

Sweetening Ontologies cont'd

Elisabetta Jezek
Università di Pavia, Italy

By applying the Corpus Pattern Analysis procedure (CPA, Hanks 2004) to the analysis of concordances for ca 1000 English, Italian and Spanish verbs conducted with the aim of acquiring their most recurrent patterns, intended as corpus-derived argument structures with specification of the expected semantic type for each argument position (i.e. [[Human]] attends [[Event]]), we compiled a list of about 220 semantic types obtained from manual clustering and generalization over sets of lexical items found in the argument positions in the corpus (details of the Italian project in Jezek 2012).

These types look very much like conceptual / ontological categories for nouns but should instead be conceived as semantic classes, as they are induced by the analysis of selectional properties of verbs. They are language-driven, and reflect how we talk about entities in the world. As such, despite the obvious correlations, they differ from categories of entities defined on the basis of ontological axioms, such as those of DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering), which, despite “aiming at capturing the ontological categories underlying natural language and human common sense” (Gangemi et al. 2002) does not base category distinctions on systematic observation and clustering of language data.

In my presentation, I will report the preliminary results of the experiment of aligning the type inventory to the categories of DOLCE, with the aim of verifying how semantic classes obtained through pattern-based corpus analysis differ from categories which are defined on the basis of axiomatization. Also, I will discuss the opportunity to enhance the taxonomic structuring of our list using the OntoClean methodology (Guarino and Welty, 2009), which was also exploited for the development of DOLCE. Finally, I will highlight the mutual benefit of the experiment, and confirm the advantages of keeping the lexical level separated from the ontological level in language resource building (Oltamari et al. 2013).

References

- Gangemi, A. Guarino, N., Masolo, C. Oltamari A., Schneider L. et al. (2002). Sweetening Ontologies with DOLCE. In Gómez-Pérez A. and V.R. Benjamins (eds.) Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02), Ontologies and the Semantic Web, Berlin, Springer-Verlag, 166-181.
- Guarino, N. and C. Welty. 2009. An overview of OntoClean. In Staab, S. and R. Studer (eds.) Handbook on Ontologies (second edition), Berlin, Springer-Verlag, 201-220.
- Hanks, P. (2004) Corpus Pattern Analysis. In Williams, G. and S. Vessier (eds.) Proceedings of the Eleventh EURALEX International Congress, Lorient, France, 87-98.
- Jezek, E. (2012) Acquiring typed predicate-argument structures from corpora, in Bunt H. (ed.) Proceedings of the Eighth Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation ISA-8, Pisa, October 35, 2012, 28-33.
- Oltamari, A. Vetere, G. Chiari, I. Jezek, E. Zanzotto, F.M. Nissim, M. Gangemi, A. (2013) “Senso Comune: A collaborative Knowledge Resource for Italian”. In: Iryna Gurevych and Jungi Kim (eds.). The People’s Web Meets NLP: Collaboratively Constructed Language Resources, Berlin-Heidelberg, Springer, 45-68.