

Towards Revised System of Verb Wordnet Relations for Polish

Agnieszka Dziob, Maciej Piasecki, Marek Maziarz,
Justyna Wieczorek, and Marta Dobrowolska-Pigoń

G4.19 Research Group, Wrocław University of Science and Technology, Wrocław, Poland
{agnieszka.dziob,maciej.piasecki}@pwr.edu.pl

Abstract. A revised model for the description of verbs in plWordNet - a large wordnet of Polish is presented in the paper. The model has originated from the previous one (the versions 2.0 and 3.0) as a combination of simplifications and expansions. The simplifications were motivated by the collected experience of the previous model implementation for almost 18 thousand Polish verbs. The proposed expansions are meant to improve the description of verbs and support provided by plWordNet for various applications.

1 Underestimated Wordnet Part

Verbs are mostly described by fewer relations in wordnets and the verb subnetwork expresses lower density providing less information. In this paper we claim that:

- a richer structure of verb relations can be introduced into a wordnet,
- rich verb relations in a wordnet can supplement valency-based model in a way which is attractive for many applications.

Our goal is to outline a model for the description of verbs in plWordNet which consists of: features (e.g. verb class, aspect) and lexico-semantic relations and is strongly based on the plWordNet general model. We are going to confine the discussion only to the Polish language and plWordNet¹ (plWN - a large wordnet for Polish [30]), but in a perspective enabling more general conclusions. The presented work builds on the previous model proposed for plWN 2.0 [28] and used also in the version 3.0 [30], as well as the experience collected from its implementation for 17,391 verb lemmas. During this work it became clear that the model must be refined in order to improve its applicability in daily wordnet editing practice and consistency of the wordnet editor decisions.

Six verb relations were proposed for Princeton WordNet (PWN) [34]: *synonymy*, *antonymy*, *inclusive entailment* (proper inclusion), *troponymy* (coextensiveness), *cause* and *presupposition*. Later, since PWN 1.5, inclusion and presupposition have been combined into *entailment*.

GermaNet identified troponymy with *hyponymy* [22]. PWN's inclusive entailment was renamed *subevent* in EWN [1], and 'entailment' is only a label for presupposition (backward along the timeline) [15]. Causality is cross-categorial [23]. EWN's set is similar to GermaNet's [48]. Hyponymy is PWN's troponymy. Cause includes PWN's

¹ <http://plwordnet.pwr.edu.pl>

presupposition [47]. EWN also introduced *near-synonymy* (for semantically close co-hyponyms), *near-antonymy*, as well as *cross-categorical synonymy*, *antonymy* and *hyponymy*.

There are 19 verb relations in p1WN 3.0, 6 of them have derivational origin. Table 1 contrasts our selected verb relations with those of PWN, EWN and GermaNet. A list of p1WN 4.0 verb relations is presented in Table 2 and discussed in Sec. 3-4.

2 Constitutive Relations and Features

2.1 Synset Definition

The p1WordNet model, cf [29; 30] is based on *lexical units* (LUs, i.e. triples: *lemma*, *part of speech*, *sense identifier*) as basic building blocks. LUs are grouped into *synsets* on the basis of sharing *constitutive relations* and *features*, i.e. two LUs belong to one synset if and only if they share targets of the constitutive relations and values of the constitutive features. Constitutive relations are selected lexico-semantic relations that are *enough frequent*, express relatively *high sharing factor* and originate from the *tradition of lexicography* and *wordnets* and are relatively *well understood*, e.g. hyper/hyponymy, holo/meronymy, iterativity, distributivity. Constitutive features *constrain* the shape of the system of lexico-semantic relations (e.g. *aspect* or *adjective* and *verb classes*) or express some general *pragmatic conditions* on the use of lexical meanings (e.g. *stylistic register*). Constitutive features are often referred to in the definitions of the lexico-semantic relations, e.g. hyper/hyponymy is defined only for adjectives of the same class or inter-register synonymy links LUs of non-compatible lexical registers, cf [30].

Summing up this very brief p1WN model description, see also [29], lexico-semantic relations are the main means of description. They all can be traced back to the language data. *Synset relations* are a kind of abbreviations representing sets of constitutive relations held between members of two synsets. The synonymy relation is to some extent derived from the constitutive relations and features that define synsets.

2.2 Aspect

Aspect is one of the fundamental grammatical verb categories and has gained a lot of attention in the literature, cf. e.g. [4; 8; 47]. Comrie describes it as “the internal temporal constituency of the situation” [4]. In Slavic languages it is marked morphologically by affixes and - in some cases - by ablaut, cf. e.g. [3; 4; 19; 24; 35; 49]. According to the most widespread opinion among slavists, grammatical aspect is a way in which a speaker envisages the situation he is speaking about. Thus the *imperfective aspect* classifies the situation represented by a verb as ongoing, while the *perfective aspect* informs that the represented situation is delimited in time and described as complete in relation to its elements [24]. Every Polish verb may be either perfective, or imperfective, if only we ignore ca 150 bi-aspectual verbs [31].² Concerning *lexical aspect* (*Aktionsarten*) there are many verb pairs with opposition

² The bi-aspectual verbs are ambiguous with respect to the morphological form, and the actual aspect is determined by the context of use [41].

extended beyond imperfective/perfective by additional features of the represented situation, e.g., the impf. verb *robić* ('~to do, to be doing some activity') gives birth not only to its perf. counterpart *zrobić*, but also to many other derivatives formed with non-empty prefixes, including the delimitative derivative *po-robić* 'to have been doing some activity for some time, and to have it finished' or accumulative derivative *na-robić się* 'to work until being exhausted'. pLWN covers the grammatical and lexical aspects jointly in the aspectuality relation. *Pure aspectual pairs* differs only in the grammatical aspect, while *secondary aspectual pairs* embody also the lexical aspect [28].

In pLWN 3.0 we adopted the theory of grammatical aspect that binds it with the concepts of telicity/atelicity in such a way that only telic verbs were allowed to form pure aspectuality pairs. Since telicity depends heavily on syntactic context [46], we took into account and compared only primary uses of verbs [29], cf [24,35]. However, the identification of the verb pairs was very laborious, because it required corpus-based analysis of many examples. Moreover, it was difficult to maintain the consistency of editors' decisions. For instance, the pair *gubić*_{impf.} / *zgubić*_{perf.} 'to lose something' looks like an ordinary pure aspectual pair, but the close inspection of a corpus reveals that *gubić* has the restriction that it rarely occurs with direct objects expressed with countable NPs in singular (in cursive meaning), which is a clear signal of atelicity [46; 8]. For many verbs such semantic-syntactic patterns are not so clear with either telic or atelic contexts being dominant, e.g., the primarily atelic verb *jeść*_{impf.} 'to eat' (being a couple of *zjeść*_{perf.}) occurs in many telic contexts (like *jeść jabłko* 'to eat the apple'). Thus, in the model proposed here, we have resigned from the telicity condition and based the category on the analysis of definitions and paraphrases. Thus secondary aspectuality was restricted mainly to those verb pairs that are overtly linked derivationally by a non-empty affix, mostly a prefix.

Aspect is an intriguing problem for Slavic wordnet designers [13; 20; 21; 36; 37]. Some put pure aspectual pairs into one synset, e.g. BulNet [20] or CzechWordNet [37], following the model of PWN (in English the similar distinction of progr./perf. is inflectionally marked, and not derivational in nature). We treat the aspect in an opposite manner as an important constitutive feature that shapes our verb relations [29].

The rationale for this decision is following: if we took a closer look into Polish dictionaries, we could notice that verb definitions (paraphrases) almost always include superordinate LU (*genus proximum*) in the same aspect as the verb being defined (*definiendum*). For example, the impf. *robić* 'to do' gets the impf. *wykonywać* '~ to execute' as a *genus*, while the perf. *zrobić*_{perf.} 'to do' - the perf. *wykonać* '~ to accomplish'. This is, of course, not a coincidence: you cannot freely substitute in a context a verb of a given aspect with a superordinate verb in an opposite aspect. This is due to the fact, that although a perf. verb implies its impf. counterpart, the reverse implication is rarely true (what is in progress not necessarily would be completed [42]).³ This suggests natural restriction of the hyper-/hyponymy relation only to LUs of the same aspect. Consequently, if two LUs differ in their constitutive relations they cannot belong to the same synset, see Sec. 2.1. Bi-aspectual verbs have mostly two hypernyms: imperfective and perfective and the verb hypernymy tree forks, e.g.

³ The exception are parametric pairs, e.g., *rosnąć*_{impf.} - *urosnąć*_{perf.} 'to grow up'.

mianować 'to nominate' can mean both *nadać* [tytuł] 'to grant [a title]' but also *nadawać* [tytuł] 'to be grating [a title]'.

Pure aspectual pairs that we disunite into separate synsets are interlinked solely by pure aspectuality relation. The meaning differences in the case of secondary aspectuality are even more significant and results also in additional lexico-semantic relations between the pair elements.

Aspect is also correlated with other verb relations, e.g. multiplicativity:iterativity is observed only for imperfective verbs and tests for meronymy are constrained by aspect, see Sec. 3.

2.3 Verb Classes

In pLWN 3.0 verbs were organised into a hierarchical system of classes derived from the proposal of Vendler [45] and Laskowski [24] for Polish. Classes and subclasses were represented by the top level verb hypernym synsets. However, the classification was fine-grained, multi-level and appeared to be difficult in consistent use. Moreover, the representation of classes as top synsets made shaping the structure a challenging task, e.g. all co-hyponyms had to belong to one single fine-grained class. For pLWN 4.0 we decided to significantly simplify the classification and its representation. While still being close to Vendler [45], we use only two top classes: *state* and *dynamic* verbs.

State verbs are imperfective atelic and durative verbs that represent situations which are stable and do not include any change during their time span. State verbs can represent [24]: 1) localisation or position in space, (e.g. *leżeć* 'to lie', *stać* 'to stand', *znajdować się* '~to be in some place', *otaczać* 'to surround'), 2) possession of permanent material features like mass, temperature, state of matter (e.g. *ważyć* 'to weigh', *mierzyć* '~to be of particular size', *świecić* '~to produce light'), 3) relationships between entities, both material and non-material, like to be a part of something, to belong to someone/something, to be in a space (e.g. *ograniczać* 'to delimit', *równać się* 'to equal to', *składać się* 'to consist of'), 4) mental states, emotional, sense experience (e.g. *odczuwać*, *czuć* both meaning 'to feel', *wiedzieć* 'to know', *doświadczać* 'to experience', *kochać* 'to love', *widzieć* 'to know'), 5) other static situations (e.g. *żyć* 'to live', *spać* 'to sleep', *spodziewać się dziecka* 'to be pregnant, lit. to expect a child').

State verbs are all imperfect and are linked with their perfective counterparts - dynamic verbs - by the secondary aspectuality relation.

Dynamic verbs include all former auxiliary classes of pLWN 3.0 [28]: 1) *distributive* verbs (to do something by many agents or in relation to many objects, e.g. *powykańczyć się* '~to get exhausted, finished, plural sub., about many things', *powyłamywać* '~to be breaking off many things/parts'), 2) *accumulative* (to do something to such an extent that it is enough, e.g. *najeść się* 'to be eating until an agent is full', *spracować się* '~to be working so long that it is enough'), 3) *perdurative* (to be doing something during particular or limited time, e.g. *przepracować* 'to be working during some time', *przewegetować* '~to be vegetating during a particular time'), 4) *delimitative* (to be doing or to happening for some time or to some extent, e.g. *pobiegać* '~to be running a little', *popadać* '~to be raining a little').

All other dynamic verbs are grouped in 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 verbs⁴ (i.e. representing punctual or instantaneous events), e.g. *mrugać_{impf}* (from *mrugnąć_{perf}*) ‘to flicker’, representing multiple changes, d) imperfective causative verbs (expressing occurrence of a new situation), e) processive (expressing gradual transition to a new state), f) inchoative (representing introductory phase of a new situation) and g) limitative (representing ending of some situation), and also g) all other imperfective verbs that represent situation changing due to actions of entities involved or with respect to them, e.g. *ić* ‘to walk_{imp}’, *biegać* ‘to run_{imp}’, *plakać* ‘to cry_{imp}’.

The proposed system is much less fine-grained and more shallow - two main classes and 5 subtypes - than the one of pLWN 3.0 (9 main classes, 4 auxiliary classes). In pLWN 3.0 only some of these verb classes had direct influence on the system of verb relations. Thus the former auxiliary classes became subclasses of the dynamic verbs. Their direct influence on the relation structure was limited only to distributive verbs related to the distributivity relation, see Sec. 4.3.

The main reasons for distinguishing state verbs were the cross-categorical synset relation called *state* [28], see Sec. 3.4, imperfective aspect of state verbs and also the secondary aspectuality which links state verbs with their dynamic counterparts.

3 Synset Relations

3.1 Hypernymy Hierarchical Structures

In PWN *troponymy* relation was used instead hyponymy for verbs [9], and described as “to V1 is to V2 in some particular manner”. It is called also “a manner relation” and links temporally co-extensive verbs. However, EWN has opted for verb hyponymy which can be adapted to the manner relation [48]. GermaNet identified troponymy with hyponymy [22]. Maziarz et al. [29] compared the ways of defining nouns and verbs in dictionaries and argued for the use of hyper/hyponymy relations for verbs in pLWN 3.0. It links verbs of the same semantic *class*, *aspect* and compatible *register*:

Derwojedowa et al. [5] argued that there is place for both relations: troponymy and hyponymy in pLWN due to the specific morphology of Polish verbs. Many verbs are derived by prefixes from their derivational bases that seem to be their hypernyms in the same time. Troponymy could be a relation signalling a derivational association between a more specific and more general verb. However, starting from pLWN 2.0 such association started to be described by secondary aspectuality relation. Derwojedowa et al. [5] and Piasecki et al. [38] noticed the necessity of having two verb relations expressing specification of the meaning as a result of the rich semantics of verb prefixes in Polish. In GermaNet these two tendencies are combined [15; 16] by putting on the upper levels of the verb hierarchy synsets that exemplify the meaning of verb prefixes. In Polish, the semantics of the verb prefixes is more

⁴ Semelfactive verbs as perfective verbs are included into actions, see the point a).

complex and its more in depth description is planned for further research.

p1WN 3.0 verb hierarchy resembles the one from PWN 3.1 where top levels are populated by artificial synsets (non-lexicalised)⁵ that are similar to 40 “top-level synsets” in PWN [11]. We plan to rebuild it in such a way that the upper levels of the hierarchy will be also synsets grouping hyponymy branches, but constructed in a bottom-up direction. The reconstruction work will be started from the most specific verbs in the bottom levels and proceed to upper levels. The top synsets will result from the language data and potential needs for more comprehensible organisation of the tree in the case of artificial synsets. As in the case of nouns, we will not combine all verb synsets into one single-rooted tree.

Semantic class and aspect are kept as constitutive features that determine the hypernymic trees for two reasons. First, the simplified set of classes is easier in use for wordnet editors and next the class assignment facilitates identification of hypernyms. Second, state verbs are all imperfective, while their perfective derivatives are in fact secondary aspectuality pairs and express additional semantic difference. Division into classes supports consistent organisation of the verb synsets.

3.2 Associations between Situations

Besides hyper/hyponymy, the proposed verb constitutive relations (i.e. synset relations) can be divided into 3 groups. The first group includes lexico-semantic relations expressing associations between situations that are related to the timeline to some extent. It includes: *presupposition* and *preceding* (backward relations), *meronymy* (co-occurrence of two situations), *inchoativity* (beginning of a situation), *causality* (resulting in a situation) and *processuality* (a state resulting). They are only slightly amended since p1WN 3.0. If we observe a situation represented by a verb from the perspective of a point t_0 in the present time, then causality, processuality and inchoativity refers to a point t_1 in the future resulting somehow from the verb situation, *presupposition* and *preceding* refer to a point t_{-1} in the past, as the verb situation is influenced by what had happened around t_{-1} , while meronymy refers to the present time period and a co-occurring situation, cf [28].

In p1WN 3.0 model all these relations had many subtypes related to properties of LUs linked (e.g. aspect). According to our experience, definitions and substitution tests have been simplified by parameterizing them by the feature values. WordnetLoom [39] application has been extended in such a way that on the basis of the aspect attribute of a LU, an appropriate version of a test is generated and presented to the editor.

Table 1. Selected verb lexico-semantic relations in four wordnets

PWN	GermaNet	EWN	p1WN
synonymy (V-V)	synonymy (V-V)	synonymy (V-V)	synonymy (V-V)
antonymy (V-V)	antonymy (V-V)	antonymy (V-V, N, Adj, Adv)	antonymy (V-V)
troponymy (V-V)	hyponymy (V-V)	hyponymy (V-V)	hyponymy (V-V)

⁵ Representing very fine-grained division into semantic classes.

entailment - proper inclusion (V-V)	subevent (V-V)	subevent (V-V)	meronymy (V-V)
entailment - backward presupposition (V-V)	entailment	cause <i>factial</i> (V-V, N, Adj)	presupposition (V-V, N, Adj, Adv)
entailment (V-V)	entailment	cause <i>non-factial</i> (V-V, N, Adj)	preceding (V-V, N, Adj, Adv)
cause (V-V)	cause (V-V, N, Adj)	cause (V-V, N, Adj)	causality (V-V, N, Adj, Adv)
troponymy (V-V)	_____	manner (V-Adv)	manner (V-Adv)
_____	selectional restrictions (V-N)	_____	subject (V-N)
_____	selectional restrictions (V-N)	_____	object (V-N)
_____	_____	role inclusion (V-N)	role inclusion (V-N)

Except meronymy (V-V) and inchoativity (V-V,N) all other relations can link verbs to any other part of speech. The system of 6 situation association relations can seem to be complex. In PWN, most of them are covered by entailment with cause as the only exception [9]. A more fine grained description was proposed in EWN [48], see Table 1. However, some relations are signalled derivationally, more fine-grained types helps to write more precise definitions and substitution tests and all of them are relatively frequent, see statistics from pLWN 3.0 in Tab. 2. Due to the above mentioned problems with the model of pLWN 3.0 we can expect that the presented numbers are not complete.

Inchoativity [28] links: V-V,N, see Tab. 2, where the first verb represents an initial phase of a situation represented by the second element [24]. Inchoativity is signalled derivationally, i.e. the second verb LU is a derivational basis for the first, e.g. *rozplakać się* `~to start crying` -inch.→*plakać* `to be crying`, but due to its semantic regularity was expanded to LU pairs not linked derivationally, e.g. {*aktywować się* 1 `to activate itself`} -inch.→ {*funkcjonować* 1 `to function_{imp}`, *działać* 3 `to work_{imp}`, *chodzić* 4 `~to work_{imp}`, ... , *pracować* 3 `to work_{imp}`, ...}, in all these cases *aktywować się* 1 means `to start X-ing`. In pairs with nouns, the noun names a situation whose beginning is represented by the verb, e.g. *rozplakać się* `~to start crying` -inch.→ *placz_{noun}* `cry`.

Processuality [28] can be described as `to become or to be becoming`. It is often signalled by a derivational association, but it was also expanded to a synset relation, e.g. *czerwienieć* 1= `to be becoming` *czerwony* 1 `red` or *zmieniać się* 1 `to be changing itself/yourself` = to be becoming `inny 1 `different`.

Causality is used in PWN [33] to link only verbs, in EWN is cross-categorical. pLWN causality is defined for dynamic verbs (as in EWN), and the effect, intentional or unintentional, can be represented by a dynamic or static verb, as well as any other PoS [28]. For instance *ogrzać* 1 `to warm up` means: 1) `to cause smth` *zagrzać się* 1 `to have become warm by itself`, 2) `to cause heat [*ciepło* 1]`, 3) `to cause that smth. is warm [*ciepły* 1]`, 4) `to cause that it is warm` [*spowodować, że jest ciepło* 1]`.

EWN cause relation encompasses also *cause non-factial* relations. In pLWN this type of relationship is expressed by *preceding* relation and the stricter cases of necessity by **presupposition** [28]. Presupposition is close to logical presupposition and informs about the necessity of earlier occurrence of some situation, e.g. *żyć* 1 `to live`

←pres.- *umrzeć* 1 `to die`, because if someone has died, he had to live earlier, the same if he has not died; thus *umrzeć* 1 presupposes *żyć* 1.

Preceding relation represents a possibility that one situation happens before the other one, e.g. *siedzieć* `to sit`, *stać* `to stand` ←prec- *położyć się* `to have laid down`, if someone has laid down, he could earlier sit or stand, but both precede the former.

Because during construction of plWN 3.0 we could observe problems with consistent selection between causality and presupposition/preceding, we propose here to introduce two subtypes of both: *identical* (e.g. *martwy* `dead` ←pres- *zmartwychwstać* `to resurrect`; *żona* `a wife`, *mąż* `a husband` ←prec.- *rozwiść się* `to divorce`) and *non-identical subjects* (e.g. *zabić* `to kill` ←pres.- *żyć* `to live`; *minimum* `a minimum`, *maksimum* `a maximum` ←prec.- *przekroczyć* `to overcome`). The latter subtype of presupposition and preceding resembles causality, but in opposite direction, and we assume that both can be use in parallel but in opposite directions.

Table 2. Verb relations in plWordNet 4.0.

Relation	POSS	Example	No(3.0)
inter-register synonymy	V-V	<i>pieprzyć</i> [vulgar] `~to speak nonsense` → <i>mówić</i> `to speak`	2016
aspectuality - pure - secondary	V-V	<i>napisać</i> `to write _{perf} ` - <i>pisać</i> `to write _{imperf} ` <i>popisać</i> `~to write _{perf} a little` - <i>pisać</i> `to write _{imperf} `	26558
hyponymy	V-V	<i>płynąć</i> `to swim` → <i>przemieszczać się</i> `to move`	22680
inchoativity	V-V, N	<i>rozplakać się</i> `to start crying` → <i>plakać</i> `to cry`	403
processuality	V-N, Adj, Adv	<i>owdowieć</i> `to become a widow or widower` → <i>wdowa</i> `a widow`, <i>wdowiec</i> `a widower`	729
causality	V-V, N, Adj, Adv	<i>suszyć</i> `to cause that smth. is drying` → <i>schnąć</i> `to dry`	1773
presupposition	V-V, N, Adj, Adv	<i>umrzeć</i> `to die` presupposes <i>żyć</i> `to live`	167
preceding	V-V, N, Adj, Adv	<i>siedzieć</i> `to sit` precedes <i>wstać</i> `to stand up` (but <i>wstać</i> does not presuppose <i>siedzieć</i>)	169
meronymy	V-V	<i>przełykać</i> `to swallow` is an integral part of situation <i>jeść</i> `to eat`	2829
holonymy	V-V	<i>jeść</i> `to eat` is a typical situation including <i>przełykać</i> `to swallow`	2888
multiplicativity - iterativity - distributivity	V-V	<i>grywać</i> `~to play a little from time to time` → <i>grać</i> `to play` <i>pokraść</i> `~to steal much` → <i>ukraść</i> `to steal`	542
state	V-V, N, Adj, Adv	<i>czekać</i> `to wait` → <i>gotowy</i> `ready`	93
antonymy - complementary - proper	V-V	<i>oddać</i> `to give back` - <i>zatrzymać</i> `to keep` <i>zwiększać</i> `to increase` - <i>zmniejszać</i> `to reduce`	2348
converseness	V-V	<i>sprzedać</i> `to sell` - <i>kupić</i> `to buy`	112
role inclusion - subject	V-N	<i>ocieniać</i> `to shade` ← <i>cień</i> `a shade`	1357

- instrument		<i>solić</i> ‘to salt’ ← <i>sól</i> ‘salt’	
- result		<i>portretować</i> ‘to make portrait’ ← <i>portret</i> ‘a portrait’	
- location		<i>garażować</i> ‘to keep in garage’ ← <i>garaż</i> ‘a garage’	
- object		<i>bębnić</i> ‘play drums’ ← <i>bęben</i> ‘drum’	
- time		<i>nocować</i> ‘to spend a night’ ← <i>noc</i> ‘a night’	
- indefinite		<i>akumulować</i> ‘to accumulate’ ← <i>akumulacja</i> ‘an accumulation’	
subject	V-N	<i>rżnąć</i> ‘to neigh’ → <i>koniowate</i> ‘equine’	0
object	V-N	<i>wzuwać</i> ‘to put on shoes’ → <i>but</i> ‘a shoe’	0
manner	V-Adv	<i>podsmazyć</i> ‘to fry a little’ → <i>trochę</i> ‘a little’	0
circumstance	V-N	<i>dobijać</i> ‘to reach a shore’ → <i>brzeg</i> ‘a shore’	0
derivationality	V-V, N, Adj, Adv	<i>uwznioślić</i> ‘to get exalted’ → <i>wzniosły</i> ‘exalted’	281

Meronymy (holonymy) was defined in PWN and EWN only for nouns. Because p1WN verb hyponymy does not cover cases in which a situation is an element of a larger, more general situation, verb meronymy was introduced to express such associations [28] with two subtypes of sub-situation and accompanying situation. As this distinction was not very clear in many cases, for p1WN 4.0, we have removed both subtypes and put emphasis on the necessary simultaneous co-occurrence of two situations. A verb meronym represents a more narrow situation, e.g. *chrapać* 2 ‘to snore’ -mero→ *spać* 1 ‘to sleep’. Holonymy is a reverse relation, but is not automatically defined.

3.3 Multiplicativity

Multiplicativity relation describes repetition of some state or activity, and performing an activity by many subjects or on many objects. The first subtype is called *iterativity*, and the second *distributivity*, cf [28; 49]. Iterativity links imperfective verbs, while distributivity perfective. Both are expanded also to verbs pairs that come from the same root, but which are not direct derivatives. Multiplicativity relation is more specific than hyponymy and excludes it. Multiplicative verbs are linked with their derivational bases of different aspects by the lexical relation (non-synset) of secondary aspectuality (e.g. *nakupować* 1 ‘to buy_{perf} many things’ -sec. asp.- *kupować* 1 ‘to buy_{imp}’).

3.4 Cross-categorial relations

Constitutive relations are the primary means of defining synsets and we are still looking for possibilities to increase the number of their instances and to increase the network’s density. Moreover, cross-categorial relations are valuable for many wordnet applications, including WSD. **State** relation, introduced in p1WN 2.0, in p1WN 4.0 is restricted to the state verbs. It represents being in some state, e.g. *ograniczać* 3 ‘to delimit_{imp}’ means ‘to be a border [*granica* 1]’, *czwienić się* 1 ‘to be red’ -state→ *czwony* 1 ‘red’, *czuwać* ‘to stay_{imp} awake’-state→ *światłomie*

`consciously`.

We propose four new cross-categorial relations in pLWN 4.0. **Manner** relation describes a verb by a link to an adverb describing a manner in which the activity is performed:

Jeżeli ktoś lub coś X-wał(o), to znaczy, że Z-wał(o) Y-owo. (gdzie X jest hiponimem Z).

`If smn/smth has X, it means that he/it has Z Y[ADV]', where X is a hyponym of Z'
e.g. *podgotować* 'to cook_{perf} a little' → *trochę* [Adv] 'little [Adv]';

Circumstance describes a verb by referring to an adverbial realised by a simple prepositional phrase and links the verb to the noun which is the semantic head of the prepositional phrase:

Jeżeli ktoś lub coś X-wał(o), to znaczy, że Z-wał(o) Y[Prep]. (gdzie X jest hipo. Z.)

`If smn/smth has X, it means that has Z Y[Prep]`

e.g., *dopłynąć* 'to swim_{perf} to some point/place' -circum. → *brzeg* 'a bank', where the phrase *do brzegu* 'to the bank' is an unexpressed element.

The prepositional phrase is an unexpressed element of the relation.

Subject and **object** are conceptually generalised from the morpho-semantic role relation (see Sec. 5.2) It was motivated by functional grammar associating semantics with syntax [18], e.g. subject and object relations in [6]. They characterise semantic roles of the semantic subject and object [43]. Inclusion of predicate-argument roles as definition elements can result in their proliferation and subjectivity in adding them. That is why, we limited the number of instances per verb to 3. However, we aim at adding only very characteristic links, so smaller number, better. Such links must originate from meaning definitions.

Subject: *Jeżeli mowa o tym, że ktoś lub coś X-ował(o), to ten ktoś lub to coś pewnie był(o) Y-kiem.* 'If we say that smn/smth X, then this smn/smth is certainly Y'

e.g. *muczeć* 'to moo' → *krowa* 'a cow',

Object: *Jeżeli mowa o tym, że X-owano kogoś lub coś, to ten ktoś lub to coś pewnie był(o) Y-kiem.* 'If we say that smn/smth was X, then this smn/smth had to be certainly Y' e.g. *wzuwać* 'to put shoes on' → *but* 'a shoe'.

4 Lexical Relations

Relations that do not fulfill requirements for constitutive relations are described on the level of LUs, not synsets. This type includes first of all semantic oppositions (like in PWN and most other wordnets), but also a large number of derivationally motivated relations (that is common for Slavic wordnets), called also morpho-semantic relations.

4.1 Oppositions

Opposition relations have not changed since pLWN 3.0, cf [28; 38]: **pure antonymy**, **complementary antonymy**, and **converseness**. All are restricted to verbs. Converseness, considered in PWN, but not included, cf [9] links verbs representing the same situation, but from different viewpoints and with the reverse positions of arguments in the semantic valency frame, e.g. *sprzedać* 1 'to sell' - *kupić* 1 'to buy'.

4.2 Morpho-semantic relations

Derivational relations mostly express some lexical meaning change, but derivational morphemes are often very ambiguous, cf [40]. The idea of morpho-semantic relations was proposed, e.g. [12] to overtly describe these semantic links, and used in wordnets, e.g. [48]. Morpho-semantic relations are especially numerous in Slavic wordnets, e.g. [20]. In EWN relations of this type does not need to be strictly derivationally motivated that is similar to our expansion of some morpho-semantic relations, namely: inchoativity, multiplicativity, processuality, and state to synset relations, see Sec. 4. In addition a lexico-semantic relation of **role** cf. [14] was introduced in plWN 2.0 [28], and has been preserved unchanged. Role is described only if a noun is derived from a verb (*role inclusion* describes derivation of a verb from a noun), and has 7 subtypes (see Tab. 2). In plWN 4.0 role will be registered only for imperfective verbs, as most perfective verbs are derived from imperfective. Such a solution should help to improve consistency in adding role instances and limits its artificial proliferation.

5 Potential Applications and Conclusions

The modified and extended system of verb relations will lead to a denser network and richer description of verbs. Cross-categorial links are especially important for WSD in relation to all PoSs. WSD methods based on activation spreading across the wordnet network express increasing performance with the increasing number of words from the given text context that are connected in the network.

Relations describing associations between situations in Sec. 4.2 form a kind of a system. On the basis of our former experience, we are convinced that they can be defined with good consistency and they are relatively frequent. Moreover, the current description, see Tab. 2, is definitely not complete. So, such relations should facilitate calculating graph-based text similarity, recognition of textual entailment, text semantic relations or paraphrases, as well as even some forms of reasoning.

The former, plWN 3.0, verb model was complex, and some problems have appeared in its application in wordnet development. Some elements of the model (e.g. classes and relations) were not interpreted in a way consistent enough, many relation subtypes have not been described with good coverage. Our applications of plWN (e.g. WSD, text similarity) revealed a need for further increase of the verb network density, especially in relation to cross-categorial links, e.g. there are significant differences in WSD quality depending on the density of the local wordnet subnetworks around different verbs.

The proposed modifications in the plWN verb model are more an evolution than revolution. In most cases we try to simplify the model, but in a careful way, from the perspective of the wordnet development process. We try to firmly base new proposed relations on the analysis of the language data. References to the intuition of linguists are inevitable, but not the core of the definitions. We try to perceive plWN as a multi-purpose, semi-formal description of the Polish lexical semantics systems, but necessarily built on a large scale and focused on range of applications in NLP.

Acknowledgments

This work was co-financed as a part of the investment in the CLARIN-PL research infrastructure (www.clarin-pl.eu) funded by the Polish Ministry of Science and Higher Education.

References

1. Alonge, A.: Definition of the links and subsets for verbs, EuroWordNet Project LE4003, Deliverable D006. Pisa (1996), <https://www.illc.uva.nl/EuroWordNet/docs/D006.ai>.
2. Apresjan, J.: *Semantyka leksykalna: synonimiczne środki języka*. Zakład Narodowy im. Ossolińskich, Wrocław (2000).
3. Cockiewicz, W. (2007). Na peryferiach aspektu. *LingVaria* 2(4), 9-25 (2007).
4. Comrie, B.: *Aspect: An introduction to the study of verbal aspect and related problems* (Vol. 2). Cambridge university press, Cambridge (1976).
5. Derwojedowa, M., Zawisławska, M.: Relacje leksykalne w polskiej i czeskiej bazie WordNet. In: Rudnik-Karwatowa, Z. (eds.), *Z polskich studiów slawistycznych, seria XI, Językoznawstwo*, Warszawa (2007).
6. Dik, S. C.: *Functional grammar* (Vol. 7). Foris, Dordrecht (1981).
7. Dimitrova, T., Tarpomanova, E., and Rizov, B. Coping with derivation in the Bulgarian WordNet. In *Proceedings of the Seventh Global Wordnet Conference (GWC 2014)*, pages 109–117, (2014).
8. Dowty, D.: The Effects of Aspectual Class on the Temporal Structure of Discourse: Semantics or Pragmatics? *Linguistics and Philosophy* 9(1), 37-61 (1986).
9. Fellbaum, C.: A semantic network of English verbs. In: Fellbaum (eds.), *WordNet: An electronic lexical database*, MIT Press, Cambridge (1998).
10. Fellbaum, C.: *WordNet*. Blackwell Publishing Ltd., Hoboken (1998a).
11. Fellbaum, C.: WordNet. In: Poli, R., Healy, M., Kameas, A. (eds.), *Theory and applications of ontology: Computer applications*. Springer, Heidelberg (2010).
12. Fillmore, C. J. (1967). The case for case. In Bach, E. and Harms, R.T. (eds.). *Universals in Linguistic Theory*, pp. 1-91, New York (1967).
13. Fišer, D.: A multilingual approach to building slovene wordnet. In *Proceedings of the workshop on A Common Natural Language Processing Paradigm for Balkan Languages held within the Recent Advances in Natural Language Processing Conference RANLP* (Vol. 7, p. 26) (2007).
14. Grzegorzczkova, R., Puzynina, J.: Rzeczownik. In: Grzegorzczkova, R., Wróbel H., Laskowski R. (eds.), *Gramatyka współczesnego języka polskiego. Morfologia 2*, Warszawa (1998).
15. Hamp, B., Feldweg, H.: Germanet-a lexical-semantic net for german. In: Vosen, P., Adriaens, G., Calzolari, N., Sanfilippo, A., Wilks, Y., *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pp. 9-15, Madrid (1997).
16. Hoppermann, C., Hinrichs, E.: Modeling prefix and particle verbs in GermaNet. www.aclweb.org/anthology/W14-0107 (2017/04/09).
17. Karolak, S.: *Arguments sémantiques contre la distinction: aspect / modalité d'action*, In: *Studi Italiani di Linguistica Teorica e Applicata*, anno XXII, 1993, No 2, pp. 255-284 (2003).
18. Kiklewicz, A.: *Podstawy składni funkcjonalnej*. Wydawnictwo UWM. Olsztyn (2004).
19. Kiklewicz, A.: Finitywny (teleologiczny) model aspektualności: założenia teoretyczne. *Prace Filologiczne L*, pp. 59-82 (2005).

20. Koeva, S.: Derivational and morphosemantic relations in Bulgarian wordnet. *Intelligent Information Systems* 16, 359-369 (2008).
21. Koeva S., Krstev C., Vitas D.: Morpho-semantic Relations in Wordnet – a Case Study for two Slavic Languages. In: Tanács, A., Csendes, D., Vincze, V., Fellbaum, Ch., Vossen, P. (eds.) *Global Wordnet Conference (GWC'08). Global Wordnet Conference*, 2008, pp. 239-253, Szeged, (2008).
22. Kunze, C.: Semantics of verbs within GermaNet and EuroWordNet. In: *Proceedings of 11th European Summer School in Logic, Language and Information*, pp. 189-200, Utrecht (1999).
23. Kunze, C., Lemnitzer, L.: GermaNet – representation, visualization, application. In: *Proceedings of LREC 2002, main conference, volume V*, pp. 1485–1491, Las Palmas (2002).
24. Laskowski, R.: Kategorie morfologiczne języka polskiego–charakterystyka funkcjonalna. In: Grzegorzczkowska, R., Wróbel H., Laskowski R. (eds.), *Gramatyka współczesnego języka polskiego. Morfologia 1*, Warszawa (1998).
25. Lis, M.: Annotation scheme for multimodal communication: Employing plWordNet 1.5. In: *Proceedings of the Formal and Computational Approaches to Multimodal Communication Workshop. 24th European Summer School in Logic, Language and Information (ESSLI 2012)*, (2012).
26. Lyons, J.: *Semantics*. CUP, Cambridge (1977).
27. Marcińczuk, M., Oleksy, M., Bernaś, T., Kocoń, J., Wolski, M.: Towards an event annotated corpus of Polish. *Cognitive Studies| Études cognitives* 15, 253-267 (2015).
28. Maziarz, M., Piasecki, M., Szpakowicz, S., Rabięga-Wiśniewska, J., Hojka, B.: Semantic Relations between Verbs in Polish Wordnet 2.0. *Études Cognitives/Studia Kognitywne* 11, 183-200, (2011).
29. Maziarz, M., Piasecki, M., Szpakowicz, S.: The chicken-and-egg problem in wordnet design: synonymy, synsets and constitutive relations. *Language Resources and Evaluation* 47, 769-796, (2013).
30. Maziarz, M., Piasecki, M., Rudnicka, E., Szpakowicz, S., Kędzia, P.: (2016) plWordNet 3.0 - a Comprehensive Lexical-Semantic Resource. In: Calzolari, N., Matsumoto, Y., Prasad, R. (eds.), *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16*, pp. 2259-2268, Osaka (2016).
31. Mędak, S.: *Słownik form koniugacyjnych czasowników polskich*. Universitas, Kraków (1997).
32. Młynarczyk, A.: *Aspectual Pairing in Polish*. LOT, Utrecht (2004).
33. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K. J.: Introduction to WordNet: An on-line lexical database. *International journal of lexicography* 3(4), 235-244 (1990).
34. Miller, G., Fellbaum, Ch. (eds.): *WordNet – An Electronic Lexical Database*. The MIT Press, Cambridge (1998).
35. Paducheva, E. V.: Taxonomic categories and semantics of aspectual opposition. *Temporal reference, aspect and actionality* 1, 71-89 (1995).
36. Pała, K., Smrż, P.: Building czech wordnet. *Romanian Journal of Information Science and Technology* 7(2-3), 79-88 (2004).
37. Pała K., Hlaváčková D.: Derivational relations in Czech WordNet. In: *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies (ACL '07)*, pp. 75-81, Association for Computational Linguistics, Stroudsburg, PA (2007).
38. Piasecki, M., Broda, B., Szpakowicz, S.: *A wordnet from the ground up*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław (2009).
39. Piasecki, M., Marcińczuk, M., Ramocki, R., Maziarz, M.: *WordNetLoom: a WordNet*

- development system integrating form-based and graph-based perspectives. *International Journal of Data Mining, Modelling and Management* 5(3), 210-232 (2013).
40. Piasecki, M.; Ramocki, R. & Maziarz, M.: Recognition of Polish Derivational Relations Based on Supervised Learning Scheme In Calzolari, N.; Choukri, K.; Declerck, T.; Dogan, M. U.; Maegaard, B.; Mariani, J.; Odijk, J. & Piperidis, S. (Eds.) Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA), pp. 916-922 (2012).
 41. Perlin, J.: Ile jest we współczesnej polszczyźnie czasowników dwuaspektowych?. *Linguistica Copernicana*, 3(1), 165-172 (2010).
 42. Rozwadowska, B.: Initial Boundary and Telicity in the Semantics of Perfectivity. In: Kosta, P., Błaszczak, J., Frasek, J., Geist, L., Żygis, M. (eds.) *Investigations into Formal Slavic Linguistics*. Berlin: Peter Lang, pp. 859-872 (2003).
 43. Rudyk, A.: O subiekcje semantycznym w rosyjskich zdaniach z predykatywnym wartościującym pod względem emocjonalnym w porównaniu z językiem polskim. *Acta Universitatis Lodziensis. Folia Linguistica Rossica*. (2011).
 44. Šojat, K., and Srebačić, M.: Morphosemantic relations between verbs in Croatian WordNet. In Proceedings of the Seventh Global WordNet Conference, pp. 262–267, (2014).
 45. Vendler, Z.: Verbs and times. In: Vendler, Z. (eds.) *Linguistics in philosophy*, Ithaca, Cornell University Press, pp. 97-121, New York (1957).
 46. Verkuyl, H.J. (2005): Aspectual Composition: Surveying the Ingredients. In: Verkuyl, H.J., de Swart, H., van Hout, A. (eds.), *Perspectives on Aspect [Studies in Theoretical Psycholinguistics, v. 35]*, Springer, pp 19-39, (2005).
 47. Vossen, P., Bloksma, L., Alonge, A., Marinai, E., Peters, C., Castellon, I., Rigau, G.: Compatibility in interpretation of relations in EuroWordNet. In: Vossen, P. (eds.), *EuroWordNet: A multilingual database with lexical semantic networks*, pp. 81-112, Springer Netherlands, Dordrecht (1998).
 48. Vossen, P.: EuroWordNet. General document. University of Amsterdam, Amsterdam (2002).
 49. Wróbel, H.: Czasownik. In: Grzegorzczkova, R., Wróbel, H., Laskowski, R. (eds.), *Gramatyka współczesnego języka polskiego. Morfologia 2*, Warszawa (1998).