

Semantic typology as an approach to mapping the nature-nurture divide in cognition

Jürgen Bohnemeyer, Department of Linguistics, University at Buffalo

jb77@buffalo.edu

www.acsu.buffalo.edu/~jb77

*Distributed under the Creative Commons Attribution Non-Commercial Share Alike license:
<http://creativecommons.org/licenses/by-nc-sa/3.0/> cc by-nc-sa 2010 Juergen Bohnemeyer*

Abstract – One of the central projects of the cognitive sciences is the determination of which aspects of cognition are biologically determined – directly or mediated by neurophysiology – and species-specific and which aspects are culture-specific and learned. Semantic typology, the crosslinguistic study of semantic categorization in natural languages, plays a key role in this project. The task of mapping the nature-nurture divide in semantic categorization can be compared to the task of mapping early human migrations to their reflexes in the contemporary gene pool. Progress in semantic typology has been hampered by a longstanding bias in the cognitive sciences in favor of postulates of underlying uniformity and innateness, but also by the inherently collaborative nature of semantic typology and the multi-faceted training it requires. Funding institutions such as the National Science Foundation can and should play a key part in correcting this situation.

1. A challenge question for the cognitive sciences: Mapping the nature-nurture divide in cognition – There is no more foundational question in the cognitive sciences than the question of the nature-nurture divide in cognition: which aspects of human cognition are learned, the sum of individual experience and cultural transmission, and which are innate, species-specific and perhaps shared with other animals? The cognitive paradigm itself, at least in its present form, requires us to assume that the development of the individual mind does not start from a blank slate, but that evolution has predisposed us to process the incoming perceptual data stream by sorting it according to categories and representational formats inscribed directly or indirectly, in ways we do not yet understand, in our genetic code.

At the same time, human cognition is far more social and cultural than any known form of animal cognition. A large portion of what we know or assume – from swimming, and perhaps walking, to the preparation of food, the making of clothes, the constructions of houses, to abstract concepts of science, politics, religion, and art – we do so neither on the basis of purely personal experience nor instinctively, because of something coded in our genes, but because this knowledge is part of our culture. Human populations have adapted to most environments of the Earth's land mass and in the process developed local conceptual categories for the different land and water forms and the different species of plants and animals they encountered. They developed different modes of subsistence, different technologies, different structures of household and social organization, different medical practices, and different belief systems and art forms. All of this became part of a vast reservoir of cultural knowledge transferred from generation to generation within each human group.

2. A scientific strategy in mapping the nature-nurture divide in cognition: Semantic typology

The question of where the province of biology in cognition ends and that of culture begins has been driving research in what we now consider the cognitive sciences since the beginnings of

modern psychology and anthropology in the late 19th century. It is a purely empirical question. Answers can only be obtained by bringing together genetic, neurophysiologic, and developmental research on individual populations with comparative research into the cognitive representation of reality across human populations. Language is a key variable in this endeavor. Absent telepathy, intergenerational transfer of cultural knowledge must rely on observable behavior, and the world's surviving 6,000 -7,000 spoken or signed natural human languages are by far the most powerful and complex systems of external, social representations that have evolved on this planet. Moreover, the internal structure of linguistic representations submits to analysis in terms of discrete categories and combinatory syntactic rules in a way the structure of thought does not.

Semantic typology is the subfield of linguistics in charge of exploring the language variable in crosscultural cognition. A subfield of linguistic typology, it is dedicated to the comparative study of semantic categories. Typology proceeds by generalizing over samples of languages which are as large as possible and as broadly varied in terms of language families, geographic areas, and cultures as possible. Semantic typology applies this research strategy to languages viewed as representational systems. Semantic typologists sort out the universal from the language-specific in how languages of a broadly varied sample represent and categorize reality.

A methodological canon for semantic typology was first explicitly stated in the 1990s by the members of what is now the Language and Cognition group at the Max Planck Institute for Psycholinguistics. This method employs non-verbal stimuli such as pictures, videos, and toys to represent the conceptual distinctions of interest. Semantic categorizations – preferred descriptions and ranges of possible descriptions – of these stimuli are collected in samples of unrelated and structurally broadly diverse languages by administering a standardized protocol to sufficiently large populations of speakers of each language. Early precursors of this method were questionnaire studies dating back as far as the 19th century. Modern pre-Max-Planck-Institute studies include the World Color Survey conducted in the 1970s. Uniform patterns in the resulting data are attributed to species-specific properties of cognition, which in turn may be interpreted as directly or indirectly – mediated by neurophysiology – biologically grounded. The underlying assumption here is that there is no genetic variation in human populations that affects cognition – so far, none has been attested. Consequently, crosslinguistic variation in a particular property of semantic categorization is interpreted as evidence that the property in question is culture-specific and learned.

In the past two decades, pioneering research in semantic typology has shattered assumptions about the uniformity of human cognition. To cite just two examples: Pederson *et al.* 1998 demonstrate an astonishing amount of variation across populations in the use of spatial reference frames – coordinate systems for the computation of spatial relations – in both language and recall memory. It had been taken for granted that the use of reference frames is not subject to cultural variation until evidence emerged in the late 1970s from Australian Aboriginal communities who use predominately absolute frames (“The saucer is east of the cup”). Pederson *et al.*'s landmark study triggered a renaissance of research on the Linguistic Relativity Hypothesis, according to which language use may influence nonlinguistic cognition. Bohnemeyer *et al.* 2007 show unexpected variation in what can be represented as a simple motion event in different languages. In many of the world's languages, motion from one point to another is obligatorily represented as a sequence of a departure event and an arrival event,

with the path traversed in between left to inference. This study also found evidence suggesting that the greatest degree of crosslinguistic uniformity may be neither in syntax nor in semantics, but rather in the mapping between the two, suggesting important implications for language evolution and language acquisition.

3. The current state of the field – The study of semantic categorization across natural languages can be compared to the ongoing efforts to relate the miniscule genetic differences across human populations to early human migratory movements. Both undertakings require a global network of collaborating institutions and an army of field researchers – in the case of the Human Genome Diversity Project, researchers taking genetic samples from human groups around the world; in the case of semantic typology, researchers collecting data on semantic categorization from speakers of languages around the world. However, a global network of institutions and scholars collaborating on research in semantic typology does not currently exist in more than an embryonic state. Despite the groundbreaking achievements mentioned above, the advances of population geneticists and paleoanthropologists in the same timeframe far exceed those made by semantic typologists. The largest studies in semantic typology ever conducted, the World Color Survey and the Mesoamerican Color Survey, involved the collection of data from 112 and 116 languages, respectively – less than 2% of the world’s languages. More recent studies – such as the landmark studies mentioned above – have been conducted on small samples of less than three dozen languages. Such samples support existential quantification over the languages of the world (“There are languages that...”), but not universal quantification (“All languages ...”). The vast majority of lexical meanings and meanings of inflections and function words have not been targeted by typological work at all to date, and there have been no typological studies at all in the areas of meaning composition (or formal semantics) and utterance meaning (or pragmatics).

Several factors have slowed down the crosslinguistic exploration of semantic categorization. The dominant rationalist paradigm in linguistics and the cognitive sciences has strongly favored the pursuit of universalist, innatist accounts for the past half century. The great majority of linguists and cognitive scientists work only with speakers of English and a few other European languages. Since the 1990s, the field has been undergoing an ever accelerating empiricist turn. Nevertheless, the long-standing bias in favor of uniformity has reduced interest in language diversity for so long that we today lack even the most minimal scientific records for more than half of the world’s languages – at a moment in time at which more than half of the world’s languages have fewer than 10,000 speakers and the UNESCO Atlas of the World’s Languages in Danger identifies 2,471 languages as endangered or extinct.

A second, equally long-term bias in linguistics in favor of syntactic and phonological research over work in semantics has caused advances in semantic typology to lag well behind advances in other areas of typology. In syntactic typology, it has become standard practice to derive quantitative generalizations from databases containing information on hundreds of languages. For semantic typology, such databases are virtually unavailable to date. The World Atlas of Language Structures (WALS, Haspelmath *et al.* 2008), the largest typological database every compiled, provides data on 2,650 languages. Of the 142 features mapped in WALS, only eight deal with issues of lexical semantics, with half of these providing information on basic color terms drawn from the World Color Survey. The larger WALS maps of syntactic features contain

data from more than a thousand languages. The creation of such large databases has benefited from available language descriptions focusing mostly on syntactic and phonological phenomena and neglecting semantics. Bringing semantic typology up to speed with syntactic typology, producing WALS-style databases on semantic categorization, therefore requires the collection of primary data in field work, in the communities in which the target languages are spoken.

Data collection for semantic typology in the field faces two further challenges: it requires training in typology, field linguistics, semantics, and experimental methods - a demanding and to date quite rare profile. Plus, semantic field work can only be carried out by experts on the field languages, and there is a natural limit to the number of languages any researcher can have an expertise in at a sufficient level. Consequently, semantic typology is generally a collaborative effort by a group of researchers. And the possibility of such collaborative research is presently severely limited due to lack of institutional support.

4. Semantic typology and NSF – The environment in which the goal of mapping the divide between biology and culture in semantic categorization can be met does not yet exist. There is no global network of collaborating institutions. The only institution that currently supports sustained research in semantic typology on a world-wide scale – albeit so far still only in the form of studies on fewer than three dozen languages each – is the Max Planck Institute for Psycholinguistics. The National Science Foundation can and should be a core part of the response to this situation. NSF can fund research in semantic typology both through awards to individual researchers and through the PIRE and SLC programs. A historic example of funding large-scale research in semantic typology through an award to individual researchers is the World Color Survey. Currently, the project *Spatial language and cognition in Mesoamerica* (MesoSpace; Award #BCS-0723694) demonstrates the feasibility of this funding model in semantic typology. MesoSpace is funded through an award to the author and has brought together 18 field researchers from nine institutions in Mexico and the U.S. for a comparative study of the representation of space in 15 indigenous languages of Mexico and Central America. A follow-up proposal extending the project to languages and populations outside Mesoamerica is under review. MesoSpace also shows the potential of awards of this type for training students in the skill set required for semantic typology. 11 of the members were graduate students at the start of the project. Implementing and carrying out projects at a scale that goes beyond that of MesoSpace requires the kind of international multi-institution collaborations the PIRE program is ideally suited to funding. Within the U.S., SLC awards could be used to build the institutional collaborations needed for semantic typology.

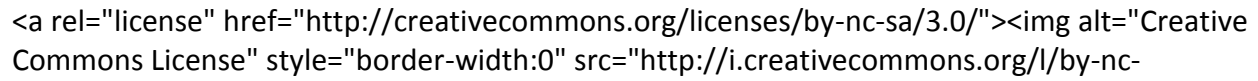
5. Summary – Determining the boundary between biology and culture in cognition is one of the central projects of the cognitive sciences. Cataloging the semantic categories of the world's languages is an indispensable part of this project. NSF is in a unique position to make large-scale research in semantic typology possible.

References

Bohnenmeyer, J., Enfield, N. J., Essegbey, J., Ibarretxe, I., Kita, S., & F. K. Ameka (2007). Principles of event segmentation in language: The case of motion events. *Language* 83(3): 495-532.

Haspelmath, M., Dryer, M. S., Gil, D., & B. Comrie (eds.) (2008). The World Atlas of Language Structures Online. Munich: Max Planck Digital Library. Available online at <http://wals.info/index>. Accessed on 9/29/2010.

Pederson, E., Danziger, E., Wilkins, D., Levinson, S., S. Kita & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language* 74: 557-589.

<http://creativecommons.org/licenses/by-nc-sa/3.0/>
Semantic typology as an approach to mapping the nature-nurture divide in cognition by Juergen Bohnemeyer is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](http://creativecommons.org/licenses/by-nc-sa/3.0/).