Semantic map geometry: two approaches

Joost Zwarts (Radboud University Nijmegen & Utrecht University)

Common to all semantic map approaches is the idea of a 'geometric' layout of meanings, which represents graphically how meanings (or functions) of words (or grams) are related to each other. Where does this geometry come from? In most semantic map applications, the geometry emerges *a posteriori* from the linguistic data, in an inductive way, either by constructing the smallest graph of meanings in which every word covers a connected subgraph, or by applying statistical scaling techniques. However, it is also possible to work in the opposite direction, from an *a priori* geometry or grid of meanings, deducing relations that can be tested against linguistic data. The colour space offers the classical example of such a language-independent geometry of meanings (Gärdenfors 2000). The prepositional network of Lakoff (1987) and the reciprocal lattice of Dalrymple et al. (1998) can also be interpreted as conceptual spaces of this type.

The point of this paper is that we need *both* approaches, complementing each other. Often, a data-driven approach is the only way to get some idea about how a set of meanings hangs together. It is both a powerful heuristic and an important check on misguided a priori assumptions about a particular meaning space. However, the approach also has its limitations.

1 A semantic map should not be the theoretical endpoint. We want to know why the meanings are distributed in a particular way, but it actually turns out to be difficult to make the step from a data-driven semantic map to a semantic model of the underlying conceptual space. This is even harder when statistical mapping methods are applied. By using an exclusively inductive approach, the semantic map approach runs the risk of broadening the gap with semantic theories, both from the formal and cognitive paradigm. We therefore need to work from the other end too: define a geometry on the basis of particular semantic assumptions and study the cross-linguistic mapping of such a geometry.

2 One of the exciting things about semantic maps is that they could embody a non-classical, but constrained theory of categorization, thanks to the connectivity (convexity, contiguity) property. However, some small-scale maps show a distribution of data that can easily be captured in terms of necessary and sufficient features (as I will show for the modality map of Van der Auwera and Plungian 1998 and the A – S – P map of Croft 2001). If we want to show that regions on a semantic map are really more than classical feature bundles, a model of the underlying semantic geometry is inevitable.

3 In the data-driven mapping approach the important connectivity hypothesis is part of the methodology itself and as a result its validity can only be studied indirectly. There is no room for principled exceptions to connectivity, unless we already have some idea about what meanings are non-adjacent on independent semantic grounds (as I will illustrate with the modality map). A purely data-driven approach can not recognize the individual exceptions and working in the opposite direction is more fruitful here.

4 The a priori approach allows us to separate two roles of non-discreteness in mapping, which can be obscured in scaling methods. There can be non-discreteness in the conceptual space itself (the famous cups and saucers of Labov), but this should be distinguished from the non-discreteness that results from the way linguistic data distribute over a discrete geometry of meanings. I will show that a 'three-dimensional' map, in which words are not regions but *hills*, helps us to give a proper place to this distinction.

At the moment there are no good examples (apart from colour terminology) where large amounts of linguistic data are tested against an extensive 'a priori' conceptual space, but I will present a range of examples of smaller scale that suggest the direction in which this work might go, involving prepositions, clothing items and birds' names.

A data-driven, inductive approach to semantic maps has serious limitations, but, at the same time, a purely theory-driven, a priori approach to semantic maps does not work either. It is only when we are willing to go back and forth between semantic modeling and linguistic data that we can hope to gain insight in the way languages divide up spaces of meanings into words and grammatical markers.