SPECIAL ISSUE ARTICLES

Deep dives into big data

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

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

The six articles in this special issue 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,

https://doi.org/10.1075/rcl.00023.jan
Review of Cognitive Linguistics 17:1 (2019), pp. 1–6. issn 1877-9751 | e-issn 1877-976x

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 issue). 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 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 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 special issue 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 the 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 attest 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 the differences in the 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 identifies 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 issue 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: Constructicon development across languages*. Amsterdam: John Benjamins.

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

Address for correspondence

Wei-lun Lu
Language Center, Faculty of Medicine Division
Masaryk University
Kamenice 5 (Building A15)
62500 Brno
Czech Republic
wllu@med.muni.cz

Biographical notes

Laura A. Janda (PhD 1984 UCLA) is currently Professor of Russian Linguistics at UiT the Arctic University of Norway, where she directs the CLEAR (Cognitive Linguistics: Empirical Approaches to Russian) research group and has won awards for both teaching and research. She is a past president of the International Cognitive Linguistics Association and serves on the boards of numerous scholarly journals in linguistics and Slavic studies. Her research focuses primarily on the semantics of Russian grammatical categories, particularly aspect and case, although she has also published on topics involving other Slavic languages as well as Spanish and North Saami. Janda is engaged in developing research-based language teaching methods and digital materials, including the Russian Constructicon (https://spraakbanken.gu.se/karp/#?mode=konstruktikon-rus) and the Strategic Mastery of Russian Tool (https://uit-no.github.io/smartool/).

Naděžda Kudrnáčová is Associate Professor at the Department of English and American Studies, Faculty of Arts, Masaryk University, Brno. Her research interests lie mainly within the fields of cognitive semantics, lexicology and the interface between syntax and semantics.

Wei-lun Lu is Assistant Professor at the Language Center (Medical Division) of Masaryk University in Brno, Czech Republic. His research expertise is corpus-based cultural and cognitive linguistics, with a research focus on metaphor, viewpoint and grammar.