Visualizing Prepositional Source Encoding in the NT Greek Gospels: A Similarity Semantics Approach using Multidimensional Scaling

Hanne M. Eckhoff (University of Oslo), Peter de Swart (University of Groningen), and Olga Thomason (University of Georgia)

In this paper we discuss how an empirical-contrastive approach can facilitate the semantic analysis in a single language. Our empirical domain is that of the prepositional encoding of Source meanings – a relatively understudied area of the spatial domain – in the four canonical gospels of the Greek New Testament. Our main goal is to identify factors that influence the choice of spatial markers (e.g. the choice between Greek *ek=GEN*, *apo=GEN*, *para=GEN*, and *exo=GEN*). Moreover, we want to reevaluate the relationship between the Source domain, on the one hand, and the Location/Direction domains, on the other, as it has been proposed that in many Indo-European languages the latter are closer, whereas Source is kept more distinct (Stolz 1992, Luraghi 2003).

Our data consist of the prepositional encodings in the four canonical gospels of the Greek New Testament and their parallel translations into Latin (L), Gothic (Goth) and Old Church Slavic (OCS) as made available in the PROIEL corpus developed at the University of Oslo, Norway (see Haug et al. 2009).[1] In our analysis we rely on the notion of *similarity semantics* (cf. Cysouw forthcoming, Wälchli forthcoming): the more similar two meanings are, the more likely they are to be expressed by the same form in any language. Thus, if two situations are consistently coded by the same marker in each language they should express semantically very similar concepts. If some languages code them in the same way but others differently they are less similar. Following the approach in Wälchli (forthcoming) we use multidimensional scaling (MDS) to visualize similarity by means of two-dimensional maps whereby semantic similarity is translated into proximity on the map (much alike traditional semantic maps as in Haspelmath 1997, 2003). A clear advantage of this computational approach is that semantic categories are not imposed on the data before the analysis, but emerge from the interpretation of the analysis itself.

Comparing Gk prepositional expressions and their translations we arrive at a general picture of how the spatial domain is structured in these languages. In the single-language analysis we superimpose the prepositions used in each situation onto this map and arrive at a picture of how Gk encodes this meaning space and what part of the meaning space is taken up by each preposition (Fig. 1). The map suggests some structure in this domain, namely, one can see three distinguished clouds corresponding to the three traditional spatial domains of Location (bottom), Direction (top right) and Source (top left), and clear patterns in the transitional areas. Zooming in on the Source domain, we find that the Gk PPs that are typically seen as Source markers (*ek=GEN*, *apo=GEN*, *para=GEN*, and *exo=GEN*) don't show a clear-cut pattern making prediction of prepositional choice difficult. A multifactorial exploration investigating factors such as animacy, definiteness, verbal semantics, syntactic relationship does suggest certain tendencies. For instance, *apo=GEN* is regularly chosen if a complement designates a human being and is indefinite. In this respect *apo=GEN* acts like a counterpart of *para=GEN* which tends to be used with definite nouns naming humans. *Ek=GEN*, by contrast, is a preferred PP if the governed phrase is definite and marks something non-concrete.

Finally, looking at the translational equivalents of the Gk Source markers in L, Goth and OCS, we find stable correspondences but also correlates with constructions that are not typically associated with the Source domain. For instance, Gk *ek=GEN* commonly marks location 'near, at' in combination with words naming sides in addition to its Source meanings. Translations of locational usages of *ek=GEN* show that Goth is parallel to Gk and reserves its Source marker *af=DAT*: However, L and OCS use a locational marker to express this meaning, *ad=ACC* and *o=ACC*, respectively. These latter Ps expresses a variety of locational and directional meanings, but their semantics do not extend to the denotation of Source. Such examples strongly suggest that the

[1] http://foni.uio.no:3000/, http://www.hf.uio.no/ifikk/proiel/corpus.html

boundaries between the domain of Location and the domain of Source are less strict than previously assumed.

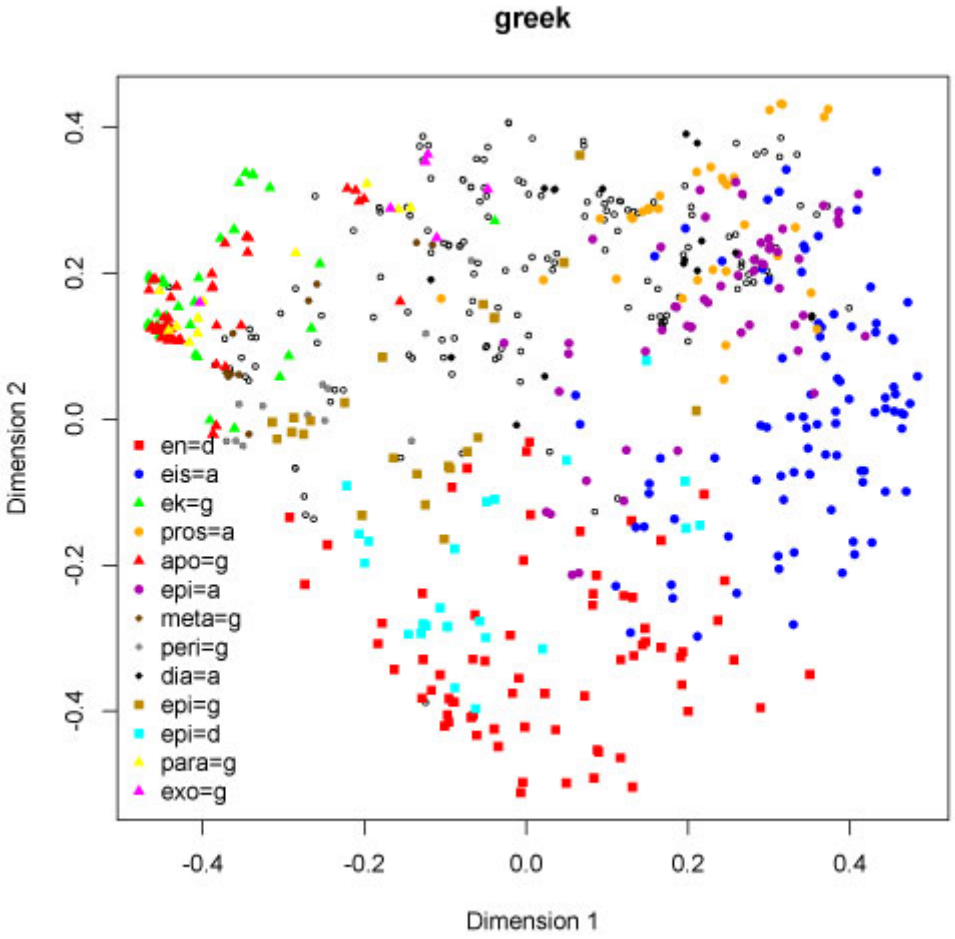

Figure 1 *The Gk prepositional domain based on the corresponding translations in Gk, L, Goth and OCS (only most frequent prepositions are shown). Shapes indicate typical spatial semantics: squares=Location, circles=Direction, triangles=Source (open dots are unclassified).*

References

Cysouw, Michael. Forthcoming. Semantic maps as metrics on meaning. *Linguistic Discovery.*

Haspelmath, Martin. 1997. *Indefinite pronouns.* Oxford: Oxford University Press

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

Haug, Dag T. T., Marius L. Jøhndal, Hanne M. Eckhoff, Eirik Welo, Mari J. B. Hertzenberg, Angelika Müth. 2009. Computational and Linguistic Issues in Designing a Syntactically Annotated Parallel Corpus of Indo-European Languages. *Traitement Automatique des Langues* 50/2, 17–45.

Luraghi, Silvia. 2003. *On the Meaning of Prepositions and Cases: The Expression of Semantic Roles in Ancient Greek*. Amsterdam: John Benjamins.

Stolz, Thomas. 1992. *Lokalkasussysteme. Aspekte einer strukturekken Dynamik.* Wilhelmsfeld: Egert.

Wälchli, Bernhard. Forthcoming. Similarity Semantics and Building Probabilistic Semantic Maps from Parallel Texts. To be published in *Linguistic Discovery.*