Lang Resources & Evaluation (2013) 47:769–796
DOI 10.1007/s10579-012-9209-9

ORIGINAL PAPER

The chicken-and-egg problem in wordnet design: synonymy, synsets and constitutive relations

Marek Maziarz · Maciej Piasecki · Stanisław Szpakowicz

Published online: 18 April 2013

Abstract Wordnets are built of synsets, not of words. A synset consists of words. Synonymy is a relation between words. Words go into a synset because they are synonyms. Later, a wordnet treats words as synonymous because they belong in the same synset... Such circularity, a well-known problem, poses a practical difficulty in wordnet construction, notably when it comes to maintaining consistency. We propose to make a wordnet a net of words or, to be more precise, lexical units. We discuss our assumptions and present their implementation in a steadily growing Polish wordnet. A small set of constitutive relations allows us to construct synsets automatically out of groups of lexical units with the same connectivity. Our analysis includes a thorough comparative overview of systems of relations in several influential wordnets. The additional synset-forming mechanisms include stylistic registers and verb aspect.

Keywords Wordnet · *WordNet* · Synset · Lexical unit · *plWordNet* · Wordnet relations · Constitutive relations · Register · Aspect

Work financed by the Polish Ministry of Science and Higher Education, Project N N516 068637.

M. Maziarz · M. Piasecki
Institute of Informatics, Wrocław University of Technology, Wrocław, Poland

M. Maziarz
e-mail: marek.maziarz@pwr.wroc.pl

M. Piasecki
e-mail: maciej.piasecki@pwr.wroc.pl

S. Szpakowicz (✉)
School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada
e-mail: szpak@eecs.uottawa.ca

S. Szpakowicz
Institute of Computer Science, Polish Academy of Sciences, Warszawa, Poland

1 Introduction

A wordnet is a complex structure with a slightly misleading name. Far more than a "net of words", a typical thesaurus/dictionary/ontology has synsets at its bottom rather than word forms or lexemes. Synonymy is intended as the cornerstone of a wordnet, hypernymy—its backbone, meronymy—its essential glue. None of these relations, however, holds first and foremost between synsets: they are lexico-semantic relations, while a synset represents a concept. Whatever the term *concept* refers to, it is not lexical (only a single-word synset can be construed as involved in the same relations as its lone word) (Fellbaum 1998, p. 210). Quite inconveniently, to define a synset as a set of synonyms is to introduce a vexing circularity, if a synonym—as it happens so often—is defined as an element of a synset. Hypernymy fares no better: a synset may be so heterogeneous that its place in a class hierarchy is a matter of degree, not a certainty, even if a typical wordnet hypernymy tree is assumed to implement a crisp classification.

1.1 Synsets in Princeton *WordNet*

In Princeton *WordNet* (henceforth PWN), Miller et al. (1990, 1993) present a *synset* as "a set of synonyms that serve as identifying definitions of lexicalised concepts". The Authors also write that lexical "meanings can be represented by any symbols that enable a theorist to distinguish among them" (Miller et al. 1993, p. 5). Words are meant to be symbols which differentiate meaning, and the only criterion of their selection is their synonymy. The Authors emphasise that a synset, because of its members, directs a native speaker to the concept lexicalised (thus shared) by all synset members. The synset is, then, supposed to be a vehicle for a *lexicalised concept* (*ibid.*). It is sometimes defined as a set of words which refer to the same lexicalised concept—and lexicalised concepts are presented as objects described, via synsets, by "conceptual-semantic relations" (Fellbaum 1998a, p. 210).

The key element of the definition of the synset in PWN is the notion of *synonymy*. Miller et al. (1993) rely on Leibnitz's perspective on synonymy: the exchange of a word in a sentence for its synonym does not change the truth value of that sentence in its usages. Such a definition, however, severely limits the number of synonymous pairs in any natural language. That is why the Authors have proposed a weaker criterion. It is enough that truth conditions be preserved only in some contexts or usages. But now context becomes an intrinsic part of the synonymy criterion, so it must be properly described. Two problems emerge: how such a description should look, and how specific it should be. In practice, for many word pairs one can find many contexts which allow truth-preserving exchange, and many contexts which do not. The nature and granularity of contexts is left to intuition. Such synset definitions—with varying wording—are common in wordnets, and they all fall short (Pedersen et al. 2009; Tufiş et al. 2004; Koeva et al. 2004).

1.2 Synonymy in *EuroWordNet*

EuroWordNet (henceforth EWN) (Vossen 2002, p. 5) follows Miller et al. (1990) but also refers to the notion of the semantic equivalence defined at the level of word denotations:

> In *EuroWordNet*, we further mean by semantically-equivalent that two words denote the same range of entities, irrespective of the morpho-syntactic differences, differences in register, style or dialect or differences in pragmatic use of the words. Another, more practical, criterion which follows from the above homogeneity principle is that two words which are synonymous cannot be related by any of the other semantic relations defined.

Substitution tests for synonymy include a clear criterion of word exchange in some contexts. Here is a test for nouns (Vossen 2002, p. 18):

> in any sentence S where Noun1 is the head of an NP which is used to identify an entity in discourse another noun Noun2 which is a synonym of Noun1 can be used as the head of the same NP without resulting in semantic anomaly. And vice versa for Noun2 and Noun1.

It can be difficult to evaluate the equality of word denotations. It *is* difficult for highly abstract nouns and for a wide range of verbs. Vossen's *semantic anomaly* can lead to conditions on synonymy so weak that too many words are treated as synonymous. Semantic anomaly can also be absent because of a kind of textual entailment between both variants of the sentence. Synonymy can go across linguistic boundaries such as style, register or even part of speech; for the latter, a separate subtype of synonymy has been introduced in *EuroWordNet*. Significantly, the definition plays up a clear distinction between synonymy and other relations. Synonymy cannot occur in parallel with other relations for the same words.

We propose to extend this observation. Synonymy cannot be redundant and it cannot contradict other relations: two words (two lexical units, to be precise) are synonymous only if they show very similar patterns of lexico-semantic relations. We will elaborate on this idea later in this paper.

Vossen (2002) presents a wordnet as a linguistic ontology which describes concepts lexicalised in language, paying attention to detailed distinctions between fine-grained concepts. Tufiş et al. (2004, p. 10) perceive a wordnet as *a lexical-semantic network* whose nodes are synsets:

> the nodes of which represented sets of actual words of English sharing (in certain contexts) a common meaning.[1]

Miller et al. (1993) also presented synonymy as "a continuum along which similarity of meaning can be graded" and noted that only words which express mutual, equal semantic similarity can be included in one synset. Still, they refer to the rule of exchangeability of words in a context as the only means of measuring the

[1] In general, nodes in semantic networks may be labelled with abstract names, not necessarily with valid natural language tokens.

degree of semantic similarity. Borin and Forsberg (2010) based the construction of synsets for Swedish on a measure of semantic similarity among words acquired from native speakers. There is a general assumption about word synonymy and about assigning words to synsets: decisions are finely graded rather than binary. This is an attractive and realistic perspective, but it requires extensive experimental research and the participation of many language users. An alternative source of lexical knowledge can, to some degree, be automated extraction of semantic relatedness from large corpora (Piasecki et al. 2009).

1.3 Derivation and wordnets

There are other reasons, less pronounced and less universal, why the synset may not be the most felicitous choice of the bottom-most node for a wordnet. Some of those reasons are to do with the "anglocentrism" of wordnets, whose design is (naturally) deeply influenced by PWN and, to a rather high degree, by the peculiarities of English, despite a 15-year tradition of developing wordnets for other languages. In Slavic languages—the area of our immediate concern—even various inflectional forms may have different connections, whereas various derivational forms almost inevitably enter into lexical relations perhaps less central to wordnets.

Derivational phenomena *have* been tackled in PWN[2] and in EWN. EWN considers a range of *cross-part-of-speech* lexico-semantic relations (Vossen 2002). Raw derivational association of a pair of word forms is recorded in a *derived*-type relation; Vossen (2002, p. 20) also recommends that the pair be added to "some other semantic relation". Derivational pairs occur in three relations: *cross-part-of-speech* synonymy, *be-in-state/state-of* and *involved/role*; examples of the last of these relations are given for four of its eight sub-types.

All such measures notwithstanding, derivational phenomena have not been prominent in research on wordnet-building. In Slavic languages, derivational relations tend to be explicitly marked by a rich system of morphological affixes. The regularities observed at the level of word forms have lent increased importance to the description of derivational relations, for example, in wordnets for Czech (Pala and Smrž 2004), Bulgarian (Koeva et al. 2004) or Russian (Azarova 2008). The focus is gradually shifting from a systematic but simple record of derivational instances, as in Czech WordNet, to a semantic classification, as in *plWordNet* (Piasecki et al. 2010). Most derivational relations are shared with those introduced in EWN, some are even present in the less derivationally "developed" English,[3] but few are explicitly recorded in wordnets. The main difference is the change of status from a semantically secondary formal phenomenon to an important mechanism in the lexical-semantic system embodied by a wordnet. Derivational relations hold among lexical units and their word forms, so they cannot be described at the level of synsets.

This paper revisits the idea of synsets as the smallest building blocks in a wordnet structure, and defines the fundamental structural elements of a wordnet in a way

[2] See detailed studies in Miller and Fellbaum (2003), Fellbaum et al. (2009), not yet fully applied to PWN.

[3] *Femininity*, as in *actor–actress*, is a representative example.

which combines two perspectives. One perspective focusses on concept-sharing among elements of the lexicon; the other is grounded in the linguistic tradition of describing the lexicon as a system.

First, we will propose to promote the *lexical unit* to the role of the basic structural element of a wordnet, and discuss the benefits of such a decision. Next, we will analyse the consequences of the primary role of the lexical unit. We will consider both the theoretical and the practical aspect of the matter. Is a system based on lexical units linguistically more justified than a system based on synsets? Are lexical units easier to enter into a (growing) wordnet? The latter point will be illustrated by our experience with the construction of a Polish wordnet.

2 Lexical unit as the basic building block of a wordnet

We have proposed and implemented in *plWordNet* (Piasecki et al. 2009) a granularity radically different than that of a synset.[4] The nodes in the network are, for all practical purposes, lexemes, but we refer to them as *lexical units*[5] (henceforth LUs) to avoid the controversial variety of accounts for the notion of lexeme.

The idea of the LU as the centrepiece of a wordnet first arose in the practice of wordnet-building. We have found that it is equally hard to define synsets via synonymy and synonymy via synsets. We sought a manner of definition which would allow guidelines for lexicographers to be precise enough to support consistent editing decisions. The idea appears even more attractive if we consider—as pointed out in the previous section—that synonymy, hypernymy, meronymy and an assortment of other lexical relations all hold among LUs.

2.1 Constitutive wordnet relations

Lexico-semantic relations form a continuum of semantic distinctions. Their description can be easily developed down to the finest granularity of relations specific to individual pairs of LUs. Relations established in linguistics, such as hypernymy or meronymy, are based on subspaces of the continuum with fuzzy borders. Depending on the relation type, linguists agree to a varying degree on classifying word pairs as relation instances. For example, one can expect much higher agreement on hypernymy than on meronymy, even considering just one specific meronymy subtype. Even if we set problematic synonymy aside, we can perceive a wordnet as a *generalisation* of that relation continuum, with few distinctions preserved and most subtle distinctions de-emphasised. This arbitrarily-imposed coarser granularity is, at the same time, an advantage of wordnets and their drawback—if only a detailed, formally complete semantic lexicon can be available. The reality of defining wordnet relations is shaped by three concerns: that a wordnet be

[4] *plWordNet* is the largest Polish wordnet. Under construction since October 2005, in August 2012 it has reached the size of ≈150,000 lexical units and ≈110,000 synsets.

[5] Technically, we define lexical units as lemma-sense pairs, where sense is represented by a number. We assume that one lexical unit belongs to exactly one synset and the synset represents the sense.

1. suitable for the *construction* of generalisations,
2. suitable for the *application* of generalisations in NLP tasks,
3. *compatible* with other wordnets.

The last concern, clearly quite down-to-earth, acknowledges the status of wordnets as *de facto* standard lexical resources, and emphasises the importance of inter-wordnet multilingual structures—see (Vossen 2002; Vossen et al. 2008).

It is not quite feasible to perform a complete analytical assessment of the fitness of a wordnet as a generalised description of the lexico-semantic system of a natural language. At best, there can be an ongoing verification and validation in NLP tasks, given that wordnets are incessantly put to practical tests. There is a close relation between knowledge representation, notably ontologies, and the lexical system, perhaps particularly close in English.[6] Thus, what one expects of a wordnet is naturally shaped by the established paradigms of knowledge representation.

We assume, a little arbitrarily, that linguistic tradition makes wordnet-building more consistent.[7] Such tradition should inform the choice of relations, ensure that they are closely tied to language data, and guide verification. In particular, one should leverage existing linguistic resources, beginning with large unilingual dictionaries. There is perhaps a surfeit of theories of meaning. It would not do for a wordnet to favour any of them. We posit a *minimal commitment* principle: construct a wordnet with as few assumptions as possible. Such system simplicity becomes an advantage—little must be assumed to create even a very large wordnet.

Princeton *WordNet* has been pivotal in thousands of applications. Its popularity is perhaps due in equal measure to the coverage of the vocabulary and to the underlying system of lexico-semantic relations. It is not feasible to capture all of a natural language's lexical system, but the PWN project has been an eminently successful compromise between the expressive power of such a system's description and the workload required to construct that description. It is not our intention to come up with a different structural principle for new wordnets. We only aim for theoretical clarity in explaining wordnet structure and for practical gains in consistency during wordnet construction.

We have argued earlier in the paper that synonymy can be hard to define in a manner which supports the consistency of wordnet editors' decisions. On the other hand, it is the synset that every wordnet user expects. Applications have come to assume implicitly that hypernymy puts *synsets* into a hierarchy. A way out of the synset-synonymy circularity may be a definition of the synset which avoids synonymy altogether. In any case, perfect synonymy is exceedingly rare in natural languages. We expect, therefore, that synsets too express much less than near-identity of the underlying meaning. There is, we assume, a form of feature sharing among LUs, a generalisation over unavoidable specific differences between them. In keeping with the minimal-commitment principle, we also aim to determine synset membership via other relations already noted in the wordnet.

[6] Most existing ontologies, wordnets included, turn to English for labels for concepts, relations, attributes, values and so on. This tends to blur very strongly the distinction between formal abstractions and expressions in natural language.

[7] This may be so because—in the end—it is linguists who make a new wordnet happen.

We propose that, to belong to the same synset, LUs should share instances of a carefully selected subset of the relations defined in a wordnet. That is, a synset comprises those LUs which share a set of lexico-semantic relation targets. In effect, to say that synsets S_1 and S_2 are linked by relation R is to say that any pair of LUs s_1 and s_2, such that $s_1 \in S_1$ and $s_2 \in S_2$, is an instance of R. So, relations which link synsets in a wordnet can be perceived as derived from lexico-semantic relations. A synset can thus be defined principally via those relations in which its elements participate.[8]

By way of illustration, let us consider the synset {*miłość 1* 'love', *serce 6* '≈ love (*lit.* heart)', *uczucie 3* '(positive) emotion', *afekt 1* 'affection'}.[9] The synset is a hypernym of {*uwielbienie 1* 'adulation', *adoracja 2* 'adoration'}: *uwielbienie 1* is a kind of *miłość 1* and so is *adoracja 2*; *uwielbienie 1* is a kind of *afekt 1*; and so on for every pair.[10]

Thus, in order to define synsets, we need a set of lexico-semantic relations *well-established* in linguistics, *definable* with sufficient specificity and *useful in generalisation*.

Synsets and their interconnections are the centre of a wordnet from the point of view of applications. We will refer as *constitutive relations* to those relations upon which the definition of synsets can be based. Such constitutive relations are what turns a set of words into a wordnet. One can conceive of a constitutive relation R as a synset relation such that $R(s_1, s_2)$ for each member s_1 of a synset S_1 and each member s_2 of a synset S_2.

2.2 The quest for constitutive relations

We concern ourselves with those lexico-semantic relations which are well-established in linguistics. This allows us to base wordnet-building on good understanding of those relations and on existing descriptions, and promises better consistency among wordnet editors.

Research in linguistics has suggested paradigmatic relations with a central position in structuring the vocabulary of a language. Four types of relations appear to be especially important: synonymy, hyponymy / hypernymy, antonymy and meronymy / holonymy (Murphy 2010, pp. 109, 122–123), (Stede 1999, pp. 86–87), (Painter 2001, p. 80), (Collinge 1990, pp. 84–85). There are variations. Some authors do not include meronymy among such central relations (McCarthy 2003, p. 16), (Yule 2010, pp. 116–119). Others add relations, for example entailment and presupposition for verbs (Pustejovsky 2001, pp. 23–24). Whether a particular relation should be considered is a difficult decision, because there are no universal lexicographic criteria. It is obvious that paradigmatic relations vary in language

[8] The Appendix presents a formalisation of our idea of relation sharing among LUs as the reason for forming synsets.

[9] We assume that all those are *strong feelings*.

[10] Predictably, the minimal-commitment approach to synset definition was met with a challenge. The language data processed during the construction of *plWordNet* have dictated an expansion: we had to cope, among others, with semantic opposition resulting from differences in stylistic registers or from the effect of semantic verb classes and aspect. We will come back to these issues in Sections 4, 5. For now, they will stay in the background, so we can keep the presentation simple.

(Cruse 2004, p. 143). Among the attempts to put semantic relations on a firm footing, one of the finest proposals resorts to set theory. That point of view distinguishes paradigmatic relations of identity (synonymy), inclusion (hyponymy and meronymy) and exclusion: opposition (antonymy)[11] and incompatibility (co-hyponymy, co-meronymy) (Cruse 2004, pp. 148–168).

The linguistic paradigmatic relations which we have just listed are present in all wordnets. To be useful for generalisation, constitutive wordnet relations should be frequent and should describe sets of LUs systematically. This is true of most of the paradigmatic relations, with a notable exception of antonymy, which is seldom used to link synsets among wordnets.

We have named several lexico-semantic relations as likely constitutive relations in a wordnet—relations which define synsets. We will now examine them more closely, keeping in mind the concerns postulated in Section 2.1, wordnet practice, and the solutions adopted in *plWordNet*.

While wordnets follow the blueprint of Princeton *WordNet*, there are always many small and large changes. A distinguishing feature is usually how synsets are interlinked by *synset relations*.[12]

Synset relations determine a wordnet's basic structure. We assume that a synset effectively arises from the sharing of relation targets by certain LUs—considered to be this synset's members. That is why synset relations are the key factor in shaping the wordnet's ability to generalise over properties of individual LUs. The granularity and systematicity of the distinctions between LUs is determined by which synset relations are selected for a wordnet.

The verb LUs *roztłuc*, *rozbić*, *stłuc*, *zbić* 'smash$_{pf}$ (a bottle, a glass, a vase)' and *rozdeptać* 'squash$_{pf}$ with a foot (a worm, a spider)' are all the subordinates of *zniszczyć* 'destroy$_{pf}$'. If only hyponymy were available ($X \rightarrow zniszczyć$), we would merge the five LUs into one synset, because their connections would be indistinguishable in the net. In *plWordNet*, the *cause* relation links the first four LUs to the intransitive verb *stłuc się* 'break$_{pf}$' (smashing causes something to break), whereas *rozdeptać* is a holonym of *deptać* 'tread$_{impf}$' (to squash with a foot is to destroy something by treading). We thus construct two sets of synonyms, {*roztłuc*, *rozbić*, *stłuc*, *zbić*} and {*rozdeptać*}, in keeping with the linguistic intuition.

The discussion so far, in particular the three concerns about wordnet relations, suggests that the constitutive wordnet relations fit the bill. Wordnet developers can manipulate the level of generalisation by changing the set of synset relations.

2.2.1 Nouns

Let us focus on nouns for a while. Most wordnets appear to choose only a few relations to act as constitutive wordnet relations: hyponymy / hypernymy, meronymy / holonymy and synonymy (Miller et al. 1990; Vossen 2002; Hamp and Feldweg 1997;

[11] We use the term *antonymy* in a broader sense. It covers *complementaries*, *proper antonyms*, *reversives* and *converses* (Cruse 2004, pp. 164–168).

[12] They are often called *conceptual relations*, a term which we prefer to avoid—along with whatever implicit assumptions may underlie whatever theories of meaning.

Table 1 Frequency of wordnet relation instances in Princeton *WordNet* 3.1

Rank	Nouns	Rank	Verbs
1	**Hyponymy, Hypernymy (45.5 %)**	1	Derivationally related form (55.4 %)
2	Derivationally related form (22.4 %)	2	**Troponym, Hypernym (31.7 %)**
3	**Meronymy, Holonymy (13.3 %)**	3	**Verb group (4.2 %)**
4	Member of domain and domain of synset (9.1 %)	4	Member of domain (3.0 %)
5	**Type/instance (5.1 %)**	5	Antonymy (2.6 %)
6	Pertainymy (2.9 %)	6	**Also see** (1.4 %)
7	Antonymy (1.3 %)	7	**Entailment (1.0 %)**
8	**Attribute (0.4 %)**	8	**Cause (0.5 %)**
		9	Participle (0.2 %)

For nouns, we count only instances of meronymy, because holonymy and meronymy are mutual inverses

Koeva et al. 2004; Pedersen et al. 2009; Piasecki et al. 2009). Miller (1998, p. 40) calls all of them except synonymy "fundamental organizing relations". A similar picture can be found in *GermaNet* (Hamp and Feldweg 1997). All these relations are well-established in linguistics (see Section 2.1) and are frequent—see the PWN statistics in Table 1.[13] EWN adds cross-categorial relations.[14] Most of them can be perceived as constitutive, and they play an important role in distinguishing co-hyponyms (Vossen 1998, pp. 102–103). XPOS near-synonymy and XPOS antonymy, however, are often a practical tool rather than theoretically sound semantic relations (Vossen 1998, p. 105). We propose to perceive a synset as a group of words with analogous positions in a network of few, well-defined relations. A synset is, therefore, a kind of an equivalence class of LUs over synset relations. The Appendix develops this idea in a formalised way. Because synsets represent synonymy, synonymy can be reduced to the other synset relations.

The nouns *lustro* and *zwierciadło* both denote a mirror; the latter is a literary word. Both LUs are hypernyms of *lustro weneckie* 'Venetian mirror' and *tremo* 'trumeau mirror, pier glass, pier mirror'. It is natural to see *lustro* and *zwierciadło* as objects, so both are the hyponyms of *przedmiot* 'object'. Next, *szkło* 'glass' is a meronym of *lustro* and of *zwierciadło*—both objects can be made of glass. Such relation-sharing allows us to determine that *lustro* and *zwierciadło* are synonyms in Polish, and to put them into one synset.

The linguistic literature tends to treat antonymy as a basic lexico-semantic relation (see Sect. 2). Antonymy is very seldom *shared* among groups of LUs.

[13] The names of PWN relations follow the terminology listed at wordnet.princeton.edu/wordnet/man/wninput.5WN.html.

[14] Those are XPOS near-synonymy, XPOS antonymy, XPOS hypernymy, state of / be in state (noun-adjective), involved/role (verb-noun, noun-noun and variants with adverbs) (Vossen 1998). "XPOS" is meant to denote relations "across parts of speech".

Given a pair of antonyms, LUs closely semantically related to them need not be antonymous, either among themselves and in relation to the given pair. We can say that antonymy has a very low *sharing factor*, to be measured by the average size of the LU group which shares the given relation; derivational relations also have a low sharing factor. That is why antonymy is mostly described as a relation between LUs—in PWN (Miller et al. 1990; Fellbaum 1998b), in EWN (Vossen 2002, p. 24), in *GermaNet* (Hamp and Feldweg 1997), and so on. In EWN and wordnets originating from it, e.g., (Koeva et al. 2004), a special *near-antonymy* relation enables the transfer of meaning opposition to synsets—groups of LUs. Yet, EWN does not define near-antonymy directly and precisely.

2.2.2 Verbs

Sets of verbal synset relations differ across wordnets, but we can notice that they refer to a shared set of semantic associations and the differences result mainly from different partitioning of this set. Fellbaum (1998b, pp. 76–88, 220–223) describes these verbal relations:

1. *synonymy*—mutual entailment, *relation between word forms* (Miller et al. 1990, pp. 242–243),
2. *antonymy—lexical relation between word forms* (*ibid.*),
3. *inclusive entailment* (or entailment with *proper inclusion*, resembling meronymy),
4. *troponymy—coextensiveness*, instead of verbal hyponymy,
5. *cause*,
6. *presupposition*.

In practice, presupposition and proper inclusion were combined into the *entailment* relation (at least from PWN 1.5 onwards), but its frequency is still low (Table 1). The relation set in PWN 3.1 includes the assignment of nominal and verbal synsets to domains, and the grouping of verbal synsets according to the similarity of their senses. The former is similar to the classification according to stylistic registers (this will be discussed in Section 3), while the definition of the latter is too vague to analyse it as a potential constitutive relation.

Troponymy—"a manner relation" (Fellbaum 1998a, p. 213)—is described by the test "*to* V_1 *is to* V_2 *in some* (*particular*) *manner*"[15] (Fellbaum 1998b, p. 79, 285). Fellbaum's troponymy resembles hyponymy (Fellbaum 1998b, pp. 79–80).[16] Fellbaum denies the identity of nominal and verbal hyponymy on the grounds of incompatibility of nominal and verbal testing expressions and elementary

[15] For verbal hyponymy, Lyons (1977) proposed a similar test [*to X is to Y in certain way*—see also (Murphy 2003, p. 222)]. Cruse (1997) proposes to test verbal hyponymy via nominalization: *Xing is a way of Ying*, see also (Murphy 2010, pp. 116–117).

[16] The classical analytical definition stipulates that the hypernym play a vital role as a head of *definiens* (Geeraerts 2010, p. 83). In the Katzian model of hyponymy (Katz and Fodor 1963) a definition of hyponym (of any word class) includes a definition of hypernym (of the same word class) (Geeraerts 2010, pp. 105, 111). The idea of identifying a head of verbal definition with hypernym is known to the EWN authors (Vossen 1998, p. 100).

differences between semantic structure of verbs and nouns, but at the same time she emphasises the similarity of the two.[17]

GermaNet's verbal relations follow those of PWN with two exceptions: (verbal) hyponymy occurs in place of troponymy (Kunze 1999) and *subevent* relation is different from entailment. The resultative relation (*toeten* 'to kill'—*sterben* 'to die') is called a *causal/causation relation*[18] (Kunze and Lemnitzer 2010, p. 166). Meronymy remains limited to nouns, and for verbs a *subevent* relation is used, "which replaces the entailment relation of a former specification" (Kunze, 1999).[19]

EWN includes all *GermaNet* relations (Vossen 1998, p. 94) with verbal *hyponymy* and *subevent relation* ("meronymy", proper inclusion of PWN[20]). The *cause* relation is defined less strictly than in PWN.[21] The system is extended with near-synonymy (close co-hyponyms but not synonyms—a synset relation), cross-categorial relations (synonymy, antonymy and hypernymy), and near-antonymy (vague opposition) in a similar way to EWN nominal relations. EWN's system is much more elaborate than PWN's, while *GermaNet* stands between these two, but they all share the main types of lexico-semantic associations as the basis. Every system includes constitutive relations which represent hyponymy, cause and various types of entailment.

To sum up: verbal synset relations in wordnets are located in the similar subspaces of the semantic relation continuum, and are mainly based on the common properties of various forms of entailment and troponymy/hyponymy. The latter is the second most frequent (Table 1). The other relations—relatively frequent if counted together—are crucial in determining semantically motivated groupings of verbal LUs. Thus all such relations can be used as constitutive wordnet relations. That, to some degree, is the case of *plWordNet*.

3 The case of *plWordNet*

The expansion of *plWordNet* with new LUs is based on the idea of topological identity of synonyms in a complex net of words. The idea of synonymy has evolved since the première of *plWordNet* 1.0. Piasecki et al. (2009, p. 25) define the synset as a set of LUs which share *central lexico-semantic relations*: hypernymy,

[17] She entitled a paragraph devoted to troponymy as *Hyponymy among verbs*; in a few cases troponyms were called "*verb hyponyms*" in quotation marks (Fellbaum 1998b, par. 3.3.1.1).

[18] "A verbal predicate causes an event, a transition or a result, i.e. another verb, or a noun or an adjective" (Kunze 1999).

[19] *GermaNet* employed PWN's idea of entailment with one modification. Two types of entailment—"meronymic" and presuppositive—are different phenomena, but these two cases are quite distinct from each other, justifying two different relations in *GermaNet*. The relation of entailment is kept for the case of backward presupposition. Following a suggestion made in EuroWordNet (Alonge 1996, p. 43), we distinguish temporal inclusion by its characteristics that the first event is always a subevent of the second, and thus the relation is called "subevent relation" (Hamp and Feldweg 1997).

[20] Vossen also equates Fellbaum's proper inclusion entailment with *subevent*.

[21] The events need not be temporally disjoint as in PWN. It also captures presupposition (Vossen 1998, p. 109).

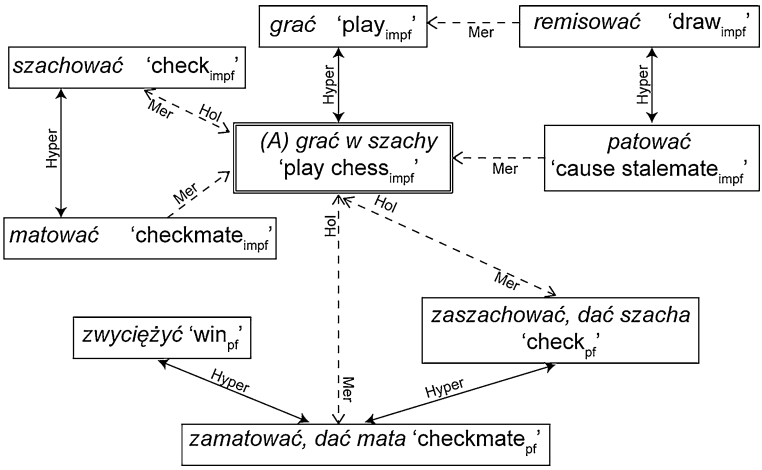


Fig. 1 Chess-playing in *plWordNet*

meronymy and holonymy. They are among the relations which we now call *constitutive*.

Most of *plWordNet*'s structure centres on hyponymy / hypernymy and on meronymy / holonymy, and fairly complex subgraphs are possible. For example, Fig. 1 shows a group of verbs related to chess: *szachować* 'check$_{impf}$', *zaszachować*, *dać szacha* 'check$_{pf}$', *matować* 'checkmate$_{impf}$', *zamatować*, *dać mata* 'checkmate$_{pf}$', *patować* 'cause a stalemate$_{impf}$'. In *plWordNet*, verbs are mainly differentiated by means of hyponymy/hypernymy and meronymy/holonymy—well enough to distinguish between most of them. All those verbs are involved in relations with a central holonym—*grać w szachy* 'play chess$_{impf}$', but they have different hypernyms. *Matować* 'checkmate$_{impf}$' has hypernyms *szachować* 'check$_{impf}$' and *zwyciężać* 'win$_{impf}$', perfective *zamatować* 'checkmate$_{pf}$' has perfective hypernyms *zaszachować* 'check$_{pf}$', *zwyciężyć* 'win$_{pf}$'. *Patować* 'cause a stalemate$_{impf}$' has a hypernym *remisować* 'draw$_{impf}$'. Both *szachować* and *zaszachować* have their own hypernyms not shown in Fig. 1. Because LUs *zamatować*, *dać mata* are involved in the same relations, they belong to the same equivalence class / to the same synset; similarly *zaszachować*, *dać szacha* are wordnet synonyms, because they share constitutive relations.

Our "topology-based" definition of the synset is supported by a specialised wordnet editor, the *WordnetLoom*, constructed for *plWordNet*. Every editing decision is preceded by the presentation of substitution tests defined for a given relation and instantiated by lemma pairs taken from two synsets under consideration. The editor can select only a subset of pairs, or even skip this step. A detailed analysis of many relation instances can be time-consuming. As a compromise, substitution tests for synonymy are also included in the *plWordNet* editor guidelines. Experienced editors can create or modify synsets without laborious tests. The final form of the definition (which may later be reviewed by the project's senior lexicographers) is the one based on relation types. The editors' work is assessed only in relation to the topology-based definition.

The *plWordNet* development environment, including *WordnetLoom*, takes the editors through the following steps when they put a new LU into *plWordNet*:

- present the user with a lemma list based on corpus frequency;
- present lemma usage examples split into sense clusters by word-sense disambiguation (Broda et al. 2010; Broda and Mazur 2011);
- present a measure of semantic relatedness between lemmas (for now, nouns and adjectives) (Piasecki et al. 2007)—this suggests potential synonyms, hyponyms, antonyms;
- suggest links to the given LU using the *WordnetWeaver* algorithm (Piasecki et al. 2012);
- check meanings in contemporary Polish dictionaries—for example, (Dubisz 2004; Bańko 2000)—encyclopaedias and Polish *Wikipedia*;
- adjust the structure of *plWordNet*, if needed—the user has this option;
- apply substitution tests to the LU, to reveal and verify possible connections to the lexical net;
- add the LU to *plWordNet* and link it to other LUs with relations;
- determine which LUs share the same constitutive relations—they are considered synonymous.

Consider the verb lemma *kąsać*$_{impf}$ 'bite', 'nip' (also about wind or cold), 'sting' (about insects). We start with automatically-generated and disambiguated usage examples, grouped under several meaning labels:

- (1) 'bite using teeth' (about animals) "(Małpy) [c]iągnęły go za włosy, **kąsały** w uszy" 'The apes pulled his hair and bit his ears'.
- (2) 'sting' (about insects) "Część niebezpiecznych owadów przedostała się już do sanatorium i **kąsają**" 'Some of the dangerous insects have already penetrated into the sanatorium and are stinging'.
- (3) 'sting, nip' (about cold, wind etc.) "mróz **kąsał** stopy" 'the cold was stinging the feet'.
- (4) 'be spiteful' (about people) "To, że są uprzejmi, nie znaczy, iż nie potrafią **kąsać**" 'That they are polite does not mean that they cannot bite'.

Next, *WordnetWeaver* generates five link proposals:

- (a) {*doskwierać*$_{impf}$ *1*, ... 'cause$_{impf}$ pain, nuisance, suffering'},
- (b) {*gryźć 2* 'bite$_{impf}$, chew$_{impf}$'},
- (c) {*ugryźć 1* 'bite$_{pf}$ into (causing wounds)'},
- (d) {*żądlić 1* 'sting$_{impf}$'},
- (e) {*ciąć 1*, *ucinać 1* 'bite$_{impf}$, sting$_{impf}$'}.

Dubisz (2004) gives these descriptions of the verb *kąsać*:

- (I) *kaleczyć zębami, ciąć żądłem; gryźć* 'injure using teeth, sting';
- (II) *o mrozie, zimie, wietrze: szczypać, powodować ból* 'about cold, winter, wind: pinch, cause pain';
- (III) *dokuczać, dręczyć* '(about malicious people or about troubles) torment'.

The three resources can be easily compared, with the following five sets of connections: (1 = b + c ≈ I), (2 = d + e ≈ I), (3 = II ≈ a), (4 ≈ III), (a ≈ III).

With all that background information, the user distinguishes five LUs:

- *kąsać 1* is acknowledged as a synonym of *gryźć 1* ' (about an animal) to bite$_{impf}$ using teeth and causing wounds' (the *Wordnet Weaver* suggested the perfective variant *ugryźć 1*)—see (c), (1) and (I);
- *kąsać 2* '(of weather conditions) bite, nip'—see (3) and (II), and there is an association with (a);
- *kąsać 3* is semantically connected with *ciąć 1*, *ucinać 3* ' (about insects) bite, sting'—see (d), (e), (2) and (I);[22]
- *kąsać 4* '(about worries) trouble'—see (a) and (III);
- *kąsać 5* 'be spiteful'—see (4) and (III).

Figure 2 (i) presents the neighbourhood of *kąsać 1* and *kąsać 3*. They are hyponyms of the same LU *kaleczyć* 'cut$_{impf}$ (up), injure$_{impf}$', distinguished from each other by a hyponym of *kąsać 3*, which is *żądlić 1* 'cut the skin with a sting'. *Żądlić* is also a hyponym of two LUs: *ciąć 1* and *ucinać 3*, both hyponyms of *kaleczyć*. The same set of constitutive relations for *kąsać 3*, *ciąć 1* and *ucinać 3* signals synonymy. Each instance of hyponymy passed *plWordNet*'s substitution tests.

Figure 2 (ii) shows that *kąsać 2* and *kąsać 4* are closely semantically related. In fact they are co-hyponyms of the same hypernym set {*doskwierać 1*, ... 'cause$_{impf}$ suffering'}. *Kąsać 2* refers to weather conditions and physical pain, *kąsać 4* to concerns, worries and mental suffering. They are not synonyms, because they are differentiated by cause relations: *kąsać 2* → *marznąć 2* '(about a man or animal) become$_{impf}$ cold' and *kąsać 4* → *martwić się 2* 'worry (intransitive)'. We do not show all six synonyms of *doskwierać 1*, but substitution tests confirmed that relations between *kąsać 2*, *kąsać 4* and all six LUs do hold.

The user attached *kąsać 5* 'be spiteful' to two synonymous hypernyms *szkodzić* 'act malevolently' and (more formal) *działać w złej wierze* 'act in bad faith'—see Fig. 2 (iii). Let us present substitution tests for the two instances of hyponymy.

Kąsać 5 and *szkodzić 1*

Jeśli ***kąsa***, *to* ***szkodzi*** 'If (he) is spiteful, then (he) acts malevolently'

Jeśli ***szkodzi***, *to niekoniecznie* ***kąsa*** 'If (he) acts malevolently, then (he) need not be spiteful'

Kąsać *to* ***szkodzić*** *w specjalny sposób* 'To be spiteful is to act malevolently in a special way'

Kąsać 5 and *działać w złej wierze 1*

Jeśli ***kąsa***, *to* ***działa w złej wierze*** 'If (he) is spiteful, then (he) acts malevolently'

Jeśli ***działa w złej wierze***, *to niekoniecznie* ***kąsa*** 'If (he) acts malevolently, then (he) need not be spiteful'

[22] These words are, in fact, synonymous, as we will explain shortly.

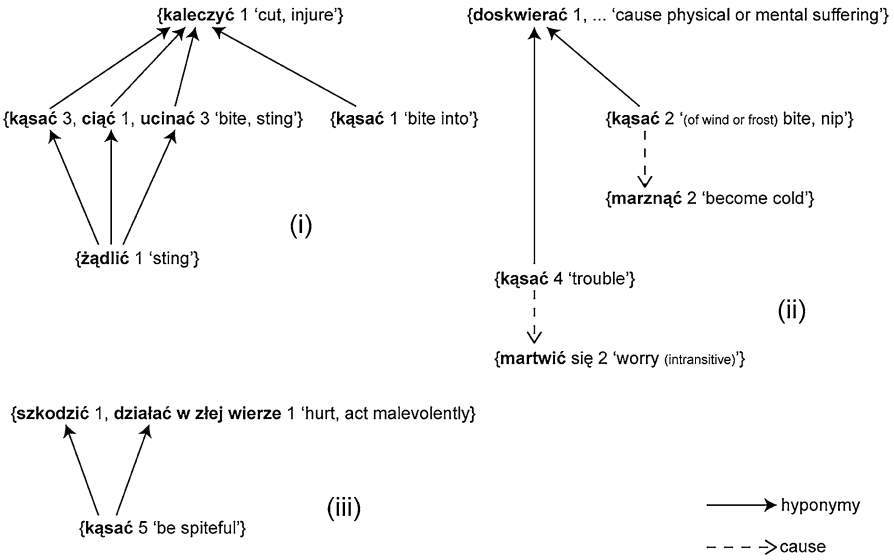


Fig. 2 (*i*) *Kąsać 1* and *kąsać 3* in *plWordNet* and their topological neighbourhood. (*ii*) Differentiation of *kąsać 2* and *kąsać 4* by cause relation. (*iii*) *Kąsać 5* as a hyponym of two LUs from the same synset

> ***Kąsać*** *to* ***działać w złej wierze*** *w specjalny sposób* 'To be spiteful is to act malevolently in a special way'

Naturally, to prove synonymy of *szkodzić* and *działać w złej wierze* we should check all relations in which the two are involved. Indeed, they both have more hyponyms and common hypernyms, not shown in Fig. 2 (iii).

3.1 *plWordNet* relation statistics

Statistical data have influenced the choice of constitutive relations for *plWordNet*. Frequently occurring relations can substantially affect the shape of a wordnet, while those much less frequent may not be conducive to maintaining homogeneity. Hyponymy, hypernymy, meronymy and holonymy are "popular": together they account for 48.4 % of wordnet relations among nouns and 30.1 % among verbs. Table 2 shows the details for *plWordNet* 1.6.

If we rule out derivational relations and inter-register synonymy (it is secondary in our model, as is synonymy; see Table 2 and a discussion in Section 4), it will appear that just a handful of remaining relations (shown in bold) can be considered constitutive.

Tables 3 and 4 compare *plWordNet* 1.6 with two Polish monolingual dictionaries, edited by Dubisz (2004) and Bańko (2010). The former, the *Great Dictionary of Synonymy* (GDS), is a dictionary of synonyms, antonyms, hyponyms/hypernyms and meronyms/holonyms. The latter, the *Universal Dictionary of Polish* (UDP), is a basic contemporary dictionary of Polish. We collected random samples of LUs in the two dictionaries and checked their relations. In GDS we counted links

Table 2 Frequency of wordnet relation instances in *plWordNet* 1.6

Rank	Nouns	Rank	Verbs
1	Derivational relations (38.5 %)	1	Derivational relations (except aspectuality) (41.7 %)
2	**Hyponymy, Hypernymy (37.6 %)**	2	**Hyponymy, Hypernymy (26.5 %)**
3	**Meronymy, Holonymy (10.8 %)**	3	Aspectuality (14.0 %)
4	Fuzzynymy (6.5 %)	4	**Meronymy, Holonymy (3.6 %)**
5	Antonymy, Converseness (3.1 %)	5	Antonymy, Converseness (3.0 %)
6	Inter-register synonymy (2.7 %)	6	Inter-register synonymy (2.4 %)
7	Other (0.8 %)	7	Fuzzynymy (2.1 %)
		8	**Causality (2.0 %)**
		9	**Processuality (0.8 %)**
		10	**State (0.1 %)**
		11	Other (1.9 %)

For nouns, we count both meronymy and holonymy, because in *plWordNet* the relations are not mutual inverses

Table 3 Frequency of verbal semantic relations in the UDP

Rank	UDP—verbs
1	**Hyponymy, hypernymy (51.6 %)**
2	Aspectuality (12.9 %)
3	**Meronymy, holonymy (9.8 %)**
4	Other derivational relations (7.7 %)
5	**State (6.7 %)**
6	**Processuality (5.2 %)**
7	**Causality (3.1 %)**
8	Inter-register synonymy (1.6 %)
9	Antonymy (1.0 %)
10	Other (0.5 %)
size	237

Size = sample size in LUs

of particular entries.[23] In UDP we worked only on definitions; we analysed the meaning of verbs in the definitions and assigned *plWordNet* relations to those verbs.[24] GDS overrepresents antonymy. In the more typical UDP, antonymy makes up $\approx$1.0 % of all relations.

[23] The dictionary consists of lemmas and their relational links. It can be seen as a "wordnet on paper", as the Editor has called it in his correspondence.

[24] UDP also has links but we chose to analyse only the text of definitions.

Table 4 Semantic relations in (Bańko 2010)

Rank	Nouns	Rank	Verbs
1	**Hyponymy**, **Hypernymy (75.6 %)**	1	**Hyponymy**, **Hypernymy (88.8 %)**
2	**Meronymy**, **Holonymy (17.8 %)**	2	Antonymy (11.2 %)
3	Antonymy (6.6 %)		
Size	816		523

Size = sample size in LUs

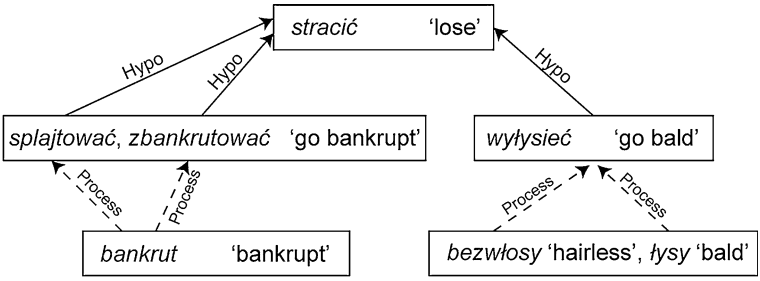


Fig. 3 Processuality as a constitutive relation

Verbal and nominal relations differ non-trivially. Nominal hyponymy and hypernymy are better defined, and more widespread. They account for 37.6 % of nominal and 26.5 % of verbal relations in *plWordNet*. Hyponymy and hypernymy make up 51.6 % of relations among verbs in UDP. It is similar for meronymy and holonymy. Meronymy is much harder to define for verbs than for nouns. Relation frequencies show that meronymy and holonymy are more popular for nouns (10.8 % in *plWordNet*, 17.8 % in GDS) than for verbs (3.6 % in *plWordNet*, 9.8 % in UDP, *none* in GDS).

It was necessary to supplement the list of constitutive verbal relations in order to make the system more efficient in differentiating verb LUs which otherwise would be grouped, unintuitively, in the same synsets. Apart from derivational relations, few lexico-semantic relations have been added: causality (2.0 % in *plWordNet*, 3.1 % in UDP), processuality (0.8, 5.2 %), state (0.1, 6.7 %), inchoativity (0.4, 0.0 %), presupposition and preceding (0.4, 0.5 %); most of them are clones of relations in PWN and EWN.[25] Together they add up to 4.0 % (*plWordNet*) or 15.5 % (UDP) of the total number of relations.

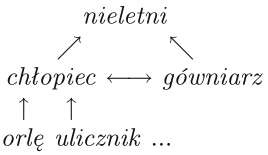
The main function of the six relations is to differentiate co-hyponyms. Verbs with identical hyponymy/hypernymy and meronymy/holonymy links belong in the same synset. Hyponymy/hypernymy and meronymy/holonymy are often insufficient to separate verbs which native speakers would never consider synonyms; see Fig. 3 for an illustration. The verbs *wyłysieć* 'go bald$_{pf}$', *zbankrutować* 'go bankrupt$_{pf}$' are

[25] Those are presupposition, cause, state and troponymy/subevent/meronymy.

hyponyms of *stracić* 'lose$_{pf}$'; they have no hyponyms, meronyms or holonyms. If processuality were not a verbal constitutive relation, these words—most unintuitively!—would have to be synonyms. We define *zbankrutować* using processuality as 'become$_{pf}$ a bankrupt', linking it with the Polish noun *bankrut*, and *wyłysieć* as 'become a bald (person)', linking it with the Polish nominalised adjective *łysy*. The verb *splajtować* 'become bankrupt$_{pf}$' shares all constitutive relations with *zbankrutować*, even processuality, so it will appear in the same synset with it.[26]

The relational paradigm of lexical semantics, as implemented in a wordnet, has an intrinsically limited expressive power. For one thing, senses are not defined in a formal language which might support inference. One can expect, however, that the structure of synset relations is a basis only for conclusions acceptable to a native speaker. A hyponym, for example, can be exchanged with any of its even remote hypernyms without causing abnormality of the given language expression—but even the most elaborate system of constitutive relations does not guarantee this property. We can observe semantic oppositions which systematically go across large parts of the lexicon and influence contextual behaviour of LUs; that includes differences in stylistic register, aspect or verb class. The topological definition of the synset based on relation-sharing does not eliminate all inappropriate LU grouping in the same synset, if they differ with respect to one of those features.

In order to illustrate the problem better, we will analyse three examples. The first example concerns nouns. The nouns *chłopiec* 'boy' and *gówniarz* '*(derogative)* youngster, squit' share the hypernym *nieletni* 'juvenile', and have no meronyms or holonyms. Their hyponyms are what makes them different: *chłopiec* has hyponyms which *gówniarz* cannot have. For example, *orlę* means approximately 'a proud, brave boy', but *gówniarz* can be neither proud nor brave; *ulicznik* 'urchin' can be paraphrased 'a boy who spends time on streets', but the definition 'a squit who spends time on streets' sounds wrong. To sum up, *chłopiec* and *gówniarz* cannot be synonyms—they have different hyponym sets. To record their intuitive semantic closeness, they are linked in *plWordNet* by inter-register synonymy, a weaker form of synonymy which precludes the sharing of hyponyms. It will be analysed in the next section.



The second example shows how verb aspect influences hypernymy/hyponymy links. The pair *pogarszać* 'worsen$_{impf}$, make$_{impf}$ worse' and *zmieniać* 'change$_{impf}$' is a proper instance of hyponymy, but the hypernym cannot be replaced by its aspectual counterpart *zmienić* 'change$_{pf}$': a perfective semantic element should not be included in an imperfective hyponymic verb.

[26] They are nearly synonymous and the difference in meaning is small: *splajtować* is slightly less formal.

Table 5 Determinants of *plWordNet*'s structure

Constitutive relations	Hyponymy, hypernymy, meronymy, holonymy and several verb-specific relations
Derived relations	Synonymy, inter-register synonymy
Constitutive features	Register; verbal semantic class

Turning to the third example, a similar dependency can be found in verb classes assumed in *plWordNet* and lexico-semantic relations. The verb *mętnieć* 'become clouded$_{impf}$' is a hyponym of *stawać się* 'become$_{impf}$'—both are accomplishments; the activity verb *nawracać się* 'convert' is a sub-ordinate verb of the activity hypernym *zmieniać się* 'change$_{impf}$ oneself' (an iterative meaning). Aspect and verb classes will be discussed in Section 5.

In order to make our relation system more consistent and accurate, we have decided to build register values and verbal semantic classes into the *plWordNet* structure. This is summarized in Table 5.[27] We refer to them as *constitutive features*, because they too influence the structure of our wordnet. To preserve lexico-semantic relations as the basic means of description, constraints related to the constitutive features were added to the relation definitions. In the following sections we will examine the identified constitutive features more closely.

4 Lexical registers

The set theory perspective does not exhaust and explain the distributional properties of the potential constitutive relations. Wordnets generally neglect the fact that a lexical unit's register strongly affects its usage. Consider geographical (dialectal) variation—quotations from (Cruse 2004, p. 59):

> It would be almost unthinkable for publicity material for tourism in Scotland to refer to the geographic features through which rivers run as *valleys*, although that is precisely what they are: the Scottish dialect word *glen* is de rigueur, because of its rich evoked meaning.[28]
>
> Nothing can be said everywhere, every time, to everyone:
>
> *Did you do it with her?* might be described as 'neutral informal'; however, *bonk* is humorous, whereas *fuck*, *screw*, and *shag* are somehow aggresively obscene (although probably to different degrees). In the same humorous-informal category as *bonk*, we find *willie* (cf. *penis*), *boobs* (cf. *breasts*), and perhaps *pussy* (cf. *vagina*).

We understand *register* as a property of text or smaller language expression. Homogeneity in language is rare. The characteristics of a text vary in many

[27] The verb-specific constitutive relations are presupposition, preceding, cause, state, processuality and inchoativity (Maziarz et al. 2011).

[28] In Princeton WordNet *glen* has its register label (domain: region—*Scotland*) and it is a *hyponym* of *valley*, because of differentiae: *glen* 'a narrow secluded valley (in the mountains)'.

dimensions: temporal (contemporary language—archaic or dated language), geographical (common language—regional varieties), socio-cultural (neutral language—language socio-linguistically marked: popular, slang, vulgar or general; also technical or scientific language—general language), formality (formal–informal), text type (poetic, literary language—general language) and many others (Svensén, 2009, p. 316). The register is sometimes defined as "a variety of language with a particular situational context, such as an occupation or social activity" (Hartmann and James 1998, p. 118). Halliday (Halliday and Hasan 1985) in his popular theory of stylistic variation of language distinguishes between field (subject matter, area of discourse), tenor (style, degree of formality) and mode of discourse (written or spoken) (Cruse 2002, p. 492), (Lipka 2002, p. 23), (Cruse 2004, p. 59).

Tests commonly used in wordnets to detect semantic relations are not immune to register differences:

> Note that these tests are devised to detect semantic relations only and are not intended to cover differences in register, style or dialect between words (Vossen 2002, p. 13).

Anomalies in our contextual tests arise simply from the fact that register is directly connected with pragmatics. Pragmatics states that propositional synonymy[29] has its limitations: words can be exchanged in a particular context to some degree of acceptability (Cruse 2004, pp. 155–156). We check interchangeability of a given pair of words in *testing* contexts (not in all contexts), but the tests often lead to nonsensical sentences. Consider an example of a synset from (Vossen 2002, p. 18):[30]

{cop, pig, policeman, police officer}

In PWN, the direct hyponyms of *policeman* include {captain, police captain, police chief}. Let us construct an EWN-style hyponymy test for *police captain* (according to Vossen (2002, p. 22)) using *pig*, a synonym of *policeman* in Vossen's proposal:

A *police captain* is a *pig* with certain properties.
It is a *police captain* and therefore also a *pig*.
If it is a *police captain* then it must be a *pig*.

Are the test expressions normal? odd? contradictory?[31]

In PWN 3.1 there still are such discrepancies. For example, an unmarked term *crossing guard* 'someone who helps people (especially children) at a traffic crossing' is a direct hyponym of an informal *traffic cop* 'a policeman who controls the flow of automobile traffic'.[32]

[29] "Propositional synonymy can be defined, as its name suggests, in terms of entailment. If two lexical items are propositional synonyms, they can be substituted in any expression with truth-conditional properties without effect on those properties." (Cruse 2004, p. 155).

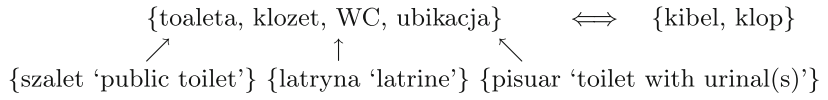
[30] Vossen proposed putting words with different registers into one synset, but practice diverges from theory. In PWN 1.5, the synset was split into two, linked with hyponymy {bull, cop, copper, fuzz, pig} => {policeman, police officer, officer}, and for good reasons! Clearly, treating pairs such as *pig* and *policeman* "democratically" may have entertaining effects.

[31] We borrow a scale of degrees of necessity from Cruse (2004, 54).

[32] Never mind the fact that a crossing guard need not always be a policeman.

The reaction to these test stimuli is not obvious—and if it is not, then what premises can guide editing decisions?

In *plWordNet*, LUs with a similar denotation but different registers will be placed differently in the net of lexico-semantic relations. Consider the series *toaleta* 'toilet', *klozet* 'toilet/WC', *WC* 'WC', *ubikacja* 'toilet', *kibel* 'bog (Br.), loo (Am.)', *klop* 'bog, loo'. Some of these are marked. The names of subclasses *szalet* 'public toilet', *pisuar* 'toilet with urinal(s)' and *latryna* 'latrine' fail the substitution tests for hyponymy with, for example, *kibel*: some test expressions will be unacceptable. The large set of toilet names must be split into two synsets, representing general language usage ('toilet') and marked units ('bog'). We use a special relation of *inter-register synonymy* (here shown as the double arrow).



We have decided to introduce lexical registers to avoid confusing our linguists, wordnet editors, with the ambiguous substitution tests.[33] The precise definition of the relation states that inter-register synonyms (a) share all constitutive relations except hypernymy and (b) differ in stylistic register. The latter is important, because the absence of different hyponyms may be accidental. (That was the case of our example: *szalet*, *pisuar* and *latryna* were put in *plWordNet* later than their hypernyms.) In order to avoid constantly rebuilding *plWordNet* structure, we decided to strengthen our wordnet with register values.

5 Semantic verbal classes and aspect

The range of lexico-semantic relations among verbs is strongly influenced by the semantic classes of verbs and by aspect. That is why both properties should play a role in determining the wordnet structure—no less than constitutive wordnet relations and registers. This is typical not only of Slavic languages but also of other branches of the Indo-European family. Consider a few entries in Cambridge Dictionary Online (Heacock 1995–2011), a traditionally organised English dictionary. The examples are motivated by Rappaport Hovav (2008, p. 38).

- The word *arrive*, a prototypical achievement verb, is defined like this: 'to reach a place, especially at the end of a journey'. This takes another achievement verb, *reach*, as a *genus proximum*.
- The stative verb *resemble* has in its definition another stative verb *be* and the phrasal verb *to look like* ('to look like or be like someone or something').
- The verb of activity *read* is defined as 'to look at words or symbols and understand what they mean'. It is not surprising that *look* also has an activity interpretation.

[33] In introducing new *inter-register* synonymy we return to the lexicographic tradition: "Some dictionaries provide synonyms and near-synonyms, marking differences between them by labels such as (form.), (vulg.), (poet.), (bibl.), etc." (Verkuyl et al. 2003, p. 302).

It is not by chance that all those words have hypernyms (=*genera proxima*) representing the same verb semantic class. In Slavic languages this property of verbs is even more pronounced because of the higher prominence of aspect. In Polish, for example, the perfective verb *napisać* 'write$_{pf}$' would never be explained by any imperfective verb, even one as semantically close as *pisać* 'write$_{impf}$'. In the Universal Dictionary of Polish (UDP) (Dubisz 2004) it is defined thus: 'nakreślić na czymś jakieś litery lub cyfry, wyrazić coś słowami na piśmie' *'draw$_{pf}$ on something letters or numbers, express$_{pf}$ something with words in writing'*.

Semantic classes do not seem to be overtly present in the criteria typically defined for wordnet development, but they have definitely been implicitly taken into account in editing decisions made in most wordnets.

It is almost impossible to analyse synonymy among Polish verbs without considering their semantic classes or aspect, especially because both are fairly interconnected. The taxonomy, presented in Table 6, is based on post-Vendlerian typologies of verbs: Polish (Laskowski 1998)[34] and Russian (Paducheva 1995). We borrowed from Vendler (1957) the names of the first four classes. Concerning aspect, states (stative verbs) are *imperfectiva tantum*; activities are *imperfectiva tantum*; accomplishments (or telic verbs) are both *imperfective* and *perfective*; achievements are *perfectiva tantum*; finally there are perfectives with additional characteristics (delimitatives, perduratives, accumulatives and distributives) which, according to Paducheva (1995), do not belong to any of the previously mentioned categories.

For synonymous and hyponymous verbs, we have introduced the requirement of the identity of aspect and semantic class. Thus verbs of achievement (which are perfective) cannot be synonyms or hyponyms of verbs of accomplishment (neither perfective nor imperfective) and *vice versa*. For example, we consider as inappropriate the lexicographic definitions from the UDP of *wylecieć* 'fly out' using *wydostać się* 'get out' as a *genus proximum*. That is because in our typology the former is an achievement and the latter is an accomplishment: *wylecieć* «o ptakach, owadach: wydostać się skądś na skrzydłach; wyfrunąć, ulecieć» '*of birds, insects: to get out of somewhere on wings; to fly out*'.

On the other hand, we consider it correct when the UDP defines an achievement *zgubić* 'to misplace' with an achievement *stracić* 'to lose':[35]

[34] English translations of Laskowski's class labels (Cetnarowska and Stawnicka 2010).

[35] Note that we try to classify LUs, not different uses. For example, the imperfective verb *budować* 'to build' may be used as follows (all examples come from the IPI PAN Corpus):

(a) A progressive meaning: Jac buduje dom pod Jaktorowem; będziemy mieli stadninę, wiesz? 'Jack is building a house near Jaktorów; we will have a stud farm, you know?'

(b) An iterative meaning: O tym, kto będzie budował domy, zadecyduje—w każdym przypadku—bankier. 'It will be the banker who decides—in each case—who will build houses'.

The verb *build* is an accomplishment if we look at its progressive meaning, but when one considers its iterative meaning it becomes an activity. In our typology, the same LU cannot mean both an accomplishment and an activity; that is why we take into account only primary uses of verbs, like Paducheva (1995, pp. 75, 77–78) and Laskowski (1998, pp. 160–161). That is to say, we try to categorise LUs—not different usages of the same LU.

Table 6 A comparison of semantic verb classes in *plWordNet* with those of Laskowski and Paducheva (modelled after Vendler)

plWordNet 1.6	Laskowski (1998)	Paducheva (1995)	Characteristics
States	States	Inherent states, atemporal properties/relations	−dynamicity
Activities	Activities and eventives	Activities and atelic processes	+dynamicity −change of state −telicity +durativity
Accomplishments	Actions and processuals	Actions proper, telic processes, unfolding actions and unfolding processes	+dynamicity +change of state +telicity +durativity
Achievements	Acts and happenings	Achievements and happenings	+dynamicity +change of state −telicity −durativity
Perfectives with additional characteristics	Uncertain position in the typology	Derived categories	+dynamicity +change of state −telicity +durativity

zgubić «dopuścić, żeby coś zginęło, pozostawić, stracić coś przez nieuwagę, niedopatrzenie» 'to let something be lost, to leave something, to lose something unintentionally, by oversight'.

We have also seen this property in examples taken from the Cambridge Dictionary Online (Heacock 1995–2011). Semantic classes (as well as aspect) affect synonymy.

Verb classes have been built into *plWordNet*'s hyponymy hierarchy. The top-level synsets, mostly non-lexical, represent imperfective *state verbs* and *activities*, perfective *achievements* and *atelic non-momentary change of state situations*, and perfective or imperfective *accomplishments*. Most verbs are linked via hyponymy to those artificial synsets or to their hyponyms.[36] Practically every verb belongs to one verb family in the hyponymic "genealogy", and two verbs can be synonyms only if they share all constitutive relations. It is therefore impossible to put verbs from different semantic classes into one synset. To ensure that it indeed never happens, we have introduced the requirement of semantic class identity between candidates for synonyms: it supplements the set of constitutive relations and register identity requirement. The three form the skeleton of *plWordNet*.

[36] The verbs *robić* 'do$_{impf}$', *zrobić* 'do$_{pf}$', *powodować* 'cause$_{impf}$', *spowodować* 'cause$_{pf}$' cannot be put into any of the classes, but their hyponyms are also linked to our main groups.

6 Conclusions

We propose to avoid the usual synset-synonymy circularity by making the synset the consequence of other elements of a wordnet's topology, rather than a fundamental building block. We introduce constitutive wordnet relations which—supplemented by aspect, register and semantic verb class—determine the structure of a Polish wordnet.

Our list of constitutive relations serves its purpose well. Nonetheless, we have had to select among more lexical-semantic relations and lexical properties which could also have been acceptable. As any informed selection, ours has been guided by objective criteria as far as possible. We need relations which allow the wordnet editor to shun the rather controversial synonymy but still indirectly capture its intuition. We want to avoid putting in one synset two words which a consensus of native speakers would never consider synonymous. The constitutive relations aptly differentiate units with a significant difference of meaning, yet do not require a continual introspection on near-identity of meaning. Instances of part-whole or subclass-superclass relations are easier to recognize and less skewed by subjectivity. In the end, we replace a less tractable relation with a carefully constructed set of more tractable relations.

We illustrate our deliberations with examples from Princeton WordNet, EuroWordNet, *plWordNet* and a few other well-known wordnets, as well as several dictionaries. The overall effect is a reduced conceptual base of our wordnet: by bypassing synonymy as a major design criterion, we have made *plWordNet* less dependent on complex semantic considerations.

No paper can be complete without a note on future plans. Here is ours: we will continue our work on *plWordNet*, both on its design (including the theory and practice of lexical-semantic relations) and on the systematic growth of its coverage.

Acknowledgments Many thanks to Dave Carter, Chris Fournier and Alistair Kennedy for a very close and attentive read of the penultimate version of this paper.

Appendix: The synset as an equivalence class

Section 2.1 introduced the idea that synset membership of LUs is based on their ability to share a set of lexico-semantic relations; see (Piasecki et al. 2009) for an earlier discussion. From this perspective, LUs in a synset cannot be distinguished if we analyse only their occurrences across the given set of *constitutive relations*. LUs belonging to one synset form an *equivalence class* with respect to those relations. We will formalise this idea, assuming—as already noted—that constitutive relations are constrained by the LUs' register, semantic class and aspect.

Table 7 Equivalence among register values for *plWordNet*, part I

$\overset{R}{=}$	Gen.	Lit.	Coll.	Slang	Vulg.	Hist.	Arch.	Old use
General use	**t**	**t**	**t**	f	f	**t**	f	f
Literary	**t**	**t**	**t**	f	f	**t**	f	f
Colloquial	**t**	**t**	**t**	f	f	**t**	f	f
Slang	f	f	f	**t**	**t**	f	f	f
Vulgar	f	f	f	**t**	**t**	f	f	f
Historical	**t**	**t**	**t**	f	f	**t**	f	f
Archaism	f	f	f	f	f	f	**t**	**t**
Old use	f	f	f	f	f	f	**t**	**t**
Dialectical	f	f	f	f	f	f	f	f
Technical	f	f	f	f	f	f	f	f
Scientific	f	f	f	f	f	f	f	f
Metaphorical	f	f	f	f	f	f	f	f
Ironic	f	f	f	f	f	f	f	f

Table 8 Equivalence among register values for *plWordNet*, part II

$\overset{R}{=}$	Dial.	Techn.	Sci.	Metaph.	Iron.
General use	f	f	f	f	f
Literary	f	f	f	f	f
Colloquial	f	f	f	f	f
Slang	f	f	f	f	f
Vulgar	f	f	f	f	f
Historical	f	f	f	f	f
Archaism	f	f	f	f	f
Old use	f	f	f	f	f
Dialectical	**t**	f	f	f	f
Technical	f	**t**	**t**	f	f
Scientific	f	**t**	**t**	f	f
Metaphorical	f	f	f	**t**	f
Ironic	f	f	f	f	**t**

The equivalence class of $z \in Y$ is $[z] = \{y \in Y \mid y \sim z\}$ where relation $\sim$ is reflexive ($v \sim v$), symmetrical (if $v \sim w$ then $w \sim v$) and transitive (if $v \sim w$ and $w \sim x$ then $v \sim x$).

Let L be a set of LUs, R be registers, A be verbal semantic classes.[37]

$$\overset{R}{=}: R \times R \longmapsto \{false, true\}$$

$$\overset{A}{=}: A \times A \longmapsto \{false, true\}$$

[37] Nouns and adjectives will have a dummy value of the verbal semantic class.

Tables 7, 8 define $\overset{R}{=}$. For the needs of $\overset{A}{=}$ we assume that the equality of semantic classes is required for most verbal constitutive relations in a wordnet: $\overset{A}{=}$ is *true* only for identical arguments. Here we follow the practice of dictionary editors—see the preceding sections—and conclusions drawn from the analysis of Polish data.

Let $f_R : L \mapsto R$, $f_A : L \mapsto A$ be surjective functions. Let $\mathcal{W} = \{w_i\}_{i \in I}$ be a set of binary *constitutive wordnet relations*:

$$w_i : L \times L \longmapsto \{false, true\}$$

We can at last define synonymy. It is binary relation

$$S : L \times L \longmapsto \{false, true\}$$

such that

$$S(x,y) \overset{df}{=} \underset{z \in L}{\forall} \underset{i \in I}{\forall} (w_i(x,z) \iff w_i(y,z)) \wedge f_R(x) \overset{R}{=} f_R(y) \wedge f_A(x) \overset{A}{=} f_A(y)$$

The synonymy relation is reflexive, symmetrical and transitive simply because the relations $\iff$ and $\overset{R}{=}$ and $\overset{A}{=}$ are. Synonymy therefore induces equivalence classes, to which we refer as synsets.

One of the effects of the proposed definition of *wordnet synonymy*—as expressed by synsets—is a sharp separation in the wordnet structure of LUs which native speakers consider closely semantically related. An example: *mężczyzna* 'man'—*facet* 'guy'. In order to keep the wordnet perspective close to native speakers' competence, we should consider a weaker form of close semantic relation going across stylistic register barriers.

That is why we introduce *inter-register synonymy*, $IRS : L \times L \longmapsto \{false, true\}$. Assume that $J \subset I$ indexes all relations except hypernymy. Then:

$$IRS(x,y) \overset{df}{=} \underset{z \in L}{\forall} \underset{j \in J}{\forall} (w_j(x,z) \iff w_j(y,z)) \wedge \neg(f_R(x) \overset{R}{=} f_R(y)) \wedge (f_A(x) \overset{A}{=} f_A(y))$$

Our formalisation may appear excessive, because a wordnet cannot really be a formal model of lexical semantics. The idea of a synset as an equivalence class, however, can be applied in wordnet development practice in a straightforward way: all decisions concerning the inclusion of a LU into a synset should be based on the analysis of potential instances of the constitutive relations.

References

Alonge, A. (1996). *Definition of the links and subsets for verbs*. EuroWordNet Project LE4003, Deliverable D006. http://www.illc.uva.nl/EuroWordNet/docs/D006.zip (unpublished).

Azarova, I. V. (2008). RussNet as a computer lexicon for Russian. In *Proceedings of Sixteenth International Conference on Intelligent Information Systems*, pp. 447–456.

Bańko, M. (Ed.) (2000). *Inny słownik języka polskiego PWN [another dictionary of Polish]*. Warsaw: Wydawnictwo Naukowe PWN.

Bańko, M. (Ed.) (2010). *Wielki słownik wyrazów bliskoznacznych [a dictionary of synonyms, antonyms etc]*. paper version. PWN.

Bhattacharyya, P., Fellbaum, C., & Vossen, P. (Eds.) (2010). Principles, construction and application of multilingual wordnets. In *Proceedings of 5th Global Wordnet Conference*. Mumbai: Narosa Publishing House.

Borin, L., & Forsberg. M. (2010). From the people's synonym dictionary to fuzzy synsets—first steps. In *Proceedings of LREC 2010 Workshop on Semantic Relations. Theory and applications*. ELRA http://spraakbanken.gu.se/personal/lars/pblctns/lrec2010-ws-fsl-fwn.pdf.

Broda, B., & Mazur, W. (2011). Evaluation of clustering algorithms for word sense disambiguation. *International Journal of Data Mining, Modelling and Management* (to appear).

Broda, B., Piasecki, M., & Maziarz, M. (2010). *Evaluating LexCSD-a weakly-supervised method on improved semantically annotated corpus in a large scale experiment*. In Intelligent Information Systems, Publishing House of University of Podlasie, Siedlce, pp. 63–76.

Cetnarowska, B., & Stawnicka, J. (2010). The verb's semantics and its compatibility with temporal durative adverbials in Polish. *Studies in Polish Linguistics, 5*, 27–50.

Collinge, N. (Ed.) (1990). *An encyclopaedia of language*. London: Routledge.

Cruse, A. (2004). *Meaning in Language. An introduction to semantics and pragmatics*. Oxford: Oxford University Press.

Cruse, D.A. (1997). *Lexical semantics*. Cambridge: Cambridge University Press.

Cruse, D. A. (Ed.) (2002). *Lexicology. An international handbook on the nature and structure of words and vocabularies, vol 1*. Walter de Gruyter GmbH.

Dubisz, S. (Ed.) (2004). *Uniwersalny słownik języka polskiego [a universal dictionary of Polish]*. electronic version 1.0. PWN.

Fellbaum, C. (1998a). A semantic network of English: The mother of all WordNets. *Computers and the Humanities, 32*, 209–220.

Fellbaum, C. (Ed.) (1998b). *WordNet—an electronic lexical database*. Cambridge, MA:The MIT Press.

Fellbaum, C., Osherson, A., & Clark, P. E. (2009). *Putting semantics into WordNet's "Morphosemantic" links, LTC2007*. pp. 350–358.

Geeraerts, D. (2010). *Theories of lexical semantics*.Oxford: Oxford University Press.

Halliday, M. A. K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. USA: Deakin University Press.

Hamp, B., & Feldweg, H. (1997). GermaNet—a lexical-semantic net for German. In *Proceedings of ACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, Madrid, pp. 9–15.

Hartmann, R. R. K., & James, G. (1998). *Dictionary of lexicography*. London: Routledge.

Heacock, P, (Ed.) (1995–2011) *Cambridge dictionaries online*. Cambridge: Cambridge University Press.

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language 39*(2), 170–210.

Koeva, S., Mihov, S., & Tinchev, T. (2004). Bulgarian Wordnet—structure and validation. *Romanian Journal of Information Science and Technology, 7*(1–2), 61–78.

Kunze, C. (1999). Semantics of verbs within GermaNet and EuroWordNet. In E. Kordoni (Ed.), *Proceedings of Workshop at 11th European Summer School in Logic, Language and Information*, pp. 189–200.

Kunze, C., & Lemnitzer, L. (2010). Lexical-semantic and conceptual relations in GermaNet. In P. Storjohann (Ed.), *Lexical-semantic relations. Theoretical and practical perspectives*. Amsterdam: John Benjamins Publishing Co.

Laskowski, R. (1998). Kategorie morfologiczne języka polskiego—charakterystyka funkcjonalna. In R. Grzegorczykowa, L. Roman, & H. Wróbel (Eds.), *Gramatyka współczesnego języka polskiego—Morfologia,* (vol. 1, 2nd edn). PWN.

Lipka, L. (2002). *English lexicology: Lexical structure, word semantics and word-formation*. Tubingen: Gunter Narr Verlag.

Lyons, J. (1977). *Semantics, vol 1–2*. Cambridge: Cambridge University Press.

Maziarz, M., Piasecki, M., Szpakowicz, S., Rabiega-Wiśniewska, J., & Hojka, B. (2011). Semantic relations between verbs in polish wordnet 2.0. *Cognitive Studies, 11* (in press).

McCarthy, M. (2003). *Vocabulary*. Oxford: Oxford University Press.

Miller, G. A. (1998). *Nouns in WordNet*. In WordNet—an electronic lexical database. Cambridge, MA: The MIT Press.

Miller, G. A., & Fellbaum, C. (2003). Morphosemantic links in WordNet. *Traitement Automatique de Langue, 44*(2), 69–80.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography, 3*(4), 235–244. ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1993). Introduction to WordNet: An on-line lexical database. ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps, unpublished, part of the set called: "Five Papers".

Murphy, M. L. (2003). *Semantic relations and the lexicon*. Cambridge: Cambridge University Press.

Murphy, M. L. (2010). *Lexical meaning*. Cambridge: Cambridge University Press.

Paducheva, E. V. (1995). Taxonomic categories and semantics of aspectual opposition. In P. M. Bertinetto (Ed.), *Temporal reference, aspect and actionality,* (vol. 1, pp. 71–89). Torino: Rosenberg & Sellier.

Painter, C. (2001). *Learning through language in early childhood*. London: Continuum.

Pala, K., & Smrž, P. (2004). Building Czech WordNet. *Romanian Journal of Information Science and Technology, 7*(1–2), 79–88.

Pedersen, B. S., Nimb, S., Asmussen, J., Sørensen, N. H., Trap-Jensen, L., & Lorentzen, H. (2009). DanNet: The challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation, 43*(3), 269–299.

Piasecki, M., Szpakowicz, S., & Broda, B. (2007). Extended similarity test for the evaluation of semantic similarity functions. In Z. Vetulani (Ed.), *Proceedings of 3rd Language and Technology Conference, Wydawnictwo Poznańskie Sp.* (pp. 104–108). z o.o., Poznań, Poland.

Piasecki, M., Szpakowicz, S., & Broda, B. (2009). A Wordnet from the Ground Up. Oficyna Wydawnicza Politechniki Wrocławskiej. http://www.site.uottawa.ca/szpak/pub/A_Wordnet_from_the_Ground_Up.pdf.

Piasecki, M., Szpakowicz, S., & Broda, B. (2010). Toward plWordNet 2.0. In Bhattacharyya et al. (Eds.), *Principles, construction and application of multilingual wordnets. Proceedings of 5th Global Wordnet Conference* (pp. 263–270). Mumbai: Narosa Publishing House.

Piasecki, M., Marcińczuk, M., Ramocki, R., & Maziarz, M. (2012). WordnetLoom: A wordnet development system integrating form-based and graph-based perspectives. *International Journal of Data Mining, Modelling and Management*. http://www.inderscience.com/browse/index.php?journalID=342.

Pustejovsky, J. (2001). *The generative lexicon*. Cambridge, MA: MIT Press.

Rappaport Hovav, M. (2008). Telicity, atomicity and the Vendler classification of verbs. In S. Rothstein (Ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect*. Amsterdam: John Benjamins Publishing Co.

Stede, M. (1999). *Lexical semantics and knowledge representation in multilingual text generation*. Dordrecht: Kluwer.

Svensén, B. (2009). *A handbook of lexicography. The theory and practice of dictionary-making*. Cambridge: Cambridge University Press.

Tufiş, D., Cristea, D., & Stamou, S. (2004). BalkaNet: Aims, methods, results and perspectives. A general overview. In D. Tufiş (Ed.), *Romanian Journal on Information Science and Technology. Special Issue on BalkaNet, vol 7(1-2)* (pp. 9–34) Germany: Romanian Academy.

Vendler, Z. (1957). Verbs and times. *The Philosophical Review, 66*(2), 143–160.

Verkuyl, H., Janssen, M., & Jansen, F. (2003). The codification of usage by labels. In P. van Sterkenburg (Ed.), *A practical guide to lexicography* Amsterdam: John Benjamins Publishing Co.

Vossen, P. (Ed.) (1998) *EuroWordNet. A multilingual database with lexical semantic networks*. Dordrecht: Kluwer.

Vossen, P. (2002). *EuroWordNet general document version 3. Tech. rep*. Amsterdam: University of Amsterdam.

Vossen, P., Agirre, E., Calzolari, N., Fellbaum, C., & Hsieh, S. K., Huang, C. R. et al. (2008). KYOTO: A system for mining, structuring, and distributing knowledge across languages and cultures. In A. Tanács, D. Csendes, V. Vincze, C. Fellbaum, & P. Vossen (Eds.), *Proceedings of Fourth Global WordNet Conference* (pp. 474–484). University of Szeged, Department of Informatics.

Yule, G. (2010). *The study of language*. Oxford: Oxford University Press.