Commentary on de Haan - Can Semantic Maps Be Built Purely Bottom-Up?

Joost Zwarts

Utrecht Institute of Linguistics OTS

Can Semantic Maps Be Built Purely Bottom-Up?

Comment on 'Building a Semantic Map: Top-Down versus Bottom-Up Approaches' by Ferdinand de Haan (2010)

Joost Zwarts
Utrecht Institute of Linguistics OTS

Methodological considerations of semantic mapping play an important role in the papers of this issue. De Haan (2010) argues for a bottom-up approach in building semantic maps. In his view, domains (like "modality" versus "evidentiality") cannot and should not be distinguished until a map has been constructed that is based on a study of the meanings of individual morphemes. That might be a useful way to avoid semantic discussions about 'where' an item belongs, but some critical notes are in order about the way such a bottom-up approach might be done without top-down considerations.[1]

In a bottom-up approach, the data underdetermine the map, as shown in the case of Dutch *moeten*. The three meanings of *moeten* allow four possible maps. De Haan chooses one possible map, without further argumentation, but the question arises on what basis the lines are drawn in such a situation, if we do not want to use *top-down semantic* information about the similarity between meanings and the domains they belong to. A purely bottom-up procedure is not possible then.

De Haan identifies and classifies meanings of modal items in a rough descriptive way ("evaluation of evidence", "strong epistemic modality", "assertion of indirect evidence", "predictive", "future"). It is not clear how we can actually do this, without having some good idea about the broader semantic domains from which these meanings are drawn. It would be more in line with a real bottom-up approach to use neutral numbers for situations to which sentences apply, without bothering about the semantic description of those situations. However, that requires a rather different, "stimulus-driven" approach, like that of Levinson and Meira (2003) or Majid et al. (2007), where the points on the map are really theory-neutral to an important extent. This is probably not what de Haan would want to do, since he is clearly interested (and rightfully so) in how a set of more abstract modal/evidential meanings hangs together.

Another potential problem with the bottom-up approach is its partiality. In his paper, de Haan presents an essentially linear arrangement of meanings ranging from purely epistemic to purely temporal, with evidential as a category in the middle:[2]

strong — evaluative — assertive — predictive — future

---

[1] De Haan writes as if an item (like English *must* or Tuyuca *-yi*) all by itself can be called epistemic modal or evidential marker. But that cannot be correct in general. We know that items can be polysemous and span different domains. If we classify anything into domains, then it is individual meanings or functions. Lexical or grammatical items can only belong to a domain in a derivative and somewhat uninteresting sense.

[2] I am taking the "geometry" of the map seriously here, ignoring the way de Haan 'warps' this semantic map so that 'future' is below 'evaluative', though still unconnected to it. What counts, of course, are the connections, not the way the map is displayed on a two-dimensional piece of paper.

As he points out, some of the items discussed have other meanings too (e.g. "obligation"), but these are left out of the picture. In general, an analysis is always partial to a certain extent (in the sense that we restrict ourselves to a particular set of meanings), but in semantic maps this partiality carries a certain risk. We don't know whether "strong" and "future" are really far apart from each other on the map or whether this is an accidental result of the way items and meanings are selected. Including the "obligation" meaning, together with some more items from different languages, might easily result in a more connective map in which 'strong' and 'future' are then more closely related and the domains of epistemic modality and future interact with each other. What this shows is that we need some idea of the constellation of meanings making up a domain (or family of domains) before (or in addition to) building its semantic map, so that the partiality of our representation results from a well-informed abstraction and not from a more accidental lack of data.

As de Haan writes in footnote 18, a "merger" of top-down and bottom-up methods has proved to be the best way to go in parsing (computational linguistics). I hope that my comments on de Haan's paper suggest that the same is true for semantic maps (typological linguistics). It is not top-down *versus* bottom-up, but rather top-down *and* bottom-up.

References

de Haan, Ferdinand. 2010. Building a semantic map: top-down versus bottom-up approaches. Linguistic Discovery, this issue.

Levinson, Stephen C. and Sérgio Meira. 2003. 'Natural concepts' in the spatial topological domain – Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. Language 79/3.485-516.

Majid, Asifa, Melissa Bowerman, Miriam van Staden and James S. Boster. 2007. The semantics of 'cutting and breaking' events: A cross-linguistic perspective. Cognitive Linguistics 18/2.133-152.

Author's contact information:
Joost Zwarts
Opleiding Taalwetenschap
Departement Moderne Talen
Faculteit Geesteswetenschappen
Universiteit Utrecht
Trans 10
3512 JK Utrecht
The Netherlands
j.zwarts@uu.nl