

PolNet 3.0 - DEMO

Zygmunt Vetulani, Grażyna Vetulani, Bartłomiej Kochanowski
Adam Mickiewicz University in Poznań, Poland

Abstract

We present the latest version of PolNet - Polish wordnet. Next step of transformation into a Lexicon Grammar of Polish. The current work consists in inclusion of verb-noun collocations and relations linking the verbal synsets to noun synsets.

Keywords: wordnet, PolNet, synonymy, synset, collocation

Introduction

The “PolNet-Polish WordNet” project started at the Adam Mickiewicz in 2006 with the objective to produce a lexicographical database inspired by the Princeton Wordnet (Miller et al., 1990). In December 2011 we have released PolNet 1.0 at the 5th Language and Technology Conference in Poznań (2011). Since that day we continue our work in order to build a Lexicon Grammar of Polish. During the demo we will present the current status of that work.

The concept

Our intention to make PolNet useful for systems with language functionality was the reason to extend the initial PolNet with verbs (initially simple, then compound verbs). This was also the first step towards the ambitious goal to transform a lexical ontology for Polish (PolNet) into a lexicon-grammar (Vetulani, Z. Obrębski, T., and Vetulani, G. 2007). In a lexicon-grammar, to describe the meaning of a predicative word one seeks to characterize the set of elementary sentences having this word as predicate. Our method of implementing this idea differs from what Gross did for French (through syntactic tables). In particular we introduce the so called semantic role relations such as Agent, Patient, Beneficient between verb and noun synsets to encode how the verbs and nouns combine to form simple sentences.

Project goal - lexical ontology

PolNet may be considered as a situational-semantic network of concepts (represented by synsets). Verb synsets may be considered as representing situations (events, states), whereas semantic roles (Agent, Patient, Beneficient,...) provide information on the ontological nature of various actors participating, actively or passively, in these situations (events, states). Abstract roles (Manner, Time,...) refer to concepts which position the situation (event, state) in time, space and possibly also with respect to some abstract, qualitative landmarks. Formally, the semantic roles are functions (in mathematical sense) associated to the argument positions in the syntactic pattern(s) corresponding to synsets. Values of these functions are ontology concepts (here in the form of noun synsets).

Granularity

Extension of the initial PolNet with simple verbs (PolNet 1.0) and verb-noun collocations (PolNet 2.0) opened up new application opportunities and motivated us to reconsider the fundamental problem of synonymy, directly related to the granularity of the wordnet. For verbs, and generally for all predicative structures, we focus on relations between the verb synsets (corresponding to predicative concepts) and noun synsets (representing object concepts), rather than on hierarchical relations, in order to show the semantic/morpho-syntactic connectivity restrictions corresponding to arguments. For these words, we propose to refine the concept of synonymy by considering valency structures. By valency structure we mean the structured information on the arguments opened by the predicative word including both semantic constraints on the arguments (semantic role values) as well as the surface morpho-syntactic and pragmatic properties of the text fillers of argument positions (like case, number, gender, preposition, register etc.) . The valency structure of a verb is one of the formal indices of meaning and should be considered as an attribute of a synset, i.e. all members of a synset should share the valency structure. Strict application of this principle results in a fine granularity of the verb section of the wordnet.

Grammatical case transformation

Extending the initial PolNet and in particular adding verb-noun collocations was not straightforward because of specific phenomena frequent in highly inflected languages (e.g. Polish) but rare in low inflected ones. Paraphrasing a sentence with a simple verb by replacing this verb by a verb-noun collocation often requires change of the argument's grammatical case. Although the simple verbs *kupić*, *nabyć*, as well as the collocation *dokonać zakupu* may all be translated into *to buy* in English, the grammatical case of the inanimate object will change from Accusative to Genitive when replacing any of the simple verbs by the collocation (*kupić/nabyć towar*(Acc) and needs to be replaced by *dokonać zakupu towaru* (Gen)). To simplify further processing, we decided to apply our definition of synonymy rigorously. This decision implies storing collocations and their corresponding single word equivalents in separate synsets, if only their valency structures are different (even if the intuitive meaning and usage seem be identical).

However, in all such cases we keep the corresponding synsets related by the transformational relations which describe the differences among their morpho-syntactic properties.

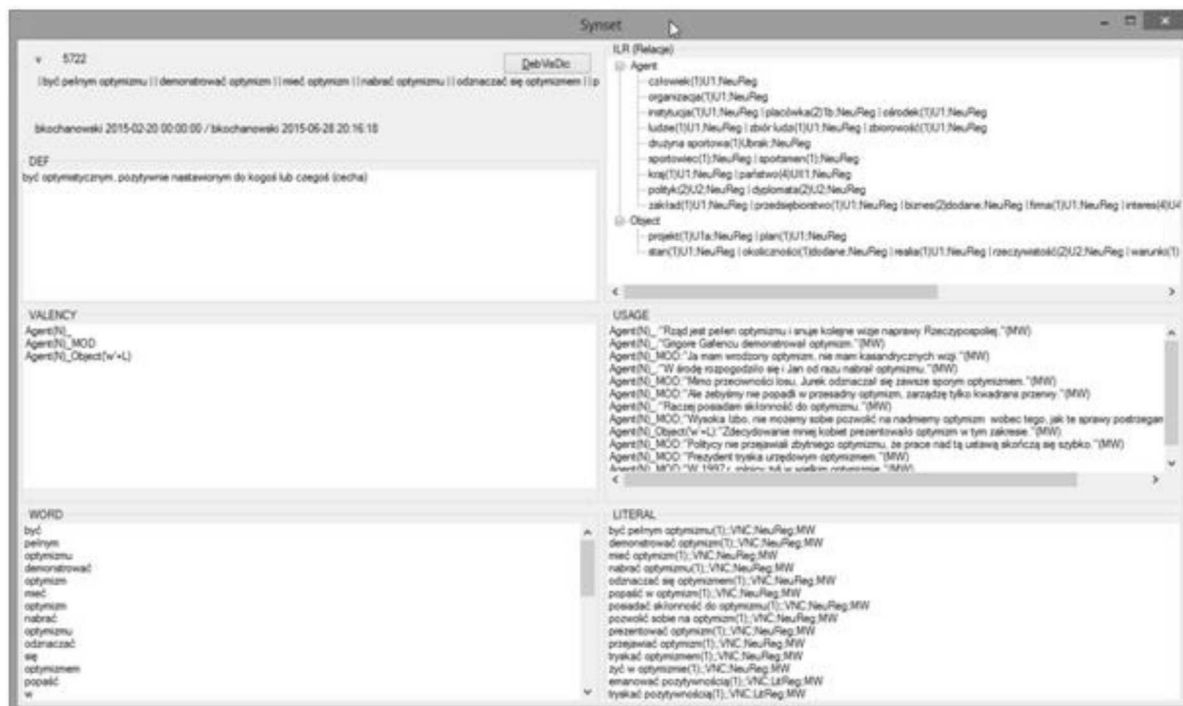


Figure 1. An example of a synset

Registers

For the present refinement of PolNet we assume that the category of language register is a part of the meaning. The totality of synsets was revised in order to split that contain different register words into register-uniform sub-

synsets. The initial synset was retracted, and all the subsynsets (with identical valency structure except for register) were introduced instead and related by the relation of synset similarity.

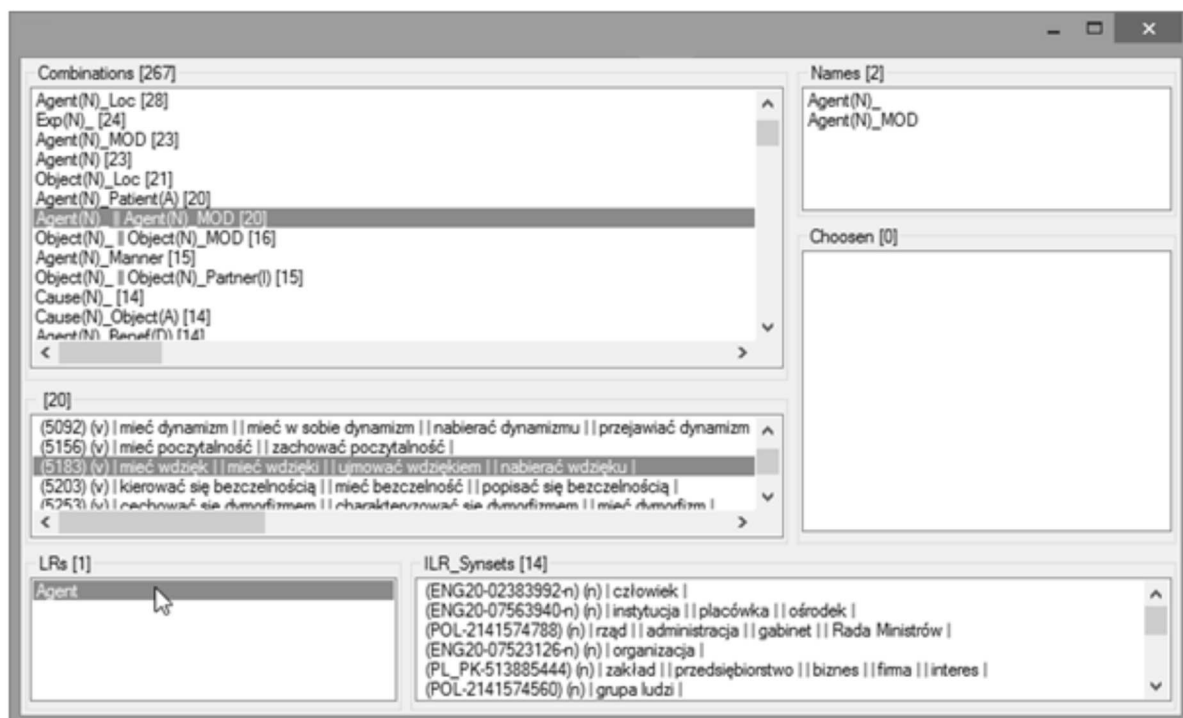


Figure 2. An example of valency assignment

Statistics

The current versions of PolNet contains:

- Nouns: 12011 synsets (10687 lemmas, 20364 word+meaning pairs)
- Verbs: 3645 synsets (5608 lemmas, 8201 word+meaning pairs)
- verb-noun collocations: 1908 synsets (4533 lemmas, 4774 word+meaning pairs)

References

G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. (1990.): WordNet: *An online lexical database*. Int. J. Lexicograph. 3, 4, pp. 235–244.

Vetulani, G. (2000). *Rzeczowniki predykatywne języka polskiego. W kierunku syntaktycznego słownika*

rzeczowników predykatywnych. (In Polish). Poznań: Wyd. Nauk. UAM.

Vetulani, G. (2012). *Kolokacje werbo-nominalne jako samodzielne jednostki języka. Syntaktyczny słownik kolokacji werbo-nominalnych języka polskiego na potrzeby zastosowań informatycznych. Część I.* (In Polish). Poznań: Wyd. Nauk. UAM.

VETULANI, Z., OBRĘBSKI, VETULANI, G. (2007): *Towards a Lexicon-Grammar of Polish: Extraction of Verbo-Nominal Collocations from Corpora*. In: *Proceedings of FLAIRS-07*, AAAI Press, Menlo Park, California, 267-268.