audience the text addresses (adults vs. children), the existing linguistic conventions in the target language, and influence from the source text.

5.1 Rationale

Connectives are considered to be optional linguistic units since their presence in a text is not a necessary condition for the establishment of a link between two word groups, clauses or sentences. Frequently, this relationship already exists, and cohesion is achieved by "the making explicit of the way that the meaning of one piece of the text relates to the meaning of another"(Christiansen 2011, p. 163). As a result, both syndetic (overtly marked) and asyndetic (unmarked)

constructions are possible. Consider the following asyndetic construction: *Mark didn't go to work today. He is ill.* In this case, there is a causal relationship between the two clauses, which can be made explicit through the use of a causal connective, such as *because*. This optionality makes connectives a good candidate for cross-linguistic examination, especially when mediation is involved in the form of translation. It has been observed that connectives tend to be regularly added or omitted by translators, and, unlike other linguistic features, this addition/omission is easy to spot (Becher 2011b). Additionally, connectives lend themselves relatively easily to cross-linguistic analysis, since, at least as far as Indo-European languages are concerned, there are similarities in the elements used to create textually important links (Rudolph 1988), but also significant differences in the way different languages use connectives (Halverson 2004). As a result, these linguistic features are interesting objects of investigation for both practising translators and translation scholars. Even though no study has ever focused on the examination of the differences in the use of connectives in English and Russian, it is likely that some marked differences exist, given that English and Russian belong to different linguistic families, and there are significant typological differences between them (e.g. morphology, sentence length, word order).

The importance of connectives is further evidenced by the number of linguistic studies on their use in different languages (Degand and Fagard 2012; Olmos and Ahern 2009; Pit 2007; Stukker and Sanders 2012; Zufferey 2012) and the considerable space that grammar books dedicate to the use of connectives, at least in English (Biber et al. 1999; Huddleston and Pullum 2002; Quirk et al. 1985). Despite connectives being a rich and significant field of investigation, translation studies has touched only very lightly on this topic. Existing studies either focus on language acquisition cross-linguistically (Granger and Tyson 1996; Lamiroy 1994; Rudolph 1996) or analyse connectives to investigate more general translation phenomena, such as explicitation (Becher 2011a; Denturck 2012; Puurtinen 2004). Typically, these studies use a reciprocal corpus consisting of both translated and non-translated texts in two languages belonging to the same genre.[1] Even though reciprocal corpora are examples of corpus data triangulation, many more combinations of corpora are possible, which are often ignored in translation studies (see Chapter 1). Without a clear framework, which would allow researchers to approach this and similar topics from new perspectives, the already limited research in the field is bound to remain constrained to the use of reciprocal corpora, which offer only a basic understanding of the phenomenon under investigation.

The present case study aims to address this gap and increase our knowledge of the cross-linguistic differences in the use of connectives, but also of the factors that influence their distribution in mediated discourse. This, in turn, will reveal some of the factors that might influence translation decisions and have an impact on the nature of translated texts. As already mentioned, connectives are often investigated in relation to explicitation in translation. However, we will not make any attempt here to interpret the data using any of the explicitation hypotheses available,[2] as it is a controversial topic in translation, which has generated a heated

debate, with some studies pointing towards its existence (Chung-ling 2008; Olohan and Baker 2000; Øverås 1998; Pápai 2004), and others disproving it (Baumgarten and Özçetin 2008; Becher 2010; Hansen-Schirra et al. 2007; Puurtinen 2004). The aim of this case study is not to propose a new theory of explicitation or try and validate existing ones—this would be a too ambitious goal given the nature of this book and the space available. Rather, the aim is to gain a more comprehensive understanding of the use of connectives in translation. Focusing on this linguistic feature, which deserves to be examined in more detail, without necessarily connecting it to wider translation phenomena, can increase our understanding of both the nature of translated texts and that of the translation activity more generally. This is not to say that corpus triangulation is not suitable for the examination of possible explicitation phenomena. On the contrary, it is perhaps with the help of such a comprehensive methodological framework that we can reach a better understanding of this and similar controversial topics. This is even more so if we consider that studies claiming to validate the Explicitation Hypothesis (Blum-Kulka 1986) have been strongly criticised by Becher (2011a) as biased by methodological problems.

5.2 General properties of connectives

Connectives perform an important cohesive and connective function (Halliday and Hasan 1976) by marking connections between units of discourse and, as a result, have been extensively studied. This has led to a large amount of variation in terminology, and terms such as *connective adjuncts* (Huddleston and Pullum 2002), *connectives* (Finch 2000), *linking adjuncts* (Carter and McCarthy 2006), and *logical connectors* (Celce-Murcia et al. 1999) have been used in English. In this case study, the term *connectives* is preferred as it is the most transparent and highlights the function of these linguistic elements, which is to mark a link between phrases, clauses or sentences. Connectives refer to all linguistic items that can mark links in a text. For example, in English, these include conjunctions (e.g. *but, although*), linking adverbials (e.g. *however, then*), prepositional phrases (e.g. *for that reason, in this case*), and fixed expressions (e.g. *that is to say*, *having said this*). Connectives have a pragmatic function, which, according to Gotti (2003, p. 107) "clarifies the purpose of the sentence that follows". In other words, the writer/speaker can control the way in which the reader/listener will interpret the discourse. The fact that the speaker/writer has some control over the discourse explains why connectives are considered optional elements and why translators tend to add or delete them in translation. As mentioned earlier, a pragmatic parameter which has been found to affect the use of connectives is that of language (Halverson 2004), and connectives are considered language-specific elements (Christiansen 2011). Also, it has been argued that connectives are used differently in different genres, with certain specialised discourses, like legal texts (Gotti 2003) favouring the use of specific connectives to avoid ambiguity.

Syntactically, connectives can be divided into coordinating and subordinating. Coordinating connectives connect elements that have the same syntactic role, as

in the example *He is arrogant and doesn't care about his friends.* Here, the two clauses can stand independently and even appear in reverse order. This is in contrast to subordinating connectives, where one element depends on the other, and their order cannot be reversed, as in the example. *He is arrogant but still cares about his friends.* Semantically, different scholars divide connectives into different categories. For example, Halliday and Hasan (1976) divide connectives into five categories: additive, adversative, causal, temporal and continuatives. Celce-Murcia et al. (1999) identify four categories by combining temporal and continuative connectives to form the sequential category, while Biber et al. (1999) identify six categories: enumeration and addition, summation, apposition, result/inference, contrast/concession, and transition. In this case study, Halliday and Hasan's original model is employed, due to its general economy and comprehensiveness. According to this taxonomy, additive connectives (e.g. *and, furthermore, similarly*) are relatively neutral and express a general addition to what has already been said. Adversative connectives (e.g. *but, in fact, instead*) signal that the new information will contradict the information already available, while causal connectives (e.g. *so, because, in this respect*) signal that what is being added should be interpreted as a consequence of what has already been said. Finally, temporal connectives (e.g. *then, soon, finally*) signal that elements are connected in time. Continuatives cannot be clearly assigned to any of these categories, but nevertheless, have a cohesive function. This category consists of miscellaneous items, such as *now* and *of course*, which are relatively common in spoken discourse.

To facilitate analysis, this case study focuses only on connectives that can be broadly mapped into the adversative category. These have been chosen over other connectives because adversative relations are considered, together with additive relations, more prototypical than causal or temporal (Christiansen 2011). Adversative connectives also lend themselves more easily to corpus analysis, since they are less vague than other connectives (Christiansen 2011). Halliday and Hasan (1976) divide adversative connectives into adversative proper, dismissal, contrastive and corrective. However, certain adversative connectives can belong to two or more categories simultaneously. For example, *nevertheless* can be either adversative proper or dismissal, depending on the linguistic and extra-linguistic context.[3] The importance of the extra-linguistic context makes it difficult to identify the correct meaning, even if the limited linguistic context of concordance lines is examined. For this reason, this case study focuses on the category of adversative connectives more generally, without making finer distinctions regarding the elements that belong to this category. In the following two sections, we will look at the specific properties of adversative connectives in English and Russian respectively.

5.3 Adversative connectives in English

Due to their central role in discourse cohesion, connectives have received considerable attention in English monolingual research, with the majority of studies

focusing on the distribution of connectives in different registers (Lei 2012; Zareva 2011) or time periods (Dorgeloh 2004; König 1985). While such studies examine a broad range of connectives, the discussion here will focus only on adversative connectives.

5.3.1 Categories

Adversative connectives in English consist of conjunctions, linking adverbials, prepositional phrases and fixed expressions. Conjunctions are tightly linked to linking adverbials but are more limited in syntactic position. Linking adverbials can "make semantic connections between spans of discourse of varying length" (Biber et al. 1999, p. 558), whereas conjunctions can only make semantic connections at or below clause level (Liu 2008).

The only coordinating adversative connective in English is the conjunction *but*, while there are many more subordinating adversative connectives. However, existing grammar books and studies on connectives in English do not provide a comprehensive list of these. Typically, some representative examples are provided, but overall different studies focus on different elements, with some focusing mostly on linking adverbials (Peacock 2010), while others are analysing conjunctions, linking adverbials and prepositional phrases (Liu 2008). Table 5.1 offers a summary of subordinating adversative connectives in English, created by combining the material available in some of the most prominent grammar books, as well as in relevant research articles in the field, together with information available in major monolingual dictionaries.

Table 5.1 Subordinating adversative connectives in English

Conjunctions	*Linking adverbials*	*Prepositional phrases*	*Fixed expressions*
albeit	actually	after all	all the same
although	admittedly	as a matter of fact	either way
only	anyhow	as against that	even if
though	anyway	at any rate	even so
whereas	despite	at the same time	even though
while	conversely	in any case/event	that may be
	however	in either case/event	then again
	nevertheless	in fact	to tell the truth
	nonetheless	in point of fact	
	notwithstanding	in reality	
	rather	in spite of	
	still	in/by comparison	
	yet	in/by contrast	
		of course	
		on the contrary	
		on the other hand	

5.3.2 Identification

While in the majority of cases connectives serve only an adversative function, some of them can have additional functions. Consider the following sentences:

(a) *I didn't want to go,* ***but*** *I did.*
(b) *He's anything* ***but*** *nice.*
(c) *It is* ***but*** *a piece of paper.*
(d) *No more* ***buts*** *and ifs.*

But acts as a coordinating adversative conjunction in sentence (a), whereas it is a preposition expressing exclusion in sentence (b), similar to *apart from*. In sentence (c), it acts as an adverb, while it is used as a noun in sentence (d).

Similarly, some of the conjunctions and linking adverbials might have additional connective functions, apart from adversative. For instance, w*hile* is used as an adversative connective in sentence (e) and as a temporal in sentence (f).

(e) ***While*** *it started out as a threat, it's since transformed into a promise.*
(f) *I waited* ***while*** *he was getting ready.*

Additionally, some linking adverbials might act as adverbs without a linking function, such as *still*, which can be used as an adversative connective (g) or as an adverb (h).

(g) ***Still,*** *it wasn't all sweet talk.*
(h) *There's* ***still*** *time.*

In order to exclude instances of non-adversative use, corpus data is manually refined, and only instances with a clear adversative function are included in the analysis.

5.3.3 Distribution

Given the range of different approaches to the study of connectives, it is difficult to make any generalisation or comparison of their frequency of use. However, some conclusions can be drawn on the use of some categories of connectives based on corpus findings. In the corpus examined by Biber et al. (1999), which consists of conversation, fiction, academic prose and news, the conjunction *but* was found to be used with a frequency of approximately 5,000 pmw (per million words), while adversative linking adverbials occurred with a frequency of approximately 1,000 pmw. Liu (2008), who analysed five genres in the British National Corpus (BNC), reports a frequency of approximately 2,000 pmw for adversative linking adverbials, which is higher to what Biber et al. (1999) have found in their corpus. This discrepancy can be explained by the fact that Biber et al. analysed only the three most frequent adversative linking adverbials, namely, *however, nevertheless, yet*, while Liu conducted a much more thorough analysis of 30 items.

Examining different genres in more detail, the conjunction *but* has been reported to be more frequent in spoken discourse (7,000 pmw) and fiction (5,000 pmw), and least frequent in academic prose (2,000 pmw), while subordinating connectives such as *though*, *in contrast*, *however* are used with approximately the same frequency in spoken discourse, fiction and academic prose (approximately 2,000 pmw), and much less frequently in news (approximately 1,000 pmw) (Biber et al. 1999). In the BNC, Liu (2008) has found that adversative linking adverbials are most frequent in academic texts (3,000 pmw), and least frequent in news (1,500 pmw), with fiction and conversation somewhere in between (2,000 pmw). Business textbooks have also been found to make relatively high use of adversative linking adverbials (1,400 pmw), while humanities textbooks use these more rarely (800 pmw). Engineering, natural sciences and social sciences textbooks stand in the middle with approximately 1,000 pmw (Biber 2006). Finally, Peacock (2010) reports a higher use of adversative linking adverbials (3,000 pmw) in research articles belonging to the humanities (from the fields of economics, linguistics, management and psychology), compared to those belonging to the hard sciences (chemistry, computer science, material science and neuroscience) (2,500 pmw). From the discussion so far, it is clear that adversative connectives are a frequent linguistic feature, a fact that underlines their important cohesive function in English, but it is in more specialised discourse, such as academic writing, that they seem to play a more significant role as cohesive devices. The use of connectives appears to be tightly linked with the communicative function of each genre, but also with the audience it addresses (e.g. compare humanities textbooks and humanities research articles).

5.4 Adversative connectives in Russian

As in English, connectives are instrumental in providing cohesion in both spoken and written discourse in Russian (Simmons 1981). According to Shvedova et al. (1980), they are auxiliary linguistic items, which help establish links between phrases, clauses and sentences, and their role is to make these links explicit. However, studies examining Russian connectives differ from respective studies of English connectives both in volume and focus. Most studies on Russian connectives focus on the specific uses of certain conjunctions, their particularities and distribution (Jasinskaja and Zeevat 2008; Kapatsinski 2009; Uryson 2000), rather than on capturing a more holistic view of the phenomenon. In that respect, it can be said that studies on Russian connectives are interested in their micro properties, rather than their macro properties. Nevertheless, it is possible to make some useful generalisations, which allow Russian and English connectives to be meaningfully compared.

5.4.1 Categories

Adversative connectives in Russian are a much more controversial topic compared to English connectives, and there is still little agreement as to how exactly

different linguistic categories might perform a connective function, what distinguishes conjunctions from other words or phrases with a connective function and how the various categories are to be labelled (Beloshapkova 1977; Ledenev 1988; Orlov and Cheremisina 1980; Priyatkina 1977; Rogozhnikova 1983; Serebryanaya 1970). Studies report that the category of connectives in Russian is populated by conjunctions, particles, adverbs, modals, adverb-modal hybrids and fixed expressions, with some linguistic items, like *a*, being reported as both conjunctions and particles, blending into a category of conjunction-particles. While the nature of adversative connectives in Russian is worth examining in detail, the aim, as well as the available space, of this chapter, do not allow for a more in-depth investigation. Since a distinction between the different linguistic categories serving a connective function is outside the scope of this case study, no distinction is made between these, and the generic term *connectives* is used.

The lack of extensive studies on the general properties of Russian connectives creates a significant gap in their categorisation. Contrary to English where existing taxonomies are perhaps one too many, in Russian there is no clear taxonomy of adversative connectives. Thus, it is not possible to argue whether Russian shares the same subcategories of adversative proper, dismissal, contrastive and corrective connectives. Adversative connectives in Russian are discussed as a single group, and any distinction between them can only be made based on their meaning (i.e. dictionary definition), which tends to be unreliable, as it does not take into account the range of different factors affecting their use. Another problem is that, even if such a categorisation existed, it would most likely not map directly onto the one available in English. For this reason, in addition to the reasons explained in Section 5.2, the focus in this case study is on the category of adversative connectives more generally.

As in English, adversative connectives in Russian can be divided into coordinating and subordinating. Regarding coordination, Russian makes use of four connectives, namely *a, da, no* and *odnako*. Thus, many more options are available in Russian for coordination compared to English, where the only option is *but*. Regarding subordination, Russian makes use of the linguistic items listed in Table 5.2.[4] If we compare the number of subordinating connectives available in Russian to that available in English, it is clear that English makes use of a larger number of connectives belonging to this category compared to Russian (i.e. 43 in English and 34 in Russian). It is interesting to examine whether the greater number of coordinating connectives in Russian and the larger number of subordinating conjunctions in English means that coordination is employed more frequently in Russian, while subordination more frequently in English (see Chapter 6).

It might be assumed by looking at the table that many synonymous connectives are available in Russian, particularly with the meaning of *but*. In reality, there are subtle differences in meaning and use between these connectives, and the choice of adversative connectives in Russian is affected by "semantic, syntactic, and stylistic factors" (Kapatsinski 2009, p. 170). For example, while *but* in English serves a number of uses, namely (a) denial of expectation (Lakoff 1971), (b) argumentative (Anscombre and Ducrot 1977), and (c) semantic opposition (Lakoff 1971) or formal contrast (Malchukov 2004), in Russian, the first two

Table 5.2 Coordinating and subordinating adversative connectives in Russian

Coordination			*Subordination*		
Russian	*Transliteration*	*English equivalent*	*Russian*	*Transliteration*	*English equivalent*
а	*a*	but	в действительности	v deystvitel'nosti	in fact
да	*da*	but	в конце концов	v kontse kontsov	after all
но	*no*	but	в отличие от этого	v otlichiye ot etogo	in contrast
однако	*odnako*	but	вернее	verneye	rather
			вместо этого	vmesto etogo	instead
			во всяком/любом случае	vo vsyakom/lyubom sluchaye	in any case
			вопреки	vopreki	despite
			впрочем	vprochem	however
			все ж таки	vse zh taki	all the same
			все же	vse zhe	yet
			все таки	vse-taki	nevertheless
			даже если/если даже	dazhe yesli/yesli dazhe	even if
			даром что	darom chto	even though
			для сравнения	dlya sravneniya	in comparison
			за всем тем	za vsyem tem	notwithstanding
			зато	zato	but
			на самом деле	na samom dele	actually
			наоборот	naoborot	conversely
			напротив	naprotiv	on the contrary
			невзирая на	nevziraya na	regardless
			несмотря на	nesmotrya na	despite
			при всем том	pri vsem tom	in spite of
			при том	pri tom	besides
			пускай	puskay	even though/even if
			пусть	pust'	even though/even if

(*Continued*)

Table 5.2 (Continued)

Coordination			*Subordination*		
Russian	*Transliteration*	*English equivalent*	*Russian*	*Transliteration*	*English equivalent*
			с другой стороны	s drugoystorony	on the other hand
			скорее	skoreye	rather
			так или иначе	tak ili inache	anyway/anyhow
			так или этак	tak ili etak	anyway/anyhow
			тем не менее	tyem ne meneye	nevertheless
			только	tol'ko	only
			точнее	tochneye	rather
			хоть	khot'	although/though
			хотя	khotya	although/though

functions can be served by *no*, whereas the third only by *a* (Jasinskaja and Zeevat 2008). The aim of the present case study is not to identify translation equivalents for each of the connectives, but rather to gain an understanding of their distribution, as well as the factors affecting it. For this reason, the English equivalents provided serve only as a guide regarding meaning.

5.4.2 Identification

When it comes to identifying adversative connectives, the most acute problem is that presented by *a* (Jasinskaja and Zeevat 2008; Jasinskaja and Zeevat 2009; Uryson 2000), which is a polysemous connective sometimes having the meaning of *but*, as in sentence (i), and sometimes that of *and*, as in sentence (j). Thus, *a* can be used as both an additive and an adversative connective. Similarly, *da* is polysemous and can be used as both an adversative and an additive connective (Kapatsinski 2009), and it can also be used as an interjection meaning *yes*.

(i) Сверху вроде корочка, **а** внутри мягкое.
(It seems to have a crust on the top, **but** inside it's soft.)
(j) Никто не в курсе, **а** кто в курсе, тот не скажет.
(Nobody knows, **and** those who do will not say it.)

Apart from connectives with multiple meanings, some linguistic items can be treated as both connectives and particles. For instance, *khotya* can be used as an adversative connective (k) or as a particle meaning *at least* (l).

(k) Дедушка тогда уже на бабушке был женат, **хотя** детей еще не было.
(Grandpa was then already married to grandma, **even though** they didn't have children yet.)
(l) Съешьте **хотя** кусочек.
(Have **at least** a little piece.)

Very similar to this is *khot'*, which in addition to the above can also be used as a particle meaning *for example* (m).

(m) Возьмите **хоть** эту страну.
(Take **for example** this country.)

Finally, *pust'* can be used both as an adversative connective, as in sentence (n), and as a particle forming the imperative similar to the English *let*, as in sentence (o).

(n) **Пусть** мир несовершенен и ты не в состоянии его исправить, но ты можешьотвечать сам за себя.
(**Even though** the world isn't perfect and you are not in a position to make it right, but you can answer for yourself.)
(o) **Пусть** придет к шести.
(**Let** him come at six.)

To distinguish when a linguistic item is used as an adversative connective and when not, corpus data are manually refined, and the immediate linguistic context is consulted. Only instances where it is clear that the linguistic item under question serves an adversative connective function are included in the analysis.

5.4.3 Distribution

Detailed corpus-based studies, which would offer quantitative data on the distribution of connectives both in the language in general, as well as in specific genres, are not available in Russian. However, there are some studies available on the distribution of individual connectives, which might serve as guides for comparisons. Regarding coordination, the most common connective is claimed to be *no*, followed by *odnako* and *da* (Kapatsinski 2009). *Odnako* seems to be preferred in written discourse and often associated with formality (Lekant et al. 1982). Conversely, *da* tends to be related to colloquial discourse (Krilova 1980; Kruchinina 1988; Shvedova et al. 1980), but even in spoken discourse it is "highly infrequent" (Kapatsinski 2009, p. 168), and *no* is preferred.

Regarding subordination, *khotya* is considered to be the most frequently used connective (Chernikova 2007), with the second place occupied by *nesmotrya na*. In terms of genres, *khotya* has been found to be particularly widespread, whereas *nesmotrya na* is preferred in academic prose and news, that is, genres where it is necessary to explicitly mark the relationship between clauses (Chernikova 2007). Other connectives that have been reported to be more prevalent in specialised written discourse are *nevziraya na*, as well as *vopreki*. On the other hand, the connectives *khot', darom chto, pust'* and *puskay* are associated with spoken discourse (Chernikova 2007), and are, thus, rather infrequent in written discourse. As a result, they are also more common in fiction, especially in dialogue. It must be noted that since detailed quantitative data on the frequency of each connective are not available, it is not possible to quantify how much more or less frequent a certain connective is compared to others.

To conclude, direct comparisons with English cannot be made, at least regarding frequency, since quantitative data on Russian connectives, as well as cross-linguistic studies between Russian and English connectives, are not available. However, there is sufficient evidence to suggest that, as in English, the use of adversative connectives in Russian depends on the communicative function of each genre, as well as on its audience. In that respect, there is no reason to assume that the distribution of connectives in Russian is homogenous across genres. Rather, it is expected that certain genres will make higher use of adversative connectives in general, as well as have a strong preference for specific connectives. At the same time, while the communicative function and audience of specific genres is expected to be similar across languages, different languages are expected to exhibit different preferences in their use of connectives, following observations made by Christiansen (2011). It is these differences that this case study aims to investigate, not only in monolingual discourse, but also in translation.

5.5 Factors affecting the use of adversative connectives

As already mentioned, previous studies examining connectives in translation employ reciprocal corpora, mainly focusing on the language direction as a possible factor affecting the distribution of connectives. Using reciprocal corpora, it is also possible to examine the influence of the source text and that of the existing conventions in the target language. However, studies employing reciprocal corpora seem to assume that the two languages under investigation will differ in their use of connectives. While this might be true in the majority of cases, it is preferable to begin by first addressing the question of whether corpus evidence suggests that there is a difference between the two languages. If the answer to this question is positive, it is then reasonable, for example, to examine the possible influence of the source texts. If the answer is negative, however, different factors might need to be considered. In other words, examining first monolingual production in two (or more) languages helps delimit which factors need to be further investigated. This integrated approach, where one stage of analysis leads to the next, can only be achieved if a data triangulation technique is employed, suggesting that the examination of this and similar linguistic phenomena can significantly benefit from triangulation.

An additional factor that can be examined in relation to the distribution of connectives is the genre to which a text belongs. The discussion of connectives in English and Russian demonstrates that, as far as monolingual production is concerned, there is a stark contrast between fiction and academic prose not only in the overall frequency of connectives, but also in the distribution of specific adversative connectives. It is worth investigating whether a similar contrast can be found in translation. However, when examining the specific context of English–Russian translation, it becomes clear that the availability of translated academic prose is rather limited, and not readily available. By contrast, non-fiction (mostly in the form of current affairs, politics, history, and popular science), which bears some similarities to academic prose, is frequently translated from English into Russian and circulates widely in Russia. Focusing on fiction and non-fiction, instead of academic prose, offers an additional advantage. Although the topics presented in fiction and non-fiction are very different, as well as the style of writing, and other aspects of the texts, the audience is quite similar, that is, the general public. This allows for a clear examination of the extent to which the subject matter, but not necessarily the audience, might affect linguistic conventions in translated texts. At the same time, while existing monolingual research has examined in some detail written discourse addressed to adults, there is no indication as to how the use of adversative connectives might be affected when the recipients of the discourse are children. Existing research in children's fiction demonstrates that it exhibits significant differences from fiction for adults both regarding language and narrative elements (e.g. plot, characters) (McDowell 1973; Nikolajeva 2002; Sunderland 2012), and it can be expected that the use of adversative connectives might also differ.

Thus, the three genres identified here (fiction, children's fiction and non-fiction) allow for a rigorous examination of the extent to which the subject matter and

the audience affect the use of adversative connectives in translation. Fiction addresses a general adult audience, favours narration and description (Rimmon-Kenan 2002), and is highly creative (Hague 2003). Children's fiction addresses primarily children, and although it shares some similarities with fiction for adults, the language employed is child oriented, and there is a much more extensive use of creativity and fantasy, while dialogue tends to be favoured (McDowell 1973). Non-fiction addresses an adult audience wanting to find out more about a specialised subject. It favours argumentation and uses a more impersonal style compared to fiction (Buss and Karnowski 2002). Thus, fiction and children's fiction address similar subject matters, but different audiences, while fiction and non-fiction deal with different subject matters, but address a similar audience.

Based on the above, at least five factors might be expected to affect the use of adversative connectives in translated texts: (a) the language direction, (b) the influence of the source text, (c) existing conventions in the target language, (d) the genre to which a text belongs and (e) the audience a text addresses. Due to practical limitations, that is availability of data, it is not possible to study all these factors in a single case study focusing on English–Russian. Specifically, it is not possible to create a reciprocal corpus of English and Russian texts consisting of different genres and addressing different audiences. Although different genres are translated into Russian, when examining the reverse direction, that is Russian into English, the majority of translations consists of fiction for adults. To take into account the maximum number of factors, and include those factors that also affect the use of adversative connectives in monolingual production, language direction is not examined in this case study. Thus, any conclusion reached in this case study is only valid regarding translation from English into Russian and not vice versa.

Based on the discussion above, it is possible to expand the broad research aim, which has been mentioned at the beginning of this chapter (i.e. the examination of the pragmatic factors affecting the use of connectives in translated texts), into a set of more specific research questions:

1) Are adversative connectives distributed differently in English and Russian non-translated texts belonging to different genres (fiction, non-fiction, children's fiction)?
2) Does the genre (fiction, non-fiction, children's fiction) to which a text belongs and the audience (adults vs. children) to which a text is addressed have an impact on the use of adversative connectives in translation?
3) How is the use of adversative connectives in translation compared to that in non-translated texts?
4) What role do source texts play?

As already explained, the first research question aims at scoping out whether it is meaningful to distinguish between the influence of non-translated texts in the target language and source texts. If the answer to the first research question suggests that there are differences, then the last question will also be addressed. In the opposite

scenario, the role of the source texts will not be examined. It should be stressed here that, since different genres are examined, it is possible that the role of the source texts will need to be considered for some genres, but not for others. The second question aims at addressing the genre and audience factors, while the third question examines whether we might be in a position to talk about a translation-specific language. The way in which the research questions interact is a first indication that data will be analysed in an integrated manner, which is an essential condition for triangulation. It also implies that sequential corpus data triangulation will need to be employed, as the analysis of one type of data is a prerequisite for the rest. Apart from corpus data triangulation, corpus method triangulation needs to be used, as it is necessary to identify meaningful, that is, statistically significant, differences, and also measure how big or small these are. The analysis focuses on the overall frequency of adversative connectives and the variation of their distribution, but also on the syntactic properties of these connectives, namely the differences in the distribution between coordinating and subordinating connectives.

5.6 Conclusion

The aim of this chapter has been to set the background for the case study of the factors affecting the distribution of adversative connectives in English–Russian translation. The close examination of previous studies on the topic in translation studies, as well as of the properties of adversative connectives in the two languages, has provided a definition of what is to be considered an adversative connective in this case study and has informed the research context and the specific research questions. We identified genre as a more interesting factor to investigate compared to the language direction since genre-related issues have acquired growing prominence in studies of connectives, at least in English. It is somewhat surprising that existing research on connectives in translation has not touched upon the issue, with only Mauranen (2000) comparing their distribution between two different genres. By employing a corpus triangulation framework, this case study aims to address this gap and inspire more similar studies in the future.

Notes

1 The only exception to this is Mauaranen's (2000) study on the use of multi-word strings (including text-internal connectors), where reciprocal corpora belonging to two different genres, namely academic science and popular science, are used to reveal differences based on genre. However, Mauaranen is interested in testing the hypothesis that cultural prestige in academic texts might determine translation decisions, rather than in specifically examining the factors that affect the distribution of these linguistic features in translated texts.
2 Readers interested in finding out more about the most recent argument regarding explicitation, and the idea of translation universals related to it, are encouraged to read Mauranen and Kujamäki (2004), Becher (2011a) and Murtisari (2014).
3 Similarly, the Russian connective *no* can have either an adversative proper meaning or a corrective one.
4 It must be noted that this list, as well as the list in Table 5.1, although detailed, is not exhaustive.

References

Anscombre, J.C. and Ducrot, O., 1977. Deux mais en Français? [Two but in French]. *Lingua*, 43(1), pp. 23–40.

Baumgarten, N. and Özçetin, D., 2008. Linguistic Variation Through Language Contact in Translation. In P. Siemund and N. Kintana, eds. *Language Contact and Contact Language*. Amsterdam and Philadelphia: John Benjamins, pp. 293–316.

Becher, V., 2010. Towards a More Rigorous Treatment of the Explicitation Hypothesis in Translation Studies. *Trans-Kom 3*, 3(1), pp. 1–25.

Becher, V., 2011a. *Explicitation and Implicitation in Translation: A Corpus-Based Study of English-German and German-English Translations of Business Texts*. University of Hamburg.

Becher, V., 2011b. When and Why Do Translators Add Connectives? A Corpus-Based Study. *Target*, 23(1), pp. 26–47.

Beloshapkova, V.A., 1977. *Sovremennyy Russkiy Yazyk [Modern Russian Language]*. Moscow: Vysshaya Shkola.

Biber, D., 2006. *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Biber, D. et al., 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Blum-Kulka, S., 1986. Shifts of Cohesion and Coherence in Translation. In J. House and S. Blum-Kulka, eds. *Interlingual and Intercultural Communication*. Tubinger: Gunter Narr, pp. 17–35.

Buss, K. and Karnowski, L., 2002. *Reading and Writing Nonfiction Genres*. Newark: International Reading Association.

Carter, R. and McCarthy, M., 2006. *Cambridge Grammar of English*. Cambridge: Cambridge University Press.

Celce-Murcia, M., Larsen-Freeman, D. and Williams, H.A., 1999. *The Grammar Book: An ESL/EFL Teacher's Course* 2nd ed. Boston: Heinle and Heinle.

Chernikova, N.S., 2007. K Svedeniyu Perevodchika: Soderzhatel'nyy i Formal'nyy Aspekty Slozhnopodchinennykh Ustupitel'nykh Konstruktsiy v Russkom i Angliyskom Yazykakh [Information for Translators: Semantic and Formal Peculiarities of Complex Concessive Constructions in Russian]. *Voronezh State University Bulletin: Linguistics and Intercultural Communication*, 1, pp. 166–171.

Christiansen, T., 2011. *Cohesion: A Discourse Perspective*. Bern: Peter Lang.

Chung-ling, S., 2008. Corpus-Based Study of Differences in Explicitation: Between Literature Translations for Children and for Adults. *Translation Journal*, 12(3).

Degand, L. and Fagard, B., 2012. Competing Connectives in the Causal Domain: French Car and Parce que. *Journal of Pragmatics*, 44(2), pp. 154–168.

Denturck, K., 2012. Explicitation vs. Implicitation: A Bidirectional Corpus-Based Analysis of Causal Connectives in French and Dutch Translations. *Across Languages and Cultures*, 13(2), pp. 211–227.

Dorgeloh, H., 2004. Conjunction in Sentence and Discourse: Sentence-Initial and Discourse Structure. *Journal of Pragmatics*, 36(10), pp. 1761–1779.

Finch, G., 2000. *Linguistic Terms and Concepts*. London: Macmillan Press.

Gotti, M., 2003. *Specialised Discourse*. Bern: Peter Lang.

Granger, S. and Tyson, S., 1996. Connector Usage in the English Essay Writing of Native and Non-Native EFL Speakers of English. *World Englishes*, 15(1), pp. 17–27.

Hague, A., 2003. *Fiction, Intuition, and Creativity: Studies in Brontë, James, Woolf, and Lessing*. Washington: Catholic University of America Press.

Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English*. London and New York: Routledge.

Halverson, S., 2004. Connectives as a Translation Problem. In K. Herald et al., eds. *Übersetzung—Translation—Traduction: An International Encyclopedia of Translation Studies*. Berlin and New York: Walter de Gruyter, pp. 562–572.

Hansen-Schirra, S., Neumann, S. and Steiner, E., 2007. Cohesive Explicitness and Explicitation in an English-German Translation Corpus. *Languages in Contrast*, 7(2), pp. 241–266.

Huddleston, R. and Pullum, G., 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Jasinskaja, K. and Zeevat, H., 2008. Explaining Additive, Adversative and Contrast Marking in Russian and English. *Revue de Semantique et Pragmatique*, 24(2008), pp. 65–91.

Jasinskaja, K. and Zeevat, H., 2009. Explaining Conjunction Systems: Russian, English, German. In A. Riester and T. Solstad, eds. *SinSpeC: Working Papers of the SFB 732 "Incremental Specification in Context."* Stuttgart: Stuttgart University, pp. 231–246.

Kapatsinski, V., 2009. Adversative Conjunction Choice in Russian (no, da, odnako): Semantic and Syntactic Influences on Lexical Selection. *Language Variation and Change*, 21(2), pp. 157–173.

König, E., 1985. On the History of Concessive Connectives in English: Diachronic and Synchronic Evidence. *Lingua*, 66(1), pp. 1–19.

Krilova, G., 1980. Ruskiyat Sŭyuz da i Negovite Funktsionalni Ekvivalenti v Khudozhestveni Prevodi na Bŭlgarski Ezik [The Russian Connective da and Its Functional Equivalents in Literary Bulgarian Translations]. *Comparative Linguistics*, 5, pp. 18–25.

Kruchinina, I.N., 1988. *Struktura i Funktsii Sochinitel'noy Svyazi v Russkom Yazyke [Structure and Functions of Coordinating Connectives in Russian]*. Moscow: Nauka.

Lakoff, R., 1971. "Ifs", "ands" and "buts" About Conjunction. In C. Fillmore and D. Langendoen, eds. *Studies in Linguistic Semantics*. New York: Holt, Rinehart and Winston, pp. 114–149.

Lamiroy, B., 1994. Pragmatic Connectives and L2 Acquisition: The Case of French and Dutch. *Pragmatics*, 4(2), pp. 183–201.

Ledenev, J.I., 1988. Nepolnoznachnyye Slova Kak Sredstva Oformleniya-Sintaksicheskikh Konstruktsiy, Svyazey i Otnosheniy [Semi-Notional Words as Means of Syntactic Constructions, Connections and Relationships]. In J.I. Ledenev, ed. *Nepolnoznachnyye Slova kak Sredstva Oformleniya v Sintaksise [Semi-Notional Words as Syntactic Means]*. Stavropol: Stavropol State Pedagogical Institute, pp. 3–19.

Lei, L., 2012. Linking Adverbials in Academic Writing on Applied Linguistics by Chinese Doctoral Students. *Journal of English for Academic Purposes*, 11(3), pp. 267–275.

Lekant, P.A. et al., 1982. *Sovremennyy Russkiy Literaturnyy Yazyk [Modern Russian Literary Language]*. Moscow: Moscow Higher School.

Liu, D., 2008. Linking Adverbials: An Across-Register Corpus Study and Its Implications. *International Journal of Corpus Linguistics*, 13(4), pp. 491–518.

Malchukov, A.L., 2004. Towards a Semantic Typology of Adversative and Contrast Marking. *Journal of Semantics*, 21(2), pp. 177–198.

Mauranen, A., 2000. Strange Strings in Translated Language: A Study on Corpora. In M. Olohan, ed. *Intercultural Faultlines: Research Models in Translation Studies I: Textual and Cognitive Aspects*. Manchester: St. Jerome Pub., pp. 119–142.

Mauranen, A. and Kujamäki, P. eds., 2004. *Translation Universals: Do They Exist?* Amsterdam and Philadelphia: John Benjamins.

McDowell, M., 1973. Fiction for Children and Adults: Some Essential Differences. *Children's Literature in Education*, 4(1), pp. 50–63.

Murtisari, E.T., 2014. Response to Becher's Two Papers on the Explicitation Hypothesis. *trans-kom*, 7(2), pp. 272–281.

Nikolajeva, M., 2002. *The Rhetoric of Character in Children's Literature*. New York: Scarecrow Press.

Olmos, S. and Ahern, A., 2009. Contrast and Propositional Attitude: A Relevance Theoretic Analysis of Contrast Connectives in Spanish and English. *Lingua*, 119(1), pp. 51–66.

Olohan, M. and Baker, M., 2000. Reporting That in Translated English: Evidence for Subconscious Processes of Explicitation? *Across Languages and Cultures*, 1(2), pp. 141–158.

Orlov, A.E. and Cheremisina, M.E., 1980. Kontaktnyye Sochetaniya Soyuzov i Chastits v Russkom Yazyke: K Postanovke Problemy [Combined Conjunctions and Particles in Russian: Problem Definition]. In M. Ivanova and M.E. Cheremisina, eds. *Polipredikativnyye Konstruktsii i ikh Morfologicheskaya Maza [Multi-Predicative Constructions and Their Morphological Basis]*. Novosibirsk: Nauka, pp. 217–218.

Øverås, L., 1998. In Search of the Third Code: An Investigation of Norms in Literary Translation. *Meta*, 43(4), pp. 557–570.

Pápai, V., 2004. Explicitation: A Universal of Translated Texts? In A. Mauranen and P. Kujamäki, eds. *Translation Universals: Do They Exist?* Amsterdam and Philadelphia: John Benjamins, pp. 143–165.

Peacock, M., 2010. Linking Adverbials in Research Articles Across Eight Disciplines. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*, 20(20), pp. 9–34.

Pit, M., 2007. Cross-Linguistic Analyses of Backward Causal Connectives in Dutch, German and French. *Languages in Contrast*, 7(1), pp. 53–82.

Priyatkina, A.F., 1977. Ob Otlichii Soyuza ot Drugikh Svyazuyushchikh Slov [On the Difference Between Conjunctions and Other Connective Words]. *Russkiy Yazyk v Shkole [Russian Language at School]*, 4, pp. 102–106.

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 and P. Kujamäki, eds. *Translation Universals: Do They Exist?* Amsterdam and Philadelphia: John Benjamins.

Quirk, R. et al., 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Rimmon-Kenan, S., 2002. *Narrative Fiction: Contemporary Poetics* 2nd ed. Abingdon and New York: Routledge.

Rogozhnikova, P., 1983. Slozhnyye Predlozheniya s Soyuznymi Chastitsami [Compound Sentences With Connective Particles]. In G.P. Uhanov, ed. *Sintaksis Predlozheniya [Sentence Syntax]*. Kalinin: Kursk State University, pp. 48–57.

Rudolph, E., 1988. Connective Relations—Connective Expressions—Connective Structures. In J. Petofi, ed. *Text and Discourse Constitution: Empirical Aspects, Theoretical Approaches*. Berlin: Walter de Gruyter, pp. 97–133.

Rudolph, E., 1996. *Contrast : Adversative and Concessive Relations and Their Expressions in English, German, Spanish, Portuguese on Sentence and Text Level*. Berlin and New York: Walter de Gruyter.

Serebryanaya, F.E., 1970. K Voprosu o Formirovanii Slozhnykh Sochinitel'nykh Soyuzov na Baze Neproizvodnykh [On the Formation of Complex Coordinating Conjunctions of the Basis of Non-Derivatives Ones]. In T.P. Lomteva and A.A. Kamyninoy, eds. *Issledovaniya po Sovremennomu Russkomu Yazyku [Modern Russian Studies]*. Moscow: Lomonosov Moscow State University, pp. 227–240.

Shvedova, N.J. et al., 1980. *Russkaja Grammatika [Russian Grammar]—Volume 1*. Moscow: Nauka.

Simmons, C., 1981. Cohesion in Russian: A Model for Discourse Analysis. *The Slavic and East European Journal*, 25(2), p. 64.

Stukker, N. and Sanders, T., 2012. Subjectivity and Prototype Structure in Causal Connectives: A Cross-Linguistic Perspective. *Journal of Pragmatics*, 44(2), pp. 169–190.

Sunderland, J., 2012. *Language, Gender and Children's Fiction*. London and New York: Continuum.

Uryson, E.V., 2000. Russkij Soyuz i Chastitsa i: Struktura Znacheniya [Russian Connective and Particle I: Meaning Structure]. *Issues in Linguistics*, 3, pp. 97–121.

Zareva, A., 2011. "And so that was it": Linking Adverbials in Student Academic Presentations. *RELC Journal*, 42(1), pp. 5–15.

Zufferey, S., 2012. "Car, parce que, puisque" Revisited: Three Empirical Studies on French Causal Connectives. *Journal of Pragmatics*, 44(2), pp. 138–153.

6 Corpus triangulation in the study of connectives

6.0 Introduction

This chapter explains the corpus data and method triangulation approach adopted for the examination of the factors related to the use of adversatives connectives in English–Russian translation. It begins by examining issues of corpus design and explains how corpus data triangulation is achieved by combining texts in different corpus configurations. It then presents a detailed account of the methods of analysis used in this case study and how these are combined to achieve corpus method triangulation. Finally, the results obtained from the corpus analysis are reported. It is only through multiple corpus triangulation (both data and method) that it is possible to examine how different factors, identified in Chapter 5, interact to form the language of translation used in various Russian genres. On the one hand, corpus data triangulation establishes which subcorpus needs to be examined during each stage of analysis. On the other hand, corpus method triangulation establishes the role played by each factor.

6.1 Corpus design

Data employed in this case study consist only of published books to facilitate corpus collection and increase comparability. The focus is on contemporary, twenty-first century, English and Russian written discourse, namely the years 2000–2015. This is mainly because translation into Russian is less likely to be ideologically censored and manipulated (Sherry 2010) after the fall of the Soviet Union and the end of the Cold War in 1991, removing an additional factor that might be responsible for specific linguistic preferences in translation. Additionally, electronic versions of books, which allow corpus researchers to avoid the laborious and time-consuming task of scanning from paper, have become readily available in more recent years. Making sure that all books (both English and Russian) have been published during the same time span makes the corpora directly comparable, avoiding cases where an English book has been translated into Russian a couple of decades after the publication of the English original. Based on the above and the research questions identified in the previous chapter, the corpus analysed for the purposes of this case study consists of the following nine components (Table 6.1).

Table 6.1 Corpus components

	Fiction	*Children's Fiction*	*Non-fiction*
non-translated English books	X	X	X
non-translated Russian books	X	X	X
translated Russian books	X	X	X

Different factors inform the size of the corpus. Books belonging to the genres of fiction, children's fiction and non-fiction can be lengthy, which facilitates the creation of a large corpus. A larger corpus is also necessary in this case if data from different authors and translators are to be captured. However, as already explained in the previous chapter, the nature of the linguistic phenomenon under investigation, that is connectives, requires close readings of the linguistic context, and manual refinement of concordance lines, which limit the corpus size. Taking all these into consideration, as well as the focus on specific genres, each of the nine corpus components is 1 million words, resulting in a 9-million-word corpus, which is a medium size for a corpus, according to Gavioli (2005).[1]

The corpus components described earlier are combined in different ways to create four smaller subcorpora which can be analysed separately to address the research questions identified in the previous chapter (Table 6.2).

Corpus data triangulation is achieved by combining texts in different corpus configurations based on the VVA typology. Firstly, by combining attributes from the *text variable*: fiction, children's fiction and non-fiction. Secondly, values from the *corpus type variable* are combined: comparable and parallel. Practically, there is a combination of two types of comparable corpora (one of non-translated texts in different languages, and one of translated and non-translated texts in the same language) and one type of parallel corpus. It is also possible to talk about a combination of values (both monolingual and bilingual comparable corpora) and attributes (both Russian and English non-translated texts) from the *languages variable*. Similarly, there is a further combination from the text variable (translated and non-translated texts). However, these combinations are tied to corpus type and, although important in achieving triangulation, are secondary here. This combination of corpora allows for detailed comparisons, as well as a close examination of how different factors might interact regarding the use of adversative connectives.

Corpus data triangulation is sequential, rather than simultaneous, as the results obtained from the analysis of one corpus inform which corpus needs to be examined next. Subcorpus A is examined to establish whether adversative connectives are used differently in Russian and English texts (Research Question 1). The subcorpus focuses on three genres, which are analysed separately, as it is possible that differences between the two languages might be observed in one genre, but not others. Subcorpus B is examined to establish whether the genre to which a text belongs, and the audience it addresses affect the use of adversative connectives in translation (Research Questions 2 and 3). It consists of three components,

Table 6.2 Corpus compilation

	Subcorpus	*Components*
A.	a comparable, bilingual (English-Russian), synchronic (2000–2015) subcorpus of non-translated texts (from the genres of fiction, children's fiction and non-fiction)	• non-translated English fiction books published in 2000–2015 • non-translated Russian fiction books published in 2000–2015 • non-translated English children's books published in 2000–2015 • non-translated Russian children's books published in 2000–2015 • non-translated English non-fiction books published in 2000–2015 • non-translated Russian non-fiction books published in 2000–2015
B.	a comparable, monolingual (Russian), synchronic (2000–2015) subcorpus of translated texts (from the genres of fiction, children's fiction, and non-fiction)	• translated Russian fiction books published in 2000–2015 • translated Russian children's books published in 2000–2015 • translated Russian non-fiction books published in 2000–2015
C.	a comparable, monolingual (Russian), synchronic (2000–2015) subcorpus of translated and non-translated texts (from the genres of fiction, children's fiction and non-fiction)	• non-translated Russian fiction books published in 2000–2015 • translated Russian fiction books published in 2000–2015 • non-translated English children's books published in 2000–2015 • translated Russian children's books published in 2000–2015 • non-translated Russian non-fiction books published in 2000–2015 • translated Russian non-fiction books published in 2000–2015
D.	a parallel, bilingual (English-Russian), synchronic (2000–2015) subcorpus (from the genres of fiction, children's fiction and non-fiction)	• non-translated English fiction books published in 2000–2015 • translated Russian fiction books published in 2000–2015 • non-translated English children's books published in 2000–2015 • translated Russian children's books published in 2000–2015 • non-translated English non-fiction books published in 2000–2015 • translated Russian non-fiction books published in 2000–2015

each focusing on texts belonging to a different genre, and results from each component are compared. Subcorpus C is examined to establish the influence of target linguistic conventions, and the possibility of a translation-specific language, focusing on the three different genres (Research Question 4). Finally,

subcorpus D is used to investigate the role of the source texts, examining each genre separately (Research Question 5). It is worth repeating here that only if the answer to the first research question suggests that there are differences, then the last question will also be addressed. In other words, subcorpus D is examined only if significant differences between Russian and English texts are observed in subcorpus A.

There is, of course, some overlap, and two subcorpora might share the same texts. For example, some of the components of the comparable monolingual Russian subcorpus of translated texts are also included in the comparable monolingual Russian subcorpus of non-translated texts, and in the parallel bilingual corpus. This overlap is evidence of corpus triangulation, and an indication that the approach adopted is integrated. Overall, the translated texts from the three different genres in Russian are compared with non-translated texts from the same genres, as well as the English source texts. Additionally, non-translated texts in Russian are compared to respective English texts (Figure 6.1).

Diversity is achieved in the corpora by including books belonging to different subcategories of each genre and selecting different authors and translators. For example, regarding non-fiction, books discussing business, history, psychology, science and environment, among others, are included, making sure that a range of different topics is covered. The same is true about the fiction corpus, where thrillers, historical novels, science fiction and romantic novels are included. Similarly, children's fiction consists of children's, as well young adult novels, following Knowles and Malmkjaer's (1996) definition of children's literature, and books belonging to fiction, horror, adventure and fantasy are included. Regarding authors and translators, attention is paid not to add the same person twice, as individual styles might affect the use of adversative connectives. Issues of currency

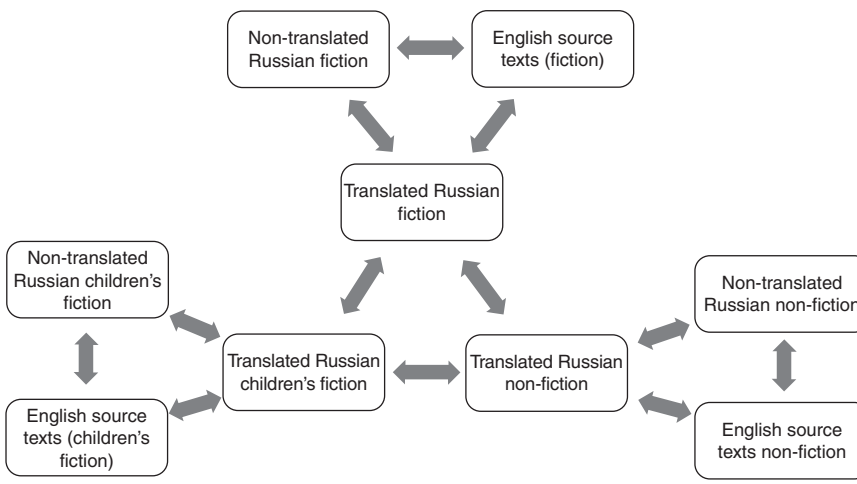

Figure 6.1 Schematic representation of corpus data triangulation

and influentialness have also been taken into consideration when designing the corpus for this study, especially for fiction and children's fiction, and an attempt was made to include as many top selling books as possible. For this, bestseller lists have been consulted, as well as rankings in sites such as www.goodreads.com. Finally, the corpus components for each genre are not internally balanced regarding the number of books, with fiction consisting of nine books for both translated and non-translated texts, children's fiction of 14 translated books, and 16 non-translated, and non-fiction of nine non-translated and 13 translated books. Individual books included in the same genre components are also different is terms of size. A detailed list of the books included in each corpus component can be found in Appendix 1.

6.2 Corpus analysis

Since a relatively large number of adversative connectives is available in each language, with some of these connectives having multiple functions, we will limit the analysis to a specific number of adversative connectives for each language, whose distribution corresponds to at least 90 per cent of the total number of adversative connectives. To identify these, corpus components belonging to the same language need to be combined, creating a large Russian subcorpus, and a smaller English subcorpus. Then, each of these subcorpora is searched for the adversative connectives listed in Sections 5.3 and 5.4 to establish their relative frequency in each language irrespective of genre. The software selected for the analysis of the corpus data is Wordsmith Tools 7.0. In particular, the Concord Tool is used to generate concordances based on word searches. A sampling technique is used (Sinclair 1999) for polysemous adversative connectives, which consists of selecting 30 random concordance lines and identifying adversative connectives in these, then selecting a different 30 concordance lines and so on until the proportion of adversative connectives out of the total number of concordance lines examined does not change with further selections of concordance lines. This sampling technique offers a close approximation of the total number of adversative connectives in the corpus. Once the list of adversative connectives with their respective approximate frequencies is completed, their proportions are calculated. Finally, the connectives, whose combined frequencies correspond to at least 90 per cent of the total number of adversative connectives, are identified (Table 6.3). This approach makes the corpus analysis more manageable, and provides more focused and detailed corpus results, without compromising on the reliability of the conclusions. Once all instances of connectives are captured, detailed tables are created, where the frequency of each connective in each corpus component and each book is documented. In total, nine such tables are created, one for each unique corpus component, which allow for the comparison across these.

Corpus method triangulation, which is sequential, is achieved by combining three types of descriptive statistics (raw frequencies, normalised frequencies and standard deviation) and three types of inferential statistics, (statistical significance, Bayes Factor and effect size). In other words, it is achieved though within-method triangulation. Where relevant, data are presented visually through

Table 6.3 Adversative connectives in Russian and English examined in this case study

	Russian		*English*	
	Connective	*Frequency*	*Connective*	*Frequency*
1.	Но (no)	35,644	but	19,797
2.	А (a)	6,310	though	1,882
3.	Однако (odnako)	3,554	although	814
4.	Хотя (khotya)	3,338	however	617
5.	на самом деле (na samom dele)	1,265	instead	610
6.	все таки (vse taki)	1,235	still	603
7.	Впрочем (vprochem)	1,130	anyway	546
8.	несмотря на (nesmotrya na)	931	yet	561
9.	Зато (zato)	824		
10.	все же (vse zhe)	817		
	Total	**55,048**		**25,430**

tables and graphs. Although it is generally considered good practice to combine quantitative and qualitative methods (especially in a book about triangulation), qualitative methods are not used here. The reason for this is to demonstrate to the readers the strengths of quantitative analyses, and to stress the fact that quantitative methods, especially inferential statistics, can help reach valid conclusions if used appropriately, thus highlighting their interpretive power, which is often ignored in corpus-based translation studies.

The descriptive measures employed include both raw and normalised frequencies, as well as standard deviation to measure dispersion across corpus components. The ratio of occurrence of adversative connectives is calculated by dividing the raw frequency of these linguistic items by the total number of words in each corpus component, similar to the approach followed by Milton and Tsang (1993), and Liu (2008). Bolton et al. argue that a word-based calculation is "fundamentally flawed" (2002, p. 172) because connectives mostly operate on the sentential level or beyond that. However, connectives might also link non-finite, dependent clauses (Chen 2006; Lei 2012), and pilot studies conducted with the corpus created for this case study indicate that it is typical for adversative connectives, especially subordinating ones, to connect elements below sentence level. Normalised frequencies, which allow comparisons across corpus components to be made, are calculated based on the ratio of occurrence per 100,000 words, following Granger and Tyson (1996). Numbers are rounded to the closest whole number, to avoid meaningless references to proportions of connectives.

Regarding inferential statistics, statistical significance is calculated using the log-likelihood test, while effect size is measured using percentage difference (%DIFF). Statistical significance is important because it helps ascertain whether the differences observed in the frequencies are not the result of coincidence, while percentage difference measures how large the differences are. Bayes Factor (BIC) is also calculated to measure the probability that a difference in frequency is due to chance. The null hypothesis (H_0) is that any differences observed in the

subcorpora are due to chance. The alternate hypothesis (H_1) is that the difference found can be attributed to a factor other than chance (e.g. influence for the source text or target language linguistic conventions). %DIFF and BIC are used only if results from the log-likelihood test indicate that differences are statistically significant, and thus there is indication that the H_0 can be refuted. The use of inferential statistics helps measure which of the factors identified (genre, audience, target linguistic conventions, source language influence) serves the most important role. It is worth repeating here that while genre and audience are addressed as part of the same research question, they can be understood as separate factors, with genre referring to fiction, children's fiction and non-fiction and audience referring to adults and children. Thus, two genres might address the same audience but focus on different subject matters (fiction, non-fiction), or they might focus on a similar subject matter but address different audiences (fiction, children's fiction).

This case study makes use of multiple triangulation by employing both corpus data and corpus method triangulation. Corpus data triangulation establishes the different stages of analysis and dictates which subcorpus will be examined at each stage, while the actual comparison is achieved through corpus method triangulation. The stages of analysis are listed below. The analysis focuses on the overall frequency of the adversative connectives listed in Table 6.3 and the variation of their distribution, but also on the syntactic properties of these connectives, namely the differences in the distribution between coordinating and subordinating connectives in each of the corpus components. Coordinating adversative connectives are *but* for English and *no*, *a* and *odnako* for Russian. All other adversative connectives in Table 6.3 are subordinating.

1. The examination of whether adversative connectives are distributed differently in English and Russian non-translated texts across different genres
2. The examination of whether adversative connectives are distributed differently across three translated genres in Russian
3. The investigation of whether any observed differences might be related to the linguistic preferences found in respective non-translated Russian genres
4. The investigation of whether any observed differences might be traced back to the English source texts

The final stage is included only if the results from the first stage suggest that there are significant differences between English and Russian in their use of adversative connectives. Each stage of analysis corresponds to the analysis of each of the subcorpora described in Section 6.1. A unique aspect of the corpus analysis, which is further evidence of the integrated approach adopted, is that the first two stages of analysis mainly generate corpus results and examine the distribution of connectives in each corpus component, while the other two stages of analysis mostly compare the results generated during the first two stages of analysis.

6.3 Corpus results

6.3.1 Linguistic preferences

The first stage of analysis involves the examination of the comparable, bilingual (English–Russian), synchronic (2000–2015) subcorpus of non-translated texts. Each of the components corresponding to fiction, children's fiction and non-fiction in English is compared to the respective component in Russian. The aim of this stage is to examine whether adversative connectives are used differently in the two languages in each of the three genres, which will guide the next stages of the analysis. The analysis of each of the six components reveals that there are differences in the distribution of adversative connectives in Russian and English texts for all three genres, with Russian texts showing a stronger preference for this linguistic feature compared to English texts (Table 6.4).

The three genres examined here can be ranked differently based on their preference for adversative connectives in Russian and English. While writers of Russian non-fiction demonstrate the strongest preference for adversative connectives, closely followed by writers of children's fiction, in English, the strongest preference is found in children's fiction, while the weakest in non-fiction. It is also evident that Russian writers make an overall higher use of adversative connectives across genres compared to English, without significant differences across genres, while in English, children's fiction clearly stands out from the other two genres.

Regarding fiction, adversative connectives are more frequent in Russian than in English texts (870 vs 667 per 100,000 words).[2] Statistical analysis reveals that this difference is significant ($p < 0.0001$), and there is very strong evidence against the H_0 (BIC = 280.31), although the effect size is small (30.4 per cent). Variation is also higher in Russian texts (SD = 185) than in English ones (SD = 153), but

Table 6.4 Distribution of adversative connectives in non-translated English and Russian texts

	Raw Frequency	*Normalised Frequency (100,000 words)*	*SD*
Non-translated English fiction	8,081/1,211,879	667	153
Non-translated Russian fiction	8,678/997,997	870	185
Non-translated English children's fiction	10,023/1,194,751	839	125
Non-translated Russian children's fiction	9,271/1,006,072	922	207
Non-translated English non-fiction	7,337/1,181,655	621	111
Non-translated Russian non-fiction	9,405/1,009,780	931	211

the difference is not particularly large in this case. A similar pattern is observed in children's fiction. Adversative connectives are employed slightly more frequently in Russian than in English texts (922 vs 839 per 100,000 words). While the statistical analysis suggests that this difference is significant ($p < 0.00001$) and that there is strong evidence against the H_0 (BIC = 27.79), effect size indicates that the difference is very small (8.96 per cent). As in fiction, a higher variation is observed in Russian (SD = 207) than English (SD = 125) texts. In this case, the difference in the variation is considerably larger than that reported for fiction. Finally, a similar pattern is found in non-fiction with Russian texts showing a stronger preference for adversative connectives than English (931 vs 621 per 100,000 words). This distribution is statistically significant ($p < 0.0001$), and there is very strong evidence against the H_0 (BIC = 669.85), while the effect size is similar to that found for fiction (33.4 per cent). A large variation is observed in Russian texts (SD = 211), while a considerably smaller one for English texts (SD = 111).

Figure 6.2 summarises the distribution of adversative connectives in each genre. As can be seen, Russian adversative connectives appear higher and also slightly further away from the line of best fit, which goes roughly through the middle of all the points, compared to English. This means that more variation is observed in Russian texts compared to English across all genres, which is also supported by the figures for standard deviation.

Based on the comparison of connectives in English and Russian in Chapter 5, it might be expected that while coordination is likely to be more frequent than subordination in both languages, English texts will make less frequent use of coordination, but more frequent use of subordination compared to Russian texts since more subordinating and fewer coordinating adversative connectives have been identified in English compared to Russian. Coordinating adversative connectives are found to be more common than subordinating adversative connectives in both languages across all genres, and in all cases, the difference is statistically significant, with very strong evidence against the H_0, and a very large effect size.[3] However, coordination is more frequent in Russian texts compared to English only in the genres of fiction and non-fiction, while Russian writers have been found to make more frequent use of subordinating adversative connectives compared to English in all three genres (Table 6.5).

Coordinating adversative connectives are less frequent in English than in Russian texts belonging to fiction (543 vs 600 per 100,000 words). A bigger difference is observed in non-fiction, where 439 coordinating adversative connectives are identified in English and 549 in Russian. The difference is statistically significant only in the case of non-fiction ($p < 0.0001$), and there is also very strong evidence against the H_0 (BIC = 226.74), although the effect size is not particularly large (44.28 per cent). The reverse pattern is observed in children's fiction, where coordination is more frequent in English than in Russian texts (667 vs 656 per 100,000 words). The difference is very small and statistical analysis confirms that it is not significant ($p > 0.05$)—thus, we cannot exclude the possibility that it has arisen by chance.

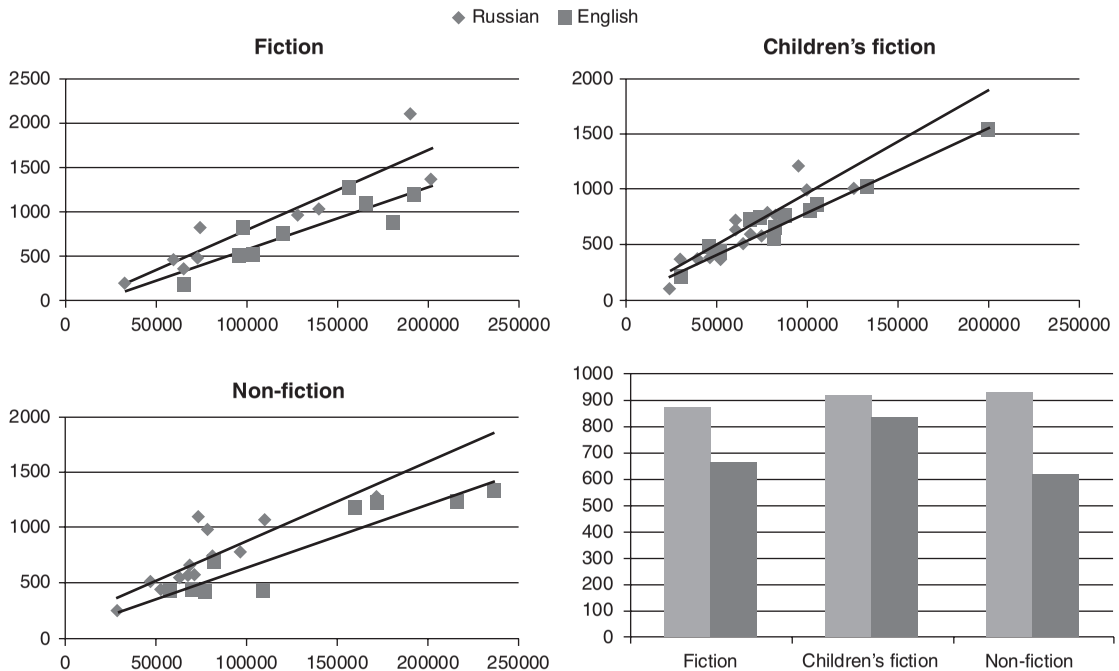

Figure 6.2 Graphic representation of the distribution of adversative connectives in Russian and English non-translated texts (per 100,000 words)

Table 6.5 Distribution of coordinating and subordinating adversative connectives in non-translated English and Russian texts

	Coordinating		*Subordinating*	
	Raw Frequency	*Normalised Frequency (100,000 words)*	*Raw Frequency*	*Normalised Frequency (100,000 words)*
Non-translated English fiction	6,577/ 1,211879	543	1,504/ 1,211879	124
Non-translated Russian fiction	7,233/ 997,997	600	1,445/ 997,997	270
Non-translated English children's fiction	8,038/ 1,194,751	677	1,940/ 1,194,751	162
Non-translated Russian children's fiction	7,751/ 1,006,072	656	1,520/ 1,006,072	266
Non-translated English non-fiction	5,182/ 1,181,653	439	2,155/ 1,181,653	182
Non-translated Russian non-fiction	7,948/ 1,009,780	599	1,457/ 1,009,780	332

Subordinating adversative connectives are less frequent in English than in Russian fiction (124 vs 270 per 100,000 words). A smaller difference is found in children's fiction—English 162 and Russian 266 per 100,000 words. The biggest difference is observed between English and Russian non-fiction (182 vs 332). However, it is only in non-fiction that the difference is found to be both statistically significant, and there is sufficient evidence against the H_0 ($p < 0.0001$, BIC = 33.71), although the effect size is small (26.39 per cent). Thus, the hypothesis that Russian texts make more frequent use of coordination than English texts can only be partially supported for the genre of non-fiction. The hypothesis that English texts might show a clearer preference for subordination compared to Russian cannot be supported, since English and Russian writers have been found to use subordination with a similar frequency, except for non-fiction where it is more frequent in Russian than English texts.

Example 6.1 illustrates the high use of coordination in Russian non-fiction, taken from the socio-political book *Проект Россия* (Project Russia) where the highest proportion of coordination is observed, that is *no, a, odnako* represent 93.73 per cent of adversative connectives used in the book. In contrast, in the English non-fiction historical book *Gulag: A History*, *but* represents only 52.6 per cent of all adversative connectives used in the book. In this book, subordination (most notably the use of *although* which represents 18.4 per cent of all connectives) is preferred (Example 6.2).

The comparison of non-translated texts in English and Russian demonstrates that there are significant differences between the two languages in the way writers employ adversative connectives, not only in the overall frequency of connectives,

Example 6.1

Но пора понять, что если болен весь организм, бессмысленно лечить руку. Если капает с потолка, надо бороться не с лужей на полу, **а** с дыркой в крыше. Чтобы спасти тонущее судно, нужно все силы бросить на заделывание пробоины. Если спасать отдельные каюты, утонут все. Действие имеет смысл, если оно скоординировано с общим планом. Действие ради действия бессмысленно. Сегодня тонет весь мир, **но** где взять людей, способных озадачиться проблемой такого масштаба? Кто сегодня хотя бы в мыслях может озадачиться не только спасением своей страны, **но** и всего мира?

(source: *Проект Россия*)

[back translation]

But it is time to understand that if the whole body is sick, it is useless to treat the hands. If it's dripping from the ceiling, one shouldn't focus on the puddle on the floor, **but** on the hole in the roof. To save the sinking ship, you need to throw all the forces to fixing holes. If you save individual cabins, everyone drowns. The action makes sense if it is coordinated with the general plan. Action for action's sake is pointless. Today the whole world is sinking, **but** where to find people capable of dealing with a problem of this magnitude? Who today would ever deal not only with the salvation of the country, **but** of the whole world?

Example 6.2

Yet although they lasted as long as the Soviet Union itself, and **although** many millions of people passed through them, the true history of the Soviet Union's concentration camps was, until recently, not at all well known. By some measures, it is still not known. Even the bare facts recited above, **although** by now familiar to most Western scholars of Soviet history, have not filtered into Western popular consciousness. "Human knowledge," once wrote Pierre Rigoulot, the French historian of communism, "doesn't accumulate like the bricks of a wall, which grows regularly, according to the work of the mason. Its development, **but** also its stagnation or retreat, depends on the social, cultural and political framework."

(source: *Gulag: A History*)

but also in terms of variation and some syntactic preferences regarding coordination and subordination in non-fiction. Even though these differences tend to be small, as demonstrated by the small effect size, there is sufficient evidence to suggest that they are not due to chance. Based on these findings, it is necessary to examine the influence of the source texts across all genres for the overall frequency of adversative connectives, but only regarding non-fiction for the syntactic patterning.

6.3.2 Genre differences

Before we study the influence of the target linguistic conventions and that of the source texts, it is necessary to investigate whether and to what extent there are differences in the use of adversative connectives in translated Russian texts across genres. For this, the comparable, monolingual (Russian), synchronic (2000–2015) subcorpus of translated texts (from the genres of fiction, children's fiction and non-fiction) is analysed. The aim is to examine whether genre and audience are parameters affecting the distribution of adversative connectives in translated texts. Results suggest that there are significant differences in the distribution of adversative connectives across genres (Table 6.6).

Different genres seem to employ adversative connectives with a different frequency, suggesting a correlation between genre and use of adversative connectives in translation, but the differences across genres are small (Figure 6.3). Children's fiction shows the strongest preference for adversative connectives (1,060 per 100,000 words), while non-fiction the lowest (792). Fiction stands somewhere between (889) the other two genres. Since fiction stands between the other two genres, we can compare the frequency of adversative connectives found in children's fiction and non-fiction to that found in fiction. This will also help demonstrate differences more clearly than conducting a statistical analysis across all three genres simultaneously. The difference between children's fiction and fiction is statistically significant ($p < 0.0001$), and there is very strong evidence against the H_0 (BIC = 137.44), but the effect size is very small (16.11 per cent). Similarly, the difference between non-fiction and fiction is statistically significant $p < 0.0001$), and there is very strong evidence against the H_0 (BIC = 41.13), but the effect size is again very small (12.22 per cent).

Regarding variation, the highest standard deviation is observed in the fiction component (SD = 184), closely followed by children's fiction (SD = 164), while the lowest is found in the non-fiction component (SD = 114). The differences

Table 6.6 Distribution of adversative connectives in translated Russian texts

	Raw Frequency	*Normalised Frequency (100,000 words)*	*SD*
Translated Russian fiction	8,978/1,009,715	889	184
Translated Russian children's fiction	10,811/1,009,996	1,060	164
Translated Russian non-fiction	7,805/985,075	792	114

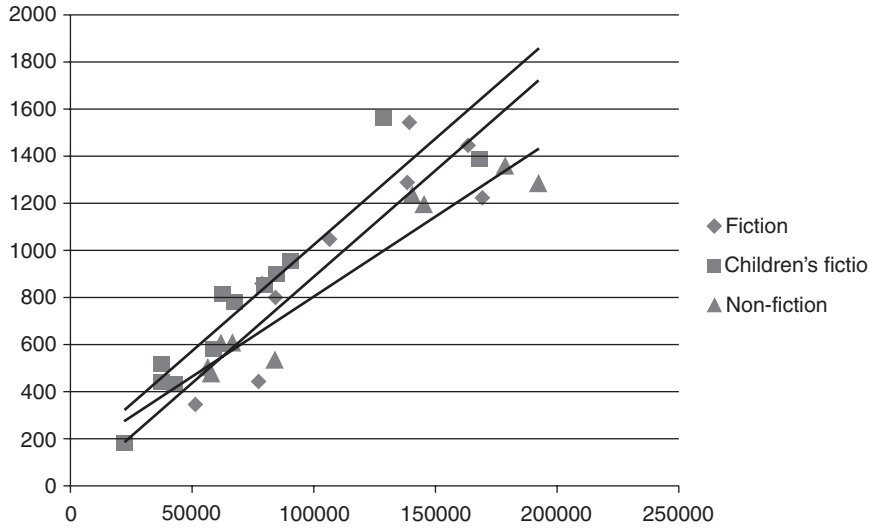

Figure 6.3 Graphic representation of the distribution of adversative connectives in translated Russian texts (per 100,000 words)

in the variation between fiction, children's fiction and non-fiction might be explained by the fact that fiction, irrespective of the audience, allows for much more creativity, compared to non-fiction, where linguistic conventions are likely to be more established. It will be possible to examine some reasons behind this variation when we compare translated texts to respective non-translated texts, as well as their English source texts.

Regarding syntactic properties, there is a clear preference for coordinating over subordinating adversative connectives across all genres (Table 6.7). The strongest preference for coordinating over subordinating adversative connectives is observed in children's fiction (805 per 100,000 words), while the lowest in non-fiction (566). Fiction stands between the other two genres (672). Thus, the results from the analysis of the syntactic properties of adversative connectives confirm those obtained from the analysis of the overall frequency of this linguistic feature regarding the ranking of genres. This preference for coordinating over subordinating adversative connectives is statistically significant for all three genres ($p < 0.0001$), with very strong evidence against the H_0 (BIC = 4,420.16 for fiction, BIC = 4,728.80 for children's fiction, BIC = 2,933.64 for non-fiction). A large effect size is observed across all three genres, with the strongest in fiction (409.92 per cent), and the smallest in non-fiction (293.05 per cent). Children's fiction stands between the other two genres (350.74 per cent).

It is perhaps to be expected that children's fiction employs adversative connections more frequently than fiction for adults, as younger readers require relationships between words and clauses to be signalled more clearly. However, it is somewhat surprising that non-fiction makes relatively low use of adversative

Table 6.7 Distribution of coordinating and subordinating adversative connectives in Russian translated texts

	Coordination		*Subordination*	
	Raw Frequency	*Normalised Frequency (100,000 words)*	*Raw Frequency*	*Normalised Frequency (100,000 words)*
Translated Russian fiction	7,506/ 1,009,715	672	1,472/ 1,009,715	217
Translated Russian children's fiction	8,848/ 1,009,996	805	1,963/ 1,009,996	255
Translated Russian non-fiction	6,222/ 985,075	566	1,583/ 985,075	227

connectives, as it is largely based on argumentation, and links between ideas need to be clearly marked, leading to the main point that the author wants to make. Similarly, it might be expected that children's fiction would strongly favour coordination since subordination might be considered as a more elaborate linguistic phenomenon, while children's fiction prefers simpler forms of argumentation compared to fiction for adults and non-fiction. For example, in the children's fiction component, the Russian translation of *Percy Jackson and the Lightning Thief* (Example 6.3) exhibits a very high proportion of coordination, namely 87.2 per

Example 6.3

Давай, Перси!—пронзительно завопила Аннабет, таща меня за запястье. — **Но** это . . . —Я знаю! —выкрикнула она. —Место, которое ты видел во сне! **Но** Гроувер свалится, если мы не удержим его! Разумеется, она была права. Беспокойство за Гроувера заставило меня снова двинуться вперед. Сатир пронзительно кричал, цеплялся за землю, впиваясь в нее ногтями, **но** крылатые туфли подтаскивали его все ближе к яме, и было очень сомнительно, что мы подоспеем вовремя.

(source: *Перси Джексон и Похититель Молний*)

[back translation]
Come on, Percy! —Annabeth screamed piercingly, pulling me by the wrist. **But** . . .
I know! she cried. —A place you saw in a dream! **But** Grover will fall if we don't keep him! Of course, she was right. Concern for Grover made me move forward again. Satyr screamed, clung to the ground, digging with his nails, **but** the winged shoes dragged him closer to the hole, and it was very doubtful that we would make it in time.

Example 6.4

Хотя горные снега и подземные воды, на первый взгляд, имеют мало общего, **на самом деле** они связаны.

(source: *Коллапс: почему одни общества выживают, а другие умирают*)

Although mountain snow and underground water, at first glance, have little in common, **actually** they are related.

cent. It is likely that coordinating English connectives are translated as coordinating connectives in Russian, but this relation will be examined in more detail in section 6.3.4. Similarly, non-fiction favours more elaborate means of expression and argumentation, which would explain why fewer coordinating adversative connectives were found in non-fiction texts compared to the other two genres. Example 6.4, which is an excerpt from the translation of *Collapse: How Societies Choose to Fail or Succeed*, illustrates the use of subordinating adversative connectives in Russian translated non-fiction. In this book 26.03 per cent of all adversative connectives are subordinating,

The analysis so far indicates that there are important differences across the three genres in translation concerning the overall frequency of adversative connectives, the degree of variation and the syntactic preferences. Even though differences are often small, there is clear and strong evidence that these are not due to chance. This is an indication that the type of publication, whether it is fiction or non-fiction, but also the target audience to which it is addressed, affect the way in which translators employ adversative connectives. In certain cases (e.g. variation), the subject matter appears to be responsible for the differences, and thus non-fiction stands out from the other two genres. In other cases (e.g. overall frequency), it is the audience that seems to be the main reason behind the differences, and thus children's fiction stands out from the other two genres. However, other parameters might also play a role, such as linguistic conventions found in respective target texts and source texts. These parameters are discussed in the following sections.

6.3.3 The role of target linguistic conventions

In the third stage of analysis, the focus is on the comparable, monolingual (Russian), synchronic (2000–2015) subcorpus of translated and non-translated texts (from the genres of fiction, children's fiction and non-fiction). The aim of the analysis is to reveal whether non-translated texts might have played a role in the distribution of adversative connectives observed in translation. The analysis reveals some interesting similarities and differences between non-translated and translated texts (Table 6.8).

Table 6.8 Distribution of adversative connectives in non-translated and translated Russian texts

	Raw Frequency	*Normalised Frequency (100,000 words)*	*SD*
Non-translated Russian fiction	8,978/1,009,715	870	185
Translated Russian fiction	8,678/997,997	889	184
Non-translated Russian children's fiction	10,081/1,009,996	922	211
Translated Russian children's fiction	9,271/1,006,072	1,060	164
Non-translated Russian non-fiction	9,045/1,009,780	931	207
Translated Russian non-fiction	7,805/985,075	792	114

In non-translated Russian texts, the strongest preference for adversative connectives is found in non-fiction, and the weakest in fiction, while in translated texts, children's fiction shows the strongest preference for adversative connectives, and non-fiction the lowest. The difference in the distribution of adversative connectives is quite small across non-translated genres, ranging from 870 to 931 connectives per 100,000 words, while the range is much wider in translated texts, that is between 792 and 1,060 connectives per 100,000 words (Figure 6.4). This simple comparison is preliminary indication that there are some important differences between translated and non-translated texts. It has been argued in the previous section that the relatively low frequency of adversative connectives in translated non-fiction is somewhat surprising. Evidence from the non-translated non-fiction component further supports this, since Russian writers seem to make extensive use of adversative connectives to convey their ideas.

Adversative connectives are used with approximately the same frequency in both translated and non-translated texts belonging to fiction (870 vs 889 per 100,000 words). The small difference in favour of translated fiction is not statistically significant ($p > 0.05$), and we can, thus, not exclude the possibility that it might have arisen by chance. The degree of variation in the translated and non-translated fiction components is also very similar (SD = 184 for translation and SD = 185 for non-translation). Unlike fiction for adults, differences are observed between translated and non-translated children's fiction. Translated texts show a stronger preference for adversative connectives compared to non-translated texts (1,060 vs 922). This difference is statistically significant ($p < 0.0001$), and there is very strong evidence against the H_0 (BIC = 97.79), but the effect size is very small (13.91 per cent). Greater variation is observed in non-translated children's fiction (SD = 207) compared to translated texts (SD = 164). This variation is also larger compared to that observed for texts belonging to fiction.

The reversed pattern is observed in texts belonging to non-fiction: a stronger preference for adversative connectives is observed in non-translated texts than translated ones (931 vs 792 per 100,000 words). This difference is statistically significant ($p < 0.0001$), and there is very strong evidence against the

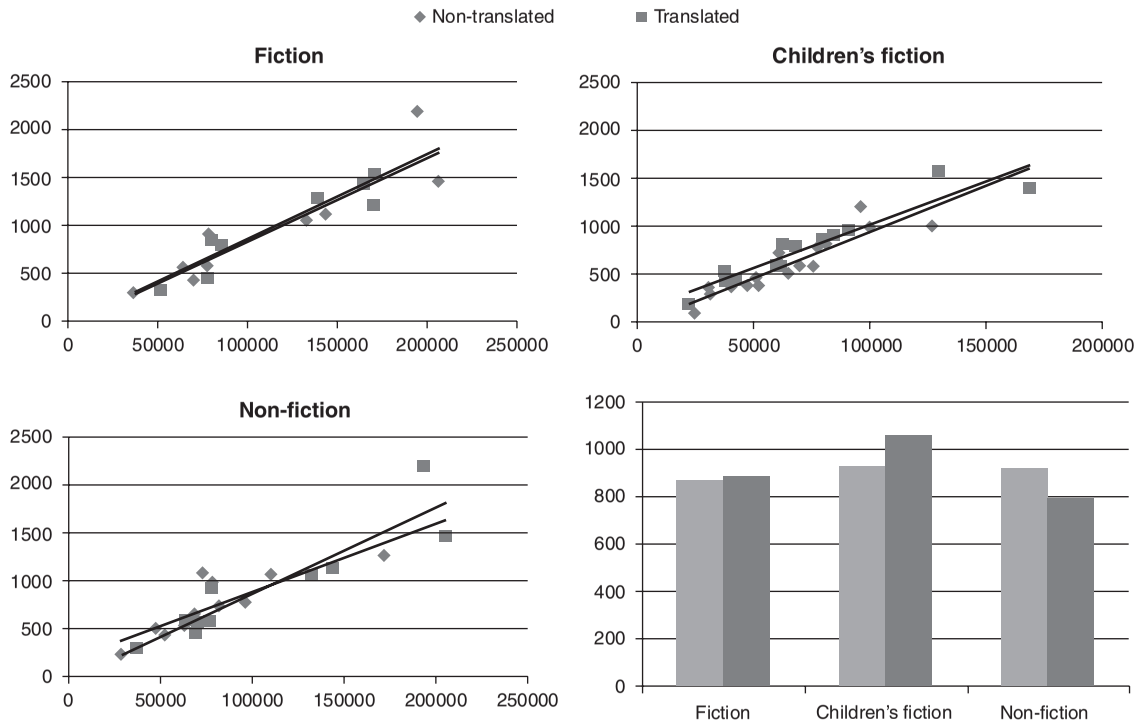

Figure 6.4 Graphic representation of the distribution of adversative connectives in non-translated and translated Russian texts (per 100,000 words)

H_0 (BIC = 97.47). However, the effect size is very small (17.55 per cent), but larger than that observed in texts belonging to children's fiction. Considerably greater variation is observed in translated texts (SD = 237) compared to non-translated ones (SD = 114). It can be argued that translators of non-fiction texts employ adversative connectives less frequently than Russian writers of such texts, but, at the same time, there is significant variation across translators. Overall, the biggest difference between translated and non-translated texts in the frequency of adversative connectives is found in non-fiction, while the smaller in fiction. Similarly, the largest degree of variation is observed in non-fiction, while the smallest in fiction.

Even though differences in syntactic preferences between English and Russian have only been observed in non-fiction, it is necessary to examine all three genres when examining the role of target linguistic conventions. This is because, if differences are observed, it will be an indication that some factor other than the influence of the target language is behind the use of coordination and subordination in translated texts, such as audience considerations. The corpus analysis suggests that while coordination is employed more frequently in translated texts, subordination is more frequently used in non-translated texts (Table 6.9).

Coordinating adversative connectives are employed more frequently in translated texts for both fiction and children's fiction compared to non-translated texts (672 vs 600 and 805 vs 656 per 100,000 words respectively). The difference is statistically significant only in the case of children's fiction ($p < 0.0001$), and there is strong evidence against the H_0 (BIC = 5.48), but the effect size is very small (12.06 per cent). The opposite pattern is observed in non-fiction,

Table 6.9 Distribution of coordinating and subordinating adversative connectives in Russian translated texts

	Coordination		*Subordination*	
	Raw Frequency	*Normalised Frequency (100,000 words)*	*Raw Frequency*	*Normalised Frequency (100,000 words)*
Non-translated Russian fiction	7,233/ 997,997	600	1,445/ 997,997	270
Translated Russian fiction	7,506/ 1,009,715	672	1,472/ 1,009,715	217
Non-translated Russian children's fiction	7,751/ 1,006,072	656	1,520/ 1,006,072	266
Translated Russian children's fiction	8,848/ 1,009,996	805	1,963/ 1,009,996	255
Non-translated Russian non-fiction	7,948/ 1,009,780	599	1,457/ 1,009,780	332
Translated Russian non-fiction	6,222/ 985,075	566	1,583/ 985,075	227

with coordinating adversative connectives being more frequent in non-translated than translated texts (599 vs 566). The difference is statistically significant ($p < 0.0001$), there is very strong evidence against the H_0 (BIC = 155.68), and the effect size, although small (24.62 per cent), is larger than that observed for children's fiction, suggesting that the difference in the case of non-fiction is slightly more marked compared to children's fiction.

Conversely, subordinating adversative connectives are employed more frequently in non-translated texts compared to translated ones across all three genres: fiction (270 vs 217), children's fiction (266 vs 255) and non-fiction (332 vs 227). The difference is statistically significant only in the case of children's fiction ($p < 0.00001$) and non-fiction ($p < 0.01$). However, it is only in the case of children's fiction that there is also very strong evidence against the H_0 (BIC = 40.27), even though the effect size is small (22.27 per cent). In the case of non-fiction, there is positive evidence in favour of the H_0 (BIC = –5.7), and the effect size is very small (10.21 per cent). Thus, the only genre in which it can be argued that target linguistic conventions have played a critical role in the use of adversative connectives is fiction. In the case of children's fiction, it is clear that translators seem to be more willing to depart from target preferences, at least as far as the syntactic properties of adversative connectives are concerned. It should be noted however that this departure is minor, as suggested by the small percentage difference. Results are inconclusive regarding non-fiction.

The comparison between translated and non-translated texts reveals that, although the language combination and direction of translation are kept the same, different patterns are observed across different genres, to such an extent that it is not possible to try and make any generalisations about translated texts from Russian into English. As far as fiction for adults is concerned, translators employ adversative connectives with the same frequency as Russian writers and thus seem to be influenced by the linguistic conventions found in non-translated Russian fiction. There is also very similar variation in the fiction components, and potential influence has been observed regarding coordination and subordination. On the contrary, in children's fiction, translators systematically employ adversative connectives with a higher frequency compared to Russian children's writers, and non-translated children's fiction seems to exert little influence on respective translated texts. There are also clear differences regarding syntactic preferences between translated and non-translated texts belonging to this genre. Finally, translators of non-fiction employ adversative connectives less frequently, make lower use of coordination, and there is a particularly clear difference regarding variation compared to Russian writers of such texts. Thus, it is only in the case of fiction that influence from target linguistic conventions is observed. To understand the differences better, it is necessary to examine the English source texts to see whether they might have also exerted some influence.

6.3.4 The role of source texts

The final stage of analysis involves the examination of the parallel, bilingual (English–Russian), synchronic (2000–2015) subcorpus (from the genres of fiction, children's

fiction and non-fiction) to investigate whether the English source texts might have played a role in the distribution of adversative connectives in Russian translated texts. The role of the source texts is examined as a separate factor for the overall frequency of adversative connectives across all genres since differences in the distribution of adversative connectives have been identified between non-translated Russian and English texts for all of these during the first stage of analysis. However, it is only examined in relation to non-fiction regarding syntactic preferences, since this is the only genre where differences between English and Russian texts have been observed during the first stage of analysis. The analysis indicates that English source texts make considerably lower use of adversative connectives compared to Russian target texts (Table 6.10).

In all three genres, adversative connectives are used more frequently in translated Russian texts compared to their English source texts: fiction (889 vs 667), children's fiction (1,060 vs 839) and non-fiction (792 vs 621) (Figure 6.5). Statistical analyses suggest that the differences are significant for all three genres.[4] Regarding variation, a slightly higher variation is observed for translated fiction (SD = 184) compared to the English source texts (SD = 153), a higher variation is observed for translated children's fiction (SD = 164) compared to the English source texts (SD = 125), while it is approximately the same for both the Russian translated texts (SD = 114) and their English source texts (SD = 111).

To further illustrate this, *Fast Food Nation* is the English title in the non-fiction subcorpus with the lowest proportion of adversative connectives (392 per 100,000 words), while *The Swerve: How the World Became Modern* has the highest proportion of these (840 per 100,000 words). It is the translations of these titles that also demonstrate the lowest and highest proportion of adversative connectives respectively in the Russian non-fiction subcorpus. Thus, there is evidence to suggest that the variation observed in the translated non-fiction subcorpus can be justified by the influence from the English source texts. No similar pattern can be found in fiction and children's fiction. Example 6.5, which is taken from

Table 6.10 Distribution of adversative connectives in English source texts and translated Russian texts

	Raw Frequency	*Normalised Frequency (100,000 words)*	*SD*
Non-translated English fiction	8,081/1,211,879	667	153
Translated Russian fiction	8,978/1,009,715	889	184
Non-translated English children's fiction	10,023/1,194,751	839	125
Translated Russian children's fiction	10,811/1,009,996	1,060	164
Non-translated English non-fiction	7,337/1,181,655	621	111
Translated Russian non-fiction	7,805/985,075	792	114

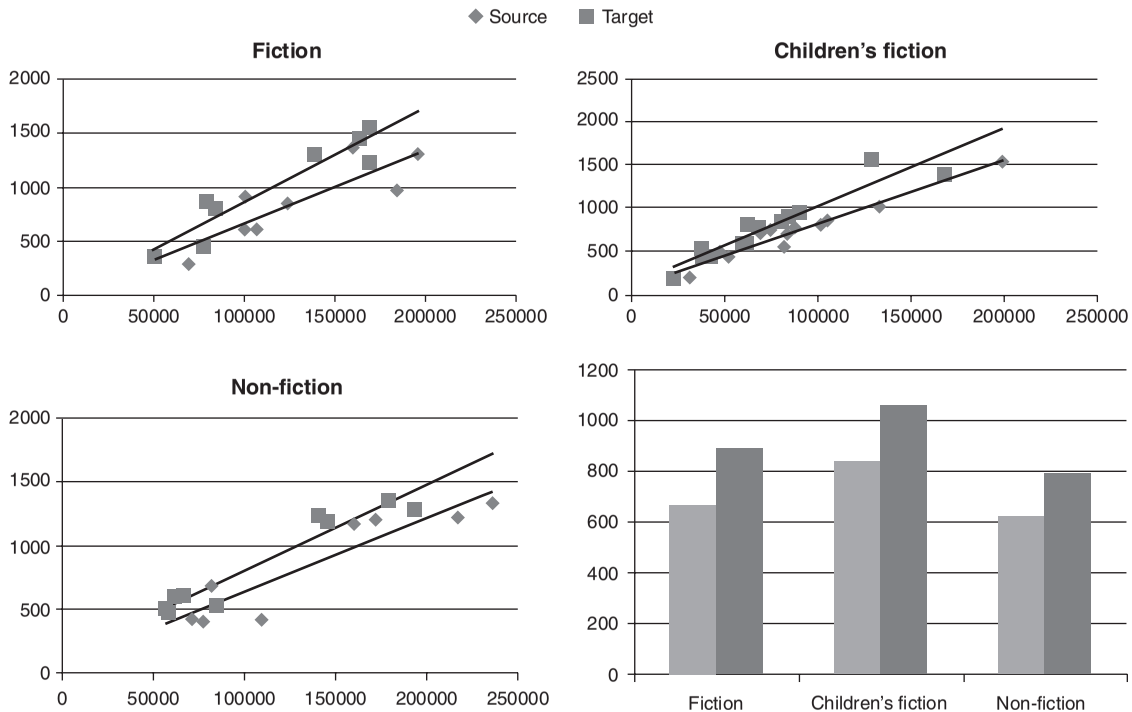

Figure 6.5 Graphic representation of the distribution of adversative connectives in English source texts and Russian target texts (per 100,000 words)

Example 6.5

I'm telling you, I don't think it was the Shax. I think the Shax was pursuing her—hunting her down for something, or someone, else.

(source: *Clockwork Angel*)

Сказал ведь, это не шакс! Возможно, где-то она и столкнулась с шаксом, от которого попыталась убежать. Может быть, демон даже погнался за ней. **Но** вот убило ее что-то другое.

(source: *Механический ангел*)

[back translation]
I told you, it's not the Shax! It is possible that she faced the Shax somewhere, from which she tried to escape. It might be that the demon chased her. **But** something else killed her.

the children's fiction component, illustrates how translators often add adversative connectives in the target text, to mark the connections between sentences. Notice also the amount of other information that the translator adds. This provides evidence that audience considerations significantly affect the use of adversative connectives in children's fiction.

The subcorpus analysis also indicates that both coordination and subordination are employed more frequently in Russian translated non-fiction compared to the English source texts. In particular, coordinating adversative connectives are employed more frequently in Russian translated non-fiction than the English source texts (566 vs 439 per 100,000 words).[5] The difference is statistically significant ($p < 0.0001$), and there is very strong evidence against H_0 (BIC = 363.88), but the effect size is small (30.57 per cent). As far as coordination is concerned, English source texts do not seem to exert any influence on translated texts.

A slightly different pattern is found regarding subordination. Subordination is more frequent in translated texts than source texts (277 vs 182).[6] The difference is statistically significant ($p < 0$ .001). However, there is not sufficient evidence against the H_0 (BIC = 0.10), suggesting that, even though there is a statistically significant difference, this is not particularly strong. At the same time, the effect size is very small (13.49 per cent). Unlike coordination, some influence might be exerted from English source texts on translated non-fiction, although quantitative data are not conclusive. This might be attributed to fact that translators often translate English subordinating connectives into Russian coordinating ones (Example 6.6).

Overall, the subcorpus analysis in this section suggests that the use of adversative connectives in translated Russian texts can be attributed to some extent to the influence from English source texts in the case of non-fiction, as far the

Example 6.6

Though perhaps there ought to be a box for people to check, or not check, on their body donor form: Okay to use me for cosmetic purposes.

(source: *Stiff: The Curious Lives of Human Cadavers*)

Однако, возможно, следовало бы предлагать донорам заполнять графу: Согласен (или не согласен) использоваться в косметических целях. (source: *Кадавр. Как тело после смерти служит науке*)

[back translation]
But, perhaps, donors should be offered a box to check: I agree (or disagree) to be used for cosmetic purposes.

variation in corpora and the distribution of subordination are concerned, but not in the case of fiction and children's fiction. Linking this finding to findings from previous stages of analysis, it is clear that the use of adversative connectives in translation is affected not only by an interplay of factors, but that these factors affect different aspects of their use (i.e. overall frequency, variation, syntactic preferences). In the remainder of this chapter, the results from the different stages of analysis are combined and discussed in more detail.

6.4 Discussion

The aim of this case study has been to examine the factors affecting the use of adversative connectives in translated Russian texts across three genres, namely fiction, children's fiction and non-fiction. Before we summarise results, it must be stressed that any differences reported are rather small, as evidenced by the small effect size, which is what makes them particularly difficult to capture. Based on the analysis of the three genres, it is possible to examine the use of adversative connectives for different subject matters (fiction and non-fiction), and audiences (adults and children). A multiple corpus triangulation approach has been adopted, where triangulation occurred both through data triangulation, and within-method (quantitative) triangulation. Results suggest that different factors can be identified, which affect different genres to a different degree and level. These factors are (a) influence from existing target language linguistic conventions, that is, the linguistic preferences found in non-translated texts; (b) source language interference, that is, influence from the English texts; and (c) other factors related to the particularities of each genre, such as audience. Thus, it is not possible to talk about a single factor affecting the use of this specific linguistic feature in translation or about a more general translation-specific language. This

could explain why previous studies on the use of connectives in translation have produced contradicting evidence as to the existence of explicitation phenomena (see Section 5.1).

Translated Russian fiction is very close to non-translated fiction in the same language, and there is clear and strong evidence suggesting there is significant influence from existing linguistic conventions found in non-translated Russian fiction. This is supported by results from the overall frequency of adversative connectives, the analysis of variation across individual titles, and the examination of syntactic preferences.

Regarding children's fiction, translated Russian texts make a much higher use of adversative connectives than both non-translated Russian and English children's fiction. This could be considered an indication that there is something particular about the translation of this genre that might have encouraged this significantly higher use of adversative connectives. What differentiates this genre from the other two is the age group of the readers. Thus, it is argued here that considerations of comprehensibility more generally have preoccupied translators, who wanted to create more accessible texts for children in particular, which is also supported by Chung-ling (2008). Translators are aware of their audience, and they adjust their translation techniques accordingly. This is also supported by the results obtained from the analysis of syntactic patterns, at least regarding coordination. In term of variation, translated children's fiction stands between non-translated Russian texts and English source texts, suggesting that the linguistic conventions found in both the source texts and non-translated texts in the same language might have exercised some influence. Overall, audience considerations seem to have played the most prominent role in the distribution of connectives in translations of this genre.

Finally, regarding non-fiction, translated texts seem to be somewhere in between Russian non-translated texts and English source texts, suggesting that some influence from both source language and Russian non-translated language might have been exerted, as far as the overall frequency of adversative connectives and the distribution of coordination is concerned. Results from the analysis of variation across corpora, however, provide support for the possible influence from the source texts, since the same variation is observed between translated texts and their source texts. Also, there is some indication of the possible influence from English source texts in the analysis of subordination, even though results are inconclusive. That translated non-fiction is affected to some extent by English linguistic conventions might be explained by the fact that non-fiction appears to be much more established in English-speaking countries, with some books becoming international bestsellers, while the same cannot be said for Russian non-fiction, which has a much smaller market. Thus, the prestige of these English bestsellers might be the reason why Russian titles show some similarities to English titles. In other words, cultural prestige might determine translation decisions, similar to what has been found by Mauranen (2000) for scientific writing.

Had corpus triangulation techniques not been employed, it would not have been possible to examine the way in which these different factors interact to form

the language of translation used in various Russian genres. On the one hand, the use of a large number of corpus components, which allow for the combination of different values, variables and attributes has facilitated the investigation of different factors, while the use of different quantitative methods has revealed meaningful patterns and has helped identify the exact role played by each of these factors. The corpus results from each stage interact, cross-fertilize ideas and provide insights. This is also evident from the fact that the first two stages of analysis focus more on generating corpus results, while the other two stages on comparing and contrasting these results. Each stage provides a partial answer to the research question, while it is only by combining the results from all four that a clear conclusion can be reached.

6.5 Conclusion

The corpus data triangulation employed in this case study allowed for meaningful comparisons to be made across different types of texts, in an effort to acquire a comprehensive understanding of the phenomenon of adversative connectives and how these are employed in translation. It has been found that a complex interplay of factors affects the use of adversative connectives in Russian translation from English, and these can be related to the influence from existing target language linguistic conventions, the source language interference, and audience, or other genre-specific, considerations. This is not the first time that corpora consisting of different types of texts are combined in the investigation of similar phenomena. However, this is the first time that this is conducted in a systematic and integrated way, taking into consideration different factors. Moreover, the corpus method triangulation and the use of descriptive and inferential statistics allowed for results to be confirmed, or contradicted, thus increasing the confidence of the conclusions, and offering a complete account of the phenomenon under investigation. It should be stressed that, while the focus is on English and Russian adversative connectives, the corpus design and methods employed can be adjusted to the study of the linguistic and discursive properties of a wide range of linguistic phenomena related to the investigation of the language of translation.

Notes

1 Since only full texts are included, and Russian translations tend to be shorter than English source texts, some of the corpus components consisting of Russian translated texts are slightly smaller than 1 million words, while those consisting of English source texts are slightly larger. This inconsistency does not affect the comparability of corpus data, as both raw and normalised frequencies are employed.
2 For brevity and clarity of argumentation, only normalised frequencies are reported in the discussion. For raw frequencies, readers should consult the detailed tables in each section.
3 English fiction: $p < 0.0001$, BIC = 3,421.44, 337.3 per cent. English children's fiction: $p < 0.0001$, BIC = 4,030.81, 314.33 per cent. English non-fiction: $p < 0.0001$, BIC = 1,272.25, 140.46 per cent. Russian fiction: $p < 0.0001$, BIC = 1,262.94,

400.55 per cent. Russian children's fiction: $p < 0.0001$, BIC = 1,700.77, 409.93 per cent. Russian non-fiction: $p < 0.0001$, BIC = 766.31, 445.50 per cent.

4 Fiction: The difference is significant ($p < 0.0001$), and there is very strong evidence against the H_0 (BIC = 338.13), but a small effect size (33.34 per cent).

Children's fiction: The difference is statistically significant ($p < 0.0001$), and there is very strong evidence against the H_0 (BIC = 294.40), but the effect size is once again small (27.56 per cent).

Non-fiction: The difference is statistically significant ($p < 0.0001$) providing evidence that it has most likely not arisen by chance. There is also very strong evidence against the H_0 (BIC = 210.19), but the effect size is once again small (27.61 per cent).

5 Raw frequencies: 5,152 coordinating adversative connectives out of 1,181,653 words for English and 6,222 out of 985,075 words for Russian translated texts.

6 Raw frequencies: 2,155 subordinating adversative connectives out of 1,181,653 words for English and 1,583 out of 985,075 words for Russian translated texts.

References

Bolton, K., Nelson, G. and Hung, J., 2002. A Corpus-Based Study of Connectors in Student Writing: Research From the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*, 7(2), pp. 165–182.

Chen, C.W., 2006. The Use of Conjunctive Adverbials in the Academic Papers of Advanced Taiwanese EFL Learners. *International Journal of Corpus Linguistics*, 11(1), pp. 113–130.

Chung-ling, S., 2008. Corpus-Based Study of Differences in Explicitation: Between Literature Translations for Children and for Adults. *Translation Journal*, 12(3).

Gavioli, L., 2005. *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins.

Granger, S. and Tyson, S., 1996. Connector Usage in the English Essay Writing of Native and Non-Native EFL Speakers of English. *World Englishes*, 15(1), pp. 17–27.

Knowles, M. and Malmkjær, K., 1996. *Language and Control in Children's Literature*. Abingdon and New York: Routledge.

Lei, L., 2012. Linking Adverbials in Academic Writing on Applied Linguistics by Chinese Doctoral Students. *Journal of English for Academic Purposes*, 11(3), pp. 267–275.

Liu, D., 2008. Linking Adverbials: An Across-Register Corpus Study and Its Implications. *International Journal of Corpus Linguistics*, 13(4), pp. 491–518.

Mauranen, A., 2000. Strange Strings in Translated Language: A Study on Corpora. In M. Olohan, ed. *Intercultural Faultlines: Research Models in Translation Studies I: Textual and Cognitive Aspects*. Manchester: St. Jerome Pub., pp. 119–142.

Milton, J. and Tsang, E.S., 1993. A Corpus-Based Study of Logical Connectors in EFL Students' Writing: Directions for Future Research. In *Studies in Lexis: Proceedings of a Seminar on Lexis Organized by the Language Centre of the HKUST, Hong Kong (Language Centre, HKUST, Hong Kong, 1993), 6–7 July 1992*. pp. 215–246.

Sherry, S., 2010. Censorship in Translation in the Soviet Union: The Manipulative Rewriting of Howard Fast's Novel the Passion of Sacco and Vanzetti. *Slavonica*, 16(1), pp. 1–14.

Sinclair, J., 1999. A Way With Common Words. In H. Hasselgård and S. Oksefjell, eds. *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam: Rodopi, pp. 157–179.

7 English–Greek language contact through translation

(Pseudo-)Clefts

7.0 Introduction

The main objective of this case study is to explore the extent to which translation might act as a site of language contact and encourage linguistic developments in the target language. The focus is on English–Greek translations of popular science articles, and in particular on cleft and pseudo-cleft constructions. The aim is to examine how these linguistic features might have developed in the Greek genre of popular science as a result of contact with English popular science translations through time. The investigation of the complex relationship between translation and language change can benefit from an approach where data from different points in time, and types of texts (translated and non-translated) are combined (data triangulation), based on the VVA typology. At the same time, different aspects of a linguistic phenomenon can be examined using different methods (method triangulation). The chapter begins by introducing the research context of popular science and why this genre is suitable for the investigation of translation as a language contact phenomenon. Since this case study presents a diachronic analysis, the time frame selected, as well as the specific points in time, are determined. It, then, presents the reasons for choosing to focus on cleft and pseudo-cleft constructions, and offers a detailed examination of how these are realised in English and Greek. This detailed examination will help us reach a clear definition of what constitutes an instance of cleft or pseudo-cleft construction. Syntactic constructions, such as those examined in this case study, are more difficult to identify using corpus-based methods compared to simple lexical units, such as the connectives examined in the previous case study. A clear definition of the feature under investigation allows for better defined search parameters and, ultimately, more reliable results. The way in which the definitions of cleft and pseudo-cleft constructions are translated into search parameters for corpus analysis is discussed in Chapter 8.

7.1 Rationale

7.1.1 Popular science

The genre of popular science is particularly appropriate for the investigation of translation as a language contact phenomenon since the influence of English on

the target language (regarding specific linguistic features) is likely to be strong, especially if the target language is less widespread in terms of popular scientific production. This is because popular science received a significant boost from the rise of English as an international language of science and has largely developed in the Anglophone world (Kaplan 2001; Tardy 2004). For instance, *Popular Science* and *Scientific American*, two of the most widely circulating and well-known popular science magazines, were established in the United States in the nineteenth century. Many non-English speaking countries would have access to publications like these through translations of English texts, and in some cases, those translations helped establish and develop the standards of popular science writing in other languages, such as Chinese (Liao 2010; 2011). Popular science writing in other languages can, therefore, be influenced by English, similarly to the way in which French academic writing has been shaped by translations of classic Latin texts (Lusignan 1986). This is also the case for the Greek genre of popular science which developed into its present form in the twenty-first century. In particular, a popular science 'boom' occurred in Greece in 2002–2003 when a number of translations from English popular science magazines, namely *Popular Science*, *Scientific American* and *New Scientist*, started to circulate widely. The peak of popular science publications in Greece is the years 2005–2007, during which a total of seven different magazines and newspaper sections featuring both translated and non-translated popular science articles were available; of these, two were established during that period.

The present case study investigates how translation might encourage linguistic developments in the target language through time and, therefore, it is necessary to define the time period that needs to be investigated. Generally speaking, the time span selected for diachronic studies should be large enough "to allow for significant changes but small enough to rule out the possibility of reversals and retrograde movements: we might say from a minimum of half generation to a maximum of two" (Labov 1981, p. 177). Although this claim was made about phonetic change, it can be used as a guide to most diachronic studies. Typically, a generation, that is 20 years, is considered an adequate time span for any language change to occur, although it has been argued that different types of change require a different time span of investigation (Mair 2009). Regarding popular science publications in Greece, the most recent point in time that can be studied, that is for which there is adequate material available, are the years 2009–2010. Thus, the two points in time that are selected for this case study are the years 1990–1991 and 2009–2010, covering a total time span of 20 years. Although a longer period, perhaps two generations or more, would be more appropriate for the study of syntactic change, the genre of Greek popular science publications, particularly as far as translations are concerned, is relatively new. A time span of 20 years is thus the largest that can be studied at this point based on the availability of data. However, when dealing with diachronic development where the aim is to examine how a specific factor or event might have had an impact on the language, it is also useful to include an additional point in time, i.e. the year(s)

when this factor or event started being relevant. Regarding popular science in Greece, this additional point is the years 2003–2004. Thus, a third point is added to the diachronic corpus for these years.

7.1.2 Linguistic features

Two linguistic features that are good candidates for the investigation of linguistic change in the genre of popular science are cleft and pseudo-cleft constructions.[1] One of the key features of popular science texts is the presentation of new information to readers regarding scientific and technological advancements. Popular science texts are likely to exploit linguistic means to change perspective and attract readers' attention. Two such linguistic means available to English are cleft and pseudo-cleft constructions, which allow for emphasis to be placed on specific constituents of the sentence. For example, in the case of cleft constructions, the emphasis is typically placed on the theme (see Firbas 1992 for a Functional Sentence Perspective analysis of language and theme/rheme), as in the sentence, *It is oxygen that the brain needs.* Here, the new information is *oxygen*, which is foregrounded. Similarly, in the case of pseudo-cleft constructions, emphasis is typically placed on the rheme, which represents new information, as in the sentence, *What the brain needs is oxygen*. Here, the new information is again *oxygen*, but this time it is placed in sentence final position (see also Section 7.2).

In languages with rich morphological systems and a free word order, such as Greek, emphasis can be placed on particular constituents with greater ease by simply manipulating word order (De Beaugrande and Dressler 1981) than in languages with a relatively fixed word order, such as English. For instance, when the object of the construction needs to be emphasised, Greek tends to manipulate word order and simply move the object to initial position. Thus, a Subject–Verb–Object sentence such as Ο εγκέφαλος θέλει οξυγόνο (The brain needs oxygen) can be easily transformed into an Object–Verb–Subject construction such as Οξυγόνο θέλει ο εγκέφαλος (Literally: Oxygen needs the brain). Such a construction closely corresponds to what is characterised as a cleft construction in English, since *oxygen* is foregrounded. It is for this reason that cleft and pseudo-cleft constructions are not considered to form part of the Greek syntax, and they are not mentioned in standard grammars of Greek (e.g. Holton et al. 1997; Klairis and Babiniotis 2005; Triantafyllidis 1999; Tsopanakis 1994). However, the flexible word order of Greek means that constructions that correspond to English cleft and pseudo-cleft can be replicated. Karanasios (2008) and Apostolou-Panara (1999) argue that Greek has taken over cleft and pseudo-cleft constructions from English (and possibly French), where they are extensively used because of the fixed word order of these languages. They argue that this process has resulted in a new structural pattern being introduced in Greek, which has led to the introduction of cleft and pseudo-cleft constructions. Nonetheless, it seems that even though these constructions are possible in Greek, they are, according to Sifianou (2006), unnatural and infrequent, suggesting that they verge on ungrammaticality.

7.1.3 Research aims

The present case study aims to explore whether and how cleft and pseudo-cleft constructions are used in Greek in a genre that is likely to encourage the use of such constructions. It also examines the role that translations from English might have played in the way in which these constructions have developed in Greek in the context of popular science. Thus, it aims to address the following research questions:

1) What changes in the frequency and patterning of cleft and pseudo-cleft constructions can be observed in non-translated Greek popular science articles over a 20-year period (1990–2010)?
2) To what extent can any observed changes be attributed to influence from English?
3) To what extent can these changes be linked to translated Greek popular science articles?

The first research question aims at capturing any diachronic developments in the genre of Greek popular science regarding the frequency and patterning of cleft and pseudo-cleft constructions. On the one hand, previous studies (Apostolou-Panara 1999; Karanasios 2008) argue that these constructions have become more frequent in Greek as a result of contact with English. On the other hand, the different possibilities available in English and Greek for the syntactic realisation of both cleft and pseudo-cleft constructions (see Section 7.2) suggest that it is possible that changes might be observed in the way these constructions are realised in Greek as a result of contact with English. If changes are observed in relation to either the frequency or the patterning of cleft and pseudo-cleft constructions, then the differences between Greek and English non-translated articles need to be examined to confirm whether any of the observed changes might be attributed to the frequency and patterning of cleft and pseudo-cleft constructions in English texts (Question 2). If there is evidence to suggest that these changes can be traced back to English, then the role of translations is addressed. The approach adopted in this case study is an integrated one, and the research questions have been designed in a stepped way so that the answer to one question points towards the question that needs to be examined next. For instance, if changes are observed only in the frequency of pseudo-cleft constructions, only their frequency will be examined in relation to translation. As a result of this integrated approach, corpus data triangulation is sequential, with one stage of analysis pointing to the next. Apart from corpus data triangulation, corpus method triangulation is also employed, which in this case combines quantitative methods for the examination of frequencies and qualitative methods for the examination of patterning.

Ultimately, this case study contributes to the understanding of the extent to which translation might act as a site of language contact, and encourage linguistic developments in the target language. Although, a number of studies use a

combination of corpora to provide evidence that translation and linguistic developments in the target language are potentially related (e.g. Baumgarten and Özçetin 2008; Becher et al. 2009; Bisiada 2013; Gellerstam 1986; House 2003; House 2006; Kranich et al. 2012; McLaughlin 2011; Musacchio 2005), such combinations tend to be ad hoc, without a clear rationale of how or why they might be combined, as well as the benefits that this combination might offer (see Chapter 2). This case study aims to address this gap by following a more structured approach to the combination of corpus data and methods, thus forming the basis on which future studies on this, and similar topics, can rest.

7.2 Cleft and pseudo-cleft constructions in English and Greek

Cleft and pseudo-cleft constructions are two linguistic features that share some characteristics, which allow them to be studied in conjunction. The most important of these is that they share the same basic principle, that is, they are formed "by dividing a more elementary clause into two parts"(Huddleston and Pullum 2002, p. 1414), each with its own verb phrase, and with one part featuring a foregrounded element. Perhaps as a result of their similarities, some disagreement exists regarding their categorisation. Some scholars treat them as variants of the same category and distinguish between *it-clefts* and *wh-clefts* in English (e.g. Declerck 1984; Prince 1978), while others argue that they form separate categories (e.g. Collins 1991; Higgins 1979; Gundel 1977; Quirk et al. 1985). According to the latter, pseudo-cleft constructions differ from cleft constructions in that they can be "completely accountable in terms of the category of main and subordinate clause" (Quirk et al. 1985, p. 1387). In other words, the identification of a main and a subordinate clause is more easily achieved with pseudo-cleft constructions than with cleft constructions, since they resemble to a greater extent constructions that feature subordinate relative clauses. In this case study, Collins'(1991) approach is adopted, according to which the categories of cleft and pseudo-cleft constructions are examined separately. Collins also identifies a variety of structural possibilities for pseudo-cleft constructions, which are pertinent to this case study (Figure 7.1).

Cleft and pseudo-cleft constructions in English are a field of study that has received considerable, detailed attention, especially as far as the semantic and pragmatic properties of the constructions are concerned (Collins 1991; Gundel 1977; Prince 1978; Van Riemsdijk and Williams 1986). Such constructions have always been considered an important means of placing emphasis on different constituents of the sentence, since English word order does not allow for free movement of constituents and, thus, form part of English syntax (Biber et al. 1999; Huddleston and Pullum 2002; Quirk et al. 1985). Corpus-based studies of English cleft and pseudo-cleft constructions (Collins 1991) indicate that although such constructions are not very frequent in English, they do constitute regular patterns of English. Cleft constructions occur with a frequency of 5.7 per 10,000 words in written language, whereas pseudo-cleft constructions are slightly less frequent, with a frequency of 4.1 per 10,000 words. Overall, both

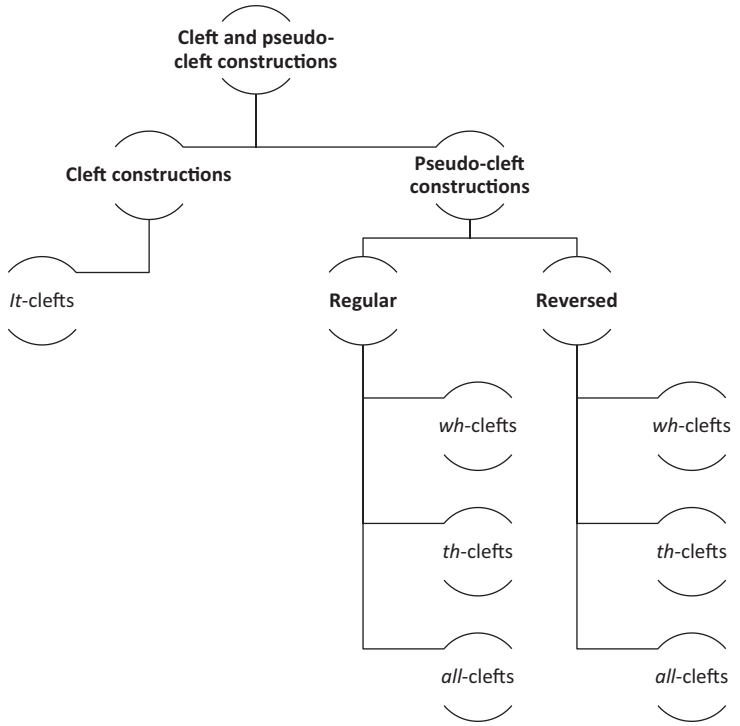

Figure 7.1 Categorisation of cleft and pseudo-cleft constructions

constructions occur with a frequency of 9.8 per 10,000 words in English both spoken and written discourse. These proportions are rather small, which makes such constructions marked—in most cases a basic Subject–Verb–Object construction with an unmarked theme will be adopted. However, these constructions play an important role in the language as there are few other means of placing emphasis on different constituents of a sentence.

Unlike English, cleft and pseudo-cleft constructions in Greek constitute a field of study that has received limited attention, and no detailed study of the frequency of cleft and pseudo-cleft constructions has ever been conducted in Greek. This might be explained by the fact that such constructions, although possible in the language, do not seem to occur as frequently as they do in other languages such as English. In what is possibly the first account of Greek cleft and pseudo-cleft constructions, Veloudis (1979) discusses some of the general properties of these constructions and employs examples from Greek, albeit he treats cleft and pseudo-cleft constructions as possible realisations allowed by Greek syntax, rather than typical constructions that characterise the syntactic patterning of the language. Linguistic interest in these constructions only emerged in the late 1990s

(Alexiadou 1999; Alexiadou and Giannakidou 1998; Iatridou and Varlokosta 1998), but these studies deal with the general language-independent properties, either semantic or pragmatic, of cleft and pseudo-cleft constructions, rather than discuss the specific syntactic properties of such constructions in Greek.[2] In that sense, they do not differ considerably from Veloudis' (1979) first account. Nevertheless, they reflect an increased interest in cleft and pseudo-cleft constructions, which might be explained by the increased contact between Greek and English in more recent years, the latter being a language where such constructions are more typical (Apostolou-Panara 1999; Karanasios 2008). The only account regarding the frequency of these constructions in Greek can be found in Sifianou, who observes that cleft and pseudo-cleft constructions in Greek "are more marked and, therefore, less frequent [than in English] because thematisation is achieved simply by fronting clause elements" (2006, p. 163). She also argues that cleft constructions sound "rather unnatural" in Greek, whereas pseudo-cleft constructions sound "very unnatural" and are "rather infrequent" (Sifianou 2006, p. 163). Taking into consideration such observations, as well as the fact that there are many possibilities in Greek for placing emphasis on different constituents of a sentence, mainly simply by manipulating word order, it is expected that the frequency of cleft and pseudo-cleft constructions is relatively low in Greek, and generally lower than in English.

7.2.1 Cleft constructions

A cleft construction in English consists of a copular verb, typically the verb *is*, a noun phrase that forms the foregrounded element, followed by the conjunction *that* and the rest of the sentence:

(a)	*It is*	*magnesium*	*that*	*the scientists are looking for.*
	copular verb	noun phrase	that	rest of the sentence

The copular verb may take any form of the verb *be*, or appear with a modal verb. Thus, sentence (b) is a possible variant of sentence (a). The foregrounded element can also be a subordinate clause or a prepositional phrase (Quirk et al. 1985), as in sentence (c).

(b)	*It was not*	*magnesium*	*that*	*the scientists are looking for.*
	copular verb	noun phrase	that	rest of the sentence
(c)	*It was*	*because he was ill*	*that*	*he didn't go to work.*
	copular verb	subordinate clause	that	rest of the sentence

Apart from *that*, relative pronouns such as *which*, *who*, *whom*, and the relativisers *when* and *where* can also be used. It is also possible for the relative pronoun to be omitted (Biber et al. 1999; Van Riemsdijk and Williams 1986). The sentences presented in Table 7.1 are all valid cleft constructions in English, although some are more marked than others.

Table 7.1 Possible cleft constructions in English

Pronoun/Relativiser	*Cleft Construction*
which	It was the final experiment which provided the solution.
who(m)	It was John who killed him.
whose	It's Mary whose ring I lost.
when	It was last year in Denmark when she got married.
where	It is in the lab where the experiments take place.
ø	It was John I gave it to.

Corpus studies of the London-Lund Corpus and the Lancaster-Oslo/Bergen Corpus, consisting of half a million and a million words respectively, have shown that cleft constructions featuring *that* are considerably more frequent (representing 64.2 per cent) (Collins 1991) than all other structural possibilities. Constructions involving *who* and the zero relative pronoun are also fairly common, representing 12.1 per cent and 14.4 per cent respectively (Collins 1991). However, the corpora analysed by Collins included both spoken and written language. It is possible that when written language is examined separately, more prototypical instances of the construction might be favoured and the zero pronoun construction is likely to be avoided, since, as noted by Declerk, such constructions are most typically found in "familiar English" (1981, p. 141) and are, in the strict sense of the term, ungrammatical. In written language, Quirk et al. argue that even *whom* and *which* are "only marginally possible" (1985, p. 1387). Since instances of the zero relative pronoun are not typical of written language, and they are also difficult to capture using corpus-based methods of analysis, these are excluded from the present case study.

Similar to English, cleft constructions in Greek consist of a copular verb, which is typically the verb είναι (is), a noun phrase, which is the foregrounded element, and the rest of the sentence, which is normally introduced with the conjunction που (that):

(d)	Είναι	το μέγεθος	που	κάνει τον εγκέφαλο εξυπνότερο.
	(It is	size	that	makes the brain smarter.)
	copular verb	noun phrase	που	rest of the sentence

The copular verb can take any form of the verb *be* in Greek, which results in over ten different verb types when the rich morphology of the language is taken into account. According to Karanasios (2008), the foregrounded element can be a prepositional phrase or even an embedded clause if it is causal, final or temporal, but not if it is conditional or adversative. Finally, the rest of the sentence can be introduced with the relative pronoun ο οποίος (who/which) or its feminine and neuter counterparts (η οποία and το οποίο respectively) (Karanasios 2008). For example, sentence (d) can be rewritten as sentence (e), although the latter might be considered more marked.

(e)	Είναι	το μέγεθος	το οποίο	κάνει τον εγκέφαλο εξυπνότερο.
	(It is	size	which	makes the brain smarter.)
	copular verb	noun phrase	relative pronoun	rest of the sentence

A slightly different syntactic realisation is also possible for Greek cleft constructions; this is a construction where the copular verb and the noun phrase occur in reversed positions, with the cleft sentence beginning with the noun phrase, as in sentence (f). This syntactic patterning is not possible in English and is a reflection of the flexible word order of Greek.

(f)	Το μέγεθος	είναι	που	κάνει τον εγκέφαλο εξυπνότερο.
	(Size	is	what	makes the brain smarter.)
	NP	copular verb	που	rest of the sentence

Despite the detailed definition of cleft constructions provided in this section, it must be noted that it is not always easy to distinguish between cleft constructions and constructions that employ a relative clause. The problem arises from the fact that *that* in English and που (that) in Greek can also be used to introduce relative clauses that specify the preceding noun phrase and the overall structure of the two types of construction is similar (Quirk et al. 1985). Declerk (1981) provides an extensive list of the criteria that can be used to distinguish between cleft and relative constructions. He proposes ten different criteria, not all of which are, however, relevant for the purposes this case study, and thus only the two most pertinent criteria are discussed here. Although these criteria have been developed with English in mind, they can be also applied to Greek.

The first point of difference between cleft and relative constructions involves the *it* element. Declerk argues that in anaphoric sentences *it* has "a deictic or anaphoric reference, . . . while the [*it*] of a cleft sentence is a formal subject that has no referent" (1981, p. 139). To identify the function of *it*, it is necessary to consult the immediate linguistic context surrounding the sentence, and in particular the preceding clause. Secondly, cleft constructions can typically alternate with pseudo-clefts, while relative clauses cannot. For example, the sentence *It is magnesium that the scientists are looking for* can be turned into *What the scientists are looking for is magnesium*. By contrast, a sentence such as *It is a mystery that has puzzled scientists*, although it can be turned into *What has puzzled scientists is a mystery*, conveys a rather different semantic meaning that is not related to the cleft construction. The difficulties associated with identifying cleft constructions suggest that apart from formal criteria, a close reading of the surrounding linguistic context is necessary.

7.2.2 Pseudo-cleft constructions

Pseudo-cleft constructions can be divided into three main types based on what introduces the foregrounded element: *wh*-clefts, *th*-clefts, and *all*-clefts.

Additionally, based on the position at which the foregrounded element appears they can be classified into regular and reversed (Figure 7.1). In regular pseudo-cleft constructions, the noun phrase that constitutes the foregrounded element appears at sentence final position, while in reversed pseudo-cleft constructions, it appears at sentence initial position.

The most typical pseudo-cleft constructions are *wh*-clefts. In English, basic *wh*-cleft constructions are usually introduced with *what*, followed by a clause, a copular verb, (which may include any form of the verb *be*), and a noun phrase that constitutes the foregrounded element:

(g)	*What*	*we are looking for*	*is*	*water.*
	what	clause	copular verb	noun phrase

Collins (1991) identifies *who*, *where*, *when*, *why* and *how* as possible candidates for introducing pseudo-cleft constructions. However, Quirk et al. (1985) question the wide usage of pseudo-clefts with items other than *what*. In particular, they argue that "[c]lauses with *who, when*, and *where* are sometimes acceptable . . . [and] [c]clauses introduced by *whose*, *why*, and *how* do not easily enter into the pseudo-cleft construction at all" (Quirk et al. 1985, p. 1388). Corpus studies, on the other hand, have shown that although *what* is the most frequent type of *wh* element, *why* has a frequency of 12.1 per cent and *how* of 3.5 per cent out of all pseudo-cleft constructions found in the London-Lund and Lancaster-Oslo/Bergen corpora (Collins 1991). Thus, even though *what* is most likely to be used in written language, all *wh* elements need to be considered in any definition of pseudo-cleft constructions in English.

Greek pseudo-cleft constructions which could be mapped onto *wh*-clefts typically consist of the demonstrative pronoun αυτός (this one) or its feminine (αυτή) and neuter counterpart (αυτό), followed by the conjunction που (that), a clause, a copular verb, which is typically the verb είναι (is), and the foregrounded element, which is usually a noun phrase. Sentence (h) is a typical Greek pseudo-cleft construction.

(h)	Αυτό	που	μας λείπει	είναι	η εμπειρία.
	(The one	that	we lack	is	experience.)
	demonstrative pronoun	που	clause	είναι	noun phrase

Apart from αυτός, αυτή, αυτό, the place of the demonstrative pronoun can be occupied by εκείνος, εκείνη, εκείνο (that one),[3] as in sentence (i). The difference between the two demonstrative pronouns does not always lie in issues of proximity, but is more closely related to discourse preferences.

(i)	Εκείνος	που	με βοήθησε	ήταν	ο αδερφός μου.
	(The one	who	helped me	was	my brother.)
	demonstrative pronoun	που	clause	είναι	noun phrase

Pseudo-cleft constructions might also be introduced by a lexical head. Collins (1991) argues that such constructions, that is *th*-clefts, are quite common in English and account for approximately 25 per cent of all pseudo-clefts analysed in the London-Lund corpus and the Lancaster-Oslo/Bergen Corpus. Similarly, other possibilities for the initial position of a pseudo-cleft in Greek are noun phrases like το πράγμα (the thing) and το μέρος (the place) (Veloudis 1979). Table 7.2 offers examples of all the possible realisations of *th*-clefts in English and Greek.

The final type of pseudo-cleft constructions identified by Collins (1991) is *all*-clefts. *All*-clefts in English are constructions that include an *all* element in initial position, while the rest of the construction resembles a pseudo-cleft:

(j) *All the brain needs is oxygen.*

Similarly, in Greek, some scholars also identify ό,τι (all) as a possible candidate for sentence initial position (Iatridou and Varlokosta 1998; Giannakidou 2000):

(k) Ό,τι ήπιε ήταν νερό.
(All she drank was water.)

All of the above types of pseudo-cleft constructions, that is *wh*-clefts, *th*-clefts and *all*-clefts, can have reversed forms. For instance, reversed *wh*-clefts in English consist of the noun phrase, which is the foregrounded element, followed by the copular verb (typically *is*), *what*, and the rest of the sentence. In Greek, such constructions consist of the noun phrase, which is the foregrounded element,

Table 7.2 th-clefts in English and Greek

Lexical head	*English*	*Greek*
thing(s)—το πράγμα	The thing he gave her was a ring.	Το πράγμα που με εντυπωσίασε περισσότερο ήταν το πείσμα της. (The thing that impressed me the most was her stubbornness.)
reason—ο λόγος	The reason I cancelled was because I was ill.	Ο λόγος που ήρθε ήταν για να παρακολουθήσει το φεστιβάλ. (The reason s/he came was to watch attend the festival.)
way—ο τρόπος	The way you should go is through the city centre.	Ο τρόπος που μου μίλησε ήταν άσχημος. (The way he spoke to me was bad.)
place—το μέρος	The place we visited is the coffee shop.	Το μέρος που συναντήθηκαν ήταν το κυλικείο. (The place they met was the cafeteria.)
one(s)	The one I like is John.	–
time	The time I hate the most is the dry season.	–

followed by the copular verb (typically είναι (is)), the demonstrative pronoun (e.g. αυτός (this one)), the conjunction που (that) and the rest of the sentence. Table 7.3 offers examples of some of the possible realisations of reversed pseudo-cleft constructions in English and Greek. In Greek, it is also possible to reverse the order of the noun phrase and the copular verb, which reflects the flexible word order of the language. However, it must be stressed that these possible realisations vary considerably regarding markedness, especially in Greek. Thus, although these constructions might be syntactically allowed, some of these are likely to be considered unnatural, as suggest by Sifianou (2006).

According to corpus findings (Collins 1991), certain types of pseudo-cleft constructions appear more frequently in reversed forms, namely, English pseudo-cleft constructions that feature *why*, *how*, *where* and *when* can only be found in reversed form. Moreover, approximately 55 per cent of all pseudo-cleft constructions in spoken and written English tend to appear in reversed form. No similar information is available regarding the frequency of reversed constructions in Greek.

Finally, there are some problems when it comes to the distinction between Greek cleft and pseudo-cleft constructions. One such instance is cleft constructions in which the foregrounded element occupies initial position. Sentence (l), which is a cleft construction, can easily be turned into a reversed pseudo-cleft with the addition of a demonstrative pronoun, as in sentence (m).

(l) Το μέγεθος είναι που κάνει τον εγκέφαλο εξυπνότερο.
(Size is what makes the brain smarter.)
(m) Το μέγεθος είναι **αυτό** που κάνει τον εγκέφαλο εξυπνότερο.
(Size is **the one** that makes the brain smarter.)

Although in both cases the back translation is a pseudo-cleft construction in English, in Greek, only the latter can be considered a pseudo-cleft construction according to the formal criteria presented above. Whether the former sentence is an elliptical pseudo-cleft construction, that is one where the demonstrative pronoun is omitted, or a different realisation of a cleft construction allowed by the flexible word order of Greek is debatable. What seems to be important for the interpretation of a construction as pseudo-cleft is the presence of a pronoun. For this case study, only formal syntactic criteria are employed, and thus sentences

Table 7.3 Reverse pseudo-cleft constructions in English and Greek

	English	*Greek*	
***wh*-cleft**	A car is what they're looking for.	Ένα αυτοκίνητο είναι αυτό που ψάχνουν.	Είναι ένα αυτοκίνητο αυτό που ψάχνουν.
***th*-cleft**	A car is the thing they're looking for.	Ένα αυτοκίνητο είναι το πράγμα που ψάχνουν.	Είναι ένα αυτοκίνητο το πράγμα που ψάχνουν.
***All*-cleft**	A car is all they're looking for.	Ένα αυτοκίνητο είναι ό,τι ψάχνουν.	Είναι ένα αυτοκίνητο ό,τι ψάχνουν.

such as (l) are considered cleft constructions unless there is strong evidence to indicate otherwise.

7.3 Conclusion

This chapter introduced the research context for the investigation of translation as a language contact phenomenon focusing specifically on the Greek translation of English popular science articles. For the diachronic analysis, the years 1990–1991 and 2009–2010 are selected because they capture a 20-year period including the closest cut-off point based on the availability of data. The years 2003–2004 are selected as an additional point in time to be analysed because translations of popular science articles started to circulate more widely in Greece during those years. Cleft and pseudo-cleft constructions are identified as suitable candidates for the examination of the extent to which new syntactic patterns have been introduced in the target language, and their properties in both English and Greek have been examined in detail. Although the relationship between translation and language contact has already been addressed in previous studies, results from these studies tend to be inconsistent, partly because, despite the fact that corpus-based methods are used, the types of corpora and the methods used to analyse these vary considerably, which limits the replicability and comparability of these studies. The present case study aims to address these shortcomings by offering a detailed account of how both data and method triangulation are achieved and their potential advantages. This, together with the results from the corpus analysis, is presented in the next chapter.

Notes

1 In a previous study (Malamatidou, 2016), in which, however, I do not explicitly make use of corpus triangulation, I found that influence from English has affected the use of the passive voice in Greek popular science articles.
2 For example, Iatridou and Varlokosta argue that a difference between cleft and pseudo-cleft constructions is that pseudo-cleft constructions involve "an ordinary copular sentence with a free relative [clause] in one of the copular positions and a phrase in the other copular position modifying that free relative" (1998, p. 3) and provide examples from English, Greek and German.
3 Although αυτός and εκείνος are translated in English as 'this one' and 'that one' respectively, and the resulting equivalent English construction is introduced with *the one*, the pseudo-cleft construction is classified as a *wh*-cleft, since no lexical head is employed.

References

Alexiadou, A., 1999. Specificational Pseudo-Clefts as Lists. In K. Shahin, S. Blake, and E.-S. Kim, eds. *The Proceedings of the Seventeenth West Coast Conference on Formal Linguistics*. Stanford: Stanford Publications, pp. 1–16.

Alexiadou, A. and Giannakidou, A., 1998. Specificational Pseudo-Clefts and the Semantics of Lists. *ZAS Working Papers on Linguistics*, 10, pp. 1–21.

Apostolou-Panara, A.-M., 1999. Γλωσσική Μεταρρύθμιση και Εξωτερικός Δανεισμός: Η περίπτωση της Νέας Ελληνικής [Language Reform and External Borrowing:

The Case of Modern Greek]. In *Συνέδριο για την Ελληνική Γλώσσα 1976–1996: Είκοσι χρόνια από την καθιέρωση της Νεοελληνικής (Δημοτικής) ως επίσημης γλώσσας [Conference on the Modern Greek Language 1976–1996: Twenty Years From the Establishment of Modern Greek (Demotic) as an Official l]*. Athens: Tsiveriotis, pp. 333–341.

Baumgarten, N. and Özçetin, D., 2008. Linguistic Variation Through Language Contact in Translation. In P. Siemund and N. Kintana, eds. *Language Contact and Contact Language*. Amsterdam and Philadelphia: John Benjamins, pp. 293–316.

Becher, V., House, J. and Kranich, S., 2009. Convergence and Divergence of Communicative Norms Through Language Contact in Translation. In K. Braunmüller and J. House, eds. *Convergence and Divergence in Language Contact Situations*. Amsterdam and Philadelphia: John Benjamins, pp. 125–152.

Biber, D. et al., 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Bisiada, M., 2013. Changing Conventions in German Causal Clause Complexes: A Diachronic Corpus Study of Translated and Non-Translated Business Articles. *Languages in Contrast*, 13(1), pp. 1–27.

Collins, P., 1991. *Cleft and Pseudo-Cleft Constructions in English*. Abingdon and New York: Routledge.

De Beaugrande, R. and Dressler, W., 1981. *Introduction to Text Linguistics*. London: Longman.

Declerck, R., 1984. The Pragmatics of It-Clefts and Wh-Clefts. *Lingua*, 64(4), pp. 251–289.

Declerk, R., 1981. Pseudo-Modifiers. *Lingua*, 54, pp. 135–163.

Firbas, J., 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.

Gellerstam, M., 1986. Translationese in Swedish Novels Translated From English. In L. Wollin and H. Lindquist, eds. *Translation Studies in Scandinavia: Proceedings From the Scandinavian Symposium on Translation Theory (SSOTT) II, Lund 14–15 June, 1985*. Lund: CWK Gleerup, pp. 88–95.

Giannakidou, A., 2000. Crosslinguistic Semantics and the Study of Greek. *Journal of Greek Linguistics*, 1, pp. 223–261.

Gundel, J., 1977. Where Do Cleft Sentences Come From? *Language*, 53(3), pp. 543–559.

Higgins, F.R. (Francis R.), 1979. *The Pseudo-Cleft Construction in English*. New York: Garland Pub.

Holton, D., Mackridge, P. and Philippaki-Warburton, I., 1997. *Greek: Comprehensive Grammar of the Modern Language*. Abingdon and New York: Routledge.

House, J., 2003. English as a Lingua Franca: A Threat to Multilingualism? *Journal of Sociolinguistics*, 7(4), pp. 556–578.

House, J., 2006. Covert Translation, Language Contact, Variation and Change. *SYNAPS: A Journal of Professional Communication*, 19, pp. 25–47.

Huddleston, R. and Pullum, G., 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Iatridou, S. and Varlokosta, S., 1998. Pseudoclefts Crosslinguistically. *Natural Language Semantics*, 6, pp. 3–28.

Kaplan, R., 2001. English: The Accidental Language of Science? In U. Ammon, ed. *The Dominance of English as a Language of Science: Effects on Other Languages and Language Communities*. Berlin: Mouton de Gruyter, pp. 3–26.

Karanasios, G., 2008. Δίπτυχη Πρόταση [Cleft Sentence]. Available at: www.greek-language.gr.

Klairis, C. and Babiniotis, G., 2005. *Γραμματική της Νέας Ελληνικής: Δομολειτουργική-επικοινωνιακή [Modern Greek Grammar: Structural-Functional-Communicative]* 3rd ed. Athens: Ellinika Grammata.

Kranich, S., House, J. and Becher, V., 2012. Changing Conventions in English-German Translation of Popular Scientific Texts. In K. Braunmüller, C. Gabriel, and B. Hänel-Faulhaber, eds. *Multilingual Individuals and Multilingual Societies*. Amsterdam and Philadelphia: John Benjamins, pp. 315–334.

Labov, W., 1981. What Can Be Learned About Change in Progress From Synchronic Description? In D. Sankoff and H. Cedergren, eds. *Variation Omnibus*. Edmonton: Linguistic Research, pp. 177–200.

Liao, M.-H., 2010. Influence of Translations on Non-Translations: Popular Science as a New Genre in Taiwan. *Perspectives: Studies in Translatology*, 18(2), pp. 127–140.

Liao, M.-H., 2011. Interaction in the Genre of Popular Science. *The Translator*, 17(2), pp. 349–368.

Lusignan, S., 1986. *Parler Vulgairement: Les Intellectuels et la Langue Française aux XIIIe et XIVe Siècles [Talking Vulgarly: Intellectuals and the French Language in the Thirteenth and Fourteenth Centuries]*. Paris and Montreal: Vrin and ler Presses de l'Université de Montréal.

Mair, C., 2009. Corpora and the Study of Recent Change in Language. In A. Lüdeling and M. Kytö, eds. *Corpus Linguistics: An International Handbook*. Berlin: Mouton de Gruyter, pp. 1109–1125.

Malamatidou, S. (2016). Understanding Translation as a Site of Language Contact: The Potential of the Code-Copying Framework as a Descriptive Mechanism in Translation Studies. *Target*, 28(3), 399–423.

McLaughlin, M., 2011. *Syntactic Borrowing in Contemporary French: A Linguistic Analysis of News Translation*. Oxford: Legenda.

Musacchio, M.T., 2005. The Influence of English on Italian: The Case of Translations of Economic Articles. In G. Anerman and M. Rogers, eds. *In and Out of English: For Better, for Worse?* Clevedon: Multilingual Matters, pp. 71–96.

Prince, E., 1978. A Comparison of Wh-Clefts and It-Clefts in Discourse. *Language*, 54, pp. 883–906.

Quirk, R. et al., 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Sifianou, M., 2006. *Discourse Analysis: An Introduction*. Athens: Hillside Press.

Tardy, C., 2004. The Role of English in Scientific Communication: Lingua Franca or Tyrannosaurus Rex? *Journal of English for Academic Purposes*, 3(3), pp. 247–269.

Triantafyllidis, M., 1999. *Νεοελληνική Γραμματική [Modern Greek Grammar]*. Athens: OEDV.

Tsopanakis, A., 1994. *Νεοελληνική Γραμματική [Modern Greek Grammar]*. Thessaloniki: Kyriakidis.

Van Riemsdijk, H. and Williams, E., 1986. *Introduction to the Theory of Grammar*. Cambridge: MIT Press.

Veloudis, I., 1979. Παρατηρήσεις Πάνω στα Γενικά Χαρακτηριστικά των Διπτύχων και Ψευδο-Διπτύχων Προτάσεων [Remarks on the General Characteristics of Cleft and Pseudo-Cleft Sentences]. In *Research Yearbook of the School of Philosophy*. Thessaloniki: University of Thessaloniki, pp. 41–56.

8 Triangulating language contact through translation

8.0 Introduction

The final chapter of the book explains how corpus data and method triangulation are achieved regarding the examination of the relation between translation and language change in the case of cleft and pseudo-cleft constructions in Greek popular science articles. The way in which the definitions of cleft and pseudo-cleft constructions provided in the previous chapter are translated into detailed search parameters for corpus analysis is presented here, as well as a stepped approach to the investigation of translation as a language contact phenomenon. The largest part of the chapter is dedicated to reporting the results obtained from the examination of the corpus. The focus is on the frequency and patterning of cleft and pseudo-cleft constructions and results from the examination of one subcorpus point towards what needs to be examined in the other subcorpora.

8.1 Corpus design

Although popular science can take many different forms (e.g. as an Internet-based product, part of a newspaper, separate magazine, TV or radio programme, book), in order to delimit the category of popular science, and allow for a more in-depth analysis of data, this case study focuses only on articles appearing in print news media (magazines and sections of newspapers). Print news media allow for material to be captured relatively easily from a range of different sources within the given time frame of the diachronic study. Specifically, Greek and English non-translated articles are available for all three points in time analysed here (i.e. 1990–1991, 2003–2004, 2009–2010). Translated Greek popular science articles are only available for 2003–2004 and 2009–2010 since translations from English were very scarce in 1990–1991. Based on the research questions established in the previous chapter, the corpus analysed for the purposes of this case study consists of the eight components presented in Table 8.1:

Table 8.1 Corpus components

	1990–1991	*2003–2004*	*2009–2010*
non-translated Greek popular science articles	X	X	X
non-translated English popular science articles	X	X	X
translated Greek popular science articles		X	X

The corpus is altogether rather small, as it consists of 800,000 words. This is due to a number of factors, one main practical consideration being that the availability of Greek data, especially in machine-readable form, is limited, and articles tend to be short, ranging from 500 to 2,500 words. Moreover, material is taken from a specific genre in a specific language pair, making this a specialised corpus focusing on popular science in English and Greek, and specialised corpora tend to be smaller in size (Gavioli and Zanettin 1997). Finally, the linguistic features examined require close readings of the linguistic context in which they are produced to distinguish, for example, between cleft constructions and constructions that employ a relative clause (see Chapter 7). It is generally argued that corpora employed in morphosyntactic studies, which tend to be analysed manually, can be justifiably smaller than those employed in lexical studies (Biber et al. 1998; Givón 1995; Hundt and Leech 2012). Taking these factors into account, 100,000 is considered an adequate size for each of the corpus components presented above.[1] The corpus components are combined in different ways to create four smaller subcorpora which are analysed separately to address the research questions identified in the previous chapter. Table 8.2 offers a detailed description of corpus compilation.

Corpus data triangulation is achieved by combining texts in different corpus configurations based on the VVA typology (see Chapter 3). Firstly, by combining values from the *corpus type variable*: comparable, parallel and reference. Although three corpus types are combined, these are, in fact, four, since the comparable corpora belong to different categories (non-translated texts in two languages, and translated and non-translated texts in the same language). Secondly, corpus data triangulation is achieved by combining values from the *time variable*: synchronic and diachronic. Similar to this, triangulation occurs in terms of *time attributes*, that is, by examining different points in time. It has been argued that longitudinal studies are not considered as triangulated since they are interested in examining how the phenomenon changes over time, rather than focusing on commonalities across time (Kimchi et al. 1991), which is considered to be the aim of triangulation by some scholars (Denzin 1970; Jick 1979; Webb et al. 1981). However, convergence should not be considered the only, or indeed the

Table 8.2 Corpus compilation

	Subcorpus	*Components*
A.	a reference, monolingual (Greek), diachronic (1990–2010) subcorpus of non-translated texts (from the genre of popular science)	• non-translated Greek popular science articles published in 1990–1991 • non-translated Greek popular science articles published in 2003–2004 • non-translated Greek popular science articles published in 2009–2010
B.	a comparable, bilingual (Greek-English), diachronic (1990–2010) subcorpus of non-translated texts (from the genre of popular science)	• non-translated Greek popular science articles published in 1990–1991 • non-translated Greek popular science articles published in 2003–2004 • non-translated Greek popular science articles published in 2009–2010 • non-translated English popular science articles published in 1990–1991 • non-translated English popular science articles published in 2003–2004 • non-translated English popular science articles published in 2009–2010
C.	a comparable, monolingual (Greek), diachronic (2003–2010) subcorpus of non-translated and translated texts (from the genre of popular science)	• non-translated Greek popular science articles published in 2003–2004 • non-translated Greek popular science articles published in 2009–2010 • translated Greek popular science articles published in 2003–2004 • translated Greek popular science articles published in 2009–2010
D.	a parallel, bilingual (English-Greek), diachronic (2003–2010) corpus (from the genre of popular science)	• non-translated English popular science articles published in 2003–2004 • non-translated English popular science articles published in 2009–2010 • translated Greek popular science articles published in 2003–2004 • translated Greek popular science articles published in 2009–2010

most important, aim of triangulation. Divergent results can help us acquire a more complete picture of the phenomenon under investigation (Redfern and Norman 1994; Fielding and Fielding 1986) (see Chapter 2). Thus, diachronic studies are considered as triangulated according to the methodological framework presented in this book. Additionally, the time frame selected for analysis presents a unique characteristic, since it involves three points in time, instead of two as is typical in diachronic studies. Triangulation also occurs through a combination of values from the *text variable* (non-translated and translated) and by combining attributes from the *languages variable*, as far as the non-translated

texts are concerned (English and Greek). Although these two combinations are not particularly new in corpus-based translation studies, it is the first time that these are achieved following a comprehensive and flexible corpus typology. These two combinations are, however, linked to corpus type and are considered secondary, albeit important.

Corpus data triangulation is sequential since the results obtained from one corpus point towards which corpus needs to be examined next. Subcorpus A is examined to establish whether there is any development in non-translated Greek texts over time (Research Question 1). It is worth repeating here that the focus is on both the frequency and patterning of cleft and pseudo-cleft constructions. If any development is observed, subcorpus B, which is divided into three synchronic components each focusing on different years (i.e. 1990–1991, 2003–2004, 2009–2010), is examined for two reasons both of which are related to Research Question 2. Firstly, to exclude the possibility that similar diachronic developments are also observed in English, which would suggest a more general tendency of languages towards these constructions. Secondly, to identify any differences in the use of cleft and pseudo-cleft constructions in Greek and English, especially in earlier years, which might explain the diachronic development in Greek. If differences are observed, subcorpora C and D are used to investigate the role of translation in encouraging these linguistic developments (Research Question 3). Subcorpus C is analysed to identify whether Greek non-translated articles share any similarities with Greek translated articles from English. Subcorpus D is examined to reveal any potential similarities between the source and target texts. Both corpora are divided into synchronic components each focusing on different years, which offer information on whether these similarities become stronger over time. This corpus design demonstrates how diachronic corpora can consist of synchronic components, allowing for complex linguistic analyses on different levels.

Naturally, there is some overlap between corpora, which is strong indication that corpus data triangulation occurs in an integrated manner. Thus, a component of one corpus might be a component of another corpus. For instance, all texts included in subcorpus A are also part of subcorpus B, while the English texts of subcorpus B are also included in subcorpus D. Based on this corpus design, comparisons are made both vertically (across different years) and horizontally (across different texts) (Figure 8.1).

Diversity is achieved by including articles from the majority of the different popular science publications that were available during 1990–1991, 2003–2004 and 2009–2010. Although diversity is not particularly difficult to achieve, the fact that corpora are triangulated means that issues of comparability also need to be taken into consideration. To facilitate comparability, particularly for diachronic corpora, publications that had a continuous circulation between 1990 and 2010 are selected where possible[2] (Table 8.3). However, this has not always been possible. For instance, the Greek editions of *Popular Science* and *Scientific American* ceased publication in 2007 and 2008

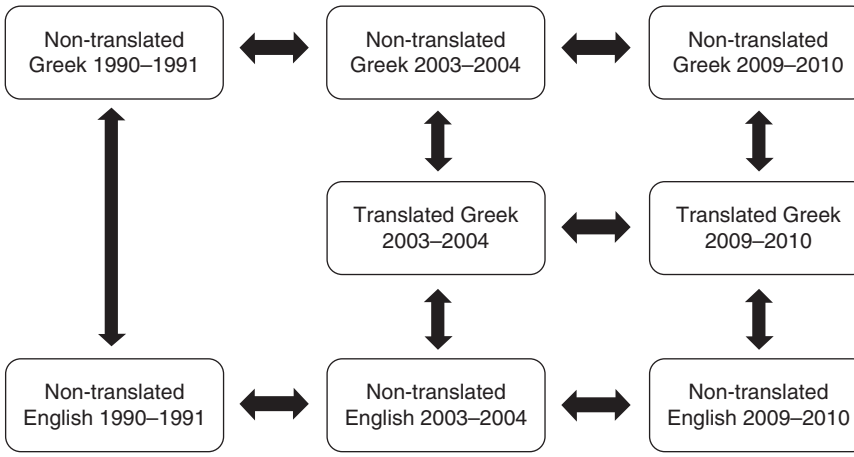

Figure 8.1 Schematic representation of corpus data triangulation

Table 8.3 Popular science publications in the corpus

Publication	*1990–1991*	*2003–2004*	*2009–2010*
Non-translated Greek articles			
Periscopio tis Epistimis	X	X	X
To Vima	X		
Ta Nea	X		
Vima Science		X	X
Translated Greek articles			
Vima Science		X	X
Popular Science		X	
Scientific American		X	
English articles			
New Scientist	X	X	X
Popular Science	X	X	
Scientific American	X	X	

respectively.[3]This affected the creation of the corpus component consisting of Greek translated articles published in 2009–2010, which, therefore, consists of texts taken from only one publication, i.e. *Vima Science*. This is a good example of how practical considerations, regarding the context of the study, affect corpus design. The full list of texts included in each corpus component can be found in Appendix 2.

Practical considerations further affect corpus design, resulting in material in the corpus not always being representative of the years they aim to capture. Limited availability of translated material meant that not enough translated articles could be identified for the period 2009–2010. This inevitably meant a smaller corpus component and less reliable findings. To overcome this problem and maintain the balance in the corpus, the time span of the corpus component consisting of translated articles published in 2009–2010 was extended to also capture the year 2008. However, for purposes of consistency, the component is referred to as covering the years 2009–2010, since most articles included in it were published during these years. Specifically, only 20 of the 100 articles in the parallel, bilingual (English–Greek), synchronic (2009–2010) subcorpus were published in 2008, which corresponds to approximately 10 per cent of its total size.

It has not been possible to acquire data on the circulation of popular science magazines in Greece, although it was possible to get such data for newspapers. Thus, issues of currency and influentialness could not be taken into account, and the corpora created are not proportional in that sense. Instead, an attempt is made to acquire equal amounts of texts from different sources, and approximately 33,000 words are taken from each publication for each corpus component. The only exceptions to this are the corpus component consisting of Greek translated articles published in 2009–2010 and the component of their English source texts, which are larger for reasons explained above. Also, the corpus components consisting of articles from *Popular Science* published in 2003–2004 in both Greek and English are smaller than 33,000 words by approximately 35 per cent due to practical consideration related to the accessibility of Greek data. As a result, the corpus components of translated Greek popular science articles published in 2003–2004 and of non-translated English articles published in 2003–2004 are slightly smaller than 100,000 words by approximately 10 per cent. This inconsistency does not affect the comparability of corpus data, as both raw and normalised frequencies are reported.

8.2 Corpus analysis

The software selected for the analysis of the corpus data is Wordsmith Tools 7.0. In particular, the Word List tool is used to generate statistics on the size of the each corpus component, that is number of sentences, while the Concord Tool is used to generate concordances based on elaborate word search queries that

include context words as well as context horizons. The search parameters are informed by the discussion of the definition and properties of cleft and pseudo-cleft constructions in English and Greek in Chapter 7, and separate search queries are used for cleft and pseudo-cleft constructions. Search queries are further refined with sample corpus analyses. For cleft constructions, the word search query includes all the possible types of the verb *be* in both languages, including instances of modality, to identify all occurrences of cleft constructions. The context words consist of the conjunctions used to introduce the second part of the cleft construction (i.e. that, wh*, που, οποί*). Sample analyses of the corpora indicated that a context horizon of R7 (seven words to the right), with the limit set at sentence break point, is adequate for the second part of a cleft construction to be captured in both languages. Wildcards capture instances of contracted negation and, thus, help identify the maximum number of cleft constructions with the minimum number of search items. For instance, *it is** captures both instances of *it is*, and *it isn't*.

Contrary to clefts, pseudo-cleft constructions present a variety of structural possibilities in both languages. This case study focuses on *wh*-clefts, *th*-clefts and *all*-clefts, as well as their reversed possibilities, all of which are considered to be subcategories of pseudo-cleft constructions (Collins 1991), and a separate search is conducted for each type of pseudo-cleft construction. As in the analysis of cleft constructions, complex word search queries, context words and context horizons are used to identify all relevant instances. The search query includes all the possible types of the verb *be* in both languages and wildcards are incorporated to capture the maximum number of pseudo-cleft constructions with the minimum number of search items. Context words include the different linguistic items used to introduce the foregrounded element. Sample analyses of the corpora indicated that a context horizon of L10 (ten words to the left) with the limit set at sentence breakpoint is adequate for regular pseudo-cleft constructions, whereas R4 (four words to the right) is considered adequate for their reversed form. Table 8.4 summarises the search parameters for cleft and pseudo-cleft constructions in English and Greek. Once concordances are generated, they are carefully examined and manually refined, since the linguistic context needs to be consulted to decide whether an identified construction is a (pseudo-)cleft.

Once all instances of cleft and pseudo-cleft constructions are identified, corpus analysis can begin, and corpus method triangulation can be implemented. The methodology employed in this case study involves a combination of within-method and between-method triangulation. Within-method triangulation is achieved by combining descriptive and inferential statistics, while between-method triangulation is achieved by combining quantitative and qualitative analyses of the concordances. Quantitative methods are employed to examine changes in the frequency of cleft and pseudo-cleft constructions, while qualitative methods are employed to capture changes in their patterning. Within-method triangulation is sequential, as raw frequencies are required for

Table 8.4 Search parameters for cleft and pseudo-cleft constructions in English and Greek

	English	*Greek*
***It*-cleft**		
Search words:	it is*/it's/it was*/it will be/ it will not be/it won't be/ it can* be/it could*/ it must*/it might*/ it should*/it would*/it may	είμ*/είσ*/είν*/ ήμ*/ήσ*/ήτ*
Context words:	that/wh*	που/οποί*
Context horizon:	R7	R7
***Wh*-cleft**		
Search words:	is*/was*/were*/will be/will not be/ won't be/can* be/could*/must*/ might*/may*/should* be/would* be	είμ*/είσ*/είν*/ ήμ*/ήσ*/ήτ*
Context words:	wh*/how	αυτ*/εκείν*/οποί*
Context horizon:	L10/R4	L10/R4
***Th*-cleft**		
Search words:	is*/was*/were*/will be/will not be/ won't be/can* be/could*/must*/ might*/may*/should* be/would* be	είμ*/είσ*/είν*/ ήμ*/ήσ*/ήτ*
Context words:	thing*/one*/place*/time*/reason*/ way*	πράγμα*/μέρ*/ λόγ*/τρόπ*
Context horizon:	L10/R4	L10/R4
***All*-cleft**		
Search words:	is*/was*/were*/will be/will not be/ won't be/can* be/could*/must*/ might*/may*/should* be/would* be	είμ*/είσ*/είν*/ ήμ*/ήσ*/ήτ*
Context words:	all	ό,τι
Context horizon:	L10/R4	L10/R4

statistical analyses, while between-method triangulation is simultaneous, as one type of analysis is not a prerequisite for the other, although the two are inevitably linked.

Regarding descriptive statistics, the concordance lines are counted and, for the raw frequency, the results are compared to the total number of sentences in the corpus component under investigation. Although there is a tendency in corpus linguistics to use the number of words in a corpus as the basis against which comparisons are made, this is not a strong basis when it comes to morphosyntactic features. For meaningful comparisons, it is important that both the feature under investigation and the basis against which it is compared belong to

the same category (e.g. lexis or morphosyntax) (see also Chapter 4). Thus, for this case study, the number of sentences is considered a more meaningful concept than the number of words in each corpus component. Apart from raw frequencies, normalised frequencies are also calculated. Because the focus of this study is on syntactic features which are expected to be quite low, 10,000 sentences have been selected as the common base for the comparison. Numbers are rounded to the closest whole number to avoid meaningless references to proportions of sentences.

Regarding inferential statistics, statistical significance is calculated using the log-likelihood test, while effect size is measured using the Effect Size for Log Likelihood (ELL). Statistical significance is important because it helps ascertain whether the differences observed in the frequencies are not the result of coincidence, while effect size measures how large the differences are. The null hypothesis (H_0) is that the differences observed are due to chance. The alternate hypothesis (H_1) is that the differences observed can be attributed to a factor other than chance (e.g. the influence from English through translation). ELL is used instead of percentage difference (%DIFF) because in some cases more than two corpus components need to be compared and %DIFF can only be used for pairwise comparisons. Bayes Factor (BIC) is also calculated to measure the probability that a difference in frequency is due to chance. For diachronic comparisons, percentage change is also calculated. It is important to note here that ELL and BIC are used only if results from the log-likelihood test indicate that differences are statistically significant, and thus there is indication that the H_0 can be refuted.

Overall, two types of descriptive statistics (raw frequency and normalised frequency) and three types of inferential statistics (statistical significance, effect size, and Bayes Factor) are used. These constitute the combination of quantitative methods in the analysis of the corpora, thus achieving within-method triangulation. Since the proportions of cleft and pseudo-cleft constructions are expected to be quite low, it might not be possible to observe clear patterns regarding a change in their frequency. For this reason, it is necessary to also focus on their patterning for which a qualitative analysis of concordance lines is required, where the syntactic properties of each construction are closely analysed. Thus, both within and between-method triangulation is employed not so much to confirm results obtained using different methods, but rather to complement results, paying particular attention to instances where contradictions might occur. The different methods are treated like pieces of a puzzle, which, when put together, offer a (more) complete picture of the phenomenon under investigation.

In summary, this case study makes use of multiple triangulation and employs both corpus data and corpus method triangulation, while corpus method triangulation occurs through both within and between-method triangulation. However, methodological triangulation does not occur independently of data triangulation. These are combined in an integrated manner, and it would have been impossible to reach an answer to the set of research questions identified in

the previous chapter without benefiting from both of these. Specifically, corpus data triangulation establishes the different stages of analysis based on the different corpora available, while corpus method triangulation achieves the actual comparison among these. The three stages of analysis are the following:

1 The investigation of whether and to what extent cleft and pseudo-cleft constructions have changed over time in Greek non-translated popular science articles regarding their frequency and patterning
2 The examination of whether any observed changes might be a result of influence from English
3 The examination of whether translations might have encouraged these changes

During each stage of analysis, corpus findings are analysed using both quantitative and qualitative methods as explained above, and comparisons across corpus components and subcorpora are made. Each stage corresponds to the examination of one of the subcorpora identified earlier. Only for the last stage, two subcorpora are used, namely the comparable, monolingual (Greek), diachronic (2003–2010) subcorpus of translated and non-translated texts (corpus C) and the parallel, bilingual (English–Greek), diachronic (2003–2010) subcorpus (corpus D). Each stage analyses a larger number of words than the previous one and involves more elaborate comparisons—a further indication that the approach is integrated.

8.3 Corpus results

8.3.1 Linguistic development in Greek

The first stage of analysis involves the examination of the reference, monolingual (Greek), diachronic (1990–2010) subcorpus of non-translated popular science articles. The subcorpus is divided into three components based on the years the articles have been published, that is, 1990–1991, 2003–2004 and 2009–2010. Each component is analysed separately, and comparisons are drawn. Regarding frequency, the diachronic analysis reveals that the frequency of cleft and pseudo-cleft constructions in Greek non-translated texts has increased slightly in 20 years, but proportions remain low (Table 8.5).

Overall, the frequency of both cleft and pseudo-cleft constructions has increased in Greek non-translated popular science articles by 93.1 per cent: from 29 per 10,000 sentences in 1990–1991 to 44 in 2003–2004 and 56 in 2009–2010 (Figure 8.2).[4] This increase is, however, not statistically significant ($p > 0.05$). The quantitative data are not conclusive in this case, and although there seems to be a significant increase in the distribution of these constructions diachronically, we cannot exclude the possibility that this is due to chance. This could be related

Table 8.5 Frequency of cleft and pseudo-cleft constructions in non-translated Greek popular science articles

	Clefts		*Pseudo-clefts*		*Total*	
	Raw Frequency	*Normalised Frequency (10,000 sentences)*	*Raw Frequency*	*per 10,000 sentences*	*Raw Frequency*	*Normalised Frequency (10,000 sentences)*
1990–1991 Greek (non-translated)	2/4,194	5	10/4,194	24	**12/4,194**	**29**
2003–2004 Greek (non-translated)	3/4,312	7	16/4,312	37	**19/4,312**	**44**
2009–2010 Greek (non-translated)	4/3,895	10	18/3,895	46	**22/3,895**	**56**

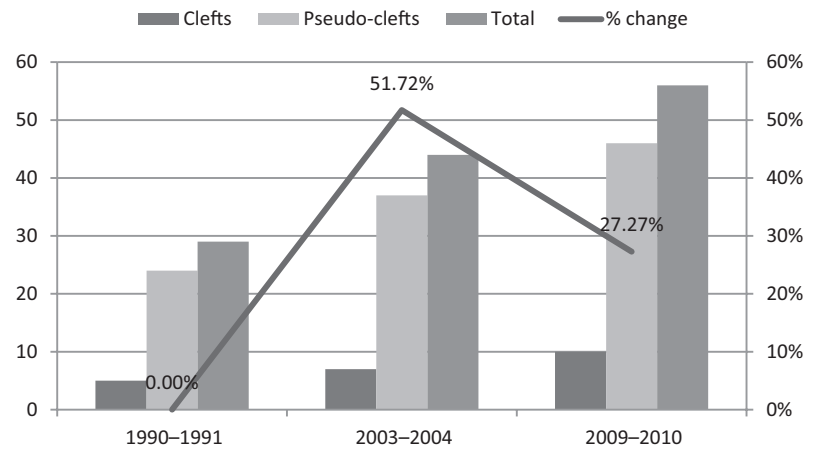

Figure 8.2 Graphic representation of the frequency of cleft and pseudo-cleft constructions in non-translated Greek popular science articles (per 10,000 sentences)

to the small proportion of cleft and pseudo-cleft constructions identified in the corpus, which does not easily lend to analysis using inferential statistics.

Examining cleft and pseudo-cleft constructions separately will help establish whether the different categories have changed diachronically in the same way. The examination of cleft constructions in Greek non-translated popular science articles reveals that their number has increased from 5 per 10,000 sentences in 1990–1991 to 7 in 2003–2004 and 10 in 2009–2010. Overall, the proportion of cleft constructions has doubled (100.0 per cent increase) in the 20-year period analysed here. However, statistical analysis indicates that the diachronic increase is not statistically significant ($p > 0.05$) and, thus, we cannot exclude the possibility that it has arisen by chance. Respectively, the use of pseudo-cleft constructions has increased from 24 per 10,000 sentences in 1990–1991 to 37 in 2003–2004 and 46 in 2009–2010. Overall, their frequency has approximately doubled (91.7 per cent increase) between 1990 and 2010. However, as with cleft constructions, statistical analysis reveals that the increase is not statistically significant ($p > 0.05$).

While there does not seem to be a significant change in the frequency of cleft and pseudo-cleft constructions throughout the years, some changes can be noticed in their syntactic properties. The two cleft constructions identified in Greek non-translated articles published in 1990–1991 have a very similar patterning and place emphasis on the noun phrase not only by foregrounding it through a cleft construction, but also by adding a demonstrative pronoun, as can be seen in Examples 8.1 and 8.2. The use of the demonstrative pronoun places adequate emphasis on the noun phrase without the need for a cleft construction.

Example 8.1

Η διαφορά αυτή ακριβώς είναι που οδηγεί στα σχετικά συμπεράσματα για τη φυσιολογικη ή μη λειτουργεία της μέσης.

(source: *ToVima 1990*)

[near-literal translation]
Exactly this difference is that leads to the relevant conclusions about the normal function of the waist.

Example 8.2

Και αυτή η παράλειψη είναι που θέτει σε πραγματικό κίνδυνο την υγεία της.

(source: *Τα Νεα*1990)

[near-literal translation]
And this omission is that puts her health at real risk.

As a result, the fact that a cleft construction has been chosen here makes these sentences very marked. We can also notice that in both these sentences the copular verb and the foregrounded element occur in reversed positions compared to a typical cleft construction.

In 2003–2004, none of the cleft constructions identified employed a demonstrative pronoun for additional emphasis, as was the case in 1990–1991, while for the first time a cleft construction is employed where the copular verb and the foregrounded element do not occur in reversed positions (Example 8.3). It is also interesting to note that in 2003–2004, cleft constructions start employing negation.

In 2009–2010, a demonstrative pronoun is employed in two out of four instances of cleft constructions identified. The copular verb and the foregrounded element occur in reversed positions in half of the instances, while negation appears only once. This variation suggests that although the frequency of cleft constructions is low, they are now better incorporated into the language, entering a wider range of syntactic constructions. In particular, the fact that regular positions are selected for the copular verb and the foregrounded element might suggest an influence from English, where this is the only possibility.

Examining the patterning of pseudo-cleft constructions qualitatively, we can observe that, as with clefts constructions, there is considerably more variation in

Example 8.3

Δεν είναι το χρήμα που κάνει τον κόσμο να γυρίζει, αλλά οι ενοχές.

(source: *Focus 2003*)

[near-literal translation]
It is not money that makes the world go round, but guilt.

recent years. For instance, in 1990–1991, only *wh*-clefts are employed, with the majority of these appearing in reversed form, as in Example 8.4.

In 2003–2004, the majority of *wh*-clefts appear in regular form, indicating a tendency towards what might be considered a more prototypical pattern in English. More importantly, during these years, *th*-clefts, both regular and reversed, are used for the first time. However, only one lexical head is employed, namely *ο λόγος* (the reason) (Example 8.5).

A similar distribution of the different pseudo-cleft constructions is observed in texts produced in 2009–2010, with both *wh*-cleft and *th*-clefts, in regular and

Example 8.4

Το ρομάντσο είναι αυτό που σκέπτονται οι περισσότεροι όταν αναφέρεται ο συνδυασμός διάθεσης και φαγητού.

(source: *Ta Nea*1991)

[near-literal translation]
Romance is what most people think about when referring to the combination of mood and food.

Example 8.5

Ο λόγος που ξεχωρίζουν είναι αυτή καθαυτή η μορφή τους.

(source: *Focus*2003)

[near-literal translation]
The reason they stand out is precisely their form.

reversed forms. In all instances of *th*-clefts, the lexical head *ο λόγος* (the reason) is employed, as in 2003–2004. As with cleft constructions, pseudo-cleft constructions enter a wider range of syntactic constructions from 2003–2004 onwards.

8.3.2 Influence of English

The next stage of analysis involves the examination of the comparable, bilingual (Greek–English), diachronic (1990–2010) subcorpus of non-translated popular science articles. Specifically, the results from the analysis of the English texts are compared to the results from the Greek texts obtained during the previous stage of analysis. Two comparisons are made, one diachronic and one synchronic. The diachronic comparison aims at establishing whether the patterning of cleft and pseudo-cleft constructions have changed in English texts, which would suggest that some reason, other than language contact, might be behind the development in both languages. The synchronic comparison, with a separate comparison for each point in time (1990–1991, 2003–2004, 2009–2010), aims at establishing whether there are any linguistic differences between English and Greek texts in the use of cleft and pseudo-cleft constructions, which might justify the changes in the patterning of cleft and pseudo-cleft constructions in Greek non-translated texts in more recent years. If no differences are observed across the years, or if differences are observed across all time periods, this will suggest that the changes in the patterning observed in Greek articles cannot be related to English. If differences are mostly observed in articles produced in 1990–1991, compared to 2003–2004 and 2009–2010, this can be considered an indication that Greek articles have come closer to patterns found in English popular science.

Even though no change in the frequency of these constructions has been observed during the previous stage of analysis, frequency is also examined in relation to English texts to establish whether there is any potential for change. In other words, if differences in the frequency are observed between English and Greek texts, this would suggest that through the years, and if contact with English becomes more intense, a change in Greek popular science articles is possible. Alternatively, if no difference is observed, this will suggest that English popular science articles do not have the potential of introducing change in respective Greek texts regarding the frequency of cleft and pseudo-cleft constructions. Based on the corpus findings, a slight change in the frequency of cleft and pseudo-cleft constructions is identified in English, while these constructions are employed more frequently in English than in Greek texts during all three points time (Table 8.6).

Diachronically, English cleft and pseudo-cleft constructions have slightly increased from 59 per 10,000 sentences in 1990–1991 to 70 in 2003–2004 and 76 in 2009–2010. Overall, their frequency in English texts has increased by 28.9 per cent between 1990 and 2010.[5] However, any difference in the distribution of cleft and pseudo-cleft constructions through the years is not statistically significant ($p > 0.05$), which suggests that the observed increase is most probably a result of the inherent variability in the corpus data.

Table 8.6 Frequency of cleft and pseudo-cleft constructions in non-translated English and Greek popular science articles

	Clefts		*Pseudo-clefts*		*Total*	
	Raw Frequency	*Normalised Frequency (10,000 sentences)*	*Raw Frequency*	*Normalised Frequency (10,000 sentences)*	*Raw Frequency*	*Normalised Frequency (10,000 sentences)*
1990–1991 English	9/4,762	19	19/4,762	40	**28/4,762**	**59**
1990–1991 Greek (non-translated)	2/4,194	5	10/4,194	24	**12/4,194**	**29**
2003–2004 English	11/4,949	22	24/4,949	48	**35/4,949**	**70**
2003–2004 Greek (non-translated)	3/4,312	7	16/4,312	37	**19/4,312**	**44**
2009–2010 English	16/4,944	32	22/4,944	44	**38/4,944**	**76**
2009–2010 Greek (non-translated)	4/3,895	10	18/3,895	46	**22/3,895**	**56**

Synchronically, cleft and pseudo-cleft constructions are employed more frequently in English than in Greek popular science articles in 1990–1991 (59 vs 29 per 10,000 sentences), 2003–2004 (70 vs 44) and 2009–2010 (76 vs 56). The diachronic analysis reveals that the difference between English and Greek decreases through time by 55.6 per cent (Figure 8.3). The difference is statistically significant only in 1990–1991 ($p < 0.05$), but there is positive evidence in favour of the H_0 (BIC = –4.39), and the effect size is very small (ELL = 0.00018). English texts seem to have some limited potential of introducing change, but given that no statistically significant development in the frequency of cleft and pseudo-cleft constructions has been observed in the monolingual diachronic Greek subcorpus, we cannot convincingly argue that Greek texts have moved closer to English popular science articles in more recent years. The patterning of cleft and pseudo-cleft constructions is a more promising area for investigation of potential influence from English. For this reason, only the results of the qualitative analysis are reported in the remainder of this section.

As has already been noted, a very specific type of cleft constructions is identified in Greek non-translated articles published in 1990–1991, while a much wider range of constructions is observed in more recent years. It is worth reminding here that cleft constructions cannot appear in reversed form in English. The fact that Greek non-translated articles start making use of regular cleft constructions only in 2003–2004, that is after translations from English start circulating widely, is an indication that this change in their patterning might have been influenced by English. To confirm this, it is necessary to examine the Greek translations of the English articles (see section 8.3.3). A qualitative analysis of cleft constructions

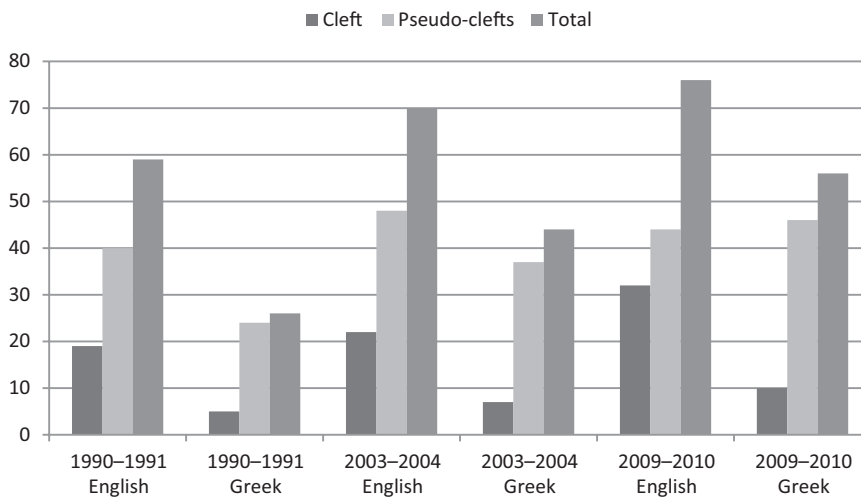

Figure 8.3 Graphic representation of the frequency of cleft and pseudo-cleft constructions in non-translated English and Greek popular science articles (per 10,000 sentences)

also reveals that these appear in negative constructions in English articles published during all three points in time. Once again, the fact that negative cleft constructions only appear in non-translated Greek popular science articles from 2003–2004 onwards is an indication that preferred English patterns might have influenced their use.

As with cleft constructions, the quantitative analysis suggests that there are some substantial differences in the patterning of pseudo-cleft constructions between English and Greek, especially in 1990–1991. Specifically, Greek employs only *wh*-clefts in 1990–1991, typically in reversed form, whereas English makes use of *wh*-clefts (Example 8.6), *th*-clefts (example 8.7), in both regular and reversed forms, and *all*-clefts in regular form (Example 8.8). *th*-clefts are used with a range of lexical heads, such as *the one, the reason, the thing* and *the time*. Compared to Greek non-translated articles, English articles show considerably higher syntactic variation regarding pseudo-cleft constructions across all points in time. In subsequent years, Greek starts employing *th*-clefts, but not *all*-clefts, and pseudo-clefts start appearing in regular form, as seen during the previous stage of analysis.

The existence or non-existence of certain types of constructions in the different corpus components is highly revealing and can inform us about the development

Example 8.6

What came floating out were thousands of tiny, beautifully preserved fossil flowers.

(source: *New Scientist* 2008)

Example 8.7

The reason Greenway and Durrant-Whyte's drones don't need maps is that they make their own.

(source: *New Scientist 2003*)

Example 8.8

All he can do is insist on the following three points.

(source: *Scientific American*1990)

of cleft and pseudo-cleft constructions diachronically. In this case, three important conclusions can be reached. Firstly, no significant diachronic developments regarding patterning are observed in English popular science articles. Secondly, Greek popular science articles started employing a much wider range of cleft and pseudo-cleft constructions in more recent years compared to 1990–1991 possibly imitating English patterns. Thirdly, different cleft and pseudo-cleft types seem to be first observed in English texts and only later, if at all, in Greek texts, indicating that they might originate in English (e.g. *all*-clefts). Although we cannot argue with certainty that the frequency of cleft and pseudo-cleft constructions in non-translated Greek articles has been affected by English, we can hypothesise that the fact that Greek cleft constructions are employed in a wider range of syntactic contexts in more recent years could be a result of an English influence, where similar variety is observed.

8.3.3 The role of translation

The last stage of analysis involves the examination of two subcorpora: the comparable, monolingual (Greek), diachronic (2003–2010) subcorpus of non-translated and translated popular science articles and the parallel, bilingual (English–Greek), diachronic (2003–2010) subcorpus of popular science articles. Where relevant, the texts are first compared for each period separately (synchronic analysis) and then results from different time periods are contrasted (diachronic analysis). Even though during the previous stages of analysis there was insufficient evidence of a change in the frequency of cleft and pseudo-cleft constructions in Greek, which could be traced back to English, the analysis of the frequency in relation to translated texts is reported here mainly for the purpose of completeness and to offer additional confirmation that no significant change is observed. At the same time, to keep the discussion concise, frequency data are reported only about the total frequency of both cleft and pseudo-cleft constructions. More attention is given to the qualitative analysis of the patterning of these constructions.

Regarding the overall frequency, the corpus analysis reveals that similar proportions of cleft and pseudo-cleft constructions are observed in translated and non-translated Greek popular science articles, while translated texts employ these constructions less frequently than their English source texts (Table 8.7).

Overall, translated texts employ cleft and pseudo-cleft constructions more frequently than non-translated texts in 2003–2004 (60 vs 44 per 10,000 sentences) and 2009–2010 (66 vs 56 per 10,000 sentences). Diachronically, the difference between the two types of texts decreases by 46.8 per cent (Figure 8.4). However, these differences are not statistically significant ($p > 0.05$) and, thus, we cannot exclude that they have arisen by chance. Additionally, the total frequency with which cleft and pseudo-cleft constructions are employed in translated articles over the years has increased by only 10 per cent, which has not been found to be statistically significant ($p > 0.05$). Based on these quantitative measures, cleft and pseudo-cleft constructions are used with the same low frequency in both translated and non-translated texts across the years.

Table 8.7 Frequency of cleft and pseudo-cleft constructions in non-translated and translated Greek popular science articles, and the English source texts

	Clefts		*Pseudo-clefts*		*Total*	
	Raw Frequency	*Normalised Frequency (10,000 sentences)*	*Raw Frequency*	*Normalised Frequency (10,000 sentences)*	*Raw Frequency*	*Normalised Frequency (10,000 sentences)*
2003–2004 Greek (non-translated)	3/4,312	7	16/4,312	37	**19/4,312**	**44**
2003–2004 English	11/4,949	22	24/4,949	48	**35/4,949**	**70**
2003–2004 Greek (translated)	3/3,864	8	20/3,864	52	**23/3,864**	**60**
2009–2010 Greek (non-translated)	4/3,895	10	18/3,895	46	**22/3,895**	**56**
2009–2010 English	16/4,944	32	22/4,944	44	**38/4,944**	**76**
2009–2010 Greek (translated)	1/4,417	2	28/4,417	63	**29/4,417**	**66**

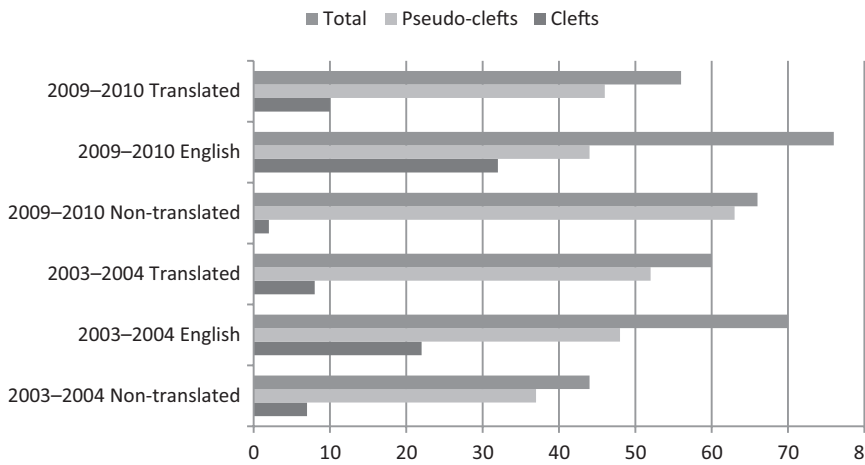

Figure 8.4 Graphic representation of the frequency cleft and pseudo-cleft constructions in non-translated and translated Greek popular science articles and their English source texts (per 10,000 sentences)

Translated texts employ cleft and pseudo-cleft constructions less frequently than their English source texts in 2003–2004 (35 vs 70 per 10,000 sentences) and 2009–2010 (66 vs 76 per 10,000 sentences). The difference between the two types of texts remains fairly stable through the years (Figure 8.4). The differences observed are not statistically significant ($p > 0.05$). These findings confirm that, regarding the frequency of cleft and pseudo-cleft constructions, no diachronic development has been introduced in Greek popular science articles through translation from English.

The very similar results obtained from the comparison between translated and non-translated texts and translated and source texts suggest that translated popular science articles do not differ significantly from either their source texts or respective non-translated texts. However, there is one aspect of the frequency of cleft and pseudo-cleft constructions that offers valuable insight into how translated texts employ these constructions. It might be expected that, given that there is no significant difference in the frequency of cleft and pseudo-clefts constructions between translated and source texts, their frequency is identical or at least very similar in the two types of texts. This is clearly not the case for cleft constructions, while regarding pseudo-cleft greater similarities are observed in 2003–2004 than 2009–2010.

A closer look at concordance lines reveals that English cleft constructions are not always translated as cleft constructions in Greek popular science articles. In fact, there is only one cleft construction in the English source texts that has been translated as a cleft in Greek. Conversely, there are also cases where a cleft construction is used in the translated text but does not occur in the English source text, as in Example 8.9.

Example 8.9

As anyone who has struggled with their weight knows, losing pounds in the short term is relatively easy—the real challenge is keeping them off.

(source: *New Scientist* 2004)

Όπως πολύ καλά γνωρίζουν τα υπέρβαρα άτομα που έχουν ακολουθήσει πολλές δίαιτες κατά τη διάρκεια της ζωής τους, δεν είναι τόσο η απώλεια των περιττών κιλών που αποτελεί δύσκολη διαδικασία, όσο η διατήρηση του βάρους μετά την υποβολή σε δίαιτα.

(source: *Vima Science* 2004)

[back translation]
As overweight people who have followed many diets in their life know very well, it is not so much the loss of excess weight that is difficult, but rather keeping the same weight after submitting to a diet.

Example 8.9 offers a very clear indication of the way in which the translated text is edited to meet the linguistic conventions of the target audience. The target sentence is more elaborate (almost two times longer), expressing the meaning of the source sentence explicitly and using rather formal vocabulary (e.g. *loss of excess weight*, *submitting to a diet*). A very typical structure in Greek that is used to add emphasis, that is δεν είναι τόσο . . . όσο (it is not so much . . . but), is combined with a cleft construction, resulting in additional emphasis. The reason the translator decided to place that much emphasis on this sentence is not clear, but it is obvious that he or she made an effort to communicate the meaning of the original sentence as explicitly as possible.

Examining the patterning of pseudo-cleft constructions more closely, it appears that in most cases, pseudo-clefts in Greek translated texts originate in similar constructions employed in the English source texts, as in Example 8.10. In some cases, cleft constructions in English have been translated as pseudo-cleft constructions in Greek, as in Example 8.11. These examples provide evidence that contrary to cleft constructions, pseudo-cleft constructions found in English are more easily rendered into similar pseudo-cleft constructions in Greek. This is possibly due to the fact that they are also more frequently used in Greek compared to cleft constructions (see Section 8.3.1). However, there are still cases where English pseudo-cleft constructions have not been rendered as such in Greek, and vice versa. Thus, even though no change in the frequency is observed, it appears that pseudo-cleft constructions carry more potential for such change to occur in the future.

In the previous section, it was found that cleft constructions where the foregrounded element follows the copular verb have been identified for the first time

Example 8.10

What came floating out were thousands of tiny, beautifully preserved fossil flowers.

(source: *New Scientist* 2008)

Αυτό που παρέμενε στην επιφάνεια του ηθμού ήταν χιλιάδες μικροσκοπικά εξαιρετικά καλά διατηρημένα απολιθώματα άνθεων.

(source: *Vima Science*2009)

[back translation]
What remained on the surface of the filter were thousands of tiny, well-preserved fossil flowers.

Example 8.11

It's washing that causes the problem.

(source: *New Scientist*2008)

Αυτό που τους δημιουργεί πρόβλημα είναι το πλύσιμο.

(source: *Vima Science* 2009)

[back translation]
What causes them trouble is washing.

in Greek non-translated articles produced during 2003–2004. Similar constructions are found in translated articles produced during the same period, as in Example 8.9, which is repeated here for ease of reference. It is also interesting to note that translated articles published in 2003–2004, also employ cleft constructions with negation.

In translated articles published in 2009–2010, only cleft constructions where the foregrounded element follows the copular verb are employed, and all instances of these also employ negation. The use of a demonstrative pronoun in cleft constructions, which has been identified as a specific type of cleft constructions employed in Greek non-translated articles, does not appear in translated articles produced at either point in time examined here. These observations indicate that

Example 8.9

Όπως πολύ καλά γνωρίζουν τα υπέρβαρα άτομα που έχουν ακολουθήσει πολλές δίαιτες κατά τη διάρκεια της ζωής τους, δεν είναι τόσο η απώλεια των περιττών κιλών που αποτελεί δύσκολη διαδικασία, όσο η διατήρηση του βάρους μετά την υποβολή σε δίαιτα.

(source: *Vima Science* 2004)

[near-literal translation]
As overweight people who have followed many diets in their life know very well, it is not so much the loss of excess weight that is difficult, but rather keeping the same weight after submitting to a diet.

translated texts make use of more prototypical types of cleft constructions, possibly as a result of influence from English where cleft constructions cannot appear with the foregrounded element at sentence-initial position. The use of these constructions might then spread from translated to non-translated texts, resulting in more variation regarding cleft constructions in non-translated Greek popular science articles.

If we examine the concordance lines of pseudo-cleft constructions qualitatively, we observe some differences in the use of pseudo-cleft constructions in Greek translated and non-translated texts, which are more revealing than their similarities. Although there is similar variation in the different types of pseudo-cleft constructions between translated and non-translated texts, with both *wh*-clefts and *th*-clefts employed, translated articles also employ *all*-clefts, albeit only in 2009–2010 (Example 8.12).

Additionally, in translated texts, only *wh*-clefts appear in both regular and reversed form, while *th*-clefts appear only in regular form, and *all*-clefts only in reversed in both 2003–2004 and 2009–2010. A further difference can be found in the use of *th*-clefts, where although non-translated articles use only the lexical head ο λόγος (the reason), translated texts prefer the lexical head το πράγμα (the thing) in both 2003–2004 and 2009–2010.

It is interesting that non-translated and translated texts employ different lexical heads, suggesting that in the case of *th*-clefts, while the general pattern might be borrowed from translated texts, once the construction enters the language it interacts with native elements to serve the linguistic needs of the target language. However, the small proportion of *th*-clefts does not allow for reliable conclusions to be reached. The qualitative analysis suggests that no significant differences are observed in translated texts diachronically, while non-translated texts seem to imitate translated ones and not vice versa. In other words, there is no instance of a pattern found in non-translated texts that is not identified in translated texts

Example 8.12

Μελέτες έχουν δείξει ότι ένα περιβάλλον που επιδεικνύει φροντίδα για την ανατροφή του και η απασχόληση, μαζί με τον γονιό, με παιχνίδια όπως το κρυφτό, οι κύβοι και τα τουβλάκια, τα παιδικά τραγουδάκια ή η ταξινόμηση σχημάτων είναι ό,τι χρειάζεται ένα παιδί για να αυξήσει τον δείκτη ευφυΐας του και να αποκτήσει ενδιαφέρον για τη μάθηση.

(source: *Vima Science* 2009)

[near-literal translation]
Studies have shown that an environment that demonstrates care for its upbringing and the engagement, with the parent, with games such as peek-aboo, building blocks, nursery rhymes or shape sorting is all a child needs to increase its IQ level and acquire interest in learning.

as well, while the reverse has been observed, notably with *all*-clefts. Similarly, no pattern has been observed in translated texts, which has not been observed in the English source texts. For example, *all*-clefts are employed in English texts during all three points in time, they appear in translated texts in 2009–2010, but are not found in non-translated Greek articles.

8.4 Discussion

The aim of this case study has been to examine whether the use (i.e. frequency and patterning) of cleft and pseudo-cleft constructions has changed through time in Greek non-translated popular science articles, and to what extent any observed developments might be attributed to contact with English through translation. A multiple corpus triangulation approach has been adopted, where triangulation occurred through a combination of both different corpora and different methods in their investigation.

Regarding the frequency of cleft and pseudo-cleft constructions, there is not sufficient evidence to suggest that these have become more frequent in non-translated Greek popular science articles as a result of contact with English through translation. There is also no clear indication that their frequency might be different between English and Greek, which would imply a potential future development. Finally, translated Greek popular science articles do not differ significantly from either their English source texts or Greek non-translated popular science articles. However, it is important to note that the very small number of identified patterns, especially for Greek cleft constructions, might also be responsible for the lack of statistical significance in the results. This is a possible indication that, while small corpora might generally be appropriate for the examination

of morphosyntactic features (Biber et al. 1998; Givón 1995; Hundt and Leech 2012), for these specific features a larger corpus than the one employed in this case study might be necessary.

Although Greek non-translated popular science articles do not seem to be influenced by the frequency with which cleft and pseudo-cleft constructions are employed in English, they appear to be influenced by the variation of their syntactic patterning, and there is a development in the range of syntactic realisations of Greek cleft and pseudo-cleft constructions, as the qualitative analysis seems to suggest. In particular, the unique type of cleft construction, with a demonstrative pronoun, is used less frequently after 2003–2004. During those years, regular cleft constructions also start being employed, as well as constructions including negation. These syntactic patterns are also observed in English texts, and in their Greek translations, indicating some influence from English through translation. Regarding pseudo-cleft constructions, the preferred syntactic realisation in non-translated Greek articles in 1990–1991 is reversed *wh*-clefts. Since 2003–2004, *th*-clefts are also used, and regular forms become more frequent. Similar patterns are also observed in English source texts and their Greek translations. Generally, there is no pattern observed in non-translated articles published in 2003–2004 and 2009–2010, which has not been observed in translated articles published during the same periods. In turn, there is no pattern observed in translated articles published in 2003–2004 and 2009–2010, which has not been observed in their English source texts. Thus, based on the qualitative analysis, we can hypothesise that the patterning of cleft and pseudo-cleft constructions in Greek has changed diachronically and that this change has been encouraged to an extent by the contact with English through translation.

While it is not possible to claim with any certainly that translation might play a role in linguistic developments in the target language, at least as far as the use of cleft and pseudo-cleft constructions in Greek popular science articles is concerned, this should not detract the value of the present case study, which showcases how multiple corpus triangulation might be employed and its advantages are. Firstly, by employing corpus data triangulation and cross-examining different corpora of translated and non-translated articles in two different languages, it has been possible to exhaustively examine all possible relations among these, to an equal level of detail. This combination allows for a broad range of comparisons to be made, which investigate the factors that might explain any similarities or differences in the data. Without such a combination, it would not have been possible to provide a coherent answer to the set of research questions. Secondly, corpus method triangulation allowed us to identify patterns in the data, by scrutinising data from different perspectives. Corpus method triangulation has proven particularly useful in this case study since a large number of corpora have been examined and a range of patterns have been identified. To illustrate how each method has been indispensable, we can consider what each has offered, and what conclusions would have been reached, had it not been employed. For example, if only descriptive statistics had been used in this case study, we could have argued that cleft and pseudo-clefts constructions are employed more frequently in Greek in

recent years and that this increase in frequency can be related to the influence of English through translation. However, the situation is not so straightforward, and it is inferential statistics that provided this information. Further, the use of qualitative methods suggests that while significant differences in the frequency of these constructions might not be observed, some changes in their patterning could be attributed to contact with English through translation. This demonstrates that corpus method triangulation can occur when different methods produce contradicting results. As with the previous case study, although the focus is on a specific language pair, genre and linguistic features, the corpus design and methodology can be applied to the study of translation as a language contact phenomenon in a broad range of languages, genres and linguistic features.

8.5 Conclusion

As far as the genre of popular science is concerned, previous arguments (Karanasios 2008; Apostolou-Panara 1999) suggesting that cleft and pseudo-cleft constructions have become more frequent due to contact with English cannot be adequately supported, at least for any observable change between 1990 and 2010. The possible changes that have been observed in the 20-year period examined are related to the patterning of these constructions, rather than to their frequency. It is, however, likely that a change in the frequency of these constructions occurred before 1990 and that, once the frequency has been established, a change took place at a different level. However, direct translations of English popular science articles would not have played any significant role in this case, since they did not exist before 2003, and thus this frequential change must have reached the language of popular science articles through other means. While the aim here is not to confirm or contradict findings of previous studies on the topic, it is clear that the use of corpus triangulation approaches has allowed us to analyse data exhaustively and shed light on new aspects of translation and its relation to language change, which would not have been revealed otherwise. In that sense, the conclusions reached by previous studies are not right or wrong, when compared to those reached by this case study, but simply incomplete.

Notes

1 Greek translated popular science articles are often edited in terms of content before being published, a process that results in target texts being shorter than their source texts. As a result, the corpus components consisting of English articles are fairly larger than 100,000 words. This discrepancy does affect the comparability of corpus data, as both raw and normalised frequencies are reported.
2 For example, in the case of the newspaper *To Vima*, the 1990–1991 one-page section titled *Pros to Avrio* (To the Future) later developed into a much longer section of the newspaper specifically dedicated to popular science issues, titled *Vima Science*, which includes both non-translated and translated articles from *New Scientist*.
3 This is due to the marked decline in the publication of translated popular science articles in Greece, which is particularly noticeable in 2009. The main factor behind this decline is probably the socioeconomic situation of the country during that period. Since translation projects, as well as obtaining the legal rights for translation,

are costly, translations were reduced to a minimum. Ultimately, all Greek editions of Anglophone publications stopped circulating and a considerably smaller number of translated articles were included in *Vima Science*. An additional reason for the decline could be the establishment of the genre of popular science in Greece. Over time and initially with the help of translated publications, popular science became established as a genre in Greece, and a sufficient volume of non-translated articles could be produced, thus reducing the need for translations.

4 For brevity and clarity of argumentation, only normalised frequencies are reported in the discussion. For raw frequencies, readers should consult the detailed tables in each section.

5 It is interesting that contrary to what has been argued by Collins (1991), pseudo-cleft constructions are more frequent than cleft construction in English popular science.

References

Apostolou-Panara, A.-M., 1999. Γλωσσική Μεταρρύθμιση και Εξωτερικός Δανεισμός: Η περίπτωση της Νέας Ελληνικής [Language Reform and External Borrowing: The Case of Modern Greek]. In *Συνέδριο για την Ελληνική Γλώσσα 1976–1996: Είκοσι χρόνια από την καθιέρωση της Νεοελληνικής (Δημοτικής) ως επίσημης γλώσσας [Conference on the Modern Greek Language 1976–1996: Twenty Years From the Establishment of Modern Greek (Demotic) as an Official l]*. Athens: Tsiveriotis, pp. 333–341.

Biber, D., Conrad, S. and Reppen, R., 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Collins, P., 1991. *Cleft and Pseudo-Cleft Constructions in English*. Abingdon and New York: Routledge.

Denzin, N.K., 1970. *The Research Act: A Theoretical Introduction to Sociological Methods* 1st ed. Chicago: Aldine Pub. Co.

Fielding, N.G. and Fielding, J.L., 1986. *Linking Data*. Thousand Oaks, CA: SAGE Publications.

Gavioli, L. and Zanettin, F., 1997. Comparable Corpora and Translation: A Pedagogic Perspective. In G. Aston, L. Gavioli, and F. Zanettin, eds. *Proceedings of Corpus Use and Learning to Translate*. Available at: www.federicozanettin.net.

Givón, T., 1995. *Functionalism and Grammar*. Amsterdam and New York: John Benjamins.

Hundt, M. and Leech, G., 2012. "Small is Beautiful": On the Value of Standard Reference Corpora for Observing Recent Grammatical Change. In T. Nevalainen and E. Closs Taugott, eds. *The Oxford Handbook of the History of English*. Oxford: Oxford University Press, pp. 175–188.

Jick, T.D., 1979. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4), pp. 602–610.

Karanasios, G., 2008. Δίπτυχη Πρόταση [Cleft Sentence]. Available at: www.greek-language.gr.

Kimchi, J., Polivka, B. and Stevenson, J.S., 1991. Triangulation: Operational Definitions. *Nursing Research*, 40(6), pp. 364–366.

Redfern, S.J. and Norman, I.J., 1994. Validity Through Triangulation. *Nurse Researcher*, 2, pp. 41–56.

Webb, E.J. et al., 1981. *Nonreactive Measures in the Social Sciences*. Boston: Houghton Mifflin.

Conclusion

With this section, we conclude the overview of corpus triangulation and highlight, what could be considered, the most significant contributions of this book. If this book has accomplished nothing else, I hope it has given some indication of the considerable research potential of triangulation techniques. Although corpus triangulation will not be able to answer all questions in translation studies, it is hopefully a step towards the advancement of the field of corpus-based translation studies. Therefore, this book should be treated as an introduction to corpus triangulation, and there is certainly much scope for its future development.

Following from the Introduction, when corpus-based methods were first introduced into translation studies, the discipline underwent what is described by Olohan (2004, p. 1) as "early teenage angst, seeking to develop its own corpus-related image while coming to terms with other self-centred preoccupations". On a similar vein, although almost a decade later, Zanettin (2012, p. 206) noted that "corpus-based translations studies are still at an early stage of development, and have not yet risen to fully meet the promises foreshadowed by new methodologies". These statements can be treated as a call to maturity for corpus-based translation studies, to develop its own methods, based on its unique needs and preoccupations, and not simply adopt those handed down by linguistics. Maturity will, undoubtedly, not come overnight; after all the full development of corpus-based methods "represents a long-term investment for the field of Translation Studies" (Tymoczko 1998, p. 658). The corpus triangulation framework might not be the definite sign of maturity for the discipline, but it certainly assists in its coming-of-age, and there are many ways in which this book contributes to this.

Perhaps the most obvious contribution of this book is that it offers a novel and much-needed framework for the analysis of translation-related phenomena, which allows for the combination of different corpora (e.g. comparable, parallel, synchronic, diachronic) and/or different methods of analysis (e.g. quantitative and qualitative). This new methodological framework allows for existing questions in translation studies to be answered more confidently and provides answers to new questions. Naturally, it will not be able to offer answers to all questions, but any answers will be as comprehensive as possible. In other words, following Risjord (2000), the ultimate aim of corpus triangulation is to leave fewer questions unanswered and fewer answers unquestioned. The case studies presented

in this book have shown how corpus triangulation can offer detailed insight into translation-related phenomena, at points challenging existing assumptions.

Another significant contribution is the emphasis the book places on integration as a distinguishing characteristic of triangulation. This helps distinguish between 'real' triangulation and ad hoc combinations of corpus data and methods attempted in translation studies. Integration is also an important research parameter which can increase rigour and reduce the risk of bias. Research in corpus-based translation studies, and translation studies more generally, can also benefit by the importance given to both convergence and divergence, thus challenging existing assumptions that triangulation can only be performed to increase validity (i.e. to establish convergence). Especially the second case study (Chapters 7 and 8) on the relationship between translation and language change, where the results from the quantitative and qualitative analysis pointed to different directions, clearly demonstrates how divergence in the results can increase our understanding of a phenomenon and generate hypotheses.

On a more practical level, the present book offers corpus-based translation studies a new typology for the description of corpora, which is based on the idea of variables, values and attributes (VVA typology). Its significant descriptive potential derives from its flexibility and comprehensiveness, and, while this has been purposefully created to inform the corpus data triangulation framework, it can be easily adopted in any corpus-based project and offers a common and clear terminology to be used across corpus-based translation studies. At the same time, the model of corpus method triangulation developed in this book stresses the importance of employing not only both qualitative and quantitative methods of corpus analysis, but also combining different quantitative methods, not least inferential statistics, which tend to be neglected in translation studies. To demonstrate the potential of quantitative methods, the first case study (Chapters 5 and 6) on the language of translation makes use of only quantitative methods and, yet, reaches insightful conclusions that provide adequate answers to the research questions. This is not to say that qualitative methods should not be used in corpus-based research, but rather that we should acknowledge the importance of quantitative methods, which, if combined in an integrated manner, can offer significant insight into a phenomenon.

The two case studies described in this book also demonstrate that corpus triangulation has not been developed with specific research questions, texts, features or languages in mind, but is a truly universal framework with wide applicability. Specifically, although limited in number, the case studies manage to focus not only on two distinct translation-related phenomena, but also foreground research into both literary and non-literary texts. Moreover, they apply the triangulation framework to the examination of texts produced at different time periods, successfully combining diachronic and synchronic corpora. Triangulation has also been shown to be suitable for the examination of different linguistics features operating at the word (i.e. connectives) and sentence level (i.e. cleft and pseudo-cleft constructions). Finally, two different language pairs are examined, each with its own idiosyncrasies.

It might appear, however, that this book foregrounds a corpus design where translated, non-translated and source texts are combined resulting in a similar number of components for each corpus. This similar corpus design can be said to have come about by coincidence, given that the focus in the first case study is on different genres, while in the second on different time periods. Three elements are selected for each case study (i.e. three genres and three points in time), which is related to the image of the triangle (see Chapter 2), resulting in similarities in corpus design. Moreover, due to the limited availability of English translations of children's and non-fiction books, it has not been possible to create a reciprocal corpus for the first case study (Chapters 5 and 6), which would allow for the investigation of the translation direction as a possible factor affecting the use of connectives, but would also result in a larger number of corpus components and a more complex corpus design. However, we need to recognise that a number of different combinations are possible, which can include more than three elements, resulting in the image of the crystal mentioned in Chapter 2. It is possible that, at some point in the future, as corpus triangulation techniques develop further, we will see more varied corpus designs, which might also come closer to that crystal image.

Finally, a clear indication that the corpus triangulation framework developed in this book contributes towards the advancing of corpus-based translation studies is that it is perhaps the first time that translation studies is not making use of tried and tested methods developed by corpus linguistics, but instead develops its own methodology. Not only that, but it is also the first time that the roles can be reversed and corpus linguistics can adopt this new methodological framework for the study of language. However, it is not just corpus linguistics that can benefit from this triangulation framework. Any discipline where corpora have traditionally been used, or where they might be used in the future, can rely on corpus triangulation principles to increase its understanding of the phenomena it focuses on. For instance, stylistics might use a corpus triangulation design, similar to that described in Chapter 6, to investigate features related to specific writers or translators. By the same token, the corpus design described in Chapter 8 would be of relevance to historical linguistics. Even though longitudinal studies are typically considered as not being triangulated since they do not focus on convergence (Kimchi et al. 1991), as argued in Chapter 2, and elsewhere in the book, triangulation is interested in both convergence and divergence, which is also supported by Denzin in his updated account of triangulation (1989). Literary studies might also use a similar corpus design to examine the diachronic development of literary norms and styles. Finally, as corpus linguistics finds applications in the social sciences (e.g. Dayrell and Urry 2015; Hardaker and Mcglashan 2016), the potential of corpus triangulation can be further expanded; in a sense returning to the discipline where triangulation, as we understand it today, originates. What makes corpus triangulation so utterly exciting is that it offers considerable scope for fertilisation of other disciplines and there is no reason why this framework should be limited to corpus-based translation studies. We must remember, however, that combining corpus-based methods with methods from other disciplines has always

been, and remains, a challenge (McEnery and Hardie 2012). Perhaps now that corpus-based translation studies has slowly started embracing maturity, it can let go of (most) its self-centred preoccupations, and start considering more carefully how it can reach out to other disciplines.

References

Dayrell, C. and Urry, J., 2015. Mediating Climate Politics. *European Journal of Social Theory*, 18(3), pp. 257–273.

Denzin, N.K., 1989. *The Research Act: A Theoretical Introduction to Sociological Methods* 3rd ed. Englewood Cliffs: Prentice Hall.

Hardaker, C. and Mcglashan, M., 2016. Real Men don't Hate Women: Twitter Rape Threats and Group Identity. *Journal of Pragmatics*, 91, pp. 80–93.

Kimchi, J., Polivka, B. and Stevenson, J.S., 1991. Triangulation: Operational Definitions. *Nursing Research*, 40(6), pp. 364–366.

McEnery, T. and Hardie, A., 2012. *Corpus Linguistics.* Cambridge: Cambridge University Press.

Olohan, M., 2004. *Introducing Corpora in Translation Studies.* London and New York: Routledge.

Risjord, M.W., 2000. *Woodcutters and Witchcraft: Rationality and Interpretive Change in the Social Sciences.* Albany: SUNY Press.

Tymoczko, M., 1998. Computerized Corpora and the Future of Translation Studies. *Meta: Journal des traducteurs*, 43(4), p. 652.

Zanettin, F., 2012. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies.* Manchester: St. Jerome.

Appendix 1

Non-translated English fiction books

Title	*Year*	*Author*	*Words*
No Country for Old Men	2005	Cormac McCarthy	69,360
White Teeth	2000	Zadie Smith	169,930
The Help	2009	Stockett Kathryn	159,957
The Kite Runner	2003	Khaled Hosseini	107,287
Oryx and Crake	2003	Margaret Atwood	101,284
Middlesex	2002	Jeffrey Eugenides	196,084
Water for Elephants	2006	Sara Gruen	100,031
American Gods	2001	Neil Gaiman	184,126
Atonement	2001	Ian McEwan	123,820
Total			**1,211,879**

Non-translated Russian fiction books

Title	*Year*	*Author*	*Words*
Санькя	2006	Захар Прилепин	76,633
Кысь	2001	Татьяна Толстая	69,395
Рыба: История одной миграции	2006	Петр Алешковский	63,753
Ж.Д.	2006	Дмитрий Быков	193,492
Священная книга оборотня	2004	Виктор Пелевин	77,719
Путь Мури	2007	Илья Бояшов	36,759
Каменный мост	2009	Александр Терехов	205,308
2017	2006	Ольга Славникова	131,977
Белый Шанхай	2010	Эльвира Барякина	142,961
Total			**997,997**

Non-translated English children's fiction books

Title	*Year*	*Author*	*Words*
The Lightning Thief	2005	Rick Riordan	87,587
The End!	2006	Lemony Snicket (pen name of Daniel Handler)	51,817
Harry Potter and the Deathly Hallows	2007	J. K. Rowling	198,512
The Boy in the Stripped Pyjamas	2006	John Boyne	46,689
Coraline	2002	Neil Gaiman	30,745
Clockwork Angel	2010	Cassandra Clare	132,577
Looking for Alaska	2005	John Green	69,220
The Girl Who Circumnavigated Fairyland in a Ship of Her Own Making	2011	Catherynne M. Valente	81,818
Gathering Blue	2000	Lois Lowry	48,014
Miss Peregrine's Home for Peculiar Children	2011	Ransom Riggs	85,479
The One	2014	Kiera Cass	74,337
The Maze Runner	2009	James Dashner	101,168
The Evolution of Calpurnia Tate	2009	Jacqueline Kelly	81,444
Divergent	2011	Victoria Roth	105,344
Total			**1,194,751**

Non-translated Russian children's fiction books

Title	*Year*	*Author*	*Words*
Астровитянка	2008	Николай Горькавый	126,046
Две кругосветки	2012	Елена Ленковская	52,196
Жизнь среди людей	2015	Алиса Рекунова	95,601
Кузя, Мишка, Верочка и другие ничейные дети	2011	Татьяна Губина	65,004
Асино лето	2014	Тамара Михеева	50,629
Самая младшая	2014	Лариса Романовская	60,772
Большой Кыш	2006	Мила Блинова	75,415
Три сказки об Италии	2008	Светлана Лаврова	40,039
Класс коррекции	2007	Екатерина Мурашова	30,306
Дом, в котором . . . Курильщик	2009	Мариам Петросян	69,328
Здесь вам не причинят никакого вреда	2006	Андрей Жвалевский, Игорь Мытько	61,068
Кадын—владычица гор	2011	Анна Никольская	24,531
Часодеи. Часовой ключ	2011	Наталья Щерба	77,550
Место Снов	2006	Эдуард Веркин	99,719
Маг на два часа	2009	Тамара Крюкова	31,015
Король хитрости	2004	Дмитрий Емец	46,853
Total			**1,006,072**

Non-translated English non-fiction books

Title	*Year*	*Author*	*Words*
Fast Food Nation	2001	Eric Schlosser	108,989
Freakonomics	2005	Steven D. Levitt Stephen J. Dubner	58,487
Blink: The Power of Thinking Without Thinking	2005	Malcolm Gladwell	70,867
The Evolution of God	2009	Robert Wright	159,933
Stiff: The Curious Lives of Human Cadavers	2003	Mary Roach	77,090
The Emperor of All Maladies	2010	Siddhartha Mukherjee	171,807
Collapse: How Societies Choose to Fail or Succeed	2005	Jared M. Diamond	216,558
The Swerve: How the World Became Modern	2011	Stephen Greenblatt	81,871
Gulag: A History	2003	Anne Applebaum	236,053
Total			**1,181,655**

Non-translated Russian non-fiction books

Title	*Year*	*Author*	*Words*
История отмороженных в контексте глобального потепления	2007	Александр Никонов	81,503
Глэм-капитализм	2008	Дмитрий Иванов	28,531
Геополитика: Как Это Делается	2014	Николай Стариков	71,988
Повелительное наклонение истории	2010	Олег Матвейчев	110,104
Сумма биотехнологии	2015	Александр Панчин	78,231
Суперобъекты. Звезды размером с город	2015	Сергей Попов	47,220
Цепная реакция. Неизвестная история создания атомной бомбы	2013	Олег Фейгин	52,482
Почему языки такие разные. Популярная лингвистика	2010	Владимир Плунгян	73,111
Рождение сложности	2010	Александр Марков	96,238
Гравитация	2013	Александр Петров	63,219
Закат империи доллара и конец “Pax Americana”	2003	Андрей Кобяков,Михаил Хазин	67,540
Проект Россия	2006	Various authors	68,197
Цунами 2010-х годов	2008	Максим Калашников	171,416
Total			**1,009,780**

Translated Russian fiction books

Title	*Year*	*Translator(s)*	*Words*
Старикам тут не место	2009	Валерий Минушин	51,012
Белые зубы	2005	Мария Мельниченко	138,660
Прислуга	2010	Мария Александрова	139,416
Бегущий за ветром	2008	Сергей Соколов	77,565
Орикс и Коростель	2004	Наталья Гордеева	79,569
Средний пол	2003	Мария Ланина	163,437
Воды слонам!	2007	Мария Фаликман	84,236
Американские боги	2009	Вадим Михайлин	169,217
Искупление	2004	Ирина Доронина	106,603
Total			**1,009,715**

Translated Russian children's fiction books

Title	*Year*	*Translator*	*Words*
Перси Джексон и похититель молний	2009	ВладимирСимонов	79,572
Конец!	2007	НаталияРахманова	42,439
Гарри Поттер и Дары Смерти	2007	Майя Лахути, Сергей Ильин	167,999
Мальчик в полосатой пижаме	2016	Елена Полецкая	37,349
Коралина	2009	Евгений Кононенко	22,288
Механический ангел	2012	Л. Я. Соловьева	128,849
В поисках Аляски	2012	Юлия Федорова	62,275
Девочка, которая объехала Волшебную Страну на самодельном корабле	2014	Беленкович Владимир	59,199
В поисках синего	2015	Сергей Петров	37,164
Дом странных детей	2012	Елена Боровая	79,446
Единственная	2014	Ирина Тетерина	67,394
Бегущий в Лабиринте	2014	Дмитрий Евтушенко	90,672
Эволюция Кэлпурнии Тейт	2015	Ольга Бухина, Галина Гимон	60,837
Дивергент	2014	Александра Киланова	84,513
Total			**1,019,996**

Translated Russian non-fiction books

Title	*Year*	*Translator*	*Words*
Нация фастфуда	2016	А. Логвинская	84,267
Фрикономика	2016	Я. Лебеденко	58,214
Озарение. Сила мгновенных решений	2010	Вячеслав Логвинов	56,652
Эволюция бога	2012	Ульяна Валерьевна Сапцина	140,404
Кадавр. Как тело после смерти служит науке	2011	Татьяна Мосолова	66,630
Царь всех болезней. Биография рака	2013	М. Виноградова	145,477
Коллапс: почему одни общества выживают, а другие умирают	2010	О. Жаден, А. Михайлова, И. Николаев	192,788
Ренессанс. У истоков современности	2013	И. Лобанов	61,906
ГУЛАГ. Паутина Большого террора	2015	Леонид Мотылев	178,737
Total			**985,075**

Appendix 2

Non-translated Greek popular science articles 1990–1991

Title		*Date*	*Words*
Περισκόπιο της Επιστήμης (Periscope of Science)			
1.	Αναζητώντας την πηγή της νεότητας	05/1991	3,940
2.	Γεύση: Μια πολυδιάστατη αίσθηση	10/1990	2,143
3.	Ένας υπερεπιταχυντής για την Ευρώπη	12/1991	2,114
4.	Ενέργεια από τον ήλιο	03/1991	6,249
5.	Η νόσος των τρελών αγελάδων	06/1991	2,571
6.	Ηλεκτρονική ολογραφία χωρίς ηλεκτρομαγνητικούς φακούς	12/1991	1,029
7.	Κοπωση: Η ασθενεια της εποχης	06/1990	3,373
8.	Μαγνητουδροδυναμικός κινητήρας: Για πλοία θα καταργήσει την έλικα;	06/1991	1,275
9.	Ο μηχανισμός της δίψας	02/1991	2,757
10.	Οθόνες υγρών κρυστάλλων	06/1990	3,337
11.	Οι μελλοντικές εξελίξεις στο αυτοκίνητο	09/1991	2,575
12.	Το Μικροσκόπιο πρωτονίων	09/1991	1,171
Total			**32,534**
Το Βήμα (To Vima)—Προς το Αύριο (Pros to Avrio)			
1.	"Εξυπνα" υφάσματα	14/10/1990	172
2.	"Εξυπνο" ευρωπαϊκό σπίτι	15/12/1991	473
3.	"Κάρτες" αντί για το χρήμα	01/07/1990	312
4.	"Λογικές" οικιακές συσκευές	10/06/1990	395
5.	"Υδρόβιος" τοκετός	07/04/1991	190
6.	Brick, η επανάσταση στους υπολογιστές	16/06/1991	474
7.	Cameo, ο ηλεκτρονικός πυροσβέστης	21/04/1991	419
8.	Το αστέρι των ναυτικών πεθαίνει	15/07/1990	298
9.	Το ντιζάιν κατά της κλειστοφοβίας	03/06/1990	360
10.	Αλλάξτε κανάλι με ένα στυλό	30/09/1990	170
11.	Ανατομικό πληκτρολόγιο	21/10/1990	135
12.	Αντιηλιακό πουκάμισο	07/07/1991	163
13.	Αόρατοι ηλεκτρονικοί κώδικες	17/11/1991	274
14.	Αρκούδες για θεραπεία	29/09/1991	161

(*Continued*)

(Continued)

Title		*Date*	*Words*
15.	Αρωματικές παγίδες εντόμων	24/03/1991	336
16.	Ασύρματο μεγάφωνο: επανάσταση στη στερεοφωνία	25/08/1991	293
17.	Αυτόματος πιλότος για αυτοκίνητα	21/10/1990	95
18.	Αυτόματος πιλότος νοικοκυριού	20/10/1991	389
19.	Βακτηρίδια για περισσότερο χιόνι στις χιονοδρομικές πίστες	30/09/1990	287
20.	Βακτηρίδια για τη διάσωση των αρχαίων	02/12/1990	295
21.	Βακτηριδιοκτόνο ύφασμα	28/04/1991	344
22.	Βλαβερό το χλώριο του νερού	04/11/1990	216
23.	Βλέποντας τους γαλαξίες απο τον . . . βυθό του ωκεανού!	18/08/1991	357
24.	Βρέθηκε το ελιξίριο της νεότητας;	09/09/1990	369
25.	Βρείτε το χαμένο σας παιδί με κομπιούτερ	26/05/1991	275
26.	Γαλάζια νερά χωρίς καρχαρίες	08/12/1991	271
27.	Γλυπτική με λέιζερ	25/11/1990	300
28.	Γονίδια από αγριόχορτα εναντίον της ξηρασίας!	05/08/1990	399
29.	Γονίδιο κατά του πάγου	11/08/1991	158
30.	Γυαλιά χάι τεκ	23/12/1990	320
31.	Διάγνωση καρκίνου σε λίγα λεπτά	20/05/1990	345
32.	Διάγνωση λουμπάγκο	04/11/1990	292
33.	Δορυφορική βιντεοθήκη	16/09/1990	283
34.	Δορυφόροι “συνεννοούνται” με λέιζερ	12/05/1991	331
35.	Εγγραφή στο βίντεο με έναν αριθμό	24/02/1991	275
36.	Εμβόλιο για τον διαβήτη;	12/05/1991	125
37.	Ένα βήμα πιο κοντά στο τεχνητό αίμα	02/06/1991	225
38.	Ένα ένζυμο ένοχο για τον διαβήτη	23/09/1990	126
39.	Ένα ρομπότ για το DNA	04/11/1990	250
40.	Ένας ιός φράσσει τις αρτηρίες	07/10/1990	167
41.	Ένας μεταφραστής τσέπης ομιλεί και με προφορά	19/08/1990	302
42.	Ένας πολυμήχανος εκτυπωτής	18/11/1990	188
43.	Ενέσεις χωρίς φόβο	22/12/1991	88
44.	Εξέταση ούρων για Aids	09/12/1990	172
45.	Έξυπνα τζάμια	30/09/1990	138
46.	Έξυπνα υλικά	08/07/1990	354
47.	Έτοιμο τυρί όποτε εσείς θέλετε	28/07/1991	305
48.	Ευνουχισμένες ντομάτες και θηλυκό ρύζι	18/11/1990	343
49.	Ζωγραφίζοντας στην τηλεόραση	25/11/1990	231
50.	Η γενετική θεραπεία είναι πλέον γεγονός	23/09/1990	237
51.	Η γοητεία του ωαρίου	21/04/1991	240
52.	Η μαγική κάψουλα	02/12/1990	123
53.	Η μαγική πρίζα	23/09/1990	174
54.	Η τέλεια τέχνη της πλαστογραφίας	17/06/1990	307
55.	Η τρισδιάστατη τηλεόραση είναι γεγονός	21/10/1990	234
56.	Θα ψωνίζουμε και . . . τηλεοπτικώς	25/08/1991	215
57.	Ιπτάμενα τηλέφωνα στην Ευρώπη	16/12/1990	218
58.	Ιπτάμενα υποβρύχια	09/06/1991	351
59.	Καθαρίστε το σπίτι σας με σκούπα ρομπότ	30/09/1990	174

Title		*Date*	*Words*
60.	Και “φορετοί” υπολογιστές	24/11/1991	446
61.	Και μαλλιά του . . . σωλήνα	11/11/1990	149
62.	Και οικολογικά εντομοκτόνα	28/07/1991	247
63.	Και οικολογική και μιας χρήσεως	16/06/1991	301
64.	Και τώρα σκάφη με . . . φτερά	16/12/1990	136
65.	Κασετόφωνο DAT σε μέγεθος γουόκμαν	14/10/1990	145
66.	Κατασκευή τεχνητού δέρματος	28/04/1991	385
67.	Κλιματισμός στο δάπεδο	31/03/1991	201
68.	Κόψτε το τσιγάρο με ένα έμπλαστρο	23/09/1990	124
69.	Λάδι από τη θάλασσα	21/04/1991	228
70.	Μαγειρέψτε με υπολογιστή	14/10/1990	179
71.	Μεγάφωνο από βακτηρίδια	07/04/1991	95
72.	Μια “φανταστική πραγματική” δοκιμή	23/06/1991	276
73.	Μια εικόνα από χίλιες λέξεις	23/09/1990	216
74.	Μια πυραμίδα από γυαλί υψώνεται στο Έβερεστ	12/08/1990	442
75.	Μικρόβιο παράγει καύσιμα από σκουπίδια	31/03/1991	149
76.	Μόλυνση και τρύπα του όζοντος	17/11/1991	137
77.	Μπιμπερόν που αλλάζει χρώματα	31/03/1991	171
78.	Μπισκότο αναισθητικό	09/12/1990	138
79.	Μπουφάν με θερμόμετρο	23/12/1990	170
80.	Νανοτεχνολογία και νανοσυσκευές	08/12/1991	572
81.	Νέα μηχανή ντίζελ	31/03/1991	128
82.	Ξεβάφουν τα μπλου τζιν με . . . ευγενή ένζυμα	16/06/1991	255
83.	Ξεναγός στο σούπερμάρκετ	09/12/1990	159
84.	Ο βιολογικός κύκλος της χολέρας	08/09/1991	110
85.	Ο καφές δεν βλάπτει	21/10/1990	88
86.	Οι μαϊμούδες δε θα μιλήσουν	18/11/1990	134
87.	Οι τυφλοί βλέπουν θέατρο	09/12/1990	184
88.	Οπτικές ίνες κατ’ οίκον	11/11/1990	136
89.	Όταν τα πλαστικά θα . . . φυτρώνουν στα χωράφια!	18/08/1991	315
90.	Πατάς ένα κουμπί και έρχονται τα ψώνια	07/07/1991	168
91.	Πατάτα	21/10/1990	395
92.	Πλαστικοί μαγνήτες	07/07/1991	185
93.	Πολύστροφη οδοντόβουρτσα	31/03/1991	231
94.	Πρόβατα που κουρεύονται μόνα τους	12/05/1991	234
95.	Προφυλακτικό για γυναίκες	17/11/1991	199
96.	Πώς μπορείτε να κατασκευάσετε ένα Σύμπαν	02/09/1990	339
97.	Ραντάρ εναντίον εγκλήματος	11/11/1990	166
98.	Ρομπότ για επικίνδυνες αποστολές	31/03/1991	271
99.	Σάκχαρο για τη συντήρηση τροφίμων	05/05/1991	112
100.	Σε λίγα χρόνια διακοπές στον Άρη	13/10/1991	565
101.	Σκουπιδότοποι στα βάθη των ωκεανών	27/01/1991	351
102.	Στα φύλλα το μυστικό της ενέργειας	17/03/1991	253
103.	Στείρωση στα . . . φυτά	08/12/1991	104
104.	Συνταγή για . . . δεινοσαύρους!	11/08/1991	473
105.	Συσκευή παράγει ινσουλίνη	07/07/1991	176
106.	Τεστ για τον καρκίνο	30/06/1991	262
107.	Τηλεοράσεις τοίχου σε . . . φωτογραφία	11/08/1991	314

(*Continued*)

(Continued)

Title		*Date*	*Words*
108.	Το αυτοκίνητο του μέλλοντος	30/09/1990	177
109.	Το ελικόπτερο που . . . βλέπει	28/10/1990	291
110.	Το ιερό δισκοπότηρο της ανοσολογίας	10/11/1991	508
111.	Το κιλό χάνει βάρος	22/07/1990	452
112.	Το πρόγραμμα πληροφορικής Bliss	28/10/1990	193
113.	Το σπασμένο χέρι κινείται	07/04/1991	186
114.	Το σύγχρονο σπίτι είναι πλαστικό!	19/05/1991	307
115.	Το τέλος των λαμπτήρων	10/03/1991	408
116.	Τρένα χωρίς μηχανή	09/12/1990	171
117.	Τυπώστε τη δική σας εφημερίδα	30/06/1991	384
118.	Τυρί χωρίς χοληστερίνη	25/11/1990	82
119.	Τώρα και ιπτάμενο αυτοκίνητο!	10/02/1991	302
120.	Τώρα καρπούζια χωρίς κουκούτσια	27/12/1991	256
121.	Τώρα κομπιούτερ για στρατιώτες	15/09/1991	346
122.	Υπερηχητική μεζούρα	11/11/1990	232
123.	Υποδόρια αντισύλληψη διαρκείας	13/01/1991	225
124.	Υποθαλάσσιος κατάσκοπος της μόλυνσης	22/12/1991	169
125.	Υπολογιστές χωρίς πληκτρολόγιο	30/09/1990	223
126.	Υπολογιστής για λίγους	02/12/1990	135
127.	Υπολογιστής μανεκέν	21/10/1990	126
128.	Υπολογιστής με . . . μύτη λαγωνικού	07/07/1991	585
129.	Φάρμακο για τα πεδία των μαχών	24/02/1991	241
130.	Φάρμακο κατά του πανικού	25/11/1990	83
131.	Φρέσκα τριαντάφυλλα για . . . έξι μήνες	25/11/1990	178
132.	Φωτισμένα χρωμοσώματα	03/11/1991	145
133.	Φωτογραφία που ρετουσάρεται μόνη της	09/12/1990	242
134.	Φωτογραφίες χωρίς . . . φιλμ	06/10/1991	261
135.	Χάρακας με λέιζερ	11/11/1990	105
136.	Χειρουργικά εργαλεία μιας χρήσεως	28/07/1991	192
Total			**33,932**

Τα Νέα (Ta Nea)

1.	"Φ" όπως φωτογραφία	22/02/1991	1,120
2.	PC όπως πιστό σκυλί	10/03/1990	969
3.	Ανώδυνη έγχυση φαρμάκου χωρίς ένεση	11/04/1991	660
4.	Βιταμίνη Α	07/11/1991	1,011
5.	Γιατρειά από τη μάνα γη	03/10/1991	1,232
6.	Διαμάντια . . . σαν αληθινά	17/08/1990	1,130
7.	Ελληνικές γόβες από κομπιούτερ	26/04/1991	735
8.	Ένας κομπιούτερ που ακούει	15/06/1990	818
9.	Ένας υπολογιστής για τα πιο τρελά όνειρα	16/01/1991	956
10.	Η γενετική υπηρέτης μας	19/09/1991	1,201
11.	Η ταχυκαρδία θεραπεύεται με "νυστέρι"!	26/09/1991	1,280
12.	Κάθε επάγγελμα μια ασθένεια	24/11/1990	1,116
13.	Καταζητείται μαύρος εκδικητής	16/03/1991	1,769
14.	Κυτταρομετρία: Νέα μέθοδος για την πρόγνωση της πορείας της νόσου	4/04/1991	906
15.	Με οθόνη και . . . μολύβι	25/05/1990	579

Title		*Date*	*Words*
16.	Μετακινώντας τα σύνορα της ζωής	03/01/1991	1,079
17.	Μια έξυπνη καρτα για όλες τις δουλειές	15/06/1991	1,105
18.	Οι αληθινές . . . ψεύτικες εικόνες	02/03/1991	1,189
19.	Οι εξετάσεις με μέτρο!	05/07/1990	1,293
20.	Όταν το τηλέφωνο χτυπάει στην τσέπη του σακακιού μας	01/12/1990	686
21.	Πόλεμος λόγω . . . βιντεοδίσκων!	12/01/1990	758
22.	Πονοκέφαλος	12/07/1990	2,453
23.	Προεμμηνορροϊκό σύνδρομο	16/08/1990	1,277
24.	Τα "βουβά" ανευρύσματα	25/04/1991	1,112
25.	Τεχνητά μέλη με μικροϋπολογιστή	08/03/1990	1,283
26.	Τεχνολογία: απειλή αλλά και μόνη ελπίδα	24/08/1991	1,267
27.	Τηλέφωνο τσέπης	23/03/1990	1,020
28.	Το μέλλον θα είναι ευκρινές	11/09/1990	858
29.	Τροφές, τα μόνα φάρμακα	15/03/1990	775
30.	Τρώγοντας μόνο, δεν έρχεται η . . . ερωτική όρεξη	08/08/1991	623
31.	Φως στις ασθένειες εκ φωτός	26/04/1990	1,220
Total			**33,480**
Grand total			**99,946**

Non-translated Greek popular science articles 2003–2004

Title		*Date*	*Words*
Περισκόπιο της Επιστήμης Periscope of Science			
1.	Αϋπνία: Που οφείλεται και πως μπορεί να αντιμετωπιστεί	07/2004	4,444
2.	Βιοφωτόνια: Το φως των οργανισμών	05/2003	4,052
3.	Γονιδιακή θεραπεία	07/2003	5,654
4.	Έξυπνο σπίτι: Η τεχνολογία στην καθημερινή ζωή	09/2004	3,365
5.	Η αναβίωση της ατμομηχανής	01/2003	4,120
6.	Ιπτάμενα αυτοκίνητα: Η φαντασία γίνεται πραγματικότητα	05/2004	3,679
7.	Τα μυστήρια της δομής του DNA	03/2003	3,604
8.	Υποβρύχια . . . από μπετόν	04/2003	6,001
Total			**34,919**
Βήμα Science (Vima Science)			
1.	E-mail από εξωγήινους	17/10/2004	1,906
2.	Hi-tech επανάσταση στα αυτοκίνητα	06/07/2003	1,120
3.	Γενετικά όπλα εναντίον ασθενειών	14/11/2004	1,078
4.	Διαδίκτυο με γεύση και οσμή	06/06/2004	1,084
5.	Εις υγείαν των γυναικών	20/07/2003	1,428
6.	Εκτυπωτές με ζωντανό μελάνι	18/05/2003	929
7.	Εμβόλια κατά του καρκίνου	08/06/2003	1,491

(*Continued*)

(Continued)

Title		*Date*	*Words*
8.	Η επανάσταση της ιντερνετικής τηλεφωνίας	05/09/2003	1,189
9.	Η καρδιοχειρουργική έτοιμη για νέο άλμα	02/05/2004	1,785
10.	Η Πράσινη Επανάσταση	19/10/2003	1,869
11.	Θεραπεύοντας τον "πυλώνα" του σώματος	21/03/2004	1,426
12.	Καλά και άσχημα νέα για τον προστάτη	09/03/2003	2,009
13.	Να πάρω κινητό στο παιδί μου;	09/11/2003	1,923
14.	Οι αντιδράσεις στην τεχνολογική επέλαση	15/08/2004	2,423
15.	Οι προφήτες της αρχιτεκτονικής	03/08/2003	1,641
16.	Όταν η επιστήμη ταξιδεύει με το Έντερπράιζ	27/04/2003	1,989
17.	Πιείτε μια . . . κάμερα στην υγειά σας	20/06/2004	1,872
18.	Πυρηνική ενέργεια από ένα . . . ποτήρι	12/12/2004	1,087
19.	Τα ναυάγια σπάνε τη σιωπή τους	08/02/2004	2,039
20.	Υπολογιστές με . . . αισθήματα	16/02/2003	1,902
21.	Φύση και παιδαγωγική σκέψη	06/06/2004	1,071
Total			**33,261**
Focus			
1.	2025: Οι νέες θεραπείες	02/2003	1,886
2.	Αδύνατη μέχρι θανάτου	05/2003	1,819
3.	Αεριούχα αινίγματα	03/2003	1,224
4.	Αχ, αυτός ο ιδρώτας	08/2004	1,051
5.	Γιατί νιώθεις ενοχές;	02/2003	1,610
6.	Γίναμε αόρατοι!	10/2003	1,020
7.	Εθισμένοι χωρίς ναρκωτικά	04/2004	1,249
8.	Η μαγεία των κεραμικών	06/2003	1,739
9.	Θεματικά πάρκα	11/2004	961
10.	Οι ημέρες της δημιουργίας	10/2003	1,401
11.	Όταν τα φάρμακα ωφελούν κατά λάθος	01/2003	1,246
12.	Πού είναι η έξοδος;	01/2003	1,211
13.	Πρωτεομική	11/2003	1,138
14.	Πωπω, βήχας!	11/2004	798
15.	Σε καραντίνα!	11/2003	928
16.	Σμήνη ρομπότ	11/2003	1,351
17.	Στο κέντρο της Γης	05/2003	1,344
18.	Συναγερμός! Το ήλιο τελειώνει	09/2003	808
19.	Τα "μπαούλα" του 2000	05/2004	1,115
20.	Τα μυστικά της φονικής πνευμονίας	06/2003	1,152
21.	Τα χρώματα και τα μυστικά τους	08/2004	1,195
22.	Το μέλλον είναι πλαστικό	09/2004	852
23.	Το στοιχείο μηδέν	09/2003	813
24.	Υπό κατασκευήν	08/2003	1,450
25.	Φορέστε το μέλλον	04/2004	1,556
26.	Χικ . . . χικ . . . χικ!	09/2003	1,162
27.	Η αντιύλη; Όχι μία, αλλά τρεις!	05/2004	1,238
Total			**33,317**
Grand total			**101,497**

Non-translated Greek popular science articles 2009–2010

Title		*Date*	*Words*
Περισκόπιο της Επιστήμης Periscope of Science			
1.	Αϋπνία	02/2009	8,093
2.	Αφαλάτωση	05/2009	7,343
3.	Εναέριες ανεμογεννήτριες: Αξιοποιώντας τους ανέμους μεγάλου υψομέτρου	05/2010	6,183
4.	Τα όρια της ανθρώπινης ταχύτητας: Υπάρχει τέλος στα παγκόσμια ρεκόρ;	01–02/2010	5,532
5.	Υπολειμματικά όργανα: Είναι πράγματι άχρηστα;	04/2010	6,423
Total			**33,574**
Βήμα Science (Vima Science)			
1.	3DTVπροσδοκίες και παρενέργειες	28/03/2010	1,890
2.	Αλλαγή φύλου με ένα γονίδιο	17/01/2010	1,943
3.	Αλμα από το κινητό στον υπολογιστή	05/04/2009	1,153
4.	Αόρατοι εχθροί στα νοσοκομεία	15/02/2009	2,320
5.	Διάδραση και διασκέδαση	20/09/2009	1,123
6.	Εικόνα με εκατομμύρια λέξεις	14/06/2009	1,358
7.	Εννέα μήνες που κρατάνε χρόνια	28/11/2010	2,733
8.	Έρευνα στη ρίζα του αυτισμού	27/06/2010	1,966
9.	Η ανατομία του στολισμένου δέντρου	25–27/12/2009	1,935
10.	Κοινωνία ώρα . . . 3D	01/08/2010	2,003
11.	Κοινωνίες πιθήκων και ανθρώπων	0/11/2009	1,130
12.	Ναι, υπάρχει ελληνική ατομική ώρα	29/03/2009	1,959
13.	Ο θαυμασμός και ο οίκτος περνούν απ' το . . . στομάχι	19/07/2009	1,891
14.	Όλος ο κόσμος σε μια ταμπλέτα	03/10/2010	1,699
15.	Παιχνίδια με τη ζωή . . .	11/01/2009	2,018
16.	Πέφτοντας από τα σύννεφα	05/07/2009	1,942
17.	Πόσο μας απειλούν οι νάνοι	02/05/2010	1,749
18.	Τα μυστικά της υπερ-μακροβιότητας	11/07/2010	2,167
19.	Το μυστικό στην ουρά της σαύρας	12/04/2009	1,762
20.	Τρισδιάστατο άλμα στο σινεμά	13/12/2009	1,161
Total			**35,902**
Focus			
1.	"Χάκερ, χτυπά!"	12/2010	868
2.	Α Παγκόσμιος κυβερνοπόλεμος	10/2009	1,748
3.	Δεσμώτες των ονείρων	09/2010	1,290
4.	Δύσκολοι καιροί για παιδιά	01/2009	2,148
5.	Εγώ Εύα, εσύ;	02/2010	3,885
6.	Έρχεται η εποχή του petabyte	03/2010	1,952
7.	Ζήσε την ψηφιακή 6η αίσθηση	10/2009	983
8.	Η "επαυξημένη πραγματικότητα"	10/2010	1,867
9.	Η δίαιτα των σπηλαίων	03/2010	2,280

(*Continued*)

(Continued)

Title		*Date*	*Words*
10.	Ο Ήρων, οι Άραβες και τα Ρομπότ	11/2009	1,374
11.	Οι εφευρέτες της ανάστασης	11/2010	619
12.	Συντροφιά γεμάτη εκπλήξεις	03/2009	3,015
13.	Τι συμβαίνει όταν . . .	03/2010	3,012
14.	Το GPS των σχέσεων	12/2010	1,748
15.	Το υπεράνθρωπο σούπερ μυαλό	07/2009	1,334
16.	Το χέρι μου τρελάθηκε!	09/2010	1,202
17.	Τρίχες με μέλλον	08/2009	1,251
Total			**30,576**
Grand total			**100,052**

Translated Greek popular science articles 2003–2004

Title	*Date*	*Words*
Popular Science GR		
1. Βρίσκεται στα γονίδια το μυστικό της γενεαλογίας;	01/2004	3,822
2. Η Sally έχει 2 μητέρες + 1 πατέρα	03/2003	3,785
3. Η θεραπεία του μικρού μου αδερφού	04/2004	4,393
4. Μυαλό στη μηχανή	02/2004	4,498
5. Ο άνθρωπος που μπέρδεψε την κοπέλα του με ρομπότ	09/2003	3,376
6. Το GPS μου θα σε βρει . . . όπου κι αν είσαι!	11/2004	3,418
Total		**23,292**
Scientific American GR		
1. Εγκέφαλε, θεράπευσον σεαυτόν	01/2004	4,342
2. Ένας πρόγονος για μας	11/2003	4,497
3. Ζωύφια στον εγκέφαλο	10/2003	1,840
4. Η ψυχαγωγία ξεπερνά τα σύνορα	12/2003	3,083
5. Καταπολεμώντας το στρες	03/2004	5,662
6. Οι κεραίες γίνονται έξυπνες	03/2004	4,545
7. Στρατός από μικρά ρομπότ	03/2004	3,142
8. Το πλέγμα: υπολογιστικές υπηρεσίες δίχως όρια	10/2003	5,014
Total		**32,125**
Βήμα Science (Vima Science)		
1. "Κλαδέψτε" τους υπολογιστές σας	13/6/2004	1,053
2. Αυτοκίνητα έτοιμα προς . . . απογείωση	29/6/2003	1,115
3. Ένας ξένος μέσα στο σώμα μας	21/12/2003	2,081
4. Η "ζούγκλα" των εντέρων και τα (χρήσιμα) μυστικά της	23/5/2004	1,985
5. Η γλώσσα μας είναι παράσιτο	2/2/2003	2,024
6. Η δύναμη της μουσικής	7/3/2004	2,587
7. Η μηχανή που διαβάζει (και γράφει) τη σκέψη	23/3/2003	976
8. Νέα θεραπεία για την παχυσαρκία	9/5/2004	2,641
9. Νέο όπλο κατά του καρκίνου	5/10/2003	1,819

Title	*Date*	*Words*
10. Ο πολιτισμός είναι στη φύση μας	31/10/2004	2,542
11. Οι μπαταρίες του αύριο είναι . . . νάνοι	11/7/2004	1,194
12. Ποιος οδηγεί αυτό το αυτοκίνητο;	30/11/2003	1,904
13. Ραδιόφωνο χωρίς όρια	8/8/2004	1,128
14. Ρομπότ που ζουν μόνα τους	6/7/2003	1,200
15. Τα σήματα που προκαλούν ημικρανία	31/8/2003	2,854
16. Τι είναι η κρυπτογράφηση;	10/8/2003	2,180
17. Το αδύνατο σημείο του καρκίνου	25/7/2004	1,787
18. Το διαδίκτυο αλλάζει την Κίνα	28/11/2004	1,086
19. Το κύτταρο που μας κάνει ανθρώπους	11/7/2004	1,889
Total		**34,045**
Grand total		**89,462**

Translated Greek popular science articles 2009–2010

Title	*Date*	*Words*
Βήμα Science (Vima Science)		
1. +4oC Υποδεχτελιτε μιαν άλλη Γη	01/03/2009	2,385
2. Déjà vu Η σκοτεινή πλευρά της οικειότητας	10/05/2009	2,294
3. Άγνωστη Γη: Τα επτά μεγαλύτερα μυστήρια του πλανήτη μας	26/04/2009	4,950
4. Από τα ατομικά στα οπτικά ρολόγια	29/03/2009	1,016
5. Ασθένειες . . . αγνώστου ταυτότητος	23/08/2009	1,004
6. Αυτά τα burgers είναι ίδια	26/07/2009	2,087
7. Γη ώρα μηδέν	15/11/2009	2,848
8. Γιατί οι παγετώνες δεν κρατούν για πάντα;	25/07/2010	2,047
9. Διαλέγουμε ηγέτες που πραγματικά χρειαζόμαστε	07/06/2009	2,390
10. Εγώ, ο ιός. Γιατί είμαστε άνθρωποι μόνο κατά το ήμισυ!	21/02/2010	2,143
11. Ένας εγκέφαλος για όλες τις ηλικίες	09/08/2009	4,187
12. Ευτυχείτε! Είναι μεταδοτικό	11/01/2009	3,117
13. Η Βαβέλ είναι παντοτινή	22/08/2010	2,507
14. Η επιστήμη πίσω από τη φαντασία	18/05/2008	2,113
15. Η ζωή αρχίζει στα 100	13/09/2009	2,319
16. Η μυστική ζωή του εγκεφάλου	08/03/2009	2,009
17. Η οδός της . . . μικρότερης αντίστασης	12/04/2009	2,270
18. Η τηλεόραση πέθανε. Ζήτω η τηλεόραση!	04/05/2008	2,050
19. Ηλεκτρισμός από τα έγκατα της Γης	10/08/2008	854
20. Καθαρή ενέργεια από το διοξείδιο του άνθρακα	03/08/2008	1,592
21. Καλύτερη ζωή με πράσινη χημεία	17/10/2010	2,257
22. Κάτω τα χέρια απ' το τιμόνι!	18/04/2010	2,191
23. Κλικ στην e-γεια	09/11/2008	1,133
24. Λουκέτο στην παιδική χαρά & Δικτυωθείτε!	04/05/2008	1,398
25. Μαθήματα αρχιτεκτονικής από τους τερμίτες	2803/2010	1,912
26. Με τη δύναμη της σκέψης	04/05/2008	2,017

(*Continued*)

(Continued)

Title		*Date*	*Words*
27.	Μεταλλαγμένοι σολομοί κολυμπούν προς το πιάτο μας!	26/09/2010	973
28.	Ναρκωτικά: από την παρανομία στο . . . φαρμακείο	02/09/2010	1,135
29.	Ο Δρόμος προς το μη-σύμπαν	24/02/2008	2,154
30.	Ο θρίαμβος των λουλουδιών	15/03/2009	1,750
31.	Ο προφήτης της πολιτικής	11/04/2010	2,107
32.	Όλα τα λεφτά στο κόκκινο	06/09/2009	1,938
33.	Όταν το λίπος προστατεύει!	04/04/2010	1,164
34.	Παγωμένα βλέμματα και ζεστή καρδιά	29/11/2009	1,184
35.	Παραδοσιακά παιχνίδια στην . . . πρίζα!	05/07/2009	1,534
36.	Πενταλοκίνητη εξερεύνηση	01/03/2009	1,838
37.	Ποια εξαφανισμένα είδη θα μπορούσαν να "αναστηθούν";	18/01/2009	1,389
38.	Πολυσύμπαν	14/03/2010	2,626
39.	Πρέπει να καταργήσουμε το κρέας;	19/09/2010	2,446
40.	Σαρακοστιανές μέδουσες και κέικ από πλαγκτόν	05/04/2009	2,115
41.	Τα μαθηματικά του Χόλιγουντ	28/02/2010	1,035
42.	Τα μέσα των πολιτών	04/05/2008	2,310
43.	Ταξίδι στην καρδιά του ηφαιστείου	15/11/2009	1,052
44.	Τι νιώθει ο σύγχρονος άνθρωπος;	31/01/2010	2,132
45.	Το θεώρημα της πίτσας	17/01/2010	1,095
46.	Το τέλος της γνώσης;	14/02/2010	2,305
47.	Τοξικά κοκτέιλ καθημερινής χρήσης . . .	16/03/2008	2,293
48.	Τροπικοί: η κοιτίδα της ζωής	16/05/2010	1,710
49.	Υπόγειες σεισμικές βολίδες σαρώνουν τον πλανήτη	04/10/2009	1,293
50.	Φως στα μυστικά του εγκεφάλου	08/08/2010	3,970
Total			**100,638**

Non-translated English popular science articles 1990–1991

Title		*Date*	*Words*
Popular Science EN			
1.	High-intensity Headlight	07/1990	1,838
2.	Fuzzy Logic	07/1990	1,620
3.	Night and day solar	02/1990	1,043
4.	How can we win the war against garbage	10/1990	3,906
5.	Coming for cars: Smart Glass	12/1991	1,531
6.	Back to the future	06/1991	1,304
7.	Laptops go cellular	09/1991	2,684
8.	Home Theaters	11/1991	2,356
9.	Memory chips that don't forget	04/1990	2,600
10.	Digital Holography	01/1991	1,875
11.	Robot Insects	03/1991	2,535
12.	Speeders beware: It's the photo and laser cops	09/1990	2,109
13.	Space Spies	03/1990	3,023
14.	Searching for the cosmic connection	04/1991	5,277
Total			**33,701**

Title	*Date*	*Words*
Scientific American EN		
1. Suspension Feeding Vertebrates	03/1990	4,127
2. What Causes Diabetes?	07/1990	4,624
3. Ocean Acoustic Tomography	10/1990	3,532
4. Aspirin	01/1191	4,635
5. The Silicon Retina	05/1991	3,076
6. Civil Liberties in Cyberspace	09/1991	3,759
7. Could a Machine Think?	01/1990	4,973
8. Antichaos and Adaptation	08/1991	4,993
Total		**33,719**
New Scientist		
1. Nicotine on the brain	03/11/1990	2,423
2. Seeing colours in their true light	11/08/1990	2,604
3. Another life for electronics	24/02/1990	2,604
4. The power of negative matter	17/03/1990	2,803
5. A cool solution to global warming	12/05/1990	3,028
6. Kingdoms in turmoil	23/03/1991	3,359
7. Why small is sometimes sexy	07/09/1991	2,811
8. On the saving of the species	19/01/1991	2,039
9. Building Babbage's dream machine	29/06/1991	2,530
10. The spread of cancer in the human body	21/07/1990	3,750
11. The first gene on Earth	09/11/1991	3,130
12. The World of Liquid Crystals	18/05/1991	3,423
Total		**34,504**
Grand total		**101,924**

Non-translated English popular science articles 2003–2004

Title	*Date*	*Words*
Popular Science EN		
1. Every step you take	10/2004	3,490
2. Mind over machine	02/2004	3,927
3. My little brother on drugs	04/2004	4,038
4. Putting the gene back in genealogy	12/2004	4,364
5. Sally has 2 mommies and 1 daddy	01/2003	3,273
6. The man who mistook his girlfriend for a robot	08/2003	3,506
Total		**22,598**
Scientific American EN		
1. An army of small robots	11/2003	2,673
2. Ancestor to Call Our Own	01/2003	4,103
3. Antennas get smart	07/2003	3,887
4. Brain repair yourself	09/2003	3,962

(*Continued*)

(Continued)

Title		*Date*	*Words*
5.	Bugs in the brain	03/2003	1,733
6.	Digital entertainment jumps the border	03/2003	3,181
7.	Taming stress	09/2003	5,665
8.	The grid: Computing without bounds	04/2003	3,906
Total			**29,110**
New Scientist			
1.	A game of chance	07/06/2003	2,761
2.	Behind the mask	19/07/2003	2,582
3.	Born to trade	18/09/2004	2,306
4.	Fat buster	10/04/2004	2,435
5.	Give it some gas	19/06/2004	2,654
6.	Hack out the useless extras	05/06/2004	818
7.	Hear my voice	22/02/2003	1,910
8.	Hit cancer where it hurts	03/07/2004	2,642
9.	Is migraine all in the mind?	21/06/2003	2,763
10.	It's a jungle in there	24/04/2004	2,534
11.	Now who's in the driver's seat?	08/11/2003	2,273
12.	Radio sans frontiers	10/07/2004	2,621
13.	The 'blog' revolution sweeps across China	24/11/2004	1,931
14.	The cell that makes us human	19/06/2004	2,670
15.	The language bug	18/01/2003	2,489
16.	The power of music	29/11/2003	3,992
17.	The stranger within	15/11/2003	2,576
18.	Uncharted territory	31/05/2003	2,481
19.	When jeep meets jump-jet	14/06/2003	2,538
Total			**46,966**
Grand total			**98,674**

Non-translated English popular science articles 2009–2010

Title		*Date*	*Words*
New Scientist			
1.	Brain imaging monitors effect of movie magic	08/09/2010	955
2.	I, virus: Why you're only half human	29/01/2010	2,289
3.	Seven unsolved medical mysteries	16/12/2008	914
4.	A measure for the multiverse	03/03/2010	2,425
5.	Across the ocean in a pedal-powered submarine	28/01/2009	1,812
6.	Better living through green chemistry	10/03/2010	2,829
7.	Born believers: How your brain creates God	04/02/2009	2,251
8.	Déjà vu: Where fact meets fantasy	25/03/2009	2,370
9.	Digital doomsday: the end of knowledge	02/02/2010	2,373
10.	Five emotions you never knew you had	13/01/2010	2,285
11.	Follow me: The origins of leadership	12/09/2008	2,139

	Title	*Date*	*Words*
12.	For sustainable architecture, think bug	22/02/2010	1,681
13.	How 'citizen journalists' are transforming the news	15/03/2008	2,058
14.	How to survive the coming century	25/02/2009	2,872
15.	How your friends' friends can affect your mood	30/12/2008	2,704
16.	Icy stares and dirty minds: Hitch-hiking emotions	09/09/2009	1,392
17.	Iron Man: The science behind the fiction	01/05/2008	2,030
18.	Jellyfish sushi: Seafood's slimy future	04/03/2009	2,359
19.	Language lessons: You are what you speak	26/05/2010	2,296
20.	Living world: Why the tropics are hotbeds of evolution	21/04/2010	1,896
21.	Look, no hands: Cars that drive better than you	06/04/2010	1,901
22.	Meltdown: Why ice ages don't last forever	24/05/2010	2,511
23.	Next generation of video games will be mental	13/03/2008	1,704
24.	Nine games computers are ruining for humanity	18/05/2009	2,298
25.	Obesity: Food kills, flab protects	10/3/2010	1,037
26.	Our changing brains	06/04/2009	3,482
27.	Petal power: How flowering plants conquered the world	29/10/2008	2,243
28.	Picking our brains: Nine neural frontiers	30/03/2010	3,373
29.	Plan to pierce heart of urban monster volcano	04/11/2009	888
30.	Post-human Earth: How the planet will recover from us	30/09/2009	2,421
31.	Psychoactive drugs: From recreation to medication	01/09/2010	1,346
32.	Secrets of the centenarians: Life begins at 100	07/09/2009	2,360
33.	Seismic boom: Breaking the quake barrier	29/07/2009	2,294
34.	Solved: The mathematics of the Hollywood blockbuster	18/02/2010	510
35.	Super clocks: More accurate than time itself	04/02/2009	2,452
36.	Ten extinct beasts that could walk the Earth again	07/01/2009	1,885
37.	The calorie delusion: Why food labels are wrong	15/07/2009	2,292
38.	The future of television is online	15/03/2008	2,118
39.	The hunt for the Un-universe	25/01/2008	2,534
40.	The online doctor will see you now	08/11/2008	1,050
41.	The perfect way to slice a pizza	09/12/2009	1,705
42.	The predictioneer: Using games to see the future	17/03/2010	2,501
43.	The secret life of the brain	05/11/2008	2,328
44.	Toxic cocktail	01/09/2008	2,626
45.	Transgenic fish swimming towards a plate near you	15/09/2010	1,052
46.	Turning CO2 back into hydrocarbons	03/03/2008	1,425
47.	Unknown Earth: Our planet's seven biggest mysteries	28/09/2008	4,191

(*Continued*)

(Continued)

Title		*Date*	*Words*
48.	Veggieworld: Why eating greens won't save the planet	20/07/2010	2,260
49.	Web 3.0: Playing it safe with our data	14/03/2008	993
50.	Who needs coal when you can mine Earth's deep heat?	17/07/2008	1,188
51.	Winners wear red: How colour twists your mind	28/08/2009	2,405
Total			**105,303**

Index