Towards a truly bilingual wordnet - plWordNet 2.0.

Ewa Rudnicka
Wrocław University of Technology
Marek Maziarz
Wrocław University of Technology
Maciej Piasecki
Wrocław University of Technology

The paper focuses on the results of the first stage of mapping between two huge lexico-semantic networks: plWordNet and Princeton WordNet. The mapping not only led to the construction of a big bilingual lexico-semantic resource, but also allowed to trace a number of contrasts in the content and structure of the two networks. Though many of them obviously concern English and Polish lexis (e.g. lexical and cultural gaps, different structuring of information), other are contingent on different methodology applied in the creation of both networks, and these are going to be our main interest here.

Wordnets are a kind of a huge electronic thesauruses, yet of a much more advanced structure, since they incorporate not only synonymy relations among words, which are grouped into synonym sets called synsets, but also semantic relations between senses, thus synsets themselves. These relations involve synonymy, antonymy, hyponymy/hypernymy, meronymy/holonymy, and fuzzynymy, to name the major ones. The first ever wordnet was created in 1980s at Princeton University and was an unprecedented attempt to integrate data gathered in the existing lexical resources such as traditional and electronic dictionaries as well as gained from corpora into one huge database (cf. Fellbaum 1998). The creation of Princeton WordNet stimulated further work in the area and subsequently similar databases have been created both for European and non-European languages, and among them for Polish - plWordNet (cf. Piasecki et al. 2009, Maziarz et al. 2012). The natural forthcoming step was to build a multi-lingual database called EuroWordNet (cf. Vossen 2002), within which several interlinked wordnets for European languages were created. Most of them were constructed on the basis of the so called 'transfer method' that is by a kind of translation of Princeton WordNet structure. Independently, plWordNet was developed by applying a unique corpus-based method.

Since plWordNet was not linked to EuroWordNet, the aim of our project is to map plWN on Princeton WordNet. Contrary to the previous translation-based approach we have decided to adopt a novel perspective that is to link two independent systems, not just merely translate PWN. The main challenge of this task resides in different philosophical, theoretical and methodological assumptions that lay behind the construction of PWN and plWN. Still, our idea is to link PWN and plWN synsets via hierarchically ordered, inter-lingual relations such as synonymy, hyponymy, meronymy, hypernymy, holonymy, near-synonymy and fuzzynymy. This is accomplished by means of a mapping procedure and is supported by a semi-automatic system capitalizing on the existing bilingual resources (cf. Rudnicka et al. 2012). So far the process of mapping has revealed a number of contrasts boiling down to lexico-grammatical and lexico-semantic differences between English and Polish, different methodology used for the construction of two networks and different theoretical assumptions concerning their structure. These specifically involve the structure and role of lexical unit and synset, the type and application of lexico-semantic relations, the topology and content of the networks, and the use of different relations to code the same conceptual dependencies.

References:

Fellbaum, Ch. (ed). 1998. *WordNet: An Electronic Lexical Database.* MIT Press: Cambridge, Massachusets.

Maziarz M., Piasecki M., Szpakowicz S. (2012) Approaching *plWordNet* 2.0. Proceedings of the 6th Global Wordnet Conference, Matsue, 9-13th January, 2012, Japan (*Accepted for publication*)

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) Mapping plWordNet onto Princeton WordNet. A report available at http://nlp.pwr.wroc.pl/plwordnet/download/?lang=pl

Vossen, P. (ed). 2002. *EuroWordNet. General Document.* Amsterdam.