Lexical evolution, cognition, and computation

Yang Xu (yangxu@cs.toronto.edu)
Department of Computer Science & Cognitive Science Program
University College, University of Toronto
Barbara C. Malt (barbara.malt@lehigh.edu)
Department of Psychology, Lehigh University
Mahesh Srinivasan (srinivasan@berkeley.edu)
Department of Psychology, University of California, Berkeley

Keywords: lexicon; semantic systems; polysemy; language evolution; cognitive modelling; language phylogeny

Human languages evolve over time, manifested in changes in phonology, morphology, syntax, and word meaning. Compared to other components of language where there are rich formal characterizations of change, evolution of word meaning remains poorly understood. How do semantic systems in specific domains such as color or number evolve over time? Further, are there overarching principles that emerge across domains and that underlie the evolution of the lexicon at large? And can we advance the characterization and evaluation of these principles by exploiting emerging resources, e.g., those from the digital humanities, using formal methodologies, e.g., those from the computational disciplines? The goal of this symposium is to address these questions.

We explore these questions by bringing together researchers from cognitive and developmental psychology, linguistic semantics and evolutionary linguistics, cultural anthropology, and computer science. The presentations will draw on state-of-the-art progress and perspectives on empirical and computational approaches to lexical evolution, with an emphasis on the evolution of word meanings. They consider influences on lexical evolution ranging from human perceptual and cognitive capacities to cultural needs, communicative pressures, and the fact that an evolving system at time t must derive from what was present at time $t-1$ (e.g., Evans & Levinson, 2009; Malt & Majid, 2013). The talks will stimulate discussion of future opportunities and challenges and promote cross-disciplinary collaborations at the intersection of lexical evolution, cognition, and computation.

The symposium will address these issues at two different scales. The first set of talks will examine how individual semantic systems have evolved over time. Two of these will focus on semantic domains of perceptual experience: smell (Asifa Majid) and color (Susanne Vejdemo). The third will address number, a domain that contributes uniquely to human culture and is driven less by sensory input and more by cultural needs (Andrea Bender and Sieghard Beller). The second set of talks examines evolution of the lexicon more broadly. One presentation will take a computational approach to understanding how recurring patterns of polysemy arise across languages (Yang Xu, Barbara Malt, and Mahesh Srinivasan). The other will use computational phylogenetic approaches to illuminate semantic change (Michael Dunn).

Evolution of semantic systems

1. The smell lexicon (Asifa Majid, Radboud University)

It has long been presupposed that language just cannot express our olfactory experiences. Plato himself said: "the varieties of smell have no name, ... but they are distinguished only as painful and pleasant". Recent crosscultural investigation shows this generalization is not true. For example, the hunter-gatherer Jahai who reside in the tropical rainforests of the Malay Peninsula have around a dozen terms to refer to different types of odor qualities, and are able to talk about smells with the same ease as colors (Majid & Burenhult, 2014). The Jahai are not the only ones to have an elaborate lexicon for smell. This observation raises new questions about what factors might give rise to olfactory language. In this talk, I review recent work that shows that subsistence (e.g., being a hunter-gatherer; Majid & Kruspe, 2018), as well as ecology (O'Meara, Smythe Kung, & Majid, under review) shape olfactory language, and its use. Together this work points to the importance of marrying studies of the lexicon to those of culture and use.

2. Color words (Susanne Vejdemo, University of Stockholm)

The proposal (Berlin & Kay, 1969) that there is a largely predictable evolutionary sequence according to which languages partition and name color space has engendered fruitful and contentious discussion for half a century. What motivates the universal tendencies in color naming? Computational analysis of large amount of experiment data from hundreds of languages have shown that the naming and categorization of color space occurs according to optimum perceptual partitions (e.g., Regier, Kay, & Khetarpal 2007). Psychophysical constraints are thus one factor in why this semantic system often develops the way it does in languages, but far from the only explanatory factor. Large computational studies are best married to smaller, detail-oriented studies that investigate how the change actually proceeds in the system, and how lexical competition works in small time-scale perspectives of a few generations. This reveals additional constraints and tendencies in the domain. I will discuss experimental results from seven Germanic languages (Vejdemo et al., 2015) and from two generations of Swedish speakers involved in an ongoing lexical and semantic replacement process in the color

domain (Vejdemo, 2017), as well as new data from an elicitation study on the stabilization of color word vocabulary in children and young adults.

3. Numeral systems (Andrea Bender & Sieghard Beller, University of Bergen)

Counting is central to human cognition, based on mathematical principles that abstract regularities from concrete instances. Still, the way in which these are expressed is culturally variable and changes over time. A case in point are the numeral systems that are part of almost every language. While the present-day systems within a family of related languages typically originate from a common ancestor (i.e., the number system in the ancestor language), they exhibit striking diversity. Based on an overview on what is already known about the evolution of numeral systems and which questions have remained open, in this talk, we intend to retrace how present-day peculiarities may have emerged and evolved out of single systems by highlighting the mathematical principles behind the changes, their cognitive underpinnings, and the cultural conditions that may have triggered them.

Evolution of the lexicon

4. Polysemy (Yang Xu, University of Toronto; Barbara C. Malt, Lehigh University; Mahesh Srinivasan, University of California, Berkeley)

Human language relies on a finite lexicon to express a potentially infinite set of ideas. A key result of this tension is that words become polysemous over time: A single word can be extended to express multiple different senses, e.g., *face* may refer to "body part", "expression", or "surface of an object". Certain patterns of polysemy tend to recur across languages; that is, the same set of senses is labeled by a single word form, despite variations in language genealogy, geography, climate, and culture. We ask how it is that, over time, languages come to share many patterns of co-lexification. Building on recent work on the evolution of polysemy (Ramiro, Srinivasan, Malt, & Xu, in press), we examine the perspective that the cross-linguistic frequency distribution of shared polysemy reflects a historical drive toward efficiency in communicating and learning meanings. We test our hypothesis computationally using a large database of digitized lexicons from the world's languages. Results indicate that semantic associativity predicts the frequency with which senses are colexified across languages, and it does so better than several alternative variables. This outcome is consistent with the view that cross-linguistically recurring patterns of co-lexification arise from a historical process of word sense extension that tends to minimize cognitive effort.

5. Language phylogenies (Michael Dunn, Uppsala University)

The study of semantic change has focused on relatively constrained areas of the lexicon, and it is rare to find investigations of processes of semantic change broader than change in the meaning of individual terms. Larger scale studies of organization of semantic domains and the lexicon in turn tend to neglect the temporal dimension. But the quantitative phylogenetic analysis of language has come of age, and there is now available a well-equipped toolbox of so-called 'phylogenetic comparative methods', a family of quantitative methods which use evolutionary models of historical relatedness to interrogate processes of change in structure. While these methods were developed with biological questions in mind they have been used in anthropology and linguistics as well. Phylogenetic comparative methods have been developed for many purposes, including the reconstruction of ancestral states, the detection of functional dependencies, and for testing hypotheses about evolutionary pathways, amongst others. I will show how the combination of the phylogenetic approach with semantic typology promises make tractable the study of large scale change in semantic systems, and shed new light on questions of universals, directionality of change, and the functional motivation of lexical-semantic organization. A proof of concept will be shown, demonstrating how cognitive biases in semantic representation can be shown to have a cumulative effect on the structure of the entire Indo-European language family.

References

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. University of California Press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–448.

Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition*, *130*, 266–270.

Majid, A., & Kruspe, N. (2018). Hunter-gatherer olfaction is special. *Current Biology*.

Malt, B. C., & Majid, A. (2013). How thought is mapped into words. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*, 583–597.

O'Meara, C., Smythe Kung, S., & Majid, A. (under review). Olfaction, ideophones, and sound symbolism in huehuetla tepehua within the larger totonac-tepehua context.

Ramiro, C., Srinivasan, M., Malt, B. C., & Xu, Y. (in press). Algorithms in the historical emergence of word senses. *Proceedings of the National Academy of Sciences*.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*, 1436–1441.

Vejdemo, S. (2017). *Triangulating perspectives on lexical replacement: From predictive statistical models to descriptive color linguistics*. Unpublished doctoral dissertation, Department of Linguistics, Stockholm University.

Vejdemo, S., Levisen, C., van Scherpenberg, C., Næss, Å., Zimmermann, M., Stockall, L., ... others (2015). Two kinds of pink: Development and difference in germanic colour semantics. *Language Sciences*, *49*, 19–34.