How variation in sampling changes semantic maps built on a comparison of parallel text data in the domain of motion events (verb stems and case/adpositions)

This paper deals with non-implicational semantic maps, built automatically using classical multi-dimensional scaling from a direct comparison of parallel text data (the Gospel according to Mark) in the domain of motion events (verb stems and case/adpositions) in more than 100 languages from all continents in more than 300 parallel clauses, and investigates how robust the result is if different biased smaller samples of (a) selected languages or (b) selected clauses are taken as a basis for the analysis. For implicational semantic maps, Haspelmath (2003: 217) claims that "[e]xperience shows that it is generally sufficient to look at a dozen genealogically diverse languages to arrive at a stable map that does not undergo significant changes as more languages are considered." This paper explores what are the actual differences in the results if the method is applied to subsamples of a dozen or more languages of particular continents (Africa, Eurasia, Oceania, the Americas) or particular language families (Indo-European, Austronesian, Niger-Congo). In the same way it is investigated what differences can result if the sample of clauses is manipulated. It is found that the selection of analytical primitives (the objects represented on the semantic map representing the functions compared cross-linguistically) is at least equally important as the sampling of languages.

The theoretical background of the approach taken here is briefly summarized below: Both implicational and non-implicational semantic maps (for the latter cf. Cysouw submitted, Wälchli submitted) rely on the single principle that cross-linguistically recurrent identity in form reflects similarity in meaning. This principle rests—implicitly or explicitly—on a theory of similarity semantics. Similarity semantics is concerned with the network of similarity and dissimilarity relationships emerging from the cumulative pairwise comparison of meanings and is particularly appropriate for comparing the meanings of contextually-embedded concrete utterances without having to assume an arbitrary set of primitive semantic units, since it operates without any reference to the notion of semantic identity (identical meanings, if there is such a thing in context, are treated as maximally similar meanings). Even if radically different, similarity semantics shares with Fregean truth semantics its indirect approach to meaning. It is not argued that similarity semantics is the only way how humans process meaning; rather the various forms of semantic map approaches are empirical methods to investigate how far we can go with similarity semantics alone, based on a minimal set of basic entities, similarly, e.g., as those assumed in Locke's (1690-1714, book iv, ch. i) theory of knowledge. Similarity semantics is also compatible with the structuralists' (de Saussure, Hjelmslev) concept of meaning as a continuous mass, analogous to the continuous phonetic space. Like phonemes, semantic categories of particular languages categorize particular areas of the continuous semantic space and typology is an indirect method to reconstruct the underlying semantic space, which cannot be measured directly unlike the articulatory-acoustic space in phonetics.

In semantic map approaches one has to distinguish strictly between underlying theoretical assumptions, such as outlined above, and aspects of the practical method applied. Most practical approaches to semantic maps (including the one applied here) assume that cross-linguistically identified "translation-equivalent" functions are identical, where they are in fact only similar. The semantic map approach works in practice to the extent that cross-linguistically identified functions are more similar than the functions of particular languages to be compared. Since cross-linguistically different categorization patterns can be distinguished in semantic maps only to the extent that there is a sufficiently high resolution of analytic primitives, it is crucial to match functions to be identified cross-linguistically as sharply as possible, which is best done by matching concrete examples. This is also necessary because concepts are at least partly based on exemplary knowledge in natural languages (Goldberg 2006: 229; Marty 1908: 530).

References

Cysouw, Michael (submitted). Building semantic maps: the case of person marking. In M. Miestamo, & B. Wälchli (eds.), *New Challenges in Typology: Broadening the Horizons and Redefining the Foundations*.

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

Haspelmath, Martin (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The New Psychology of Language* 2: 211-242. Mahwah, NJ: Lawrence Erlbaum.

Locke, John (1975 [1690-1714]). *An essay concerning human understanding*. edited by P.H.Nidditch. Oxford: Clarendon.

Marty, Anton (1908). *Untersuchungen zur Grundlegung der allgemeinen Grammatik und Sprachphilosophie*. Erster Band. Halle: Niemeyer.

Wälchli, Bernhard (submitted). Constructing semantic maps from etic parallel text data.