

Registers in the System of Semantic Relations in plWordNet

Marek Maziarz

Maciej Piasecki

Ewa Rudnicka

Institute of Informatics

Wrocław University of Technology

Wrocław, Poland

mawroc@gmail.com

maciej.piasecki@pwr.wroc.pl

ewa.rudnicka78@gmail.com

Stan Szpakowicz

Institute of Computer Science

Polish Academy of Sciences

Warsaw, Poland

&

School of Electrical Engineering

and Computer Science

University of Ottawa

Ottawa, Ontario, Canada

szpak@eecs.uottawa.ca

Abstract

Lexicalised concepts are represented in wordnets by word-sense pairs. The strength of markedness is one of the factors which influence word use. Stylistically unmarked words are largely context-neutral. Technical terms, obsolete words, “officialese”, slangs, obscenities and so on are all marked, often strongly, and that limits their use considerably. We discuss the position of register and markedness in wordnets with respect to semantic relations, and we list typical values of register. We illustrate the discussion with the system of registers in plWordNet, the largest Polish wordnet. We present a decision tree for the assignment of marking labels, and examine the consistency of the editing decisions based on that tree.

1 Introduction

A dense network of lexico-semantic relations is the feature that best differentiates a wordnet from other types of dictionaries and thesauri. Wordnets are organised into synsets and lexical units (LUs), whose meaning is crucially determined just by such relations. The inventories of relations, usually based on the findings in lexical semantics, seem largely comparable across wordnets, but specific definitions and strategies of applications vary. Wordnets also vary in the amount of typical dictionary information encoded. An apt example of such variation is the treatment of stylistic registers, as well as broad semantic domains, to which a given synset or LU belongs.

Lexico-semantic relations are the principal determinant of lexical meaning in plWordNet. Marking often constrains those relations; some of them cannot hold between LUs of incompatible registers. Semantics can constrain derivationally based relations, *e.g.*, femininity is limited to the nouns denoting animals and humans. Among verbs there are also relations limited to the particular aspect or verb class, like hyponymy or inchoativity (Maziarz et al., 2013, section 4).

Registers, semantic domains and verb classes are attributes. It is far from obvious how to put attributive information into an inherently relational structure of a wordnet. Princeton WordNet (PWN) represents semantic domains by synsets in two roles: elements of the lexical system and meta-information which characterises those elements. To associate a synset with a domain, a domain relation links it with another synset which represents that domain. Attributes of LUs can also be represented by sub-dividing those LUs into classes. This mechanism has been present in PWN from the beginning. There are, *e.g.*, separate bases for parts of speech or semantic domains, represented by the so-called lexicographers’ files.

Registers have a major role to play in shaping the structure of plWordNet. We will continue the practice of making registers figure in the definitions of lexico-semantic relations, but we will analyse them, and introduce register values into plWordNet, very systematically. To begin with, we need an appropriate set of marking labels. It should streamline the description of lexico-semantic relations, facilitate future plWordNet ap-

plications, and ensure the consistency of the linguists' decisions.

Section 2 of the paper recaps the model of the semantic relation system in pWordNet. Section 3 serves as an overview of related work insofar as it contributes to our intended use of the marking labels for the enrichment of the wordnet-based description of lexical meaning. Section 4 presents the details of the markedness labelling in pWordNet. Section 5 reports on a small, carefully arranged experiment meant to determine how consistent marking can be expected given a precise procedure in the form of a decision tree. Section 6 offers a few conclusions based on our experience, and briefly discusses our expectations for the ongoing development of pWordNet.

2 Constitutive relations and registers

A wordnet is founded on synonymy. Its basic unit, the synset, is a group of *lexical units* (LUs).¹ Although synonymy is undoubtedly key, wordnets vary as to how it is defined and applied. The creators of PWN (Miller et al., 1993; Fellbaum, 1998) adopt a very strict definition of synonymy usually attributed to Leibniz,² but realistically make it context-dependent. The effect is a take on synonymy which is linguistically satisfying but insufficiently accurate: the wordnet authors' intuition largely dictates what LUs go into a synset. Moreover, a synset is often understood as a set of synonymous LUs, while synonymous LUs are understood as elements of the same synset. Such circularity is hard to make operational.

The interdependence of the notions of synonymy and synset, and the subjectivity of authorial judgement, can be avoided. Maziarz et al. (2013) propose a different perspective. The LU, rather than the synset, becomes the basic structure in pWordNet. As Vossen (2002) notes, the central relations – synonymy, hyponymy, hypernymy, meronymy and holonymy – are lexical: they hold between words, not between concepts. A PWN synset denotes a *lexicalized concept*, and *conceptual* relations link synsets, but those relations have a lexico-semantic origin. Our model derives synset content and synonymy from a carefully constructed set of *constitutive relations* between

LUs. The construction is discussed in (Maziarz et al., 2013); the focus of this paper is on the properties which help constitutive lexico-semantic relations determine synsets.

The constitutive relations in pWordNet are hyponymy, hypernymy, meronymy and holonymy, plus verb-specific relations of presupposition, preceding, cause, processuality, state, iterativity and inchoativity (Maziarz et al. (2011) discuss the details), and adjective-specific relations of value (of an attribute), gradation and modifier (Maziarz et al., 2012b). They are supplemented by *constitutive features*: verb aspect and semantic class, and register.

A wordnet describes lexical meaning primarily via semantic relations, so it is important for a constitutive relation to be fairly widespread in the network. A high degree of sharing among groups of LUs is necessary because a constitutive relation underlies grouping LUs into synsets. It also helps if a constitutive relation is well established in linguistics: linguists who are wordnet editors will encounter fewer misunderstandings. Finally, a constitutive relations which accords with the wordnet practice will make for better compatibility among wordnets (Maziarz et al., 2013).

Verbs of different aspect participate in different lexico-semantic relations, *e.g.*, a hypernym cannot be replaced by the other element of its aspectual pair. The value of aspect thus constrains selected verb relations (Maziarz et al., 2012a). Semantic verb classes also restrict links for some verb relations. The verb classification is based on a Vendlerian typology. A hierarchy of verb classes has been implemented in pWordNet as a hypernymy hierarchy of *artificial lexical units*, each naming a different class. Verbs in a given class are hyponyms of the corresponding artificial LU.

Stylistic registers have been introduced into pWordNet relation definitions; they appear in guidelines and in some substitution tests. With every editing decision a linguist must recognise the registers of the LUs and synsets to be linked. The marking labels represent pragmatic features of LU usage, so it seems natural to have register values encoded explicitly.

A synset in pWordNet is a set of lexical units which are connected to the rest of the network by the same set of instances of constitutive relations, and have compatible values of the constitutive features. Note how this definition does *not* refer to

¹We understand the *lexical unit* informally as a lemma-sense pair.

²“Two words are said to be synonyms if one can be used in a statement in place of the other without changing the meaning of the statement.”

synonymy. Once synset membership has been decided, its elements are understood to be in the relation of synonymy.

It now becomes crucial to recognise accurately the connectivity afforded by the constitutive relations. Linguists who build the wordnet are assisted by conditions in the definitions of relations (such conditions often refer to registers and semantic classes) and by substitution tests. Vossen (2002) discusses tests for semantic correspondence, which did not take into account the differences in register or usage, often essential for the possibility of contextual interchangeability.

Lexical units which have nearly the same sense but significantly differ in register are put into separate synsets, but the proximity is not lost: those synsets become linked by inter-register synonymy. That relation is weaker than synonymy with respect to sharing. Synsets linked by inter-register synonymy share a hypernym, but *not* hyponym sets, and clearly have different register values.

Consider an example: *komputer* ‘computer’ has an obsolete inter-register synonym *mózg elektroniczny* ‘electronic brain’. Figure 1 shows hyponymy to *urządzenie elektroniczne* ‘electronic device’, which is shared.³ There is, however, a hyponym *komputer cyfrowy* ‘digital computer’, a specialist term which should not be linked to the obsolete term for a computer.⁴ The terms *komputer* and *mózg elektroniczny* have the same denotation but different linguistic contexts of use.

The model and the development of plWordNet comply with a form of minimal commitment principle: make as few assumptions as possible about the construction process. First of all, the model avoids references to theories of cognition and specific theories of lexical semantics. By minimising the theoretical underpinning and grounding all editing decisions on the language data observable in a corpus, we try to focus on the lexical system regardless of the reasons why it is organised as it is. We thus hope to make the wordnet theory-neutral and ready for use in a wide range of applications.

Minimal commitment does not preclude a *mapping* to an ontology. Such a mapping supplements the linguistic dependencies recorded in the wordnet with a theoretical interpretation: the cogni-

³In plWordNet, a hyponymy link from X to Y means “X is a hyponym of Y” rather than “X has a hyponym Y”.

⁴The substitution test “If it is *a digital computer*, then it must be *an electronic brain*.” sounds distinctly funny.

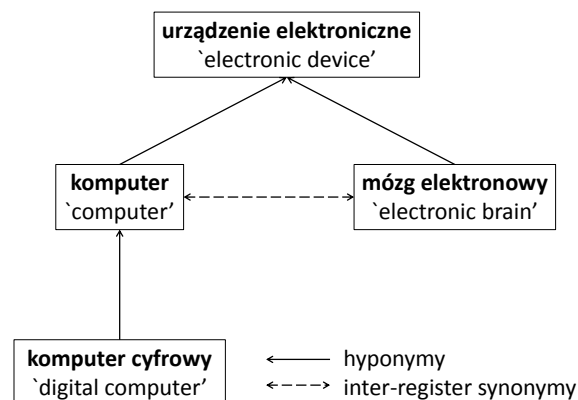


Figure 1: Inter-register synonymy between LUs from different registers.

tive principles of the ontology. The wordnet describes lexico-semantic relation of varying, possibly complex, background and origin, while the ontology mapping shows a possible relation between the lexical system and the internal cognitive structure of concepts. A potential plus is the possibility of considering different ontologies as the mapping target, and so different interpretations of the lexical dependencies.

3 Markedness in lexicography and in Princeton WordNet

Svensén (2009) notes that lexicographers refer as *marked* to the part of the vocabulary with additional pragmatic features which narrow the usage to a specific context or group of speakers. Such distinction includes, but is not limited to, different stylistic registers. Svensén adopts the classification of “diasystematic marking in a contemporary general purpose dictionary” (Hausmann, 1989), organised along 11 criteria: time, place, nationality, medium, socio-cultural, formality, text type, technicality, frequency, attitude and normativity. An unmarked centre and a marked periphery⁵ are established for each criterion. The main peripheries include “archaism-neologism; regionalism, dialect word; foreign word; spoken-written sociolects; formal-informal; poetic, literary, journalese; technical language; rare; connoted; and incorrect”. In a dictionary, the location of a lexical item in a periphery is signalled by a label, *e.g.*, *arch* ‘archaic’, *AmE* ‘American English’ or

⁵or peripheries, because for some criteria there can be more than one periphery

poet ‘poetic’.

Wordnets vary with respect to the ways and degree of coding markedness. PWN signals markedness with a special DOMAIN - MEMBER OF A DOMAIN relation with three sub-types: TOPIC, REGION and USAGE. It can be established between synsets in the same grammatical category or between categories. The subtype names correspond to the criteria of marking (Hausmann, 1989). Surprisingly, noun synsets play the role of specific labels within particular subtypes. PWN 3.0 has 438 labels pertaining to DOMAIN TERM TOPIC, 166 labels to DOMAIN TERM REGION, and 29 labels to DOMAIN TERM USAGE.

A closer look at the specific label instances within the selected domains shows that some of them belong to different peripheries of Svensén (2009) / Hausmann (1989). The TOPIC domain includes such labels as, e.g., ‘archeology’, ‘Arthurian legend’ or ‘auto racing’. The USAGE domain includes, e.g., ‘archaism’, ‘African American Vernacular English’ and ‘colloquialism’. REGION seems to be built most consistently: in principle it concerns dialectal names. It could thus be treated as an equivalent of the ‘regionalism-dialect word’ periphery. Yet, some of those links signal only geographical membership, but not dialectal variation. Consider, for example, the relation DOMAIN TERM REGION between {Polynesia} and {Austronesia}. It is clearly not the case that *Polynesia* is a dialect word used mainly in, or coming from, Austronesia.

Polish lexicography distinguishes groups of marking (register) labels not unlike those we showed above: diachronic, stylistic, emotional, terminological (professional, scientific), diastratic, diatopic (geographical), diafrequential (Dubisz, 2006; Engelking et al., 1989). The consistency of marking is low. Lexicographers point out mistakes and dubious decisions in the dictionary-making process (Kurkiewicz, 2007).⁶ Not only

⁶Consider *metal* ‘one listening to heavy metal music’ and *wywiad* ‘interview’. Dubisz (2006) labels the former *youth language*, the latter – *journalism*. Żmigrodzki (2012) assigns *music* to the former and no label to the latter.

This is not only the malady of Polish lexicography. In English and German dictionaries, words also carry assorted register labels. Svensén (2009, p. 316) notes: “Different dictionaries may use different labels, and the categories represented by the labels may have different ranges in different dictionaries. Moreover, there may be differences in labelling practice, so that, in one dictionary, fewer or more lexical items are regarded as formal or informal, correct or incorrect, etc., than in another one (Hausmann 1989: 650).”

do dictionaries label the same lexical units differently (Engelking et al., 1989), but the label lists vary significantly (*exemplum* (Dubisz, 2006) and (Kurkiewicz, 2007)). There also are too many labels (ca. 20-30 main and more than 100 secondary categories), so it is virtually impossible to mark the semantico-pragmatic constraints with any degree of consensus.

Several sets of criteria have emerged during the lexicographic debates in Poland. We find the set proposed by Buttler and Markowski (1998) to be the most interesting. Three semantico-pragmatic features are posited: *official*, *specialist*, *emotional* (or *emotionally marked*, or *expressive*). Their +/- (present/absent values) define a space in which all language variants or styles can be placed. Thus, general language could be characterised by {–*official*, –*specialist*, –*emotional*}, and literary style by {+*official*, –*specialist*, –*emotional*}.

4 Registers in plWordNet

Although in plWordNet 2.0 registers did influence relations, they were not introduced explicitly. In order to gain high consistency, we have decided to mark labels explicitly, and to create detailed guidelines for the lexicographers.

The set of plWordNet marking labels is inspired by Buttler and Markowski (1998) and by Kurkiewicz (2007). As does the Great Dictionary of Polish (Kurkiewicz, 2007; Żmigrodzki, 2012), we aim to lower the overall number of labels by about an order of magnitude. In the end, we have distinguished nine marking labels, with general (unmarked) language as the tenth register.⁷

- **obsolete** – this label marks LUs which are outdated, typically used only by elderly or (rarely) middle-aged people;
- **regional** – LUs from a dialect, well known to (but not used by) almost all Poles;
- **terminological** {+off, +spec, –emo} – LUs used by specialists, scientists, engineers,

For example, *Oxford English Dictionary* (Simpson, 2013) equips the word *malady* with the label *literary*, while *Cambridge Dictionaries Online* (Heacock, 1995 2011) consider it *formal*. The word *freak* is *informal* in (Simpson, 2013), but has no label (!) in (Heacock, 1995 2011).

⁷We abbreviate the three features from (Buttler and Markowski, 1998) as *off* = *official*, *spec* = *specialist*, *emo* = *emotional*.

and generally professionals;

- **argot/slang** $\{-off, +spec, +emo\}$ – LUs used by a particular social group or a small/local community;
- **literary** $\{+off, -spec, \pm emo\}$ – this label marks high style vocabulary, especially LUs used only in literature or in speeches;
- **official** $\{+off, -spec, -emo\}$ – LUs used on official and formal occasions, mainly in communication between citizens and representatives of state institutions;⁸
- **vulgar** $\{-off, -spec, +emo\}$ – crude vocabulary, LUs with very restricted acceptable usage;
- **popular** $\{-off, -spec, +emo\}$ – LUs which might be used in a familiar context, but normally not acceptable in other situations;
- **colloquial** $\{-off, -spec, +emo\}$ – vocabulary used informally, in a free style, but with low acceptability in official situations;
- **general** $\{\pm off, -spec, -emo\}$ – LUs which could be used virtually in every situation.

To help plWordNet editors maintain consistency, we have designed a series of substitution tests in the form of a decision tree. The editor systematically inspects the semantic features $\pm spec$, $\pm off$ and $\pm emo$ for a given LU, as well as more specific pragmatic features. The tree appears in Figure 2. Consider Example 1 (the prerequisite is italicised, the actual test is set in roman):

Example 1 (*regional*)

Test. *The LU `pyra` ‘potato’ may have equivalents in other regions of Poland or in general language.* The Poles know the LU `pyra` and recognise it as regional.⁹

The test is applied right after the diachronic criterion (Figure 2, `obs`). If the prerequisite and the test proper both hold, the LU `pyra` is marked as

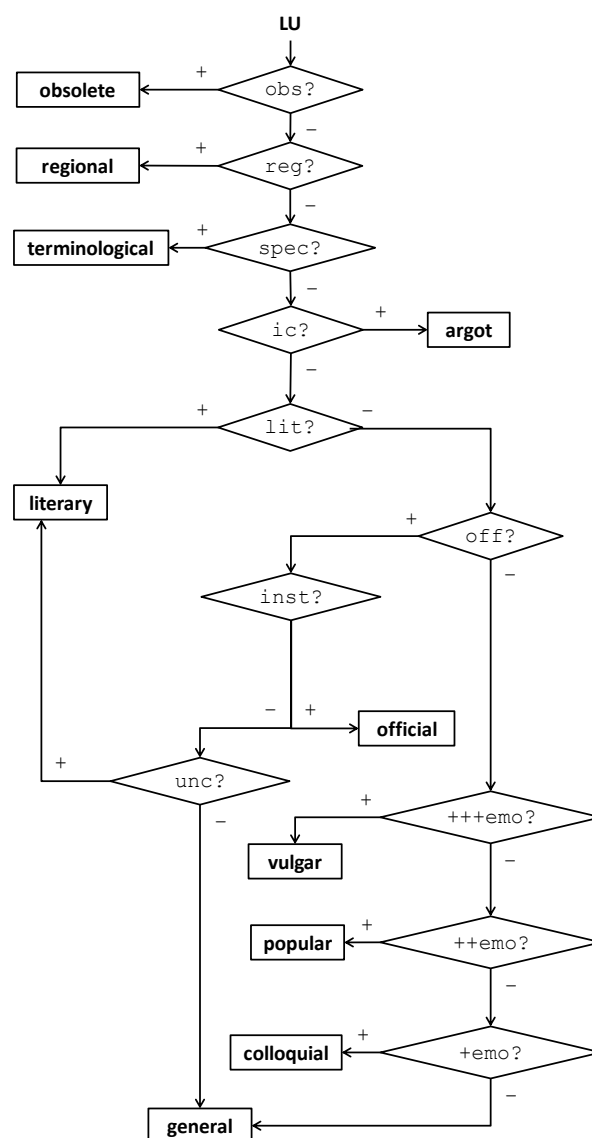


Figure 2: Substitution tests for markedness in plWordNet. Legend: `+++emo` = emotionally marked LU unacceptable in most situations (that includes vulgarity), `++emo` = emotionally marked LU acceptable only in familiar situations, `+emo` = emotionally marked LU acceptable in some familiar situations and when talking to strangers, `ic` = LU from a slang or argot, `inst` = LU used only in communication with state institutions, `lit` = LU used only in literature, `obs` = used by the elderly, `off` = language suitable for official situations, `reg` = regional LU, `spec` = LU used only by specialists, `unc` = LU unsuitable for common communication.

regional. The test fails if either part disagrees with the plWordNet editor’s intuition.

⁸Such language develops around any bureaucracy.

⁹It is used in Greater Poland.

Example 2 shows a two-step test: consider a (possibly) vulgar noun first in an unofficial situation of talking to a stranger, and then in a very official situation.

Example 2 (vulgar)

Test 1. *Imagine that you meet a stranger in the street and talk a while. You have just used the LU skurwiel ‘son of a bitch’. Your interlocutor will most likely think that you are crude.*

Test 2. *Imagine yourself in the middle of a very official or public situation (you are in the presence of an elder, your superior, president of the Polish Republic, a professor, a bishop, or you are being interviewed on TV news). You have just used the LU skurwiel ‘son of a bitch’. Your interlocutor – or TV viewer – will most likely think that you are crude.*

The substitution tests are applied in a cascade of filters. An LU which passes through all filters must land in the final bin – the general register.

5 The stability of the substitution tests

To ensure that the marking labels introduced in Section 4 can be applied with sufficient consistency, we examined the inter-rater agreement between two plWordNet editors who independently marked a sample of LUs. They were given a document with detailed guidelines and complete tests, and a spreadsheet with 385 noun LUs randomly drawn from plWordNet (a simple random sample, proper names and gerunds excluded).

Figure 3 presents the histograms of the counts of marking labels in the 385-LU sample. The most frequently assigned registers are terminology, general language, and literary and colloquial styles. These four account for more than 90% of the sample. Both editors found terminology to be the most frequent register, and neither found the *vulgar* label necessary. If we were to extrapolate, we could venture a broad guess on the approximate distribution of register values of LUs in plWordNet:

- $\frac{2}{5}$ in the terminology register,
- $\frac{1}{3}$ in the general register,
- $\frac{1}{6}$ in the literary style,

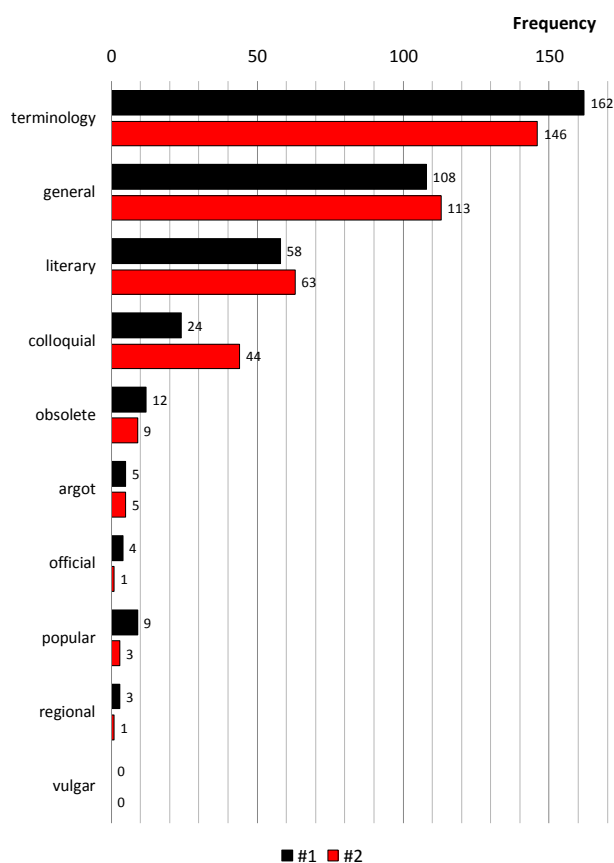


Figure 3: Counts of stylistic register values in a 385-LU sample from plWordNet, with two raters.

- $\frac{1}{12}$ in the colloquial style,
- $\frac{1}{12}$ in the remaining registers.

The annotators are in reasonable agreement, as measured by Cohen’s kappa: $\kappa = 0.645$ with the confidence interval 0.586-0.722 (Table 1).¹⁰ According to Landis and Koch (1977, p. 165), the confidence interval covers two values of agreement strength: moderate and substantial.

It is commonly assumed that only $\kappa \geq 0.8$ guarantees reliable results in computational linguistics, and κ in 0.67-0.8 is tolerable. Reidsma and Carletta (2007) show that this rule of thumb does not always work. Sometimes lower κ makes the results reliable, sometimes even $\kappa \geq 0.8$ does not suffice. The authors recommend checking whether differences between annotators are systematic or random,¹¹ so we have decided also to put our data

¹⁰The confidence interval was calculated by simple percentile bootstrap (DiCiccio and Efron, 1996; DiCiccio and Romano, 1988) suitable for Cohen’s κ (Artstein and Poesio, 2008).

¹¹The former is a real problem for computational methods,

label system	Cohen's κ	confidence interval of κ	p -value of χ^2 test
10 labels	0.645	0.586-0.722	0.03962
5 labels	0.722	0.657-0.785	0.02686

Table 1: Inter-rater agreement of two annotators assigning marking labels to nouns from plWordNet. Confidence intervals are calculated by the percentile bootstrap method, $n = 10000$ resamplings, $\alpha = 0.05$. P -values are calculated for the χ^2 tests of independence. The 10-label system was described in Section 4. The 5-label system equates compatible labels, as described in this section.

through a non-parametric χ^2 test of independence. The p -value is 0.03962, so we do not reject the null hypothesis that the plWordNet editors' choices are distributed similarly at 1% significance level.

The Cohen's κ value will increase if there are fewer marking labels. One fairly obvious way of doing that is to consider as *compatible* those marking label bins whose definitions are close; see the decision tree in Figure 2:

- general \approx literary \approx colloquial,
- official \approx terminology \approx argot,
- vulgar \approx popular.

This boosts Cohen's kappa to 0.722 with a very good confidence interval of 0.657-0.785. Now the κ is in the area of substantial agreement of Landis and Koch. The χ^2 test for the new labelling system again leads us to the fortunate assumption that distributions of editor choices are similar at 1% significance level (so none of the editors has any bias). Fewer labels, narrow and high inter-rater agreement, but somewhat less information. . .

6 Conclusions

The model proposed for plWordNet bases the grouping of lexical units (LUs) into synsets on *constitutive relations*. In order to match the language data even more accurately, we enriched the definitions of some of the semantic relations. We added constraints which refer to verb aspect, verb semantic classes and registers. Those features play a central role in shaping the wordnet relation structure, so we named them *constitutive features*.¹² the latter it not a threat.

¹²It is *attributive* information in an inherently *relational* system, but there is no contradiction. This information only

Registers appear to be particularly important, because they characterise all parts of speech covered by plWordNet, and they link the pragmatics of usage in a simple manner with the lexico-semantic description in the relational paradigm. That is why registers in plWordNet will now explicitly characterise LUs.

A review of the linguistic study of registers has suggested a set of ten registers, including the default unmarked register. We have also designed rules for register identification in the form of a decision tree, and made them a mandatory element of the guidelines for wordnet editors. We ran an annotation experiment in which two linguists independently assigned register values to a representative sample. We conclude that LUs can be given register labels with acceptable inter-annotator agreement.

Our wordnet model follows the minimal commitment principle. We only consider a small set of homogeneous and quite carefully specified basic notions. The whole system of semantic relations and synsets in a wordnet is directly derived from the linguistic lexico-semantic relations and from language data. The structure of the wordnet is closer to language facts, because it is derived from the lexico-semantic relations between LUs which can largely be observed directly in corpus data. That is why the adopted wordnet model facilitates semi-automated wordnet expansion using knowledge extracted from corpora. The systematic introduction of registers allows us to take into account elements of pragmatics without giving up the conceptual simplicity of the model.

Acknowledgment

Co-financed by the Polish Ministry of Education and Science, Project CLARIN-PL, and the European Innovative Economy Programme project POIG.01.01.02-14-013/09.

References

- Ron Artstein and Massimo Poesio. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34(4):555–596.
- Danuta Buttler and Andrzej Markowski. 1998. Słownictwo wspólnoodmianowe, książkowe i potoczne współczesnej polszczyzny [The general, bookish and colloquial vocabulary of contemporary Polish].

helps constrain semantic relations, which remain the principal vehicle for the description of lexical meaning.

- Język a Kultura [Language and Culture]*, 1:179–203.
- Thomas J. DiCiccio and Bradley Efron. 1996. Bootstrap Confidence Intervals. *Statistical Science*, 11(3):189–212.
- Thomas J. DiCiccio and Joseph P. Romano. 1988. A Review of Bootstrap Confidence Intervals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(3):338–354.
- Stanisław Dubisz. 2006. Wstęp [introduction]. In Stanisław Dubisz, editor, *Uniwersalny słownik języka polskiego PWN. Wersja 3.0 (elektroniczna na CD) [A universal dictionary of Polish. Version 3.0 (electronic on CD)]*. Polish Scientific Publishers PWN.
- Anna Engelking, Andrzej Markowski, and Elżbieta Weiss. 1989. Kwalifikatory w słownikach – próba systematyzacji [Qualifiers in dictionaries – an attempt to systematise]. *Poradnik Językowy [Language Guide]*, pages 300–309.
- Christiane Fellbaum, editor. 1998. *WordNet – An Electronic Lexical Database*. The MIT Press.
- Franz Josef Hausmann. 1989. Die Markierung im allgemeinen einsprachigen Wörterbuch: eine Übersicht. In Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand, and Ladislav Zgusta, editors, *Wörterbücher. Ein internationales Handbuch zur Lexikographie*, volume 5.1, pages 649–657. De Gruyter.
- Paul Heacock, editor. 1995-2011. *Cambridge Dictionaries Online*. Cambridge University Press.
- Juliusz Kurkiewicz. 2007. Kwalifikatory w wielkim słowniku języka polskiego [Qualifiers in the great dictionary of Polish]. In Piotr Żmigrodzki and Renata Przybylska, editors, *Nowe studia leksykograficzne [New lexicographic studies]*. Wydawnictwo Lexis.
- J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174.
- Marek Maziarz, Maciej Piasecki, Stanisław Szpakowicz, Joanna Rabiega-Wiśniewska, and Bożena Hojka. 2011. Semantic Relations between Verbs in Polish Wordnet 2.0. *Cognitive Studies*, 11.
- Marek Maziarz, Maciej Piasecki, and Stan Szpakowicz. 2012a. An Implementation of a System of Verb Relations in plWordNet 2.0. In Christiane Fellbaum and Piek Vossen, editors, *Proc. 6th International Global Wordnet Conference*, pages 181–188, Matsue, Japan, January. The Global WordNet Association. (www.globalwordnet.org/gwa/proceedings/gwc2012.pdf).
- Marek Maziarz, Stanisław Szpakowicz, and Maciej Piasecki. 2012b. Semantic Relations among Adjectives in Polish WordNet 2.0: A New Relation Set, Discussion and Evaluation. *Cognitive Studies*, 12.
- Marek Maziarz, Maciej Piasecki, and Stanisław Szpakowicz. 2013. The chicken-and-egg problem in wordnet design: synonymy, synsets and constitutive relations. *Language Resources and Evaluation*, 47(3):769–796.
- George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. 1993. Introduction to WordNet: an on-line lexical database. Unpublished, one of “Five Papers” ([ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps](http://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps)).
- Dennis Reidsma and Jean Carletta. 2007. Reliability measurement without limits. *Computational Linguistics*, 1(1):1–8.
- John Simpson. 2013. *Oxford English Dictionary*. Oxford University Press. (www.oed.com/).
- Bo Svensén. 2009. *A handbook of lexicography: the theory and practice of dictionary-making*. Cambridge University Press.
- Piek Vossen. 2002. EuroWordNet General Document Version 3. Technical report, University of Amsterdam.
- Piotr Żmigrodzki, editor. 2012. *Wielki słownik języka polskiego: Zasady opracowania [A great dictionary of Polish: The principles of compilation]*. Institute of Polish Language, Polish Academy of Sciences.