Wei-lun Lu,
Naděžda Kudrnáčová
and Laura A. Janda (eds.)

Corpus Approaches to Language, Thought and Communication

Corpus Approaches to Language, Thought and Communication

Benjamins Current Topics

Special issues of established journals tend to circulate within the orbit of the subscribers of those journals. For the Benjamins Current Topics series a number of special issues of various journals have been selected containing salient topics of research with the aim of finding new audiences for topically interesting material, bringing such material to a wider readership in book format.

For an overview of all books published in this series, please see *benjamins.com/catalog/bct*

Volume 119

Corpus Approaches to Language, Thought and Communication
Edited by Wei-lun Lu, Naděžda Kudrnáčová and Laura A. Janda

These materials were previously published in *Review of Cognitive Linguistics* 17:1 (2019).

Corpus Approaches to Language, Thought and Communication

Edited by

Wei-lun Lu
Masaryk University

Naděžda Kudrnáčová
Masaryk University

Laura A. Janda
University of Tromsø

John Benjamins Publishing Company

Amsterdam / Philadelphia

Table of contents

Deep dives into big data

Best practices for synthesis of quantitative and qualitative analysis in Cognitive Linguistics

Laura A. Janda,[1] Naděžda Kudrnáčová[2] and Wei-lun Lu[2]
[1] UiT The Arctic University of Norway | [2] Masaryk University

The six articles in this volume are exemplary studies that profile the current state-of-the art in cognitive linguistics, namely the synthesis of quantitative and qualitative linguistic analysis. This introduction is an opportunity to take stock of where cognitive linguistics started out, what kinds of approaches have been developed, and how we have arrived at a synthesis in which empirical exploration informs the interpretation of language phenomena.

In the 1980s, cognitive linguistics sprang from the rejection of the assumption made in generative linguistics that language-related cognition is separate from cognition in general. This rejection included a series of corollaries that were needed to buttress that assumption, such as the existence of language universals, a "language module" in the brain, underlying forms, and *poverty of the stimulus*, the idea that linguistic input is insufficient to support language acquisition (Chomsky, 1980). Cognitive linguistics rests instead on the more conservative assumption that all language phenomena can be explained in terms of general cognitive mechanisms (Langacker, 1987, 1991a, 1991b), and seeks to explain linguistic behaviors in terms of what is independently established by psychologists and neurologists about brain functions. Thus instead of narrowing the task of a linguist to the investigation of an internal grammar (such as *langue, competence, i-language*, etc.) that cannot be directly observed, cognitive linguistics opened the way for the study of language *use* (such as *parole, performance, e-language*, etc.).

In addition, cognitive linguistics is a *usage-based* framework, which views language as an aggregate of usage events. This perspective, which continues to be a driving force in cognitive linguistics, has motivated a series of tendencies within our framework. One early tendency was to investigate the structure of linguistic meaning, modeled after research in psychology showing that human beings organize concepts in terms of prototypes and radial categories. Following this line of thought, metaphor, metonymy, and blending got considerable attention for their role in structuring radial categories (Lakoff & Johnson, 1980; Lakoff,

1987; Fauconnier & Turner, 2002). As a result of such development, the 1980s and 1990s were characterized by the study of how meaning is grounded in the shared human experience of bodily existence, and how this experience is incorporated in image schemas and their extensions. Crucial during this period also was the exploration of how languages differ from each other. While there is a repertoire of basic experiences that are shared (GRAVITY, SYMMETRY, FIGURE-GROUND, COUNT-MASS, SOURCE-PATH-GOAL), they motivate, rather than determine, language phenomena, accounting for the ubiquity of cross-linguistic variation (Croft, 2001). If we take COUNT-MASS for example, the general notion that there are some items of realia that exist as units as opposed to others that are substances is a part of the grammar of every language. But the boundary between count and mass can be very different in different languages. In Russian, *gorox* 'peas' and *izjum* 'raisins' are grammatically singularia tantum words that refer to substances, although in English the same realia are treated as plural countable items, and of course there are languages such as Yucatec Mayan and Chinese where the difference between count and mass is signaled not by plural but by the use of classifier constructions (see for example, Dosedlová and Lu, this volume). For reasons like these, motivation has come to be recognized as more important than prediction in the framework of cognitive linguistics.

Cognitive linguistics takes the central role of meaning in language seriously, and links meaning directly to form, namely as "symbolic units" pairing a phonological pole with a semantic pole as defined by Langacker (1987, p. 58). In keeping with the discovery that the basic units of language are neither those that are smallest (such as phonemes) nor those that are largest (such as discourse), the symbolic unit that has emerged as the most common focus of study is the *construction* (Goldberg, 1995, 2006; Croft, 2001): any conventionalized pairing of form and meaning. Construction grammar has become a core pursuit in cognitive linguistics. Language is understood to be composed of constructions, at various levels of complexity, and researchers are now describing languages in terms of *constructicons* (Lyngfelt, Borin, Ohara, & Torrent, 2018).

The usage-based perspective of cognitive linguistics has always been data-friendly, poised to take advantage of the digital resources and statistical software that have seen enormous expansion in the age of big data since the turn of the twenty-first century. Cognitive linguists now routinely turn to corpora to extract data, identify trends, and feed statistical models. Experimental studies are also on the rise, often inspired by or carried out in tandem with corpus studies. Quantitative analysis has become an essential tool.

When cognitive linguists face research questions today, they have an assortment of ways to address them. For many languages, they can fetch large quantities of examples from corpora of millions or billions of words that have been tagged

for the purposes of linguistic research. Even some of the world's smaller languages have electronic corpora (for example, the KORP corpus of North Saami, a language spoken by only 20,000 people, currently contains over 32 million words, or the NTU Corpus of Formosan Languages, reported in Su, Sung, Huang, Hsieh, and Lin [2008]). Tagging facilitates corpus-based work on construction grammar by making it possible to track the behaviors not just of words, but of constructions. And the traditional methods of probing the internal structure of radial categories via metaphor, metonymy, and blending persist, now enhanced by data extraction tools that make it possible for the linguist to strategically target the most valuable material for in-depth analysis.

This volume showcases studies in which researchers take deep dives into material that emerges from modern digital corpora and apply methods of analysis of constructions and meaning structure from cognitive linguistics. Gathered below here are brief synopses of those contributions.

Laura Janda's study discusses the relevant aspects of the quantitative turn in cognitive linguistics, with comprehensive scope and richly informative content. Janda surveys the history of the quantitative turn (based on the articles published in *Cognitive Linguistics*, the flagship journal of the field, from its inaugural volume in 1990 to the volume in 2017) and identifies factors whose confluence has facilitated the quantitative turn: the usage-based model of language in the cognitive linguistics framework, the advent of electronic language resources, and the development of statistical software. Janda's article also provides an analytical comparative overview of quantitative methods in cognitive linguistics research, and attends to the relationship between them and introspection. In addition, it provides a perspicacious and useful discussion of the opportunities and dangers that the quantitative turn poses, and delineates the possible future development of quantitative methodology. This article will be of interest not only for cognitively oriented linguists but also for linguists adhering to a variety of theoretical approaches.

Vladan Pavlović explores the use of N1 V (*for*) + N2 + *to*-infinitive constructions in American English, using data from two massive corpora, the Corpus of Historical American English (COHA) and the Corpus of Global Web-based English (GloWbE). The author argues that the patterns observable in the data result from an interplay between the semantics of the constructions, the lexical semantics of the main verbs, and the dominant communicative style. In order to test the claim, the study compares synchronic data for American English, British English, Indian English and Hong Kong English on the basis of GloWbE. The analysis is innovative, combining insights from cognitive linguistics, verbal semantics and models of cross-cultural communication, and brings convincing evidence on the usefulness of massive corpora in linguistic research.

Kudrnáčová's article contributes to a hitherto relatively unexplored area, a fine-grained cross-linguistic analysis of differences in manner of motion verbs. Based on data retrieved from InterCorp, a synchronic parallel translation corpus, Kudrnáčová looks into the differences in the construal of walking between the English verb *walk* and its nearest Czech counterparts, i.e. *jít* and *kráčet*. Despite their apparent commonalities, the verbs in question do not construe the most prototypical type of human locomotion in the same way. As opposed to *jít*, both *walk* and *kráčet* foreground the segmentation of the movement into individual kinetic quanta. Nevertheless, while *kráčet* bears reference to the actor's experiential self and is endowed with an evaluative potential, this possibility is not available for *walk* or *jít*. The contribution, in other words, shows how the Czech language lacks an exact semantic counterpart of *walk*.

Drawing on data excerpted from the Czech National Corpus and the Balanced Corpus of Contemporary Written Japanese, Petra Kanasugi focuses on differences between Czech and Japanese in adnominal modification. While Czech tends to utilize adjectives for both classification and qualification, Japanese tends to express classification by compounding and to use a whole range of parts of speech for qualification. The author observes that part of speech membership thus often differs between the Czech and Japanese rendering of the same referential content. The author argues that parts of speech have schematic meaning which contributes to conceptualization and, further, that differences in part of speech membership result in different tendencies in meaning extension and the degrees of abstractness of expressions. Specifically, Czech adjectives in adnominal modification are more abstract and schematic while Japanese verbs in adnominal modification are more concrete.

Dosedlová and Lu's study examines the near-synonymy of different classifiers within one language. Drawing on data retrieved from the zhTenTen corpus, a corpus of simplified Standard Chinese built via web-crawling, this article provides a cognitive analysis of the semantic functions of Mandarin plant classifiers *kē* and *zhū*. The authors argue that the different constructional profiles of the two classifiers reflect different construals of partially overlapping conceptual contents invoked by the classifiers in question. They observe that the classifier *zhū* tends to modify objects of smaller size, but of larger quantity, which is not characteristic of *kē*. Accordingly, they conclude that the construal invoked by [QUANTIFIER]–[ZHU]–[NOUN] provides a higher resolution, and a more granular view of the scene linguistically elaborated, whereas [QUANTIFIER]–[KE]–[NOUN] does not share that preference.

Based on data retrieved from the NTU Corpus of Spoken Chinese, the study by Hsieh and Su investigates the use of *xiangshuo* 'think' as a complement-taking mental predicate in Taiwan Mandarin conversation. This study is innovative in

the scope of analysis and in testing out multiple theoretical frameworks, facilitating an approach to the issue from a broader perspective. The authors adopt the Interactional Construction Grammar approach, which incorporates interactional factors into Construction Grammar analysis to account for patterns that involve interpersonal functions and global contexts. They present the co-occurrence patterns of this verb with different subjects, and identify three sequential patterns in which *xiangshuo* most frequently occurs, including account-giving, contrast-projecting and involvement-constructing. The authors argue that the distributional patterns of subjects and particles that recurrently collocate with *xiangshuo* can be explained only by taking into account the sequential context and interactional function.

From the collection of papers, one theme is obvious: approaching language use in different contexts from different perspectives, each of the contributions in this volume presents its own unique take on the intertwined relationship between language, thought and communication, but however different these papers are, each of them makes a valid point in how a corpus method helps shed new light on an old issue, reflecting the usage-based nature of cognitive linguistic research.

References

Chomsky, N. (1980). *Rules and representations*. Oxford: Basil Blackwell.
https://doi.org/10.1017/S0140525X00001515

Croft, W. (2001). *Radical Construction Grammar*. Oxford: Oxford University Press.
https://doi.org/10.1093/acprof:oso/9780198299554.001.0001

Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalizations in language*. Oxford: Oxford University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
https://doi.org/10.7208/chicago/9780226471013.001.0001

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar*. Vol. I: *Theoretical prerequisites*. Stanford: Stanford University Press.

Langacker, R. W. (1991a). *Concept, image, and symbol: The cognitive basis of grammar*. Berlin: Mouton de Gruyter.

Langacker, R. W. (1991b). *Foundations of Cognitive Grammar*. Vol. II: *Descriptive application*. Stanford: Stanford University Press.

Lyngfelt, B., Borin, L., Ohara, K., & Torrent, T. T. (Eds.). (2018). *Constructicography: Construction development across languages*. Amsterdam: John Benjamins. https://doi.org/10.1075/cal.22

Su, L. I., Sung, L., Huang, S., Hsieh, F., & Lin, Z. (2008). NTU corpus of Formosan languages: A state-of-the-art report. *Corpus Linguistics and Linguistic Theory*, 4(2), 291–294. https://doi.org/10.1515/CLLT.2008.012

Quantitative perspectives in Cognitive Linguistics

Laura A. Janda
UiT The Arctic University of Norway

As a usage-based approach to the study of language, cognitive linguistics is theoretically well poised to apply quantitative methods to the analysis of corpus and experimental data. In this article, I review the historical circumstances that led to the quantitative turn in cognitive linguistics and give an overview of statistical models used by cognitive linguists, including chi-square test, Fisher test, Binomial test, t-test, ANOVA, correlation, regression, classification and regression trees, naïve discriminative learning, cluster analysis, multi-dimensional scaling, and correspondence analysis. I stress the essential role of introspection in the design and interpretation of linguistic studies, and assess the pros and cons of the quantitative turn. I also make a case for open access science and appropriate archiving of linguistic data.

Keywords: quantitative methods, statistical models, corpus data, experimental data, usage-based approach

1.   Introduction

The quantitative turn in cognitive linguistics is a force to reckon with. In this article, I track the history of our quantitative turn, which has been facilitated by a confluence of three factors: the usage-based nature of the cognitive linguistics framework, the advent of electronic archives of language data, and the development of statistical software. I give an overview of the types of statistical models cognitive linguists are turning to, illustrated by the kinds of research questions that are being asked and answered using quantitative tools. I also discuss the opportunities and dangers that we face now that we have taken our quantitative turn.

2.   What brought about the quantitative turn?

A survey of articles published in the journal *Cognitive Linguistics* gives us a per-spective on the quantitative turn in cognitive linguistics. Figure 1 presents the distribution of articles in the journal from its inaugural volume in 1990 through the most recent complete volume in 2017, according to whether or not they presented quantitative studies.[1]

percent quantitative articles
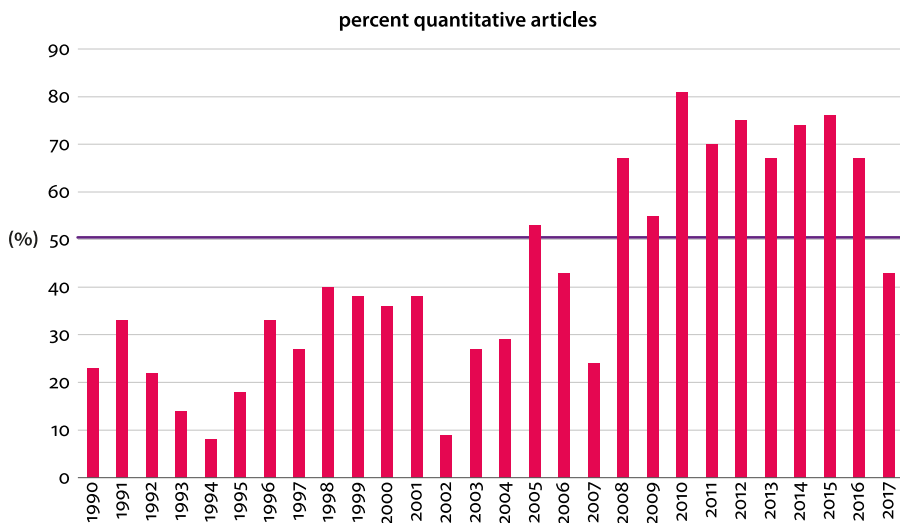
Figure 1. Percentage of articles presenting quantitative studies published in *Cognitive Linguistics* 1990–2017

Figure 1 reports percentages of quantitative articles for each year. A thick line marks 50% to make this visualization clearer. On the basis of this distribution we can divide the history of *Cognitive Linguistics* into two eras: 1990–2007 – when most articles were not quantitative; and 2008–2017 – when most articles were quantitative. In 1990–2007, twelve out of eighteen volumes had 20–40% quanti-tative articles. The lowest points were 1994, with one out of twelve articles, and 2002, with one out of eleven articles. 2005 reached in the other direction, with ten out of nineteen articles. In the year 2017, only 43% of articles were quantita-tive, partly due to the fact that there was a Special Issue that year on multimodal

---

1. This survey includes only articles proper, excluding review articles, book reviews, overviews, commentaries, replies, and squibs. For the purpose of this survey we define a "quan-titative article" as an article in which a researcher reports numbers for some kind of authentic language data or discusses statistical methods.

communication, and within that special issue, five of seven articles (71%) did not involve quantitative analysis. Leaving aside 2017, the publication of quantitative articles seems to be leveling off at a rate above 60%.

Quantitative articles have always been with us; no year has ever been without quantitative studies. Three quantitative articles appeared already in the very first volume of *Cognitive Lingustics*: Goossens (1990) (with a database of metaphorical and metonymic expressions), Delbecque (1990) (citing numbers of attestations in French and Spanish corpora), and Gibbs (1990) (presenting experimental results). However, since 2008 quantitative studies have dominated the pages of our flagship journal.

This survey gives an approximate indication of the trend toward quantitative studies in our scholarly output. It also shows us that cognitive linguistics has always engaged in quantitative studies, yet there is no reason to expect quantitative studies to entirely eclipse non-quantitative studies either. I do not mean to imply that there is a dichotomy between quantitative vs. non-quantitative studies. A variety of valuable types of studies require no quantitative analysis, such as descriptive linguistics, theoretical works, and overviews of the state of the art. Conversely, an ideal quantitative study relies on linguistic description, expands our theoretical framework, and thus contributes to the state of the art. Thus in a sense, quantitative studies depend on and ideally integrate non-quantitative components, though the reverse is not necessarily true.

Although this survey is based on a single journal, *Cognitive Linguistics* is the signature journal of our field and it reflects the recent history of cognitive linguistics as a whole. Evidence from conferences and textbooks devoted to quantitative studies points in the same direction. Since 2002 there have been six biannual meetings of Quantitative Investigations in Theoretical Linguistics, a conference series devoted to statistical analysis of language data predominantly from the point of view of cognitive linguistics. QITL has grown over the years from a workshop with only a dozen speakers to a three-day event. Three of the authors of the five textbooks on the use of statistical methods in linguistics that I cite in Section 2.3 have close ties to cognitive linguistics: Harald Baayen, Stefan Gries, and Natalia Levshina.

How did we reach the quantitative turn? As is usually the case with historical developments, there was no single cause, but rather a combination of factors that pushed and pulled cognitive linguistics in this direction. Pushes have come from the theoretical framework of cognitive linguistics, which has proved to be fertile ground for developing research questions that rely on analysis of observed data. Pulls have come from the attraction of vast data resources and the access to sophisticated tools for their analysis.

2.1   A usage-based model of language is data-friendly

Cognitive linguistics is a usage-based model of language structure (Langacker, 1987, p. 46; 2013, p. 220). In other words, we posit no fundamental distinction between "performance" and "competence", and recognize all language units as arising from usage events. Usage events are observable, and therefore can be collected, measured, and analyzed scientifically (Glynn, 2010, pp. 5–6). In this sense, cognitive linguistics has always been a "data-friendly" theory, with a focus on the relationship between observed form and meaning. Linguistic theories that aim instead to uncover an idealized linguistic competence have less of a relationship to the observation of usage, though there are of course notable exceptions.[2]

Even the question of what constitutes data in linguistics is controversial, and largely dependent upon the theory that one uses. Some researchers refer to constructed examples and individual intuitions as data, while others prefer to use corpus attestations or observations from acquisition or experiments. Introspection certainly plays an important role in linguistic analysis and indeed in the scientific method in general (cf. Section 3.2), but reliance on introspection to the exclusion of observation undermines linguistics as a science, yielding claims that can be neither operationalized nor falsified (cf. Section 4.2). It may seem attractive to assume that language is a tightly ordered logical system in which crisp distinctions yield absolute predictions, but there is no a priori reason to make this assumption, and usage data typically do not support it. Instead we find complex relationships among factors that motivate various trends in the behavior of linguistic forms. A usage-based theorist views language use as the data relevant for linguistic analysis, and this gives cognitive linguistics a natural advantage in applying quantitative methods, an advantage that we have been steadily realizing and improving upon over the past quarter century.

It is crucial to distinguish between the linguist's own introspection about data (perhaps augmented by introspection solicited from a few colleagues) and the systematic elicitation of the intuitions of naïve informants under experimental conditions, which is a legitimate scientific method that normally involves quantitative analysis. The difference is that whereas the linguist's introspection does not necessarily yield reliable, replicable results, the elicitation of native speakers' intuitions can yield such results. Introspection on the part of linguists can present numerous problems in that there are disagreements between linguists (cf. Carden & Dieterich, 1980; Cowart, 1997); their intuitions about mental phenomena are often inaccurate (Gibbs, 2006); and last but not least, linguists' intuitions may be

---

2. For overviews of the use of corpus linguistics across various theoretical frameworks, see Joseph (2004) and Gries (2009).

biased by their theoretical commitments (Dąbrowska, 2010). Even if we put aside the issue of whether a linguist can report viable intuitions about language data, it is a fact that a linguist is an individual speaker, and there is abundant evidence that different speakers of the same language have different intuitions about linguistic forms. Given the fact of inter-speaker variation, it is more reasonable to assume that there is not just one model, but instead many models of the grammar of a given language (Dąbrowska, 2012; Günter, 2014; Barth & Kapatsinski, 2017). Every speaker, linguist or not, has to some extent a unique experience with the use of his or her native language, and a usage-based theoretical framework is well equipped to accommodate this fact.

2.2   Advent of electronic language resources

Recent history has impacted the practice of linguistics through the development of language corpora and statistical software. Today we have access to balanced multi-purpose corpora for many languages, often containing hundreds of millions of words, some even with linguistic annotation. Modern corpora of this kind became widespread only a little over a decade ago, but have already become the first resource many linguists turn to when investigating a phenomenon. Many languages have national corpora, and open corpora are being built, providing free access not only to the linguistic forms and annotation in the interface, but also to the code itself, facilitating further exploration of data. A free resource that has attracted linguists is the Google Books Ngrams Corpus, which has a function that charts the frequency of words and phrases in a few of the world's largest languages. In addition to corpora of written language, spoken corpora are becoming available, and some resources are even multimodal. For example, the UCLA NewsScape Library is an archive of billions of words in several languages, along with associated sound and images captured from television newscasts.

The attraction of all this data is predictably compelling, particularly for linguists who view usage events as linguistic data. It is no surprise that a large portion of the quantitative studies undertaken by cognitive linguists have involved the analysis of corpus data, either alone or in comparison with experimental results.

2.3   Advent of analytical tools

At approximately the same time that electronic corpora emerged, statistical software likewise became widely available. Thus linguists have at their disposal the means to explore the structure of complex data. The tool of choice for cognitive linguists is primarily "R" (R Development Core Team, 2010), which is open-source,

supports UTF-8 encoding for various languages, and has a programming package, "languageR", specially developed by Harald Baayen for linguistic applications.

A natural place to turn to for inspiration in the use of analytical tools is computational linguistics.[3] Computational linguistics has of course been around since the 1950s, and computational linguists have considerable expertise in digital exploration of language data. However, the goals of cognitive linguistics and computational linguists have traditionally differed significantly due to the theoretical focus of cognitive linguistics (though there is good potential for collaboration, cf. Section 4.1). Therefore, in addition to drawing on the capacities of computational linguistics, we have looked for leadership to other disciplines that also deal with human behavior but took the quantitative turn earlier, in particular psychology (in addition to sociology and economics).

We linguists are still in a formative period where we have not yet settled on a set of best practices for use of statistical methods. A pioneering work in bringing statistical methods to linguists was Butler's (1985) textbook. But ten years ago, this textbook was out of print and there were very few alternatives. Since cognitive linguistics took its quantitative turn in 2008, several texts have been published, such as Baayen (2008), Johnson (2008), Larson-Hall (2010), Gries (2013), and Levshina (2015). These books, together with scholarly works, are helping to establish norms for the application of statistical models to linguistic data and analysis. However the field of statistics is itself in a state of considerable flux, particularly in the area of nonparametric models (especially relevant for us, since linguistic data is usually nonparametric; see Section 3.1.2), adding an extra challenge for cognitive linguists as relative late-comers to quantitative analysis.

3.   What does the quantitative turn bring us?

An introduction to statistical methods goes beyond the scope of this article and is better addressed by the textbooks cited above, so I will give only a bird's eye view, sprinkled with illustrative examples of how cognitive linguists are applying such methods. The scope of this overview is restricted to tracking some trends and discussing the relationship between quantitative methods and introspection.

---

3. See, for example, the journal *Computational Cognitive Science* at http://www.computational cognitivescience.com/.

3.1 Quantitative methods in cognitive linguistics

The goal of this section is to illustrate how quantitative methods are being used in cognitive linguistics and to identify some methods that are likely to stand the test of time. All statistical models are subject to assumptions and limitations concerning the nature of the data that need to be carefully observed and many models also facilitate the measurement of effect sizes which should be applied wherever possible, but since these issues are covered in textbooks, neither of them will be addressed in detail here.

3.1.1 *Is A different from B? Chi-square test, Fisher test, Binomial test, t-test, ANOVA*

The main idea of this set of tests is to find out whether there are significant differences between two (or more) measured phenomena. Just because two numbers are different does not mean that there is a statistically significant difference between them. This set of tests aims to discover whether there is sufficient reason to reject the "null hypothesis". The null hypothesis is the default position according to which there is no difference between the measured phenomena. If the null hypothesis is true, the observed difference can be accounted for by random fluctuations in samples taken from a larger population of observations in which there is no difference. If the null hypothesis is rejected, the observed difference is unlikely to be accounted for by such fluctuations.

Languages often give speakers choices, for example the choice between: (A) the ditransitive (*read the children a story*), and (B) the prepositional dative (*read a story to the children*) constructions in English. Corpus or experimental data might reveal a pattern such that there is more use of A in one environment (X) than in another environment (Y). But is the difference between the measurements of A and B a significant difference? In other words, is there reason to believe that there is a real difference between the frequency of A and B, or might the difference we observe be just a matter of chance (the null hypothesis)? A chi-square test can tell us the probability that the observed difference is significant. Chi-square tests have been used, for example, to test differences between the two English constructions listed above (Stefanowitsch, 2011a; Goldberg, 2011), the difference between physical and metaphorical understanding of English *path* vs. *road* (Falck & Gibbs, 2012), and the difference in the use of SVO constructions between a child and his mother (Theakston, Maslen, Lieven, & Tomasello, 2012).

While a chi-square test can give an overall evaluation of whether there is something significant in a matrix of numbers, the Fisher test is useful when trying to find exactly which of those numbers deviates significantly from the overall distribution of the matrix. The Fisher test was brought to the attention of

cognitive linguists by Stefanowitsch and Gries (2003, 2005) in collostructional analysis, where the point was to find out which words (such as *disaster, accident*) were more or less attracted to constructions (such as *an N waiting to happen*).[4] This application of the Fisher test has since come under criticism (Bybee, 2010, pp. 97–101; Baayen, 2011, p. 15; Schmid & Küchenhoff, 2013; Küchenhoff & Schmid, 2015),[5] primarily for the use of numbers on very different scales (especially when some of these numbers are estimated rather than actual numbers), and for the use of the p-value as a measure of collostruction strength. However, when used on actual (not estimated) numbers of low values (tens or hundreds rather than tens of millions), the Fisher test is a useful way to probe the relationships among values in a matrix.[6]

If you know the overall distribution of a phenomenon, a binomial test can tell you whether the frequency of that phenomenon in your sample is significantly different from that in the overall distribution. Gries (2011) compared the frequency of alliterations in the British component of the International Corpus of English (the ICE-GB, here taken to reflect the overall distribution of alliteration in English) with the frequency of alliteration in lexically-specified idioms such as *bite the bullet* (as opposed to *spill the beans* with no alliteration). The binomial test showed that the frequency of alliteration in English idioms is indeed significantly higher than in English overall.

If two groups of items (for example, two different semantic groups of lexemes – let's call them A and B) each get a set of scores (for example, acceptability scores), those two sets of scores will probably overlap. If the means of scores of the two groups are different, how do we know whether there is a significant difference between group A and group B? In other words, how do we know whether the difference in means is likely to reflect a real difference, or just chance variation in a situation where A and B actually behave the same in a larger sample? A t-test can handle a simple comparison of two groups. ANOVA ("analysis of variance"), which is an extension of the t-test, compares the between-group variation in scores with the within-group variation in scores, making it possible to compare more than two groups or more than one variable across the groups. Dąbrowska, Rowland, and Theakston (2009) wanted to investigate the nature of long-distance dependencies such as *Who₁ did Mary hope that Tom would tell Bill*

---

4. Collostructional analysis a family of quantitative (statistical) corpus-linguistic methods (namely that it comprises colexeme analysis, (multiple) distinctive collexeme analysis, and covarying collexeme analysis. See Stefanowitsch (2013) for further details.

5. See also Gries' responses to this criticism in Gries (2014, 2015).

6. A relevant example of the application of the Fisher test is presented here: http://emptyprefixes.uit.no/semantic_eng.htm.

that he should visit _____*₁ ?* Dąbrowska et al. (2009)'s hypothesis was that spontaneously produced long-distance dependencies follow the lexically specific templates *WH do you think S-GAP?* or *WH did you say S-GAP?*, where S-GAP is a subordinate clause with a missing constituent, and the majority of the remaining attestations are minimal variations on these patterns. They conducted an experiment in which children and adults were asked to repeat long-distance dependencies that did vs. did not follow the lexically specific templates. An ANOVA analysis showed that children rely on lexically specific templates as late as age 6, and that even adults are more proficient with long-distance dependencies that match the templates. These results support the usage-based approach, according to which children acquire lexically specific templates and make more abstract generalizations about constructions only later, and in some cases may continue to rely on templates even as adults.

3.1.2   *What factors are associated with A? Correlation, regression, mixed effects regression, classification and regression trees, naïve discriminative learning*

Suppose you want to find out what factors contribute to a given phenomenon, such as reaction time in a word-recognition task. The reaction time (A), termed the dependent variable in this example, may be related to various other phenomena such as frequency, length, and morphological complexity (B, C, D, etc.), known as independent variables. Correlation and regression are a family of models that can be used to explore such relationships.

Correlation refers to the degree of relationship between two variables, such that the stronger the correlation, the better we are able to predict the value of one variable given the value of the other. Let's say, for example, that we want to explore the relationship between the corpus frequency of a word and reaction time in a word-recognition experiment. A likely outcome would be that there is a correlation, such that the higher the frequency of a word, the shorter the reaction time, and thus it is possible to fit a line to a plot of data where one variable (frequency) is on the x-axis and the other variable (reaction time) is on the y-axis. If there is a correlation, given the frequency of a word it is possible to use the slope and intercept of the line to predict the reaction time, and conversely, given the reaction time associated with a word it is possible to predict its frequency.

Notice that the prediction goes both ways. A big caveat with correlation is that prediction is not the same as causation: an association between frequency and reaction time does not necessarily mean that higher frequency causes shorter reaction times (or the converse). Even if you can use the value of B to predict the value of A with 100% accuracy, correlation tells you only that there is a relationship, not that B causes A. However, linguists are not immune to the temptation

to assume causation when correlation is found; for a survey of correlation in relation to this problem, see Ladd, Roberts, and Dediu (2015). Another problem with interpreting correlation is that an apparent association between variables A and B might well be caused by other variables that have not been taken into account. The larger the dataset, the easier it is to find spurious relationships such as a positive correlation between linguistic diversity and traffic accidents (overlooking more telling factors such as population size and GDP; see Roberts & Winters, 2013).

Correlation has been used in a wide variety of studies. For example, in a study of long-distance dependencies, Ambridge and Goldberg (2008) found a correlation between the backgrounding of a clause (measured by a negation test) and the difficulty of extracting a clause (measured by the difference between acceptability in questions vs. declaratives), such that verbs like *know* and *realize* behaved very differently from verbs like *think* and *believe*. In a study of Polish prefixed verbs, Kraska-Szlenk and Żygis (2012) discovered a correlation between the reported morphological transparency of a prefixed verb and its acceptability rating by experiment participants.

A regression analysis allows you to consider the relationship between a dependent variable (A) and a set of independent variables (factors associated with A). Linear regression is based upon the same calculations as correlation, since the line of best fit in a correlation is the regression line, defined by the regression equation. Because the correlation is generally not perfect, there is a difference between the predicted values and the actual values, and this difference is referred to as the "residual error". The standard error of estimate (which is an estimate of the standard deviation of the actual scores from the predicted scores) gives us a measure of how well the regression equation fits the data. Because regression is based upon the same calculations as correlation, it also inherits the same drawbacks, namely that: by default it assumes a linear relationship (though this can be modified), it cannot tell us anything about causation, and any association that we find might actually be the result of other variables that have not been taken into account.

Regression models come in a variety of types and all involve the prediction of a dependent variable based upon one or more independent variables (also called predictors). Ideally the independent variables should be independent not just of the dependent variable, but also of each other (thus avoiding what is called "collinearity").

In logistic regression (named after the logistic function used to divide all values into a categorical choice between two levels), the dependent variable has only two values, and this is particularly useful for linguistic phenomena that involve a choice between two forms. The goal of a logistic regression model is to predict the probability that a given value (for example, initial vs. final position) for the

dependent variable will be chosen. If the dependent variable has an ordered set of more than two values (such as the values low, medium, and high acceptability), it is possible to use an ordinal regression model. The use of regression, and in particular logistic regression, has become fairly common in cognitive linguistics. For example, Diessel (2008) tested the hypothesis that there is an iconic relationship between the position of a temporal adverbial clause (which can come before or after the main clause) and the order of the event reported in the adverbial clause as prior, simultaneous, or posterior to the event in the main clause. In other words, the prediction is that a speaker is more likely to produce *After I fed the cat, I washed the dishes* than *I washed the dishes after I fed the cat*. Diessel constructed a logistic regression model to explore the relationship between the position of the adverbial clause (initial vs. final) as the dependent variable (the factor that is being predicted), and as independent variables conceptual order (iconicity), meaning, length, and syntactic complexity.

Mixed effects models are regression models that can take into account "random effects", which are the effects introduced by individual preferences. Mixed effects models are commonly used in experimental studies where random effects account for the behavior of individual stimuli and/or participants, and such models make it possible to arrive at generalizations that go beyond a specific sample of speakers or data. Random effects are relevant when we need to cope with what are called "repeated measures", such as in an experiment where multiple measurements are taken from each participant. In a word-recognition task where each participant responds to a set of words, some participants will be faster in general than others, so the baseline speed of each participant needs to be taken into account as a random effect. Random effects are opposed to fixed effects, which have a fixed set of values such as those for sex and age for experimental participants or tense, number, and person for verbs. For example, lexemes might act as random effects in a model, since they can have individual patterns of behavior. Janda, Nesset, and Baayen (2010) and Nesset and Janda (2010) applied a mixed effects model to a historical change underway in Russian verbs; in this model the individual verbs are a random effect since each verb has its own tendencies in relation to the ongoing change: some verbs use more of the innovative forms while others tend to resist innovative forms. In a study of the relative success of anglicisms in Dutch, Zenner, Speelman, and Geeraerts (2012) treated the concept expressed as a random effect, along with a number of fixed effects: relative length of anglicisms vs. Dutch equivalents, lexical field, era of borrowing, "luxury borrowing" (when a Dutch equivalent exists) vs. necessary borrowing (when there is no Dutch equivalent), era of borrowing, concept frequency, date of measurement, register, and region.

Regression models rest on assumptions that are often violated by linguistic data. Linear regression is a parametric model, which means that it tests hypotheses about population parameters. In other words, this type of model assumes that data should follow the bell curve of what statisticians call a normal distribution. Corpus data is however usually highly skewed, thus rendering linear regression less appropriate. Logistic regression assumes that all of the combinations of the various levels of all variables should be represented in the dataset. However, linguistic data often involves systematic gaps where certain combinations of the relevant variables are necessarily absent. There are at present at least two alternatives to regression models that offer the advantage of being nonparametric tests that also do not require all levels of variables to be observed in the dataset: classification and regression trees and naïve discriminative learning.

The classification and regression tree model ("CART"; Strobl, Tutz, & Malley, 2009) uses recursive partitioning to yield a tree showing the best sorting of observations separating the values for the dependent variable. Figure 2 shows an example of a CART tree from Baayen, Endresen, Janda, Makarova, and Nesset (2013), showing the behavior of the Russian verb *gruzit'* 'load' with respect to two grammatical constructions: the "goal" construction, as in *load the truck with hay* vs. the "theme" construction, as in *load the hay onto the truck.*
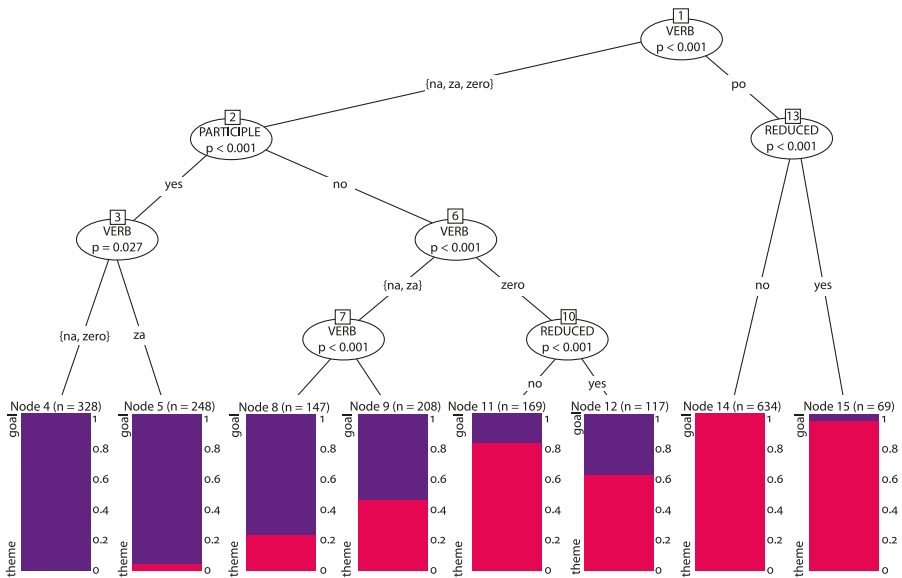

Figure 2. CART tree for Russian *gruzit'* 'load' from Baayen et al. (2013)

The terminal nodes at the bottom of the tree show the number of examples in each node ("n=") and plot the distribution of theme vs. goal uses for those examples. The top node of the tree (node 1) takes the entire dataset and makes the cleanest first division by finding the independent variable that is most effective at separating the goal uses from the theme uses, namely VERB: the 'load' verb prefixed in *na-*, *za-* or without prefix (the left branch) prefers goal use (represented by the purple bars in the terminal nodes) more than when prefixed in *po-* (the right branch), where theme use (scarlet bars in terminal nodes) is strongly preferred. On the right side at node 13, the *po-*prefixed verb forms are further sorted into reduced constructions (yes), where a few goal uses are attested (purple in node 15) vs. full constructions (no), where only theme uses are attested (node 14). Most of the goal uses appear to the left, where we see that at node 2 the most important factor is whether the verb form is a participle (yes) or not (no): nearly all these examples are goal uses, though a few theme uses are found for the *za-*prefixed verb (scarlet in node 5).

A CART tree can literally be understood as an optimal algorithm for predicting an outcome given the predictor values, and Kapatsinski (2013, p. 127) suggests that from the perspective of a usage-based model, each path of partitions along a classification tree expresses a schema, in the Langackerian sense (Langacker, 2013, p. 23), since it is a generalization over a set of instances. For example, in Figure 2, node 11 is a generalization over 169 examples in which finite (non-participial) unprefixed (zero) forms of Russian 'load' in full (not reduced) constructions show a strong tendency (over 80%) for theme use.

Naïve discriminative learning (Baayen, 2011; Baayen, Milin, Filipovic Durdjevic, Hendrix, & Marelli, 2011) is a quantitative model for how choices can be made between rival linguistic forms, making use of a system of weights that are estimated using equilibrium equations, modeling the usage-based experience of a speaker. Both CART and naïve discriminative learning offer means for measurement of the importance of variables and validation of results. A CART random forest analysis uses repeated bootstrap samples drawn with replacement from the dataset such that in each repetition some observations are sampled and serve as a training set and other observations are not sampled, so they can serve for validation of the model and for measurement of variable importance. Naïve discriminative learning partitions the data into ten subsamples, nine of which serve as the training set, reserving the tenth one to serve for validation. This process is repeated ten times so that each subsample is used for validation.

Baayen et al. (2013) test the performance of regression against classification tree and naive discriminative learning models across four datasets and find that the three models perform very similarly in terms of accuracy and measurement of the relative importance of variables.

3.1.3 *What is the structure of relationships among a group of items? Cluster analysis, multi-dimensional scaling, correspondence analysis*

A given linguistic item, for example, a lexeme, might be measured in many different ways, yielding an array of data; and a group of lexemes could then each have an array. The linguist might want to ask: which of these items are more similar to others, and how can these items be grouped? Cluster analysis, multi-dimensional scaling, and correspondence analysis take as input arrays of data associated with a set of items and use various mathematical techniques to arrange the items into a "space" of two or more dimensions.

Janda and Solovyev (2009) approached the relationships within two sets of Russian synonyms, six words meaning 'sadness', and five words meaning 'happiness', by measuring the relative frequency distribution of the grammatical constructions for each word in a corpus. The output of a hierarchical cluster analysis shows us which nouns behave very similarly as opposed to which are outliers in the sets. These results largely confirm the introspective analyses found in synonym dictionaries, and point to asymmetries between metaphorical uses of grammatical constructions and concrete ones.

Multidimensional scaling has been used in various ways in cognitive linguistics, for example to map out the functions of grammatical case in Slavic languages (Clancy, 2006) and to map the relations of aspect and expressions for spatial location (Croft & Poole, 2008; see also Janda, 2009).

Eckhoff and Janda (2014) used correspondence analysis to measure distances between verbs according to the frequency distributions of their grammatical forms, yielding a sorting that suggests that there was indeed a difference in behavior between perfective and imperfective verbs in Old Church Slavonic.

3.2 Role of introspection

There should be a healthy balance between introspection and observation in any scientific inquiry. Introspection is the source of inspiration for hypotheses, which are then tested via observation. When it comes to analysis, introspection is indispensable in order to interpret the results and understand what they mean for both theory and facts of language. The data do not speak for themselves; we need introspection in order to understand what they mean. The critical eye of introspection is necessary to ferret out suspicious results and alert us to problems in design and analysis. Whereas theory should of course be informed by data, theoretical advances are typically born through introspection.

Introspection is irreplaceable in the descriptive documentation of language. In fieldwork, a linguist interacts with speakers and posits the structure of a gram-

mar based on a combination of observations and insights. The foundational role of descriptive work and reference grammars is not to be underestimated, for without this background we would have no basis for stating any hypotheses about language at all.

4.   Where does the quantitative turn lead us?

Like any journey, taking the quantitative turn both opens up new opportunities and exposes us to new perils. It is worth taking stock of the pros and cons of this situation.

4.1   Opportunities

The most obvious advantage to taking the quantitative turn is of course the opportunities we gain to discover structures in linguistic data that would otherwise escape our notice. In addition, we can bolster the scientific prestige of our field and foster greater accountability and collaboration.

It is essential for the legitimacy of our field to secure and maintain the status of linguistics as a science. In applying quantitative measures we are developing linguistics as a discipline, following psychology and sociology in bringing the scientific method best known from the natural sciences to the fore. Stefanowitsch (2011b) urges cognitive linguists to adopt a more scientific outlook, to gain a higher degree of methodological awareness and to restrict our models to linguistic constructs and hypotheses that can be operationalized and falsified. Cognitive linguists are on the leading edge in terms of implementing data analysis in the context of a theoretical framework and we may well have a historic opportunity now to show leadership not only within cognitive linguistics, but in the entire field of linguistics. We can establish best practices in quantitative approaches to theoretical questions.

One important step we can take as a community is to make a commitment to publicly archive both our data and the statistical code used to analyze it. This will help to move the field forward by providing standards and examples that can be followed. In so doing, we can create an ethical standard for sharing data, stimuli, and code in a manner explicit enough so that other researchers can access the data and re-run our experiments and statistical models. Publicly archived linguistic data and statistical code have great pedagogical value for the community of linguists. As anyone who has attempted quantitative analysis of linguistic data knows, one of the biggest challenges is to match an appropriate statistical model to a given dataset. Access to examples of datasets and corresponding models will

help us all over the hurdle of choosing the right models for our data. We can advance more efficiently if we pool our efforts in a collective learning experience. In many cases, funding agencies require researchers to share their data, adding further motivation for public archiving of data. Ultimately, the most important reason for making data publicly accessible stems from the basic principles of the scientific method, namely that scientific findings should be falsifiable and replicable. Researchers should be held accountable for their findings and only findings that can be replicated can be considered valid. One good option for linguists is the Tromsø Repository of Language and Linguistics ("TROLLing" at https://dataverse.no/dataverse/trolling), a professionally managed, free and open international archive of linguistic data and statistical code built on the Dataverse platform from Harvard University.

As cognitive linguists become more familiar with quantitative methods, the opportunity for joining forces with computational linguists also increases. We can bring to the table valuable descriptive analyses and theoretical perspectives that can enrich collaboration in the building of better natural language processing and language technology applications.

4.2   Dangers

There are at least two types of dangers lurking just beyond the quantitative turn. One involves over-reliance on quantitative methods, and the other involves various kinds of misuse or neglect of data. In the face of these dangers, we can lose sight of the bigger picture of our theoretical principles and values.

If taken too far, quantitative research runs the risk of triviality and fractionalization of the field. It is very easy for researchers to be seduced by fancy equipment and sophisticated software to the point that these receive more attention than relevant linguistic principles. The most harmless negative outcome of this situation are shallow studies that do little or nothing to advance the field because they involve number-crunching without any real linguistic or theoretical goal. The potential outcome is a cognitive linguistic version of "cargo cult science"[7] in which linguists perform empty rituals of calculations in hopes of conjuring up publishable results.

More problematic is the substitution of "quantitative" for "empirical" and "scientific" in the minds of researchers. The use of quantitative methods in a

---

7. This term is used by Feynman (1992) to compare inept scientists to "cargo cult" south sea islanders, who, after experiencing airlifts during WWII, constructed mock runways manned by mock air traffic controllers, in hopes that this would cause more airplanes to land and bring them cargo.

study does not make it better or necessarily any more empirical or scientific than language documentation or qualitative analysis. Confusion of these concepts could result in the marginalization of many of the traditional endeavors of linguists that could then be disadvantaged in the selection of works presented at conferences and in publications. We thus risk erosion of the core of our field, linguistic description and theoretical interpretation, which are also the source for research hypotheses. As Langacker stated in 2015, "linguistic investigation is a highly complex and multifaceted enterprise requiring many kinds of methods and expertise"[8] and these various kinds of expertise should ideally be mutually supportive.

In the age of big data, it becomes far too easy to find results simply because as the number of observations increases toward infinity (or just millions and billions), statistical tests are able to find effects that are infinitesimally small and therefore meaningless. To some extent, this can be corrected for by the use of effect sizes as a check on results. However, Kilgarriff (2005) argues that since languages do not behave in a random fashion, the use of statistics to test null hypotheses is perhaps misguided to begin with. There will always be some patterns in linguistic data. The linguist's job is to bring enough insight to the enterprise to know what is worth looking for and to distinguish between results that have a real impact on the advancement of our science and those that don't.

Focus on big data analysis also threatens to marginalize languages themselves. Only a tiny fraction of the world's languages have the resources to support large corpora, experimental studies, and comprehensive language technology coverage. The quantitative turn has the potential to exacerbate the existing imbalance between the few languages that many linguists study and the majority of languages that are largely ignored.

We should not engage in an arms race to find out who can show off the most complex statistical models. It is usually the case that the simplest model that is appropriate to the data is the best one to use, since the results will be most accessible to readers. Sometimes the structure of the data dictates a more complex model, but very complex models carry with them the disadvantage that they are well understood only by the statisticians who developed them. Overuse of "black box" methods will not enhance the ability of linguists to understand and communicate their results.

Wherever numbers are involved, there is a temptation to misrepresent them. Most academic fields in which researchers report statistical findings have experienced scandals involving fudged data or analyses, and current pressures to pub-

---

8. Quoted from Langacker's presentation at the "Theory and Method" panel at the International Cognitive Linguistics Conference in 2015.

lish present an incentive to falsify results in hopes of impressing reviewers at a prestigious journal. Data sharing and best practices (cf. Section 4.1) can help us to protect our field from this kind of dishonor. While transparency does not guarantee integrity, it does make some kinds of fraud easier to detect, and it always improves the quality and depth of scholarly communication.

Major corporations such as Google, Amazon, Apple, and Facebook, along with hacking and spyware operations and state governments, have access to massive quantities of human language data. The lure of developing mining techniques via language analysis is part of what Kelly (2010) terms the "technium", the collective of archives and devices that constitute an organism-like system with a powerful momentum. This technology is advancing rapidly, and like it or not, we as linguists are contributing to it by improving our understanding of languages. This development is unstoppable; our only defense is to keep as much of it as possible in the public domain rather than behind clandestine corporate, state, and criminal firewalls.

5.   Conclusion

Since about 2008, cognitive linguistics has shifted its focus, and is now dominated by quantitative studies. On balance, the quantitative turn is a hugely positive step forward since it puts powerful new tools into the hands of cognitive linguists. Time always brings changes, and changes always bring challenges, but in this case the pros clearly outweigh the cons. Our field can gain in terms of scientific prestige and precision and collaboration. We can show leadership in best practices and the norming of application of statistical models to linguistic data. At the same time, I hope we can retain a humble attitude of respect for our venerable qualitative and theoretical traditions, which we should continue to nurture. If anything, we need qualitative and theoretical insights now more than ever in order to make sense of all the data at our command because those insights are the wellspring for hypotheses and the yardstick for interpretation of results.

Acknowledgements

Approaches to Russian) research group at UiT The Arctic University of Norway for comments on a previous draft of this article.

References

Ambridge, B. & Goldberg, A. E. (2008). The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics*, 19, 357–389. https://doi.org/10.1515/COGL.2008.014

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Baayen, R. H. (2011). Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics*, 11, 295–328. https://doi.org/10.1590/S1984-63982011000200003

Baayen, R. H., Milin, P., Filipovic Durdjevic, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118, 438–482. https://doi.org/10.1037/a0023851

Baayen, R. H., Endresen, A., Janda, L. A., Makarova, A., & Nesset, T. (2013). Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics*, 37(3), 253–291. https://doi.org/10.1007/s11185-013-9118-6

Barth, D., & Kapatsinski, V. (2017). A multimodel inference approach to categorical variant choice: Construction, priming and frequency effects on the choice between full and contracted forms of am, are and is. *Corpus Linguistics & Linguistic Theory*, 13(2), 213–260. https://doi.org/10.1515/cllt-2014-0022

Butler, C. (1985). *Statistics in Linguistics*. Oxford: Blackwell.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Carden, G., & Dieterich, T. G. (1980). Introspection, observation and experiment: An example where experiment pays off. *Journal of the Philosophy of Science Association*, 2, 583–597. https://doi.org/10.1086/psaprocbienmeetp.1980.2.192612

Clancy, S. J. (2006). The topology of Slavic case: Semantic maps and multidimensional scaling. *Glossos*, 7, 1–28. http://www.seelrc.org/glossos/issues/7/

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgements*. Thousand Oaks, CA: Sage Publications.

Croft, W., & Poole, K. T. (2008). Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. *Theoretical Linguistics*, 34, 1–37. https://doi.org/10.1515/THLI.2008.001

Dąbrowska, E. (2010). Naive v. expert competence: An empirical study of speaker intuitions. *The Linguistic Review*, 27, 1–23. https://doi.org/10.1515/tlir.2010.001

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, 2, 219–253. https://doi.org/10.1075/lab.2.3.01dab

Dąbrowska, E., Rowland, C., & Theakston, A. (2009). The acquisition of questions with long-distance dependencies. *Cognitive Linguistics*, 20, 571–597. https://doi.org/10.1515/COGL.2009.025

Delbecque, N. (1990). Word order as a reflection of alternate conceptual construals in French and Spanish: Similarities and divergences in adjective position. *Cognitive Linguistics*, 1, 349–416. https://doi.org/10.1515/cogl.1990.1.4.349

Diessel, H. (2008). Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, 19, 465–490. https://doi.org/10.1515/COGL.2008.018

Eckhoff, H. M., & Janda, L. A. (2014). Grammatical profiles and aspect in Old Church Slavonic. *Transactions of the Philological Society*, 112(2), 231–258. https://doi.org/10.1111/1467-968X.12012

Falck, M. J., & Gibbs, R. W. (2012). Embodied motivations for metaphorical meanings. *Cognitive Linguistics*, 23, 251–272. https://doi.org/10.1515/cog-2012-0008

Feynman, R. (1992). *Surely you're joking, Mr. Feynman!* London: Vintage.

Gibbs, R. W. (1990). Psycholinguistic studies on the conceptual basis of idiomaticity. *Cognitive Linguistics*, 1, 417–452. https://doi.org/10.1515/cogl.1990.1.4.417

Gibbs, R. W. (2006). Introspection and cognitive linguistics: Should we trust our own intuitions? *Annual Review of Cognitive Linguistics*, 4, 135–151. https://doi.org/10.1075/arcl.4.06gib

Glynn, D. (2010). Corpus-driven Cognitive Semantics: Introduction to the field. In D. Glynn & K. Fischer (Eds.), *Quantitative methods in Cognitive Semantics: Corpus-driven approaches* (pp. 1–42). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110226423.1

Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, 22, 131–153. https://doi.org/10.1515/cogl.2011.006

Goossens, L. (1990). Metaphtonymy: The interaction of metaphor and metonymy in expressions for linguistic action. *Cognitive Linguistics*, 1, 323–342. https://doi.org/10.1515/cogl.1990.1.3.323

Gries, S. T. (2009). What is corpus linguistics? *Language and Linguistics Compass*, 3, 1–17. https://doi.org/10.1111/j.1749-818X.2009.00149.x

Gries, S. T. (2011). Phonological similarity in multi-word units. *Cognitive Linguistics*, 22, 491–510. https://doi.org/10.1515/cogl.2011.019

Gries, S. T. (2013). *Statistics for Linguistics with R.* Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110307474

Gries, S. T. (2014). Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: Some necessary clarifications. In N. B. Gisborne & W. Hollmann (Eds.), *Theory and data in Cognitive Linguistics* (pp. 15–49). Amsterdam: John Benjamins. https://doi.org/10.1075/bct.67.02gri

Gries, S. T. (2015). More (old and new) misunderstandings of collostructional analysis: On Schmid & Küchenhoff (2013). *Cognitive Linguistics*, 26, 505–536. https://doi.org/10.1515/cog-2014-0092

Günter, F. (2014). Form, meaning and cognition: Language- and speaker-specific variation in linguistic and non-linguistic forms of interaction with spatial scenes. PhD Dissertation, Ludwig-Maximilians-University.

Janda, L. A. (2009). What is the role of semantic maps in cognitive linguistics? In P. Stalmaszczyk & W. Oleksy (Eds.), *Cognitive approaches to language and linguistic data: Studies in honor of Barbara Lewandowska-Tomaszczyk* (pp. 105–124). Frankfurt am Main: Peter Lang.

Janda, L.A. (2013). Quantitative methods in *Cognitive Linguistics* . In L.A. Janda (Ed.), *Cognitive Linguistics: The quantitative turn. The Essential Reader* (pp. 1–32). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110335255.1

Janda, L.A. (2017). The quantitative turn. In B. Dancygier (Ed.), *The Cambridge handbook of Cognitive Linguistics* (pp. 498–514). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316339732.032

Janda, L.A., Nesset, T., & Baayen, R.H. (2010). Capturing correlational structure in Russian paradigms: A case study in logistic mixed-effects modeling. *Corpus Linguistics and Linguistic Theory*, 6, 29–48. https://doi.org/10.1515/cllt.2010.002

Janda, L.A., & Solovyev, V.D. (2009). What constructional profiles reveal about synonymy: A case study of Russian words for SADNESS and HAPPINESS . *Cognitive Linguistics*, 20, 367–393. https://doi.org/10.1515/COGL.2009.018

Johnson, K. (2008). *Quantitative methods in linguistics*. Malden, MA: Blackwell.

Joseph, B. (2004). On change in Language and change in language. *Language*, 80, 381–383. https://doi.org/10.1353/lan.2004.0132

Kapatsinski, V. (2013). Conspiring to mean: Experimental and computational evidence for a usage-based harmonic approach to morphophonology. *Language*, 89, 110–148. https://doi.org/10.1353/lan.2013.0003

Kelly, K. (2010). *What technology wants*. New York: Viking Press.

Kilgarriff, A. (2005). Language is never, ever, ever random. *Corpus Linguistics and Linguistic Theory*, 1, 263–276. https://doi.org/10.1515/cllt.2005.1.2.263

Kraska-Szlenk, I., & Żygis, M. (2012). Phonetic and lexical gradience in Polish prefixed words. *Cognitive Linguistics*, 23, 317–366. https://doi.org/10.1515/cog-2012-0010

Küchenhoff, H., & Schmid, H.J. (2015). Reply to "More (old and new) misunderstandings of collostructional analysis: On Schmid & Küchenhoff" by S. T. Gries. *Cognitive Linguistics*, 26, 537–547. https://doi.org/10.1515/cog-2015-0053

Ladd, D.R., Roberts, S.G., & Dediu, D. (2015). Correlational studies in typological and historical linguistics. *Annual Review of Linguistics*, 1, 221–241. https://doi.org/10.1146/annurev-linguist-030514-124819

Langacker, R.W. (1987). *Foundations of cognitive grammar*. vol. 1: *Theoretical prerequisites*. Stanford: Stanford University Press.

Langacker, R.W. (2013). *Essentials of cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Levshina, N. (2015). *How to do linguistics with R*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.195

Nesset, T., & Janda, L.A. (2010). Paradigm structure: Evidence from Russian suffix shift. *Cognitive Linguistics*, 21, 699–725. https://doi.org/10.1515/cogl.2010.022

R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Roberts, S., & Winters, J. (2013). Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PLOSone*, 8(8), e70902. https://doi.org/10.1371/journal.pone.0070902

Schmid, H.J., & Küchenhoff, H. (2013). Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics*, 24, 531–577. https://doi.org/10.1515/cog-2013-0018

Stefanowitsch, A. (2011a). Constructional preemption by contextual mismatch: A corpus-linguistic investigation. *Cognitive Linguistics*, 22, 107–129. https://doi.org/10.1515/cogl.2011.005

Stefanowitsch, A. (2011b). Cognitive linguistics as a cognitive science. In M. Callies, W. R. Keller, & A. Lohöfer (Eds.), *Bi-directionality in the cognitive sciences* (pp. 296–309). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/hcp.30.18ste

Stefanowitsch, A. (2013). Collostructional analysis. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of Construction Grammar* (pp. 290–306). Oxford: Oxford University Press.

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, 8, 209–243. https://doi.org/10.1075/ijcl.8.2.03ste

Stefanowitsch, A., & Gries, S. T. (2005). Covarying collexemes. *Corpus Linguistics and Linguistic Theory*, 1, 1–43. https://doi.org/10.1515/cllt.2005.1.1.1

Strobl, C., Tutz, G., & Malley, J. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14, 323–348. https://doi.org/10.1037/a0016973

Theakston, A. L., Maslen, R., Lieven, E. V. M., & Tomasello, M. (2012). The acquisition of the active transitive construction in English: A detailed jyoti case study. *Cognitive Linguistics*, 23, 91–128. https://doi.org/10.1515/cog-2012-0004

Zenner, E., Speelman, D., & Geeraerts, D. (2012). Cognitive Sociolinguistics meets loanword research: Measuring variation in the success of anglicisms in Dutch. *Cognitive Linguistics*, 23, 749–792. https://doi.org/10.1515/cog-2012-0023

Massive corpora and models of cross-cultural communication styles in Cognitive Linguistics

The case of the N1 V (*for*) N2 *to*-infinitive construction in English

Vladan Pavlović
University of Niš

The paper addresses a pair of construction variants: the N1 V *for* N2 *to*-infinitive construction and its version without *for*, in the cases when both are possible with the same main verb (*I need <u>for</u> him to leave / I need you to stop*). It aims to point to the importance of the use of massive e-corpora in gaining better insight into the given construction pair. It also aims to test the hypothesis that the obtained quantitative data from such corpora can at least partially be accounted for by the interplay of: (1) the differences in the semantics of the two construction variants (based on combining relevant cognitive-linguistic insights), (2) the differences in the lexical semantics of the main verbs, and (3) extra-linguistic factors dealt with by models of cross-cultural communication styles. The paper thus argues for a tighter integration of cognitive-linguistic insights and a social-interactional perspective on language phenomena.

Keywords: construction variants, principle of no-synonymy, massive e-corpora, models of cross-cultural communication styles, recontextualizing approach to language

1. Introduction

The paper focuses on the N1 V (*for*) N2 *to*-infinitive construction in English, i.e. the one where the same main verb is followed by the N2 *to*-infinitive part, which can be: (1) either preceded by *for*: *He wished <u>for</u> her to kiss him goodnight*, or (2) not preceded by *for*: *He wished her to kiss him goodnight*.

The given constructional variants were chosen for two reasons. Firstly, such syntactically quite similar construction variants are relevant from the point of view of Cognitive Linguistics in general, and Construction Grammar in particular, especially in view of the principle of no-synonymy, namely the standpoint that "if two constructions are syntactically distinct, they must be semantically and / or pragmatically distinct" (Goldberg, 1995, p. 67; for a recent criticism of the given principle see Uhrig, 2015). And secondly, they are usually studied separately rather than together regardless of the theoretical approach (cf. Cuyckens & De Smet, 2007; Davies, 2013; De Smet, 2007; Fanego, 2007; Fischer, 2000; Huddleston & Pullum, 2002; McFadden, 2008; Radford, 2004; Wagner, 2000).

The paper has two major aims. The first one is to show that reliance on massive e-corpora may help gain additional insight into the two construction variants, which was not possible before with relatively limited e-corpora such as those used in cognitive(-functional) analyses of the given pair of constructions, especially when it comes to diachronic data. For example, Cuyckens and De Smet (2007) and De Smet (2007) explore the spread of the *for…to*-infinitive (rather than the infinitive where *for* is optional in front of the N2 element) in the English system of verbal complementation (as in *I'm waiting for them to arrive, She would like for him to settle*). The two authors conclude that the given construction came into existence through reanalysis of the *for*-phrase as part of an infinitive clause and state the mechanisms responsible for the propagation of *for…to*-infinitives as complements to many verbs (for details see *ibid.*). More importantly for the purposes of this paper, they rely on relatively limited corpora – the LOB (Lancaster – Oslo/Bergen) corpus, the FLOB (Freiburg-LOB) corpus, and the CLMET (the Corpus of Late Modern English Texts), containing approximately 12 million words in total, with massive electronic corpora, such as those belonging to the Brigham Young University suite of corpora, still not being available at the time of their publication. Consequently, albeit quite valuable, the given sources lack some of the insights that the latter can provide in the analysis of the pair of constructions in question. The paper thus establishes the quantitative data regarding the frequency of the appearance of the verbs that can be used as the main ones in each of the construction types (*ask, beg, expect, hate, wish*, etc.) from the diachronic perspective, based on the data retrieved from the COHA (Corpus of Historical American English), a massive e-corpus spanning the period between 1810 and 2009. This paper thus also aims to at least partly act upon Davies' (2013) view that robust data from large corpora of English can shed light on shifts in verbal syntax in ways that might not be possible with smaller ones, and to extend the analysis of the given pair of constructions this author initiates in his paper (cf. Davies, 2013, pp. 62–66).

And secondly, the paper argues for a tighter integration of the insights gained in Cognitive Linguistics, on the one hand, and of those obtained from the theories dealing with the wider social context, namely the theories of cross-cultural and intercultural communication styles. It thus attempts to provide a qualitative interpretation of the obtained quantitative results by relying on a combination of linguistic and extralinguistic factors, such as: (1) the differences in the semantics of the two construction types (based on combining insights of Lakoff and Johnson [1980], and Wierzbicka [1988]), (2) the differences in the lexical semantics of the main verbs, and (3) the dominant communication style that has developed in the USA during the last two centuries, as exemplified by the data from the COHA. Namely, the paper puts forward the hypothesis that the observed trends in the frequency of diachronic appearance of semantically different main verbs in the two given construction types in the COHA corpus, as representative of the USA linguistic data, at least partly correlate with the dominant communication style that has developed in the given country over the two centuries that this corpus covers. In that sense, the paper relies on three models of cross-cultural communication styles: Hall's model of high and low context cultures (Hall, 1976), Lewis's model of cultural types (the linear-active, the reactive and the multi-active ones; Lewis, 2005, 2006[1996]), and Hofstede's value dimensions, primarily those of power distance and individualism (Hofstede, 1980, 1991). In order to further test the given hypothesis, the paper compares the synchronic data for American English, British English, India English and Hong Kong English, on the basis of GloWbE (Corpus of Global Web-based English), yet another massive corpus, containing about 1.9 billion words of text from twenty different English-speaking countries. The paper thus makes a case for the application of the results of these and similar models in the field of Cognitive Linguistics, as an approach that embraces an experiential and usage-based (socio-pragmatic) view of meaning, and, more generally, as a recontextualizing approach to language (Geeraerts & Kristiansen, 2014). In that sense, this paper represents acting upon Croft's observation that "in order to be successful, cognitive linguistics must go 'outside the head' and incorporate a social-interactional perspective on the nature of language" (Croft, 2009, p. 1). It also represents acting upon Hilpert's observation that the study of constructional variation across different groups of speakers represents one of the current frontiers of Construction Grammar as "there simply is not a lot of research" in the area (Hilpert, 2014, p. 194). Some of the works that have already taken a similar approach in Cognitive Linguistics in general and Construction Grammar in particular, are Gries (2001, 2002, 2003) (which address English particle placement by taking into account the extralinguistic variable of written vs. spoken language), Gries and Stefanowitsch (2004) (which investigates not only cognitive but also cultural factors determining co-varying verbal

collexemes in the *into*-causative construction, exemplified by *She once <u>tricked me into drinking</u> decaf*), Wulff, Stefanowitsch, and Gries (2007) (which investigates the extent to which the meaning potential of the *into*-causative construction is variety-specific, i.e. specific to British as opposed to American English), Grondelaers (2008) (which addresses regional and diachronic constraints on cognitive explanations of the Dutch *er*-presentative). For an overview of literature dealing with constructional variation from the perspective of cognitive sociolinguistics, see also Hollmann (2013). This paper thus continues and elaborates on the general – recontextualizing – approach to language adopted in the given sources, namely that of attempting to address sociocultural factors in addition to cognitive-linguistic ones in the analysis of linguistic phenomena.

2.   Models of inter-cultural and cross-cultural communication styles

This section will outline three models of cross-cultural communication styles: Hall's model of high and low context cultures, Lewis's model of cultural types, and Hofstede's value dimensions, primarily those of individualism and power distance. They were chosen as they can be considered complementary.

As indicated above, in addition to American English, the paper also addresses the data from British, India and Hong Kong English (all the variants thereby exemplifying four communication styles that can be compared based on the given models). This is done for the following reason. Namely, once the paper has examined whether or not the observed trends in the frequency of diachronic appearance of semantically different main verbs in the two given construction types in the COHA corpus, as representative of the USA linguistic data, at least partly correlate with the dominant communication style that has developed in the given country over the two centuries, it was considered worthwhile to analyze whether similar correlations can be established in other varieties of English, i.e. whether reliance on communication styles can indeed help to at least partially account for the obtained data regarding the synchronic use of the two construction variants (and the different main verbs used in them) across the four varieties. This also justifies reliance on the variants of English spoken in different parts of the world (with American and British English coming from the Western, and India and Hong Kong English coming from the Eastern hemisphere).

It should be stressed that all three authors in their presentation of the communication styles pertaining to various cultures, naturally, write only about broad tendencies and generalizations rather than about any strict divisions between / among various categories they introduce.

Hall has introduced the notion of *high and low context cultures* (Hall, 1976, pp. 85–116 *et passim*). In his view, a high context message indicates a rather implicit meaning, which is "either in the physical context or internalized in the person", and little information is included in the "coded, explicit, transmitted part", and vice versa for a low context message (Hall, 1976, p. 91).

Lower context cultures, such as the USA and the UK, are generally characterized by the following: (1) they are culturally heterogeneous and tend to emphasize individualism, independence, and self-reliance; (2) the communication in them is generally clear, explicit, straightforward, direct and to the point, personal and informal, and factual rather than emotional; (3) they tend to use more exaggerated expressions of emotion, to communicate with more explicit and detailed information, and to employ elaborated communication codes (because where people share little common knowledge or value individuality above group identification, detailed elaboration becomes essential to avoid misunderstanding); in addition, in such cultures one can take the words that someone speaks much more literally; (4) they tend to write down and formalize information, and make it accessible widely; (5) they change rapidly and drastically.

In addition to the USA and the UK (England especially), lower-context cultures include but are not limited to Australian, Dutch, English Canadian, Finnish, German, Israeli, New Zealand, Scandinavian and Swiss culture.

On the other hand, *higher* context cultures are characterized by the following: (1) many things are left unsaid, letting the culture explain; (2) words and word choice are very important since a few words can communicate a complex message very effectively within a high context culture group (but less effectively outside such a group); (3) they are present in the countries with low racial diversity, high collectivism, with greater stability, with a strong sense of tradition and history, so that such cultures are relational, intuitive, and contemplative; (4) they have restricted communication codes and require listeners to share a great deal of common perspectives to understand the implicit meaning of the conversation; (5) they place a high value on interpersonal relationships, with group members making a very close-knit community; (6) their members communicate more through inference and implied ideas, so that much remains unsaid in a conversation.

Higher-context cultures include but are not limited to: African, American Indian, Arabic, Chinese, French Canadian, Greek, Japanese, Korean, Latin American, Portuguese, Russian, Spanish, Turkish, South and West Slavic culture, as well as the one present in Southern United States.

The cultures of India and Hong Kong, which are also important for our purposes in this paper, can be situated in-between the two poles, as they combine both high and low culture elements (Nishimura, Nevgi, & Tella, 2008). This once again confirms the fact that a cultural context does not rank as "high"

or "low" in an absolute sense, but can be placed somewhere on the continuum between the two.

Lewis (2005, 2006[1996]) has introduced the notion of *cultural categories of communication*, and classified cultures into *linear-active, reactive* and *multi-active* ones (Figure 1).
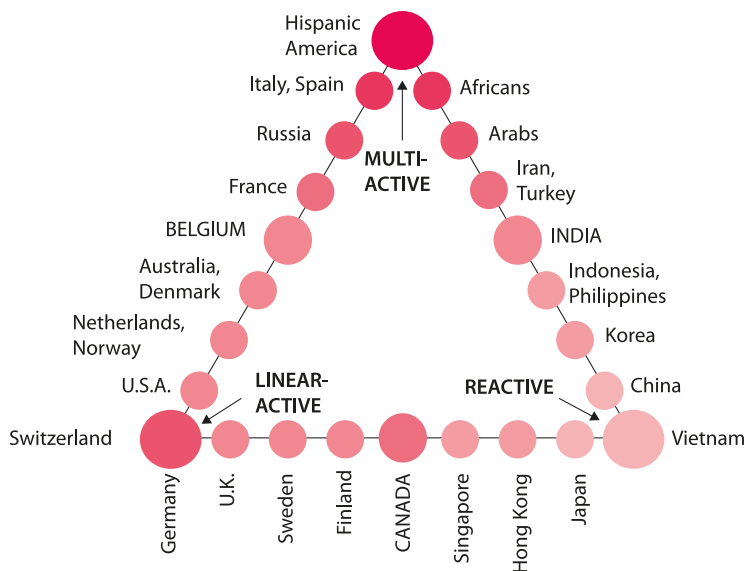
Figure 1. Cultural types (Lewis, 2005, p. 89)

According to the given author, people belonging to *linear-active* cultures are generally calm, factual and decisive planners. They prefer straightforward, direct discussion, and they talk and listen in equal proportions. They are task-oriented, highly organised and prefer doing one thing at a time. They stick to facts and figures that they have obtained from reliable sources. *Multi-actives* are warm, emotional, loquacious and impulsive. They like to do many things at a time. They often talk in a roundabout, animated way. It is typical of them to speak and listen at the same time, leading to repeated interruptions. They are uncomfortable with silence and seldom experience it among other multi-actives (Lewis, 2005, pp. 70, 89). *Reactives* are courteous, outwardly amiable, accommodating, compromising and good listeners. Their cultures are called 'listening cultures'. Reactives prefer to listen first, in order to establish both their own and the other's position. They often seem slow to react after a presentation or speech, and when they speak up, it is without clear signs of confrontation (Lewis, 2005, pp. 70–71). The paper will comment on the relative position of the USA, the UK, India and Hong Kong in view of the given cultural categories of communication soon.

Hofstede (1980, 1991) has introduced the notion of (five) *value dimensions* that a culture can be characterized by (see also: https://geert-hofstede.com /national-culture.html). Here we will focus on two of them – *power distance* and *individualism vs. collectivism.*

Power distance refers to the degree to which the less powerful members of a society accept and expect that power is distributed unequally. The fundamental issue here is how a society handles inequalities among people. People in societies exhibiting a large degree of power distance accept a hierarchical order in which everybody has a place, and which needs no further justification. In societies with low power distance, people strive to equalise the distribution of power and demand justification for inequalities of power.

Individualism can be defined as a preference for a loosely-knit social framework in which individuals are expected to take care of only themselves and their immediate families. Its opposite, collectivism, represents a preference for a tightly-knit framework in society in which individuals can expect their relatives or members of the group they are a part of to look after them in exchange for unquestioning loyalty.

Figure 2 presents the index of power distance and individualism in the four countries relevant for this paper. The data were obtained from the following links: https://geert-hofstede.com/united-states.html, https://geert-hofstede.com/united -kingdom.html, https://geert-hofstede.com/india.html, and https://geert-hofstede.com/hong-kong.html).
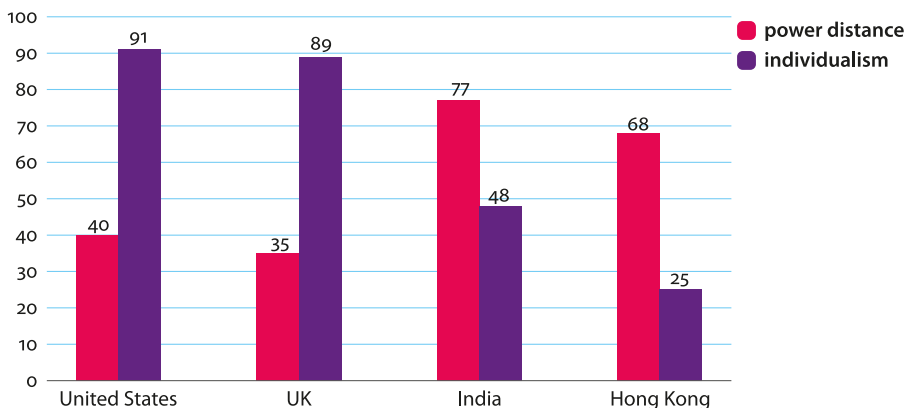

Figure 2. Index of power distance and individualism in select English-speaking countries

We will now sum up where the three given models of inter-cultural and cross-cultural communication styles situate the four countries relevant for the purposes of this paper. The USA can be labelled a low-context, linear-active culture with

just slight elements of multi-active cultures, with high individualism, and relatively low power distance. The UK can be labelled as a low-context, linear-active culture, with just slight elements of reactive cultures, with high individualism and relatively low power distance. India is closer to a low context culture while containing certain high context cultural features (Nishimura et al., 2008, p. 783), it is in-between multi-active and reactive cultures, and is characterized by much lower individualism, and much higher power distance than in the USA and UK. Hong Kong is similar to India, but it is also much closer to the reactive pole and is characterized by much lower individualism. Generally, the USA and the UK cultures, on the one hand, are more alike, just like the India and Hong Kong cultures, on the other hand. All these differences among the respective cultures will be relied on in the interpretation of the results obtained.

Other papers in linguistics that are based on at least one of those models include Ivorra Pérez (2014) and Prykarpatska (2008). Criticism of the given models (especially those of Lewis and Hofstede) is available in Piller (2007).

3.   The method and the corpus

First, a list was made of the verbs that can appear as the main ones in both construction variants. This was done by using the following search syntax in the COHA: *[vv*] for [p*] to [v*]* and *[vv*] [p*] to [v*]*, and then singling out the verbs that can appear in both patterns.

Secondly, the COHA was additionally used to establish the quantitative data regarding the frequency of appearance of such verbs in each of the constructions types from the diachronic perspective. Thus obtained quantitative data were then classified into groups (and subgroups) on the basis of the frequency of both construction variants and the trends in those frequencies diachronically with individual main verbs.

Thirdly, as indicated above, this has then served to present a partial qualitative interpretation of thus obtained quantitative data by relying on: (1) the differences in semantics of the two construction types (based on combining cognitive-linguistic insights), (2) the differences in the lexical semantics of the main verbs, and (3) the dominant communication style in the USA (the rationale thereby being that the obtained diachronic quantitative data may serve as a partial indicator of the trends that have contributed to the development of the dominant USA communication style in the way the given models of intercultural and cross-cultural communication define it today). The GloWbE corpus was additionally used at this stage for the reasons also outlined above.

Both massive corpora used (the COHA and the GloWbE) belong to the Brigham Young University suite of corpora created by M. Davies (Davies, 2012; Davies & Fuchs, 2015). The former (http://corpus.byu.edu/coha/) is the largest structured corpus of historical English, containing more than 400 million words in more than 100,000 texts which date from the 1810s to the 2000s (Davies, 2012). This makes it much larger than the corpora used in Cuyckens and De Smet (2007) and De Smet (2007), mentioned above, and allows for many types of searches that would not be possible otherwise. The latter (http://corpus.byu.edu/glowbe/) contains about 1.9 billion words of text from twenty different countries where English is spoken. This corpus allows retrieving data for each regional variety separately as well as in comparison with the remaining varieties. As indicated above, for the purposes of this paper, this corpus was used for a total of four regional varieties of English for reasons also already presented.

The search syntax used in the COHA in step 2 and in the GloWbE in step 3 outlined above consisted of the main verb established in the procedure presented in step 1 and of additional relevant operators, as in the following example with the verb *ask*: *ask.[v*] for [p*] to [v*]* and *ask.[v*] [p*] to [v*]*. Using such a search syntax has enabled the exclusion of the constructions that superficially look like either of the two constructional variants in question (*Abba Eban, the Israeli Ambassador to Washington, did not ask for permission to make a statement to the Council*), as well as the inclusion of the examples where one of the construction variants occurs at various levels below the sentential one (*Lucy spoke in a low tone, not wishing him to overhear*). The pronoun (i.e. the [p] operator) was used in the N2 slot above, as this has also helped yield more precise results than searching for other forms in the given slot.

4.   Data and discussion

The verbs appearing in both construction variants and the relevant examples are presented in Table 1. The quantitative data extracted from the COHA show that the [–*for*] variant generally occurs (far) more frequently than the [+*for*] one. According to the criteria outlined above, those data can be divided into two groups.

Group 1 is characterized by the following: (a) the [+*for*] variant has a very low number of tokens (up to two dozen for the entire period the COHA corpus encompasses), usually with two or more decades when it was not used at all, and (b) the cumulative [–*for*] variant frequency can be hundreds or even thousands with the same main verb in the same period.

The examples from this group can be subdivided into three subgroups.

Table 1. The [+*for*] and [−*for*] construction variants with the relevant main verbs

Verb	[+*for*]	[−*for*]
Ask	*For years the Interior Department has asked <u>for</u> legislation to keep the public lands out of the hands of thieves*	*We do not ask Congress to vote money for our railways, or canals, or turnpikes*
Beg	*He begs <u>for</u> you to come*	*I don't know how many of them I heard of who begged Europeans to take care of their families or their valuables*
Expect	*It is too much for him to expect <u>for</u> you to be untouched by other lips*	*Just don't expect me to do the dishes*
Hate	*If she's doing well and wants to keep on with it, I'd hate <u>for</u> her to stop her lessons for that long*	*I hate anybody to obstruct my work*
Intend	*I never intended <u>for</u> her to grow up in any city like Pittsburgh*	*In giving these illustrations of beautiful and of picturesque scenes, we have not intended them to be understood in the light of exact models for imitation in Landscape Gardening*
Like	*We'd like <u>for</u> you to take early retirement.*	*We're all friends out here and I'd like you to meet some of them*
Love	*"But you know how much we'd love <u>for</u> you to join us in Atlanta," continued Alison.*	*She loved everyone to be free and everything to be easy*
Mean	*His tone was gruff, though he didn't mean <u>for</u> it to be*	*He'd meant it to be a secret*
Need	*I need <u>for</u> you to tell me what happened*	*You need someone to take care of you.*
Prefer	*For my part, I would have preferred <u>for</u> everyone to know me as a journalist and minister who had it all together*	*If you prefer it to be as thick as gravy, you may add 1–2 teaspoons of corn-starch*
Want	*Lester gets on very well with us and we do not want <u>for</u> you to give him a hard time in any respect*	*I want you to know what you should do when I push a given button*
Wish	*I wouldn't wish <u>for</u> anyone to live through it again*	*Do you mean to say you wish me to treat my family and guests as though they were under suspicion?*

The first subgroup includes the examples where the frequency of the [−*for*] variant is generally on an increase throughout the period the COHA covers. This is the case with the verbs *need, expect*, and *want* (the pairs of examples from (1) through (3) respectively):

(1) a. *But I need <u>for</u> him to know my name, too*
    b. *That's why I need you to come back home*

(2) a. *It is too much for him to expect <u>for</u> you to be untouched by other lips*
    b. *I expect you to keep a certain dignified standard that reflect the status of that school*

(3) a. *but I said I want <u>for</u> us to have some time at the park*
    b. *I never wanted anyone to know.*

The frequency of appearance of the [+*for*] and [−*for*] variants with the three given verbs is presented in Figures 3 to 5 below.

Figure 3. Frequency of use of the [+*for*] and [−*for*] variants with the verb *need* in the COHA
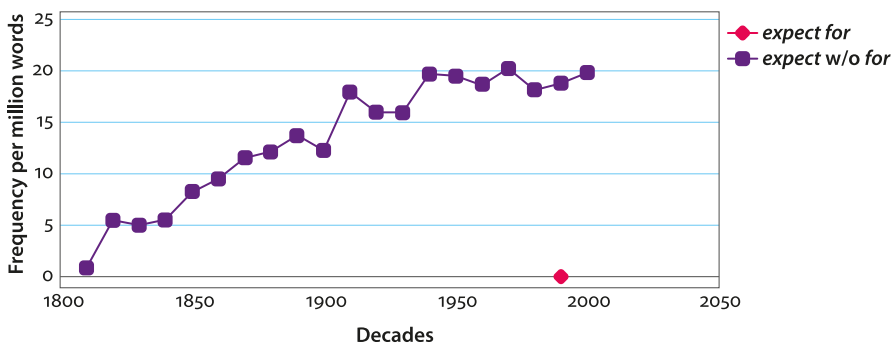
Figure 4. Frequency of use of the [+*for*] and [−*for*] variants with the verb *expect* in the COHA

The second subgroup within Group 1 includes the examples where the frequency of the [−*for*] variant is generally on a (sharp) decrease after a period of
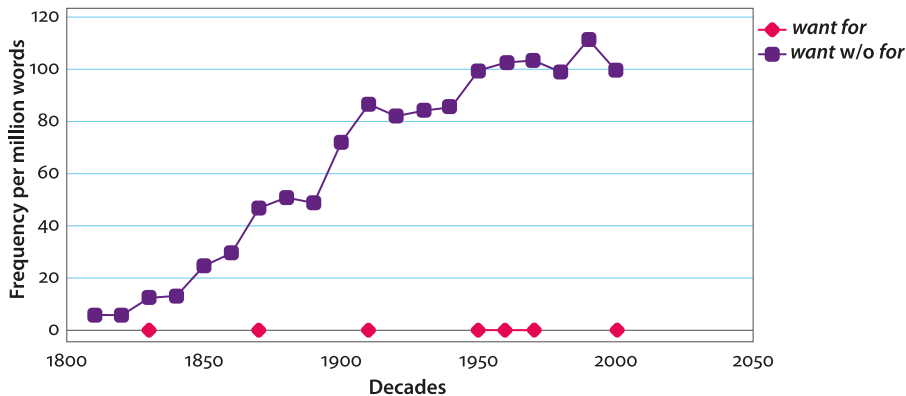
Figure 5. Frequency of use of the [+*for*] and [−*for*] variants with the verb *want* in the COHA

increase. This is the case with the verbs *beg* and *wish* (the pairs of examples under (4) and (5) respectively):

(4) a. *That person is just begging <u>for</u> someone to come up and hint around for some answers*
     b. *After my lesson, I begged him to let me stay and sleep on the deck of his ship*

(5) a. *I wouldn't wish <u>for</u> anyone to live through it again*
     b. *I wish you to remember and treasure up some things I have already said to you.*

The frequency of appearance of the [+*for*] and [−*for*] variants with the two given verbs is presented in Figures 6 and 7.

Figure 6. Frequency of use of the [+*for*] and [−*for*] variants with the verb *beg* in the COHA
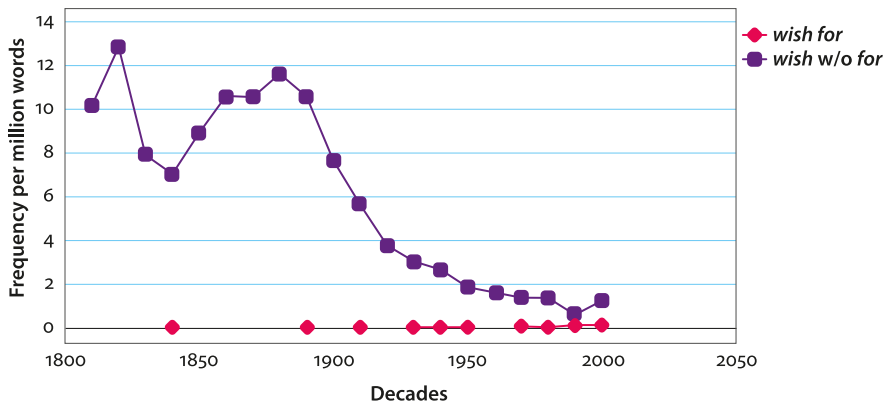
Figure 7. Frequency of use of the [+*for*] and [−*for*] variants with the verb *wish* in the COHA

The third subgroup within Group 1 includes the examples where the frequency of the [−*for*] variant is generally on a (marked) increase and then a slight decrease. This is the case with the verbs *ask* and *like* (the pairs of examples under (6) and (7) respectively).

(6) a. *She asked <u>for</u> them to close their eyes and pretend they were in the other-world*
     b. *I must ask you to extinguish it*

(7) a. *I'd like <u>for</u> you to meet his family*
     b. *I would like you to stop over in Passau*

The frequency of appearance of the [+*for*] and [−*for*] variants with the two given verbs is presented in Figures 8 and 9.
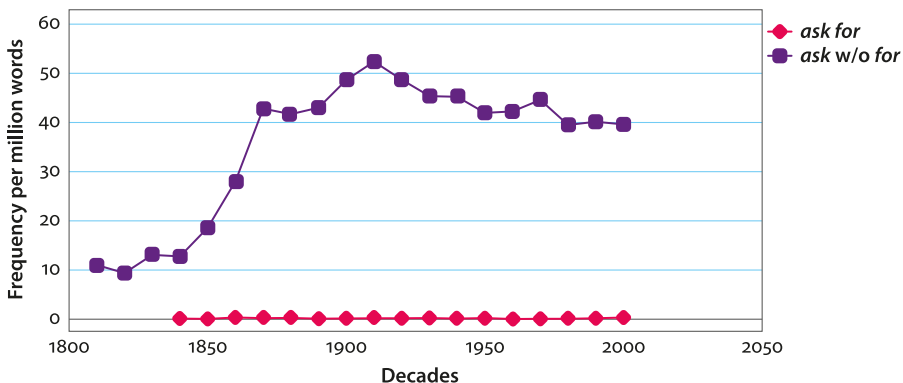
Figure 8. Frequency of use of the [+*for*] and [−*for*] variants with the verb *ask* in the COHA
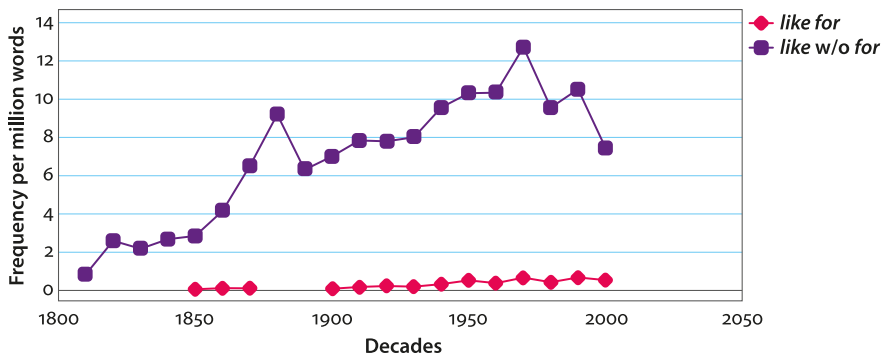
Figure 9. Frequency of use of the [+*for*] and [−*for*] variants with the verb *like* in the COHA

Group 2 includes the examples where both construction variants have a low frequency (up to 2.5 tokens per million words). This is the case with the verbs *hate, love, prefer, mean* and *intend* (pairs of Examples (8) through (12)).

(8) a. *I'd hate <u>for</u> her to stop her lessons for that long*
     b. *I'd hate you to spend the best years of your life eating sawdust*

(9) a. *"But you know how much we'd love <u>for</u> you to join us in Atlanta," continued Alison*
     b. *I'd love you to visit*

(10) a. *For my part, I would have preferred <u>for</u> everyone to know me as a journalist and minister who had it all together*
     b. *If you prefer it to be as thick as gravy, you may add 1–2 teaspoons of corn-starch*

(11) a. *Perhaps Senator Clinton did not intend <u>for</u> us to learn much about her life*
     b. *Instead, we intend them to illustrate the content uses on which social studies teachers can build instruction*

(12) a. *I'm so sorry! I didn't mean <u>for</u> you to get hurt!*
     b. *I meant it to be like a concert, to get the whole house rocking*

The frequency of appearance of the [+*for*] and [−*for*] variants with the five given verbs is presented in Figures 10 through 14.
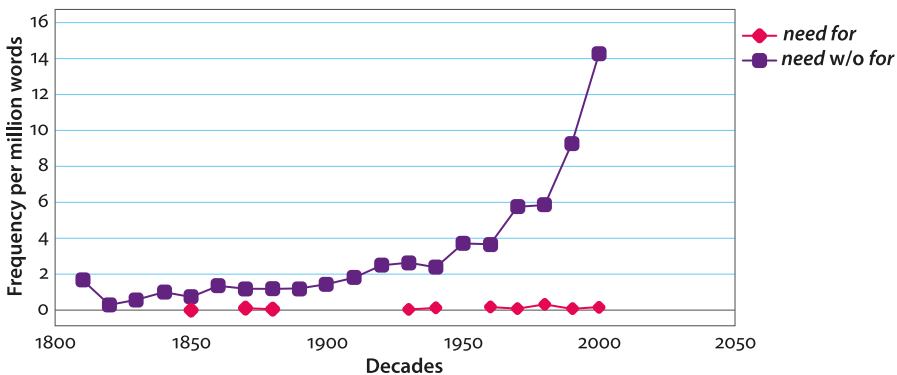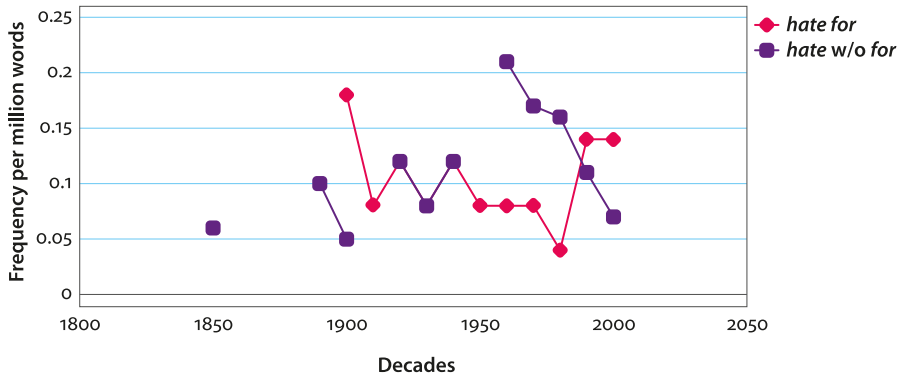
Figure 10. Frequency of use of the [+*for*] and [−*for*] variants with the verb *hate* in the COHA
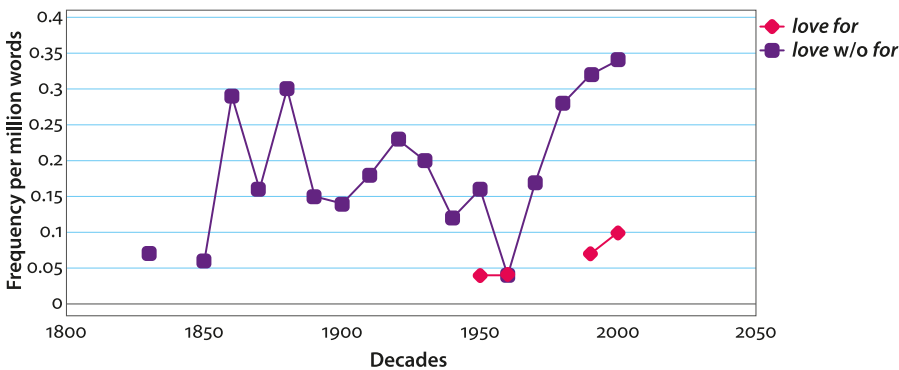
Figure 11. Frequency of use of the [+*for*] and [−*for*] variants with the verb *love* in the COHA
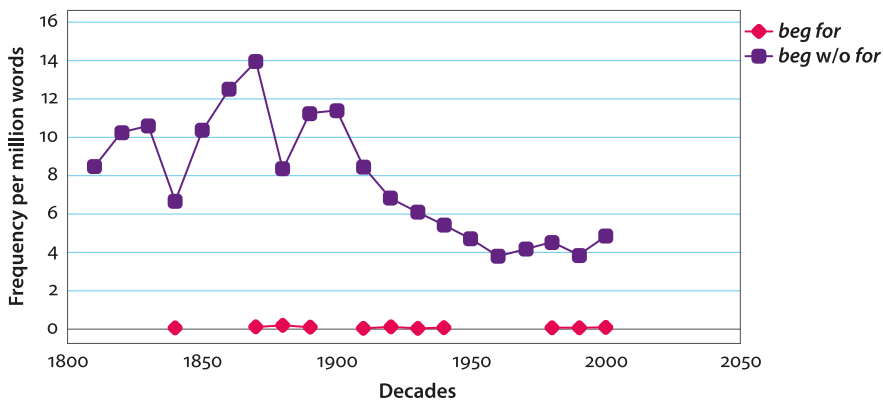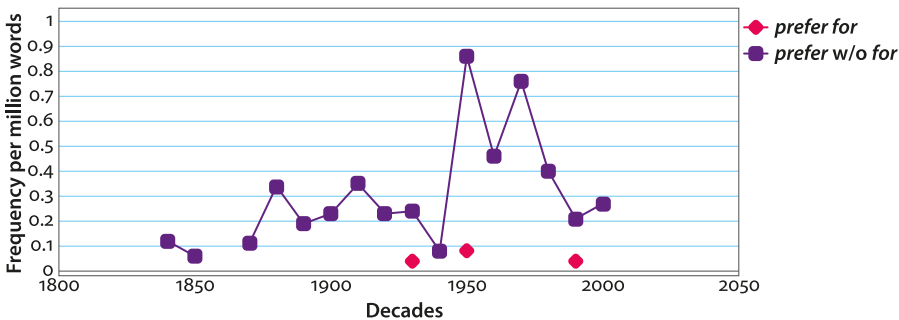

Figure 12. Frequency of use of the [+*for*] and [−*for*] variants with the verb *prefer* in the COHA
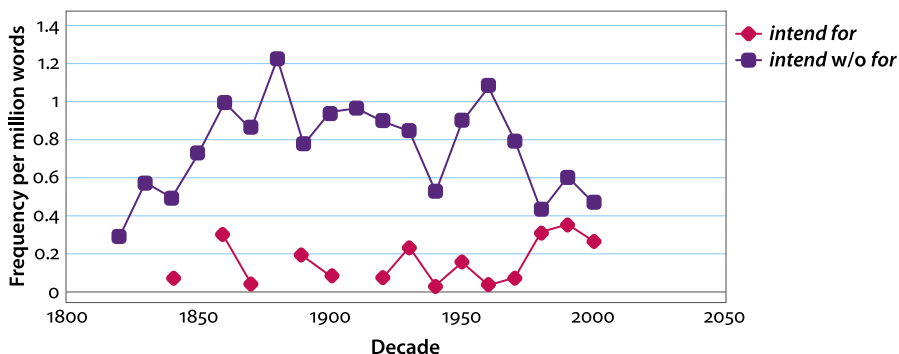
Figure 13. Frequency of use of the [+*for*] and [−*for*] variants with the verb *intend* in the COHA
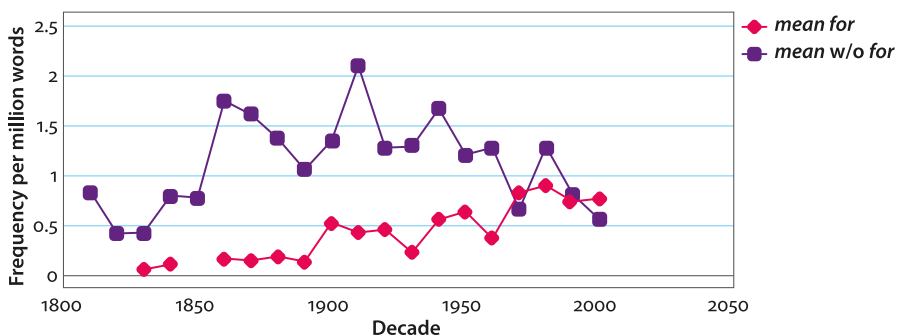
Figure 14. Frequency of use of the [+*for*] and [−*for*] variants with the verb *mean* in the COHA

As indicated above, the paper will now attempt to show that the presented quantitative data may at least partially be accounted for by the interplay of: (1) the differences in the semantics of the two construction variants (based on combining relevant cognitive-linguistic insights), (2) the differences in the lexical semantics of the main verbs, and (3) extra-linguistic factors, in particular the dominant communication style that has developed in the USA during the last two centuries, as exemplified by the data from the corpus used. The discussion to follow, while being far from an exhaustive one, might thus provide a partial insight into the logic of the results above and is an invitation for further discussion.

In the account of the semantic differences between the two construction variants, the paper will rely primarily on Lakoff and Johnson (1980, pp. 126–138), where the two authors analyze how metaphor can give meaning to syntactic form. They claim that "if the meaning of form A affects the meaning of form B, then the

closer form A is to form B, the stronger will be the effect of the meaning of A on the meaning of B" (Lakoff & Johnson, 1980, p. 129). In addition, it will also rely on Wierzbicka's (1988, pp. 111–132) treatment of the differences between *for to* and *to* clauses within her approach to the semantics of English complementation.

In the [+*for*] variant (Examples (13) through (16) below), the N1 and the N2 elements are more distant from one another than in the [−*for*] one (Examples (17) through (20)), if only for the presence of *for*:

(13)   *We do not want for you to give him a hard time in any respect*

(14)   *She asked for them to close their eyes and pretend they were in the otherworld*

(15)   *I need for him to know my name*

(16)   *Grandfather wished for her to kiss him [but he knew it would not happen]*
(Wierzbicka, 1988, p. 132)

(17)   *I want you to know what you should do when I push a given button*

(18)   *I must ask you to extinguish it*

(19)   *I need you to come back home*

(20)   *Grandfather wished us to kiss him goodnight [and we had to comply].*
(Wierzbicka, 1988, p. 132)

The presence of *for* in the former group of examples, and the consequent greater physical distance between the N1 and N2 elements, indicates a "weaker strength of effect" or a more "indirect" relation between the referents of those units. In other words, this makes such examples more like pleas, in which there is attenuation of the "force" of a verb because of pragmatic reasons – the need to be politer and not too obtrusive. In addition, in the case of emotion verbs, such as *like, love* and *wish*, the use of which can count as polite itself, the [+*for*] variant implies an extra degree of politeness. This may also explain the absence of examples such as **I ordered for somebody to do something*, as direct ordering is incompatible with this variant. In addition, this variant is less confident than the variant without *for* ("tentative *for…to*" vs. "confident *to*") and is typically associated with the non-first-person perspective – it is natural for people to be less confident about future actions of other people rather than of their own (Wierzbicka, 1988, p. 132).

On the other hand, in the latter group, which exemplifies the [−*for*] variant, the N1 and the N2 elements are closer to one another, which implies a "greater strength of effect" or a more "direct" relation between their referents, rendering such examples more like orders rather than pleas (for the problems involved in using the terms *directness* and *indirectness*, see Wierzbicka, 2003, p. 88).

When the obtained quantitative data from the COHA and the given insights based on cognitive-linguistic sources are combined with the designation of the dominant USA communication style presented above, as well as with the semantics of the main verbs in the construction pair in question, the following conclusions can be put forward. Firstly, the [+*for*] variant, i.e. the "less direct" variant, has a very low number of tokens, whereas the [−*for*] (the "more direct") variant, has a frequency that can be hundreds or even thousands of tokens with the same main verb in some period. Secondly, the frequency of the [−*for*] variant is generally on an increase with *expect, need*, and *want*, namely with the verbs that may be said to show directness and straightforwardness. Thirdly, the frequency of the [−*for*] variant is generally on a (sharp) decrease after a period of increase with the verbs *beg* and *wish*. In other words, in a culture that generally values individualism, independence and self-reliance, people seem to avoid *begging/wishing* (*for*) anyone to do anything, but apparently take proactive steps for something to happen (they *expect/want somebody to do something* rather than *beg/wish somebody to do something*). Fourthly, both construction variants with the verbs denoting high emotional involvement (*hate, love, prefer*) have a low frequency, as opposed to the use of the verb *like*, which shows less emotional involvement and is used far more frequently. Fifthly, the verb *ask*, while generally a very frequent verb (it is the second most frequent verb in our corpus), is, similarly to the verb *like*, generally also on a slow decrease, often in the same time periods when the data show an increase in the use of verbs *need, expect* and *want*. In addition, the verb *want* is the most frequent verb in our corpus, its frequency in the [−*for*] variant being approximately 2.5 times higher in the 1990s than that of the verb *ask* at the same time. This once again shows that the verbs which show attenuation of "directness" because of the pragmatic need to be politer and not too obtrusive (like the verb *ask*), are generally on a slow decline frequency-wise when compared to the verb *want* (as well as *need* and *expect*), which are characterized by no such "attenuation of directness". Eventually, the verbs of intention (*mean, intend*) have a somewhat higher (but still generally low) frequency when compared with emotion verbs. In other words, communication merely about intentions rather than concrete actions in realizing goals (exemplified especially by the use of verbs such as *want, need* or *expect*), appears not to be practiced very much in the USA in view of the data from the given corpus based on American English. All such conclusions seem to corroborate the above view of the USA as a low-context, linear-active culture with just slight elements of multi-active cultures (in Lewis' terms), with high individualism, and relatively low power distance (in Hofstede's terms), and with the dominant communication style that has developed there as one that is generally clear, explicit, straightforward, direct and to the point, personal and informal,

factual rather than emotional, and that emphasizes individualism, independence, and self-reliance (in Hall's terms).

In order to additionally check such conclusions, the paper relies on the data from the GloWbE corpus for the four regional variants of English listed above. The hypothesis thereby was that the variant without *for* (the more "direct" one, i.e. the one that shows greater "strength of effect" of the referent of N1 on the referent of N2) is more frequent in American English and British English than in India English and Hong Kong English, and that this may be the case with the verbs such as *expect* and *want*, rather than *ask*, because of the semantics of these verbs. Such a hypothesis is in keeping with the generalized description of the communication styles in the four given regional varieties of English outlined above. For example, it can be expected that examples such as *I want you to go there immediately* would be used less frequently in cultures that are closer to the high context ones and to the reactive pole, and that are characterized by relatively low individualism and relatively high power distance (which progressively holds true for India and Hong Kong, in that order), rather than in cultures that can generally be described as low context, linear-active ones, with relatively high individualism and relatively low power distance (USA and UK).

The data in Figure 15 below seem to corroborate such a hypothesis – the verbs *want* and *expect* are indeed used more frequently in American English and British English rather than in the other two variants, while the verb *ask* is used just slightly more frequently in India English and Hong Kong English than in American English and British English (the given three verbs are also the most frequently used verbs in the four varieties in the given construction variant).
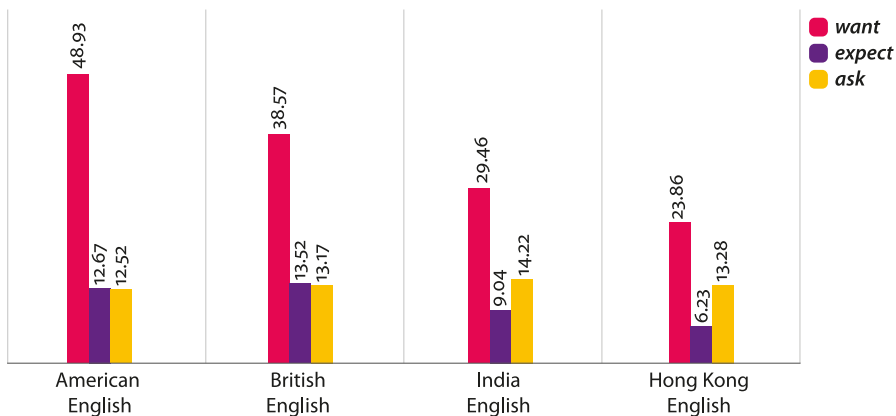
Figure 15. Frequency of use of the [*–for*] construction variant in 4 regional varieties of English (GloWBe)

At the end of this section, there is an additional point to stress regarding the e-corpora used. Cuyckens and De Smet (2007) point out that the overall frequency of the [+*for*] variant has increased during the last three hundred years (this includes verbs like *wait, arrange*, and *yearn*, which do not alternate with the [−*for*] variant). Our additional search of the COHA corpus shows that, while this is true, increase in frequency (in American English) is very slow (see Figure 16). In addition, while the frequency of the [+*for*] variant is not very high even in the two e-corpora used, as the data given above for a subset of verbs appearing in this variant show, that frequency is still much higher than in the LOB and FLOB corpora that De Smet used (there is only one instance of the given construction variant in the former and none in the latter corpus). This is yet another instance of how massive e-corpora can help build on the insights available in the literature.
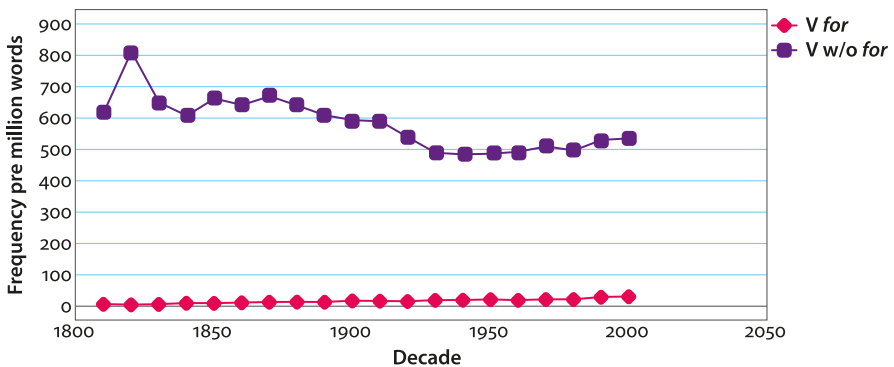
Figure 16. Frequency of use of the [−*for*] and the [+for] construction with any main verb (COHA)

5.   Conclusions

On the basis of the relevant cognitive-linguistic sources used in the interpretation of the differences between the two construction variants explored here, the paper has first provided evidence in support of the principle of no-synonymy, namely the standpoint that if two constructions are syntactically distinct, they must be semantically and/or pragmatically distinct.

The paper has then argued for a tighter integration of the insights gained from the cognitive-linguistic perspective, on the one hand, and of those obtained from the theories dealing with the wider social context, namely the theories of cross-cultural and intercultural communication styles, on the other hand. It has thus shown that there are indeed some correlations between the dominant

generalized communication style that has developed in the USA and the trends in the diachronic data regarding the frequency of use of the semantically different main verbs in the two construction variants. Additionally, synchronically speaking, it has also posited some correlations between the frequency of appearance of semantically different main verbs in the four different regional variants of English and the dominant generalized communication styles used in those cultures. Naturally, *correlations* are not to be mistaken for *cause-effect relations*, so the interpretation of the data given above in view of the aforementioned theories deserves much more research and is an invitation for further discussion. Quite importantly, the paper has continued the tendency adopted in some cognitive-linguistic sources (listed in the Introduction) to incorporate a social-interactional perspective on language phenomena into cognitive-linguistic research, i.e. to view Cognitive Linguistics as an approach that embraces an experiential and usage-based (socio-pragmatic) view of meaning, and as a recontextualizing approach to language (Geeraerts & Kristiansen, 2014), which in the case of this paper was done by reliance on the three models of inter-cultural and cross-cultural communication styles.

Eventually, the paper has also made a case for the use of massive e-corpora because insights made based on the data they provide can help build on the already established ones.

Acknowledgements

References

Croft, W. (2009). Toward a social cognitive linguistics. In V. Evans & S. Pourcel (Eds.), *New directions in cognitive science* (pp. 395–420). Amsterdam: John Benjamins. https://doi.org/10.1075/hcp.24.25cro

Cuyckens, H., & De Smet, H. (2007). *For…to*-infinitives from Early to Late Modern English. In J. Pérez-Guerra, D. González-Álvarez, J.L. Bueno-Alonso, & E. Rama-Martínez (Eds.), *'Of varying language and opposing creed': New insights into Late Modern English* (pp. 77–102). Frankfurt am Main: Peter Lang.

Davies, M. (2012). Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora*, 7(2), 121–157. https://doi.org/10.3366/cor.2012.0024

Davies, M. (2013). Recent shifts with three non-finite verbal complements in English: Data from the 100-million-word *Time* corpus (1920s–2000s). In B. Aarts, J. Close, G. Leech, & S. Wallis (Eds.), *The verb phrase in English: Investigating recent language change with corpora* (pp. 46–67). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139060998.004

Davies, M., & Fuchs, R. (2015). Expanding horizons in the study of World Englishes with the 1.9-billion-word Global Web-based English Corpus (GloWbE). *English World-Wide*, 36(1), 1–28. https://doi.org/10.1075/eww.36.1.01dav

De Smet, H. (2007). *For…to*-infinitives as verbal complements in Late Modern and Present-Day English: Between motivation and change. *English Studies*, 88, 67–94. https://doi.org/10.1080/00138380601042766

Fanego, T. (2007). Drift and the development of sentential complements in British and American English from 1700 to the present day. *Linguistic Insights: Studies in Language and Communication*, 28, 161–235.

Fischer, O. (2000). Grammaticalisation: unidirectional, non-reversible?: The case of *to* before the infinitive in English. In O. Fischer, A. Rosenbach, & D. Stein (Eds.), *Pathways of change: Grammaticalization in English* (pp. 149–169). Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.53.08fis

Geeraerts, D., & Kristiansen, G. (2014). Cognitive Linguistics and language variation. In J. Littlemore & J. Taylor (Eds.), *The Bloomsbury companion to Cognitive Linguistics* (pp. 202–217). London/New York: Bloomsbury Academic.

Goldberg, A. (1995). *Constructions: A Construction Grammar approach to argument structure.* Chicago: University of Chicago Press.

Gries, S. T. (2001). A multifactorial analysis of syntactic variation: Particle movement revisited. *Journal of Quantitative Linguistics*, 8, 33–50. https://doi.org/10.1076/jqul.8.1.33.4092

Gries, S. T. (2002). The influence of processing on grammatical variation: Particle placement in English. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.), *Verb-particle explorations* (pp. 269–288). Berlin/New York: Mouton de Gruyter. https://doi.org/10.1515/9783110902341.269

Gries, S. T. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement.* London/New York: Continuum Press.

Gries, S. T., & Stefanowitsch, A. (2004). Co-varying collexemes in the *into*-causative. In M. Achard & S. Kemmer (Eds.), *Language, culture and mind* (pp. 225–36). Stanford, CA: CSLI Publications.

Grondelaers, S. (2008). National variation in the use of *er* 'there': Regional and diachronic constraints on cognitive explanations. In G. Kristiansen & R. Dirven (Eds.), *Cognitive sociolinguistics: Language variation, cultural models, social systems* (pp. 153–203). Berlin/New York: Mouton de Gruyter. https://doi.org/10.1515/9783110199154.2.153

Hall, E. (1976). *Beyond culture.* New York: Doubleday.

Hilpert, M. (2014). Corpus-based approaches to constructional change. In T. Hoffman & G. Trousdale (Eds.), *The Oxford handbook of Construction Grammar* (pp. 457–475). Oxford: Oxford University Press.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* Beverly Hills, CA: Sage.

Hofstede, G. (1991). *Cultures and organizations*. Berkshire, UK: McGraw-Hill.

Hollmann, W. B. (2013). Constructions in cognitive sociolinguistics. In T. Hoffman & G. Trousdale (Eds.), *The Oxford handbook of Construction Grammar* (pp. 491–509). Oxford: Oxford University Press.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316423530

Ivorra Pérez, F. M. (2014). Cultural values and digital discourse: An intercultural communication approach to the transactional discourse of Spanish and US sales websites. *Journal of Intercultural Communication*, 36, 50–76.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lewis, R. D. (2005). *Finland, cultural lone wolf*. Yarmouth, ME: Intercultural Press.

Lewis, R. D. (2006) [1996]. *When cultures collide: Managing successfully across cultures* (revised edition.) London: Nicholas Brealey.

McFadden, T. (2008). Overt subjects of infinitives and *for-to* in the history of English. Paper presented at the 10th Diachronic Generative Syntax Conference. Cornell University, 7 August 2008. Retrieved from: https://www.academia.edu/2849136/Overt_subjects_of _infinitives_and_for-to_in_the_history_of_English?auto=download in August 2016.

Nishimura, S., Nevgi, A., & Tella, S. (2008). Communication style and cultural features in high/low context communication cultures: A case study of Finland, Japan and India. In A. Kallioniemi (Ed.), *Renovating and developing subject didactics: Proceedings of a subject-didactic symposium in Helsinki on Feb. 2, 2008* (pp. 783–796). Helsinki: University of Helsinki.

Piller, I. (2007). Linguistics and intercultural communication. *Language and Linguistic Compass*, 1(3), 208–226. https://doi.org/10.1111/j.1749-818X.2007.00012.x

Prykarpatska, I. (2008). Why are you late? Cross-cultural pragmatic study of complaints in American English and Ukrainian. *Revista Alicantina de Estudios Ingleses*, 21, 87–102. https://doi.org/10.14198/raei.2008.21.05

Radford, A. (2004). *English syntax: An introduction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511841675

Uhrig, P. (2015). Why the principle of no synonymy is overrated. *Zeitschrift für Anglistik und Amerikanistik*, 63(3), 323–337. https://doi.org/10.1515/zaa-2015-0030

Wagner, S. (2000). Depends how long you want for it to take: *For/to* clauses in present-day spoken British English. *Arbeiten aus Anglistik und Amerikanistik*, 25(2), 191–211.

Wierzbicka, A. (1988). *The semantics of grammar*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.18

Wierzbicka, A. (2003). *Cross-cultural pragmatics*. Berlin/New York: Mouton de Gruyter. https://doi.org/10.1515/9783110220964

Wulff, S., Stefanowitsch, A., & Gries, S. T. (2007). Brutal Brits and persuasive Americans: Variety-specific meaning construction in the *into*-causative. In G. Radden, K. M. Köpcke, T. Berg, & P. Siemund (Eds.), *Aspects of meaning construction in lexicon and grammar* (pp. 265–81). Amsterdam: John Benjamins. https://doi.org/10.1075/z.136.17wul

Contrastive semantics of human locomotion verbs

English *walk* vs. Czech *jít* and *kráčet*

Naděžda Kudrnáčová
Masaryk University

This paper is a contribution to a hitherto unexplored area in English-Czech contrastive semantics. It examines differences in the construal of walking, the most prototypical type of human locomotion. Based on the data from *InterCorp*, a synchronic parallel translation corpus, it presents a cognitive oriented analysis of the semantics of English *walk* and its nearest Czech counterparts, i.e. *jít* and *kráčet*. Despite their apparent commonalities, the verbs in question do not construe walking in the same way. In contrast to *jít*, the construal of a motion situation in *walk* and *kráčet* involves focus on leg movements and bodily position, amounting to a marked segmentation of the motion into individual quanta. Focus on leg movements and verticality of the body is even more pronounced in *kráčet*, which can then serve an evaluative function; such a possibility is not open for *walk*. *Walk* thus occupies an intermediate position between the two verbs.

Keywords: contrastive semantics, parallel translation corpus, verbs of human locomotion, quantization of motion

1. Introduction

Verbs of human locomotion have been the subject of an enormous body of research. Yet to date relatively little attention has been paid to a cross-linguistic contrastive analysis that would involve a fine-grained description of the semantics of verbs encoding walking, the most prototypical type of human self-agentive locomotion. A minute contrastive analysis of the lexical semantics of English and German descriptive verbs provided in Snell-Hornby (1983), now a classic in its field, does not include a description of the lexical semantics of *walk* and *gehen*, its German counterpart, on account of the fact that these verbs are not high in the specificity and complexity of their meanings and as such do not class

among descriptive verbs. The neglect paid to a fine-grained inter-language comparison of the semantics of verbs encoding walking may be due, at least in part, to the fact that human locomotion provides "a domain where the exemplars are broadly, if not absolutely, shared across cultures" (Malt et al., 2014, p. 109). Numerous cross-linguistic studies have not attended to fine-grained differences in manner of motion verbs and have concentrated, instead, on the compilation and macro-level comparison of inventories of manner of motion verbs, demonstrating that basic manners of motion (particularly human gaits such as walking and running) tend to be lexicalized across languages, whereas their modifications are not always lexicalized (e.g., Wienold, 1995; Slobin, 2000, 2006; Malt et al., 2008; Akita, 2017).

Nevertheless, there exist two notable, quite recent studies that attend to a cross-linguistic comparison of basic manner of motion verbs in a fine-grained, detailed way. The study by Slobin, Ibarretxe-Antuñano, Kopecka, and Majid (2014), providing a minute analysis of manners of human gait in several languages, pays attention to specific aspects of the actor's physical/mental self and to aspects of the motor pattern (e.g., speed, steps) conveyed by additional specifications. Their research, based on elicited descriptions of a variety of manners of walking and running from speakers of five languages, has shown that speakers of verb-framed languages (French, Spanish, Basque) tend to add manner to neutral motion verbs, whereas speakers of satellite-framed languages (English, Polish) tend to elaborate expressive verbs by using additional modifications. Nevertheless, the verb *walk* and its Polish equivalent *chodzić* were an exception: they both received additional specifications of motor pattern and/or of the actor's inner state or attitude. With two exceptions, all motion events that have been sampled were unidirectional. Interestingly, therefore, Polish *iść*, a unidirectional (determinate) counterpart of non-unidirectional (indeterminate) *chodzić*, is not mentioned in the paper. Slavic systems of motion verbs distinguish motion in a single direction and motion not in a single direction (describing repeated or habitual motion, motion including multiple destinations, random or aimless motion, or ability, cf. Janda, 2006). Verbs encoding the former type of movement are termed unidirectional (determinate), verbs encoding the latter type of the movement are termed non-unidirectional (indeterminate, multi-directional). The distinction determinacy – indeterminacy applies to a limited set of imperfective verbs that lexicalize basic types of motion.

The in-depth study by Goddard, Wierzbicka, and Wong (2016) has offered the Natural Semantic Metalanguage explications of the semantics of *walk* and *run*, and their nearest German counterparts (*laufen, gehen, zu Fuss gehen* and *rennen*). It has demonstrated that the conceptual semantics of these verbs differ noticeably, especially with respect to manner of motion. The details of the study are deferred

to Section 3, which provides a brief survey of the descriptions of *walk* proposed in the literature.

The aim of this paper is to contribute to a fine-grained cross-linguistic comparison of the semantics of the English verb *walk* and its nearest Czech counterparts, *jít* and *kráčet. Chodit*, an indeterminate counterpart of the determinate *jít*, is only addressed in certain situations, to provide a clearer picture of the construal of movement in *jít* and *kráčet*.

2.   Data, method, and research question

The analysis is based on the data retrieved from *InterCorp*, a synchronic multilingual parallel translation corpus, compiled at Charles University in Prague, version 10, released in December 2017. The corpus was accessed using the KonText interface. I have drawn on a sub-corpus of English fiction and non-fiction texts (24, 872, 842 tokens), aligned with their Czech counterparts. English was either the language of the original or a translation from Czech. Translations from languages other than English or Czech were excluded. The inclusion of non-fiction texts has proved beneficial, especially with regard to *kráčet*, which is, in the more recent Czech dictionaries (e.g., *Slovník spisovné češtiny pro školu a veřejnost* [A dictionary of standard Czech for school and home], 2009), given as a literary variant of *jít* (the four-volume *Slovník spisovného jazyka českého* [A dictionary of standard Czech], 1960, adduces *kráčet* as a synonym of *jít*, i.e. it does not mention its literary status). In certain cases, *kráčet* was, especially in non-fiction texts, resorted to not on account of its literary status but on account of its capacity to unequivocally encode walking.

To get tokens of *walk* in correspondence to *jít* and *kráčet*, two CQL queries were used: in the first, the query for *walk* [lemma="walk" & tag="V.*"] corresponded to the query [lemma="jít" & tag="V.*"] in the Czech section, in the second to [lemma="kráčet" & tag="V.*"]. The searches yielded 2,386 occurrences of the lemma *jít* and 1,092 occurrences of the lemma *kráčet*. All the results were then manually checked, for two reasons. First, the texts in the InterCorp parallel translation corpus are aligned at the level of sentences, which means that *jít* (or *kráčet*) in a given Czech sentence need not be a counterpart of *walk* in the corresponding English sentence. Such non-correspondence instances were therefore excluded. Second, since metaphorical walking falls outside the scope of the investigation, all instances of metaphorical walking were excluded as well. As a result, 2,055 occurrences of *jít* corresponding to *walk* and 975 such occurrences of *kráčet* were submitted to analysis.

A preliminary analysis of the data showed that in certain cases the use of *jít* as a counterpart of *walk* would blur the intended meaning and that *kráčet*, therefore, was used as a more specific, i.e. a more explicit term. A closer look revealed that the reason lies in the fact that, as opposed to *walk* and *kráčet, jít* backgrounds the movements of the legs and feet, thus backgrounds the boundaries between individual steps. In other words, *jít* suppresses the segmentation of the movement into individual quanta, and thus is not suited to act as a translation counterpart of *walk* in all situations (more details are given in Section 4). This then posed the question: are there overt, explicit manifestations of the differences in the quantization of the movement between the verbs in question? A subsequent analysis of the data revealed that the differences can be found in the domain of speed (Section 5), bodily posture (Section 6) and directionality of movement (Section 7). It has also become apparent that marked focus on the movements of the legs and on the vertical position of the body in *kráčet* endows the verb with the potential to refer to one specific aspect of the human actor's self and that in *walk* reference to the actor's self is considerably weaker, while *jít* is neutral in this respect (Section 8). *Kráčet* can then function as an evaluative verb, while this possibility is not open for *walk* and *jít* (Section 9).

3.   A brief survey of the descriptions of *walk* in the literature

In the cognitive linguistic research, manner of motion verbs have been subject to analysis at a predominantly macro-level. In Talmy's (e.g., 1985, 2000) typology of lexicalization patterns, the most salient patterns are those that conflate motion with manner or path. As pointed out by Slobin (2006, p. 62), Talmy's specification of manner was rather unclear, covering a number of dimensions, including motor pattern, often combined with speed (as in *walk*) or force dynamics (e.g., *step*), attitude (e.g., *stroll*) or instrument (e.g., *ski*). From a cognitive perspective, *walk* thus received a more or less coarse-grained treatment. In Slobin (1997, p. 459), *walk* is mentioned as belonging to Tier 1 verbs, including "neutral, everyday" manner verbs (Tier 2 includes manner verbs which are "more expressive or exceptional"). A more detailed description of *walk* has been provided by Langacker (1991, pp. 290–291). He briefly enumerates the conceptual components of *walk* (volitionality, muscular exertion, motion of the legs, force-dynamic interaction with the ground, and resultant change of location), without elaborating on them any further.

The most detailed descriptions of the semantics of *walk* have been provided by Nida (1975), and, most notably, by Goddard et al. (2016). Nida (1975, pp. 79–81) identifies the following components of *walk*: (1) Environment – supporting sur-

face, (2) Source of energy – animate being, (3) Use of limbs as means of propulsion – all the limbs which are normally in contact with the supporting surface, (4) Points of contact with a surface – the extremities of the limbs, (5) Nature of contact with the surface – continuous contact, by one and then another limb, (6) Order of repeated contact between limbs and surface – alternating, (7) Directional orientation – indeterminant. The fullest existing account of the conceptual semantics of *walk* has been provided by Goddard et al. (2016), using the Natural Semantic Metalanguage (NSM). In the NSM programme (e.g., Goddard & Wierzbicka, 2014), semantic primes are universal concepts that cannot be paraphrased in simpler terms without invoking circularity; no strict separation is thus seen between a semantic and conceptual analysis. The semantic template of *Someone X is walking* is explicated as follows:

(1)   The Lexicosyntactic Frame:
         someone X is doing something somewhere for some time
         because of this, this someone is moving in this place during this time as this
              someone wants

(That is, *walk* includes an actor, a locational grounding – distinct from the goal – and a durational component. Further, it encodes "the concurrent localised movement under the control of the actor", Goddard et al. [2016, p. 314].)

(2)   The Prototypical Scenario:
         often when someone does this in a place, it is like this:
         –   this someone wants to be in another place after some time
         –   this other place is not very far from the place where this someone is

(In addition to specifying the distance between the starting and the resulting location, the Prototypical Scenario makes clear that walking, just as any other effortful human activity, is conceptualised with reference to prototypical motivations. The Lexicosyntactic Frame and the Prototypical Scenario are common to all agentive intransitive verbs of locomotion in English.)

(3)   The Manner section:
         when someone does this, this someone does something with the legs many
              times
         because of this, this someone's legs move many times in the same way
         when this is happening, this someone's feet touch the ground at many times
              in many places

         it happens like this:

- at some time, one foot touches the ground somewhere for a short time
- during this time the other foot moves for a short time above the ground
- after this, it touches the ground in another place, at this time it is in front of this someone's body
- after this, the other foot moves in the same way

because of this, after this, this someone's body is in front of the place where it was before

(The Manner section specifies that the actor acts with his legs, describes the details of the motor pattern involving alternating, recurring movements of the legs and points of contact of the feet with the ground, and, lastly, specifies direction.)

(4)   The Potential Outcome specifies the result of the activity of walking:
      if someone does this for some time, after this, this someone can be far from the place where this someone was before

Both Nida's feature analysis and the NSM explication attend to the repeated movements of the legs and repeated contact of one of the feet with the ground. Nevertheless, neither of the explications refers to the bodily posture. Admittedly, movements of the legs and feet play a crucial role in the walking movement. However, since the legs are vertical extensions of the torso, attention to their movements evokes reference to the whole body. As will be shown, foregrounding the movements of the legs and feet in *walk* and *kráčet* amounts to foregrounding the involvement of the whole body. In *jít*, by contrast, the relative backgrounding of the movements of the legs and feet leads to backgrounding the reference to the body.

The semantic template for German *gehen* (roughly, 'go/walk'), a parallel to Czech *jít*, differs from that proposed for *walk* in the following respects. In the Lexicosyntactic Frame, the possibility of reaching a desired destination is added. The Manner section does not include the specifics of the motor pattern and only indicates that the legs are actively involved in the movement carried out by people who want to change location (Goddard et al. 2016, p. 328):

when someone does this, for some time this someone does something with the legs many times
like people often do when they want to be somewhere else after some time

As can be seen, the explication for *gehen* is broad enough to cover motion situations like *zur Arbeit gehen* ('go to work').

4.   Quantization of motion in *walk, jít* and *kráčet*: A preliminary account

Walking is the most common gait used in everyday life (e.g., Malt, Gennari, & Imai, 2010, p. 38). The prototypicality of walking is consequential in that it abolishes the necessity to have a verb that would unequivocally encode the specific physical pattern of this type of movement (a fact also noted by Vidaković, 2012, p. 244). Talmy (2000, p. 65) mentions Emai, in which the main verb with co-event conflation can be used for self-agentive motion, but only for a manner other than walking. Walking can thus be represented by highly polysemous verbs (e.g., English *go*, German *gehen*, Czech *jít*, Slovak *ísť*, Russian *idti*). As is well known, the degrees of the generalness of these verbs vary. To compare Czech with Slovak and Russian, *jít* displays, in its motion senses, a more restricted semantic applicability than Slovak *ísť* or Russian *idti* in that it cannot encode a vehicle movement, a fact especially noteworthy with respect to Slovak, the closest relative of Czech (on the semantics of Russian *idti* cf., e.g., Rakhilina, 2004; Nesset, 2010; on the semantics of Czech *jít* cf. Saicová Římalová, 2010; on the semantics of Slovak *ísť* cf. Horecký, 1999).

To repeat, highly polysemous *jít* can be used to designate walking owing to the prototypical, i.e. default status of this type of movement. *Kráčet*, by contrast, unequivocally encodes walking (note that this verb is derived from the noun *krok* 'step', originally designating a movable part of the body; Machek, 2010). These facts are consequential in that they have an impact on the choice of verb. Cf. the following two non-default situations with *walk*, which markedly deviate from the standard walking situation and which employ *kráčet*:

(1)   Already several of the guests were lying down in the dishes, and the soup ladle <u>was walk</u>ing up the table towards Alice's chair, and beckoning to her impatiently to …
Někteří hosté leželi v mísách a polévková naběračka <u>kráčela</u> po stole k Alenčinu křeslu a zlostně jí kynula, ať…

(2)   Stretching his legs out behind him, he <u>walked</u> his feet up the glass wall until he was almost horizontal.
Natáhl nohy a „<u>kráčel</u>" jimi po skleněné stěně nahoru, až se dostal téměř do vodorovné polohy.

In these scenarios, *kráčet* is resorted to for explicitation purposes. In (1) *kráčet* unequivocally implies that the ladle carries out "walking", i.e. a movement segmented into discrete quanta, whose boundaries are determined by alternating contacts of an entity's part(s) with the surface. In (2) *kráčet* is used for the same reason. If *jít* were used here, the sentence would, in actual fact, mean that the actor moved his legs in such a way that they were in constant, uninterrupted

contact with the wall (similarly as in *Jan šel prstem po mapě* 'John went a-finger along a-map' meaning "John ran his finger along the map"). As is apparent, due to the suppression of the segmentation of the motion into discrete kinetic quanta, *jít* cannot be appropriately used in non-prototypical situations involving a markedly segmented movement.

Now consider a situation which is also non-prototypical but which – as opposed to (1) and (2) – includes "real" walking, i.e. a bipedal translocational movement executed by a human actor:

(3)   If you <u>walk</u> straight ahead in a spherically shaped space, you'll find, like Magellan, that sooner or later you return to your starting point …
Jestliže <u>budete kráčet</u> rovnou za nosem ve sférickém vesmíru, dříve nebo později zjistíte, stejně jako Magellan, že jste se vrátili do výchozího bodu …

In default walking, as mentioned, the boundaries between individual steps are delimited by (alternating) contact of the feet with the supporting surface. However, as will be shown in greater detail further in the text, *jít* gives less attention to the moving body, including movements of the legs and feet than *kráčet*. This means, among other things, that it gives less attention to the contact of the feet with the surface, i.e. to the boundaries between individual steps. In (3) the translator thus opted for *kráčet*, which clearly evokes the notion of a markedly segmented movement (walking), even though carried out in an environment lacking a supporting plane. The use of *jít* (though not excluded – recall that *jít* does not encode vehicle movement) would background the segmentation of the movement into individual steps (and, consequently, foreground the directionality of the movement). The strong associative link between *jít* and prototypical walking situations seems to be the reason why translators use *kráčet* to translate *walk* in non-prototypical situations, such as those including a non-prototypical surface (walking on water) or non-prototypical actors, e.g., human figures (representing what may be termed "bodily forms") or uniforms (uniforms can "walk" due to the metonymic link to their human possessors).

It appears, as a first approximation, that the main aspect of meaning differentiating between *walk* and *kráčet* on the one hand and *jít* on the other is the character of the quantization of the movement, i.e. the segmentation of the movement into individual quanta. As this article continues, additional evidence in favour of this observation will be provided.

5.   Speed in relation to the character of the quantization of motion

As shown by Vulchanova and Martinez (2013), speed is the most distinctive feature in verbs designating motion along the horizontal axis (*walk, run* and *crawl*). It is the structural feature to which speakers attend most (Phelps & Duman, 2012). Accordingly, it is taken as the main feature differentiating between *walk* and *run* – hence the frequent descriptions of the semantics of *walk* as expressing slow, or 'normal' speed (e.g. Faber & Mairal Usón, 1999, p. 109). Nevertheless, although high in manner salience, speed has, from a non-perceptual point of view, a secondary, derived status in that it is determined by the physical properties of the movement – one can thus walk both slowly and quickly (e.g. Miller & Johnson-Laird, 1976, p. 551).

As the percentages in Table 1 show, *kráčet* appears as a counterpart of *walk* modified by specifications of speed twice more often than *jít*:

Table 1. Specifications of speed with *jít* and *kráčet*

	jít (sum total of tokens: 2055)		*kráčet* (sum total of tokens: 975)	
	tokens	%	tokens	%
slow speed	51	2.48	66	6.91
high speed	47	2.29	34	3.49
total	**98**	**4.77**	**100**	**10.4**

Specifications of slow speed modifying *walk*: *slowly, slow, with slow steps, at the same slow pace*. Specifications of high speed modifying *walk*: *quickly, briskly, fast, swiftly, rapidly, at a rapid clip, with a quick, determined step*.

The reason for the differences in the readiness of these verbs to function as translation counterparts of *walk* complemented by specifications of speed should be sought in the differences in the degree of the quantization of the movement as encoded in these verbs. As opposed to *jít, kráčet* renders the movement as segmented into individual steps. Each step, being bounded on each side by the contact of one of the feet with the surface, represents a delimited unit, a discrete quantum of walking. Foregrounding the segmentation of the movement into individual steps thus involves foregrounding the boundaries between them. The intervals between individual steps are then perceived as more pronounced, which is the reason why *kráčet* appears as a counterpart of *walk* modified by specifications of speed twice more often than *jít*. Phelps and Duman (2012, p. 861) also take perception of speed in human locomotion as related to the segmentation of the move-

ment into steps – more specifically, to cadence, i.e. to the number of steps per unit of time.

The observation appealing to the pronounced intervals between individual steps as the factor motivating the use of *kráčet* as a counterpart of *walk* with descriptions of speed receives additional support from the fact that, as the data in Table 1 show, descriptions of slow speed appear with *kráčet* more frequently than descriptions of high speed.

It should be pointed out that the same patterning of the movement is also encoded in *walk*. As Goddard et al. (2016, p. 320) observe, the most salient word associated with *walk* is *step*, whose meaning is focused "on the idea of one foot being in stable contact with the ground or other surface".

By way of concluding this sub-section, it is noteworthy to mention that in Czech the phrases employing the noun 'step', namely, *jít krokem* 'walk a-step:INSTR', *kráčet krokem* 'walk a-step:INSTR' and *běžet krokem* 'run a-step:INSTR' all denote a slow movement, and thus attest to the strong implicative link between slow speed and the segmentation of the movement into individual steps.

6.   Bodily posture

As we have seen, *walk* and *kráčet* profile the movements of the legs. Nevertheless, one must resist the instinct to reduce the elements profiled in the semantics of these verbs to leg movements only. Admittedly, translocation in walking is effected owing to the movements of the legs. However, walking also includes an upright posture (a fact often adduced as part of the semantics of *walk*, e.g. Boas, 2008, p. 35). At this point, it is important to realize that the lower limbs support the body weight and help maintain the upright posture (e.g. Maki & McIlroy, 1997). That is, what also plays a role is the position of the legs with respect to the trunk. The legs as the trunk's lower extremities are vertical "extensions" of the vertical torso. The position of the legs with respect to the trunk then explains why, in the semantics of *walk* and *kráčet*, focus on (the movement of) the legs amounts to pronounced focus on their verticality, including the verticality of the whole body. Symptomatically then, *kráčet* is used as a translation counterpart of *walk* modified by descriptions of bodily posture twice more often than *jít* – cf. Table 2:

Table 2. Specifications of bodily posture with *jít* and *kráčet*

jít (sum total of tokens: 2055)		*kráčet* (sum total of tokens: 975)	
tokens	%	tokens	%
12	0.58	11	1.13

Specifications of bodily posture modifying *walk*: *upright, erect, erectly, straight, unbowed, hunched, bent double, half bowed, bent down, stooping, stiffly, as tall as possible.*

Consider two illustrative examples – (4) and (5) with *stiffly* modifying *jít* and *kráčet*, respectively:

(4) For Liesel was out of her chair and <u>walking</u> slowly, stiffly, towards the front of the room.
     Protože Liesel už vstala ze židle a pomalu, celá ztuhlá, <u>šla</u> před tabuli.

(5) He walked stiffly and without coordination.
     <u>Kráčel</u> toporně a nemohl ovládnout své pohyby.

By way of illustration, consider also (6) with *walk* and *kráčet*, explicitly involving the whole body. Note that the second, italicised instance of *walk* is translated by means of (underlined) *jít*. For stylistic reasons, Czech generally avoids the use of the same expression in close vicinity.

(6) He was slender, a cat in the strength of his muscles, and he <u>walked</u> as Linday had seen no man *walk*, effortlessly, with all his body, seeming to lift the legs with supple muscles clear to the shoulders.
     Strang byl štíhlý, silou svých svalů podobný kočce, a <u>kráčel</u> tak, jak Linday ještě neviděl žádného člověka <u>jít</u>, bez námahy, celým tělem, jako by pružným svalstvem zvedal nohy až k ramenům.

7. Directionality of path

7.1 Prepositional path phrases

Like English, German or Russian, Czech is a satellite framing language, i.e. it expresses manner of motion in the main verb and path in the satellite (for the typological classification of motion event constructions, cf. Talmy, 1985, 2000). In Czech, information about the path can be encoded in prepositional phrases, directional adverbs or verbal prefixes (or, less frequently, in path verbs, e.g.

(*za)mířit* 'head for'). Verbal prefixes thus fulfil a twofold function: they encode aspectual distinctions and in some cases they convey path information. To give an example, the verb *vyjít* (with the prefix *vy-* 'out' attached to the stem *jít*) used in the walking sense can either mean 'start to walk' or it can mean 'walk from the inside'. In this latter case, however, the verb must be complemented by a prepositional phrase or a directional adverb denoting the reference object or place, respectively: e.g. *Jan vyšel z pokoje* 'John out-walked from a-room' means "John walked out of the room", and *Jan vyšel ven* 'John out-walked out' means "John walked out" (Czech constructions of this type are instances of "a double framing construction, in which the path or framing expression is expressed twice, once as a detached satellite and once as part of the verb"; Croft, Barðdal, Hollmann, Sotirova, & Taoka, 2010, p. 208).

Although *jít* and *kráčet* are unprefixed, imperfective verbs, they can be used not only with atelic path phrases to encode motion events which do not include a spatial end-point but also with telic path phrases to encode events which include a spatial end-point. Certainly, the latter situation can also be expressed by means of a perfective verb, which contains unequivocal information about the presence of a final quantum of the motion – that which involves reaching the spatial end-point. (To give an example, the perfective prefix *do-* in the verbs *dojít* and *dokráčet* conveys the meaning 'finish to X'.) Nevertheless, an imperfective verb in a telic situation retains its processuality, which adds to the vividness and dynamicity of the motion situation. The information whether the end-point was reached is then borne by the context. The twelve most frequent prepositional path phrases complementing *walk* which is translated by means of *jít* and *kráčet* are presented in Tables 3 and 4.

As is clear, *jít* is most frequently used to translate *walk* occurring with the *to*-phrase – more than twice more often than *kráčet*. This fact has an obvious explanation. As discussed above, *jít* backgrounds the transitions between individual steps (i.e. backgrounds the quantization of the motion), which enables this verb to foreground the orientation towards achieving a spatial goal. Quite expectedly, the goal-directedness of *jít* also manifests itself in the fact that this verb is almost invariably used (in 19 out of 22 cases) to translate *walk* in situations in which the goal of the motion is represented by an institutional object (e.g., *jít do školy / do kostela* 'walk to school / to church') or by a purpose (e.g. *jít na oběd / na schůzku* 'walk to lunch / to an appointment'). The corpus only contains 3 such situations with *kráčet*, namely, *kráčet do práce / na bohoslužby / na mši* 'walk to work / into a religious service / to Mass'. It is perhaps symptomatic that the latter two situations include a purpose closely linked to the human actor's experiential self (the inclusion of the human actor's experiential self in *kráčet* will be dealt with in Section 8).

Table 3. Most frequent prepositional path phrases complementing *walk* translated by means of *jít*

jít
(sum total of tokens: 2055)

prepositions	tokens	%
to	354	17.23
down	85	4.14
towards	78	3.80
along	78	3.80
through	50	2.43
into	34	1.65
past	32	1.56
across	30	1.45
up	20	0.97
on	12	0.58
in	8	0.40
over	6	0.29

Table 4. Most frequent prepositional path phrases complementing *walk* translated by means of *kráčet*

kráčet
(sum total of tokens: 975)

prepositions	tokens	%
toward(s)	108	11.08
to	82	8.41
along	61	6.26
down	58	5.95
through	38	3.90
across	26	2.67
on	14	1.44
in	13	1.33
up	12	1.23
past	11	1.13
over	7	0.71
into	6	0.61

As opposed to *jít, kráčet* places weight on the movements of the legs and feet (nevertheless, as we have seen, focus on the legs and feet necessarily leads to focus on the whole body). It markedly segments the movement into individual quanta, which weakens the orientation of the motion towards achieving a spatial goal. Symptomatically then, *kráčet* appears most often as a translation counterpart of *walk* complemented by the *toward(s)*-phrase (cf. Table 4), which does not include the reference object or place as its part (Jackendoff, 1983, p. 165). As argued for by Jackendoff (1991) and Piñón (1993), *toward* (*směrem k* in Czech) represents a partitive counterpart of *to* (*k* or *do* in Czech). We can thus say that the partitive semantics of the preposition *směrem k* involves the processual construal of the movement, which is in harmony with processual (because markedly sequential) *kráčet*.

The difference between more goal-directed *jít* and more processual *kráčet* stands out clearly when the two verbs are used as translation counterparts of *walk* in comparable situations, e.g. in walking to the door. In (7) with *jít*, the movement is directed straight to the spatial goal, while in (8) with *kráčet*, the movement is accompanied by specifications of the details of its progression:

(7) Patrick swung his feet from the other side of the bed, and <u>walked</u> to the door. It was closed.
Patrick přehodil nohy přes postel, slezl a <u>šel</u> ke dveřím. Byly zavřené.

(8) She <u>walked</u> very slowly and erectly towards the door.
Velice pomalu a vzpřímeně <u>kráčela</u> ke dveřím.

Processuality as involved in *kráčet* predisposes this verb to be frequently used as an equivalent of *walk* complemented by the *along-* and *down*-phrases, which involve the movement of a figure along an unbounded, linear extent (Talmy, 2000, p. 215). In Jackendoff and Landau (1992, p. 116), *going along X* is treated as involving "moving parallel to the axis of X". We can put Talmy's and Jackendoff and Landau's treatments together and say that path phrases employing the prepositions *along* and *down* encode the axial linearity of the path, hence naturally invite the processual *kráčet* as a translation counterpart of *walk*. Consider two comparable situations, namely, situations including walking along a road. In (9) with *kráčet*, the movement is presented as a process that has longer duration and is accompanied by additional actions of the actor. In (10) with *jít*, the movement is presented as a (simple) fact that occurred.

(9)   She <u>walked</u> along the Cromwell Road; and every room she passed and peered into seemed to be a dining-room where … At last she reached the Natural History Museum.
<u>Kráčela</u> Cromwellovou ulicí; každý pokoj, do něhož z chodníku nahlédla, jí připadal jako jídelna, kde … Konečně došla k Muzeu přírodních věd.

(10)   A dream it was that drew me here. In it, I <u>was walking</u> along a country road, that was all.
Sen to byl, sen mě sem přivedl. <u>Šel jsem</u> v něm po venkovské silničce, to bylo celé.

7.1.1   *Relative suppression of directionality in* in- *and* on-*path phrases*

Motion situations like walking along the road (/ along the street / across the square / through the forest / along the pavement, etc.) encode directionality, whereas their variants with the locative prepositions *in* and *on* (walking on the road / in the square / in the forest / on the pavement, etc.) to a certain degree suppress directionality. In these latter situations, the path is presented as a place in which the movement is set.[1] Not surprisingly, the (relative) suppression of directionality in these situations invites the use of *kráčet* as a translation counterpart of *walk*. In these situations, *kráčet* is used three times more often than *jít*. Cf. (11) with *kráčet* and (12) with *jít*:

(11)   I <u>was walking not</u> in a place where no one had walked before but in a place…
<u>Nekráčel jsem</u> místem, kde by nikdo přede mnou nešel, nýbrž místem, kde ….

(12)   We turned down a side street where there were no lights and <u>walked</u> in the street.
Odbočili jsme do postranní neosvětlené uličky a <u>šli jsme</u> po ní.

7.2   Prepositionless path phrases

The absence of a preposition (*walk the path / the streets / the road / the beach / the shore / the decks of the speedy frigate*, etc.) backgrounds the directionality of the path. It thus does not come as a surprise that *walk* with prepositionless path phrases is, in terms of percentage, translated by means of less goal-directed *kráčet* three times more often than by means of goal-directed *jít*:

---

1.   The (relative) suppression of directionality in these situations would seem to support Johnson's view that paths are inherently non-directional, "though we tend to impose directionality upon them, depending on our view or location" (Johnson, 1987, p. 122). Nevertheless, Clausner and Croft (1999, pp. 16–17) counter-argue that it is the asymmetry of the starting point and the end point in the path image schema that attributes directionality to the path. In view of the possibility to localize a translocational movement in a place, it seems that the directionality of the path is a scalar phenomenon.

Table 5. Prepositionless path phrases complementing *walk* translated by means of *jít* and *kráčet*

	tokens	%
jít	10	0.49
kráčet	14	1.44

By way of illustration, consider (13) with *kráčet* used to translate *walk* in the path phrase *walk this shore*:

(13)   It is enough to <u>walk</u> this shore and scuff our way on sandy symbols without caring whether that strand strangles us. Still, as the warmth cools upon my leg, I shiver …
       Stačí <u>kráčet</u> po tomto břehu a brouzdat se nohama v písečných symbolech, aniž bychom se starali, zda nás to nepodemele. Jak mi teplo na noze vychládá, otřesu se …

This example aptly illustrates that attention to the movements of the legs and feet in *kráčet* brings with itself the potential not only to background the goal-directedness of the movement but also to direct more attention to the force-dynamic contact of the feet with the surface. (Note that *walk* complemented by path phrases which employ the noun *surface* representing the ground – *walk on the surface of the sea (of a lake / of the earth), walk on a horizontal surface* – is invariably translated by means of *kráčet*.) Certainly, paths are not construed as receivers of energy (e.g. Langacker, 1987). Nevertheless, since prepositionless path phrases iconically abolish the distance between the actor and the path, they acquire the potential to highlight the physical contact between them.

It is worth noting in this connection that the corpus includes three prepositionless path phrases which denote markedly narrow paths: in two of them, *walk* is translated by means of *jít* (*walk a very straight line – jít po nějaký hrozně rovný čáře* 'walk along some awfully straight line', and *walk a tightrope – jít po laně* 'walk along a-rope') and in one of them, *walk* is translated by means of *kráčet* (*walk the wire – kráčet po drátě* 'walk along a-wire'). Admittedly, one would need a larger amount of data to draw safer conclusions but it seems that the crucial factor motivating the use of *jít* with these paths is their extreme narrowness. They represent a very low amount of two-dimensional space that is in direct contact with the feet. Since *jít* directs less attention to the movements of the legs than *kráčet*, it directs less attention to the force-dynamic contact of the feet with the surface. Therefore, a movement carried out on a very narrow stretch of space more naturally invites the use of *jít*. This argumentation receives support from the fact that when the path is represented not by a two-dimensional piece of space but by a line (i.e.

one-dimensional space), *kráčet* would be inappropriate. The phrase *walk a very straight line* is therefore translated by means of *jít*: *jít po nějaký hrozně rovný čáře* 'walk along some awfully straight line'. It should also be noted that the possibility to use *walk* to encode movement along a one-dimensional path attests to the fact that, in comparison with *kráčet*, the force-dynamic contact of the feet with the surface is profiled to a relatively lesser degree.

By way of concluding this sub-section on Czech translation equivalents of *walk* with prepositionless path phrases, mention should also be made of the use of *kráčet* to translate *walk* in motion situations involving multiplicity of paths, i.e. multiplicity of motion events. Under standard circumstances, these situations exclude the use of determinate (i.e. directed) *jít* and only admit *chodit*, its indeterminate (i.e. non-directed) variant. As observed by Janda (2006, p. 190), determinate motion verbs express goal-oriented completable situations, whereas indeterminate motion verbs express inherently noncompletable situations. Nevertheless, although *kráčet* does not class among indeterminate verbs, it can (admittedly very infrequently – the corpus contains 5 such situations) appear in such scenarios:

(14) People <u>walked</u> the narrow cobbled streets, went in and out of shops with baskets and string bags …
Lidé tu <u>kráčeli</u> úzkými dlážděnými uličkami, vcházeli do obchodů nebo z nich vycházeli s košíky a síťovými taškami …

The reason why *kráčet* is admitted into motion situations prototypically reserved for indeterminate *chodit* apparently lies in the fact that it foregrounds the quantization of the movement, and in this way acquires the potential to background its goal-directedness.

8.   The human actor's experiential self in *kráčet*

As we have seen, *kráčet* foregrounds the two focal physical components of walking, namely, movements of the legs and upright position of the body. A closer look at the verb's semantics in motion situations involving a human actor reveals that the marked profiling of these aspects endows this verb with the potential to bear reference to one specific qualitative aspect of walking, namely, the human actor's experiential self. (In this connection, one can speculate whether the literariness of this verb, adduced in the more recent Czech dictionaries, comes from the presence of precisely this aspect of meaning.)

In other words, owing to the intrinsic, inseparable link between the body and the mind, profiled attention to the body and its movements brings with itself the

potential to highlight the link between the movement and the human actor's inner self. More specifically, the actor is presented as a self-aware being, conscious of himself and of executing the movement. As Sheets-Johnstone (2012, p. 40) writes:

> In regarding ourselves and parts of ourselves as objects in motion, we experience ourselves – and conceive and speak of our bodies – in terms of the speed, range, force, and direction of movement, of initiating this movement "now" at this moment, and of terminating this movement "now" at this moment … When we observe our own movement in this way, we precisely *perceive* it, perceive it as *a force or effort put forth in time and in space*, a force or effort we are controlling or trying to control every step of the way.   (emphasis in original)

Kráčet renders this internal, first-person perspective from an observational, third-person perspective: it presents the movement through the lens of the human actor as an executor of the movement and its experiential perceiver. In this way it imposes an internalizing perspective on the movement and, in doing so, it brings the actor more in focus. *Kráčet* is thus typically used to translate *walk* in situations which include descriptions of the state of the human actor (as in (15)) or of his appearance (as in (16)):

(15)   She left the café, and as she <u>walked</u> along the Common she felt the distance widen between her and another self, no less real, who was walking back towards the hospital.
       Vyšla z kavárny. Když <u>kráčela</u> po Common, cítila, jak se rozevírá vzdálenost mezi ní a jejím druhým já, o nic méně reálným, které kráčelo nazpět k nemocnici.

(16)   Dean <u>walked</u> through with his arms hanging zombie-like at his sides, his mouth open, his eyes gleaming, and conducted a ragged and holy tour that …
       Dean přes ně <u>kráčel</u> s rukama bezvládně spuštěnýma podél těla, s pusou otevřenou, s hořícíma očima a vedl tu krkolomnou posvátnou pouť, která …

In *walk*, reference to the inner self of the human actor is weaker than in *kráčet* because it only follows from the position of the mover as a volitionally acting agent, executing a movement that has specific manner characteristics – note that in Slobin et al. (2014, p. 723), *walk* is referred to as "minimally expressive". *Jít* is neutral with respect to the human actor's inner self, apparently owing to the relative backgrounding of the body and its movements. It presents the movement in a matter-of-fact, objective-like way, in this way imposing an externalizing perspective on the movement:

(17)   A few minutes later, she caught sight of him through the large front windows, <u>walking</u> through the parking lot, his head cocked to one side, talking into a cell phone, carrying nothing …

> *O několik minut později ho zahlédla skrz širokou výlohu, jak jde přes parkoviště, hlavu nakloněnou na stranu, mluví do mobilního telefonu a nic nenese …*

As mentioned, a good way to illustrate the difference between *kráčet* and *jít* is to consider their use in comparable situations; cf. (18) with *kráčet* and (19) with *jít*:

(18)   They maneuvered Becky toward the double swinging doors leading to the OR corridor. Kim walked alongside. Becky still had a grip on his hand.
Vymanévrovali s Becky dvoukřídlými lítačkami na chodbu, vedoucí k operačnímu sálu. Kim kráčel vedle vozíku. Becky stále ještě svírala jeho ruku.

(19)   He thought that this waggon must be going to the fine town of London; so he took courage, and asked the waggoner to let him walk with him by the side of the waggon.
Řekl si, že takové spřežení určitě jede do velikého města Londýna. Sebral odvahu a zeptal se kočího, zda by ho nechal jít s ním vedle vozu.

As is evident, *kráčet* in (18) imposes an internalizing, i.e. more subjective-like, actor-centred perspective on the movement. It makes reference to the actor's experiential self and, in doing so, it brings attention to the relation of the actor's self to the self of the other participant in the situation, Becky. As Gibbs (2006, p. 28) puts it, people as experiencers "enter into specific relations to other experiencers". In Example (19) with *jít*, the movement of the human actor and the other co-moving entity (the waggon) is presented in an objective-like, externalizing way. It is rendered in purely spatial terms, as a configuration of two co-moving entities.[2]

Since *kráčet* presents the movement as if from within, through the lens of the human actor, it is intrinsically linked to the actual execution of the movement. This explains why *kráčet* is only exceptionally used to encode a movement that is yet to happen – the corpus only contains 2 such situations. (It should be added, however, that *kráčet* can freely be used in the future tense when used metaphorically.)

By way of concluding this sub-section, mention should be made of animal actors. To begin with, one can say that walking is an anthropocentric motion (e.g. Vulchanova & Martinez, 2013). Needless to add, this observation is a reflection of our human, i.e. anthropocentric bias in the interpretation of reality (on anthropocentricity of meaning in natural language cf. esp. Wierzbicka, 1988). As regards the walking of animals (like goats or horses), both *jít* and *kráčet* can be used.

---

2.   As one of the reviewers suggests, *jít* and *kráčet* may be seen as viewpoint tools in Czech.

Nevertheless, *kráčet* seems to be, most probably on account of its reference to the human actor's experiential self, a marked term in the animal domain:

(20)   … and as I walked … to the pump for water to wash myself, the horse, the goat, the German shepherd, and the cat <u>walked</u> gravely behind me. I turned and looked at them, and they looked at me …
    Když jsem šel k pumpě pro vodu, abych se umyl, <u>kráčel</u> za mnou vážně koník, koza, vlčák a kočka … otočil jsem se a díval jsem se na ně a oni se dívali na mne …

Given the specific nature of animal agentivity, it is difficult to say whether the use of *kráčet* is meant to indicate the involvement of the animal's self (note the use of *gravely*; as observed by Kudrnáčová (2013, pp. 80–91), animal actions may be linguistically rendered as involving a certain degree of self-awareness on the part of the animal actor) or whether, on account of its inherent actor-centredness, *kráčet* simply brings the animal actor more in focus. This question is thus open for further research.

9.   *Kráčet* used as an evaluative verb

As discussed thus far, *kráčet* does not carry any information that would signal an alternation of the physical pattern of the movement as is encoded in *walk* and *jít*. It only foregrounds the physical aspects of walking (movements of the legs and feet, i.e. segmentation of the motion into individual steps, and bodily posture). In the mental domain, it includes reference to the experiential self of the human actor. *Kráčet* can thus be accompanied by specifications like *neohrabaně* 'clumsily' (*heavily*), *nohy trochu vláčel* 'feet he-dragged a little' (*dragging his feet a little*), *shrbeně* 'with a stoop' (*half-bowed*), or *jedna těžce kulhala* 'one heavily limped' (*one of them limping heavily*). The reason why the translator opted for *kráčet* should apparently be sought in the high degree of manner salience of this verb (coupled with reference to the human actor's experiential self), which makes this verb predisposed to present the movement in a more dynamic, vivid way (on manner salience in relation to vividness in translation cf. esp. Slobin, 2005).

It is important to add that in dictionaries, *kráčet* is also given as one of the translation equivalents of *stride* and non-military *march*. These verbs have an evaluative status in that they point to a specific state of the actor (which may be roughly glossed over as pronounced deliberateness) as is reflected in a markedly regular and vigorous kinetic pattern involving longer steps and a pronounced upright posture. Snell-Hornby (1983, p. 135) describes *stride* as "an energetic, purposeful mode of walking, with long, swift steps, usually of a male agent" and non-

military *march* as "a regular, purposeful, energetic walk similar to that of *stride*, but with less emphasis on the steps taken" (Snell-Hornby, 1983, p. 136). Cf. (21) with *stride* and (22) with non-military *march*, both translated by means of *kráčet*:

(21) Lamia <u>strode</u> confidently ahead of them.
      Lamie rozhodně <u>kráčela</u> před nimi.

(22) He <u>marched</u> up the hill at quite a pace, but whatever propelled him suddenly gave out …
      Rázným krokem <u>kráčel</u> do kopce, ale síla ženoucí ho vpřed se znenadání vytratila …

The question of why *kráčet* is endowed with the potential to become associated with a deliberate, markedly self-conscious actor and a markedly regular kinetic pattern including longer steps and a pronounced upright posture invites, in the light of the description of the semantics of the non-evaluative sense of *kráčet* offered in Sections 5 and 6, an obvious answer. As we have seen, the profiling of the verticality of the legs as the body's lower extremities involves the profiling of the vertical axis of the body. This fact is consequential from a perceptual point of view. As verified by experimental research (Charras & Lupiáñez, 2010), the foregrounding of an axiality of an entity more likely evokes its prolonged rather than short length. This means that profiling an upright posture, i.e. profiling the axiality of the body, carries with itself the potential to profile the body's length, including the length of the legs. This naturally evokes the idea that the steps are long because, in the process of placing one foot in front of the other, they represent the width between the "wings" of the pendulum – hence the reference to long steps in the evaluative sense of *kráčet*. Now, taking longer steps naturally requires a higher energy output and a higher amount of control over the movement, which explains why *kráčet*, when used as an evaluative verb, bears reference to the energetic execution of the movement.

Apart from these physical aspects, there is obviously another aspect of meaning that enables *kráčet* to impose an evaluative perspective on the movement and serve as one of the translation equivalents of *stride* and non-military *march*. As discussed in the preceding section, non-evaluative *kráčet* presents the movement as if from within, through the lens of the executor of the movement, i.e. refers to the human actor's experiential self. The experiential self may be seen as a form of self-consciousness, and from here it is only a step to its pronounced form, namely, deliberateness.

10.   Conclusion

The contrastive analysis provided in this article has demonstrated that English *walk* and its nearest equivalents in Czech, i.e. *jít* and *kráčet*, display differences in their semantic structure. In contrast to *jít*, the construal of a motion situation in *walk* and *kráčet* involves focus on leg movements (and bodily position), which amounts to a marked segmentation of the motion into individual kinetic quanta. The profiling of leg movements and verticality of the body is even more pronounced in *kráčet*, which endows this verb with the potential to bear reference to the human actor's experiential self. In *walk*, reference to the actor's self only follows from the agentivity of the actor and the specifics of the manner pattern. Reference to the actor's experiential self coupled with marked focus on leg movements and verticality of the body enables *kráčet* to function as an evaluative verb, while such a possibility is not open for *walk* and *jít*. *Walk* thus occupies an intermediate position between the two verbs. In other words, Czech lacks an exact semantic counterpart of *walk*.

The analysis has demonstrated the usefulness of parallel translation corpora in linguistic analysis (for the critical analysis of using parallel corpora in linguistic research cf. Martinková, 2014). As pointed out by Rojo and Ibarretxe-Antuñano (2013, p. 20), transfer in translation process "is not just a mere 'relocation' of some linguistic meanings from one source language into a target language" but involves "a whole decoding process that unveils all the conceptual meaning contained in the concepts, the contexts and the constructions used". The parallel translation corpus has helped uncover aspects of meaning of the verbs under analysis that would probably escape one's attention, thus confirming that multilingual corpora "give new insights into the lexis of the languages compared – insights that are likely to be missed in studies of monolingual corpora" (Altenberg & Granger, 2002, p. 8).

It should be added, however, that a full explication of the differences between these verbs could not be given in the confines of this article. Needless to add, further research is required to include a contrastive analysis of English *go* used in a walking sense. Last but not least, perfective forms of *jít* and *kráčet* also await further investigation.

References

Akita, K. (2017). The typology of manner expressions: A preliminary look. In
     I. Ibarretxe-Antuñano (Ed.), *Motion and space across languages: Theory and applications*
     (pp. 39–60). Amsterdam: John Benjamins. https://doi.org/10.1075/hcp.59.03aki

Altenberg, B., & Granger, S. (2002). Recent trends in cross-linguistic lexical studies. In
    B. Altenberg & S. Granger (Eds.), *Lexis in contrast: Corpus-based approaches* (pp. 3–48).
    Amsterdam: John Benjamins. https://doi.org/10.1075/scl.7.04alt

Boas, H.C. (2008). Towards a frame-constructional approach to verb classification. In
    E.S. Acevedo & F.J.C. Rodríguez (Eds.), *Grammar, constructions, and interfaces. Special
    issue of Revista Canaria de Estudios Ingleses*, 57, 17–47.

Charras, P., & Lupiáñez, J. (2010). Length perception of horizontal and vertical bisected lines.
    Psychological Research, 74(2), 196–206. https://doi.org/10.1007/s00426-009-0243-1

Clausner, T.C., & Croft, W. (1999). Domains and image schemas. *Cognitive Linguistics*, 10(1),
    1–31. https://doi.org/10.1515/cogl.1999.001

Croft, W., Barðdal, J., Hollmann, W., Sotirova, V., & Taoka, C. (2010). Revising Talmy's
    typological classification of complex event constructions. In H.C. Boas (Ed.), *Contrastive
    studies in Construction Grammar* (pp. 201–235). Amsterdam: John Benjamins.
    https://doi.org/10.1075/cal.10.09cro

Faber, P.B., & Mairal Usón, R. (1999). *Constructing a lexicon of English verbs*. Berlin: Mouton
    de Gruyter. https://doi.org/10.1515/9783110800623

Gibbs, R.W. (2006). *Embodiment and cognitive science*. New York: Cambridge University
    Press.

Goddard, C., & Wierzbicka, A. (2014). *Words and meanings: Lexical semantics across domains,
    languages, and cultures*. Oxford: Oxford University Press.

Goddard, C., Wierzbicka, A., & Wong, J. (2016). "Walking" and "running" in English and
    German: The conceptual semantics of verbs of human locomotion. *Review of Cognitive
    Linguistics*, 14(2), 303–336. https://doi.org/10.1075/rcl.14.2.03god

Horecký, J. (1999). Sémantické a derivačné pole slovies *ísť* a *chodiť* [Semantic and derivative
    fields of the verbs *ísť* and *chodiť*]. *Slovenská reč*, 64(4), 202–208.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (1991). Parts and boundaries. *Cognition*, 41(1–3), 9–45.
    https://doi.org/10.1016/0010-0277(91)90031-X

Jackendoff, R., & Landau, B. (1992). Spatial language and spatial cognition. In R. Jackendoff
    (Ed.), *Languages of the mind: Essays on mental representation* (pp. 99–124). Cambridge,
    MA: MIT Press. https://doi.org/10.7551/mitpress/4129.003.0007

Janda, L. (2006). Totally normal chaos: The aspectual behavior of Russian motion verbs.
    Harvard Ukrainian Studies, 28(1–4), 183–193.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and
    reason*. Chicago: University of Chicago Press.
    https://doi.org/10.7208/chicago/9780226177847.001.0001

Kudrnáčová, N. (2013). *Caused motion: Secondary agent constructions*. Brno: Masaryk
    University.

Langacker, R.W. (1987). *Foundations of cognitive grammar. Vol. 1: Theoretical prerequisites*.
    Stanford: Stanford University Press.

Langacker, R.W. (1991). *Foundations of cognitive grammar. Vol. 2: Descriptive application*.
    Stanford: Stanford University Press.

Machek, V. (2010). *Etymologický slovník jazyka českého* [An etymological dictionary of the
    Czech language]. Prague: Nakladatelství Lidové noviny.

Maki, B.E., & McIlroy, W.E. (1997). The role of limb movements in maintaining upright
    stance: The "change-in-support" strategy. *Physical Therapy*, 77(5), 488–507.
    https://doi.org/10.1093/ptj/77.5.488

Malt, B.C., Gennari, S., Imai, M., Ameel, E., Tsuda, N., & Majid, A. (2008). Talking about walking: Biomechanics and the language of locomotion. *Psychological Science*, 19(3), 232–240. https://doi.org/10.1111/j.1467-9280.2008.02074.x

Malt, B.C., Gennari, S., & Imai, M. (2010). Lexicalization patterns and the world-to-words mapping. In B.C. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 29–57). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195311129.003.0003

Malt, B.C., Ameel, E., Imai, M., Gennari, S.P., Saji, N., & Majid, A. (2014). Human locomotion in languages: Constraints on moving and meaning. *Journal of Memory and Language*, 74, 107–123. https://doi.org/10.1016/j.jml.2013.08.003

Martinková, M. (2014). K metodologii využití paralelních korpusů v kontrastivní lingvistice [On the methodology of using parallel corpora in contrastive linguistic research]. *Naše řeč*, 97(4–5), 270–285.

Miller, G., & Johnson-Laird, P. (1976). *Language and perception*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674421288

Nesset, T. (2010). Metaphorical walking: Russian *idti* as a generalized motion verb. In V. Hasko & R. Perelmutter (Eds.), *New approaches to Slavic verbs of motion* (pp. 343–359). Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.115.19nes

Nida, E.A. (1975). *A componential analysis of meaning: An introduction to semantic structures*. The Hague: Mouton.

Phelps, K., & Duman, S. (2012). Manipulating manner: Semantic representations of human locomotion verbs in English and German. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34, 857–862.

Piñón, C.J. (1993). Paths and their names. In K. Beals, G. Cooke, D. Kathman, S. Kita, K. McCullough & D. Testen (Eds.), *Papers from the 29th Regional Meeting of the Chicago Linguistic Society* (pp. 287–303). Chicago: Chicago Linguistic Society.

Rakhilina, E.V. (2004). There and back: the case of Russsian 'go'. *Glossos*, 5, 1–33. [https:// slaviccenters.duke.edu/sites/slaviccenters.duke.edu/files/media_items_files/5rakhilina .original.pdf]

Rojo, A., & Ibarretxe-Antuñano, I. (2013). Cognitive linguistics and translation studies: Past, present and future. In A. Rojo & I. Ibarretxe-Antuñano (Eds.), *Cognitive linguistics and translation: Advances in some theoretical models and applications* (pp. 3–30). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110302943.3

Saicová Římalová, L. (2010). *Vybraná slovesa pohybu v češtině: Studie z kognitivní lingvistiky* [Selected verbs of motion in Czech: A study in cognitive linguistics]. Prague: Karolinum.

Sheets-Johnstone, M. (2012). Fundamental and inherently interrelated aspects of animation. In A. Foolen, U.M. Lüdtke, T.P. Racine & J. Zlatev (Eds.), *Moving ourselves, moving others: Motion and emotion in intersubjectivity, consciousness and language* (pp. 29–55). Amsterdam: John Benjamins. https://doi.org/10.1075/ceb.6.02she

Slobin, D.I. (1997). Mind, code, and text. In J. Bybee, J. Haiman & S.A. Thompson (Eds.), *Essays on language function and language type: Dedicated to T. Givón* (pp. 437–467). Amsterdam: John Benjamins. https://doi.org/10.1075/z.82.24slo

Slobin, D.I. (2000). Verbalized events: A dynamic approach to linguistic relativity and determinism. In S. Niemeier & R. Dirven (Eds.), *Evidence for linguistic relativity* (pp. 107–138). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.198.10slo

Slobin, D. I. (2005). Relating narrative events in translation. In D. D. Ravid & H. B. Shyldkrot (Eds.), *Perspectives on language and language development: Essays in honor of Ruth A. Berman* (pp. 115–129). Dordrecht: Kluwer. https://doi.org/10.1007/1-4020-7911-7_10

Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59–81). Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.66.05slo

Slobin, D. I., Ibarretxe-Antuñano, I., Kopecka, A., & Majid, A. (2014). Manners of human gait: A cross-linguistic event-naming study. *Cognitive Linguistics*, 25(4), 701–741. https://doi.org/10.1515/cog-2014-0061

Slovník spisovné češtiny pro školu a veřejnost [A dictionary of standard Czech for school and home]. (2009). Eds. J. Filipec, F. Daneš, J. Machač & V. Mejstřík. Prague: Academia.

Slovník spisovného jazyka českého [A dictionary of standard Czech]. (1960). Eds. B. Havránek, J. Bělič, M. Helcl, A. Jedlička, V. Křístek & F. Trávníček. Prague: Nakladatelství Československé akademie věd.

Snell-Hornby, M. (1983). *Verb-descriptivity in German and English: A contrastive study in semantic fields*. Heidelberg: Carl Winter Universitätsverlag.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description. Vol. 3: Grammatical categories and the lexicon* (pp. 57–149). Cambridge: Cambridge University Press.

Talmy, L. (2000). *Towards a cognitive semantics*. Cambridge, MA: MIT Press.

Vidaković, I. (2012). 'He walked up the pole with arms and legs': Typology in second language acquisition. In L. Filipović & K. M. Jaszczolt (Eds.), *Space and time in languages and cultures* (pp. 233–261). Amsterdam: John Benjamins. https://doi.org/10.1075/hcp.36.14vid

Vulchanova, M., & Martinez, L. (2013). A basic level for the encoding of biological motion. In C. Paradis, J. Hudson & U. Magnusson (Eds.), *The construal of spatial meaning: Windows into conceptual space* (pp. 144–168). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199641635.003.0009

Wienold, G. (1995). Lexical and conceptual structures in expressions for movement and space: With reference to Japanese, Korean, Thai, and Indonesian as compared to English and German. In U. Egli, P. E. Pause, C. Schwarze, A. von Stechow & G. Wienold (Eds.), *Lexical knowledge in the organization of language* (pp. 301–340). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.114.14wie

Wierzbicka, A. (1988). *The semantics of grammar*. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.18

Parts of speech membership as a factor of meaning extension and level of abstraction

Comparison of Czech adjectives and Japanese verbs in adnominal modification

Petra Kanasugi
Charles University

Czech and Japanese show formal differences in adnominal modification. Czech tends to utilize adjectives for both classification and qualification purposes whereas Japanese tends to express classification by compounding and to use a whole range of parts of speech for qualification. As a result, part of speech membership often differs between the Czech and Japanese rendering of the same referential content. It has been shown that parts of speech dispose of schematic meaning which contributes to conceptualization. Based on the results of corpus analysis, I argue that the difference in part of speech membership results in different tendencies in meaning extension and ultimately in different meaning of the two counterparts, Czech adjectives are more abstract and schematic, while Japanese verbs are more concrete.

Keywords: part of speech, adnominal modification, adjective, verb, meaning extension, metonymy, metaphor, Czech, Japanese

1. Introduction

Japanese is an agglutinative, SOV language. Its agglutinative nature and the historical fact of having accepted Chinese characters, which brought to Japanese an additional set of very freely combinable morphemes, result in a very large vocabulary (dictionary Nihon kokugo daijiten (2000–2002) lists approximately 600,000 entries)[1] particularly in the category of nouns.[2] Moreover, vocabulary is further flex-

---

1. Comparable Czech dictionary *Příruční slovník jazyka českého* (1935–1957) has 250,000 entries.

2. According to a lexicological survey (Nomura, 2011), more than 80% of vocabulary consists of nouns (see Okimori, Kimura, Tanaka, & Chin, 2011).

ibly extended due to high productivity of word-forming processes and openness to borrowing from other languages (Kageyama & Kishimoto, 2016, pp. 11–12). The relative ease of new word coinage and readiness for extensive borrowing may in effect lessen the pressure for meaning extension in case of individual lexical items.

However, meaning extension is not solely a matter of individual lexical items but also a matter of grammatical constructions (Croft, 2006; Sullivan, 2013, etc.). The present study also focuses on one structural difference between Czech and Japanese, specifically the difference in the range of parts of speech, which can acquire the function of attribute[3] and explores modes of meaning extension in differing patterns. Japanese, as many SOV languages, allows adnominal modification by verbs, whereas the Czech language does not. The basic aim is to illustrate how this structural trait effects the extension of meaning and thus the degree of abstractness of a given word.

2.   The target structural difference and its implications

According to Štícha, the adnominal modification (i.e. congruent adnominal modification) in the Czech language typically takes the form of an adjective,[4] but it can also take the form of a noun, a numeral or a pronoun (Štícha, 2013, p. 725). However, numerals and pronouns in Czech grammatical theory are defined primarily semantically and as such form a non-basic part of speech class (Štícha, 2013, p. 82). Their morpho-syntactic behavior in the given construction is basically consistent with that of adjectives and nouns.

The Japanese language offers a wider range of formal options for adnominal modification. There are nominal attributes (this category traditionally subsumes numerals and pronouns) and adjectival attributes, which parallel their Czech counterparts. There are two categories of adjectives in Japanese (so-called *keiyōdōshi* and *keiyōshi* or NA adjectives and I adjectives), the former leaning towards *nouny* and the latter to *verby* characteristics (Uehara, 1998, p. 94). In addition to these options, there are also attributes consisting of adnominal adjectives

---

3. Moreover, the discussion is limited to the cases rendered in Czech by congruent adnominal modification because the different patterns of incongruent modification in Czech are often rendered in Japanese by a dependent clause, rather than by an attribute.

4. Adjectives are a morphologically considerably varied group, there are un-derived so-called qualitative adjectives, deverbal adjectives, denominal adjectives and even deadjectival adjectives. The deverbal adjectives fall into two groups: paradigmatic (formed productively from one of four possible verbal forms) or formed by suffixation. (For detailed discussion see Štícha, 2013, pp. 196–212).

and verbs. Adnominal adjectives form a closed class category which consists of deictic items (*kono*[5] (this), *sono* (that) …), clusters which have undergone universation (*arayuru* (all), *ikanaru* (which)…), and individual items which ended up in this category through various historical reasons (*ōkina* (big), *okashina* (funny)…) (Nihongo Kijutsu Bunpō Kenkjūkai, 2013, p. 110). However, it is safe to say that the relevance of this category for our purposes is rather small due both to its closed class character and to its very limited use in nominal modification.[6] The present study focuses on the verbal attributes as they are both more prevalent and, unlike adnominal adjectives, form an open class category. It should also be stressed that these verbal attributes are not a second option to *standard* adjectives. Verbal attributes are often the non-marked and typical way of expressing the relevant meanings. Synonymous onomatope are available in some cases, but it is the onomatopoeias which are the stylistically and otherwise marked version.

There is another crucial difference between Czech and Japanese attributes. The Czech language utilizes the same structure (adjective + noun) both for qualification (predicating construction) and classification purposes (domain construction), whereas Japanese tends to separate these two functions. The results of a previous corpus study (Kanasugi, 2016) show that qualification in Japanese is expressed predominantly by modification but classification tends to be expressed by compounding.[7] Compounding as such is for agglutinative Japanese very natural, according to statistics Sino-Japanese compounds (*kango*) form approximately 49.7% of all vocabulary and use of compound words in magazines exceeds both in types and tokens 40% (Okimori et al., 2011, pp. 40–41). Moreover, compound nouns are besides Sino-Japanese morphemes formed also from native,

---

5. Throughout this paper, conceptual structures are written in small caps. Lexical items are in italics; examples are glossed based on the Leipzig Glossing Rules using the following set of grammatical tags: 1 – first person, 2 – second person, 3 – third person, ABL – ablative, ACC – accusative, AUX – auxiliary noun, CAUS – causative, CLF – classifier, COMP – comparative, CONC – concessive, COND – conditional, COP – copula, DAT – dative, GEN – genitive, GER – gerundium, INF -infinitive, INS -instrumental, LOC – locative, MIM – mimetic, NEG – negative, NOM – nominative, NONPST – non-past, PASS – passive, PERF – perfect, PL -plural, POL – polite, POLCOP – polite copula, POTEN – potential, PST – past, QUOT – quotative, REFL – reflexive, SFP – sentence final particle, SG – singular, SIMUL – simultaneous, TOP – topic, VOL – volitional. Information about grammatical gender and number is limited to instances where judged necessary, otherwise it is abbreviated or expressed by the form of the English equivalent.

6. The two frequently used members of the class, i.e. deictic expressions *kono* and *sono* function as grounding rather than modifying element.

7. The study included 687 Czech classifying and 692 qualifying ADJ + NOUN constructions. These were found to correspond to noun phrase in 12%, respectively 55%, to Japanese native compounds in 64%, respectively 24%, to Sino-Japanese compounds 15%, respectively 11% and to words derived by suffixation in 10% for both types.

Table 1. Parts of speech in adnominal modification in Czech and Japanese

Attribute form	Czech language	Japanese language
Adjective (in Japanese: I and NA adjectives) (in Czech: including some non-basic parts of speech pronouns and numerals)	*ohnutý hřebík* (a bent nail)	*nagai bō* (long stick)
	její promluva (her utterance)	*rippana bō* (mighty stick)
	třetí přání (third wish)	
Noun (in Japanese: including pronouns and numerals) (in Czech: including some pronouns and numerals)	*řeka Amazonka* (river Amazon)	*kawa no Amazon* (river Amazon)
	jeden příběh (one story)	*kare no hatsuwa* (his utterance)
	ten muž (that man)	*mittsu no kibō* (three wishes)
Verb	–	*magatta kugi* (a bent nail)
Adnominal adjectives	–	*sono hanashi* (that thing)
		chīsana ibiki (a small snore)
		iwayuru mondai (so-called problem)

foreign-origin and mimetic morphemes and there are apart from N-N also V-N, ADJ-N compound patterns (for details see Kageyama & Saito, 2016). The proclivity to compounding can be interpreted as a demonstration of language iconicity as one concept is in this manner expressed by one word.

The above-mentioned differences lead to a situation where the predicating construction NP takes on different forms, specifically an adjective + noun in Czech and a verb + noun in Japanese (see the shaded cells in Table 1).

The nominal phrase as such subsumes three kinds of symbolic relationships which participate in the construction of its meaning. First is the relationship between the phonological form and its semantic content (so-called lexical meaning) which is structured by the second symbolic relationship – namely the syntactic relationship between the two constituents – and the third symbolic relationship – namely between the part of speech category and its schema. Lexical meaning is not unproblematic because different languages are far from segmenting the same semantic fields in easily comparable fashion. However, since the

Table 2. Noun classification and qualification in Czech and Japanese

	Czech language	**Japanese language**
Classification	*bílý pásek* (a white belt)	*shiroobi* (a white belt in karate etc.)
	vařené vajíčko (a boiled egg)	*yudetamago* (a boiled egg as a kind of dish)
Qualification	*bílý pásek* (a white belt)	*shiroi obi* (a white sash)
	vařené vajíčko (a boiled egg)	*yudeta tamago* (a boiled egg)

focus of the study is rather structural, we will focus on the latter two schematic structures.

The importance of meaning which is brought about by the structural organization of the string can be easily demonstrated on the case of numeral classifier use in Japanese. Each of the two sentences below, (1) and (2), which differ only in the position of the numeral classifier, expresses a clearly different meaning:

(1) *Kinō   ichijikan no horā eiga o   mita.*
     yesterday one-hour GEN horror movie ACC see:PST
     I watched an hour long horror movie yesterday.

(2) *Kinō   horāeiga   o   ichijikan mite   ita.*
     yesterday horror-movie ACC one-hour see:GER be:PST
     I was watching a horror movie for one hour yesterday.

The structural position of the numeral classifier in front of the noun in (1) enforces the interpretation of the numeral as a quality-ascribing modification. This relationship between the noun and its modification is in Czech enforced by congruence. A strong tendency to express this relationship overtly by congruence can be demonstrated by the vast number of denominal and deverbal adjectives in the Czech language. We can say that there is a strong tendency in the Czech language for the part of speech schema and the nominal modification schema to be in mutual accord.

The schematic meaning of part of speech membership is discussed in detail by Langacker (1987, 1991, 2008, 2015, etc.). It is best demonstrated by an example of a simple noun phrase schema such as *hubená koza* (scrawny goat) in Czech and corresponding *yaseta yagi* in Japanese. The Japanese version uses the verbal modification *yaseta*, which is a past perfect form of the verb *yaseru* (to lose weight),

while the Czech version uses the adjective *hubený*[8] (slim). The schema differs accordingly with the Czech version placing the trajector (the goat) on the slimness scale as being slimmer than average, whereas the Japanese version constructs the goat as a trajector in the last stage of the process of losing weight.
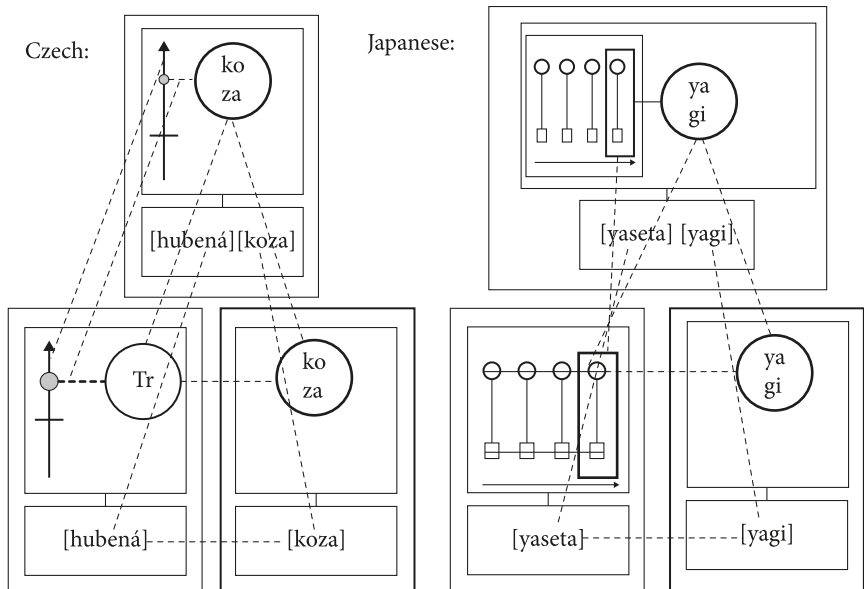
Figure 1. Construction of nominal phrase with adjectival and verbal modification

The prototypical meaning related to the attribute is a lasting quality of the modified noun. In other words, the two schematic meanings (structural schema and part of speech schema) are in accord in the case of the Czech phrase but seem to work rather against each other in the case of the Japanese phrase, since the verbal schema in Japanese presupposes a process and a completed change.

Psycholinguistic research supports Langacker's claim that part of speech categories are in essence semantic. Imai et al. (2008) specifically addresses the verbal category and shows that three-year-old Japanese children in some contexts, and five-year-old Japanese children in all contexts, understand the meaning of verbal category and are able to map and extend novel verbal vocabulary.[9]

---

8. The adjective is etymologically related to the verb *hubnout* which is imperfective and thus unlike *yaseru*, does not profile completion of change.

9. Young children shown a situation where an unknown object was manipulated in an unfamiliar fashion were capable of mapping the noun to the object and the verb to the activity. In

Another interesting psycholinguistic insight by Maass, Politi, Karasawa, and Suga (2006) indicates that Japanese native speakers (university students) in context of describing other people where a choice between adjectives and verbs was possible, consistently preferred the verbal modification to its adjectival counterpart. Maass et al. (2006) link this tendency to preference for a concrete, context-dependent expression.[10] In any case, the results of the study point to the conclusion that for Japanese speakers verbal attributes are not just a systematic necessity but a meaningful and important choice.

3.   Relevant previous research on meaning and meaning extension

The present analysis subscribes to an understanding of meaning and meaning extension as outlined by the Dynamic Usage-Based Model (Langacker, 2000, pp. 91–146). Specifically, the understanding that the meaning is extended when a target is within an acceptable degree of strain recognized as matching the instance of prototypical meaning. "Extension can thus be thought of as recognition achieved at the cost of invoking a schematized version of the categorizing structure, one whose coarser-grained specifications are satisfied by the target." (Langacker, 2000, p. 102). Accordingly, adding further extensions inevitably leads to the forming of an increasingly abstract schematic structure.

   The mechanisms of such meaning extension as elaborated by Croft (2006, pp. 269–302) are domain highlighting in the case of metonymy and domain mapping in the case of metaphor. In other words, in the case of metonymy, the domain which is secondary for the original meaning is primary for the extended meaning; in the case of metaphor, the target domain is conceptualized using the structure provided by the source domain. Conceptual metaphor research (Lakoff, Johnson, Kövecses, etc.) has pointed out that metaphors are not an issue pertaining merely to literary language but also a very common cognitive tool. Grady (1997), further developing conceptual metaphor theory, has convincingly argued that there are basically two kinds of mechanisms which lead to metaphor. The first is based on a recurring experience of correlation between two physical or perceptual facts or subjective responses within so-called primary scenes. This

---

other words, even three-year-old Japanese children have some appreciation of verbal schema and some expectations as to what kind of meaning should be carried by a verb.

10. It should be pointed out that construction grammar approaches (Langacker, 2008; Goldberg, 2006) persuasively link such repetitive use under a given set of circumstances with formation of correlations between surface forms and semantic interpretations which constitute basis of knowledge of language.

mechanism leads to primary metaphors which are witness to cognitive similarities across humankind and are broadly shared in many languages. The other type of metaphor is based on a similarity which is figured out and not directly experienced and thus more culturally specific.

Metonymy has also been a subject of vibrant research. Here, it seems especially pertinent to point out Dirven's (2002) study, which proposes the classification of metonymy based on the cline of conceptual closeness and conceptual distance. The study proposes three kinds of metonymy: linear, conjunctive and inclusive: "a linear metonymy is a type of language use where the intended referent is named by a conceptual category that has a different, but closely related, referential mass than the common expression(s) used for the intended referent ... [conjunctive metonymy] while keeping the original meaning of an expression, entails a necessary and systematic extension of this original meaning ... a third type [inclusive metonymy] always receives a figurative interpretation" (Dirven, 2002, p. 107). Besides metonymy classification, Dirven proposes in the same paper the literal-figurative continuum in the figure below, which illustrates the above-mentioned types of metonymy in a given order followed by metaphor.
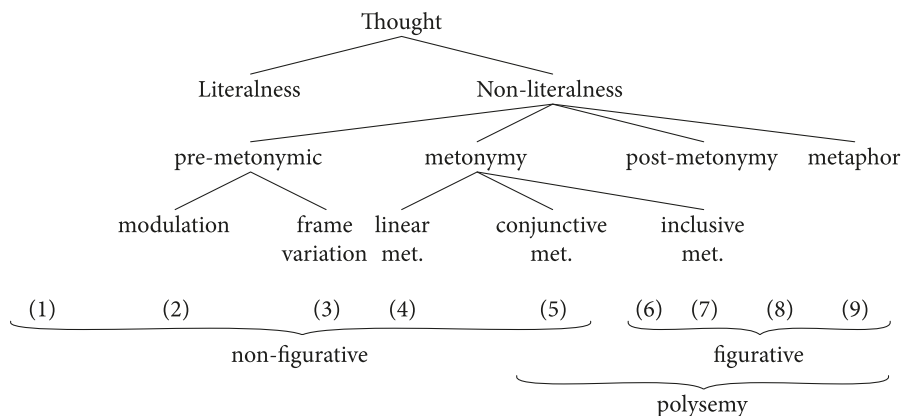
Figure 2. The literalness – Non-literalness continuum (Dirven, 2002, p. 107)

Günter Radden also assumes a continuum from literalness to metonymy and metaphor, but fittingly points out that primary metaphors are, due to the correlation of experience and the juxtaposition of both target and source concept in the same primary scene, metonymic in character. This is why, he then proceeds to re-interpret correlation by asserting that "correlation is (thus) a fundamentally metonymic relationship" (Radden, 2002, p. 95).

4.   Research question, methodology and results

We have seen in Section 1 that there is a systematic difference between Japanese and Czech adnominal modification which consists of different structural possibilities for part of speech realization of attributes. As a result, there are cases of noun phrases with the same referent attributes which are in Czech and in Japanese languages rendered differently – by adjectives in Czech and verbs in Japanese.

We have also seen that parts of speech are connected with a certain schema. Adjectives typically place their trajector on a particular quality scale and the verbs (all of the analyzed verbs are so-called verbs of change (*henka dōshi*)) fit the trajector with the profiled state resulting from a change. The psycholinguistic research has convincingly shown that part of speech meaning (schematic meaning) is psychologically real and speakers have a good grasp of it from a very young age. Moreover, there are also studies connecting verbs to concrete meanings and adjectives to abstract meanings.

On the other hand, we have also seen that the attribute position within the noun phrase itself carries certain schematic meaning which then influences the interpretation of the given attribute.

The above leads to two basic questions:

1. Is the verbal schema in the case of Japanese verbal attributes constant across various meanings of a given word, or does it undergo some modulation due to its fulfilling the syntactic position typical for adjectives?
2. Is the meaning of Czech and Japanese attributes realized by different parts of speech the same or do the processes of meaning extension in the two cases differ? In other words, is the suggested tendency of verbs to have a concrete meaning and adjectives to have an abstract meaning true in the context of noun phrase?

To be able to propose an initial answer to these research questions, a corpus study and meaning analysis were conducted. The Czech language data was excerpted from the Czech National Corpus (CNK), with the researched string being the given adjective (in all genders and both numbers) and any noun. Idiomatic uses where the attribute has a classifying function (e.g. *white wine*, *White House*, etc.) were manually checked and removed from the retrieved data. The Japanese data were excerpted from the Balanced Corpus of Contemporary Written Japanese (BCCWJ). The researched string had up to three variations: the given verb in past perfect (so-called TA form) followed by a noun, the given verb in present imperfect (so-called TE-IRU form) followed by a noun and, where the form is available as it is in the case of the Japanese equivalent of *thick* (*futoi*), an adjective followed by a noun pattern was also included. Both verbal forms were searched since they

are functionally equivalent and can be freely substituted when expressing states (Suzuki, 1982, pp. 402–403).

Both corpora used are balanced corpora of written language, though the CNK is approximately three times larger than the BCCWJ, which is reflected in the number of retrieved target structures. However, even the smaller Japanese corpus is balanced and representative, which gives us sufficient confidence that all relevant meanings in use are covered. All uses were analyzed in context and individual uses were classified as either literal or non-literal. The non-literal uses were further classified as a certain kind of metonymy or metaphor (with the inclusive metonymy not applicable). The detailed categorization system is outlined by Table 3 below. However, with regards to the figurativeness of language, Radden's (2002) argumentation which interprets primary metaphors as non-figurative language was adopted.

Table 3. Classification of retrieved structures

Literal use	
Non-literal use	Linear metonymy
Non-figurative	
	Primary metaphor
Figurative	Conjunctive metonymy
	Similarity metaphor

5.   Results

The numbers of retrieved structures are summarized in Table 4.

The following subsections will present examples of individual uses of the target forms. The examples are listed according to literal – figurative continuum. Further commentary and comparison of the Czech and Japanese results follows after the examples and in the general discussion at the end of the section.

Table 4. Summary of excerpted structures

Target attribute CZ	Number of hits	Target attributes JPN	Number of hits
mokrý + N (wet + N)	2455	*nureta* + N	602
		nureteiru + N	52
suchý + N (dry + N)	4067	*kawaita* + N	660
		kawaiteiru + N	25
živý + N (life + N)	7518	*ikita* + N	568
		ikiteiru + N	404
mrtvý + N (dead + N)	5134	*shinda* + N	2527
		shindeiru + N	28
tlustý + N (thick + N)	2620	*futoi* + N	1163
		futotta + N	163
		futotteiru + N	99

5.1 *Mokrý* versus *nureta, nureteiru* (wet)

5.1.1 *Detected types of uses for the Czech adjectival attribute* mokrý *(wet):*

Literal use: containing or covered with fluid

(3) a. *Koberec,   který   se   dal   setřít   **mokrým** hadrem.*
carpet.nom which ref could.be wipe:inf wet:ins   cloth:ins
A carpet, that could be wiped by a wet cloth.

Linear metonymy: the visual aspect of wetness highlighted

(3) b. *Zvládne   **mokrý** efekt   a   navíc   hřívu*
manage:3sg wet:acc effect.acc and moreover shock.of.hair:acc
ochrání   před   vysušením.
protect:3sg against drying:ins
It can manage a glossy effect and what's more it protects your hair from
drying out.

Linear metonymy: the auditive quality brought about by wetness highlighted

(3) c. *V tu chvíli   se   kapela   odmlčela   a   ozvalo   se*
in that moment refl band:nom went.quiet:3sg and sound:3sg.pst refl
*tlumené,   **mokré**   mlasknutí.*
muffled:nom wet:nom smack.of.lips.nom
At that moment, the band went quiet and a muffled wet smack of lips was
heard.

Linear metonymy: the tactile quality brought about by wetness highlighted

(3) d. *Vzbudil   ji **mokrý** dotyk   hladkého   jazyka   a*
      woke.up:3SG her wet:NOM touch.NOM smooth:GEN tongue:GEN and
      horečnatý   dech   nasáklý   vůní   trávy.
      feverish:NOM breath.NOM soaked:NOM scent:INS grass:GEN
      The wet touch of a soft tongue and a feverish breath smelling of grass
      woke her up.

Linear metonymy: the sensation of wetness highlighted

(3) e. *Opravdové, poctivé,   **mokré**   slzy.*
      real:NOM.PL honest:NOM.PL wet:NOM.PL tears:NOM
      Actual, honest, wet tears.

Linear metonymy: the quality of SMEARING OR BEING STILL SOFT due to containing liquid highlighted

(3) f. *Ukotvil   ji do **mokrého** betonu   a   spojovací*
      anchor:3SG.PST her into wet:GEN cement:GEN and connecting.ACC.PL
      díly   základny svařil [...]
      parts:ACC base:GEN weld:PST
      He anchored it into the wet cement and welded the connecting parts of
      the base [...]

Linear metonymy: the quality of HAVING HIGH PRECIPITATION highlighted

(3) g. *Jenže byl tenkrát **mokrý** rok   a   voda   v žumpě*
      but   was then   wet:NOM year.NOM and water:NOM in cesspool:LOC
      hodně stoupla.
      a lot   rise:PST
      But it was a wet year back then and the water in the cesspool has risen a
      lot.

Primary metaphor: primary metaphor AFFECT IS MOISTURE – EMOTIONALLY MOVED

(3) h. *[...] teprve jsem uviděl,   že   maminka   má **mokré***
          only   then see:1SG.PST that mummy:NOM has wet:ACC.PL
      oči.
      eyes:ACC.PL
      [...] only then I noticed, that Mom has wet eyes.

Primary metaphor: primary metaphor AFFECT IS MOISTURE – SEXUALLY AROUSED

(3) i. *Objevuje se v mnoha **mokrých** snech téhle ulice.*
appear:3SG REFL in many wet:LOC.PL dreams:LOC this:GEN street:GEN
She stars in many of the wet dreams of this street.

Conjunctive metonymy: RELATED TO SEAS, RIVERS, ETC.

(3) j. *Nedobytné obrnĕné koule [...] ze sebe*
unconquerable:ACC.PL armed:ACC.PL balls:ACC from themselves
*vytvářejí živočichové všech kontinentů i **mokrého***
create:3PL animals:NOM all:GEN.PL continents:GEN even wet:GEN
živlu.
element:GEN
On every continent and even in waters (wet element) [...] there are animals which can turn themselves into unconquerable armed balls [...]

Conjunctive metonymy: USING WATER

(3) k. *Jestli máte rádi vodu a všemožné **mokré** atrakce,*
if like:2PL water:ACC and various:ACC.PL wet:ACC.PL attractions:ACC
berte to jako zajímavý tip na návštĕvu.
take it as interesting:ACC tip.ACC for visit:ACC
If you enjoy water and different water attractions, here is an interesting tip for holidays.

Conjunctive metonymy: RELATED TO ALCOHOL OR ITS CONSUMPTION

(3) l. *[...] dvojice kamarádů z **mokré** čtvrti*
[...] couple:NOM friends:GEN from wet:GEN quarter:GEN
nezvládla lávku [...]
not.manage:3SG.PST bridge:ACC
[...] the couple from the wet quarters did not manage to cross the bridge [...]

5.1.2 *Detected types of uses for the Japanese verbal attribute* nureta *(wet):*

Literal use: CONTAINING OR COVERED WITH FLUID

(4) a. ***Nureta** taoru o torikaete iru uchi ni, [...]*
wet towel(s) ACC change:GER be:NONPAST while DAT
While I was changing the wet towels, [...]

Linear metonymy: the visual aspect of wetness highlighted

(4) b. ***Nureta** iro   dake ni   nari,   sore mo miru   aida ni   kawaita.*
         wet   color only DAT become that also see:NONPAST while DAT dry:PST
         It became just a wet-looking color and even that dried up while (he) was
         watching.

Linear metonymy: the auditive quality brought about by wetness highlighted

(4) c. *Gutcha gutcha   to,   **nureta** oto   ga   mimi ni   tsuita.*
         plopping.sound.MIM QUOT wet   sound NOM ear(s) DAT come.to:PST
         The plopping wet sound came to my ears.

Linear metonymy: the tactile quality brought about by wetness highlighted

(4) d. *Te   o   tsuku to   jittori   to   **nureta** kankaku, [...]*
         hand(s) ACC reach COND damp.MIM QUOT wet   feeling
         When my hand touched it, I felt a damp wet feeling, [...]

Primary metaphor: primary metaphor AFFECT IS MOISTURE – EMOTIONALLY
MOVED

(4) e. *[...], Shinobu wa **nureta** hitomi   o   matatakaseta.*
                 Shinobu TOP wet   pupuil(s) ACC blink:CAUS:PST
         Shinobu [...] kept blinking her wet eyes.

Primary metaphor: primary metaphor AFFECT IS MOISTURE – SEXUALLY AROUSED

(4) f. *"Aaa" omowazu,   Reiichi kara **nureta** aegi ga   koboreochiru.*
         Aaa   instinctively Reiichi ABL wet   sigh NOM spill:NONPST
         "Aaa" Rei'ichi let out a gasping wet sigh.

5.1.3 *Detected types of uses for the Japanese verbal attribute* nureteiru *(wet):*

Literal use: CONTAINING OR COVERED WITH FLUID

(5) *Tentō tenraku   shiyasui basho, [...], **nureteiru** yuka [...]*
     fall(s) tumble(s) do:easily place   wet   floor
     Places where one can easily trip or fall: [...], a wet floor [...]

Table 5. Summary of Czech and Japanese counterparts of WET

	Meaning	CZ *mokrý*	JP *nureta*	JP *nureteiru*
Literal use	CONTAINING/COVERED WITH LIQUID	(3a)	(4a)	(5)
Linear metonymy	RELATED TO LITERAL VISUAL QUALITY	(3b)	(4b)	
	RELATED TO LITERAL SOUND QUALITY	(3c)	(4c)	
	RELATED TO LITERAL HAPTIC QUALITY	(3d)	(4d)	
	STRESS ON WETNESS SENSATION	(3e)		
	STILL SMEARING OR SOFT	(3f)		
	WITH ABUNDANT PRECIPITATION	(3g)		
Primary metaphor	EMOTIONALLY MOVED	(3h)	(4e)	
	SEXUALLY AROUSED	(3i)	(4f)	
Conjunctive metonymy (figurative)	RELATED TO SEAS, RIVERS, ETC.	(3j)		
	USING WATER/IN WATER	(3k)		
	RELATED TO ALCOHOL CONSUMPTION	(3l)		

5.1.4 *Summary and discussion of Czech and Japanese expressions of WET*

The literal meaning WET is shared by all three forms (3a), (4a), (5). The Japanese verbal adjective *nureteiru* does not display any non-literal uses. Its shorter counterpart *nureta* appears in linear metonymy and primary metaphor extensions. These uses are very concrete in their accentuating some aspect of wetness or water: its appearance in (4b), a sound connected to wetness (4c), and its tactile quality (4d). The primary metaphor uses (4e), (4f) correspond to the AFFECT IS MOISTURE metaphor (Grady, 1997, p. 294).

The Czech adjective *mokrý* exhibits, besides the linear metonymy and primary metaphor extensions detected in the case of *nureta*, three additional linear metonymic uses (3e)–(3g). The first one stresses the all-encompassing sensation of wetness (*wet water, wet tears,* etc.) and the second one is used for materials which change their consistency from a liquid to a non-liquid while SMEARING OR BEING STILL SOFT (*soft cement, wet paint*, etc.). The third one expresses HIGH PRECIPITATION. There are also three conjunctive metonymies, the use in (3j) BELONGING TO

WATERS (*rivers, seas,* etc.), the use in (3k) activities USING WATER, and the use in (3l) for RELATED TO ALCOHOL CONSUMPTION. These uses are figurative and were not detected in the BCCWJ.

5.2   *Suchý* versus *kawaita, kawaiteiru* (DRY)

5.2.1   *Detected types of uses for the Czech adjectival attribute* suchý *(dry):*

Literal use: FREE FROM LIQUID

(6) a. *Naopak   silnice   mokrá   po   dešti   je mnohem*
on the contrary road:NOM wet:NOM after rain:LOC is much_more
„hlučnější" než **suchá**   *vozovka, […]*
noisier   than dry:NOM road:NOM
On the contrary, a road wet after rain is much noisier than a dry road, […]

Linear metonymy: the auditive quality related to dryness highlighted

(6) b. *Ten* **suchý**   *praskot   zesílil.*
that dry:NOM cracking.NOM grew_stronger
The dry cracking sound grew stronger.

Linear metonymy: the tactile quality related to dryness highlighted

(6) c. **Suchý**   *olej   vmasírujte do postižených   míst.*
dry:ACC oil.ACC massage_in into affected:GEN.PL places:GEN.PL
Massage the dry oil into the affected areas.

Linear metonymy: the quality of LACKING AIR HUMIDITY highlighted

(6) d. *Taková je bilance   možných   následků* **suchého** *vzduchu v*
such   is balance:NOM possible:GEN:PL effects:GEN dry:GEN air:GEN in
bytě.
apartment:LOC
That is the list of possible effects of dry air in your apartment.

Linear metonymy: the quality of LACKING PRECIPITATION highlighted

(6) e. *[…] současné* **suché**   *počasí   ovlivňuje zdroje   velice*
current:NOM dry:NOM weather.NOM affect:3SG sources:ACC very
kvalitní   pitné   vody […]
quality:GEN drinking:GEN water:GEN
[…] the current dry weather affects sources of high quality drinking water […]

Linear metonymy: the quality of LACKING INNER MOISTURE highlighted

(6) f. *Na věnečku   jsou svíčky […]   i   **suché**   květy.*
         on wreath:LOC are candles:NOM.PL and_also dry:NOM:PL flowers:NOM
         There are candles[…] and also dried flowers on the wreath.

Primary metaphor: AFFECT IS MOISTURE – NOT SHOWING EMOTIONAL INVOLVE-
MENT

(6) g. *"Přesně tak, mylorde," odpověděl arcibiskup se   **suchým** úsměvem.*
         exactly so   my lord   answer:PST archbishop with dry:INS smile:INS
         "Exactly, my lord," the archbishop replied with a dry smile.

Conjunctive metonymy: LACKING FAT OR MOISTURE

(6) h. *Pečivo a **suchá**   masa, […] , „uvízly" […]*
         pastry and dry:NOM:PL meats:NOM stuck:PL
         […] pastries and dry meat "stuck" […]

Conjunctive metonymy: LACKING GRAVY, BUTTER ETC.

(6) i. *[…] je lepší jíst   **suché**   brambory   ve Zlíně   než biftek*
         (it) is better to_eat dry:ACC.PL potatoes:ACC in Zlin:LOC than steak:ACC
         v Gottwaldově.
         in Gottwaldov:LOC
         […] it is better to eat dry potatoes in Zlin than a steak in Gottwaldov.

Conjunctive metonymy: SCRAWNY

(6) j. *Přednosta   stanice,   **suchý**   mužík   s   červenou čepicí […]*
         master:NOM station:GEN dry:NOM man:NOM with red:INS   hat:INS
         The station master, a scrawny man with a red hat […]

Conjunctive metonymy: LACKING ALCOHOL OR ITS CONSUMPTION

(6) k. *Pacientům závislým   na alkoholu   se   doporučuje*
         patients:DAT addicted:DAT to alcohol:LOC REFL is_recommended
         *tzv.   „**suchá**   domácnost", […]*
         so_called dry:NOM household:NOM
         Patients with an alcohol addiction are advised to keep a dry household,
         […]

Similarity metaphor: LACKING COMMENTARY ETC.

(6)   l.   *Zkrátíme   to na **suchá**   čísla.*
           will_shorten:1PL it to dry:ACC.PL numbers:ACC
           We will limit (ourselves) to dry numbers.

Similarity metaphor: LACKING WIT, BORING

(6)   m.   *Do škatulky **suchého** kravaťáka, […] Jan Bárta jaksi   nezapadá.*
           to box:GEN dry:GEN suit_boy:GEN   Jan Bárta somehow doesn't_fit
           Jan Barta does not really fit in the category of a boring suit guy.

5.2.2   *Detected types of uses for the Japanese verbal attribute* kawaita *(dry):*

Literal use: FREE FROM LIQUID

(7)   a.   ***Kawaita** fukin de kurumi 1 jikan gurai   shitsuon   ni*
           dry   cloth INS wrap   1 hour approximately room-temperature LOC
           oku.
           place:NONPST
           Wrap (it) in a dry cloth and leave it at room temperature for one hour.

Linear metonymy: the auditive quality related to dryness highlighted

(7)   b.   *Seki   wa gohogoho   to   iu **kawaita** seki.*
           cough TOP hacking.MIM QUOT COP dry   cough
           The cough is a hacking dry cough.

Linear metonymy: the tactile quality related to dryness highlighted

(7)   c.   *Sarasara   to   **kawaita** yuki ga hō   o   kasumeta.*
           powdery.MIM QUOT dry   snow NOM cheek(s) ACC brush:PST
           Dry snow brushed against (her) cheeks.

Linear metonymy: the quality of LACKING AIR HUMIDITY highlighted

(7)   d.   *Sawayakana **kawaita** kūki no naka de Manjirō wa me   o*
           refreshing   dry   air GEN midst LOC Manjirō TOP eye(s) ACC
           tojiru.
           close:NONPST
           Manjiro closes his eyes in the fresh dry air.

Linear metonymy: the quality of LACKING PRECIPITATION highlighted

(7)   e.   *Tōkyō demo Ōsaka demo, **kawaita** tenki   datta […]*
           Tōkyō even Ōsaka even   dry   weather COP:PST
           It was dry weather even in Tokyo and Osaka […]

Primary metaphor: AFFECT IS MOISTURE – NOT SHOWING EMOTIONAL INVOLVE-MENT

(7) f. *Kotae wa jimutekina **kawaita** kōchō datta.*
    answer TOP official   dry   tone COP:PST
    The answer was in an official cool tone.

5.2.3 *Detected types of uses for the Japanese verbal attribute* kawaiteiru *(dry):*

Literal use: FREE FROM LIQUID

(8) ***Kawaiteiru** kami ni   tsukeru to, [...]*
    dry   hair DAT apply   COND
    When you apply it to dry hair, [...]

5.2.4 *Summary and discussion of Czech and Japanese expressions of* DRY

Table 6. Summary of Czech and Japanese counterparts of DRY

	Meaning	CZ *suchý*	JP *kawaita*	JP *kawaiteiru*
Literal use	FREE FROM LIQUID	(6a)	(7a)	(8)
Linear metonymy	RELATED TO LITERAL SOUND QUALITY	(6b)	(7b)	
	RELATED TO LITERAL HAPTIC QUALITY	(6c)	(7c)	
	LOW HUMIDITY	(6d)	(7d)	
	WITH LOW PRECIPITATION	(6e)	(7e)	
	INNER MOISTURE DRIED OUT	(6f)		
Primary metaphor	NOT SHOWING EMOTIONAL INVOLVEMENT	(6g)	(7f)	
Conjunctive metonymy figurative	LACK OF FAT (IN MEAT)	(6h)		
	WITHOUT SAUCE, BUTTER, ETC.	(6i)		
	SCRAWNY	(6j)		
	ABSENCE OF ALCOHOL	(6k)		
Similarity metaphor	WITHOUT ANY COMMENTARY, ETC.	(6l)		
	BORING	(6m)		

The literal meaning DRY is shared by all three forms (6a), (7a), (8). There were not detected any non-literal uses of Japanese *kawaiteiru*. Its shorter counterpart *kawaita* exhibits linear metonymy and primary metaphor uses. These uses are accentuating some aspect of dryness or lack of water: a related sound (7b),

a haptic feeling (7c), a lack of humidity (7d) or precipitation in (7e). The primary metaphor use corresponds to the AFFECT IS MOISTURE metaphor (Grady, 1997, p. 294).

The Czech adjective *suchý* disposes of all uses described for *kawaita* plus several additional ones. There is an additional linear metonymic extensions (6f) expressing a LACK OF INNER MOISTURE (*dried flower, hay*, etc.). There are also several conjunctive metonymy extensions, all of which are figurative. The extension illustrated by the Example (6h) expresses a LACK OF FAT OR JUICE in meat, which is typically connected with the quality of being hard to swallow, and Example (6j) illustrates the same LACK OF FAT but in (usually older) people (*a scrawny man, a slim woman*). The extension illustrated by Example (6i) expresses an ABSENCE OF GRAVY OR GREASE (*dumplings, potatoes, bread,* etc.) The last metonymic use (6k) refers to a LACK OF ALCOHOL OR ITS CONSUMPTION. Moreover, two kinds of similarity metaphor (6l), (6m) were detected. The former implies the meaning WITHOUT ANY COMMENTARY, while the latter implies BORING.

5.3   *Živý* versus *ikita* (LIVE)

Detected types of uses for the Czech adjectival attribute *živý* (live):
Literal use: HAVING LIFE

(9) a. *[…] na sobě   má laciný   oblek   a   v kufru   mu*
          on himself has cheap:ACC suit.ACC and in suitcase:LOC he:DAT
          kdáká   živá   slepice.
          cluck:3SG live:NOM hen:NOM
          […] (he) has a cheap suit on and a live hen clucking in his suitcase.

Linear metonymy: part of entity having life highlighted

(9) b. *Nikdy nezapomněl   na   ten pocit,   když mu   otec   skrz*
          never hasn't_forgotten about that feeling:ACC when he:DAT father through
          živé   maso   protahoval průhledný nylonový   vlasec.
          live:ACC flesh:ACC pulled   clear:ACC nylon:ACC fishing_line.ACC
          He never forgot the feeling when his father pulled the clear nylon fishing line through his living flesh.

Linear metonymy: the quality of BEING MADE OF ENTITIES HAVING LIFE highlighted

(9) c. *Čtyřkilometrový*   ***živý***   *řetěz*   *prošel*
four-kilometers-long:NOM live:NOM chain:NOM walked_through
centrem   města.
center:INS city:GEN
The four-kilometer living chain walked through the city center.

Linear metonymy: area OCCUPIED BY ENTITIES HAVING LIFE highlighted

(9) d. ***Živá***   *krajina*   *je umrtvena, aby se*   *[...] stala*
live:NOM landscape:NOM is slain   in   order to   become
obětištěm   *jatek*   *silničního provozu.*
offering_ground:INS slaughter:GEN road:GEN traffic:GEN
The living landscape is slain [...] to become an offering ground for the
slaughter of road traffic.

Primary metaphor: ACTIVELY PRACTICED, USED

(9) e. *Bez   církve   a   kněží   a   **živé**   tradice   by [...]*
without churh:GEN and clergy:GEN.PL and live:GEN tradition:GEN would
liturgie musely   být neustále   vymýšleny znovu [...]
liturgy have_to be contuinally invented   again
Without the church, the clergy and the living traditions, [...] the liturgy
would have to be continually rediscovered [...]

Primary metaphor: ABOUNDING WITH LIFE

(9) f. *Agátka je hodně **živé**   dítě.*
Agatha is very   live:NOM child:NOM
Agatha is a very lively child.

Conjunctive metonymy: AUTHENTIC

(9) g. *Jeho povídky   ze   života venkovských   lidí   čerpaly*
his   stories:NOM from life:GEN country:GEN:PL people:GEN drew
*látku   i   **živý**   jazyk,   odpozorovaný z*
content:ACC as_well_as live:ACC language:ACC observed   from
mluvené   řeči   řemeslníků, [...]
spoken:GEN language:GEN craftsmen:GEN
His short stories inspired by life of country folk drew on living language
observed in speech of craftsmen, [...]

Conjunctive metonymy: VIVID

(9) h. *[…] náhle   se   mu vybavila dávná, ale stále až   bolestně*
         suddenly REFL him recall:PST old:NOM but still even painfully
         živá   vzpomínka   z   dětských   let.
         live:NOM memory:NOM from childhood:GEN.PL years:GEN
         […] all of a sudden a very old but still painfully vivid childhood memory
         came to his mind.

Similarity metaphor: BRIGHT

(9) i. ***Živé***   *barvy   a   ostré   kontrasty, […]*
         live:NOM:PL colors:NOM and sharp:NOM:PL contrasts:NOM
         Bright colors and sharp contrasts, […]

Similarity metaphor: JOYFUL

(9) j. *Ze schodů   k nim   dolehl **živý**   smích.*
         from stairs:GEN to they:DAT reached live:NOM smile.NOM
         Lively laughter was heard from the staircase.

Similarity metaphor: being under (electric) power

(9) k. *[…] dochází k obnažení   izolace   a   vylezou **živé***
         leads   to stripping:DAT insulation:GEN and pop out live:NOM:PL
         dráty.
         wires:NOM
         […] the insulation is stripped down and the live wires pop out.

5.3.1 *Detected types of uses for the Japanese verbal attribute* ikita *(live):*

Literal use: HAVING LIFE

(10) a. *[…] **ikita** hito   no kao da to   iu   koto wa hakkiri wakatta.*
          live   person GEN face COP QUOT COP AUX TOP clearly understood
          […] I knew right away that it is the face of a living person.

Linear metonymy: part of entity having life highlighted

(10) b. *[…] **ikita** niku o   heikide   hikisakaseru.*
          live   flesh ACC indifferently tear_up:CAUS:NONPAST
          […] without blinking have **live** meat torn apart.

Primary metaphor: ACTIVELY PRACTICED, USED

(10)   c.   *[...] "**ikita** chishiki"   wa, hibi   no   seikatsu – "genjitsu" no   naka   ni*
               live   knowledge TOP daily GEN life   reality   GEN inside DAT
           arimasu.
           exist:POL:NONPST
           [...] natural knowledge is in the reality of everyday life.

Conjunctive metonymy: AUTHENTIC

(10)   d.   *[...] kyōkasho   de   manabenai   **ikita** eigohyōgen [...]*
               textbook(s) INS learn:POTEN:NEG:NONPST live   English-expression(s)
           [...] authentic expressions which you cannot learn from a textbook [...]

5.4   Detected types of uses for the Japanese verbal attribute *ikiteiru* (live):

Literal use: HAVING LIFE

(11)   a.   ***Ikiteiru** kaseki to   iwareru   kabutogani*
           live   fossil   QUOT call:PASS:NONPST horseshoecrab
           A horseshoe crab is also called a living fossil.

Linear metonymy: LIVING A MEANINGFUL LIFE

(11)   b.   *Kore wa   masani **ikiteiru** ningen dearu.*
           this   TOP truly   live   human COP:NONPST
           This is truly a living human.

5.4.1   *Summary and discussion of Czech and Japanese expressions of* LIVE

Table 7. Summary of Czech and Japanese counterparts of LIVE

	Meaning	CZ *živý*	JP *ikita*	JP *ikiteiru*
Literal use	HAVING LIFE	(9a)	(10a)	(11a)
Linear metonymy	A PART OF AN ENTITY WHICH HAS LIFE	(9b)	(10b)	
	MADE OF ENTITIES WHICH HAVE LIFE	(9c)		
	PLACE WITH LIVING ORGANISMS	(9d)		
	LIVING A MEANINGFUL LIFE			(11b)
Primary metaphor	ACTIVELY PRACTICED, USED, ETC.	(9e)	(10c)	
	ABOUNDING WITH LIFE	(9f)		
Conjunctive metonymy	AUTHENTIC	(9g)	(10d)	
	VIVID	(9h)		
Similarity metaphor	BRIGHT	(9i)		
	JOYFUL	(9j)		
	HAVING (ELECTRIC, ETC.) POWER	(9k)		

The literal meaning LIVE is shared by all three forms (9a), (10a), (11a). The Japanese form *ikiteiru* disposes of only one non-literal use: LIVING AN AUTHENTIC, FRUITFUL LIFE (11b). Its shorter counterpart *ikita* disposes of literal, linear metonymy, primary metaphor and conjunctive metonymy uses. The linear metaphor usage includes a common extension from the entire unit to some of its parts (10b). The primary metaphor use (10c) corresponds to the ACTIVITY IS LIFE metaphor (Grady, 1997, p. 289). The figurative conjunctive metonymy extension expresses the trait AUTHENTIC (10d).

The Czech adjective *živý* comprises all the above mentioned uses and adds a number of various extensions. For one, there are linear metonymy extensions (9c), (9d), where the former use expresses the quality of being MADE OF LIVING ORGAN-ISMS (a chain of living people creating a living fence etc.), and the latter use is a common extension from the quality of an entity (animals, plants, etc.) to an area where such an entity resides (landscape). The primary metaphor ACTIVITY IS LIFE has two concrete renditions in the Czech language, one identical with *ikita* (9e), (10c) and the other expressing an ABUNDANCE OF LIFE or ENERGETIC CHARACTER (usually used about children) (9f). There is also another conjunctive metonymy VIVID or PRECISE (about memories, descriptions, etc.) in (9j). There are also similarity metaphor uses such as BRIGHT about colors (9i) and JOYFUL (about sounds, or appearances) (9j) and HAVING (ELECTRIC) POWER (9k).

5.5   *Mrtvý* versus *shinda* (DEAD)

5.5.1   *Detected types of uses for the Czech adjectival attribute* mrtvý *(dead):*

Literal use: DEPRIVED OF LIFE

(12) a. ***Mrtvé   ryby   poslali   na rozbor […]***
dead:ACC.PL fish:ACC.PL sent:3PL.PST to analysis.ACC
(They) sent the dead fish for analysis […]

Linear metonymy: the quality of RESEMBLING BEING DEPRIVED OF LIFE highlighted

(12) b. *Chvíli   zůstal   stát,   vyhublý,   s   **mrtvým** pohledem.*
while:ACC stay:3SG to_stand skinny:NOM with dead:INS look:INS
He stopped for a while, skinny with a dead look.

Linear metonymy: area deprived of living organisms highlighted

(12) c. *Objevují se náhle   a   překvapivě v **mrtvé**   krajině.*
appear:3PL suddenly and surprisingly in dead:LOC landscape:LOC
They appear suddenly and unexpectedly in the dead landscape.

Primary metaphor: LACKING (PHYSICAL) ACTIVITY

(12) d. **Mrtvé**   *město*   *ožilo*   *vzácnou*   *radostí, [...]*
          dead:NOM town:NOM become_alive:3SG.PST precious:INS joy:INS
          The dead town came alive with precious joy, [...]

Primary metaphor: NOT ACTIVELY USED, PRACTICED, etc.

(12) e. *[...] rekonstruují*   *dva tisíce*   *let*   **mrtvé**   *náboženství [...]*
          are_reconstructing:3PL two thousand years dead:ACC religion:ACC
          [...they] are reconstructing a two thousand years dead religion [...]

Similarity metaphor: NOT PRODUCTIVE

(12) f. *Přesto*   *nejde o*   **mrtvý**   *čas,*   *ač*   *tak*   *na žadatele*
          however isn't   about dead:ACC time:ACC even though to applicant:ACC
          může působit.
          may   give_impression
          However, it is not wasted time, even though it may make that impression
          on the applicant.

Similarity metaphor: BROKEN

(12) g. *[...] kývla*   *ke stolu,*   *kde*   *lež* **mrtvý**   *mechanismus [...]*
              beckoned:3SG to table:DAT where lay dead:NOM mechanism:NOM
          [...] (she) beckoned to the table, where the dead mechanism lay [...]

5.5.2 *Detected types of uses for the Japanese verbal attribute* shinda *(dead):*

Literal use: DEPRIVED OF LIFE

(13) a. *[...]* **shinda** *neko no yume o   mite   ita   no da.*
              dead   cat   GEN dream ACC see:GER be:PST AUX COP
          [...] I had a dream about a dead cat.

Linear metonymy: area deprived of living organisms highlighted

(13) b. *Dōse*   **shinda** *kawa ya umi da kara, yogoshitatte   shō*
          anyway dead   river and sea COP so   make_dirty:CONC measure(s)
          ga   nai [...]
          NOM exist:NONPST
          The river and sea are dead anyway, who cares if we pollute them some
          more [...]

Primary metaphor: LACKING (PHYSICAL) ACTIVITY

(13) c. *[…] saichiku   sareta   no da ga, **shinda** toshi to   iu*
            rebuilding do:PASS:PST AUX COP but dead   city   QUOT COP
        inshō   ga   tsuyoi.
        impression NOM strong:NONPST
        […] it was rebuilt, but the impression of a dead city is still strong.

Primary metaphor: NOT ACTIVELY USED, PRACTICED, etc.

(13) d. *Watashitachi no   sedai   ni-totte no   hakozen wa, mohaya kakko no*
            1PL   GEN generation DAT   GEN hakozen TOP already past   GEN
        ***shinda** shūzoku deatta.*
        dead   custom COP:PST
        For our generation, [using a] *hakozen* was already a dead custom.

Primary metaphor: NOT PRODUCTIVE

(13) e. *[…] suriru ni   toboshii, **shinda** hanashi ni   natte   shimaigachi*
            thrill DAT scarce,   dead   story   DAT become:GEN PERF:AUX
        desu […]
        POLCOP:NONPAST
        […] it tends to turn into a dead story lacking in thrill.

5.5.3 *Detected types of uses for the Japanese verbal attribute* shindeiru *(dead):*

Literal use: DEPRIVED OF LIFE

(14) a. *Kare wa **shindeiru** wakamono o   minagara, chiisaku tsubuyaita.*
        he   TOP dead   youth   ACC look:SIMUL little   sigh:PST
        When he saw the dead young people, he gave a quiet sigh.

Linear metonymy: quality RESEMBLING BEING DEPRIVED OF LIFE highlighted

(14) b. *Ikinagara ni   shite   **shindeiru** kankaku ga   boku ni-totte wa   anshin*
        live:SIMUL DAT do:GER dead   feeling NOM 1SG DAT   TOP relax
        dekiru.
        do.POTEN:NONPST
        It is a relief for me to be dead [=numb] while alive.

5.5.4   *Summary and discussion of Czech and Japanese expressions of DEAD*

Table 8. Summary of Czech and Japanese expressions of DEAD

	Meaning	CZ *mrtvý*	JP *shinda*	JP *shindeiru*
Literal use	DEPRIVED OF LIFE	(12a)	(13a)	(14a)
Linear metonymy	RESEMBLING THE QUALITY EXPRESSED BY THE LITERARY USE	(12b)		(14b)
	AREA DEPRIVED OF LIVING ORGANISMS	(12c)	(13b)	
Primary metaphor	NO (PHYSICAL) ACTIVITY	(12d)	(13c)	
	NOT ACTIVELY USED, PRACTICED, ETC.	(12e)	(13d)	
Similarity metaphor	NOT PRODUCTIVE	(12f)	(13e)	
	BROKEN	(12g)		

The literal meaning of DEAD is shared by all three forms (12a), (13a), (14a). The Japanese form *shindeiru* disposes of only one non-literal use – RESEMBLING THE STATE OF BEING DEAD. Its shorter counterpart *shinda* disposes of linear metonymy, primary metaphor and similarity metaphor uses. The linear metaphor usage includes a common extension from the quality of a target to an area where such targets reside (13b). The primary metaphor use corresponds to the ACTIVITY IS LIFE metaphor (Grady, 1997, p. 289). The figurative similarity metaphor expresses the characteristic of NOT BEING PRODUCTIVE, of something NOT BEARING THE RESULTS as it should (13e).

The Czech adjective *mrtvý* comprises all the above mentioned uses as well as an additional similarity metaphorical extension with BROKEN.

5.6   *Tlustý* versus *futotta, futotteiru* and *futoi* (THICK)

The last analyzed item differs from the previous four in that the Czech adjective *tlustý* corresponds to two separate lexical items in Japanese – verbal forms *futotta* and *futotteiru* and adjective *futoi*. Having two counterparts obviously results in smaller overlap of the individual lexical items in the two languages. However, the comparison was included since the two Japanese items differ only in the part of speech membership, which gives us an opportunity to consider the effect of part of speech schema on meaning extension directly.

5.6.1 *Detected types of uses for the Czech adjectival attribute* tlustý *(thick):*

Literal use: WIDE

(15) a. *Takovou situaci   znázorňuje na obr. 4.18 **tlustá**   čára*
          such:ACC situation:ACC depicts   on pict. 4.18 fat:NOM line:NOM
          pokračující tenčí   částí.
          continuing thiner:INS part:INS
          Such a situation is depicted in picture 4.18 with the thick line changing
          into a thinner one.

Literal use: THICK

(15) b. *Z   osikových   houštin   tu   vystupuje   **tlustá** roura [...]*
          from aspen:GEN:PL groves:GEN here is_coming_out fat:NOM pipe:NOM
          There is a thick pipe coming out of the aspen grove here [...]

Literal use: DEEP

(15) c. *[...] celá   oblast je pokryta desítky metrů **tlustým** ledovcem o rozloze*
                whole area   is covered tens   meters fat:INS glacier:INS of area:LOC
          asi   160 km².
          about 160 km²
          [...] the whole area of about 160km² is covered by a glacier (which is) tens
          of meters thick.

Linear metonymy: the quality of BEING FATTY, STOUT about living organisms
highlighted

(15) d. *[...] když se otočil, před ním stála **tlustá** jeptiška s   laskavou*
                when turned:3SG before him stood fat:NOM nun:NOM with kind:INS
          tváří [...]
          face:INS
          [...] when he turned, there was a fat nun with a kind face [...] standing in
          front of him, [...]

Linear metonymy: the quality of BEING WARM highlighted

(15) e. *[...] nosily   **tlusté**   nátělníky   a   vlněné   spodky.*
                wear:3PL.PST fat:ACC:PL undershirts:ACC and wool:ACC:PL briefs:ACC
          [...] they wore warm undershirts and wool briefs.

Linear metonymy: LONG

(15) f. *Popisem   dekontaminace   všech   takových*
description:INS decontamination:GEN all:GEN:PL such:GEN:PL
*látek   by   vznikla **tlustá**   kniha [...]*
substances:GEN would arise   fat:NOM book:NOM
Describing the decontamination of all kinds of such substances would
make a thick book [...]

Linear metonymy: the quality of BEING STRONG highlighted

(15) g. *Starý   pán   to   ovšem   neviděl,   měl   moc **tlusté***
old:NOM man.NOM it.ACC however didn't_see have:3SG.PST too fat:ACC
brýle   a   moc špatné   oči [...]
glasses:ACC and too bad:ACC.PL eyes:ACC
The old man certainly could not see it; he had too thick of glasses and too
poor of eyes [...]

Linear metonymy: area occupied with many fat people highlighted

(15) h. *Clarksdale, město   v jednom z   **nejtlustších**   okresů   v*
Clarksdale town:NOM in one:LOC from fattest:GEN.PL counties:GEN in
***nejtlustším** státě [...]*
fattest:LOC state:LOC
Clarksdale – a town in one of the fattest counties of one of the fattest states
[...]

Conjunctive metonymy: GREASY, FATTY

(15) i. ***Tlustému** vepřovému a   tučné   kachně   se   vyhněte úplně.*
fat:DAT.SG pork:DAT.SG and fatty:DAT duck:DAT REFL avoid   altogether
Avoid fatty pork and greasy duck altogether.

Conjunctive metonymy: RICH, EXPENSIVE

(15) j. *Přestaň, osvobozená Římanko, nebo budeš   dlužit   stolu   **tlustý***
stop   liberated Roman   or   will:2SG owe:INF table:DAT fat:ACC
poplatek.
fee:ACC
Stop it, you liberated Roman, or otherwise you will pay a fat fee to every-
one at the table.

5.6.2 *Detected types of uses for the Japanese verbal attribute* futotta *(fat):*

Literal meaning: FATTY, STOUT about living organisms

(16) *[...] **futotta** kamisan ni   tetsudawasete [...]*
fat   wife   DAT help:CAUS:GER
[...] (his) fat wife forced him to help [...]

5.6.3 *Detected types of uses for the Japanese verbal attribute* futotteiru *(fat):*

Literal meaning: FATTY, STOUT about living organisms

(17) *Futotteiru rikishi   no   koto*
      fat   sumo_wrestler GEN AUX
      Regarding the huge sumo wrestler

5.6.4 *Detected types of uses for the Japanese verbal attribute adjectival attribute* futoi *(thick):*

Literal use: WIDE

(18) a. *[…] futoi moji   de* ⌜*toki wa kane   nari!*⌟ *to   kakarete*
          fat   letter(s) INS time TOP money COP   QUOT write:PASS:GER
      ita.
      be:PST
      […] the bold lettering read "time is money".

Literal use: THICK

(18) b. *[…] futoi eda   ga   kiriotosareru […]*
          fat   branch(es) NOM cut:off:PASS:NONPST
      […] the thick branches are cut off […]

Linear metonymy: the quality of being FATTY, STOUT (about living organisms) highlighted

(18) c. *[…] futoi hito   no 2 kiro to yaseteru hito   no 2 kiro de   wa*
          fat   person GEN 2 kilos and slim   person GEN 2 kilos COP.GEN TOP
      naiyō   ga   chigaimasu […].
      content NOM differ:POL:NONPST
      […] 2 kilos in a fat person and 2 kilos in a slim person are a different game.

Conjunctive metonymy: DEEP

(18) d. *[…]* ⌜*memo no yōi   o   futoi koe ga   meijita.*
          note(s) GEN preparation ACC fat   voice NOM order:PST
      […] "Prepare notes", commanded the deep voice.

Conjunctive metonymy: SOLID, STRONG

(18) e. *[…] sore kurai   no **futoi** shinkei*
           that approximate_amount GEN fat   nerve(s)
           mottenaito,   jānarisuto wa   yatte
           have:GER:NEG:NONPST:COND, journalist TOP do:GER
           ikenai […]
           go_on:POTEN:NEG:NONPST
           […] you could not do journalism if you do not have such nerve […]

Similarity metaphor: LOW, BRAZEN

(18) f. **Futoi** *hito   wa, osoraku manzoku   suru made tanomu janai*
           fat   person TOP probably be satisfied:NONPST till   ask   or   not
           deshōka?
           POLCOP:VOL:SFP
           Brazen people would just keep asking till satisfied, wouldn't they?

5.6.5 *Summary and discussion of Czech and Japanese expressions of THICK*

Table 9. Summary of Czech and Japanese counterparts of THICK

	Meaning	CZ *tlustý*	JP *futotta*	JP *futotteiru*	JP *futoi*
Literal use	WIDE (2 DIMENSIONAL OBJECTS)	(15a)			(18a)
	THICK (3 DIMENSIONAL TYPICALLY ROUND OBJECTS)	(15b)			(18b)
	DEEP (3 DIMENSIONAL TYPICALLY SQUARE OBJECTS)	(15c)			
Linear metonymy	FAT, STOUT (LIVING ORGANISMS)	(15d)	(16)*	(17)*	(18c)
	WARM (ABOUT CLOTHES)	(15e)			
	LONG (ABOUT BOOKS ETC.)	(15f)			
	STRONG (ABOUT GLASSES ETC.)	(15g)			
	AREA WHERE MANY FAT PEOPLE LIVE	(15h)			
Conjunctive metonymy figurative	FATTY, GREASY	(15i)			
	RICH (ABOUT EVENTS, THINGS, ETC.)	(15j)			
	SOLID, STRONG (NERVES)				(18e)
	LOW (ABOUT VOICE)				(18d)
Similarity metaphor	BRAZEN				(18f)

In this case, the literal meaning is a rather complicated issue. The adjective *fat* in both languages has the dimensional meaning THICK as its literal meaning. However, the verbal adjectives begin with the literal meaning STOUT/FAT (marked by * in Table 7). No non-literal uses for either of the verbal attributes were detected. The Japanese adjective *futoi* is narrower in its literal meaning, since it does not include the dimension of depth (the adjective *atsui* is used instead) and is also rather marginal for expressing width (the adjective *habahiroi* is preferred). The linear metonymy extension expressing stoutness is also marginal (instead, verbal forms are prototypical here). There are conjunctive metonymies which expresses the strength or stability of someone's nerves and lowness as the quality of someone's voice. The similarity metaphor detected expresses the meanings of and BRAZEN as a quality trait.

The fact that the literal meaning of the Czech adjective *tlustý* is wider than any of the three Japanese counterparts is reflected by a higher number of linear metaphors drawing on the literal uses. There are meaning extensions such as STRONG for glasses or lenses, WARM for pieces of clothing, FATTY for meats and RICH for events or things like checks or purses. There were no metaphorical uses detected in the corpus analyzed.

5.7   Summary of results and general discussion

The overall summary of the results is in Table 10, where the numbers refer to the examples above.

In general, we can conclude that in every single instance, Czech adjectives showed quantitatively greater and more varied meaning extensions. The highest degree of correspondence of meaning extension in the two languages is in the primary metaphor category where meaning extensions match almost perfectly. There is a very good correspondence in the category of linear metaphors in some cases (*wet, dry, dead*) but other instances show a high inconsistency between the extensions in the two languages (*thick, live*). The most pronounced and frequent differences were detected in the conjunctive metonymy and similarity metaphor uses, which both qualify as figurative language use. Japanese extensions are extremely rare in these categories. There are only two figurative extensions among the ten verbal forms examined. On the other hand, the only Japanese adjective analyzed (*futoi* – THICK) displays three figurative extensions.

Returning to the presented research questions, I would like to focus first on the role of the verbal schema in meaning extension. Since there is no definitive proof linking the verbal meaning extension with non-figurative meanings (even though the comparison with adjectives in case of *thick* gives us a telling hint), let us focus on differences in meaning extension within the category of linear

Table 10. Summary of results

		linear metonymy						primary metaphor		conjunctive metonymy			similarity metaphor		
wet	mokrý (CZ)	3b	3c	3d	3e	3f	3g	3h	3i	3j	3k	3l			
	nureta (JP)	4b	4c	4d				4e	4f						
	nureteiru (JP)														
dry	suchý (CZ)	6b	6c	6d	6e	6f		6g		6h	6i	6j	6k	6l	6m
	kawaita (JP)	7b	7c	7d	7e			7f							
	kawaiteiru (JP)														
live	živý (CZ)	9b	9c	9d				9e	9f	9g	9h		9i	9j	9k
	ikita (JP)	10b						10c		10d					
	ikiteiru (JP)				11b										
dead	mrtvý (CZ)	12b	12c					12d	12e				12f	12g	
	shinda (JP)		13b					13c	13d				13f		
	shindeiru (JP)	14b													
thick	tlustý (CZ)	15d	15e	15f	15g	15h				15i	15j				
	futotta (JP)	16													
	futotteiru (JP)	17													
	futoi (JP)	18c											18e	18d	18f

metonymy. The first instance WET shows some correspondence between Czech and Japanese, but there are three extensions which are not shared by the Japanese

modification. The first one is the use stressing the sensation of wetness. It is typically used in Czech with nouns referring to liquids, such as *water* or *tears*. This extension is not only lacking in Japanese but Japanese speakers are confused when they hear the extension even in English (personal experience). The form *nureta* or *nureteiru* implies, due to its verbal schema, the aspect of change. However, water cannot change from dry to wet, since water is wet at all times. That is why such an extension is unthinkable in Japanese. The following EXTENSION STILL SMEARING OR WET is used about paint, cement, plaster, etc. The logic is, in essence, very similar to the case of a wet sensation. The combination of the verb *nureta* with the noun *cement* is possible, but it would be interpreted that the surface of the cement is getting wet by rain or other liquid. These verbal forms suppose a dynamic repeatable change and thus preclude this meaning extension. This can also be said about the following extension of HIGH PRECIPITATION. The most frequent type of noun following this extension in Czech is a noun with a temporal meaning (*year, week, day,* etc.). These are obviously not material entities, but rather temporal ones. The verbal schema has as its prototypical trajector a thing, which is why a combination with such nouns is difficult. The extensions in the instance of *thick* are difficult to compare because most of Czech linear metaphors build on one of the literal meanings, which is not shared by the Japanese counterpart (*deep* or *thick* about box-like objects). The comparison of Czech and Japanese versions of *live* are therefore more noteworthy. The verb *ikiru* normally takes a subject, which is prototypically human, animal, or plant. It is hardly possible to combine this prototypical trajector with abstract "non-living" things such as *chain* or a given *place*. The remaining instances of both *dry* and *dead* show a very high degrees of correspondence between the Czech and Japanese linear metonymies. The only case where there is no correspondence in the case of *dry* is the meaning extension of *dried* about plants, where the inner moisture is dried out. Japanese speakers instead prefer the specialized separate verb (*kareru*), which also strongly implies lack of life. The results of this very limited analysis of linear metonymy across five target attributes point to the strong role of the verbal schema and a tendency to keep the schema intact, even in the position which is prototypically occupied by adjectives disposing of a very different schema.

The proposed answer for the second research question has been discussed in the remarks above. We have shown that there is a strong tendency for the meaning extension of Japanese verbal attributes to be limited to the non-figurative level, which is not the case for Czech adjectives. The difference is visible in the case of the past perfect form (TA form) and very clear in the case of the present imperfect form (TE-IRU form). It follows from the above that the meaning of Czech adjectives is much more schematic and abstract, because unlike its Japanese verbal counterpart, it must accommodate all the various uses described.

Therefore, despite the very limited scope of the presented research, we have been able to discern certain tendencies between Czech adjectives and Japanese verbs in adnominal modification. The difference in the breadth of meaning extensions seems to be linked to the difference in the part of speech membership and the related schema. However, it remains to be confirmed whether or not this is only a broader tendency of the Japanese vocabulary in general. Such research would be interesting not only for confirmation of the presented results but also for possible links to linguistic relativity research.

References

Croft, W. (2006). The role of domains in the interpretation of metaphors and metonymies. In D. Geeraerts (Ed.), *Cognitive Linguistics: Basic readings* (pp. 269–302). Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110199901.269

Dirven, R. (2002). Metonymy and metaphor: Different mental strategies of conceptualization. In R. Dirven & R. Pörings (Eds.), *Metaphor and metonymy in comparison and contrast* (pp. 75–111). Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110219197.75

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.

Grady, J. (1997). Foundations of meaning: Primary metaphors and primary scenes. UC Berkeley, Dissertation.

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development*, 79, 979–1000. https://doi.org/10.1111/j.1467-8624.2008.01171.x

Kageyama, T., & Kishimoto, H. (Eds.). (2016). *Handbook of Japanese lexicon and word formation.* Berlin: Mouton de Gruyter. https://doi.org/10.1515/9781614512097

Kageyama, T., & Saito, M. (2016). Vocabulary strata and word formation processes. In T. Kageyama & H. Kishimoto (Eds.), *Handbook of Japanese lexicon and word formation* (pp. 1–50). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9781614512097-005

Kanasugi, P. (2016). Hyōzōsei no Kanten kara mita Chekogo to Nihongo to no Hikaku – Chekogo niokeru "Keiyōshi + Meishi" no Kōzō o megutte. In *Hikaku Nihongaku kjóiku kenkjú sentá kenkjú hókoku* (pp. 142–147). Tokyo: Ochanomizu University.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar.* Vol. 1: *Theoretical prerequisites.* Stanford: Stanford University Press.

Langacker, R. W. (1991). *Foundations of Cognitive Grammar.* Vol. 2: *Descriptive application.* Stanford: Stanford University Press.

Langacker, R. W. (2000). *Grammar and conceptualization.* Berlin: Mouton de Gruyter.

Langacker, R. W. (2008). *Cognitive Grammar: An introduction.* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195331967.001.0001

Langacker, R. W. (2015). On grammatical categories. *Journal of Cognitive Linguistics*, 1, 44–80.

Maass, A., Karasawa, M., Politi, F., & Suga, S. (2006). Do verbs and adjectives play different roles in different cultures?_ A cross-linguistic analysis of person representation. *Journal of Personality and Social Psychology*, 90(5), 734–750. https://doi.org/10.1037/0022-3514.90.5.734

Nihongo Kijutsu Bunpō Kenkjūkai. (2013). *Gendai Nihongo Bunpō*. Vol. 1. Tokyo: Kurosio Publishers.

Okimori, T., Kimura, Y., Tanaka, M., & Chin, R. (2011). *Zukai Nihon no Goi*. Tokyo: Sanseidō.

Radden, G. (2002). How metonymic are metaphors? In R. Dirven & R. Pörings (Eds.), *Metaphor and Metonymy in Comparison and Contrast* (pp. 407–434). Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110219197.407

Sullivan, K. (2013). *Frames and constructions in metaphoric language.* Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/cal.14

Suzuki, S. (1982). *Nihongo bunpō, keitairon*. Tokyo: Mugishobō.

Štícha, F. et al. (2013). *Akademická gramatika spisovné češtiny*. Prague: Academia.

Uehara, S. (1998). *Syntactic categories in Japanese: A cognitive and typological introduction.* Tokyo: Kurosio Publishers.

The near-synonymy of classifiers and construal operation

A corpus-based study of 棵 *kē* and 株 *zhū* in Chinese

Aneta Dosedlová and Wei-lun Lu
Masaryk University

This paper investigates the near-synonymy of classifiers, using Chinese *kē* and *zhū* as illustration. We find that in [QUANTIFIER] – [CLASSIFIER] – [NOUN], the two classifiers have overlapping semantic prototypes due to their similar behavioral profiles. However, despite a shared functional core, the two classifiers diverge in terms of which part of PLANT to profile. In particular, *zhū* highlights parts of PLANT that are small and vulnerable, such as flowers and seedlings. In addition, SMALL is another important conceptual characteristic exclusively associated with *zhū*, which gives it a distinctive set of peripheral members to include in that particular linguistic category, including MOLD, BACTERIUM, and even BIOLOGICAL AND CHEMICAL SUBSTANCE. Another important difference is the quantifier that precedes, where *kē* tends to occur with lower numbers (typically under 10), while *zhū* with higher numbers (typically over 1,000). Accordingly, we conclude that [QUANTIFIER] – [ZHU] – [NOUN] tends to invoke a higher-resolution construal.

Keywords: classifier, construal, quantifier, synonymy, Chinese

1.   Classifiers in Chinese

Classifiers are a subtype of noun categorization system that is of special grammatical significance in human language, found generally in languages of East Asia, Southeast Asia, Australia and the Americas (Aikhenvald, 2000). Classifiers are morphemes which characterize nouns and mostly originate from common nouns with concrete meanings. Moreover, in some languages, many or most classifiers take the same form as ordinary nouns (Goddard, 2011).

Classifiers have been reported to reflect cultural values (Allan, 1977) and to operate based on cognitive principles (Lakoff, 1987), with a typical example extensively addressed being Dyirbal.

Chinese falls under the category of numeral classifier languages, and has been reported to be a typical one. In Chinese, classifiers occur in the following composite constructional schema [QUANTIFIER] – [CLASSIFIER] – [NOUN] (Jiang, 2017). Within the category of numeral classifiers, different authors (Myers, 2000; Tai, 1994) introduce division into two subtypes: *sortal (specific) classifiers* which typically form a huge open class with many members and *general classifiers* such as 个 *ge* in Chinese. General classifiers can usually take any noun.

On the other hand, sortal classifiers are inherently partitive, emphasizing individual instances (Erbaugh, 2006). They refer to a particular category of nouns based on their inherent semantics, such as 棵 *kē*, which has a meaning connected to plants and trees. As we have mentioned, classifiers originate from ordinary nouns with concrete meanings, which means that these specific classifiers are sensitive to the semantic features of the noun categorized. Consider (1) below and the anomaly of (1a).

(1)   一   棵 花
       yī   kē huā
       one CL flower
       'one flower.'

(1)   a. ?一   棵 花瓶
          yī   kē huāpíng
          one CL vase
          'one vase'   (constructed)

The contrast can be explained by the semantic characterization of *kē* being applicable only to plants and trees, which makes (1) perfectly acceptable, as the noun modified 花 *huā* 'flower' falls under the category of plants and trees. However, (1a) is unacceptable because the noun 花瓶 *huāpíng* 'vase' refers to a type of a container rather than to a kind of plant.

There have been various studies on classifiers in Chinese and their cognitive functions, but no research that we are aware of discusses the near-synonymy of different classifiers within one single language, which constitutes a gap in literature that needs to be filled. The near-synonymy of Chinese classifiers can be evidenced by the interchangeability of 棵 *kē* and 株 *zhū* in (2)/(2a) and (3)/(3a).

(2) 爸爸　买了　　两棵　　　　巨大的　圣诞树
    bàba mǎi-le liǎng-kē   jùdà-de shèngdànshù
    father buy-PFV two-CL: plant big-LK   Christmas tree
    'Father bought two huge Christmas trees.'

(2a) 爸爸　买了　　两株　　　　巨大的　圣诞树
    bàba mǎi-le liang-zhū   jùdà-de shèngdànshù
    father buy-PFV two-CL: plant big-LK   Christmas tree
    'Father bought two huge Christmas trees.'　　　　　　(constructed from (2))

(3) 森林　有　　大概　一万多株　　　　　　　树
    sēnlín yǒu dàgài yíwàn-duō-zhū   shù
    forest have about ten thousand-more-CL:plant tree
    'The forest has about more than ten thousand trees.'

(3a) 森林　有　　大概　一万多棵　　　　　　　树
    sēnlín yǒu dàgài yíwàn-duō-kē   shù
    forest have about ten thousand-more-CL:plant tree
    'The forest has about more than ten thousand trees.'　　(constructed from (3))

As is clear from (2)/(2a) and (3)/(3a), the replacement of *kē* with *zhū*, and vice versa, does not dramatically alter the meaning of the sentences. However, the nuance created by the replacement of the classifiers is not very clear – for instance, according to the *Oxford Chinese dictionary* (2010), *kē* and *zhū* both refer to inanimate nouns, specifically to the category of trees and plants, but no further semantic information is provided therein. Nor does any thesaurus of Chinese include the two classifiers as synonyms, which has created a headache not only for linguists but also learners of the language. Therefore, the semantic similarity of classifier constructions constitutes the main issue that we will pick up on in the present paper.

As has been mentioned, the usage of classifiers represents the nature of categorization in human cognition. There are many classifiers which can be used with different nouns either to focus on specific characteristics of the noun, or the noun has more characteristics which can relate to more than one classifier. This makes the study of the classifier system even more valuable, as it provides us with a window to how a classifier language enables (or makes) its speakers to perceive and to categorize the world (Lakoff, 1987). However, it has also been noted that the choice of a classifier is usually determined by a number of factors, such as the language proficiency or whether the speaker is familiar with the object – sortal classifiers are used most often for the first mention of new objects, and the second mention often degrades to the general classifier (Erbaugh, 2006).

Chinese has many specific classifiers which focus on more than one characteristic, so the categories of nouns they refer to may cross-cut each other, which gives rise to overlapping use, or the near-synonymy, of classifier constructions within a language. There are also classifiers which focus on similar characteristics of nouns, which allows for an open choice between alternative construals of this particular noun. This is the hypothesis that we will work with in the present paper.

2.   Near-synonymy in cognitive linguistics

In cognitive linguistics, a lexical *construction* (Goldberg, 1995), or a word in the traditional linguistic sense, may bear a similar meaning with another construction. In such view, the similarity or the difference between the lexical meanings of synonymous constructions is reflected in the *behavioral profile* or *constructional profile* of the target constructions under investigation (Divjak & Gries, 2009; Glynn, 2014; Hanks, 1996; Janda & Soloyev, 2009). Following that, absolute synonymy does not really exist, because absolute synonyms would be interchangeable in *any* context without change of the meaning. Accordingly, two lexical constructions are near-synonymous when they are interchangeable only in certain contexts (with nuances of their lexical meanings) and are sometimes stylistic variants (Hirst, 1995; Hirst & Edmonds, 2002).

In cognitive linguistics, near-synonymy research has covered various different parts of speech, most of which focus on verbs (Divjak & Gries, 2009; Su & Lu, 2009), few on nouns (Janda & Soloyev, 2009; Liu, 2013) and on adjectives (Glynn, 2014). Divjak and Gries (2009), for instance, analyze five verbs in Russian and English meaning 'begin' using the behavioral profile approach. Based on data extracted from the ICE-GB and the Uppsala Corpus, they find that the difference between the semantic structure of *begin* and *start* is lexical, that between *stat'* and *načinat'/načat'* is aspectual, and that between *načinat'sja/načat'sja* and *načinat'/načat'* is about argument structure. Su and Lu (2009), for another instance, investigate the synonymous pair *worry* and *bother* in English based on search results from the British National Corpus and discusses the different construals invoked by the verbs in terms of whether the effect is also included in the encoding of the event along with the emotion of the experiencer. Janda and Soloyev (2009) study Russian nouns that denote HAPPINESS and SADNESS using data extracted from Russian National Corpus and Biblioteka Maksima Moškova. They find that although the nouns denoting sadness share a similar range of constructions, their distributions within the range are rather different. In particular, *pečal'* and *toska* are mathematically closer and so are *xandra* and *melanxolija*, with

grust' and *unynie* being the outliers. As for the nouns of happiness, *naslaždenie, radost'*, and *udovol'stvie* pattern similarly, with the first two being the closest synonyms. *Vostorg* is a behavioral outlier and *likovanie* falls between *vostorg* and the others. However, in the sea of literature, although near-synonymy research has covered various different parts of speech in cognitive linguistics, no study that we are aware of has investigated synonymous classifiers within a language.

Construal is another important issue in cognitive linguistics, which maintains that the meaning of a linguistic expression lies not only in the conceptual content invoked but also in the way the content is *construed* (Langacker, 2008, p. 55). For instance, Liu's (2013) triangulate study on the two synonymous sets *authority/power/right* and *duty/obligation/responsibility* reveal that lexical salience and the language user's construal operation are key factors in choice out of a synonymous set of nouns. In particular, the author finds that although a writer/speaker typically uses a noun collocation with the highest salience in a given context, a less salient synonymous item may be selected if the speaker has a different communicative intention that needs to be expressed, which results in a construal operation that competes with the factor of salience. However, the notion of construal has not been extensively applied to the study of near-synonymy, which constitutes part of the topic that the present study aims to pursue.

3.   Scope and method of analysis

The aim of this paper is to describe and analyze the semantic functions of specific numeral classifiers *kē* and *zhū* under a corpus-based cognitive linguistic framework, with the assumption that the different constructional profiles of the two classifiers in question reflect the different construals of partially overlapping conceptual contents invoked by the use of the two classifiers. Another assumption is that the two classifiers, although overlapping in their semantic functions, invoke slightly different construals. The ideas will be explored using authentic examples extracted from a corpus.

In order to investigate language in its natural habitat, we collected a representative sample of authentic use of *kē* and *zhū* from the zhTenTen corpus accessed through Sketch Engine, a corpus of simplified Standard Chinese built up via webcrawling. We manually went through and analyzed the first five hundred tokens of each classifier generated by a random search in Sketch Engine.[1]

---

1.   According to Sinclair (2004), it takes 20 tokens to have an idea of the meaning of a word not specially ambiguous and 50 tokens for an average word. Our sample size is 10 times the suggestion, which we believe is more than enough.

As we have mentioned, both classifiers refer to a similar category of nouns of PLANT, which leads us to treat them as synonymous constructions interchangeable in certain contexts. Interchanging *kē* and *zhū* as classifiers without changing the sense of a sentence or statement is possible with certain groups of nouns referring directly to plants or trees. While analyzing the nouns modified by each classifier, we also study whether there are nouns that exclusively occur with only one of the two classifiers. In the study, we analyze whether the nouns the classifiers modify fall square into the category of plants and trees, and, if not, what exactly happens in our sample. We also compare nouns modified by both classifiers and discuss what the differences mean at the conceptual and the cognitive level.

By using the corpus data, we try to answer the following series of questions: as classifier constructions that are semantically similar, do *kē* and *zhū* differ in their constructional profiles, as scholars working on other languages and other parts of speech have proposed (e.g., Divjak & Gries, 2009; Janda & Soloyev, 2009)? If so, in what way do *kē* and *zhū* differ, and does the difference tell us anything that is special about classifiers (compared to other parts of speech)?

We examine the nouns collocating with *kē* and *zhū* by putting them into semantic categories. Categories are further divided based on whether the nouns share a specific character representing a concrete group of nouns (e.g. the noun 树 *shù* 'tree' represents the group of nouns containing this character and referring to the category of trees such as 一棵木瓜树 *yī-kē-mùguāshù* 'one-CL-papaya tree', giving further taxonomical/biological information of the tree, or 一棵圣诞树 *yī-kē-shèngdànshù* 'one-CL-Christmas tree', which is a more detailed functional description of the tree). We work with the hypothesis that examining nouns collocating with these classifiers may help better understand the semantic features of each classifier and may also help uncover the possible mechanism that governs the selection between *kē* and *zhū*.

The subject of research is authentic data gathered from the zhTenTen corpus in Sketch Engine. A corpus in modern linguistics can be described as a finite-sized body of machine-readable text, sampled in order to be maximally representative of the language variety under consideration (McEnery & Wilson, 2001). Sketch Engine is a fourth-generation corpus analysis tool that allows users to analyze a wide range of corpora for lexical and lexico-grammatical patterns (McEnery & Hardie, 2012). Sketch Engine is a monitor corpus, which means that it is an open entity, with texts being constantly added to it. That allows linguists to trawl a stream of new texts looking either for the occurrence of new words or for newly emergent meanings of existing words. The procedure of looking up for concrete words in a corpora is *lemmatization*, a process of reduction of the words in a corpus to their respective lexemes – the stem word form that one looks up when consulting in a dictionary (McEnery & Wilson, 2001). The

subject of this research is data collected by looking up for collocations of two numeral classifiers *kē* and *zhū*. In total, the first 500 concordance lines of the lemma *kē* and *zhū* were extracted from the zhTenTen corpus in Sketch Engine. However, among the concordance lines, there were tokens of *kē* and *zhū* used as common nouns, which did not fall in the scope of our investigation and had to be excluded from the analysis. As a result, 479 examples of the lemma *kē* and 372 examples of the lemma *zhū* were analyzed.

Examples (2) and (3), repeated here for readers' convenience, illustrate the near-synonymy of *kē* and *zhū* in the sense of 'classifier for trees and plants'.

(2)　爸爸　买了　　两棵　　　　巨大的　圣诞树
　　　bàba mǎi-le liǎng-kē 　jùdà-de shèngdànshù
　　　father buy-PFV two-CL:plant big-LK   Christmas tree
　　　'Father bought two huge Christmas trees.'

(3)　森林　有　大概　一万多株　　　　　　树
　　　sēnlín yǒu dàgài yíwàn-duō-zhū 　　shù
　　　forest have about ten thousand-more-CL:plant tree
　　　'The forest has about more than ten thousand trees.'

A comparison of (2) and (3) and their constructed counterparts, (2a) and (3a), reveals a close relationship between meaning and use of both classifiers. In these examples, the classifiers are interchangeable and there is no noticeable change in the meaning of the sentence when the other classifier is used. However, with a corpus-based analysis, we will investigate the constructional profile of the two classifier constructions and identify the differences, in order to better understand their semantic functions in the natural habitat.

In Section 4, we discuss the nouns modified by the classifiers, and in Section 5, the numeral that precedes the classifiers.

4.   *Kē* and *zhū* as classifiers

In 4.1, we analyze the nouns modified by *kē* in our data, and in 4.2 those modified by *zhū*. Percentage in the corpus is also given.

4.1   The nouns modified by *kē*

In the corpus, we identify the category of plants, trees or parts of them as the largest group of nouns modified by *kē*. In addition to that, there are also extensions

from the typical usage of *kē* modifying nouns objectively having very little, or even nothing, to do with PLANT, such as fungi, cigarette, baton and coral.

The most common noun of this category is the word 树 *shù* 'tree', used only as a single word (两棵树 *liǎng-kē-shù* 'two-CL-trees') or its subtype, such as 一棵木瓜树 *yī-kē-mùguāshù* 'one-CL-papaya tree', 一棵枣树 *yī-kē-zǎoshù* 'one-CL-jujube tree', or 一棵桃树 *yī-kē-táoshù* 'one-CL-peach tree'), which accounts for 60.75% of the total usage in the corpus. The second largest group includes different kinds of plants which do not share much in common, such as 仙人掌 *xiānrénzhǎng* 'cactus', 玉米 *yùmǐ* 'corn' or 郁金香 *yùjīnxiāng* 'tulip', and forms a large collection of various kinds of plants, vegetables, fruits, flowers and seeds, accounting for 12.32% of usage in the category. The above two groups together represent more than half of the nouns modified by *kē*.

However, in addition to the two major groups of usages, there are two minor groups of usages that do not typically invoke the concept of PLANT. One of them is MUSHROOM, where there is the occurrence of 青冈菌 *qīnggāng jùn* 'Matsutake mushroom', a specific kind of a wild edible mushroom.[2] Here, note that MUSHROOM does not seem to be a typical member of the concept PLANT, though it scientifically belongs there and can be seen as a peripheral member of that category, which we believe is comparable to the case of PENGUIN or OSTRICH in BIRD (Rosch, 1978). In addition, in our data, there are two other occurrences of nouns not directly connected to the category of PLANT. The first one is 香烟 *xiāngyān* 'cigarette', even though it is not really kind of plant, the cigarette itself is made of tobacco, a type of plant product. The other noun is 指挥棒 *zhǐhuī bàng* 'baton', which is made of wood. We believe that the use of the two examples are motivated by the property of being LONG and THIN, a conceptual overlap that they share with the category prototype of TREE. In addition, the two nouns are connected to the category, as both are products made of certain parts of a plant.

In our data, there is one more interesting usage that is worth discussing, which is 珊瑚 *shānhú* 'coral', the scientific taxonomy being ANIMAL. We claim that the choice of Chinese classifying coral using a classifier of PLANT is not unfounded, as it is stationary as plants are, which is a typical characteristic of PLANT. Here we see that although corals are *biologically* considered ANIMAL, the *linguistic*, and folk taxonomic, choice of a classifier may not necessarily conform to that and can be culture-specific.[3]

---

2. The character 菌 *jùn is* a polysemous word having both meanings of 'mushroom' and 'bacterium'. However, as the following discussion will show, the word develops different meanings when occurring with different modifiers and with different classifiers.

3. The semantic radical of 珊瑚 *shānhú* 'coral' reveals interesting discrepancy between grammar and the writing system in Chinese. The radical of both of the characters is not a typical

As the classifier *kē* is defined prototypically as a classifier for plants and trees, it is not surprising that most of the nouns collocating with *kē* are nouns from this category. Note that there are oddballs that do not conform to the typical conception, which however reflects the culture-specific nature of linguistic categorization (corals as PLANT in the mind of Chinese speakers). Table 1 summarizes the data discussed above, with the percentage in the corpus also given.

Table 1. Nouns that occur in the [QUANTIFIER] – [KE] – [NOUN] construction

	Noun	Tokens	% (individual)	% (category)
Prototypical plants	树 *shù* 'tree'	291	60.75	99.16
	Specific kinds of plants	59	12.32	
	菜 *cài* 'vegetable'	28	5.85	
	木 *mù* 'wood'	20	4.18	
	苗 *miáo* 'seedling'	23	4.8	
	植 *zhí* 'plant'	17	3.55	
	花 *huā* 'flower'	2	0.42	
	草 *cǎo* 'grass'	35	7.31	
Mushroom	菌 *jùn* 'mushroom'	1	0.21	0.21
Others	香烟 *xiāngyān* 'cigarette'	1	0.21	0.63
	指挥棒 *zhǐhuī bàng* 'baton'	1	0.21	
	珊瑚 *shānhú* 'coral'	1	0.21	
		479	100.00	100.00

In 4.2, we discuss the case of *zhū* as a classifier in the zhTenTen corpus.

4.2   The nouns modified by *zhū*

In this section, we discuss the data of *zhū* extracted from the zhTenTen corpus. The sample size for *zhū* is slightly smaller than that of *kē*, as there were more tokens of *zhū* used as common nouns. A total of 372 examples is analyzed from the first 500 hits of zhTenTen.

In our data, the main group of nouns modified by *zhū* is the one of plants and trees. Nouns consisting of the character 树 *shù* 'tree' represent the most frequent sub-group in the category of plants, which accounts for 47.3% of all the occur-

---

one for PLANT (⁺⁺), but instead the radical of 玉 *yù* 'jade', highlighting the cultural association of being precious of corals.

rences. In the sample, what comes second is a part of plant 苗 *miáo* 'seedling', a part of plant small in size and vulnerable, taking 14.3% of all the occurrences. Another part of the plant that follows is 花 *huā* 'flower', which also invokes the concept of VULNERABLE. In addition to the prototypical members that are related to PLANT and TREE, the first sub-group of nouns which is not connected to either plants or trees is the category of MICRO-ORGANISM. The category consists of words containing either 霉 *méi* 'mold', as in 毛霉 *máo-méi* 'mucor (lit. hair-mold)', 青霉 *qīng-méi* 'penicillium (lit. green-mold)', 曲霉 *qū-méi* 'aspergillus (lit. ferment-mold)', 木霉 *mù-méi* 'tolypocladium inflatum (lit. wood-mold)', or 菌 *jùn* 'bacterium', as in 白腐真菌 *bái-fǔ-zhēnjùn* 'white-rot fungus', 有机磷细菌 *yǒujī-lín-xìjùn* 'organophosphorus-degrading bacterium'.

Mold and bacteria are all micro-organic entities that biologically fall under the category of PLANT, though quite peripheral ones.[4] Similar to MUSHROOM as a peripheral member of PLANT in Chinese and the classic BIRD example (Rosch, 1978), through the microscope of the Chinese classifier system, again we witness how the classifier acts as a folk categorizing system that subsumes MOLD and BAC-TERIUM as a type of PLANT, which is not linguistically elaborated in the *grammar* in most other linguistic categorizing systems of the world.

However, note that the usage extension of *zhū* to MOLD and BACTERIUM is nei-ther random nor purely driven by biological taxonomy. The extension in usage actually has its root in which part of the conceptual domain of PLANT is high-lighted by the use of *zhū*. As introduced previously, the second frequent part of PLANT modified by *zhū* is 苗 *miáo* 'seedling', which invokes the concept of SMALL. Mold and bacteria are relatively small organisms that are either visually station-ary or too small to be seen. Therefore, SMALL seems to be the key to the category extension of *zhū* from prototypical exemplars of PLANT to peripheral and less fre-quent common nouns of that linguistic category.

There are other peripheral members in the linguistic category of *zhū* that invoke the concept of SMALL, but have nothing to do with PLANT. These common nouns are chemical or biological substance, such as 法氏囊病病毒 *fǎshìnángbìng-bìngdú* 'infectious bursal disease-virus' and 抗体 *kàngtǐ* 'antibody', as in 单株抗体 *dān-zhū-kàngtǐ* 'monoclonal antibody (lit. single-CL-antibody)' both being a kind of protein, or 对硝基苯酚 *duì-xiāojī-běn-fēn* 'P-nitrophenol', a kind of a chemical compound. These nouns do not seem to share much in common but their micro-scopic size. From the distribution of *zhū* with *miáo, méi, jùn* and with *bìngdú, kàngtǐ* and *duì-xiāojī-běn-fēn*, we see that SMALL is at least an important concept

---

4. In the past, bacteria were classified as a type of PLANT constituting the class *Schizomycetes*, though in modern microbiology, bacteria are considered to constitute their own domain.

invoked by the classifier *zhū*, even if not a central one shared by all the members of that linguistic category.

The distribution of the nouns modified by *zhū*, including the percentages, is summarized as Table 2.

Table 2. Nouns that occur in the [QUANTIFIER] – [ZHU] – [NOUN] construction

	Nouns	**Tokens**	**% (individual)**	**% (category)**
Plants	树 *shù* 'tree'	176	47.3	95.5
	Specific kinds of plants	36	9.7	
	苗 *miáo* 'seedling'	53	14.3	
	木 *mù* 'wood'	17	4.6	
	植 *zhí* 'plant'	23	6.2	
	花 *huā* 'flower'	35	9.4	
	草 *cǎo* 'grass'	15	4.0	
Micro-organism	霉 *méi* 'mold'	8	2.2	2.7
	菌 *jùn* 'baterium'	2	0.5	
Others	Biological substance	5	1.3	1.8
	Chemical substance	2	0.5	
		372	100	100

Below, we compare the sets of nouns modified by *kē* and *zhū*, with a view to coming up with a conceptual-semantic division of labor of the two classifiers in Chinese.

4.3   A comparison of the nouns modified by *kē* and *zhū*

As we have discussed in 4.1 and 4.2, *kē* and *zhū* both frequently collocate with nouns of PLANT and TREE, which means that the two are truly semantically overlapping. In particular, both classifiers collocate most often with *shù* 'tree', with more than 60% of the nouns with *kē* and over 47% with *zhū*. Given the highest frequency in both samples, *shù* should be considered the prototypical common noun that occurs in both linguistic categories. This means that the two categories have largely overlapping prototypes, and that the two classifiers in such prototypical circumstances are indeed interchangeable.

Numbers of tokens of each classifier among the other groups within the category of plants differ. Nouns that are a type of *huā* 'flower' more often collocate with the classifier *zhū* than *kē,* indicated by the percentage difference of almost

9% (with the former being 9.41% and the latter only 0.42%). A similar case is with the group of *miáo* 'seedling'. In our samples, *miáo* occurs much more often with the classifier *zhū* (14.25%, as opposed to 4.80% with *kē*). Given the conceptual fact that the words *huā* and *miáo* are smaller and more tender parts of PLANT, we may generalize that *zhū* is more often associated with a specific meaning of 'smaller (part) of plant' than *kē*.

Another important observation to make is that, although both classifiers take the noun 菌 *jùn*, the lexical construction has completely different meanings (though possibly related as polysemy) when occurring with the different classifiers. When it occurs with *kē, jùn* means 'mushroom (a peripheral member of the category PLANT)', while when it occurs with *zhū, jùn* has the meaning of 'bacterium', which is even much less prototypical of PLANT given the microscopic physical size of the entity referred to.

In addition, another important difference between the two classifiers is the fact that only *zhū* may co-occur with nouns denoting biological or chemical substance, entities with a microscopic physical size, thus revealing an important functional-semantic division between the two synonymous classifier constructions.

One last difference between *kē* and *zhū* is that the word 菜 *cài* is only found with the classifier *kē*, which has a meaning of 'a head of (cabbage)', as in 一棵白菜 *yī-kē-báicài* 'one-CL-cabbage'. But given the relatively larger size of a cabbage (in comparison to mold and bacteria), it is natural for the noun *báicài* to exclusively occur with *kē* rather than *zhū*.

In addition to a difference in the nouns modified, the two classifiers have another meaningful difference in the quantity of the noun verbalized in the scene.

5.   The quantifier in [QUANTIFIER] – [CLASSIFIER] – [NOUN]

In literature, no study that we are aware of discusses the relationship between the quantifier and the classifier. However, in our language samples, there is an obvious difference in the numerals used before the classifiers (in the [QUANTIFIER] – [CLASSIFIER] – [NOUN] schema), which we will present and discuss below.

In our sample of *kē*, more than 70% of all numerals are a number lower than 10, with the most frequent being 一 *yī* 'one', which accounts for more than half of all the tokens (232 out of 428, or 54.21%).[5] Numerals above 10 represent only about 27% of all hits.

---

5.   The reason why the number of analyzed numerals differ from the number of analyzed nouns is that there were cases of demonstratives such as 这 *zhè* 'this' or 那 *nà* 'that', which were not quantifiers.

On the other hand, in our sample of *zhū*, more than half of the numerals are numbers higher than 1,000 (52.94% of the total 340 tokens), which shows the tendency of *zhū* to classify nouns representing higher quanta. Another piece of evidence for that argument is the comparatively less frequent occurrence of the numeral *yī* 'one' with *zhū*, accounting only for 81 tokens out of 340 (23.82%), which is less than half of *kē*.

The co-occurrence of numerals with *kē* and *zhū* is summarized in Table 3 below.

Table 3. The distribution of quantifiers in [QUANTIFIER] – [CLASSIFIER] – [NOUN]

Range	Numerals	Classifier *kē* Tokens	Classifier *kē* %	Classifier *zhū* Tokens	Classifier *zhū* %
X ≤ 10	一 *yī* 'one'	232	72.43	81	37.64
	两 *liǎng* 'two'	25		23	
	三 *sān* 'three'	18		15	
	Between 4 and 10	35		9	
X ≤ 100	百 *bǎi* 'hundred'	7	8.88	3	4.13
	Others	31		11	
X ≤ 1000	千 *qiān* 'thousand'	2	6.31	1	5.29
	Others	24		17	
X > 1000	千 *qiān* 'thousand'	3	12.38	5	52.94
	万 *wàn* 'ten thousand'	17		58	
	亿 *yì* 'hundred million'	1		8	
	Others	33		109	
		428	100.00	340	100.00

From the table, we see that the distribution of numerals is polarized – Numerals of the classifier *kē* are mostly represented by the numbers lower than 10 (more than 72% of all hits) which is almost two times more than numerals of *zhū* (37.64%). On the contrary, numerals of *zhū* most frequently go above 1,000 (more than 50%). In comparison with numerals of *kē, zhū* has over four times more numerals out of this interval.

6.   When it all adds up: The construal invoked by [QUANTIFIER] – [KE/ ZHU] – [NOUN]

As we have seen in the corpus-based analysis of the two classifiers, both *kē* and *zhū* mostly collocate with nouns of the category of PLANT and TREE, which shows their overlapping distribution and functional similarity. Both classifiers collocate most frequently with *shù* 'tree' or constructions that invoke a similar concept. In our samples, more than 60% of the analyzed nouns of *kē* and over 47% of *zhū* modify *shù*. As it has been already mentioned, *kē* and *zhū* also act as common nouns that have a meaning of 'trunk of a tree', so it is not unnatural for both classifiers to have TREE in the prototypical scenario where the [QUANTIFIER] – [CLASSIFIER] – [NOUN] is a part of. That constitutes the shared functional-conceptual core of the two classifiers at issue.

In addition to the shared distribution, the two classifiers each has its own distributional preferences. In our samples, we see that nouns with *huā* 'flower' more frequently collocate with the classifier *zhū* than *kē*. A similar case is the group of words that contain *miáo* 'seedling', which collocate more often with the classifier *zhū* in our corpus. We see that both *huā* and *miáo* are small or vulnerable parts of PLANT, which means that *zhū* has developed its own meaning of, or conceptual association with, "smaller part of a plant". On the other hand, the group of vegetables containing the character *cài* 'vegetable' is only used with the classifier *kē*, which gives *kē* a meaning of 'a head (of cabbage)'. There is also a difference in the meaning of *jùn* when it co-occurs with *kē* and *zhū* – when *jùn* occurs with *kē*, it has the meaning of 'mushroom', but when it occurs with *zhū*, it develops a different, though related, meaning of 'bacterium'. The difference in the spatial dimension between 'mushroom' and 'bacterium' serves as a piece of converging evidence that SMALL is indeed part of the conceptual semantics of *zhū*. Another important piece of evidence for *zhū* to have SMALL as part of its conceptual content comes from its co-occurrence with nouns of biological and chemical substance.

Another important behavioral difference, which has escaped the attention of previous cognitive linguistic research (and classifier research), is that two classifiers can differ in the numerals they tend to co-occur with. As discussed in Section 5, the majority of the quantity that occurs with *kē* is 1 (over 50%), and the number under 10 accounts for over 72% of the entire span of its usage. In contrast, *zhū* has a very different preference – over half of the usage is over the quantity of 1,000.

If we add the above differences in the size of the modified nouns and the co-occurring numerals together, we get a full picture of how the two classifiers behave at the cognitive level – *zhū* prefers to modify nouns of a smaller size but of

a larger quantity, which is not characteristic of *kē*. Following that, the conceptual difference between the two classifiers can be visualized as Figure 1(a) and 1(b), which illustrates the respective construals invoked by *kē* and *zhū* in the constructional schema [QUANTIFIER] – [CLASSIFIER] – [NOUN].
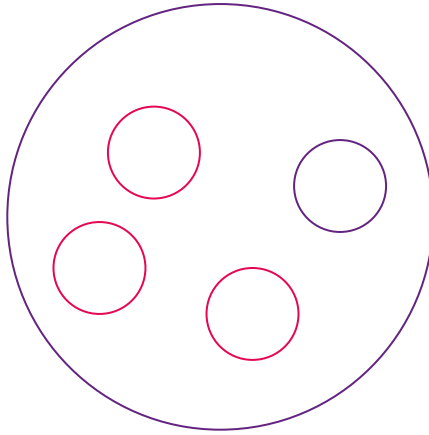
a.   [QUANTIFIER] – [KE] – [NOUN]
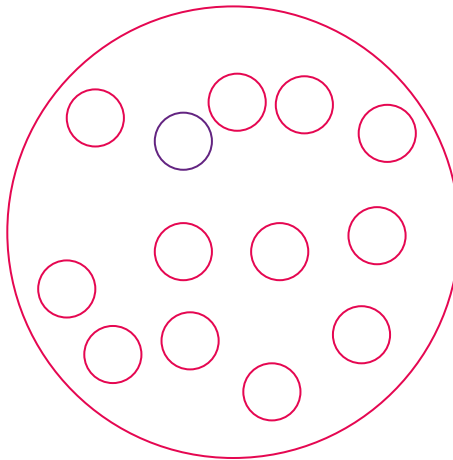
b.   [QUANTIFIER] – [ZHU] – [NOUN]

Figure 1. Visual representation of the construal of [QUANTIFIER] – [KE/ZHU] – [NOUN]

If we compare the two figures, it is obvious that the construal invoked by [QUANTIFIER] – [ZHU] – [NOUN] provides a higher-resolution, and a more granular view of the scene linguistically elaborated, whereas [QUANTIFIER] – [KE] –

[NOUN] does not share that preference. This is an important difference between the synonymous classifiers at the cognitive level that becomes available only when we go beyond the traditional research scope, i.e. the modified nouns by looking also at the co-occurring quantifier.[6]

In the present paper, we have discussed the roles played by the object noun and by the quantifier in [QUANTIFIER] – [CLASSIFIER] – [NOUN]. However, an important theoretical issue raised by our discussion is that the role played by the quantifier, which has not been taken seriously in the existing literature, is something to look into in theoretical classifier research, as it has been demonstrated that the quantifier and the object classified together invokes a holistic construal, which, when visualized, is capable of revealing a conceptual difference between synonymous classifier constructions.

From a cognitive linguistic perspective, the present paper also makes useful contribution, in the sense that the present study extends the existing research on near-synonymy to include another part of speech, and shows how construal operation is indeed a factor to consider when studying not only synonymous nouns (Liu, 2013) but also other synonymous constructions of other parts of speech.

Acknowledgements

List of abbreviations used

CL   classifier
LK   linker
PFV perfective aspect

---

6. Note that what is claimed here is based on two *separate* observations that [QUANTIFIER] – [ZHU] – [NOUN], compared to [QUANTIFIER] – [KE] – [NOUN], tends to attract quantifiers of a higher number and nouns of a smaller size in its two different slots. But it does not necessarily mean that the two sets of constructions are both attracted to, and co-occur in, the two slots of [QUANTIFIER] – [ZHU] – [NOUN], although the co-occurrence of the three constructions does look likely and can be hypothesized. However, in order to really attest such claim of co-occurrence of the three constructions, one would need to resort to the *co-varying collexeme analysis* (Stefanowitsch & Gries, 2005), which is beyond the methodological scope of the current study but is definitely an issue worth future pursuit.

References

Aikhenvald, A. Y. (2000). *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

Allan, K. (1977). Chinese classifier systems and human categorization. *Language*, 53(2), 285–311. https://doi.org/10.1353/lan.1977.0043

Divjak, D. S., & Gries, S. T. (2009). Corpus-based cognitive semantics: A contrastive study of phrasal verbs in English and Russian. In B. Lewandowska-Tomaszczyk, & K. Dziwirek (Eds.), *Studies in cognitive corpus linguistics* (pp. 273–296). Frankfurt am Main: Peter Lang.

Erbaugh, M. S. (2006). Chinese classifiers: Their use and acquisition. In H. T. Li, E. Bates, & O. J. L. Tzeng (Eds.), *Handbook of East Asian psycholinguistics: Chinese* (pp. 39–51). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511550751.005

Glynn, D. (2014). The social nature of ANGER: Multivariate corpus evidence for context effects upon conceptual structure. In I. Novakova, P. Blumenthal, & D. Siepmann (Eds.), *Emotions in discourse* (pp. 69–82). Frankfurt: Peter Lang.

Goddard, C. (2011). *Semantic analysis: A practical introduction*. Oxford: Oxford University Press.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.

Hanks, P. (1996). Contextual dependency and lexical sets. *International Journal of Corpus Linguistics*, 1, 75–98. https://doi.org/10.1075/ijcl.1.1.06han

Hirst, G. (1995). Near-synonymy and the structure of lexical knowledge. In *AAAI symposium on representation and acquisition of lexical knowledge: Polysemy, ambiguity, and generativity* (pp. 51–56). Menlo Park, CA: AAAI Press.

Hirst, G., & Edmonds, P. (2002). Near-synonymy and lexical choice. *Computational Linguistics*, 28(2), 105–144. https://doi.org/10.1162/089120102760173625

Janda, L. A., & Solovyev, V. D. (2009). What constructional profiles reveal about synonymy: A case study of Russian words for sadness and happiness. *Cognitive Linguistics*, 20(2), 367–393. https://doi.org/10.1515/COGL.2009.018

Jiang, S. (2017). *The semantics of Chinese classifiers and linguistic relativity*. London: Routledge. https://doi.org/10.4324/9781315265483

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: Chicago University Press. https://doi.org/10.7208/chicago/9780226471013.001.0001

Langacker, R. W. (2008). *Cognitive Grammar: An introduction*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195331967.001.0001

Liu, D. (2013). Salience and construal in the use of synonymy: A study of two sets of near-synonymous nouns. *Cognitive Linguistics*, 24(1), 67–113. https://doi.org/10.1515/cog-2013-0003

McEnery, T., & Hardie, A. (2012). *Corpus linguistics*. Cambridge: Cambridge University Press.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Myers, J. (2000). Rules vs. analogy in Chinese classifier selection. *Language and Linguistics*, 1(2), 187–209.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale: Lawrence Erlbaum.

Sinclair, J. (2004). Corpus and text: Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxford: Oxbow Books.

Stefanowitsch, A., & Gries, S. T. (2005). Covarying collexemes. *Corpus Linguistics and Linguistic Theory*, 1(1), 1–43. https://doi.org/10.1515/cllt.2005.1.1.1

Su, L., & Lu, W. (2009). A new look at analogous words: A corpus-based approach. In B. Lewandowska-Tomaszczyk & K. Dziwirek (Eds.), *Studies in cognitive corpus linguistics* (pp. 191–206). Frankfurt/Main: Peter Lang.

Tai, J. H. Y. (1994). Chinese classifier systems and human categorization. In M. Y. Chen & O. J. L. Tzeng (Eds.), *Interdisciplinary studies on language and language change* (pp. 479–494). Taipei: Pyramid.

Dictionary consulted

Oxford Chinese dictionary . (2010). Oxford: Oxford University Press.

Construction in conversation

An Interactional Construction Grammar
approach to the use of *xiangshuo* 'think' in spoken
Taiwan Mandarin

Chen-Yu Chester Hsieh and Lily I-Wen Su
National Taiwan University (NTU)

Although the Construction Grammar (CxG) model has yielded fruitful
findings, the role that pragmatics plays in language has not yet been fully
considered in this theoretical framework. The recent development of
spoken corpora, however, enables construction grammarians to develop a
new approach called Interactional Construction Grammar, which
incorporates interactional factors into CxG analysis to account for patterns
that involve interpersonal functions and global contexts. Adopting this
approach, the present study attempts to examine the use of a complement-
taking mental predicate *xiangshuo* in Taiwan Mandarin conversation and
analyze the co-occurrence patterns of this cognitive verb with different
subjects. We identify three sequential patterns in which *xiangshuo* most
frequently occurs, including account-giving, contrast-projecting and
involvement-constructing, and argue that only by taking into account the
sequential context and interactional function can the distribution patterns
of subjects and particles that recurrently collocate with *xiangshuo* be
explained.

Keywords: Interactional Construction Grammar, complement-taking
mental predicate, *xiangshuo* , spoken corpus, Taiwan Mandarin, sequential
environment

1. Introduction

In the past 30 years, Construction Grammar (CxG) has developed into a
full-blown model and yielded fruitful findings (Croft, 2001; Fillmore, Kay, &
O'Connor, 1988; Goldberg, 1995). One of the most important assumptions
shared among construction grammarians is that constructions are an essential

building block of grammar. In CxG, a construction is referred to as a form-meaning pairing (Goldberg, 1995), which can be realized at different levels of units, from morphemes and words to sentences and discourse (Antonopoulou & Nikiforidou, 2011). Constructions can also range from highly specific and idiomatic phrases like *let alone* (Fillmore et al., 1988) to very general and common structures like the passive voice (Croft, 2001).

Although the CxG approach to language has proven to be a very powerful, revolutionary theory, a few limitations have been identified. First, while the founders of CxG have pointed to the importance of pragmatics (Croft, 2001; Fillmore et al., 1988; Goldberg, 1995), most CxG studies either put little emphasis on this aspect or treat pragmatics in a rather narrow way (Fried & Östman, 2005; Linell, 2009; Wide, 2009).

Second, even if pragmatics is taken into account, the majority of CxG research is writing-centered (cf. Linell, 2009). Examples discussed in previous studies are predominantly constructed (cf. Fillmore et al., 1988) or based on written data (cf. Goldberg, 1995). Features of spoken language, such as turn-taking and repair-initiating are mostly overlooked, despite the attested influence of spoken interaction on lexico-grammatical patterns (e.g., Hakulinen & Selting, 2005; Selting & Couper-Kuhlen, 2001).

Third, as highlighted by Linell (2009), Fillmore (1988) in fact distinguishes between "internal syntax" and "external syntax," when discussing the mechanisms of CxG. According to Fillmore (1988, p. 36), the internal syntax of a construction describes the composition of the construction, and the external syntax refers to the larger contexts in which the construction as a whole can be used. Taking a dialogical view toward grammatical constructions, Linell (2009) suggests that while the former has received much more attention, external syntax is relatively unexplored and requires more investigation.

With the increasing availability of spoken corpora, researchers have proposed integrating CxG and Interactional Linguistics and developed a framework called "Interactional Construction Grammar" (ICxG) (Imo, 2015; Linell, 2009; Wide, 2009) to analyze the usage patterns of constructions in spoken discourse. Fried and Östman (2005), for example, study the use of pragmatic markers as constructions in spoken Czech and Solv[1] by considering three parameters including coherence, politeness, and involvement, in addition to information concerning their syntax and semantics.

Linell (2009) and Wide (2009), on the other hand, highlighted the interaction between sequences and the meaning/functional potential of a grammatical construction in interaction. Examining the x-*och*-x construction in spoken

---

1. Solv is a Swedish dialect spoken in southern Ostrobothnia.

Swedish as an example, Linell (2009) identified three general aspects that should be investigated when one adopts the ICxG approach, including the prior sequences that give rise to the construction, the sequences that ensue from the construction, and the co-occurring features that the construction prefers. Similarly, Wide (2009), analyzing the demonstrative construction in spoken dialectal Swedish (eastern Nyland), also includes the attributes of sequential position and activity in the formalized description of the construction.

Regardless of their research focus, scholars of ICxG underline the importance of conversational corpus data, which allows for the investigation of both the quantitative (e.g., frequency) and qualitative (e.g., context) sides of language. Moreover, it has been argued that insofar as conversation is the initial and primary locus of language use and speakers always need to utilize verbal and nonverbal resources to manage the ongoing interaction, examining linguistic practices in conversation is key to the understanding of how grammar is used and shaped in reality (Fox, 2007).

However, research adopting the ICxG framework is still relatively lacking in the study of Asian languages (cf. Biq, 2004; Hsieh & Huang, 2005).[2] To contribute to this line of literature, the present study sets out to examine the expression *xiangshuo* 'think' (Huang, 2003; Sanders & Uehara, 2006) as a thought-reporting construction in Taiwan Mandarin conversation, based on the data retrieved from the NTU Corpus of Spoken Chinese. The construction *xiangshuo* is chosen as the focus of this study for primarily two reasons. First, it is still an emerging pattern (Huang, 2003) and not much research has addressed its use. Second, *xiangshuo* does not seem to be as frequently used in standardized written Chinese and is primarily observed in spoken Taiwan Mandarin. The use of spoken corpus data and the ICxG approach thus allows us to better capture and explain the usage patterns of such a construction.

In what follows, we will review the previous discussions on *xiangshuo* and introduce the corpus and methodology that we adopt for the current research. We will analyze the common sequential environments in which *xiangshuo* is used and elucidate how local patterns, such as its subjects and other co-occurring expressions, are related to the sequence in which *xiangshuo* is employed. In doing so, we will illuminate how the use of conversational data helps incorporate significant interactional factors into CxG analysis, and will put forth a more consistent and comprehensive explanation for the features of the complement-taking mental predicate (CTMP) in question.

---

2. In fact, Biq (2004) and Hsieh and Huang (2005) do not explicitly refer to the ICxG approach as their analytical framework, despite their use of conversational data and their integration of concepts from CxG and Interactional Linguistics.

2.   Previous research on *xiangshuo*

To our knowledge, few have investigated the use of *xiangshuo* in Mandarin. Huang (2003) and Sanders and Uehara (2006) are the only studies we can find that distinguish between *xiang* and *xiangshuo* and analyze the differences between the two CTMPs. In what follows, we will summarize the findings of these two papers and identify a few potential research gaps.

The first study that discusses the use of *xiangshuo* is Huang's (2003) paper on the topic of complementation in Mandarin. Huang (2003) examines several complement-taking predicates in Taiwan Mandarin conversation and reveals that *xiang* 'think,' which he deems the most canonical mental verb in Chinese, is the predicate that most frequently co-occurs with a following element *shuo* 'say.' According to his report, 51.9% of the *xiang* tokens are followed by *shuo*, while less than 10% of other mental predicates are. He also notices that compared to *xiang, xiangshuo* is more often used as a deontic, rather than an epistemic, verb.[3] However, Huang (2003) seems to regard *xiangshuo* as two separate components, that is, a predicate *xiang* and a complementizer *shuo*, instead of as a lexicalized item. Drawing on the fact that *xiangshuo* is often followed by interjections, or speech act particles in his terms, he characterizes *shuo* as a "*de dicto* introducer" (Huang, 2003, p. 438), as illustrated in (1), which signals that what follows the marker are elements of speech, as opposed to "*de re*" complements, which represent elements of reality, as exemplified in (2).[4]

(1) *xiangshuo* as a *de dicto* introducer
     320 M: → (0) *wo shi xiangshuo,_*
              1SG COP think

     321   *.. ei?/*
              INT

     322   *.. wo haoxiang meiyou   kan dao you--*
              1SG seem   NEG.have see to   have

     323   *.. you zhege kemu.*
              have this.CL division
     'I thought, well, I didn't seem to see such a division (in a hospital).'

---

3. According to Huang (2003), the deontic use of a cognitive verb is related to the subject's plan or intention while the epistemic use indicates the speaker's reasoning process or source of knowledge. However, this distinction will not be made in the present study.

4. Huang (2003) does not provide any example of *xiang*(*shuo*) to directly illustrate this point. Both (1) and (2) come from the current dataset that we use.

(2) *xiang* and *de re* complements
　　176 A: → … *wo xiang ta   na   liudonglü   hen gao hoN.*
　　　　　　　　1SG think 3SG that turn.over.rate very high FP
　　177 B:　　.. *dui   a,_*
　　　　　　　　right FP
　　'A: I think the turn-over rate there is really high, right?
　　B: Yeah.'

In contrast, Sanders and Uehara (2006) treat *xiangshuo* as a lexicalized expression independent from *xiang* and adopt a survey questionnaire method to test Taiwan Mandarin (TM) speakers' intuition specifically about the acceptability of the co-occurrence of first and third-person subject pronouns with *xiang* and *xiangshuo* respectively. Comparing the results of monolingual TM speakers and bilinguals who speak both TM and Taiwan Southern Min (TSM), they note that TM speakers in general find it more appropriate to have a third-person subject pronoun with *xiangshuo* than with *xiang*, despite prescriptive Chinese grammar suggesting otherwise. They also report that monolingual and bilingual TM speakers show differences in the degree of acceptance. Drawing on these findings, Sanders and Uehara (2006) argue that such disparities may be shaped by cognitive mechanisms such as subjectivity, and sociolinguistic factors such as language contact.

In sum, these two studies identify a number of differences between *xiang* and *xiangshuo*. They are distinctive not only in the modal meaning that each of them tends to convey, but also in the type of subject and complement that they respectively prefer. These two CTMPs should thus be treated as two different constructions.[5]

Although different research methods and analytic frameworks have been adopted to approach this topic, a few gaps still remain. First, both Huang (2003) and Sanders and Uehara (2006) focus more on the quantitative aspect of the CTMP. Rarely do they look specifically into the particular environments where *xiangshuo* is deployed. Second, even when *xiangshuo* is examined in an utterance, the sequences that motivate the use and the functions that it makes possible are not discussed. Finally, while both studies pay more attention to the use of a first-person subject with *xiangshuo* or the contrast between the first and third-person pronouns, the actual subject use in a *xiangshuo*-utterance is far more complex. In fact, *xiangshuo* can also be used with a second-person subject or even without a subject, which deserves further discussion.

---

5. Although the comparison of these two CTMPs is in and of itself an interesting topic, we will not pursue it here and will focus only on *xiangshuo* instead.

In light of these gaps, our current study aims to answer the following research questions:

RQ1: What are the distribution patterns of subjects and other collocates with the CTMP *xiangshuo* in Taiwan Mandarin conversation?

RQ2: What are the sequential patterns in which *xiangshuo* is recurrently employed in talk-in-interaction?

RQ3: Is there any link between the sequential context in which *xiangshuo* is deployed and the use of a particular subject or other co-occurring expressions?

In the next section, we will introduce the corpus and methodology that enable us to find answers to these research questions.

3.   Data and methodology

The data of this study comes from the NTU Corpus of Spoken Chinese, which contains transcripts of spoken data, including face-to-face conversations, radio talk shows and telephone dialogs, up to around 15 hours in total length. All of the recorded participants are native speakers of Mandarin Chinese spoken in Taiwan. The recordings are transcribed according to the system proposed by Du Bois, Schuetze-Coburn, Cumming, and Paolino (1993). In total, 94 instances of *xiangshuo* are retrieved and constitute the basis for our analysis.

Adopting the framework of ICxG (Linell, 2009), we look specifically into the distribution patterns of subjects that co-occur with *xiangshuo* and the sequential contexts in which the cognitive verb is employed. In contrast with Huang (2003) and Sanders and Uehara (2006), we also identify the lack of an overt, same-clause subject[6] as an independent category, because not every CTMP is commonly used without an overt subject in a clause. For example, when used to introduce a clause, *xiang* 'think' alone rarely, if ever, occurs without an explicit subject in the same clause. Following Linell (2009, p. 99), we examine not only the utterance in which *xiangshuo* is used but also the adjacent sequential contexts, which is treated as the external grammar of the *xiangshuo* construction. Other recurrent collocates of *xiangshuo* are also under scrutiny.

---

6. In Mandarin, a predicate can take a zero-pronoun subject, especially when the subject is implied in or can be inferred from the context (Li & Thompson, 1981). In the current article, such a subject will be referred to as no overt (same-clause) subject or a zero-marked subject.

4.   The co-occurrence patterns of *xiangshuo* and different subjects

In this section, we will present the distribution patterns of subjects that occur in the *xiangshuo*-utterance. First, similar to many other CTMPs, *xiangshuo* exhibits a strong tendency to co-occur with an overt first-person singular pronoun (cf. Huang, 2003; Sanders & Uehara, 2006). However, as shown in Table 1 below, while 62.77% of the instances of *xiangshuo* have a first-person singular subject, other types of subjects, including zero-marked, second-person and third-person subjects, are also found.

Such a phenomenon is worth mentioning for two reasons: cognitively, it is difficult, if not impossible, to know exactly the interlocutor's inner state, and even if it is likely to know how the interlocutor thinks, it may not be appropriate to report it directly. In fact, a number of typologically distant languages constrain or even prohibit speakers from reporting non-first-person thought (Aikhenvald, 2004; Iwasaki, 1993). This shows that analyzing the co-occurrence pattern of subjects and mental verbs merely from a cognitive or semantic point of view is insufficient. We will argue later that the use of *xiangshuo* with a second-person subject in fact serves particular interactional functions and the construction is thus recurrently deployed in Taiwan Mandarin conversation regardless of the cognitive anomaly that it may imply.

Table 1. Distribution of subjects in the *xiangshuo* construction

	N	%
Zero-marked subject	13	13.83
First-person singular	59	62.77
First-person plural	7	7.45
Second-person singular	8	8.51
Third-person singular	5	5.32
Abstract noun	2	2.13
Total	**94**	**100**

Another noteworthy observation is that 13.83% of the *xiangshuo* instances do not have an overt subject in the same clause, as illustrated in (3) below, while this is rarely so in the case of *xiang* (Huang, 2003). This may be in part due to the disyllabicity of *xiangshuo*, which makes it a phonologically more acceptable unit than *xiang* in modern Mandarin conversation. This feature also reflects the weaker link between *xiangshuo* and any particular overt subject form, hinting at its status as an expression or a construction independent from any particular

subject, as opposed to (*wo*) *xiang* '(I) think' (Huang, 2003) and (*wo*) *juede* 'I think/feel' (Huang, 2003; Lim, 2011).

(3)   *xiangshuo* with a zero-marked subject
   141 B: (0) *ta   jiu shi zheyangzi   a,_*
               3SG just COP this.appearance FP
   142   *.. mingming zhidao nayiben   lan,*
              obviously know   that.one.CL bad
   143   *.. ta   hai yao   nian,*
              3SG still want read
   144   *.. juede hen qiguai.*
              feel   very strange
   145 A: ... [*ta shi*],\
              3SG COP
   146 B: → [*xiangshuo*],\
              think
   147   *.. junzi   yi   yan,*
              gentleman one word
   148   *... ji   chu si   ma   nan   zhui.*
              already out four horse difficult catch.up
   'B: He is always like that. He knows that the book is bad, but he still wants to read it. (I) find it really strange.
   A: He is-,
   B: (He may) think/(I think he is like) a promise is a promise.'

Even when the subject is an overt first-person singular pronoun, the verb *xiangshuo* does not always follow the subject immediately. Of the 59 instances in which *wo* 'I' serves as the subject, only 18 (30.51%) tokens of *xiangshuo* were produced immediately after the subject without any other particles. This serves as another piece of evidence that although the CTMP in question prefers to occur with a first-person subject, the subject and the predicate do not seem to form a prefabricated unit as *wo xiang* 'I think' (Huang, 2003) and *wo juede* 'I think/feel' (Lim, 2011) do. Moreover, the particles inserted between the subject and *xiangshuo*, such as *jiu* 'just,' *benlai* 'originally' and the aspect maker *zai*, are not randomly chosen, but show a consistent pattern, which, as we will argue, is also related to the broader sequential pattern in which the construction is used.

   Finally, and most surprisingly, *xiangshuo* can even take an inanimate abstract noun as its subject. In our data, two instances of *xiangshuo* produced by different speakers in different episodes of interaction are preceded by *yisi* 'meaning, idea, intention' as the subject of the same clause, as exemplified in (4) below. In this segment, D uses *xiangshuo* to launch his report of what the patient's wife thought

about the patient's sickness. Although the main predicate *xiangshuo* is a cognitive verb, the subject of the clause is an inanimate abstract noun *yisi* 'meaning,' instead of the cognizer being, i.e., the patient's wife. Notice that neither the speaker nor the co-participant initiates any repair. In other words, speakers in this extract seem to perceive this usage as an acceptable pattern. This shows that *xiangshuo* may be undergoing semantic bleaching and is emerging as a quotative for reported thought.[7]

(4) *xiangshuo* with an abstract-noun subject

    388 D: .. *na,*
             that

    389 → .. *na ta laopo de yisi   jiu xiangshuo,_*
             that 3SG wife GEN meaning just think

    390 .. *ni kan,*
             2SG see

    391 .. *yijing   manyan dao toushang,*
             already extend   to   head.top

    392 .. *tou ye zai tong,*
             head also ASP hurt

    'D: Well, his wife thinks that the cancer has spread to his brain, which might be causing his headaches.'

In summary, as can be seen from the above discussion, the subject pattern of *xiangshuo* is far more complex than described in previous studies. It can be used not only with first-person subjects but also with second-person, zero-marked or even abstract-noun subjects. Even in cases where the subject is first-person singular, speakers often would insert particles between the subject and the CTMP. As such, *xiangshuo* does not seem to be as closely tied to the first-person singular subject as other high-frequency CTMPs in Chinese like *xiang* 'think' and *juede* 'feel'. These patterns can hardly be accounted for based on the semantics or local syntax of the subject-verb pair. An ICxG approach is thus needed to produce a more systematic and comprehensive analysis.

---

7. One of the reviewers points out that this may be an example of metonymy, i.e., the speaker is the words that s/he speaks. This may very likely be the case, for metonymy is often the motivational force for grammaticalization (Hopper & Traugott, 2003). However, it should be noted that the combination of an abstract-noun subject and a thought-reporting construction is relatively rare in our data. Such use may be more or less conventionalized and restricted to *yisi* and *xiangshuo*.

5.   Sequential patterns of *xiangshuo* in interaction

The data from the NTU Corpus suggests that *xiangshuo* occurs most commonly in three general sequential patterns, each of which forms a social action and achieves a particular interactional purpose. These patterns include (1) giving an account in a potentially face-threatening context[8] (the account-giving sequence), (2) foreshadowing a contrast in a conversational narrative (the contrast-projecting sequence), and (3) involving the audience in taking a stance (the involvement-constructing sequence).

As indicated in Table 2 below, over half of the *xiangshuo* instances (59.57%) occur in the account-giving pattern, a finding in agreement with what is reported in the literature on quoted thought in English (cf. Couper-Kuhlen, 2007). Meanwhile, the contrast-projecting pattern constitutes the second most frequent context in which *xiangshuo* is deployed, accounting for 35.11% of the occurrences. Finally, 9.57 percent of the *xiangshuo* tokens occur in the involvement-constructing pattern, all of which contain an overt second-person pronoun or function as an imperative.

Table 2. Distribution of sequential patterns

	N	%
Account-giving	56	59.57
Contrast-projecting	29	30.85
Involvement-constructing	9	9.57
Total	**94**	**100**

Notably, each sequence is also linked to a particular set of lexico-grammatical features that co-occur with *xiangshuo*, including the verb's subject, co-occurring adverbs, modals, and other particles. These devices are not used randomly, but work in tandem to achieve particular socio-interactional ends. In other words, the form of the *xiangshuo* construction is closely tied to, and thus can only be accounted for by, its function in the broader context. In what follows, we will discuss in detail each of the sequential patterns to show that only by adopting the approach of ICxG can the usage patterns of *xiangshuo* be fully understood.

---

8. As suggested in Waring (2007), accounts can be used to address both positive and negative faces of the conversation participants.

5.1   The account-giving sequence

The sequential environment where *xiangshuo* is most commonly found is the account-giving sequence. Investigating reported speech and thought in English non-narrative spoken discourse, Couper-Kuhlen (2007) pointed out that quoted thought is often produced to form accounts for some dispreferred or accountable action. Likewise, *xiangshuo* and the utterance that it prefaces are also often used as part of an account to mitigate a potential face-threatening act (Brown & Levinson, 1987).

As shown in Table 3 below, when used in this sequence, *xiangshuo* tends to co-occur with particular subjects and lexico-grammatical resources. Regarding the subject, this sequence exhibits the greatest diversity among the three. Several of the types of subject mentioned in Section 4 can be employed, such as the first-person singular (as in (5)), the first-person plural (as in (6)), zero marked (as in (7)) and abstract nouns (as in (4) and (8)). Recurrent verb modifiers include emphatic markers like *jiu* and *shi* (as in (5)) and the progressive aspect marker *zai* (as in (6)). Insofar as the sequence is intended to present an explanation, markers indicating a causal relation, such as *yinwei* 'because' and *suoyi* 'so' are also recurrently found in this context (as in (5)).

Table 3.   The account-giving sequence

Sequence	(subject) (*shi/zai*) *xiangshuo* [account/repair initiating]
Subject	First-person singular; first-person plural; zero-marked; abstract nouns
Verb Modifier	*jiu* 'just'; the copula *shi*; the durative aspect marker *zai*
Particle	*yinwei* 'because'; *suoyi* 'so'

For example, prior to (5) below, speaker D, a doctor appearing on a radio show, addresses the call-in listener and the rest of the audience and states that parents should try not to compare their children, since there are biological differences between sexes in terms of language development. M then denies in line 88 that she was comparing her son to his older sister and launches an account in line 89 with the causal marker *yinwei* and the *xiangshuo* construction.

(5)   *xiangshuo* with a first-person singular subject in the account-giving sequence
    088 M: *wo bushi   gen   ta   bijao.*
           1SG NEG.COP with 3SG compare
    089 →   *.. yinwei wo shi xiangshuo=._*
             because 1SG COP think

090   .. *ta=you   zhege   qianli,_*
      3SG have this.CL example

091   .. *wo--*
      1SG

092   .. *wo de yisi   shi zai= qidai.*
      1SG GEN meaning COP ASP   expect

'M: I was not comparing (him to his sister). It's just that I thought he had (his sister as) an example. I mean, I had expectations for him.'

As shown in (5), reporting one's thoughts with the preface *xiangshuo* constitutes a strategy to defend oneself. This may be due to the fact that one's thoughts are a private mental state that others can hardly have access to, let alone argue against, (Barnes & Moss, 2007) and that having good intentions is often thought to justify one's deeds. The marker *shi* that precedes *xiangshuo* further enables the speaker to emphasize the following content and to contrast the speaker's thoughts with D's suggestion. The same strategy is also realized in M's use of the abstract noun *yisi* 'idea, intention' in line 92 as part of her explanation (Hsieh, 2011), which presents another thought of hers.

On the other hand, in (6) below, the *xiangshuo* construction is used to respond to a request for information and explanation. In line 341, L, a radio journalist, asks Y, a restaurant owner, if there is anything else that Y would like to talk about, implying that he should repair and elaborate more on his previous turn. Y then produces a *xiangshuo* construction to introduce their future plans for the small business that they own. Notice that in line 343, the speaker uses the first-person plural subject and the progressive aspect marker *zai* along with *xiangshuo* to preface the speaker and his co-owner's plan. Y's utterance in line 343, if taken literally, cannot be true, because his co-owner is not present in the conversation and there is no way of knowing what the co-owner "is thinking". This choice of subject and aspect allows the speaker not only to hedge the following assertion by adding uncertainty (the progressive aspect) to the reported thought (cf. Van Bogaer, 2011) and distributing the responsibility (first-person plural instead of singular subject), but to present himself as the spokesperson (cf. Scheibman, 2004). In other words, the interactional meaning of the subject and the aspect makes them appropriate for use in this grammatical construction and sequential environment.

(6)   *xiangshuo* with a first-person plural subject in the account-giving sequence

341 L: .. *chule   ni   yishang jieshao   de @.*
        besides 2SG above   introduce NOM

342 Y: ... (1)^*qiwang.*
        expect

343 → …(1.2) *women zai xiangshuo,_*
          1PL   ASP think

344   *.. ruguoshuo,_*
          if

345   *.. yihou   nenggou jixu,_*
          afterwards can   continue

346   *.. women hui jixu   zuo.*
          1PL   will continue do

347   *.. ranhou ba ta   zuo de geng hao.*
          then   BA 3SG do CSC more good

'L: In addition to what you introduced earlier, (what else would you like to talk about?)

Y: (We) expect, we are thinking if we could make the business last, we will keep working hard and try to make it even better.'

Excerpt (7) below is another example of *xiangshuo* used in the environment of explanation. In this segment, F explains to S, a book saleswoman, why she does not want to buy the books that S recommends, which may constitute a face-threatening act (Brown & Levinson, 1987). Faced with F's outright rejection, S then uses the *xiangshuo* construction to reveal her thoughts and explain the reason why F should reconsider buying the books.

(7) *xiangshuo* with a zero-marked subject in the account-giving sequence

115 F: *.. meiyou   shijian kan a.*
              NEG.have time   read FP

116   *.. suoyi wo juede youdian xinteng   a.*
          so   1SG feel a.bit   distressed FP

117 S: → (0) *qishi   xiangshuo,_*
              actually think

118   *.. na   chulai gei ni   kan,*
          take out   give 2SG read

119   *.. ni   ziji fan._*
          2SG self turn

120   *… yinwei ta   na   limian de   dongxi,_*
          because 3SG that inside ASSC thing

121   *.. jiu shi xiaopian   de   duanluo.*
          just COP small.piece ASSC paragraph

122   *… jiu kan nimen zenme ziji qu yong.*
          just see 2PL   how   self go use

'F: (The thing is) I don't have time to read it. The book will be a waste.'

S: Actually (I) was thinking about taking out the book for you to browse. You could just take a look at it. What the book contains are short passages. You can decide how to use them yourself.'

As mentioned earlier, *xiangshuo* can take the abstract noun *yisi* 'meaning/intention' as the subject. We argue that this pattern is formatted to serve a particular interactional function, that is, to provide an account and tackle a potential face-threatening act as well. In (8) below, L and J are friends in the same church, and prior to this excerpt, they had been discussing the church's plan to raise funds for constructing a new church building. Holding an opposing stance, L presents his thoughts with the pattern *wo yisi shi xianghsuo la* 'what I am thinking is' in line 508 that maybe the real-estate industry will break down soon due to the robust development of mainland China, hinting at the possibility that the church might not have to spend so much money on the purchase of land if they wait patiently. The use of the *xiangshuo* construction here functions to mitigate the potential face-threatening force that the following content may bring about.

(8) *xiangshuo* with an abstract-noun subject in the account-giving sequence
    508 L: *wo yisi   shi xiangshuo la,_*
            1SG meaning COP think   FP
    509   *.. xiangshuo,_*
             think
    510   *...(tsk) dalu   zhege jushi,_*
             mainland.China this.CL situation
    511   *.. dui-bu-dui?/*
             right-NEG-right
    512   *.. dao shihou ni   zai   guo ji   nian,_*
             to   time   2SG again pass a.few year
    513   *.. gaobuhao zhengge   dou kua   le.*
             maybe   entire-CL all break.down ASP
    'L: I mean, take a look at the current development of mainland China, right? Maybe in a few years its economy will just completely crash.'

It should be noted that the use of *xiangshuo* with an abstract-noun subject is not found in every context, but is only occasioned in the account-giving sequence so that the speaker can manage the possible dispreferred or face-threatening act. In other words, the grammatical pattern is motivated not only by the conceptual meaning of the construction but also by the speaker's interactional needs in the unfolding conversation.

5.2   The contrast-projecting sequence

Another important pragmatic function of *xiangshuo* in conversation is to fore-shadow an imminent contrast in the ensuing discourse (cf. Kim, 2014; Kärkkäinen, 2012). This particular use of *xiangshuo* is usually found in a two-part sequence as the speaker constructs a narrative. As presented in Table 4 below, the first part of the sequence is a reported thought introduced by *xiangshuo*, which sets up the background and a reference point for the later contrast, and the second part reveals the real, and usually unanticipated, outcome of the story or event. Moreover, each part of the contrast-projecting sequence prefers particular adverbials. In the first part, where *xiangshuo* is used, the CTMP often is modified by a marker indicating a previous or initial state, such as *benlai* 'originally' (Tao, 2000), while in the second part, where the contrast is revealed, particles signaling an unexpected result such as *jieguo* 'it turned out' (Chang, 2015) are often deployed.

Table 4. The contrast-projecting sequence

Sequence	Part 1: [narrative] (subject) *xiangshuo* [reported thought]
	Part 2: [contrast in the story]
Subject:	first-person singular; third-person singular, zero-marked
Verb Modifier:	*jiu* 'just'; *benlai* 'originally'
Particle	*houlai* 'afterwards' *jieguo* 'it turns out'

When used in the contrast-projecting sequence, *xiangshuo* tends to co-occur with a first- or third-person singular subject. This pattern is distinct from what we reported in Section 5.1 but more akin to previous scholars' depiction of *xiangshuo* (Huang, 2003; Sanders & Uehara, 2006) and other cognitive verbs (Iwasaki, 1993; Scheibman, 2002). That is, the CTMP shows a strong preference for the first-person singular subject, while the third-person can also take the role of the subject when needed in a narrative. This may be due to the fact that in the contrast-projecting sequence, *xiangshuo* is used to literally report a previous thought in the subject's mind in order to contradict what has happened in reality.

To illustrate, in (9), A is talking about her personal experience encountering a strange man. As both the narrator and protagonist of the story, she uses a first-person pronoun *wo* as the subject of the reporting verb *xiangshuo* to introduce her first thought about the stranger striking up a conversation with her, which was that it didn't matter if he talked to her or not. Subsequent to this reported thought, A employs *jieguo* 'consequently' to launch a clause (line 184) that discloses the eventual outcome of the story, that is, what the stranger was talking

about got weirder and crazier in the end, which contrasts dramatically with her thought at first.

(9) *xiangshuo* with a first-person singular subject in the contrast-projecting sequence

180 A: *yinwei shi   ta,_*
        because COP 3SG

181   (0) *ta   xian zhao wo jianghua,_*
            3SG first find 1SG talk

182 → … *ranhou na   shihou wo xiangshuo,_*
             then   that time   1SG think

183   .. *ta gen wo jianghua ye   mei   shenme guanxi.*
           3SG with 1SG talk   also NEG.have what   relation

184   …(3) *jieguo   ta   yue   jiang yue   qiguai de   hua.*
              consequently 3SG further say   further strange ASSC word

'A: It was him who started a conversation with me. At that point, I was thinking it didn't matter that he wanted to talk to me, but then the more he talked, the weirder things got.'

Note that the thought reported in line 183 is, in fact, not a necessary part of the narrative. However, by presenting such an inner experience, the speaker "normalizes' the initial situation (Jefferson, 2004) and makes the final outcome even more stunning and remarkable, creating the "tellability" of the narrative (Norrick, 2004).

This pattern can also be deployed in telling a story from the third-person point of view. In (10), B, a guest on a radio show, is summarizing to the audience the plot of an English movie about a fraudulent lottery ticket claim and the ensuing investigation in a small Irish town. Prior to this episode, B has been recounting the opening of the story, i.e., the main characters of the movie find out that one of their neighbors had won the lottery and, upon discovering this, had been shocked to death. They then hesitate as to whether they can take and claim the ticket. To create the tension of the story, B first characterizes the protagonist as fearful (line 343) and then uses *xiangshuo* to bring out the character's change of mind. Interestingly, the reported thought in lines 345 and 346, similar to the case in (9), also includes the idiomatic expression *mei guanxi* 'does not matter,' which seems to downgrade the seriousness of the potential consequence. Immediately after the thought is reported, the speaker uses the counter-expectation marker *jieguo* to reveal that, in contrast to what the characters had thought, the lottery company was doubtful about the claim and started to investigate the case.

(10) *xiangshuo* with a third-person singular subject in the contrast-projecting
      sequence

343 B: .. *ta= yuanxian hen haipa,_*
              3SG at.first   very scared

344 →   ... *keshi houlai   xiangshuo,_*
               but   afterwards think

345   .. *suanle,_*
             whatever

346   .. *mei   guanxi jiu jiamao.*
             NEG.have relation just fake

347   .. *jieguo=,_*
             consequently

348   .. *zhege   caijuan gongsi   ku,_*
             this-CL lottery company

349   *de   ren,_*
          GEN person

350   .. *guoran   jiu lai- lai   diaocha.*
             as.expected just come come investigate

      'B: He was very afraid at first. But then he thought 'whatever, It'll be fine If I
      just pretend to be the winner'. However, turns out that the lottery company
      does indeed send people to investigate.'

Again, we see above that *xiangshuo* not only co-occurs with particular subjects
and reported content, but also works in a more global pattern along with other
discourse markers to serve particular pragmatic functions. This schema or exter-
nal syntax of *xiangshuo* seems to be in the process of conventionalization and
has become part of the grammatical knowledge that the speaker owns and imple-
ments in everyday talk.

5.3 The involvement-constructing sequence

Finally, *xiangshuo* can also be used to involve the listener in stance-taking. To
serve this function, *xiangshuo* usually co-occurs with a second-person subject,
which as noted in Section 4, is often cognitively implausible and thus grammat-
ically unacceptable in many languages (Aikhenvald, 2004; Iwasaki, 1993). How-
ever, it should be noted that the combination of *xiangshuo* and the second-person
singular pronoun is mostly deployed in the context of making generalizations
rather than referring specifically to the addressee. This should thus be considered
an example of the impersonal use of *ni* in Mandarin conversation (Biq, 1991) com-
parable to generalized *you* in English conversation (Scheibman, 2007; Stirling &

Manderson, 2011). Due to its stance-related function, *xiangshuo* is often preceded by a modal in this sequence.

Table 5. The involvement-constructing sequence

Sequence	A: [condition] subject *xiangshuo* [generalization/evaluation]
	B: acknowledgement/response tokens
Subject	Second-person singular
Verb Modifier	Modal verbs
Particle	Question markers

For example, in (11), F and C are discussing housing issues in Taiwan and the US. Although F is explaining how one would generally move to a better house as they grow older and become financially more stable, he makes use of the second-person singular pronoun as the subject of the *xiangshuo* construction. As argued by Stirling and Manderson (2011), the use of generalized second-person singular pronouns enables the speaker to construct objectivity and solicit empathy from the addressee. In response to F's generalization, C produces an acknowledgement marker *m* in line 584 to show her attentiveness and, possibly, her agreement.

(11)   *xiangshuo* with a second-person singular subject

579 F: … *ranhou dao le   dao le   ni… zhong=nian= le   yihou,*
            then   to ASP to ASP 2SG middle.age   ASP afterwards

580   .. *na   shihou ni   yijing,_*
            that time   2SG already

581   .. *jiu jixu   gao le   yihou,*
            just deposit high ASP afterwards

582 →   .. *ni   yiding   hui xiangshuo=,_*
            2SG definitely will think

583   .. *huan   yige   bijiao hao de   me.*
            change one.CL more good NOM FP

584 C: … *m.*
              INT

'F: And as you reach middle age and you have more money in your account, you would definitely want to move to a better one.
C: Mm.'

F's stance is further strengthened by the adverb *yiding* 'definitely' and the modal verb *hui* 'will' that precede *xiangshuo* in the same line. The tendency to co-occur with a modal element like *yiding* and *hui* is characteristic of the involvement-constructing sequence, but not observed in the other two sequential patterns.

This example illustrates how the choice of the subject and the verb modifier, while partly constrained by the semantics of the verb, is largely contingent on the interactional purpose that the speaker intends to achieve in the sequential environment. The combination of the subject and the CTMP in turn influences the shape of the ongoing conversation and gives rise to particular interactional consequences (cf. Fox, 2007; Heritage, 1984).

Likewise, in (12), A and B are discussing the topic of depression on a radio show. After mentioning a few potential symptoms of depression (lines 212 to 214), B deploys the second pronoun and *xiangshuo* in line 215 to involve the interviewer as well as the audience in taking a stance on the issue while making a generalization. Notice that in this example, B also uses a modal verb *hui* 'will' alone with *xiangshuo* to introduce the reported thought. In response, A produces a few laugh tokens at line 217. Based on the generalization, B moves on to reveal more about her thoughts concerning the topic of depression and the stance she has just constructed. This instance again demonstrates how the combination of the second pronoun and *xiangshuo* is produced in a particular linguistic and sequential context to achieve the speaker's purpose.

(12) 211 B: .. *ou= ..ni- ni=,_*
              INT 2SG 2SG

 212   ... *eh you yitian   shuibuzhao,_*
              INT have one.day sleep.NEG.attained

 213   .. *huozhe you liangtian,_*
              or   have two.day

 214   ... *eh.. chi de   hen duo,_*
              INT eat CSC very much

 215 →   .. *ni   jiu hui xiangshuo,_*
              2SG just will think

 216   .. *eh [wo shibushi   youyuzheng ou].*
              INT 1SG COP.NEG.COP depression   FP

 217 A: [@@@@@].\

 218 B: .. *[zhe qishi   you   keneng shi],_*
              this actually have possible COP

 219 A: *[unh unh]./*
              INT   INT

 220 B: .. *eh- qiqifufu,_*
              INT ups.and.downs

'B: Sometimes, you can't fall asleep or you eat too much. You may think that you're depressed.
A: ((laugh))

B: That's actually quite likely to be the case,
A: mm hm.
B: (just) ups and downs.'

It should be noted that the combination of the second-person subject and a cognitive verb does not always display the pattern or serve the function that we have just presented. As shown in Huang (2003), there is usually no modal element found between the second-person subject and the CTP *kan* 'see'. On the other hand, when CTMPs like *yiwei* '(mistakenly) think' co-occur with the second-person pronoun, the subject-verb fragment usually indicates a challenging or disagreeing stance rather than involvement. The pattern of *xiangshuo* should thus be treated as predicate- and construction-specific.

6.   Conclusion

In this article, we have closely examined the use of *xiangshuo* in Taiwan Mandarin conversation. We have found that despite its preference for the first-person singular, *xiangshuo* can also co-occur with other types of subjects, ranging from the second-person singular to abstract nouns, which can hardly be accounted for by the semantics of this construction. The use of the ICxG framework thus allows us to approach this issue from a broader perspective. We identify three sequential patterns in which the *xiangshuo* construction is recurrently deployed, including account-giving, contrast-projecting and involvement-constructing. The examples show that the sequence in which *xiangshuo* is employed is related not only to its interactional functions, but also to the recurrent local patterns, including subjects, aspects, modals and co-occurring adverbials, connectors and particles.

The current article therefore carries implications for both Cognitive Linguistics in general and the study of CTMPs in particular. It demonstrates how Cognitive Linguistics models such as CxG can be enhanced by the use of data from spoken corpora and methods developed in Interactional Linguistics (cf. Zima & Brône, 2015). It also reveals how the local collocates and cognitive meanings of a construction interact with global discourse and socio-interactional factors in shaping the use of the construction and its surrounding context. Finally, the analysis presented in this paper showcases how ICxG can shed light on the grammar of linguistic devices such as CTMPs and provide a more comprehensive explanation for such grammatical practices.

Abbreviations

1SG   first-person singular
1PL   first-person plural
2SG   second-person singular
2PL   second-person plural
3SG   third-person singular
ASP   aspect marker
ASSC associative
BA   the *ba* marker in the disposal construction
CL   classifier
COP   copular verb
CSC   complex stative construction
GEN   genitive marker
INT   interjection
NEG   negative marker
NOM   nominalizer
FP   final particle

Transcription conventions

[ ]   speech overlap
/   rising pitch
\   falling pitch
_   level pitch
...   pause
^   primary accent
--   truncated word
=   lengthening
Tsk   click
@   laughter

References

Aikhenvald, A.Y. (2004). *Evidentiality*. Oxford: Oxford University Press.

Antonopoulou, E., & Nikiforidou, K. (2011). Construction grammar and conventional discourse: A construction-based approach to discoursal incongruity. *Journal of Pragmatics*, 43(10), 2594–2609. https://doi.org/10.1016/j.pragma.2011.01.013

Barnes, R., & Moss, D. (2007). Communicating a feeling: The social organization of private thoughts. *Discourse Studies*, 9(2), 123–148. https://doi.org/10.1177/1461445607075339

Biq, Y.O. (1991). The multiple uses of the second person singular pronoun *ni* in conversational Mandarin. *Journal of Pragmatics*, 16(4), 307–321. https://doi.org/10.1016/0378-2166(91)90084-B

Biq, Y. O. (2004). Construction, reanalysis, and stance: 'V yi ge N'and variations in Mandarin Chinese. *Journal of Pragmatics*, 36(9), 1655–1672. https://doi.org/10.1016/j.pragma.2003.11.009

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511813085

Chang, M. H. (2015). Two counter-expectation markers in Chinese. In J. E. Díaz-Vera (Ed.), *Metaphor and metonymy across time and cultures* (pp. 141–170). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110335453.141

Couper-Kuhlen, E. (2007). Assessing and accounting. In E. Holt & R. Clift (Eds.), *Reporting talk: Reported speech in interaction* (pp. 81–119). Amsterdam: John Benjamins.

Croft, W. (2001). *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198299554.001.0001

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45–89). Hillsdale: Lawrence Erlbaum.

Fillmore, C. J. (1988). The mechanisms of "Construction Grammar". In *Proceedings of the Fourteen Annual Meeting of the Berkeley Linguistics Society* (pp. 35–55). Berkeley: Berkeley Linguistics Society. https://doi.org/10.3765/bls.v14i0.1794

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language*, 64(3), 501–538. https://doi.org/10.2307/414531

Fox, B. A. (2007). Principles shaping grammatical practices: An exploration. *Discourse Studies*, 9(3), 299–318. https://doi.org/10.1177/1461445607076201

Fried, M., & Östman, J. O. (2005). Construction Grammar and spoken language: The case of pragmatic particles. *Journal of Pragmatics*, 37(11), 1752–1778. https://doi.org/10.1016/j.pragma.2005.03.013

Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.

Hakulinen, A., & Selting, M. (Eds.). (2005). *Syntax and lexis in conversation: Studies on the use of linguistic resources*. Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.17

Heritage, J. (1984). *Garfinkel and ethnomethodology*. New York: Polity Press.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165525

Hsieh, C. Y. C. (2011). Meaning in interaction: The use of abstract nouns and the co-construction of stance in conversation. Paper presented at the ELSJ International Spring Forum 2011, Shizuoka University, Hamamatsu City.

Hsieh, F., & Huang, S. (2005). Grammar, construction, and social action: A study of the *qíshí* construction. *Language and Linguistics*, 6(4), 599–634.

Huang, S. (2003). Doubts about complementation: A functionalist analysis. *Language and Linguistics*, 4(2), 429–455.

Imo, W. (2015). Interactional construction grammar. *Linguistics Vanguard*, 1(1), 69–77. https://doi.org/10.1515/lingvan-2015-0008

Iwasaki, S. (1993). *Subjectivity in grammar and discourse: Theoretical considerations and a case study of Japanese spoken discourse*. Amsterdam: John Benjamins.

Jefferson, G. (2004). "At first I thought": A normalizing device for extraordinary events. In G. H. Learner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 131–167). Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.125.09jef

Kim, M. S. (2014). Reported thought as a stance-taking device in Korean conversation. *Discourse Processes*, 51(3), 230–263. https://doi.org/10.1080/0163853X.2013.862479

Kärkkäinen, E. (2012). I thought it was very interesting. Conversational formats for taking a stance. *Journal of Pragmatics*, 44(15), 2194–2210. https://doi.org/10.1016/j.pragma.2012.09.005

Li, N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar.* Berkeley: University of California Press.

Lim, N. E. (2011). From subjectivity to intersubjectivity: Epistemic marker *wo juede* in Chinese. In Y. Xiao, L. Tao, & H. L. Soh (Eds.), *Current issues in Chinese linguistics* (pp. 265–300). Newcastle: Cambridge Scholars.

Linell, P. (2009). Grammatical constructions in dialogue. In A. Bergs & G. Diewald (Eds.), *Contexts and constructions* (pp. 97–110). Amsterdam: John Benjamins. https://doi.org/10.1075/cal.9.05lin

Norrick, N. R. (2004). Humor, tellability, and conarration in conversational storytelling. *Text*, 24(1), 79–112. https://doi.org/10.1515/text.2004.005

Sanders, R., & Uehara, S. (2006). Thinking about *xiang* 3 in Taiwan: Some native-speaker opinions. Paper presented at the 17th North American Conference on Chinese Linguistics.

Scheibman, J. (2002). *Point of view and grammar: Structural patterns of subjectivity in American English conversation.* Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.11

Scheibman, J. (2004). Inclusive and exclusive patterning of the English first person plural: Evidence from conversation. In M. Achard & S. Kemmer (Eds.), *Language, culture and mind* (pp. 377–396). Stanford: CSLI Publications.

Scheibman, J. (2007). Subjective and intersubjective usees of generalizations in Engilsh conversations. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 111–138). Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.164.06sch

Selting, M., & Couper-Kuhlen, E. (2001). *Studies in interactional linguistics*. Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.10

Stirling, L., & Manderson, L. (2011). About you: Empathy, objectivity and authority. *Journal of Pragmatics*, 43(6), 1581–1602. https://doi.org/10.1016/j.pragma.2010.12.002

Tao, H. (2000). Adverbs of absolute time and assertiveness in vernacular Chinese: A corpus-based study. *Journal of Chinese Language Teachers Association*, 35(2), 53–74.

Van Bogaer, J. (2011). *I think* and other complement-taking mental predicates: A case of and for constructional grammaticalization. *Linguistics*, 49(2), 295–332. https://doi.org/10.1515/ling.2011.009

Waring, H. Z. (2007). The multi-functionality of accounts in advice giving. *Journal of Sociolinguistics*, 11(3), 367–391. https://doi.org/10.1111/j.1467-9841.2007.00328.x

Wide, C. (2009). Interactional Construction Grammar: Contextual features of determination in dialectal Swedish. In A. Bergs & G. Diewald (Eds.), *Contexts and constructions* (pp. 111–142). Amsterdam: John Benjamins. https://doi.org/10.1075/cal.9.06wid

Zima, E., & Brône, G. (2015). Cognitive Linguistics and interactional discourse: Time to enter into dialogue. *Language and Cognition*, 7(4), 485–498. https://doi.org/10.1017/langcog.2015.19

Index

The studies in the present volume illustrate the current state-of-the-art in the corpus-based approach in cognitive linguistics, which seeks to motivate linguistic phenomena through the combination of quantitative and qualitative analysis. By focusing on language use in different contexts from a variety of perspectives, each of the contributions in this volume presents its own unique take on the intertwined relationship between language, thought, and communication. Thus, each article shows how a combination of quantitative and qualitative analytical techniques helps shed new light on old issues, reflecting the usage-based nature of cognitive linguistics and illustrating the explanatory adequacy of corpus-based methods. Originally published as special issue of *Review of Cognitive Linguistics* 17:1 (2019).

John Benjamins Publishing Company