Building and Extending Knowledge Fragments

Wim Peters

University of Sheffield
United Kingdom
Email: w.peters@dcs.shef.ac.uk

In this panel presentation I will contend that it is possible to extract knowledge fragments from WordNet [1] and EuroWordNet [2] that combine explicit knowledge structures already provided by the thesauri such as synonymy, hypernymy and thematic relations, and implicit information from the (Euro)WordNet's hierarchical structure and the glosses that are associated with each WordNet synset.

The initial emphasis of the work lies on the detection of patterns of figurative language use, more particularly cases of regular polysemy [3].

The work consists of three phases. First, an automatic selection process identifies candidates for instantiations of regular polysemy [4,5] in WordNet on the basis of systematic sense distributions of nouns. These systematic distributions can be characterized by a pair of hypernyms taken from the WordNet hierarchies that subsume the senses. For instance, in two of its senses 'law' falls under the pattern *profession* (an occupation requiring special education) and *discipline* (a branch of knowledge). This set of conventionalised/lexicalised figurative language use forms the basis of the building of knowledge fragments.

In the second stage, the underspecified relations that exist between the word senses that participate in patterns are further specified in an automatic fashion. This additional information is obtained by analyzing the glosses that are associated with the synsets of the word senses involved and their hypernyms. For example, the extracted pattern **person** (a human being; "there was too much for one person to do") and **language** (a systematic means of communicating by the use of sounds or conventional symbols) subsumes sense pairs of 257 words in WordNet such as *Tatar*, *Assyrian*, *Hopi*, and *Punjabi*. The analysis of the WordNet glosses yields 'speak' as a significant relation.

This explicit knowledge that can be gleaned from information implicit in glosses enriches the already existing knowledge structures of WordNet, thereby expanding its coverage as a knowledge base. Also, it forms the start of the explicit encoding of metonymic potential of words where they do not yet participate in the patterns.

In the third phase, increasingly larger knowledge frames are built up on the basis of these sense pairs. The relation triples extracted in the second stage (e.g. **person-speak-language**) form the basic building blocks of the frames.

Extension of these rudimentary frames takes place in two ways. First, the concept with which hypernyms from the regular polysemy patterns co-occur can be regarded as additional slots in a topical frame that characterizes a hypernym. For instance, the pattern **music-dance** covers words such as tango and bolero. **Music** in its turn co-occurs with a number of other concepts within the hypernym pairs that characterize the regular polysemic patterns. These concepts and the relations that have been extracted between these hypernyms form a further extension of the **music** frame.

A further extension takes the semantic context of EuroWordNet into account. From the superset of all concepts and relations that are linked to MUSIC in all eight language specific

Petr Sojka, Karel Pala, Pavel Smrž, Christiane Fellbaum, Piek Vossen (Eds.): GWC 2004, Proceedings, pp. 8–9.

wordnets the MUSIC frame is extended with this new knowledge. The resulting structure is an extended knowledge frame that, amongst others, contains the following slot fillers and relations: **person-make/accomplish-music; musician isa person; musician play music; music-accompany-activity; dancing isa activity**.

These knowledge frames can be extended with information from other resources, and be used in a variety of applications.

References

1. Fellbaum, C., ed.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge, Mass. (1998).
2. Vossen, P.: Introduction to EuroWordNet. Computers and the Humanities (Special Issue on EuroWordNet) **32** (1998) 73–89.
3. Apresjan, J.: Regular Polysemy. Linguistics **142** (1973) 5–32.
4. Peters, W., Peters, I.: Lexicalised Systematic Polysemy in WordNet. In: Proceedings of the Second International Conference on Language Resources and Evaluation, Athens, Greece. (2000).
5. Peters, W., Wilks, Y.: Distribution-oriented Extension of WordNet's Ontological Framework. In: Proceedings RANLP-2001, Tzigov Chark, Bulgaria. (2001).