

Semantic concepts in wordnet

Ewa Rudnicka and Wojciech Witkowski

Politechnika Wrocławska Uniwersytet Wrocławski

ewa.rudnicka@pwr.edu.pl wojciechw62@gmail.com

The paper focuses on semantic concepts such as synonymy, hyponymy and meronymy, central to the creation of huge electronic language resources such as wordnets. In its unique format wordnet combines the structure and content of a monolingual dictionary, extended thesaurus and lexico-semantic network. Its basic building blocks are *lexical units* (lemma and sense pairs) and *synsets* (sets of synonymous lexical units), linked by a rich inventory of lexico-semantic relations such as antonymy, hyponymy/hypernymy, meronymy/holonymy, among others (cf. Fellbaum 1998). Thus, their definitions, procedures of application and substitution tests (if present) are crucial for the structure of a wordnet. Two competitive approaches to organising lexical units into synsets (on the basis of relations) will be discussed: a psycholinguistic one depending on lexicographers' intuitions regarding the replaceability of lexical units in similar contexts advocated by the developers of *Princeton WordNet* (cf. Miller 1993) and a computational linguistic one relying on the shared set of constitutive relations among lexical units proposed by the constructors of *plWordNet* (cf. Maziarz et al. 2013b). Their advantages and disadvantages will be viewed from the perspective of wordnet construction goals. Although wordnets are built mainly as resources for natural language processing tasks, some are also meant to become models of human language lexico-semantic systems. The latter aim will be of special interest here, since Princeton WordNet and plWordNet, as the largest and built-from-scratch world wordnets, offer the possibility of the extensive lexico-semantic studies. Moreover, plWordNet has been constructed by means of a unique corpus-based method (cf. Piasecki et al. 2009). Corpora were also used in the later stages of Princeton WordNet creation. Our analysis will capitalise on the results of mapping plWordNet to Princeton WordNet which showed a variety of interesting differences in the structure of the two wordnets (cf. Rudnicka et al. 2012, Maziarz et al. 2013a). These concern sense granularity, the depth of hypernymy trees, the number of lexical units within synsets, relation density, the treatment of gender, number, countability and markedness variants, the treatment of hypernymy as well as the use of different relations to code the same conceptual dependencies (the use of hyponymy vs. meronymy). All those characteristics will be carefully examined with an eye to possible modifications of the existing models of wordnet structure as well as to regularities and discrepancies in lexico-semantic systems of English and Polish.

References:

- Fellbaum, Ch. (ed). (1998). *WordNet: An Electronic Lexical Database*. Cambridge, Massachusetts: MIT Press
- Maziarz, M., Piasecki, M., Rudnicka, E. and S. Szpakowicz. (2013a). 'Beyond the Transfer-and-Merge WordNet Construction: plWordNet and a Comparison with WordNet.' In *Proceedings of RANLP*, Hissar.
- Maziarz, M., Piasecki, M., and S. Szpakowicz (2013b). 'The chicken-and-egg problem in wordnet design: synonymy, synsets and constitutive relations'. in *Special Issue of Language Resources and Evaluation*.
- Miller, G. A., Beckwith, R., Fellbaum, Ch., Gross, D., Miller, K. (1993). 'Introduction to WordNet: An On-line Lexical Database.' *International Journal of Lexicography* 3 (4), 1990, pp. 235-244. <ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps>
- Piasecki, M., Szpakowicz, S. and B. Broda. 2009. *A Wordnet from the Ground Up*. Oficyna Wydawnicza Politechniki Wrocławskiej: Wrocław.
- Rudnicka, E., Maziarz, M., Piasecki, M., Szpakowicz, S. (2012). 'A Strategy of Mapping Polish