Empirical Approaches to Cognitive Linguistics

Empirical Approaches to Cognitive Linguistics:

Analyzing Real-Life Data

Edited by

Milla Luodonpää-Manni, Esa Penttilä
and Johanna Viimaranta

Cambridge
Scholars
Publishing

Empirical Approaches to Cognitive Linguistics: Analyzing Real-Life Data

Edited by Milla Luodonpää-Manni, Esa Penttilä and Johanna Viimaranta

This book first published 2017

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

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

ISBN (10): 1-4438-7325-X
ISBN (13): 978-1-4438-7325-3

TABLE OF CONTENTS

Part Two: Research Based on Information from Language Users

INTRODUCTION

MILLA LUODONPÄÄ-MANNI, ESA PENTTILÄ, AND JOHANNA VIIMARANTA

This book is a collection of articles that take a cognitive linguistics view on analyzing language. This means that it is founded on the idea that language is a system of conceptualized form–meaning pairings that are used in interactive situations among speakers to process and convey information. In other words, language is not regarded as a mere collection of abstract principles that form a system used for understanding and producing linguistic utterances, but it is these specific usage events that help us conceptualize the world and shape our understanding of grammar through human interaction. Language structure emerges from language use, as, for example, Langacker (1987) points out in his seminal work. Due to its usage-based nature, cognitive linguistics has always been theoretically and methodologically compatible with studying language in its actual real-life contexts. In addition to theory formulation, testing these theories in actual usage data has been considered important in cognitive linguistic since the inception of the discipline (e.g. Rudzka-Ostyn 1988; Geeraerts, Grondelaers & Bakema 1994; Herlin 1994).

The present volume grew out of two workshops organized by the Finnish Cognitive Linguistics Association (FiCLA). The workshops were entitled "Cognition, Grammar, and Corpora" and "Cognition, Meaning, and Usage", and they were arranged at The Finnish Conference of Linguistics in Turku in 2014 and in Vaasa in 2015, respectively. The aim of the two successive workshops was to discuss the various questions that corpora raise in the study of grammatical and semantic phenomena and to encourage a lively discussion on the ways corpora are, can, and may be used in the study of language. In addition to papers presented in the workshops, the collection also includes additional contributions that fit the theme. The chapters compiled in this volume put the usage-based approach into practice by investigating authentic and/or experimental data and by relying on different types of data-based approaches.

1. General orientation of the volume

1.1. Introduction to the field

As pointed out above, one of the basic assumptions of cognitive linguistics is that language is usage-based. The idea can be summed up with two key notions: meaning is use and linguistic structure emerges from use (Tomasello 2009, 69). The first of these notions can be traced back to late Wittgenstein (1953), and the second one contrasts with one of the tacit assumptions of generative linguistics: the idea that language exists as an innate, autonomous faculty in our brain. As a consequence, all our knowledge of language can be claimed to emerge from our experience of how language is actually used. This is how language is acquired and the fact must be taken into account when theorizing about language and studying its properties.

In addition to the usage-based nature of language, the cognitive approach to linguistic analysis relies on a few other basic assumptions. One of them is the idea that language is not regarded as an autonomous cognitive faculty but instead reflects the same general human principles of conceptual organization and information processing as the other human cognitive abilities (Croft & Cruse 2004, 2; Geeraerts 2006b, 4–5). This means that language should not be regarded as a separate area of cognition, but coincides with our other cognitive capacities – such as perception, attention, motor skills, and visual and spatial processing – and functions according to the same principles as they do. This idea is known as the Cognitive Commitment and was first referred to by Lakoff (1990, 40). At present, as Divjak, Levshina, and Klavan (2016, 449–451) point out, the Cognitive Commitment can be regarded as having "at least three different interpretations" depending on which aspect of cognition or which special field of cognitive science is emphasized, but we will not go further into that here.

The Cognitive Commitment closely relates to the Generalization Commitment, according to which language should not be viewed as a modular entity consisting of independent components but rather as a non-modular whole where the different aspects – phonology, morphology, syntax, semantics and pragmatics – all function according to same general principles of categorization (Lakoff 1990, 40; see also Evans & Green 2006, 27–40). Because of these two commitments, the different subfields of linguistic analysis as well as the different disciplines studying human cognition are regarded as inter-related and contributing to as well as benefitting from each other.

Another basic assumption behind the cognitive linguistics approach is the idea that grammar is conceptualization, one of the basic claims made by Langacker (1987). The claim is based on the idea that "grammar and meaning are indissociable" (Langacker 1999, 1), which places meaning in the central position of grammatical analysis implicating that grammatical analysis cannot be reduced to a mere truth-conditional correspondence between linguistic structure and the outer world. In this respect, one of the central concepts is construal, which could be defined as the way in which human beings view, construe, and conceptualize a particular situation. Instead of objectively mirroring the world, linguistic meaning is always perspectival and views the world in a particular way imposing a perspective onto it (Geeraerts & Cuykens 2007, 5). So, instead of merely reflecting the world, linguistic meaning allows us to shape it as well. Geeraerts (2006b, 4) provides a simple example of this by imagining a situation where the speaker, who is standing in the back garden, wants to refer to a bike that was left near the front door of the house. The phrases *It's behind the house* and *It's in front of the house* would both be possible, but they would indicate different perspectives. The first phrase construes the situation from the speaker's point of view, while the second one does it from that of the house. Similar alternative perspectives, not merely spatial ones, are embedded in our language, and it is one of the main tasks of cognitive linguistics to analyze these conceptualizations. Therefore, the starting point of cognitive linguistic analysis is placed on meaning and the analysis itself could be regarded as primarily semantically-oriented or semantically-founded (see e.g. Geeraerts 2006b, Geeraerts & Cuykens 2007).

The above-mentioned basic assumptions of cognitive linguistics are reflected in various ways in the studies presented in this volume. Although the articles approach linguistic analysis from different frameworks and research traditions and tackle different research questions, they all regard the usage-based nature of linguistic analysis as essential and accept the basic tenets of cognitive linguistics as tacit assumptions that underlie their endeavor to explain the mysteries behind language and cognition.

1.2. Empirical research in cognitive linguistics[1]

Embracing usage-based perspective on language has led to an increasing use of methods and data that are based on observing real-life language usage. This, together with the recent development of automatic tools that facilitate the handling of large quantities of data, has rapidly increased the use of empirical methods in cognitive linguistics, a

phenomenon sometimes referred to as the Empirical Turn (e.g. Fischer 2010; Geeraerts 2006a). Empirical methods have been the topic of many recent workshops (e.g. the two above-mentioned ones and an abundance of related workshops world-wide) and volumes (e.g. Gries & Stefanowitsch 2006; Rice & Newman 2010; Glynn & Fischer 2010; Janda 2013; Glynn & Robinson 2014, to name but a few) that have attempted to introduce new methodological innovations and to promote interdisciplinary cooperation with neighboring disciplines, such as psycho- and neurolinguistics. The aim of this book is similarly to discuss the types of empirical data and the multiple ways of analyzing such data from the cognitive perspective and, at the same time, to present innovative contemporary Finnish research to international audiences.

Although the use of empirical methods in cognitive linguistics is steadily on the increase, it is not entirely clear how we should define empirical research. There are at least two ways of understanding the concept; empirical research may be defined either in terms of the type of *evidence* or in terms of the type of (quantitative and statistical) *analysis* employed in the study. In this book, empirical research is understood in the first sense because we believe that language in all its diversity can best be studied by using the entire spectrum of modern quantitative and qualitative methods.

In cognitive linguistics literature, empirical evidence is often divided into two main categories: corpus data and experimental/elicited data (e.g. Tummers, Heylen & Geeraerts 2005, 229; Geeraerts 2006a, 23; comments of Dylan Glynn in Arppe et al. 2010, 7). Corpus data typically consist of various kinds of text corpora, with transcribed spoken text corpora and even audio corpora being increasingly available for research (e.g. The Spoken Corpus of Russian National Corpus, available online at www.ruscorpora.ru; The Longitudinal Corpus of Finnish Spoken in Helsinki, available at The Language Bank of Finland (https://www.kielipankki.fi/language-bank/); or the European Parliament Interpreting Corpus, available at http://sslmitdev-online.sslmit.unibo.it/corpora/corporaproject.php?path=E.P.I.C.). Elicited data, on the other hand, typically involve different types of speakers' responses in experimental settings or surveys that are used in linguistic studies to an increasing extent. Although survey data can be characterized as a form of *collective introspection* (Tummers, Heylen & Geeraerts 2005, 229), introspective data and fabricated examples are not typically considered as empirical evidence.[2] Instead, empirical evidence covers all *non-introspective (external) data that is independent from the conscious reflection or the production of language by the researcher*. The concept of collective introspection, however, shows that the dichotomy between

introspection and empirical research is a matter of degree rather than a categorical division. While corpus data is primarily produced for purposes other than research (and may therefore be considered the least introspective), the participants in experimental and survey settings are always aware of producing data for research purposes and, therefore, there will always be at least some introspective monitoring of language production involved. We will come back to the methodological status of introspection later on.

When empirical research is viewed this way in terms of empirical evidence, it has to be pointed out that of course there exists a long tradition of empirical research in linguistics. Fields such as sociolinguistics, psycholinguistics, field linguistics, language typology, historical linguistics, foreign language learning, and language acquisition research, in particular, have resorted to different types of elicitation tasks, corpus data and psycholinguistic tests (see e.g. Geeraerts 2006a, 35; Arppe & Järvikivi 2007a, 2007b). From the Finnish perspective, talking about empirical revolution in linguistics is especially odd because empirical study is deeply rooted in the local research tradition, and the study of regional dialects based on corpus data collected through extensive field work has been an elementary part of Finnish linguistics since the second half of the 19th century (e.g. Häkkinen 2008, 147–158; Hovdhaugen et al. 2000, 245–246). Early examples of the tradition include studies by Setälä (1883) and Latvala (1895).

Today it is recognized that the current trend towards the use of real-life authentic data may be considered as a natural extension of the (cognitive) linguistic research tradition rather than an empirical revolution (see e.g. Glynn 2014, 7). Emphasizing the novelty of the use of empirical methods in cognitive linguistics is probably related to at least two historical reasons. First, cognitive linguistics has always positioned itself at odds with the generative paradigm and favored usage-based models of language instead of purely theoretical ones. On the other hand, the current emphasis on empirical methods may be related to the fact that the methods used in the early foundational works of cognitive linguistics, such as Talmy (1986), Langacker (1987; 1988) and Lakoff (1987), were predominantly based on constructed examples, rather than corpus data, and in this respect their approach resembled that of the generative paradigm that they were supposed to oppose. Arppe and Järvikivi (2007a), for example, emphasize the fact that producing data and theory at the same time is methodologically problematic. Importantly, however, the early works of cognitive linguistics paradigm laid the theoretical foundation for the usage-based perspective on language that naturally invites empirical research and is thus directly linked to the current use of empirical

evidence: if language structure is indeed the result of language usage, it is only natural to study language usage in authentic real-life data. Although the early American studies in cognitive linguistics did not typically rely on corpus data, the study of corpus materials has been part of the European research tradition since the early days of the discipline (e.g. Rudzka-Ostyn 1988; Geeraerts, Grondelaers & Bakema 1994; Herlin 1994). This can be seen in the contributions to this volume as well, since they all offer empirical approaches to cognitive linguistics in the sense that they rely on empirical corpus data and/or data elicited in experimental settings.

Although the European setting in cognitive linguistics is different from that of the US, it is worth noticing that there is no unified European context in cognitive linguistics. Despite the dominance of English as a publishing language, important research in European cognitive linguistics is published in various national languages as well, which partly makes it inaccessible to the audiences in other (even neighboring) countries. This is especially true for such big European languages as French and Russian that have a rich national tradition in different types of cognitive and other linguistic research. The existence of such national traditions that are largely unknown outside the countries themselves and the fact that the generative tradition never had a similar standing in Europe as it did in the US gives a new perspective to the so-called Empirical Turn. The development of research ideals is usually closely related to the various national or regional research traditions. For example, in Finland, as already pointed out, the empirical research tradition has always been strong and big data sets were collected already in the 19th century. The empirical revolution is a prominent phenomenon especially in the American context, but at least in some parts of Europe, it does not constitute a radical change at all. Of course, the recent developments in data gathering and data analysis that are linked to the present emphasis on empirical research have opened up new perspectives in these research traditions as well.

Although the use of empirical evidence is considered to an increasing extent a prerequisite for usage-based linguistics, it is not always regarded as a *sufficient* condition for an empirical study. For example, in their discussion of corpora use in linguistic research, Tummers, Heylen, and Geeraerts (2005, 234) argue that usage-based analyses are divisible into two categories, corpus-illustrated and corpus-based, which differ from each other with respect to the type of empirical analyses they employ. *Corpus-illustrated* research uses corpora with the sole purpose of providing examples that prove the existence of certain phenomena. Although this can be regarded as an improvement compared to the

previous introspective approach, there are two problems related to this type of corpus use. First, the mere corpus presence of a phenomenon does not tell us anything about its frequency, which could nevertheless be statistically marginal in comparison with the alternatives. Second, in a corpus-illustrated approach, the analysis of examples is still largely intuitive, and this cannot be considered the most effective way of dealing with corpus data (ibid. 235).

As a result, Tummers, Heylen, and Geeraerts (2005, 235) suggest that what advanced corpus research needs is quantification and statistical analysis coupled with a careful operationalization of hypotheses. This type of research, referred to as *corpus-based*, gives empirical evidence a central position in the analysis and aims at revealing significant tendencies in actual language use as represented by the data. Geeraerts (2010, 72–73) maintains that empirical research necessarily involves data-driven techniques and quantitative methods, since larger amounts of data are likely to yield more reliable conclusions than a collection of a few isolated observations. Quantification together with statistical analyses are needed in order to determine whether the phenomenon observed is frequent or marginal and to check whether the observations made on the data are based on convention or pure chance. Geeraerts (ibid.) underlines, however, that quantification is *not* the core element of empirical research. Instead, what is essential is the formulation of hypotheses and their operationalization into a testable form that allows linguistics to produce more trustworthy results. For example, Stefanowitsch (2010) identifies four basic ways of defining meaning in cognitive linguistics literature (i.e. meaning as concept, meaning as proposition, meaning as reference, and meaning as context of use). Each of these has different consequences on the research design and on how meaning may be operationalized in various empirical settings. The formulation of hypotheses and the way they are reflected in theory formation might, in fact, constitute the single most important question in present day linguistics. According to Wasow and Arnold (2005), feedback from empirical findings is not reflected enough in theory formation.

The dichotomy between corpus-illustrated and corpus-based approaches is admittedly an effective tool in arguing for cognitive linguistics to make more use of the recent technical developments and large corpora in testing claims within this theoretical framework. It leaves, however, a lot of current corpus-related work outside the classification. In present day linguistics, there is an abundance of *qualitative* corpus research that does not meet the criteria for corpus-based studies presented by Tummers, Heylen, and Geeraerts (2005); still, it cannot be characterized as using

corpus data for merely illustrative purposes but rather relies on meticulous qualitative analysis based on the relevant characteristics of the data. Tummers, Heylen, and Geeraerts (ibid. 238) themselves point out that rather than a strict dichotomy this is a question of a continuum, with corpus-illustrated and corpus-based approaches representing the endpoints.

The terminology used for the spectrum of corpus-related research is varied and unestablished, and covers at least the terms *corpus-based, corpus-driven* (Tognini-Bonelli 2001)*, corpus-oriented, corpus-exemplified* (Glynn 2014), and *corpus-illustrated* (Tummers, Heylen & Geeraerts 2005). The terms *corpus-exemplified* and *corpus-illustrated* are generally used for studies that use introspection as their main form of analysis but offer some examples from authentic data to illustrate the matter. The terms *corpus-based* and *corpus-oriented* usually refer to research in which planning and starting points may rely on introspection but where corpus data are used for systematic quantitative and/or qualitative investigation of a large number of examples that could not be achieved through introspection only. The term *corpus-driven* is often reserved to the least introspective type of research that takes the corpus as its starting point in investigating "what comes out of the data" in a bottom-up manner, typically using statistical methods. In addition to these specific meanings, *corpus-based* can also be used more loosely as a general upper-level term to cover all these types of corpus-related work, and this is the way the term is mainly used in this book as well.

So, there is a current trend in cognitive linguistics towards the use of empirical methods and this trend explicitly calls for more methodological rigor in the discipline (e.g. Tummers, Heylen & Geeraerts 2005; Geeraerts 2006a; Gonzalez-Marquez 2007; Heylen, Tummers & Geeraerts 2008; Glynn & Fischer 2010). As we have seen, it has been argued that instead of constructed examples we would need to move towards using *empirical, interpersonally observable data* (Wasow & Arnold 2005, 1482) and instead of subjective, untestable claims we should focus on producing *empirical, testable results* (Glynn 2010, 2). This demand for cognitive linguistics to reduce subjectivity and increase the use of empirical methods has left many "armchair linguists" at a defensive position. Does introspection any longer play a role in the methodology of linguistics?

Methodological status of introspection in usage-based linguistics is currently under debate. Some characterize it as the least reliable method for studying the actual language use (e.g. Tummers, Heylen & Geeraerts 2005, 226). It has been shown, for example, that semantic analyses made by different researchers on the same subject tend to give different results (Sandra & Rice 1995). The intuitions of different researchers do not

necessarily coincide. Arppe and Järvikivi (2007a) point out that, even if we cannot *always* rely on intuition, this does not mean we can *never* do so. In fact, intuitive knowledge of language might also be considered empirical because it is acquired and perpetually adjusted in a process of everyday language use within a language community. Itkonen (2003, 45–46) defends this position and makes a terminological distinction between intuition and introspection, with linguistic *intuition* referring to intersubjectively shared knowledge by speakers of a language and *introspection* consisting of subjective, personal associations of a particular speaker. The intersubjective nature of language and the way it should be reflected in linguistic analysis is an important topic in present day cognitive linguistics (the so-called Social Turn, e.g. Zlatev et al. 2008; Kristiansen & Dirven 2008; Möttönen 2016). (Chapter 9 in this book explicitly addresses this problem and suggests a methodology for incorporating the intersubjective knowledge of interlocutors.) Although the distinction between intersubjective and subjective knowledge suggested by Itkonen is appealing and might offer a means to tackle the question of the methodological status of introspective evidence, we will in this book, for the convenience of the reader, go on using the term *introspection* in a fashion in which it has conventionally been used in cognitive linguistics literature (e.g. Talmy 2007). By introspection we mean a linguistic method consisting of "conscious attention directed by a language user to particular aspects of language as manifest in her own cognition" (ibid., xii).

It is worth noticing that asking for more empirical research in cognitive linguistics does not mean that the role of introspection in the methodology of linguistics is denied. On the contrary, introspection is vital for empirical research in many respects. First, theory formation requires pure conceptual analysis (Geeraerts 2006a, 34). Without insightful theoretical claims, there would not be enough material for corpus linguists and experimental researchers to test. Second, as native speakers of a language, we have a lot of intuitive knowledge on how our language works, and this is a valuable tool in formulating research hypotheses (Arppe & Järvikivi 2007a). Third, subjective decisions are typically present in data annotation because, in most cases, categories are not objectively given but consist of decisions made by the researcher(s) analyzing the material (Glynn 2010, 240). Many popular collocation-based approaches, such as collostructional analysis or constructional profiles (e.g. Stefanowitsch & Gries 2003; Gries & Stefanowitsch 2004; Janda & Solovyev 2009) are based on manual analyses of both objectifiable (grammatical) variables, such as person and number for verbs and count or mass for nouns, and non-objectifiable

variables, such as various semantic features. Fourth, results do not simply fall out of the data but always require interpretation by the researcher (Geeraerts 2010, 72). As a consequence of all this, embracing the usage-based approach to language does not in the end rule out the so-called armchair linguistics. The division of labor between linguists doing empirical research and armchair linguists is *economical* and releases resources for maximal investigation of a specific area of research.

Empirical research is a valuable part of the methodology of linguistics and offers new ways of testing hypotheses but it should not be taken as an ultimate proof for truth, nor as the only reliable scientific method. Excessive respect for empirical methods may lead to the fallacy of objectivism as opposed by Lakoff and Johnson (1980). As Glynn (2010, 240) remarks, it is a mistake to believe that (quantitative) empirical methods per se are more objective than traditional linguistic analyses. We need to remember that quantitative and qualitative studies are both theory-dependent. In all scientific disciplines, non-empirical claims are necessary for data gathering and formulating research questions, for example. The methods adopted (even when it comes to technical instruments) are also based on theoretical assumptions that determine the kinds of results one can get (Latour & Woolgar 1986 [1979]). No observation is pure but oriented by the theoretical framework of research that offers concepts for the examination of the phenomenon (Niiniluoto 1980, 224–225). As an example of such theory dependence, Stefanowitsch (2010, 359–360) gives the notion of *length*. In the study of language, the length of an utterance could be measured by using different kinds of digital or analogic devices that may give slightly different results, but it can also be measured in terms of different theoretically relevant units such as the number of phonemes, morphemes, syllables or words. There is no single correct way of measuring the length of an utterance but it is always a matter of theoretical choice. Therefore, according to modern concepts of confirmation and refutation, even corroborated hypotheses remain provisional, and negative evidence does not necessarily mean that the original hypothesis was wrong; it could just as well be that the way of testing the assumption was not appropriate (Lakatos 1970).

The occasional false conceptions on objectivity in science are problematic because they prevent recognizing the performance of different methodologies. Both those promoting the use of (quantitative) empirical methods and those defending the use of introspection have sometimes been accused of being too optimistic about the performance of the methods they have chosen. According to Geeraerts (2006a, 28), the prevalence of meaning over form that is elementary in cognitive

linguistics might impede us from advocating empirical research, since introspection is sometimes viewed as the only reliable method for the study of semantic phenomena. Wasow and Arnold (2005, 1482) warn about overreliance on introspection as a linguistic method. Introspection has its place in the methodology of linguistics but there is no reason to accord it a privileged status over other methodologies (Arppe & Järvikivi 2007a). What is equally problematic is the sometimes overenthusiastic attitude of those who promote empirical methods in linguistics but at the same time lack knowledge of the philosophy of science or the research design (Stefanowitsch 2010, 358). One solution to this problem would be to offer more training in empirical techniques, and the relationship between the different methods and the philosophy of science in linguistics. A lot of work remains to be done in order to fully understand the strengths and weaknesses related to specific methods and the aspects of language they are best equipped to investigate.

Rather than arriving at ultimate objectivity, there is a mutually beneficiary division of labor between different methods; they are adapted to different research tasks and give answers to different types of research questions. Newman (2010, 94) calls for tolerance for both qualitative and quantitative research, and many others highlight the importance of multi-methodological research (Arppe & Järvikivi 2007a, 2007b; see also the specific comments made by Arppe in Arppe et al. 2010; Gries & Divjak 2010). This perspective could not be emphasized more in the current situation where researchers are beginning to talk about the "Quantitative Turn" in cognitive linguistics (e.g. Janda 2013) and to assume that the future of cognitive linguistics belongs to quantitative methods (e.g. Geeraerts 2006a; Dabrowska & Divjak 2015). This book promotes the idea that in addition to quantitative methods we also need qualitative methods and new kinds of currently evolving empirical data to be better equipped to meet the needs of present day research. Therefore, empirical research should not be defined solely in terms of research using quantitative methods and statistical techniques. Instead, a wider perspective and the use of various types of empirical methods and data are necessary, as the articles in this book indicate.

In the present volume, a multi-methodological approach to linguistic research is promoted. The contributions to the volume consist of nine empirical case studies relying on various kinds of corpus data, experimental data or a combination of both types of empirical evidence. These case studies present both large-scale and small-scale analyses of various linguistic topics using different quantitative, qualitative and multi-methodological approaches (see Chapter 6 in this book for the advantages

of multi-methodological approaches). It is our conviction that the use of the entire spectrum of empirical methods, each of which is adapted to its own particular research topic, is the most efficient way of dealing with such a complex and fascinating phenomenon as language.

2. Contributions to the volume

The papers in this volume present a wide range of topics discussed in Finnish cognitive linguistics research ranging from transitive partitive subject construction to idiom usage, from clausal aspect to verb constructions, and from collocations to term usage without forgetting the perspectives of language development and translation studies. The languages investigated include English, Finnish, German, Russian and Spanish.

The development of information technology in the 21st century has provided us with tools for the collection and analysis of large corpora. Even though large ready-made corpora with linguistic tags exist for many languages, they are not always suitable for all linguistic research purposes and this is why researchers still often need to collect their own material. Both of these approaches have been used in the corpus-oriented studies collected here. The papers put the usage-based approach into practice by investigating authentic data from the Internet, from online corpora, or from other textual material, or they rely on data collected directly from language users. Many of the papers address the benefits and possible problems related to corpus-based studies from the cognitive perspective, while some combine other approaches to complement their corpus-oriented studies, and others exploit various types of data and methods to answer the questions posed in them. The methods used in the analysis of the data vary from distributional analysis to reaction time tests, from quantitative corpus-linguistic methods to more traditional qualitative methods, such as conceptual and semantic analysis, without forgetting the methods peculiar to cognitive linguistics, such as construction analysis and grammatical profiling. The bulk of the articles in this book could be called corpus-based, corpus-oriented, or corpus-related (Part 1), while the rest are based on data from language informants (Part 2).

2.1. Corpus-based studies (Part 1)

The first part of the volume comprises a series of articles dedicated to research based on different kinds of text corpora. The first three studies combine both qualitative and quantitative techniques, while the rest of the

chapters are dedicated to more traditional qualitative linguistic analyses. Four of the chapters are examples of actual case studies on the use of textual corpora. This section concludes with a more theoretical paper addressing methodological problems related to the use of textual corpora in the study of specific research topics, such as idioms.

The first Chapter in this collection, "Distributional semantic of the Partitive A Argument Construction in Finnish" by **Tuomas Huumo**, **Aki-Juhani Kyröläinen**, **Jenna Kanerva**, **M. Juhani Luotolahti**, **Tapio Salakoski**, **Filip Ginter**, and **Veronika Laippala**, combines both qualitative and quantitative approaches to the analysis of data that has been collected automatically from the web for the Finnish Internet Parsebank, a constantly growing large-scale corpus that at the moment of the study contained 3.7 billion tokens of morphologically and syntactically annotated Internet-Finnish. Compared to traditional corpus data, the Finnish Internet Parsebank offers certain advantages when studying an infrequent and elusive phenomenon, such as the Partitive A Construction, partly because of its size and diversity but also because of the advanced search system that is offered with the corpus. After several steps of corpus queries, the researchers obtained 1,437 potential instances of the construction that were then analyzed in terms of their argument structures, co-occurring verbs and frequencies. The quantitative analysis, in particular, opens up new methodological avenues by applying the semantic vector model to the analysis of argument structure constructions with the help of similarity-based relations.

In "Changes in figure–ground alignment in translation: Condensing information in subtitling", **Jukka Mäkisalo** and **Marjatta Lehtinen** investigate the condensation or omission of information items in subtitling. The study continues the work conducted earlier by the authors (Mäkisalo & Lehtinen 2014) and looks for more evidence to corroborate their Cognitive Retention Hypothesis: "When describing translation from a source text to a target text, it is possible to distinguish linguistic and cognitive levels, and it is the cognitive level (cognitive model of the state of affairs) that is primarily retained in translation". The method that the authors employ in tracing the changes at both linguistic and cognitive levels includes the analysis of the Figure-Ground structure from the perspective of Radical Construction Grammar. The authors conclude that the results of this study re-confirm the previous findings, but they also find some differences in information retention related to the different textual genres examined in these two studies.

In "How light can a light verb be? Predication patterns in V + NP constructions in English, Finnish, German, and Russian", **Marja**

Nenonen, Juha Mulli, Alexandre Nikolaev, and **Esa Penttilä** discuss "light verbs", i.e. verbs with little lexical meaning such as *take, give* and *put* in the four languages mentioned. The authors use both dictionary data consisting of c. 3,500 idioms (English: c. 550, Finnish: c. 830, German: c. 1,300, Russian: c. 800) and corpus data of newspaper texts from each investigated language. The analysis is both quantitative and qualitative and concentrates on analyzing the distribution between the "light" and "heavy" uses of the verbs in the four languages but also discusses the elements that determine the degree of "lightness". The authors suggest further categories that the method they have used should take into account in future studies.

In the Chapter "Expressing boundaries: A contextual approach to clausal aspect in Finnish", **Salla Nurminen** provides a detailed usage-based analysis of clausal aspect in Finnish, previously analyzed mostly on the basis of fabricated examples. To that end, the author examines 2,123 clauses from newspaper texts with the verb *olla* 'be', *tehdä* 'do, make', and *tulla* 'come, become' as the main verb from the perspective of Cognitive Grammar. Nurminen shows how the study of aspect benefits from the use of authentic corpus data and argues that, in consequence, several ideas suggested by previous studies operating with fabricated examples need to be modified. The author concludes that, in actual language use, there are multiple semantic levels of boundedness and unboundedness affecting the aspectual interpretation of a given clause, and different aspectual layers including not only clause-internal but also clause-external bounding elements need to be taken into account.

In the Chapter "The natural translation of idiomatic constructions" **Esa Penttilä**, on the basis of Internet data, discusses a fascinating phenomenon that we meet in everyday Finnish, where speakers use novel unconventional phrases that are directly translated from English into Finnish and thus mix elements of their L2 into their L1. Penttilä calls this translation-based phenomenon "concealed code switching" and ponders on whether the creators of these unconventional loan translations could be regarded as "voluntary virtual emigrants" who use their L2 to such an extent that it begins to influence their use of L1. Penttilä also points out that in real-life data it is difficult to study such a phenomenon that is not widespread in language use in general but can be significant in certain user contexts. In addition to the use of Internet data, he suggests that information on the phenomenon could possibly be gained by using experimental methods.

2.2 Research based on information from language users (Part 2)

The second part of the volume is dedicated to studies based on data from language informants. This data is gathered using survey techniques and/or relying on experimental settings. The first three chapters in this section rely on a multi-methodological approach combining either survey data with corpus data (Granvik and Luodonpää-Manni) or eye-movements data with survey and corpus data (Kyröläinen, Porretta, and Järvikivi), while the last chapter is based on observational data.

In "The choice between generic scientific terms in linguistic research articles written in Finnish" **Milla Luodonpää-Manni** introduces the basics of mixed-method analysis and discusses the use of terms meaning 'theory', 'hypothesis', 'model', 'method', 'thesis', 'approach', 'framework', 'starting point', 'point of view', 'orientation', 'manner of examination' and 'method' in the Finnish linguistic discourse. Based on a sequential analysis of material from a questionnaire for linguists and a corpus of scientific texts, Luodonpää-Manni concludes that scientific terms are ambiguous and are used differently by different users. The differences in use are related both to different but overlapping conceptualizations of the basic units of research and the traditions in different disciplines.

Anton Granvik's article in Chapter seven, "Topic-marking prepositions in Spanish: Contrasting corpus and questionnaire data in the analysis of prepositional synonymy", combines a corpus study with a questionnaire survey in an attempt to explain the possible semantic profile of four topic-marking Spanish prepositions, *de, sobre, acerca de*, and *en torno a*, all of which are regarded as more or less synonymous and interchangeable. The multinomial logistic regression analysis of randomly extracted instances from the 100 million word Corpus del español reveals clear differences between two of these prepositions, while the other two are more or less similar in both usage and distribution. With the help of questionnaire surveys that followed, it was then possible to confirm that – for native Spanish speakers – the four topic-marking prepositions form a hierarchy on the basis of how freely they are used in similar contexts.

Chapter eight, "The role of morphological verb constructions in processing Russian reflexive verbs" by **Aki-Juhani Kyröläinen, Vincent Porretta**, and **Juhani Järvikivi**, presents an experimental study where the data devised for an online lexical decision task was based on the results of frequency- and dispersion-based analyses of the Russian National Corpus. The corpus-based part of the study was based on principal component analysis, which showed that three principal components were enough to capture most of the variance in the data pointing out that in addition to the relation between the base and the reflexive verb, which has been regarded

as crucial for the morphological construction traditionally, also the structuring of the whole seems to play an important role. When the experimental data that was gathered from participants' lexical decisions on 160 reflexive Russian verbs was then analyzed with the help of generalized additive mixed-effects model, the results showed that these three components significantly modulated the measured reaction times. Viewed in the framework of construction grammar, the study thus benefits from the use of multiple methods and offers new information on the processing of Russian morphological constructions.

In the final Chapter of this volume, entitled "Development of early directives in Finnish: A usage-based approach", **Maija Surakka** examines directive constructions of a Finnish-speaking boy from two and a half to three years of age. To this end, the author analyses utterances of 'want', most of which carry a tone of directivity, and focuses on multimodal meaning structures that the child has assembled by speaking, moving, touching, reacting or demonstrating things in the usage-event. The method employed by Surakka in tracing these multimodal expressions combines the observation of the child's multimodal behavior in the context of family interaction with the perspectives offered by Cognitive Grammar, usage-based approach to language acquisition, and intersubjective nature of meaning. The author argues that such a pragma-semantic analysis enables her to examine the conceptual development that precedes explicit language acquisition and shows how the child completes the partly unelaborated meaning structures with nonverbal grounding elements.

Before we conclude our introduction, we would like to express our warmest gratitude to the board members of the Finnish Cognitive Linguistics Association (FiCLA), Markus Hamunen, Ilona Herlin, Mikko Höglund, Minna Jaakola, Juha Mulli, Helka Riionheimo, and Katja Västi, for their comprehensive cooperation and support with the publication of this volume. Furthermore, this book would not have been possible without the valuable contribution of the anonymous reviewers on each article, and we thank them for their insightful feedback. Finally, our sincere thanks go to the staff at Cambridge Scholars Publishing for their interest in this project and for their professional support and advice.

Notes

[1] We would like to address our special thanks to Markus Hamunen for his insightful comments on this section of the Introduction.

[2] For this reason, some researchers do not consider the results of experiments as empirical evidence (e.g. Sampson 2005). For a detailed commentary on Sampson's work, see Arppe and Järvikivi (2007a).

References

Arppe, Antti, and Juhani Järvikivi. 2007a. "Take empiricism seriously! In support of methodological diversity in linguistics." *Corpus Linguistics and Linguistic Theory* 3(1): 99–109. DOI: 10.1515/CLLT.2006.007.

Arppe, Antti, and Juhani Järvikivi. 2007b. "Every method counts: Combining corpus-based and experimental evidence in the study of synonymy." *Corpus Linguistics and Linguistic Theory* 3: 131–159. DOI: 10.1515/CLLT.2007.009.

Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert, and Arne Zeschel. 2010. "Cognitive corpus linguistics: Five points of debate on current theory and methodology." *Corpora* 5: 1–27. DOI: 10.3366/cor.2010.0001.

Croft, William, and D. Allan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Divjak, Dagmar, Natalia Levshina, and Jane Klavan. 2016. "Cognitive Linguistics: Looking back, looking forward. *Cognitive Linguistics* 27(4): 447–463. DOI: 10.1515/cog-2016-0095.

Evans, Vyvyan, and Melanie Green. 2006. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Geeraerts, Dirk, Stefan Grondelaers, and Peter Bakema. 1994. *The Structure of Lexical Variation. Meaning, Naming, and Context*. Berlin/New York: Mouton de Gruyter. DOI: 10.1515/9783110873061.

—. 2006a. "Methodology in Cognitive Linguistics." In *Cognitive Linguistics: Current Applications and Future Perspectives*, edited by Gitte Kristensen, Michel Archard, René Dirven, and Francisco J. Ruiz de Mendoza Ibanez, 21–50. Berlin/New York: Mouton de Gruyter.

—. 2006b. "Introduction: A rough guide to Cognitive Linguistics." In *Cognitive Linguistics: Basic Readings*, edited by Dirk Geeraerts, 1–28. Berlin/New York: Mouton de Gruyter.

Geeraerts, Dirk, and Hubert Cuykens. 2007. "Introducing Cognitive Linguistics." In *The Oxford Handbook of Cognitive Linguistics*, edited by Dirk Geeraerts and Hubert Cuykens, 3–21. Oxford: Oxford University Press.

Geeraerts, Dirk. 2010. "The doctor and the semantician." In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, edited by Dylan Glynn and Kerstin Fischer, 63–78. Berlin/New York:

Mouton de Gruyter.
Glynn, Dylan, and Kerstin Fischer, eds. 2010. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. Berlin/New York: Mouton de Gruyter.
Glynn, Dylan. 2010. "Corpus-driven Cognitive Semantics. Introduction to the field." In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, edited by Dylan Glynn and Kerstin Fischer, 1–41. Berlin/New York: Mouton de Gruyter.
Glynn, Dylan, and Justyna A. Robinson, eds. 2014. *Corpus Methods for Semantics. Quantitative Studies in Polysemy and Synonymy*. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/hcp.43.
Glynn, Dylan. 2014. "Polysemy and synonymy: Cognitive theory and corpus method." In *Corpus Methods for Semantics. Quantitative Studies in Polysemy and Synonymy*, edited by Dylan Glynn and Justyna A. Robinson, 7–38. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/hcp.43.01gly.
Gonzalez-Marquez, Monica, Irene Mittelberg, Seana Coulson, and Michael J. Spivey, eds. 2007. *Methods in Ccognitive Linguistics*. Amsterdam: John Benjamins. DOI: 10.1075/hcp.18.
Gries, Stefan Th. and Dagmar Divjak. 2010. "Quantitative approaches in usage-based Cognitive Semantics: Myths, erroneous assumptions, and a proposal." In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, edited by Dylan Glynn and Kerstin Fischer, 333–353. Berlin/New York: Mouton de Gruyter.
Gries, Stefan Th., and Anatol Stefanowitsch. 2006. *Corpora in Cognitive Linguistics. Corpus-based Approaches to Syntax and Lexis*. Berlin/New York: Mouton de Gruyter. DOI: 10.1515/9783110197709.
Fischer, Kerstin. 2010. "Quantitative methods in Cognitive Semantics. Introduction to the volume." In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, Dylan Glynn and Kerstin Fischer, 43–59. Berlin/New York: Mouton de Gruyter.
Herlin, Ilona. 1994. "Konjunktion monikäyttöisyys: *kun*." In *Näkökulmia polysemiaan*, edited by Pentti Leino and Tiina Onikki, 139–163. Helsinki: Helsingin yliopiston suomen kielen laitos.
Heylen, Kris, José Tummers, and Dirk Geeraerts. 2008. "Methodological issues in corpus-based Cognitive Linguistics." In *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*, edited by Gitte Kristiansen and René Dirven, 91–128. Berlin/New York: Mouton de Gruyter. DOI: 10.1515/9783110199154.2.91.
Hovdhaugen, Even, Fred Karlsson, Carol Henriksen, and Bengt Sigurd.

2000. *The History of Linguistics in the Nordic Countries*. Jyväskylä: Societas Scientiarum Fennica.
Häkkinen, Kaisa. 2008. *Suomen kielen historia 2: Suomen kielen tutkimuksen historia*. Publications of the Department of Finnish and General Linguistics of the University of Turku.
Itkonen, Esa. 2003. *What is Language? A Study in the Philosophy of Linguistics*. Turku: Åbo Akademis tryckeri.
Janda, Laura, ed. 2013. *Cognitive Linguistics: The Quantitative Turn. The Essential Reader*. Berlin/New York: Mouton de Gruyter.
Kristiansen, Gitte, and René Dirven, eds. 2008. *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*. Berlin/New York: Mouton de Gruyter.
Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes." In *Criticism and the Growth of Knowledge*, edited by Imre Lakatos, I., and Alan Musgrave, 91−195. Cambridge: Cambridge University Press.
Lakoff, George, and Mark Johnson 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.
Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
—. 1990. "The Invariance Hypothesis: Is abstract reason based on image-schemas?" *Cognitive Linguistics*, 1(1): 39–74.
Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar – Volume I: Theoretical Prerequisites*. Stanford University Press.
—. 1988. "A usage-based model." In *Topics in Cognitive Linguistics*, edited by Brigida Rudzka-Ostyn, 127–161. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/cilt.50.06lan.
—. 1999. *Grammar and Conceptualization*. Berlin/New York: Mouton de Gruyter.
Latour, Bruno, and Steve Woolgar. 1986 [1979]. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press.
Latvala, Salu. 1895. "Lauseopillisia havaintoja Luoteis-Satakunnan kansankielestä." *Suomi* III (12): 1–88.
Möttönen, Tapani. 2016. *Construal in Expression: Intersubjective Approach to Cognitive Grammar*. PhD diss., University of Helsinki. https://helda.helsinki.fi/handle/10138/159436.
Niiniluoto, Ilkka. 1980. *Johdatus tieteen filosofiaan – käsitteen- ja teorianmuodostus*. Helsinki: Otava.
Newman, John. 2010. "Balancing Acts: Empirical pursuits in Cognitive Linguistics." In *Quantitative Methods in Cognitive Semantics: Corpus-*

Driven Approaches, edited by Dylan Glynn and Kerstin Fischer, 79–99. Berlin/New York: Mouton de Gruyter.

Rice, Sally, and John Newman, eds. 2010. *Empirical and Experimental Methods in Cognitive/Functional Research*. Stanford: CSLI.

Rudzka-Ostyn, Brygida, ed. *Topics in Cognitive Linguistics.* Amsterdam, Philadelphia: John Benjamins. DOI: 10.1075/cilt.50.

Sampson, Geoffrey. 2005. "Quantifying the shift towards empirical linguistics." *International Journal of Corpus Linguistics* 10(1): 15–36. DOI: 10.1075/ijcl.10.1.02sam.

Sandra, Dominiek, and Sally Rice. 1995. "Network analyses of prepositional meaning: Mirroring whose mind – the linguist's or the language user's?" *Cognitive Linguistics* 6(1): 89−130. DOI: 10.1515/cogl.1995.6.1.89.

Setälä, E.N. 1883. "Lauseopillinen tutkimus Koillis-Satakunnan kansankielestä." *Suomi* II (16): 1–235.

Stefanowitsch, Anatol. 2010. "Empirical Cognitive Semantics: Some thoughts." In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, edited by Dylan Glynn and Kerstin Fischer, 355–380. Berlin/New York: Mouton de Gruyter.

Talmy, Leonard. 1986. *The Relation of Grammar to Cognition*. L.A.U.D.T.. Series A; paper no. 165. L.A.U.D. Linguistic Agency, University of Duisburg.

Talmy, Leonard. 2007. "Foreword." In *Methods in Cognitive Linguistics*, edited by Monica Gonzalez-Marquez, Irene Mittelberg, Seana Coulson, and Michael J. Spivey, XI–XXI. Amsterdam: John Benjamins. DOI: 10.1075/hcp.18.03tal.

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

Tomasello, Michael. 2009. "The usage-based-theory of language acquisition." In *The Cambridge Handbook of Child Language*, edited by Beth Levin, 69–87. Cambridge: Cambridge University Press.

Tummers, José, Kris Heylen, and Dirk Geeraerts. 2005. "Usage-based approaches in Cognitive Linguistics: A technical state of the art." *Corpus Linguistics and Linguistic Theory* 1(2): 225–261. DOI: 10.1515/cllt.2005.1.2.225.

Wasow, Tom, and Jennifer Arnold. 2005. "Intuitions in Linguistic Argumentation." *Lingua* 115(1): 1481–1496. DOI: 10.1016/j.lingua.2004.07.001.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. New York: MacMillan.

Zlatev, Jordan, Timothy P. Racine, Chris Sinha, and Esa Itkonen, eds. 2008. *The Shared Mind: Perspectives on Intersubjectivity*. Amterdam/ Philadelphia: John Benjamins. DOI: 10.1075/celcr.12.13zla.

PART ONE:

CORPUS-BASED STUDIES