

# Implementation of the Verb Model in plWordNet 4.0

Agnieszka Dziob

Maciej Piasecki

Wrocław University of Science and Technology

agnieszka.dziob|maciej.piasecki@pwr.edu.pl

## Abstract

The paper presents an expansion of the verb model for plWordNet – the wordnet of Polish. A modified system of constitutive features (register, aspect and verb classes), synset and lexical relations is presented. A special attention is given to the proposed new relations and changes in the verb classification. We discuss also the results of its verification by application to the description of a relatively large sample of Polish verbs. The model introduces a new class of relations, namely non-constitutive synset relations that are shared among lexical units, but describe, not define synsets. The proposed model is compared to the entailment relations in other wordnets, and the description of verbs based on valency frames.

## 1 Introduction

plWordNet 3.0 emo (Maziarz et al., 2016) describes 17,391 Polish verb lemmas by 31,834 *lexical units*<sup>1</sup> (LUs), and 75,643 relations. Thus, a very significant subset of Polish verbs has been covered. These numbers are also much higher than in any other wordnet, including Princeton WordNet (henceforth, PWN) (Fellbaum, 1998a). Nevertheless, plWordNet (plWN) 3.0 achieved the coverage of only ~30% of the verbs with the frequency >10 (57,969 in total<sup>2</sup>) in the *plWordNet Corpus*, i.e. 4 billion words<sup>3</sup> corpus of Polish. plWN 3.0 verbs represent only 58.9% of 29,532 verbs described in SGJP (Saloni et al., 2015) - the most comprehensive morphological dictionary of Polish. Due to a very large size of plWN Corpus

this number can be a good predictor of the expected coverage in NLP applications of plWN. It could be higher. The relation density for verbs in plWN 3.0 emo is high, but several verb lexico-semantic relations are rather infrequent<sup>4</sup>.

(Dziob et al., 2017) presented a significantly modified, new model for the description of verbs in plWN. Our goal was to apply this model in expanding plWN 3.0 by a couple of thousand Polish verb lemmas, verify the proposed relation definitions in editing practice, both from the qualitative and quantitative point of view, as well as to propose some improvements and generalisations.

## 2 Verb Model in Brief

The system of lexico-semantic relations proposed for verbs in plWordNet 4.0 (Dziob et al., 2017) is based on the plWN 2.0 model. (Maziarz et al., 2011). A pair of relations: *hypernymy* and *hyponymy* organise verbs into a hierarchy. This differentiates plWN from PWN, in which hypernymy and *troponymy* are used (Fellbaum, 1998b), but is close to the models of EuroWordNet (Vossen, 2002) and GermaNet (Kunze, 1999).

Fellbaum (1998b) argued against verb hyponymy that verbs differ from nouns and it is not possible to adapt a hyponymy test to them:

*An x is a y.*

As a consequence, troponymy in PWN “represents a special case on entailment: pairs that are always temporally coextensive and are related by entailment” (Fellbaum, 1998b). In plWN temporal co-extensiveness is expressed by two verb relations: hypernymy and meronymy, see Sec. 4. Fellbaum (1998b) defined troponymy as a *manner relation* and illustrated with a substitution test:

*To V1 is to V2 in some particular manner.*

<sup>1</sup> Lexical unit is a triple: a lemma, Part of Speech and sense id.

<sup>2</sup> However, some substantial number of these verbs can result from the errors of the morphological guesser.

<sup>3</sup> plWN Corpus 10.0 includes: ICS PAS corpus (Przepiórkowski, 2004) National Corpus of Polish (Przepiórkowski et al., 2012), Corpus of Rzeczpospolita (Weiss, 2008), Polish Wikipedia, and a large amount of texts selected from Internet with automated quality check; duplicates were automatically removed.

<sup>4</sup> See <http://plwordnet.pwr.edu.pl/wordnet/stats>

A test proposed for verb hyponymy in pLWN 2.0 correlates with the PWN troponymy test (Maziarz et al., 2011):

*to X(inf) is to Y(inf) in a special way, somehow.*

where the expression *a special way, somehow* represents a manner which is an intrinsic element of the situation definition. In order to cover this part of the definition in an explicit way *manner* relation was proposed (Dziob et al., 2017), which can be paraphrased: *X-ować to robić coś Y-owo* 'To X is to do something in an Y way'.

pLWN 1.0 included both relations: hyponymy and troponymy. However, the former was a synset relation<sup>5</sup>, while the latter was defined only for LUs and strictly related to the prefix derivational associations between members of aspectual pairs. Derwojedowa (et al., 2007) argues that there is a large group of verbs in the Polish language that are derived from such verbs that seem to be their hypernyms (i.e. expressing more general meaning than their derivatives), but of different aspect. Because it was assumed that verbs in the same hypernymy branch have the same aspect, cf (Maziarz et al., 2011), Derwojedowa (et al., 2007) proposed to use troponymy to link such verb hyponymy-like pairs in which elements differ in aspect and express some semantic addition. The use of troponymy was finally abandoned, also because its definition was very significantly different than in PWN. Instead, in order to link verbs associated by prefixal derivation such that one has a narrower meaning than the other, *secondary aspectuality* relation was introduced (Maziarz et al., 2011). It links, e.g., perfective: accumulative, distributive, and delimitative verbs with their imperfective derivational bases, like in the case of *posiedzieć* 'to keep sitting for a while' ↔ *siedzieć* 'to sit<sub>imp</sub>'.

In addition to hyponymy, which organises verbs into hierarchies, there are several more relations in pLWN that describe relationships between situations, namely: *presupposition*, *preceding*, *meronymy/holonymy*, *inchoativity*, *causality*, *processuality* and *state*.

**Presupposition** is close to the logical presupposition, expresses temporal backward relation, and signals the necessary occurrence of one situation before the other, e.g. *żywy<sub>Adj</sub>* 'alive' ← *umrzeć<sub>Verb</sub>* 'to die'<sup>6</sup>.

**Preceding** is also a temporal backward relation signalling an usual, but not necessary occurrence of one situation before the second one, it can be considered as a 'weaker variant of presupposition', e.g. *siedzieć* 'to sit' or *leżeć* 'to lie' ← *wstać* 'to stand'),

Verb **meronymy/holonymy** (not automatically reverse) express co-occurrence of two situations in the same time period, e.g. *chrapać* 'to snore' ← *spać* 'to sleep', cf (Dziob et al., 2017).

**Inchoativity** links verbs representing the beginning of a situation and this situation, e.g. *zakochać się* 'fall in love' → *kochać* 'love'.

**Causality** describes the relation between LUs representing two situations where the first (represented by a verb) results in the second (represented by V, N, Adj or Adv), e.g. *zablokować* 'to lock' → *blokada* 'lock'.

**Processuality** links a verb LU and a noun, adjective or adverb representing a state resulting from the situation represented by the verb, e.g. *zmienić się* 'to change' → *inny* 'different'.

**Multiplicativity** is a relation emphasising an aspect of repetition in the verb meaning. It signals that some situation is repeated several times or an action performed on several objects. Multiplicativity is divided into two subtypes:

- *distributivity* (perfective) representing multiple performance, e.g. *nakupić* 'to buy many things' → *kupić* 'to buy<sub>perf</sub>', and
- *iterativity* (imperfective) representing multiple repetitions, e.g. *czytywać* 'to read<sub>imp</sub> many times' → *czytać* 'to read<sub>imp</sub>'.

**State** connects state verbs with nouns, adjectives and adverbs describing states, e.g. *czzerwienić się* 'to be red' → *czzerwony* 'red'.

The next group of relations links verbs with LUs describing conditions in which situations occur.

**Circumstance** was introduced for pLWN 4.0 to link a verb representing a situation with a noun LU which is the semantic head of a prepositional phrase used to express conditions in which this situation occurs, e.g. *dopłynąć* 'to swim<sub>perf</sub> to some point/place' → *brzeg* 'a bank'.

**Manner**, added for pLWN 4.0 links a verb LU with an adverb representing a manner in which an action is performed or a state happens, e.g. *popracować* 'to work a little' → *trochę* 'a little'.

<sup>5</sup> pLWN model is based on LUs as basic building blocks. All relations are defined for LUs and synset relations are notational abbreviations for relations shared among LUs belonging to the two linked synsets, cf (Maziarz et al., 2013).

<sup>6</sup> In pLWN 4.0 model many verb relations were expanded to cross-categorial relations, see (Dziob et al., 2017)

**Object** and **subject**, introduced for pLWN 4.0, link a verb LU with noun LUs representing, respectively, an object, e.g. *obuć* ‘to put on shoe’ → *but* ‘a shoe’, and subject, e.g. *oźrebić się* ‘to foal’ → *klacz* ‘a mare’. Such noun LUs must typically occur as intrinsic elements of semantic definitions (e.g. in dictionaries) of verbs that are linked to them.

All the relations mentioned so far are synset relations, as they are shared among LUs belonging to the same synset. All of them, except *circumstance*, *manner*, *object* and *subject*, are constitutive relations, i.e. relations defining synsets. Synonymy is defined in pLWN on the basis of sharing constitutive relations by LUs, cf (Maziarz et al., 2013). The set of constitutive relations determines the structure of a wordnet.

The above listed four relations are meant to be a tool for expanded characterisation of verb meanings (e.g. for WSD). They are defined in a less strict way and do not express necessary constraints. To limit their excessive proliferation, we included sanity conditions in their definitions: if there are more than three possible instances of such a relation per one synset, than we resign from adding this relation to this synset at all. Thus, this verb characterising relations are not meant to be a tool for identifying different lexical meanings and are not constitutive relations. For instance, *jechać* ‘to ride’ can be linked by *circumstance* to *pojazd* ‘a vehicle’ or *zwierzę* ‘an animal’, but because of this we do not want to differentiate between two different meanings of *jechać*. However, as these relations are mostly shared among LUs, we represent them as synset relations. They initiate a new class of wordnet relations: supporting, non-constitutive synset relations.

As it was already mentioned, the identity of aspect is a fundamental rule in linking verbs in the hypernymy structure and, as a consequence, in grouping them into synsets. Two main aspects are morphologically distinguished<sup>7</sup> in Polish: *perfective* and *imperfective*. There is also a set of ~150 bi-aspectual verbs with the same lemma for both aspects (or ambiguous with respect to aspect) (Mędak, 1997), e.g. *nobilitować* ‘to ennoble’. In Slavic linguistics, it is used to describe the difference between the two aspects as the difference in the perspective of a subject perceiving a given situation: imperfective verb describes the situation

as lasting, while perfective describes it as finished, and besides this difference there is no other difference in the meaning of the two verbs of an aspectual pair, cf (Młynarczyk, 2004; Laskowski, 1998).

However, Młynarczyk (2004) argues that although such a definition of the aspectual verb pair is not controversial, this binary distinction does not originate from the language system as such, but it is caused by the prefixation. The derivational prefixes express semantic information beyond the mere change of the aspect. This correlates with the two types of aspectual lexico-semantic relations introduced in pLWN 2.0 (Maziarz et al., 2011): *pure* and *secondary* aspectuality - both defined as lexical relations (i.e. for LUs, not shared).

The former links pure aspectual pairs, i.e. such that two verbs in two different aspects do not differ in their meanings<sup>8</sup>, e.g. *czytać<sub>impf.</sub>* ‘to read<sub>impf.</sub>’ ↔ *przeczytać<sub>perf.</sub>* ‘to read<sub>perf.</sub>’. *Secondary aspectual* verb LU pairs are such that they express different aspects and share their derivational basis or the second is derived from the first, but the meaning of the second LU is modified beyond the aspectual difference in relation to the first, e.g. *czytać<sub>impf.</sub>* ↔ *poczytać<sub>perf.</sub>* ‘to read a little’, cf (Dziob et al., 2017).

The rest of verb lexical relation stay the same in pLWN 4.0 as in pLWN 2.0 model (Maziarz et al., 2011). The set encompasses (see also Tab. 2): **role inclusion** - a semantic association signalled by derivation of verbs from nouns - which expresses information similar to semantic roles, e.g. *bronować* ‘to harrow’ ← *brona* ‘a harrow’, *pieprzyć* ‘to pepper’ ← *pieprz* ‘a pepper’, *niańczyć* ‘to nurse’ ← *niańka* ‘a nanny’; **derivationality** representing verb links signalled by derivation, but without clear enough semantic character and not yet covered by more specific relations e.g. *hamletyzować* ‘to vacillate, to consider something pointless’ → Hamlet (PN, Shakespeare’s hero); and **antonymy** (with two subtypes), which is in pLWN a lexical relation (Piasiecki et al., 2009) and is not a constitutive relation (Maziarz et al., 2013).

PWN verb relations link only verbs (Fellbaum, 1998b), in similar way to GermaNet (Kunze, 1999). In pLWN, following EuroWordNet (Vossen 2002) verb LUs can be linked to all PoS. Modification of the verb part of pLWN 4.0 model

<sup>7</sup> I.e. A verb lemma encodes its aspect, it is not inflected with respect to aspect.

<sup>8</sup> However, more precisely, we should say that they do not significantly differ in their meanings beyond the information expressed by the aspect change.

was inspired by relations for adjectives and adverbs from pLWN 3.0, cf (Maziarz et al., 2016a, 2016b). The verb relations expanded to cross-categorical relations include: **processuality** (e.g. *anarchizować się* ‘to become<sub>Imp</sub> anarchic’ → *anarchista* ‘anarchist’ / *anarchiczny* ‘anarchic’ / *anarchicznie* ‘anarchically’), **causality** (e.g. *zmienić* ‘to change’ → *zmiana* ‘a change’ / *inny* ‘different’ / *inaczej* ‘other’), **presupposition** (e.g. *całość* ‘a whole’ / *cały* ‘whole’ ← *podzielić się* ‘to divide itself’; *jasno* ‘brightly’ ← *ściemnić* ‘to dim’), **preceding** (e.g. *dobry* ‘good<sub>adj</sub>’ / *zły* ‘bad<sub>adj</sub>’ / *dobrze* ‘good<sub>adv</sub>’ / *źle* ‘bad<sub>adv</sub>’ ← *pogorszyć się* ‘to worsen’; *mąż* ‘a husband’, *żona* ‘a wife’ ← *rozwiść się* ‘to get divorced’), **state** (e.g. *jaśnieć* ‘to shine’ → *jasny* ‘bright’, *jasno* ‘brightly’; *królować* ‘to reign’ → *król* ‘a king’), cf (Dziob et al., 2017). This expansion resulted in a significant increase of their frequency in pLWN, see Sec. 6.

### 3 Semantic Classes

The pLWN 2.0 top part of the verb hypernymy structure consisted of artificial synsets expressing verb semantic classification originating from 7 classes of Laskowski (1998): processes, actions, acts, accidents, activities, events, states, were defined on the basis of (Vendler, 1967). This classification resulted in a large number of subclasses that constrained the rest of the verb hypernymy structure.

This classification system was sophisticated and potentially useful in applications, but appeared to be very hard to be applied consistently by wordnet editors (Dziob et al., 2017), especially as the verb classes constrain verb relations in pLWN. After analysis of the editing practice and the obtained results, the classification was simplified with only two main classes left in pLWN 4.0: *state* and *dynamic* verbs. This basic division corresponds to the general linguistic tradition, cf e.g. (Vendler, 1967; Comrie, 1989, Paduceva, 1996), Polish, e.g. (Karolak, 2001; Grzesiak, 1989), and also EWN. Vossen (2002) defines dynamic verbs as:

“specific transition from one state to another (bounded in time) or a continuous transition perceived as an ongoing temporally unbounded process,”

while static verbs as

“in which there is no transition from one eventuality or situation to another, i.e. they are non-dynamic”.

pLWN 4.0 uses similar definitions for both classes, but more attention is given to detailed characterisation of subgroups of the general classes and formulation of paraphrase-based descriptions for them. As a result, state verbs in pLWN 4.0 include verbs representing: 1) **localisation** (in space): *X jest gdzieś, ma jakieś położenie, jest w jakiejś pozycji*; ‘X is somewhere, has some location, is in a location’, e.g. *znajdować się* ‘to be in some place’, *sit* ‘siedzieć’, *otaczać* ‘to surround’; 2) **possession of permanent material features**, e.g. weight or volume (*X jest jakiś, jakoś, ma jakąś cechę, coś na stałe* ‘X possesses some feature, something permanent’; e.g. *jaśnieć* ‘to shine’, *mierzyć* ‘~to be of particular size’), 3) **relationships** between entities, both material and non-material (*X pozostaje w relacji do czegoś* ‘X stays in a relation to something’; e.g. *składać się* ‘to comprise’, *należać* ‘to belong’), 4) **mental states**, emotional, sense experience (*X odczuwa coś, doświadcza czegoś* ‘X feels something, experiences sth.’; e.g. *kochać* ‘to love’, *być przy nadziei* ‘be pregnant’, *istnieć* ‘to exist’), and also the 5) group which includes all other verbs that do not express **dynamics of situation** (i.e. do not represent a change from situation X to Y).

Dynamic verbs in pLWN 4.0 are perfective verbs: 1) **distributive** (to do something by many agents or in relation to many objects, e.g. *przebadac* ‘to examine many people’), 2) **accumulative** (to do something to such an extent that it is enough; e.g. *ubawić się* ‘to amuse itself’), 3) **perdurative** (to be doing something during limited time; e.g. *przemieszkać* ‘to live during some period in a place’), 4) **delimitative** (to be doing/happening for some time or to some extent; e.g. *pomieszkać* ‘to live for short time in a place’); and also 5) **action verbs** a) all perfective and bi-aspectual, b) imperfective derivatives of accumulative, delimitative, perdurative, and distributive verbs (representing changing situations), c) imperfective derivatives of semelfactive verbs (i.e. representing punctual or instantaneous events), d) imperfective causative verbs e.g. *rozśmieszać* ‘to make<sub>Imp</sub> someone laughing’), e) **processive** (*X staje się czymś, jakoś* ‘X becomes sth, somehow’; e.g. *starzeć się* ‘to become<sub>Imp</sub> gradually old’), f) **inchoative** (*X zaczyna się, zaczyna coś robić* ‘X is starting, begins doing sth’; e.g. *położyć się* ‘to lie down’), g) **limitative** (*X przestaje być czymś, jakimś, jakoś, przestaje coś robić* ‘X stops being sth, somehow, stops doing sth.’; e.g. *wybarwiać się* ‘to lose<sub>Imp</sub> colour’) and h) all other imperfective verbs that represent situation changing due to actions of entities involved (e.g. *iść* ‘to walk’).

The subclass definitions (summarised above) are formulated in an operational way, on the basis of several substitution tests. They are referred to in relation definition and support linguists in editing. Thus, semantic class is a constitutive feature, together with stylistic register and aspect. Semantic subclasses of dynamic verbs are clearly connected to several relations that are characteristic for this class, namely: processuality, causality, inchoativity and multiplicativity. Only state verbs can participate in state relation. Other types of relations occur in both verb classes.

Verb classification is expressed by a hierarchy of *artificial LUs* (represented by singleton synsets) as in (Maziarz et al., 2011). Class assignment is done by placing a verb in an appropriate hypernymic branch, as hyper/hyponymy and synonymy (due to relation sharing) requires equality of semantic classes.

Semantic subclasses clearly refer to well-known linguistic classifications of verbs, e.g. (Levin, 1993; Fellbaum, 1998) and support wordnet editors in building hypernymic trees on the basis of semantic properties of verbs. The reduction of the number of classes (from 7 to 2) should facilitate identification of only real verb meanings and prevent introduction of non-natural and too fine-grained meanings.

#### 4 Entailment

Verb entailment relation plays an important role in PWN and GermaNet, which is defined by Fellbaum (1998b) as:

“the relation between two verbs  $V_1$  and  $V_2$  that holds when the sentence Someone  $V_1$  logically entails the sentence Someone  $V_2$ .”

In addition, Fellbaum (1998b) introduces four subtypes of entailment. In pLWN a more fine-grained division of the spectrum of verb relations is proposed, see the comparison in Table 1.

We can notice a different perspective on situations co-occurring in the same time period. In PWN it is always represented by troponymy, which is defined as a kind of entailment (see Sec. 2), while in pLWN temporal co-occurrence of situations is covered by verb meronymy. In pLWN 2.0 a dedicated subtype of sub-situation meronymy was used (Maziarz, et al., 2011) (plus *associated situation* subtype), e.g., *komunikować się* ‘to contact’ and *zadawać się* ‘to associate with sb’

- communication is a part of a relationship, but they are different situations. Verb meronymy is necessary after troponymy has been excluded from pLWN and partially exchanged with hyponymy. We observed that the distinction between sub-situation and associated situation subtypes was too subtle in practice. Thus, verb meronymy in pLWN 4.0 does not have subtypes and is described by the following test:

*Jeśli coś/ktoś X-uje, to na pewno jednocześnie Y-uje, bo X-ować można tylko Y-ując.*

‘If sb./sth. is X-ing, then it/he is surely Y-ing, as X-ing is possible only if Y-ing is performed<sup>9</sup>.’

Examples: *lunatykować* ‘to sleepwalk’ → *spać* ‘to sleep’, *nakopać się* ‘to kick so long, to be enough of it’ → *kopać* ‘to kick’.

EWN entailment	+Temporal inclusion		-Temporal inclusion	
	Co-extensiveness -troponymy	Proper inclusion	Backward presupposition	Cause
pLWN	Hyponymy, meronymy	Mero- nymy	Presupposition, preceding	Cause

Table 1: Temporal relations in PWN vs pLWN

On basis of the experience from the work on adverbs in pLWN 3.0, most verb relations of pLWN 4.0 allow for linking verbs with other PoS, including adverbs (Dziob et al., 2017). The system of relations for adverbs was derived from the one of adjectives in pLWN 3.0 (Maziarz et al., 2016b) that simplified extension of verb relations; e.g., a *processuality* link to an adjective or adverb is identified by the following tests:

*X-ować to stawać się / stać się Y-owym*

*X-ing means to be becoming/to become Y-like.*

e.g. *ochłodzić się* ‘to become cool / cooler’ → *chłodny* ‘cool’)

*X-ować to stać się / stawać się Y<sub>Adv</sub>-owo*

*X-ing to be becoming / to become Y<sub>Adv</sub>*

e.g. *ochłodzić się* ‘to become cool / cooler’ → *chłodno* ‘chilly’

#### 5 Relations Signalled by Derivation

Derivational prefixes of verbs are important semantic signal in Polish. So far, verb prefixes have been only selectively and implicitly described as correlated with relations signalled by derivations. Although, we have not yet studied this issue in a

<sup>9</sup> English gloss suggests that only verbs for which progressive forms exist can be used in this relation, but this limitation does not exist in Polish.

Table 2. Verb lexico-semantic relations in the plWordNet 4.0 model (first synset relations)

Relation	POs	Example	v3.1	G%
inter-register syn- onymy	V-V	<i>pieprzyć się</i> [vulgar] ‘to have sex’ → <i>uprawiać seks</i> ‘to have sex’	2529	25.4
hyponymy	V-V	<i>nadgryźć</i> ‘to chew a little’ → <i>ugryźć</i> ‘to chew’	29433	29.8
meronymy	V-V	<i>gryźć</i> ‘to chew’ is an integral part of situation <i>jeść</i> ‘to eat’	2311	-18,3
holonymy	V-V	<i>jeść</i> ‘to eat’ is a typical situation including <i>gryźć</i> ‘to chew’	3156	9.3
manner	V-Adv	<i>nadgryźć</i> ‘to bite a little’ → <i>trochę</i> ‘a little’	651	<i>new</i>
inchoativity	V-V, N	<i>urodzić się</i> ‘to be born’ → <i>żyć</i> ‘to live’	482	19.6
processuality	V-N, Adj, Adv	<i>ocieplać się</i> ‘to get warmer’ → <i>ciepły</i> ‘warm (adj)’, <i>ciepło</i> ‘warm (adv)’	1137	56.0
causality	V-V, N, Adj, Adv	<i>ocieplać</i> ‘to grow warm’ → <i>ocieplać się</i> ‘to get warmer’, <i>ciepły</i> ‘warm (adj)’, <i>ciepło</i> ‘warm (adv)’	3091	74.3
presupposition	V-V, N, Adj, Adv	<i>dodać</i> ‘to add’ presupposes <i>istnieć</i> ‘to be’ (no subject’s identity presupposition)	261	56.3
preceding	V-V, N, Adj, Adv	<i>rozwiść się</i> ‘to divorced’ precedes [to be] <i>żona</i> ‘a wife’ or <i>mąż</i> ‘a husband’ (subject’s identity preceding)	571	241.9
multiplicativity	V-V			
- iterativity		<i>jadać</i> ‘~to eat from time to time’ → <i>jeść</i> ‘to eat’	144	9.1
- distributivity		<i>popodgrzewać</i> ‘~to warm up many things’ → <i>podgrzać</i> ‘to warm up’	419	39.6
state	V-V, N, Adj, Adv	<i>dłużyć się</i> ‘to drag’ → <i>długi</i> ‘long (adj)’, <i>długo</i> ‘long (adv)’	176	89.2
subject	V-N	<i>ankietować</i> ‘to poll’ → <i>ankieter</i> ‘pollster’	221	<i>new</i>
object	V-N	<i>ankietować</i> ‘to poll’ → <i>ankietowany</i> ‘polled’	187	<i>new</i>
circumstance	V-N	<i>ankietować</i> ‘to poll’ → <i>kwestionariusz</i> ‘a questionnaire’	66	<i>new</i>
aspectuality	V-V		33351	25.6
- pure		<i>nadgryźć</i> ‘to bite <sub>perf</sub> a little’ - <i>nadgryzać</i> ‘to bite <sub>imperf</sub> a little’		
- secondary		<i>nadgryźć</i> ‘~to chew <sub>perf</sub> a little’ - <i>gryźć</i> ‘to chew <sub>imperf</sub> ’		
derivationality	V-V, N, Adj, Adv	<i>ocieplać</i> ‘to get warmer’ → <i>ciepły</i> ‘warm’	396	40.9
antonymy	V-V	<i>odezwać się</i> ‘to said’ - <i>przemilczać</i> ‘to left unsaid’	2530	7.6
- complementary		<i>rozbierać</i> ‘to undress’ - <i>ubierać</i> ‘to dress’		
- proper				
converseness	V-V	<i>implikować</i> ‘to imply’ - <i>wynikać</i> ‘to result’	134	19.6
role inclusion	V-N		1793	32.1
- subject		<i>gospodarować</i> ‘to farm’ ← <i>gospodarz</i> ‘a farmer’		
- instrument		<i>betonować</i> ‘to concrete’ ← <i>beton</i> ‘a concrete’		
- result		<i>filetować</i> ‘to fillet’ ← <i>filet</i> ‘a fillet’		
- location		<i>magazynować</i> ‘to store’ ← <i>garaż</i> ‘a store’		
- object		<i>lajkować</i> ‘to give a like’ ← <i>lajk</i> ‘a like’		
- time		<i>ucztować</i> ‘to feast’ ← <i>uczta</i> ‘a feast’		
- indefinite		<i>litować się</i> ‘to have pity’ ← <i>litość</i> ‘pity’		

systematic way, some associations between prefixes, meanings and lexico-semantic relations became visible.

Prefixes *do-*, *wy-* can signal situations in which an agent is accomplishing his goal, e.g. *dojść* ‘to have reached sth’, *dokopać się* ‘to have dug down

to sth’, *wysiedzieć* ‘to have continued sitting until sth happened’, *wyczekać* ‘to have continued waiting ...’. They express a relation to a goal or an end that are often implicit.

Another example is a set of prefixes expressing

a kind of manner relation in the case of delimitative verbs: *po-* and *do-*. Concerning the first, *po-* prefix means to do a little, e.g. *posiedzieć* ‘to sit a little’ (*siedzieć* ‘to sit’), *popoglądać* ‘to watch a little’ (*oglądać* to watch). The prefix *do-* signals more advanced or intensive situation, e.g. *doszkolić się* ‘to improve qualifications’ (*szkolić się* ‘to learn by himself’), *dogęszczać* ‘to thicken more (a mixture, substance etc.)’ (*zagęszczać* ‘to make thicker’).

Verbs derived by prefixes are linked by *secondary aspectuality*, e.g. *wysiedzieć* ‘to have continued sitting’ – *siedzieć* ‘to sit’ or by more specific relations, e.g. inchoativity. However, secondary aspectuality is intentionality vague, only slightly more informative than fuzzynymy, and is a way of registering LU pairs requiring deeper investigation in future. A more in depth exploration of derivational verb prefixes focused on enrichment of wordnet relations is a very interesting task to be undertaken in the future.

## 6 Implementation

plWN 3.0 includes 17,391 verb lemmas described by 31,834 LUs that should cover all meanings of the verbs. As it was declared earlier, one of the goals for plWN 4.0 is a significant expansion of the verb database. Following the corpus-based development scheme, a set of 8,000 most frequent verbs in the plWN corpus was selected that were lacking in plWN 3.0. With the help of the word2vec (Mikolov et al., 2013) model based on plWN Corpus, the selected verbs were clustered in packages of ~100 verbs each. Each package is intended to cover a limited number of topics and to be a unit of work assigned to a linguist.

So far, the number of verb lemmas in plWN has been increased to 19,272 i.e. by 11%. In parallel, we have updated the verb hypernymy structures and verb relations to a large extent. This enabled us to observe the changes triggered by the new verb model. Tab. 2 present statistics for the relations and changes in relations.

We can notice that the modification of the model resulted in the increased frequency of the following relations: processuality, causality, presupposition, inchoativity, state. In the same time the number of verb meronymy instances has decreased but this could be expected due to the more stricter definition and the remove of the ambiguous division into two subtypes (this ambiguity led to too far going interpretations).

## 7 Verb Model vs Valency Lexicon

A high quality valency dictionary with good coverage is an indispensable resource for many NLP applications. Unfortunately, its construction is very laborious and costly. plWN model defines a rich system of verb relations. The question is to what extent it can supplement a valency lexicon? Marantz (1981) argues that semantic roles are indispensable in the description of the predicate-argument structures, e.g. the *agens* role refers to the logical subject of a predicate, while the theme and patiens roles to the logical objects.

A clear reference made in the plWN verb model to the syntactic-semantic relations is aimed at improving richness of LU descriptions following Apresjan (2000) who argues that a dictionary should provide description of co-occurrences of lexico-semantic and syntactic features. In Czech WordNet (Pala et al., 2004) valency frames are added to synsets. However, we assumed in plWN that syntactic valency is not a constitutive feature of verb LUs, and does not need to be shared by synset members, so is not used to define synsets. It could be described on the level of LUs, but this is in fact done in Walenty (Hajnicz et al., 2017), a large valency lexicon of Polish. Thus, syntactic valency is not expressed in plWN, a semantic lexicon, and there are no plans for introducing it. So, this part is clearly missing, but verb arguments which are mentioned in relation definitions can be implicitly expressed in the lexico-semantic relations. As a consequence, quite a lot of information about semantic restrictions on valency arguments is hidden in plWN relations. It is partial and selective, but still can be useful.

Three relations introduced in plWN 4.0 directly evoke structure relations, namely: *subject* (referring to the semantic agent role), *object* (patient role) and *circumstance*, whose detection is based on prepositional phrases, which can correspond to other roles, for example location, result, time. As it was said in Sec. 2, *subject*, *object* and *circumstance* relations (*manner* does not link nouns) are not constitutive relations, but emphasise selected aspects of LU meanings that are common to the whole synset, and in the same time relate these aspects to the syntactic structure, e.g. *circumstance* links *brzeg* ‘a shore’ with *dobijać* ‘to reach a shore’ informs also that one of the *dobijać* predicate arguments represents location. In a similar way *object* relation links *usypiać* ‘to put down, to put to sleep, to euthanize’ with *zwierzę* ‘an animal’ and signals that one of the arguments repre-

sents animal or its hyponym. The guidelines instruct to find for these relations nouns that are located on relatively high levels of the hypernymy to describe the meaning of the verb LU, not its collocational behaviour. Linguists are also required to check if most of the hyponyms of the selected target noun fulfil the tests for this relation. In the same time the target noun should not be located too high in order to preserve meaningfulness of the link, i.e. LUs from the top level of the hypernymy hierarchy should be avoided, e.g. *byt* ‘an entity’, *istota* ‘a being’).

In Walenty semantic description is based on *selectional preferences*: “lexico-semantic dependencies between a unit which is a predicate of an utterance and units that are its arguments, that determine what kind of notions can co-occur on the subsequent valency arguments” (Hajnicz et al., 2017). Because Walenty frames have been built in relation to the pLWN LUs, selectional preferences of the Walenty entries tend to be correlated with pLWN synsets. Hajnicz (et al., 2017) aims at encompassing by selectional preferences all hyponyms of a given synset, e.g. for *rżeć* ‘to neigh’ there are two semantic frames: selectional preferences of the first restrict *agent* (“Initiator”) to *koń* ‘a horse’ (pLWN: *koń 1* ‘a horse’) and in the second to *człowiek* ‘a man’ characterising the second meaning of *rżeć* as ‘to laugh producing sound resembling neighing’. Selectional preferences in Walenty are chosen according to the frequency, i.e. in the case of *rżeć* ‘to neigh’ the editor decided that the constraint *koń* ‘a horse’ for the agents is enough frequent to be expressed in the frame; in addition, all hyponyms of *koń* ‘a horse’, e.g. *pegaz* ‘Pegasus-like’, *gniadosz* ‘a bay’, but also derivatives, i.e. diminutives e.g. *konik* ‘~a little horse’ and augmentatives, e.g. *konisko* ‘~a large, not pretty horse’ are included in the preferences. pLWN describes the *subject* link between *rżeć* ‘to neigh’ and *koniowate 1* ‘an equine’, because also zebras or giraffes are neighing (at least in Polish) and they belong to equines taxonomy together with *koń* ‘a horse’. These links can be further interpreted by explicit derivational links.

Semantic valency information can be also found in lexical relations: *role* (N-V, describing deverbal nouns) *role inclusion* (V-N, verbs derived from nouns). Both relations have 7 subtypes: agents, instrument, product, location, patients, time and indefinite subtype (Maziarz et al., 2011) that refer to thematic roles of Fillmore (1968), on the one side and to the studies on the semantics of deverbal nouns in the Polish literature, cf

(Wróbel, 2001). Both relations tell something about the selectional preferences.

For instance *solić 1* ‘to salt’ is a hyponym of *przyprawiać 2* ‘to spice’ and means ‘to spice with salt’ and is linked with *sól* ‘salt’ by *role\_inclusion:instrument* as a verb derived from a noun - a tool name. The expression *solić solą* ‘to salt with salt’ is redundant and incorrect, but one can say *przyprawiać solą* ‘to spice with salt’, where *przyprawiać 2* is linked by *role\_inclusion:instrument* to *przyprawa 1* (‘a spice’); *przyprawiać* ‘to spice’ can be done by salt or different spices - cohyponyms and cousins of *sól 1* ‘salt’. Another example can be *bokser* ‘a boxer’ linked by *role:agens* to *boksować 1* ‘to box’ (its derivational basis), which is a hyponym of *bić 4* ‘to hit, to beat’. The expression *bokser boksuje* is redundant but *bokser bije* ‘a boxer is beating’ is correct. Thus, the combination of *role/role inclusion* and verb and noun hypernymy can be used to draw conclusions about selectional preferences of the verb arguments.

Relations defined on the level of synsets go beyond the derivational associations. During the work on pLWN 4.0 we have realised that a lot of valuable semantic knowledge is not covered by strictly derivationally motivated relations. Analysis of fuzzynymy from pLWN 3.0 showed that semantic associations visible in derivations can be cautiously generalised, i.e. in a way based on strict procedure, substitution tests and guaranteeing good consistency among editors.

## 8 Conclusion

We presented an expanded verb model for pLWN, including modified constitutive features, and synset and lexical relations. Non-constitutive synset relations were introduced. They are shared among LUs in a synset, characterise important aspects of verb meaning, but are not necessary constraints for defining synsets. They seem to be a good tool for the inclusion of knowledge valuable for wordnet applications, e.g., WSD. The proposed model was verified and slightly amended on the basis of its application to a large sample of Polish verbs. The first statistical data showing the results of the proposed changes were discussed. We showed that the proposed system of relations provides information about entailment and selectional preferences. Open issues are: the relation between the defined lexico-semantic relations and relations between verb valency frames, and the extent of automatization in identification of the selectional preferences on the basis of the relations.



## Acknowledgements

Works funded by the Polish Ministry of Science and Higher Education within CLARIN-PL Research Infrastructure.

## Reference

- Iurii D. Apresian. 2000. *Systematic lexicography*. Oxford University Press on Demand, Oxford.
- Joan Bresnan. 1982. *The mental representation of grammatical relations*, volume 1. The MIT Press, Cambridge.
- Agnieszka Dziob, Maciej Piasecki, Marek Maziarz, Justyna Wieczorek, and Marta Dobrowolska-Pigoń. 2017. Towards Revised System of Verb Wordnet Relations for Polish. In *Proceedings of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets*, Galway, Ireland, 19-20 June.
- Christiane Fellbaum (ed.). 1998a. *WordNet: An electronic lexical database*. MIT Press, Cambridge.
- Christiane Fellbaum. 1998b. *A semantic network of English verbs*. In: Christiane Fellbaum (ed.), *WordNet: An electronic lexical database*. MIT Press, Cambridge.
- Charles J. Fillmore. 1968. *The case for case*. In Emmon Bach, and Robert T. Harms. (ed.). *Universals in Linguistic Theory*. Holt, Rinehart and Winston, New York.
- Jane B. Grimshaw. 1990. *Argument structure*. The MIT Press, Cambridge.
- Renata Grzegorzczkova. 1990. *Wprowadzenie do semantyki językoznawczej*. [Introduction to linguistic semantics]. PWN, Warszawa.
- Renata Grzegorzczkova. 2008. *Wstęp do językoznawstwa* [Introduction to polish linguistics]. PWN, Warszawa.
- Romuald Grzesiak. 1983. *Semantyka i składnia czasowników percepcji zmysłowej*, Zakład Narodowy im. Ossolińskich, Wrocław-Warszawa-Kraków.
- Robert T. Harms, and Emmon Bach (ed.). 1968. *Universals in Linguistic Theory*. Holt, Rinehart and Winston, New York.
- Elżbieta Hajnicz, and Bartłomiej Nitoń. 2017. Instrukcja dostępu do słownika walencyjnego Walenty za pośrednictwem programu Słowl. Institute of Computer Science PAS, URL: [http://clarin-pl.eu/wp-content/uploads/2017/05/instrukcja\\_uzytkownika\\_Walentego.pdf](http://clarin-pl.eu/wp-content/uploads/2017/05/instrukcja_uzytkownika_Walentego.pdf).
- Stanisław Karolak. 2001. *Od semantyki do gramatyki. Wybór rozpraw*. [From semantics to grammar. The choice of papers]. Sławistyczny Ośrodek Wydawniczy, Warszawa.
- Claudia Kunze. 1999. *Semantics of verbs within GermaNet and EuroWordNet*. In *Proceedings of 11th European Summer School in Logic, Language and Information*. Utrecht.
- Roman Laskowski. 1998. *Kategorie morfologiczne języka polskiego—charakterystyka funkcjonalna* [Morphological categories of Polish-functional characteristic]. In Renata Grzegorzczkova, Henryk Wróbel, Roman Laskowski (ed.), *Gramatyka współczesnego języka polskiego. Morfologia 1* [Grammar of Polish language. Morphology 1]. PWN, Warszawa.
- Beth Levin. 1993. *English verb classes and alternations: A preliminary investigation*. University of Chicago Press, Chicago.
- Alec Marantz. 1981. *On the nature of grammatical relations*. PhD Thesis. Massachusetts Institute of Technology.
- Marek Maziarz, Maciej Piasecki, Stanisław Szpakowicz, Joanna Rabeiga-Wiśniewska, and Bożena Hojka. 2011. Semantic relations between verbs in polish wordnet 2.0. *Cognitive studies*, 11:183-200.
- 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.
- Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stanisław Szpakowicz, and Paweł Kędzia. 2016a. *plWordNet 3.0-a Comprehensive Lexical-Semantic Resource*. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*, Osaka, Japan.
- Marek Maziarz, Stanisław Szpakowicz, and Michał Kaliński. 2016b. *Adverbs in plWordNet: Theory and Implementation*. In *Proceedings of the 8th International WordNet Conference — GWC 2016*, Bucharest, Romania, 27-30 January.
- Stanisław Mędrak. 1997. *Słownik form koniugacyjnych czasowników polskich* [A dictionary of polish verbs patterns], Universitas, Kraków.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. *Efficient Estimation of Word Representations in Vector Space*. CoRR, vol. abs/1301.3781.
- Anna K. Młynarczyk. 2004. *Aspectual pairing in Polish*. PhD Thesis. Utrecht University, Utrecht.
- Elena V. Padučeva. 1996. *Semantičeskie issledowanija: Semantika vremeni i vida v russkom jazyke; Semantika narrativa*. Škola Jazyki Russkoj Kultury.
- Karel Pala, and Pavel Smrž. 2004. Building czech wordnet. *Romanian Journal of Information Science and Technology*, 7(2-3):79-88.

- Maciej Piasecki, Bartosz Broda, and Stanisław Szpakowicz. 2009. *A wordnet from the ground up*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- Adam Przepiórkowski. 2004. *The IPI PAN Corpus, Preliminary Version*. Institute of Computer Science PAS.
- Adam Przepiórkowski, Mirosław Bańko, Rafał L. Górski, and Barbara Lewandowska-Tomaszczyk. (ed.). 2012. *Narodowy Korpus Języka Polskiego* [National Corpus of Polish]. PWN, Warszawa.
- Zygmunt Saloni, Marcin Woliński, Robert Wołosz, Włodzimierz Gruszczyński, and Danuta Skowrońska. 2015. *Słownik gramatyczny języka polskiego*. [Grammatical dictionary of Polish]. 3rd edition. URL: <http://sgjp.pl/>.
- Zeno Vendler. 1967. *Verbs in Times*. In Z. Vendler, *Linguistics and Philosophy*, Ithaca, New York, Cornell University Press.
- Dawid Weiss. 2008. *Korpus Rzeczypospolitej* [Corpus of text from the online edition of "Rzeczpospolita"]. Unpublished. URL: <http://www.cs.put.poznan.pl/dweiss/rzeczpospolita>.
- Henryk Wróbel. 2001. *Gramatyka języka polskiego* [Grammar of the Polish language]. Spółka Wydawnicza "Od Nowa", Kraków.