Check for updates

Editorial: The evolution of meaning: challenges in quantitative lexical typology

Gerd Carling[1]* and Annemarie Verkerk[2]

[1]Institute of Empirical Linguistics, Faculty of Linguistics and Cultural Studies, Goethe University Frankfurt, Frankfurt, Germany, [2]Department of Language Science and Technology, Faculty of Philosophy, Saarland University, Saarbrücken, Germany

Editorial on the Research Topic
The evolution of meaning: challenges in quantitative lexical typology

The Research Topic "*The evolution of meaning: challenges in quantitative lexical typology*" deals with a field of increasing interest in current linguistics: quantitative studies on the evolution of word meaning. This is part of a growing trend in linguistics in general, where a major shift toward using quantitative methods has taken place, as well as increased interest in semantic change. Outside linguistics, we may connect this quantitative turn with growing interest from the general community in areas such as large-scale corpora and artificial intelligence as well as increasing awareness of how language use matters in daily life. The current issue focuses on lexical semantics from a cross-linguistic and typological perspective, with a focus on semantic fields, polysemy, and colexification. The overarching theme is how to quantify change across disciplines, where indeed the included papers are dealing with large data sets or cross-linguistic comparison.

The Research Topic comprises eight papers of various types and content, including "Conceptual analysis," "Methods," as well as "Original research" papers. An article of the first type is by Haspelmath (published 24 October 2023), entitled "*Coexpression and synexpression patterns across languages: comparative concepts and possible explanations.*" Here, the basic concepts underlying colexification are defined and developed into further distinctions. The author widens the concept into *coexpression* and defines how this term can be used to cover both lexical and grammatical patterns, when two meanings for a form occur in different contexts. Additionally, the term *synexpression* is introduced, a term that covers the simultaneous presence of two meanings in a single form. These concepts are further described and developed, using a rich variety of examples from various languages. The terms are also compared to earlier terminology within the area of lexical typology. Most importantly, the paper indicates how the terminological distinctions can be used for defining comparative concepts that can be contrasted in typological studies. This makes the article highly useful for any typologically oriented, large-scale study on lexical or grammatical coexpression and an important theoretical contribution to the Research Topic.

A "Methods" paper has been contributed by List (16 June 2023), by the title of "*Inference of partial colexifications from multilingual wordlists.*" This article deals with a problem connected to the theoretical issue of co- and synexpression described by

Haspelmath, namely how to deal with partial (or loose) colexifications (i.e., colexifications based on parts of lexemes from compounds) in large-scale studies of colexification patterns. The article provides a hands-on description of how partial colexifications can be modeled by identifying segmentation in word-lists. This would be done by identifying parts of words as identical with other (independent) words. The papers suggests a model for the implementation and the visualization and analysis of results.

Remaining papers in the Research Topic are all of the type "Original Research Article." Two main types can be identified, either papers working with textual corpora, or papers dealing with word lists or dictionary-based databases. We start with the first type.

The study by Haug and Pedrazzini (8 December 2023), "*The semantic map of when and its typological parallels,*" uses a parallel corpus of New Testament translations to explore the semantic map of the temporal connective WHEN. Using 1,444 languages in Mayer and Cysouw's (2014) parallel corpus of Bible translations, the authors investigate WHEN and its aligned tokens to investigate whether there is a cross-linguistic distinction between *universal* WHEN and *existential* WHEN (as in German *als* and *wenn*). The study shows a cross-linguistically clear preference for a clustering of universal WHEN, but weaker evidence for other clusters.

Another corpus-based study, by Levshina et al. (28 March 2024), "*Revered and reviled: a sentiment analysis of female and male referents in three languages,*" focuses on semantic prosody and connotation in a study of pejorative nouns referring to women in relation to terms used to designate men. The study uses corpora from English, Chinese and Russian (of different stylistic types and historical periods), and applies two methods, sentiment scores from PyABSA and OpenAI's large language model GTP-3.5, to study the differentiation between words for men and women (e.g., woman—man, girl—boy). The results reveal that male referents are referred to more neutrally, and that the gender gap increases with time. However, the results have only support in the results from the sentiment score model, not when using GTP-3.5 models.

The paper "*Are categories' cores more isomorphic than their peripheries?*" by Cai and De Smet (20 June 2024) revises isomorphism, i.e., the principle that implies that a single meaning is expressed by a single form. The authors test the hypothesis that forms can code a network of related meanings that is organized in terms of prototypicality, where meanings basically monopolize the core but are more likely to come into competition with other forms in peripheral meanings. To this end, the paper looks at the variation and use of AT and WITH in the Spanish original and English translations of Cervantes' *Don Quiote* over time. Through the use of semantic maps for the different prepositions in the two languages, the results confirm that the level of variation is large and the patterns highly divergent, which supports the hypothesis that isomorphism is not a general evolutionary "attractor" in language change.

The remaining three papers deal in one way or the other with word lists or lexical databases. The paper by Fugikawa et al. (4 April 2023), "*A computational analysis of crosslinguistic regularity in semantic change,*" studies the regularity in directionality and source-target mapping by means of the database *DatSemShift*, which records thousands of meaning changes in hundreds of languages. By testing the directionality of historical semantic change between pairs of meanings, the study tests a number of hypothesized mappings, e.g., more concrete → less concrete, less valenced → more valenced, or more frequent → less frequent. The results reveal significant regularity in semantic change that is also cross-linguistically valid.

The study by Carling et al. (19 May 2023) entitled "*The evolution of lexical semantics dynamics, directionality, and drift,*" investigates a similar problem: the support for directionality in semantic change. The paper uses a phylogenetic model to reconstruct the probability of presence of colexifying meanings at ancestral nodes. By means of this model, the paper investigates the frequency of occurrence of different semantic relations, where metonomy is most frequent, followed by metaphor, specialization and generalization. Further, the study concludes that the semantic rates of change for different concepts and semantic classes are highly variable; for example, animate nouns have lower rates of change than inanimate nouns.

Last but not least, the paper "*Causal inference of diachronic semantic maps from cross-linguistic synchronic polysemy data*" by Dellert (11 January 2024) deals with how semantic maps can be approached from an algorithmic angle. The model uses a standard causal inference algorithm to reduce cross-linguistic polysemy into minimal network structures, which can explain observed polysemies.

All in all, the eight papers of the Research Topic on challenges in quantitative lexical typology give an overview of the current direction of progress within the field. By means of large-scale, quantitative methods, based either on corpora, word lists or lexical databases, we can achieve new insights into the mechanisms of lexical semantic change. The Research Topic also highlights that progress is rooted in improvements on all levels in the research procedure: at the level of theoretical innovation, where terms, distinctions, and definitions have to be reconsidered and re-established to confront new empirical and quantitative methods; at the level of methodology, where the implementation of new measurements and the use of computational and quantitative methods has to be improved to approximate the research problems more closely; and at the level of studying individual research problems, based on different types of data, both textual as well as word lists.

Author contributions

GC: Writing – original draft, Writing – review & editing. AV: Writing – review & editing, Writing – original draft.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Mayer, T., and Cysouw, M. (2014). "Creating a massively parallel Bible corpus," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* [Reykjavik: European Language Resources Association (ELRA)], 3158–3163.