COMPLEX WORDS, CAUSATIVES, VERBAL PERIPHRASES AND THE GERUND

ROMANCE LANGUAGES VERSUS CZECH (A PARALLEL CORPUS-BASED STUDY)

EDITED BY
PETR ČERMÁK
DANA KRATOCHVÍLOVÁ
OLGA NÁDVORNÍKOVÁ
PAVEL ŠTICHAUER

KAROLINUM

Complex Words, Causatives, Verbal Periphrases and the Gerund
Romance Languages versus Czech (A Parallel Corpus-Based Study)

Edited by:
Petr Čermák
Dana Kratochvílová
Olga Nádvorníková
Pavel Štichauer

Reviewed by:
Prof. Bohumil Zavadil
Dr. Jana Pešková

KAROLINUM PRESS
Karolinum Press is a publishing department of Charles University
Ovocný trh 560/5, 116 36, Prague 1, Czech Republic
www.karolinum.cz

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MŠMT
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

This work was supported by the European Regional Development Fund-Project "Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World" (No. CZ.02.1.01/0.0/0.0/16_019/0000734) and by the Charles University project Progres Q10, Language in the shiftings of time, space, and culture.

© Karolinum Press, 2020
© Edited by Petr Čermák, Dana Kratochvílová, Olga Nádvorníková and Pavel Štichauer, 2020
© Text by Leontýna Bratánková, Petr Čermák, Štěpánka Černikovská, Jan Hricsina, Jiří Jančík, Jaroslava Jindrová, Dana Kratochvílová, Zuzana Krinková, Petra Laufková, Olga Nádvorníková, Daniel Petrík, Pavel Štichauer and Eliška Třísková, 2020
Designed by Jan Šerých
Typeset by Karolinum Press
First edition

ISBN 978-80-246-4554-4
ISBN: 978-80-246-4616-9 (pdf)

Univerzita Karlova
Nakladatelství Karolinum 2020

www.karolinum.cz
ebooks@karolinum.cz

CONTENTS

1. EXPRESSIONS OF POTENTIAL PARTICIPATION, ITERATIVITY, CAUSATION, INGRESSIVITY AND ADVERBIAL SUBORDINATION IN THE LIGHT OF PARALLEL CORPORA

Petr Čermák, Dana Kratochvílová, Olga Nádvorníková, Pavel Štichauer ---- **9**
1.1 Investigation project and its history ---- **10**
1.2 Objectives and scope of the present monograph ---- **11**
1.3 Organisation of the monograph ---- **12**
1.4 Terminological remarks ---- **13**
    1.4.1 Romance languages under scrutiny and use of the term *Romance* ---- **13**
    1.4.2 Use of the terms *counterpart* and *respondent* ---- **14**

2. CORPUS DESIGN & CORPUS-BASED CONTRASTIVE RESEARCH METHODOLOGY

Olga Nádvorníková ---- **15**
2.0 Introduction ---- **16**
2.1 Corpus-based contrastive research methodology ---- **16**
2.2 Corpora used in this study ---- **21**

3. MORPHOLOGICALLY COMPLEX WORDS IN ROMANCE AND THEIR CZECH RESPONDENTS

Pavel Štichauer, Jan Hricsina, Jiří Jančík, Jaroslava Jindrová, Zuzana Krinková, Daniel Petrík ---- **25**
3.0 Introduction ---- **26**
3.1 Word-formation: complex vs simple words ---- **27**
3.2 Romance and Czech: common and different word-formation patterns ---- **27**
3.3 The typology of Czech respondents ---- **29**
    3.3.1 Typology of Czech respondents of the adjectives with the suffix *-bile/-ble/-vel* ---- **29**
    3.3.2 Typology of Czech respondents for verbs with the prefix *re-/ri-* ---- **31**
3.4 The modal suffix *-ble/-bile/-vel* ---- **32**
    3.4.1 Data elaboration and analysis ---- **33**
    3.4.2 Quantitative distribution of the types ---- **35**
    3.4.3 Discussion of various examples ---- **36**
3.5 The iterative prefix *re-/ri-* ---- **39**
    3.5.1 Data elaboration and analysis ---- **39**

3.5.2 Quantitative distribution of the types ---- **40**
3.5.3 Discussion of various examples ---- **41**
3.6 Concluding remarks ---- **43**

4. CAUSATIVE CONSTRUCTIONS IN ROMANCE AND THEIR CZECH RESPONDENTS
Petr Čermák, Dana Kratochvílová, Petra Laufková, Pavel Štichauer ---- **45**
4.0 Introduction ---- **46**
4.1 Definition of causativity and its forms of expression ---- **46**
4.2 Causativity in Romance languages ---- **48**
4.2.1 Analytic type ---- **48**
4.2.2 Synthetic type ---- **49**
4.2.3 Characteristics of the Romance construction *hacer/fare/faire/fazer* + infinitive ---- **49**
4.3 Causativity in Czech ---- **50**
4.3.1 Word-formatting causativity ---- **51**
4.3.1.1 Verbs derived from another verb ---- **52**
4.3.1.2 Verbs derived from an adjective ---- **53**
4.3.1.3 No change in the lexical basis, expressing causativity through a prefix *roz-* ---- **53**
4.3.2 Semantic causativity ---- **54**
4.3.2.1 Suppletive types ---- **54**
4.3.2.2 Causative interpretation resulting from syntax ---- **55**
4.3.3 Analytic causativity ---- **55**
4.3.3.1 Causative verbs followed by a subordinate clause ---- **55**
4.3.3.2 Causative verbs followed by a nominal syntagma ---- **55**
4.3.3.3 (Semi-)causative verbs followed by an infinitive ---- **56**
4.4 Our typology of Czech respondents ---- **57**
4.5 Methodology ---- **59**
4.6 Causative constructions in Romance – formal comparison ---- **60**
4.7 Analysis of Czech respondents ---- **62**
4.7.1 Primary Czech respondents ---- **64**
4.7.1.1 Type 3 – *shodit* type (*hacer caer / far cadere / faire tomber / fazer cair*) ---- **64**
4.7.1.2 Type 8 – *what makes you think that > proč myslíte?* ('why do you think that?') ---- **66**
4.7.1.3 Type 4 – *dát vypít* type ---- **68**
4.7.2 Secondary Czech respondents ---- **69**
4.7.2.1 Type 5 – *dohnat k slzám* type ---- **69**
4.7.2.2 Type 9 – other translation ---- **71**
4.7.2.3 Type 7 – *způsobit, že tál* type ---- **71**
4.7.2.4 Type 1 – *rozplakat* type ---- **73**
4.7.2.5 Type 2 – *posadit* type and type 6 – *způsobit tání* type ---- **73**
4.7.2.6 Type 10 – no translation ---- **73**
4.8 Conclusions ---- **73**

5. INGRESSIVE PERIPHRASES IN ROMANCE AND THEIR CZECH RESPONDENTS

Dana Kratochvílová, Jaroslava Jindrová, Pavel Štichauer, Eliška Třísková –––– **79**
5.0 Introduction –––– **80**
5.1 Verbal periphrases in Romance –––– **80**
  5.1.1 Approaches to verbal periphrases and the goal of our study –––– **81**
5.2 Aspect and Aktionsart –––– **82**
  5.2.1 Aspect and Aktionsart in Romance languages –––– **82**
  5.2.2 Aspect and Aktionsart in Czech –––– **83**
  5.2.3 Verbal periphrases and the relationship to aspect and Aktionsart –––– **84**
  5.2.4 Ingressive MoA –––– **85**
    5.2.4.1 Initial ingressivity in Romance languages –––– **85**
      5.2.4.1.1 Derivative ingressive MoA in Romance –––– **85**
      5.2.4.1.2 Analytical ingressive MoA in Romance –––– **85**
    5.2.4.2 Initial ingressivity in Czech –––– **88**
      5.2.4.2.1 Derivative ingressive MoA in Czech –––– **88**
      5.2.4.2.2 Analytical ingressive MoA in Czech –––– **91**
5.3 Corpus analysis –––– **93**
  5.3.1 Methodology –––– **93**
  5.3.2 Results of the corpus analyses –––– **95**
    5.3.2.1 Ingressive constructions expressing the mere beginning of a process –––– **95**
    5.3.2.2 Ingressive constructions expressing the beginning of a process and the notion of effort by part of the subject –––– **96**
    5.3.2.3 Ingressive constructions expressing the beginning of a process and the notions of suddenness and unexpectedness –––– **99**
    5.3.2.4 Ingressive constructions expressing the beginning of a process and the notions suddenness, abruptness and previous retention –––– **101**
    5.3.2.5 Ingressive constructions expressing the beginning of a process and the notions of suddenness, abruptness and vehemence –––– **102**
5.4 Concluding remarks –––– **103**

6. THE ROMANCE GERUND AND ITS CZECH RESPONDENTS

Olga Nádvorníková, Leontýna Bratánková, Štěpánka Černikovská, Jan Hricsina –––– **107**
6.0 Introduction –––– **108**
6.1 Morphology of the Romance gerund –––– **108**
6.2 Romance gerund as a converb –––– **109**
  6.2.1 Syntactic functions of the Romance gerund –––– **110**
  6.2.2 Semantic interpretation of the (adverbial) Romance gerund –––– **112**
6.3 Typology of Czech respondents of the Romance gerund –––– **114**
6.4 Data elaboration and quantitative analysis of the Romance gerund –––– **115**
  6.4.1 Factors influencing the frequency of the Romance gerund –––– **117**
  6.4.2 Syntactic functions of the Romance non-periphrastic gerund –––– **118**

6.5 The adverbial Romance gerund and its Czech respondents –––– **121**
  6.5.1 Semantic types of the Romance adverbial gerund and the Czech transgressive –––– **121**
  6.5.2 Czech respondents of the Romance gerund –––– **129**
    6.5.2.1 Finite verbs as respondents of the Romance gerund –––– **131**
      6.5.2.1.1 Coordinate finite clause as a respondent of the Romance gerund –––– **131**
      6.5.2.1.2 The subordinate finite clause as a respondent of the Romance gerund –––– **132**
    6.5.2.2 Nominalisations as respondents of the Romance gerund –––– **134**
    6.5.2.3 Non-finite verb forms as respondents of the Romance gerund –––– **136**
6.6 Conclusion –––– **144**

7. FORMAL EXPRESSIONS VS ABSTRACT LINGUISTIC CATEGORIES: COMING TO TERMS WITH POTENTIAL (NON-VOLITIONAL) PARTICIPATION, ITERATIVITY, CAUSATION, INGRESSIVITY AND ADVERBIAL SUBORDINATION
Petr Čermák, Dana Kratochvílová, Olga Nádvorníková, Pavel Štichauer –––– **147**
7.0 Introduction –––– **148**
7.1 Correspondences of the analysed phenomena across Romance languages –––– **149**
7.2 Czech respondents of the analysed phenomena vs systemic counterparts –––– **151**
7.3 Exploiting the parallel corpus in search of language universals and abstract categories –––– **153**

Bibliography –––– **154**

1. EXPRESSIONS OF POTENTIAL PARTICIPATION, ITERATIVITY, CAUSATION, INGRESSIVITY AND ADVERBIAL SUBORDINATION IN THE LIGHT OF PARALLEL CORPORA

PETR ČERMÁK
DANA KRATOCHVÍLOVÁ
OLGA NÁDVORNÍKOVÁ
PAVEL ŠTICHAUER

1.1 INVESTIGATION PROJECT AND ITS HISTORY

The present monograph is the result of a long-term project, which started in 2013 when an investigation group was formed by experienced teachers (Petr Čermák, Pavel Štichauer, Jan Hricsina, Jaroslava Jindrová, Zuzana Krinková, Olga Nádvorníková), and their Ph.D. (MA, in one case) students (Leontýna Bratánková, Štěpánka Černikovská, Jiří Jančík, Dana Kratochvílová, Petra Laufková, Daniel Petrík, Eliška Třísková) from the Department of Romance Studies at the Faculty of Arts of Charles University. The objective of the investigation group was to explore the possible usage of the parallel corpus InterCorp (created by the Institute of the Czech National Corpus at the very same university; for further details refer to Nádvorníková this volume) for a contrastive analysis of Romance languages and Czech, the mother tongue of all the authors. The group was comprised of students and professors of Spanish, Italian, French and Portuguese. Four structurally different phenomena, which can be found in all these languages, were selected for analysis: complex words with the suffix *-ble*/*-bile*/*-vel* and the prefix *re-*/*ri-*, causative construction *hacer*/*fare*/*faire*/*fazer* + infinitive, ingressive verbal periphrases and the gerund. The primary objective was to study the Czech respondents of these language phenomena that can be found in the InterCorp corpus, thus testing its usefulness for this kind of study and formulating conclusions regarding the systemic Czech counterparts.

In the first stage, the analysis was conducted separately for each Romance language (with a shared introduction and conclusion) and was oriented primarily to Czech readers and Czech professors and students of Romance languages, i.e. their knowledge of Czech and the knowledge of at least one of the four Romance languages was taken for granted when analysing the linguistic material and presenting the results.

The first stage was concluded in 2015 with the publication of the collective monograph *Románské jazyky a čeština ve světle paralelních korpusů* (Čermák – Nádvorníková et al. 2015), which was published in Czech.

The project entered its second phase at the beginning of 2018, with the final result being the present monograph. This stage was conducted by four members of the original investigation group, who are now professors and assistant professors at the Department of Romance Studies (Petr Čermák, Dana Kratochvílová, Olga Nádvorníková, Pavel Štichauer). Building on the results of the first phase, on the original data and the illustrative examples that were prepared by the whole investigation group, the objective was to create a new monograph which, while sharing some of the original objectives with the Czech version, would present the results in a new light.

1.2 OBJECTIVES AND SCOPE OF THE PRESENT MONOGRAPH

While the monograph representing the result of the first stage of the project was written in Czech, the present book is oriented primarily to readers whose native language is not Czech and who might only have a primary or secondary interest in this Slavic language. It thus offers a corpus-based analysis of four research topics in four Romance languages in the light of their respondents in a typologically different language. The monograph is written in English and all Czech examples are translated or glossed. The anticipated audience for this book are primarily scholars interested in at least one of the Romance languages under scrutiny (Spanish, Italian, French, Portuguese). Therefore, while we do not expect readers to be well acquainted with all four languages and we present translations for Romance examples, we do not provide exhaustive glosses for them or detailed descriptions of what is the function of the analysed phenomena within the Romance language system in general. Where necessary, we concentrate solely on important differences that can be found between the four languages.

While analysing the same phenomena, i.e. complex words, causative constructions, ingressive periphrases and the gerund, the scope and point of view of these that are presented differ notably from the original Czech monograph. In the second stage, we decided to consider these phenomena as generally Romance,[1] meaning that we considered their function in Spanish, Italian, French and Portuguese as being largely comparable (while mentioning some partial differences that were significant for our research) and then contrasted these phenomena with Czech as a whole. Thus, from the contrastive point of view, this monograph compares the representation of potential (non-volitional) participation, iterativity, causation, ingressivity and adverbial subordination in Romance and in Czech (rather than presenting partial analyses concentrating solely on one of the four languages, i.e. Spanish vs Czech, Italian vs Czech, French vs Czech and Portuguese vs Czech, as in the first stage of the research).

The second important difference is closely related to the above-presented point. Since we consider the phenomena as generally Romance, we also approach them on a more abstract level than we did in the first stage. At this point, we are not concerned primarily with the formal manifestation of the phenomena under scrutiny; we rather consider the suffix *-ble/-bile/-vel*, the prefix *re-/ri-*, the construction *hacer/fare/faire/fazer* + infinitive, the ingressive periphrases and the gerund as prototypical or "pure" expression form of abstract categories of potential (non-volitional) participation, iterativity, causativity, beginning of an action and adverbial (or circumstantial) subordination. Ranging from complex words through causatives and periphrases to the gerund, we aim to explore the extent to which the abovementioned linguistic categories are systemically encoded in Czech and on which language levels these can primarily be found. Therefore, our primary goal is to present a corpus-based contrastive analysis of these highly abstract categories and their manifestation in Czech, thus moving toward presenting the semantic notions generally attributed to them in a new light. This is based on concrete language data, rather than on intuition or formal manifestation.

---

[1] However, as in the first stage, we excluded Romance languages other than Spanish, Italian, French and Portuguese from our research.

1.3 ORGANISATION OF THE MONOGRAPH

This monograph is organised into seven sections, including the present introductory chapter, i.e. **Chapter 1**.

Chapter 2 is devoted to a detailed description of the corpus we work with and the method. It is the only chapter that has one single author, Olga Nádvorníková. In the rest of the book, this chapter is referred to as **Nádvorníková (this volume)**.

Chapters 3–6 represent the core of this monograph. As stated previously, these chapters are based on the original data and incorporate some of the observations made in the Czech version of the book. The following list briefly presents the topic of each chapter, the name of the person preparing the English version and the names of the authors of the original Czech subchapters, which have been incorporated into the new version. All contributors to the original Czech monograph are also listed as co-authors of the new version.

Chapter 3 focuses on complex words, more specifically, on the suffix *-ble*/*-bile*/*-vel* and the prefix *re-*/*ri-*, the function of these affixes and the representation of these functions in Czech. The English version was written by Pavel Štichauer, who is also the main author of the original Czech version. The authors of the original Czech subchapters referring to Spanish, French and Portuguese were Jan Hricsina (Pt., suffix *-vel*), Jaroslava Jindrová (Pt., prefix *re-*), Jiří Jančík (Fr.), Zuzana Krinková (Es., prefix *re-*) and Daniel Petrík (Es., suffix *-ble*). This chapter is referred to as **Štichauer et al. (this volume)** in the rest of the book.

Chapter 4 deals with the causative construction *hacer*/*fare*/*faire*/*fazer* + infinitive and the expression of causativity in Czech. The English version was written by Petr Čermák and Dana Kratochvílová, Petr Čermák was also the main author of the Czech version. The authors of the original subchapters referring to Italian, French and Portuguese were Petra Laufková (Fr. and Pt.) and Pavel Štichauer (It.). In the rest of the book, this chapter is referred to as **Čermák – Kratochvílová et al. (this volume)**.

Chapter 5 analyses ingressive verbal periphrases and the expression of the beginning of a process in Czech. The English version was written by Dana Kratochvílová, while the main author of the Czech version was Jaroslava Jindrová. Dana Kratochvílová was also the author of the original Czech subchapter referring to Spanish. Authors of the original subchapters referring to Italian and French were Pavel Štichauer (It.) and Eliška Třísková (Fr.). This chapter is referred to as **Kratochvílová – Jindrová et al. (this volume)** in the rest of the book.

Chapter 6 is devoted to the Romance gerund and its Czech respondents. The English version was written by Olga Nádvorníková, who is also the main author of the Czech version. The original subchapters dedicated to Spanish, Italian and Portuguese were written by Leontýna Bratánková (It.), Štěpánka Černikovská (Es.) and Jan Hricsina (Pt.). In the rest of the book, this chapter is referred to as **Nádvorníková et al. (this volume)**.

Finally, **Chapter 7** presents the conclusions and possibilities for future study in the area of contrastive corpus-based analysis and the study of abstract linguistic categories and their formal manifestation. This chapter was written by Petr Čermák, Dana Kratochvílová, Olga Nádvorníková and Pavel Štichauer and is referred to as **Čermák – Kratochvílová – Nádvorníková – Štichauer (this volume)**.

1.4 TERMINOLOGICAL REMARKS

In order to conclude this introductory chapter, we consider it important to present the most important terms that are used throughout this whole monograph and to specify the meaning we attributed to them.

1.4.1 ROMANCE LANGUAGES UNDER SCRUTINY AND USE OF THE TERM *ROMANCE*

As previously mentioned, this monograph is concerned solely with Spanish, Italian, French and Portuguese. For the sake of simplicity, we often use terms such as Romance construction, Romance prefix etc. when referring to a phenomenon we have analysed. It is important to bear in mind, that to a certain level, this is an oversimplification since we do not consider at all the Catalan and Galician language, minor Romance languages and dialects, and more importantly, we do not analyse Romanian, which displays greater structural differences from Spanish, Italian, French and Portuguese. Therefore, throughout this whole monograph, the use of the term *Romance* is identified exclusively with the four analysed Romance languages.

When referring to Spanish, Italian, French and Portuguese separately, we use the following abbreviations: Es., It., Fr. and Pt.

1.4.2 USE OF THE TERMS *COUNTERPART* AND *RESPONDENT*

We often come across situations in our analyses where the expected systemic functional counterpart of a Romance phenomenon does not appear among the prevailing translational solutions found in the InterCorp parallel corpus. Since this distinction is crucial to us, we use the term *counterpart* when discussing the theoretical Czech systemic equivalent of a specific Romance phenomenon while the term *respondent* is reserved for specific Czech translations found in the corpus. The opposition between counterpart and respondent can thus also be understood in terms of langue (typological counterpart) and parole (used respondent).

2. CORPUS DESIGN & CORPUS-BASED CONTRASTIVE RESEARCH METHODOLOGY

OLGA NÁDVORNÍKOVÁ

2.0 INTRODUCTION

Multilingual corpora strongly changed the research paradigm in contrastive studies, making it possible to base the contrastive statements not only on intuition but on large corpus data. As pointed out by Altenberg – Granger (2002, 7), bilingual and multilingual corpora have brought about a revival of interest in contrastive linguistics, since they opened up new possibilities of research, based on empirical data. According to these authors, "the information gained from corpora is both richer and more reliable than that derived from introspection" (ibid.).

Specific methods and approaches subsequently developed, e.g. bi-directional analysis ('Johansson's procedure', see Johansson 2007) or the use of 'translation counterparts as markers of meaning' (Malá 2013 and 2014). With the analysis of the overall pattern of translation correspondence, we can 'see through multilingual corpora' (Johansson 2007) and shed new light on the differences and similarities between the languages compared.

These developments would not be possible without the constitution of a rigorous methodology of the exploitation of multilingual corpora, taking into account, on the one hand, the limitations of the representativeness of these corpora in terms of size and composition, and, on the other hand, the potential specific features of the language of translation (see Nádvorníková 2017a and 2017b). This chapter first provides a brief summary of the basic methodological principles of corpus-based contrastive research (**Section 2.1**) to subsequently explain the strengths and the limitations of the corpora used in the research introduced in this book (**Section 2.2**).

2.1 CORPUS-BASED CONTRASTIVE RESEARCH METHODOLOGY

Most corpora used in contrastive corpus-based research is comprised of original, non-translated texts and the corresponding translations. These corpora are mostly called 'parallel' (see Xiao – Yue 2009, 241–242; Aijmer 2008, 276; Granger 2003, 21), with a potential distinction between unidirectional parallel corpora (i.e. containing translations only in one translation direction, e.g. from English into Norwegian and not from Norwegian into English) and bi-directional ones (i.e. comprising source and target texts in both directions of translation).[2] If in a bidirectional parallel corpus, the non-translated components have the same characteristics in terms of size and composition (and, eventually, sampling techniques), the parallel corpus may be called 'comparable'.[3]

Nevertheless, the use of the terms 'parallel corpus' and 'comparable corpus' in contrastive corpus-based research is not consistent. First, a comparable corpus cannot contain translations or cannot be multilingual. In the former, the corpora in the two (or more) languages are of the same size and composed of the same text types, but are not translations of each other.[4] In the latter, the comparable components are written in the same language but differ in specific properties: e.g. the corpus *Jerome*, comprising translated and non-translated texts in the same language – Czech (see Chlumská 2013 and 2017 and an example of its exploitation in 2.2). A similar terminological confusion can be observed in the term 'parallel': Granger (1996, 38) used the term 'parallel corpus' for corpora comparable in terms of size and composition.

In this research, we will follow the most consensual use of the aforementioned terminology, reserving the term 'parallel' for bilingual or multilingual corpora containing translationally equivalent texts (see e.g. Peters – Picchi – Biagini 2000, 74) and the term 'comparable' for corpora with the same size and composition (see also Xiao and Yue 2009, 240–241 or Aijmer 2008, 276).

However, more important issues discussed in the literature related to the use of parallel corpora in contrastive research concern methodological principles and restrictions that have to be taken in consideration while making contrastive statements on the basis of the comparison of original texts and the corresponding translations.

The first question that arises in this context is the delimitation of the units compared: what is the source item and what is its 'equivalent' in translation? The identification of the source unit and its potential counterparts requires a deep insight into their *valeur*, i.e. their position in the system of all the languages under scrutiny. In the research introduced in this book, based on the comparison of four different Romance languages and Czech, this question becomes even more pressing since the language units entering the comparison may have a different *valeur* in the source Romance languages. The gerund, for example, has a different frequency, different functions and a different position in the system of non-finite verb forms in Italian, French, Spanish and Portuguese. For this reason, a *tertium comparationis* of the cross-linguistic term

---

[2] Xiao and Yue (2009, 241) also mention multidirectional corpora where the same source text can be compared with its translations into several languages.

[3] The most prominent example of comparable parallel corpus English-Norwegian Parallel Corpus (ENPC, see Johansson 2007).

[4] See the definition of a comparable corpus in Aijmer (2008, 276): "A comparable corpus on the other hand does not contain translations but consists of texts from different languages which are similar or comparable with regard to a number of parameters such as text type, formality, subject-matter, time span, etc."

converb was suggested for its comparison with Czech (see Nádvorníková et al. this volume).[5]

The identification of the 'equivalent' of the source unit in translation has to address numerous issues. First, from the point of view of translation studies, the analysis of 'translation equivalence' at the level of only words or sentences is inaccurate since translators do not translate words or sentences but texts. In addition, the term 'equivalence' is itself questionable, as it can be understood both in a descriptive and prescriptive meaning (what corresponds to the source item or what *should* correspond, see e.g. Guidère 2011, 83). Thus, in our book, we distinguished the two meanings by using the term 'respondent' for concrete translation solutions in Czech, and by reserving the term 'counterpart' for potential systemic equivalents, see Čermák – Kratochvílová – Nádvorníková – Štichauer (this volume, **Section 1.4.2**).[6]

However, in the actual analyses of bilingual parallel concordance, a researcher has to encounter a large range of respondents, i.e. also multiple candidates to the systemic counterparts of the search unit. The crucial issue, in this case, is the distinction between the particular translation solutions and the prevailing types of respondents (*recurrent translation patterns*, see Krzeszowski 1990, 27), which potentially reveal the systemic equivalences. In fact, solid contrastive statements can only be formulated on the latter, whereas the former can be used in a study in the domain of translation studies focussed on special translation techniques (e.g. modulation or transposition, see Vinay – Darbelnet 1995) or translation quality assessment (e.g. omissions or additions).[7]

The last issue defining the usability of parallel (translation) corpora in contrastive research is related to potential specific features of the language of translated texts, different from the non-translated ones. These differences may be due to the influence of the source language (interference, shining through), but also due to the translation process itself (so-called translation universals, see Baker 1996, 176–177 for the definitions given below). The specific features of translation that are the most discussed in literature are simplification ("The idea that translators subconsciously simplify the language or message or both"; for research see e.g. Vanderauwera 1985; Laviosa 2002, or Cvrček – Chlumská 2015), explicitation ("The tendency to spell things out in translation, including, in its simplest form, the practice of adding background information", see e.g. Blum-Kulka 1986; Olohan – Baker 2000; Pápai 2004, or Nádvorníková 2017c) and normalisation ("The tendency to conform to patterns and practices that are typical of the target language, even to the point of exaggerating them", see e.g. May 1997 or Kenny 2001).[8]

---

[5] The necessity of *tertium comparationis* in contrastive linguistics is mentioned e.g. in Goddard – Wierzbicka (2008); see also Altenberg – Granger (2002, 15–18). Barlow (2008) points out that without a common basis for the comparison of the analysed phenomena, the contrastive analysis will always compare pears and apples; in the best of the cases, however, contrastive analysis compares different kinds of apples (Barlow 2008, 101).

[6] See a similar distinction in Johansson (2007, 5; translation *correspondence* vs systemic *equivalence*).

[7] Missing equivalents in translation may be due not only to the (voluntary or involuntary) omissions performed by the translator, but also to technical issues (misaligned segments). Moreover, the missing counterpart may be compensated outside the given parallel segment.

Contrastive research based on parallel (translation) corpora implemented several methodological principles designed to identify and/or avoid the influence of the specific features of translation. The basic principle is the systematic identification of the direction of translation: indeed, in a corpus of mixed directions of translation, potential sources of interference are multiplied. This principle is often combined with the bi-directional analysis, which also compares the translation respondents of a given item in the opposite direction of translation. A bi-directional analysis is especially in use in comparable corpora, where the components in all the directions of translation are comparable in size and composition (see above). We did not apply the bidirectional analysis systematically to all the topics in our analysis because the subcorpora of translations from Czech into Romance are much smaller than those in the opposite direction of translation and thus not comparable. Therefore, the bidirectional analysis was tested only in the case of the gerund, in order to establish to what extent the Czech transgressive corresponds to the Romance gerund (see Nádvorníková et al. this volume).

The specificities of parallel corpora (both in the translated and non-translated parts) can also be identified by the comparison with the corresponding monolingual reference corpora. In fact, parallel (translation) corpora, by definition, cannot be representative of the entirety of the language use, since they are limited to texts and the types of text being translated (some types of text, e.g. letters or e-mail messages, are rarely translated) or because there are more translations in one direction of translation (e.g. from English into Czech) than in another (e.g. from Czech into English), cf. Granger – Lerot – Petch-Tyson (2003, 20). For this reason, it is recommended to compare the results obtained from parallel corpora to those extracted from monolingual corpora, referential for the given languages (see e.g. Altenberg – Granger 2002, 9). However, we did not apply this procedure to our study, since a systematic comparison of the results in the four topics to five reference corpora (in Czech and in the four Romance languages) would be to go beyond the scope of this book. Nevertheless, we decided to verify the potential specificity of the language of translation in our research at least in the first topic addressed in this book: causative constructions (see Čermák – Kratochvílová et al. this volume). As explained in that chapter, the Romance causative construction (*hacer/fare/faire/fazer* + infinitive) has a wide range of types of respondents in Czech (synthetic as well as analytic, see **Section 4.3**). If the Czech translations were influenced by the source language, we could expect there to be a higher frequency of the analytic respondent *nechat* + infinitive (the closest by its form to the Romance causative constructions), in comparison with the non-translated texts. In order to test this assumption, we used the corpus *Jerome* (comparable translation corpus of Czech, see Chlumská 2013 and http://wiki.korpus.cz/doku.php/en:cnk:jerome). The corpus comprises translated and non-translated texts in equal amounts (mostly fiction, but also a subcorpus of non-fiction). The whole corpus contains 85 million tokens but also includes a smaller subcorpus (5 million tokens) balanced according to the source languages (14 languages, including the four Romance languages under study in this book).[9] The proportions of source languages in the unbalanced corpus correspond to their proportion in the Czech publishing market; consequently, English as a source language prevails. For our experiment, we used both variants of the Jerome corpus – balanced as well as unbalanced. The results of the corpus search are shown in **Table 2.1**:

---

[8] The specific language of translation is sometimes called 'translationese' (see e.g. Baker 1993 or Mauranen 1999). However, as pointed out by Chlumská (2017, 23), 'specific features of translation' and 'translationese' are not synonymous, since the latter conveys a negative evaluation.

Tab. 2.1. Comparison of the frequency of Czech causative construction nechat + *infinitive in the Jerome corpus*

Jerome corpus (*nechat* + infinitive)[10]	**Unbalanced corpus**		**Balanced corpus**	
	Non-translated texts	Translated texts	Non-translated texts	Translated texts
Size of the corpus (in tokens)	42,401,470	42,563,842	2,547,367	2,540,043
Abs.fq.	5,107	6,389	401	297
Rel.fq. (ipm)	120	150	157	117
Dice coefficient	0.22		–0.30	

Table 2.1 shows that in absolute as well as in relative frequencies (ipm), the frequency of the construction *nechat* + infinitive in translated and non-translated texts are different. In the unbalanced corpus, the frequency is higher in the translated texts, whereas in the balanced corpus, the result is the opposite. According to the chi-squared test, both differences are statistically significant (at p<.001). However, as shown in Cvrček – Kodýtek (2013), the statistical significance does not necessarily mean the statistical *relevance* (the so-called *effect size*), i.e. whether it is possible to identify a relevant factor behind it. In order to test the effect size, we used the Dice coefficient, based on the comparison of the relative frequencies:

$$\text{Dice} = 2\text{x}\ (\text{ipm1} - \text{ipm2}) / (\text{ipm1} + \text{ipm2})$$

The Dice coefficient results vary between –2 and 2, which are both extreme values signalling high relevance of the difference in frequency. However, in our analysis of

---

[9] Since the design of the corpus is synchronic, it only includes translations published after 1992. In addition, it avoids the potential influence of the authors' idiolects by limiting the number of texts written by one author to three books only. More books translated by one translator are accepted although the authors of the originals must be different.

[10] In order to reduce the amount of extraction noise and increase the comparability of the results in the two subcorpora, we used a simplified regular expression *[lemma="nechat"] [tag="Vf.*"]*, without potential elements between the verb nechat and the infinitive. Despite this limitation, we consider the results reliable.