

Connections between the semantic layer of *Walenty* valency dictionary and PLWORDNET

Elżbieta Hajnicz

Institute of Computer Science,
Polish Academy of Sciences
Warsaw, Poland

hajnicz@ipipan.waw.pl

Tomasz Bartosiak

Institute of Computer Science,
Polish Academy of Sciences
Warsaw, Poland

tomasz.bartosiak@gmail.com

Abstract

In this paper we discuss how *Walenty* is using PLWORDNET to represent semantic information. We decided to use PLWORDNET lexical units and synsets to describe both the predicate meaning and the semantic fields of its arguments. The original design decision required some further refinement caused by the structure of PLWORDNET and complex relations between arguments.

1 Introduction

Walenty, a comprehensive valency dictionary of Polish developed at the Institute of Computer Science, Polish Academy of Sciences (ICS PAS), is created to a large degree as a part of CLARIN-PL (Przepiórkowski et al., 2014a; Przepiórkowski et al., 2014b).¹ It was meant to be used both by computer programs (e.g. it is employed by two parsers of Polish, POLFIE² (Patejuk and Przepiórkowski, 2012) and Świgr³ (Woliński, 2004)) and by linguists.

The dictionary comprises above 18,000 entries (with over 101,000 schemata and 31,000 frames), including 13,000 verbs, 4,000 nouns, 950 adjectives and 200 adverbs. Therefore, nonverbal entries form 28% of the lexicon.

Walenty is composed of two main layers: syntactic and semantic. The syntactic layer was described in (Przepiórkowski et al., 2014c; Przepiórkowski et al., 2014a; Hajnicz et al., 2016b), whereas (Przepiórkowski et al., 2014b) focuses on its phraseological component. On the other hand, the semantic layer was sketched in (Hajnicz et al., 2016a).

The semantic layer of *Walenty* is strictly connected with PLWORDNET (Piasecki et al., 2009; Piasecki et al., 2016), one of two Polish wordnets.⁴

PLWORDNET describes the meaning of a lexical unit by placing this unit in a network of relations (such as synonymy, hypernymy, meronymy, etc.).

In this paper we want to focus on how semantic layer of *Walenty* was influenced by PLWORDNET and its structure.

2 Related works

There exist valency dictionaries connecting syntactic and semantic information about predicates and their arguments. The most famous is FrameNet⁵ (Fillmore et al., 2003; Ruppenhofer et al., 2006) based on a theory called Frame Semantics (Fillmore, 1976; Fillmore and Baker, 2001). It is organised around the notion of a *semantic frame* representing a situation. A semantic frame is evoked by lexical units representing corresponding meanings of words (not only verbs). Frames are lists of semantic roles called *frame elements* (FEs).

FrameNet contains about 800 hierarchically organised frames evoked by 10 000 lexical units. Frames are organised in a hierarchy which relates lexical units evoking them. Apart from a hierarchy, frames are organised into scenarios. Nevertheless, FrameNet lexical units are not related to a wordnet (in particular, Princeton WordNet, (Fellbaum, 1998; Miller and Fellbaum, 2007)) and create independent structure⁶.

Another important valency dictionary is VerbNet⁷ (Kipper-Schuler, 2005) based on the classification of verbs by Levin (1993). Each verb class in VerbNet is completely described by semantic roles, selectional restrictions on the arguments, and frames consisting of a syntactic description and semantic predicates with a temporal function. VerbNet describes about 5250 senses of 3800 verb lemmas. Each verbal sense in VerbNet may refer to a set of Wordnet senses that captured the meaning appropriate to the corresponding Levin's class

¹<http://www.clarin-pl.eu/en/>

²<http://zil.ipipan.waw.pl/LFG>

³<http://zil.ipipan.waw.pl/%C5%9Awigra>

⁴The other one is *PolNet* (Vetulani et al., 2009; Vetulani, 2014; Vetulani and Kochanowski, 2014) developed at Adam Mickiewicz University by Zygmunt Vetulani Group.

⁵<https://framenet.icsi.berkeley.edu/ndrupal/>

⁶There were several attempts to relate the resources, cf. (Cao et al., 2010).

⁷<https://verbs.colorado.edu/~mpalmer/projects/verbnet.html>

obléci¹_{pf} / **oblékat**¹_{impf} / **obléknout**¹_{pf} /
ustroit¹_{pf} / **stroit**¹_{impf}

=canbepassive yes
 =class dress-41.1.1

- 1 obléci:1 / oblékat:1 / obléknout:1
 -frame: **AG**<person:1>^{obl}_{kdo1} **VERB**
PAT<person:1>^{obl}_{komu3}
ART<garment:1>^{obl}_{co4}
 -synonym: ustroit:1 / stroit:1
 -use: prim
 -refl: obj_dat
- 2 obléci:1 / oblékat:1 / obléknout:1 /
 ustroit:1 / stroit:1
 -frame: **AG**<person:1>^{obl}_{kdo1} **VERB**
PAT<person:1>^{obl}_{koho4}
ART<garment:1>^{obl}_{do} _{zeho2}
 -synonym:
 -use: prim
 -refl: obj_ak

Figure 1: An exemplary entry of VerbaLex valency dictionary

(Dang et al., 1998; Kipper et al., 2000). Moreover, selectional restrictions are based on semantic categories labelling WordNet files. The syntactic valency information is represented by means of *LTAG* trees.

There exist several Czech valency dictionaries. Two of them, VALLEX (Lopatková et al., 2003; Žabokrtský and Lopatková, 2007) and PDT-VALLEX (Hajič et al., 2003; Urešová, 2009), are based on Functional Generative Description (Sgall et al., 1986). Despite common origins those dictionaries have been developed independently, following different approaches. While the first one tries to encompass all frames for a given lexeme, the latter is connected with Prague Dependency Treebank and has only those frames that were encountered in the corpus. In both dictionaries frames representing semantics are syntax driven, with multiple syntactic realisations of a single word meaning creating multiple (often different) frames. Nonetheless, frames are not connected to any wordnet.

A third one, VerbaLex (Hlaváčková and Horák, 2006) is connected with the Czech WordNet (Pala and Smrž, 2004; Rmbousek et al., 2017). Valency frames are connected with whole synsets, not particular lexical units. The semantic characteristic of arguments has two level representation and consists of a set of semantic roles including 40 elements from EuroWordNet top ontology (Vossen, 1998) and more precise semantic types including specific literals (lexical units) from the set of Princeton WordNet Base Concepts with relevant sense numbers. Semantic types correspond to selectional restrictions/preferences. On

the other hand, the frames are connected to Levin’s classes and hence with VerbNet.

Figure 1 presents an exemplary entry of VerbaLex. A frame corresponds to a synset containing five lexical units, but only three of them can be used in 1 as other two do not follow the same syntax.

There exist some Polish valency dictionaries as well. The most important are (Polański, 1980 1992; Świdziński, 1994). Only the first one includes semantic information, i.e. abstract selectional restrictions (cf. Figure 2, e.g. NP_A¹ has to have ‘Anim’ property, while NP_A² has to have ‘Abstr’ property). A corpus-based dictionary including some purely syntactic valency information is (Bańko, 2000).

LUBIĆ

$$NP_N \rightarrow \left\{ \begin{array}{l} NP_A^1 + \left(\left\{ \begin{array}{l} za \cap NP_A^2 \\ za \cap Ts_A, ze \cap S \end{array} \right\} \right) \\ NP_A^3 \\ zeby \cap S \\ IP \end{array} \right\}$$

NP_N → [+Hum]

NP_A¹ → [+Anim]

NP_A² → [+Abstr]

NP_A³ → $\left[\begin{array}{l} -Abstr \\ -Anim \end{array} \right] [+Abstr]$

Figure 2: Exemplary entry for the verb LUBIĆ ‘like’ in Polański’s valency dictionary

3 Basic information about the dictionary

The representation language of *Walenty* is in general universal w.r.t. parts of speech. Each lexical entry is identified by its lemma (e.g. verb GNIEWAĆ ‘irritate’ noun GNIEW ‘anger’, ‘irritation’ or adjective GNIEWNY ‘angry’, ‘irritated’).

On the syntactic level, each entry is divided into subentries according to its grammatical properties. Reflexive mark, aspect (both only for verbs), predicativity (only for adjectives and adverbs) and negativity are taken into account. For instance, the entry GNIEW has exactly one subentry **gniew** (.,), whereas GNIEWAĆ has two subentries **gniewać** (., imperf) and **gniewać się** (., imperf).

Each subentry may have any number of syntactic valency schemata⁸ assigned, each being a set of syntactic positions. A syntactic position is a set of phrase types – if two morphosyntactically different phrases may occur coordinated, they are taken to be different realisations of the same position (Szupryczyńska, 1996). Labels are used to distinguish special argument positions – subject and object (if they occur). In *Walenty* we decided that

⁸We use a term *schema* for the syntactic level representation and a term *frame* for the semantic level representation.

subject and object syntactic positions⁹ are marked only for verbs. However, there exist theories, e.g. generative ones, in which nouns, at least some of them (*derived nominals*), have (deep) subjects and objects (Chomsky, 1970). The required information can be inferred from dependencies between derivationally connected entries as both syntactic positions represent the same argument, cf. section 6. Additional label *head* was introduced in order to represent a non-local control dependency between the head of an adjective and its infinitival argument (e.g. *Szukają kompozytorów gotowych tworzyć z nimi nowoczesny teatr*. ‘[They] are looking for composers [who are] ready to create a modern theater with them.’). This matter, similarly as other issues specific for syntax of nonverbal predicates, goes beyond the scope of this article.

4 Semantic layer

The semantic layer is composed of semantic frames. Each frame is a set of semantic arguments represented as pairs ⟨semantic role, selectional preferences⟩. The set of semantic roles is presented in Figure 3 – they have colours assigned to them in a fixed way. More information about semantic roles in *Walenty* is included in (Hajnicz et al., 2016a). We assume that there cannot be two identical frames for a single entry, as otherwise there would be no way to distinguish between their meanings. This requirement does not concern frames identified by multi-word lemmas if they correspond to a different meaning.

	Initial Group	Accompanying Group	Ending Group
Main Roles	<ul style="list-style-type: none"> ■ Initiator ■ Stimulus 	<ul style="list-style-type: none"> ■ Theme ■ Experiencer ■ Factor ■ Instrument 	<ul style="list-style-type: none"> ■ Recipient ■ Result
Auxiliary Roles	<ul style="list-style-type: none"> ■ Condition 	<ul style="list-style-type: none"> ■ Attribute ■ Manner ■ Measure ■ Location ■ Path ■ Time ■ Duration 	<ul style="list-style-type: none"> ■ Purpose
Attributes	<ul style="list-style-type: none"> ■ Source 	<ul style="list-style-type: none"> ■ Foreground ■ Background 	<ul style="list-style-type: none"> ■ Goal

Figure 3: Table of *Walenty*’s roles

⁹Representation of subject and object in *Walenty* was described in (Przepiórkowski et al., 2014a).

4.1 Identification of the meaning

Each frame is connected to the meaning of a predicate. Those meanings are identified by PLWORDNET lexical units (LUs). We use PLWORDNET version 2.1, as it was the current version at the moment we started works on the semantic layer of *Walenty*.

Contrary to VerbaLex, *Walenty* frames are assigned to predicate lemmas, not to synsets. Therefore, synonyms are not related within the dictionary. This approach prevents us from overlooking some subtle differences between frames concerning selectional preferences or even presence of a particular argument (e.g. *Instrument*). The technical matter concerning potential side-effects of changes in PLWORDNET are also important.

Nevertheless, it is possible for multiple LUs to correspond to the same frame. There are three main reasons for that to happen:

1. Lexical units are derivationally connected. This includes:
 - reflexive and non-reflexive verbs, provided that they represent the same meaning (diathesis alternations, e.g. GNIEWAĆ ‘to irritate’ and GNIEWAĆ SIĘ ‘to be angry’),
 - noun and adjective derivatives of verbs (e.g. DBAĆ ‘to care’, DBAŁOŚĆ ‘a care’, DBAŁY ‘careful’ and NIEDBAŁY ‘careless’).
2. A single word describing different aspects of situation (e.g. POŻYCZAĆ can mean either ‘to borrow’ or ‘to lend’ depending on syntactic structure being a convers of itself).
3. Despite having different hypernyms, a lexical unit cannot be distinguished by semantic frame only (e.g. KOMENTOWAĆ ‘to comment’ has two lexical units in PLWORDNET – the first with hypernym KRYTYKOWAĆ ‘to criticise’ and the other with hypernym INTERPRETOWAĆ ‘to interpret’ – both taking same types of arguments, but being used in different larger contexts).

On the other hand, some lexical units may be absent in PLWORDNET. In such cases new LUs are added, indicated by capital letters instead of numbers following the lemma of an LU (wordnet standard), in order to differentiate them from the original wordnet LUs. Such new LUs are provided with glosses¹⁰ as well as potential location in PLWORDNET structure. For instance, *mleć-A* lit. ‘mill’ from Figure 5 should be a hyponym of *kręcić-4* ‘rotate’. This will facilitate including them by PLWORDNET developers.

¹⁰Original PLWORDNET LUs may have glosses in *Walenty* as well.

4.2 Selectional preferences

Arguments, identified by semantic roles, are provided with selectional preferences (Katz and Fodor, 1964; Resnik, 1993). Unlike some other dictionaries, we do not use a fixed set of qualifiers, like *abstract/concrete*, *solid/liquid/gaseous* etc. We want to be much more precise, hence we use PLWORDNET synsets (represented by LUs) and relations to represent selectional preferences. Therefore, it is *dogs* that generally BARK, we tend to DRINK *beverages* (not all *liquids*), and we prefer to use *bandages* to BANDAGE (not every *cloth*).

The selectional preferences are represented as a list of elements of the following four types (elements of different types can cooccur in the same list):

1. a PLWORDNET synset,
2. a predefined set of synsets,
3. a PLWORDNET relation to another argument,
4. a PLWORDNET relation to another synset.

The most basic way to represent selectional preferences is a direct use of PLWORDNET synsets. For instance, the frame of the verb BANDAŻOWAĆ ‘bandage’ with a strictly constrained meaning is presented in Fig. 4: *istota ludzka-1* ‘human being’ bandages *część ciała-1* ‘body part’ of *stworzenie-5* ‘creature’ by means of *bandaż-1* ‘bandage’. Contrary to VerbaLex, we use selectional preferences form a Polish wordnet, not an English one. As a consequence, no interlingual relations are required to check whether selectional preferences are satisfied in a particular sentence. However, the rich structure of PLWORDNET disallow us to use only hyponymy relation in this respect.

bandażować-1

Rama:	pewna [9873]			
Rola:	Instrument	Theme, Foreground	Theme, Background	Initiator
Preferencje selekcyjne:	bandaż-1	część ciała-1	stworzenie-5	istota ludzka-1

Figure 4: A frame for the verb BANDAŻOWAĆ with PLWORDNET selectional preferences only

In many situations, groups of PLWORDNET synsets commonly occur together in a single selectional preference. For example, both foods and drinks can be tasted or pasteurised. Similarly, both people and organisations/companies can buy, sell or store goods. What is more, people can speak about anything – objects, abstracts and situations. As such semantically connected concepts may be composed of many unrelated PLWORDNET synsets, we decided to add symbols representing such common combinations.

Table 1 lists all the predefined selectional preferences. The first column contains their labels, the second column contains their English meaning whereas the third column contains lists of corresponding PLWORDNET LUs. Such organisation of information simplifies the work of lexicographers elaborating *Walenty*, decreases its sensitivity to changes in PLWORDNET and increases the readability of the dictionary, the more so as such lists can be really long. What is most important, we can modify these lists without bothering of revising all corresponding entries. This feature has a positive impact on the cohesion of the resource.

Complicated structure of PLWORDNET (caused by specifics of Polish language) made us also introduce PLWORDNET relations to another synset as a way of representing selectional preferences. For instance, an *Instrument* for PISAĆ ‘write’ could be *a pen, a ballpen, a pencil* etc. However, in PLWORDNET their direct hypernym is *artykuł papierniczy-1* ‘writing materials’ which is evidently too wide (as it includes, e.g. ‘notebook’). They are correctly joined by the *holonymy (collection)* relation to *przybory do pisania-1* ‘writing implements’, as this term is used in Polish only in plural. This representation is equivalent to listing directly all relevant synsets, but less sensitive to changes in PLWORDNET.

For some predicates, arguments considered separately represent a wide class of entities, but actually they are closely related to each other. For instance, one meaning of MLEĆ ‘mill’ concerns objects moving their parts through some substance. For example, windmill can mill air with its sails, while water wheel can mill water with its blades (but not with sails as it has none). Classic selectional preferences tell us nothing about what can be used by those objects for milling, but we can clearly see that they have to have to be internal parts of original object. Therefore, we introduced selectional preferences determined by means of relations to another argument. Meronymy seems to be a appropriate relation here, cf. Figure 5.

mleć-A

Rama:	brak [42655]		
Rola:	Instrument, Background	Theme	Instrument, Foreground
Preferencje selekcyjne:	urządzenie-5	substancja-1	meronimia (typu część) -> [Instrument, Background]

Figure 5: Selectional preferences based on relations between arguments for the verb MLEĆ

5 Connecting both layers

In *Walenty*, syntactic and semantic valency information are represented separately. Nevertheless,

they are closely connected, but this relation is a many-to-many one. On one hand, one semantic frame can be syntactically implemented by several schemata (diathesis alternation). On the other, one schema can be used in several frames. Relating a frame and a schema we directly link semantic arguments with corresponding syntactic positions. Let us consider the verb GNIEWAĆ SIĘ ‘be angry’ / GNIEWAĆ ‘irritate’. The corresponding frame together with some schemata being its realisations are presented in Figure 6.

This is yet another difference between *Walenty* and VerbaLex. Two VerbaLex frames presented in Figure 1 differ only in the syntactic realisations of arguments. Nevertheless, the joint representation forces duplication of all information – syntactic and semantic. Moreover, lexical units involved in both syntactic realisations are connected with both frames, whereas in *Walenty* a lexical unit can label only one frame. For example, one *Walenty* frame in Figure 6 is connected to 9 verb schemata.

6 Common frames

Representation of verbs, nouns and adjectives does not differ on semantic level. What is important, derivationally connected entries of different PoSes are attached to the same frames. This is important for a correct interpretation of paraphrase. For historical reasons, this does not concern aspectual pairs.

It is worth noting that VerbNet and VerbaLex are focused solely on verbs, whereas FrameNet and PDT-VALLEX concern nouns and adjectives as well.

Let us consider the noun GNIEW ‘anger’ derivationally connected with the verb GNIEWAĆ SIĘ ‘be angry’, cf. Figure 7 (4 out of 15 schemata are visualised on the figure). Please note that the frame presented in Figures 6 and 7 is connected with the six PLWORDNET lexical units: *gniewny-1*, *gniewać-1*, *gniewać się-1*, *gniewać się-2*, *gniew-1* and *gniew-2*. This means that the frame is shared by three entries: GNIEWAĆ, GNIEW and GNIEWNY, and units representing the meaning of the current entry is written in bold.

7 Lexical units with multi-word lemmas

Walenty has a rich phraseological component (Przepiórkowski et al., 2014b). Hajnicz et al. (2016a) considers the simpler case when a lexicalised dependant does not change the meaning of a predicate and represents a fixed form of an argument (or a modifier). However, the more interesting case is when an idiomatic construction changes the meaning of the predicate, and its lexicalised dependant semantically is not an argument.

PLWORDNET contains lexical units having multi-word lemmas, and we decided to adapt this approach in *Walenty*. The semantic frame for the idiom *kraść całusa* ‘steal a kiss’ is presented in Figure 8. The fact that the frame is linked to an idiom is marked with a white rectangle with **Lemma** inside; a lexicalised dependant is marked white as well. Such phraseology appears for nonverbal entries as well¹¹. We have chosen an idiom having both verbal and nominal realisation, which is not a typical case.

LUs identifying such idioms have multi-word lemmas composed of a lemma of the main predicate (here: the verb KRAŚĆ ‘steal’) and its syntactically dependant part (here: the noun CAŁUS ‘kiss’ in accusative) in a syntactically coherent way, see Figure 8. The structure of such a lemma could be more complicated, e.g. *plakać nad rozlanym mlekiem* ‘cry over spilt milk’, cf. 9. Similarly as in the general case, such lemma can be present in PLWORDNET or added in *Walenty*.

8 Conclusions and future works

This article describes the relations between two Polish language resources PLWORDNET and *Walenty* valency dictionary. The relations appear on two levels. First, PLWORDNET lexical units are connected to each semantic valency frame as their meaning identifiers. In particular, this concerns LUs with multi-word lemmas. Moreover, synsets (represented by LUs) are used to represent selection preferences of arguments.

Walenty is based on PLWORDNET version 2.1. Therefore, one of the main future tasks is to update the connection to the current version of PLWORDNET. This will be a very complicated task due to the fact that the changes in PLWORDNET are deep, which sometimes may cause a shift of the meaning of a particular LU. We plan to apply mappings between LUs from the source and the target PLWORDNET versions and estimate their reliability comparing their neighbourhood in the net. The special attention should be paid to the LUs deleted from the PLWORDNET. On the other hand, we plan to automatically check, for all LUs added by *Walenty* developers, whether there exist relevant new PIWordNet units. The operation will be based on the synonymy/hypernymy relations. The whole procedure aims at maximal limitation of manual work.

In further future we want to connect semantically related frames of different entries in a hierarchical structure similar to hypernymy. This may involve unification of frames into a FrameNet-like hierarchy with inheritance. We are also interested in enriching the semantic layer with other semantic relations like presupposition or causation. The (morpho)syntactic level will not be influenced by these changes.

¹¹ However, most of nominal or adjectival idioms are fixed and do not open any valency positions. Such idioms are not considered in *Walenty*.

Table 1: List of predefined selectional preferences

ALL		
LUDZIE	PEOPLE	{osoba-1, grupa ludzi-1}
ISTOTY	CREATURES	{istota żywa-1, grupa istot-1}
PODMIOTY	FIRMS	{LUDZIE, podmiot-3, media-2}
WYTWÓR	ARTEFACT	{rzecz-4, wytwór-1, element-3, zbiór rzeczy-1}
JADŁO	FOOD	{pokarm-1, napój-1}
DOBRA	ESTATE	{JADŁO, mienie-1, przedmiot-1, wytwór-1, zbiór rzeczy-1}
KOMUNIKAT	COMMUNICATION	{informacja-1, wypowiedź-1}
KONCEPCJA	IDEAS	{informacja-1, wytwór umysłu-1, dzieło-2, dyscyplina-2, treść-1, zależność-3, model-1, rzecz-2, tematyka-1, struktura-2, wiedza-1, zwyczaj-1, prawo-3}
POŁOŻENIE	LOCATION	{miejsce-1, przestrzeń-1, obiekt-2}
MIEJSCE	PLACE	{lokal-1, budowla-1, rejon-1, obszar-1, państwo-1, jednostka administracyjna-1, woda-4}
OTOCZENIE	SURROUNDINGS	{powierzchnia-2, rzecz-4, wytwór-2, pomieszczenie-3, istota żywa-1}
CZAS	TIME	{chwila-1, czas-3, czas-8, godzina-3}
OBIEKTY	OBJECTS	{obiekt-2, element-3, zbiór-1}
CECHA	ATTRIBUTE	{cecha-1, zespół cech-1, atrybut-3}
CZYNNOŚĆ	ACT	{czynność-1, czyn-1}
SYTUACJA	SITUATION	{CZYNNOŚĆ, zdarzenie-2, stan-1, okoliczność-1, okoliczności-1, ciąg zdarzeń-1, działalność-1}
KIEDY	WHEN	{CZAS, SYTUACJA}
CZEMU	WHY	{CECHA, SYTUACJA, LUDZIE}
ILOŚĆ	AMOUNT	{ilość-1, rozmiar-1, rozmiar-2, jednostka-4, wielkość-6}

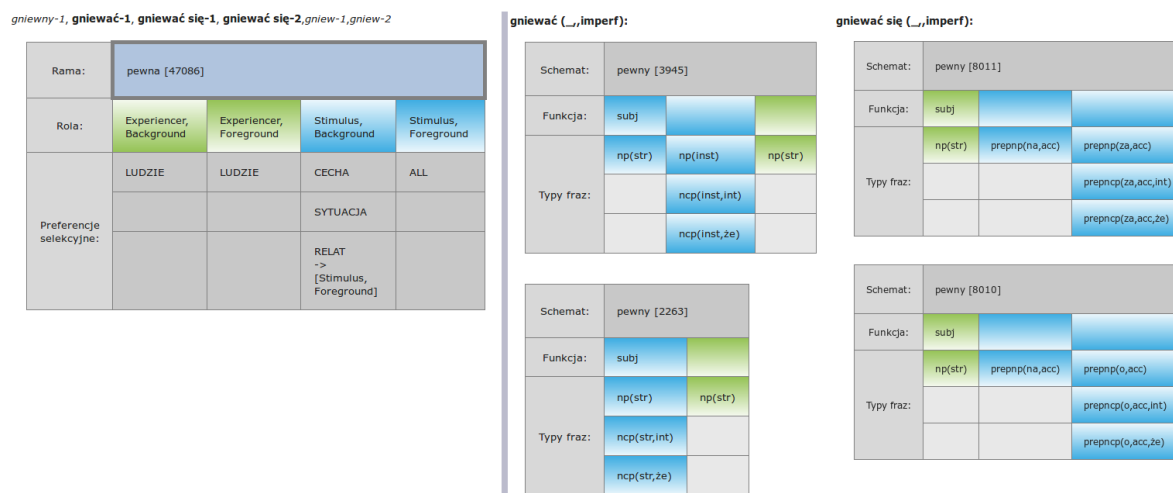


Figure 6: A screenshot with a semantic frame and schemata being its syntactic realisation

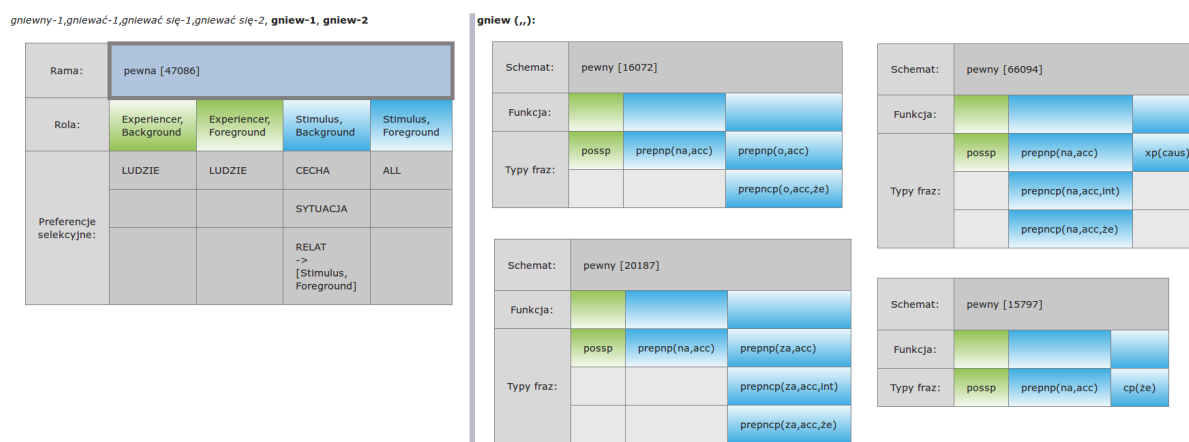


Figure 7: A screenshot with a semantic frame and schemata being its syntactic realisation form the noun perspective

(a)		(b)																			
<p><i>kradzież calusa-A, kraść calusa-A</i></p> <table border="1"> <tr> <td>Rama:</td> <td>metaforyczna [68941]</td> <td>Lemma</td> </tr> <tr> <td>Rola:</td> <td>Recipient</td> <td>Initiator</td> </tr> <tr> <td>Preferencje selekcyjne:</td> <td>istota ludzka-1</td> <td>istota ludzka-1</td> </tr> </table>		Rama:	metaforyczna [68941]	Lemma	Rola:	Recipient	Initiator	Preferencje selekcyjne:	istota ludzka-1	istota ludzka-1	<p><i>kraść (__,imperf):</i></p> <table border="1"> <tr> <td>Schemat:</td> <td colspan="2">pewny [33629]</td> </tr> <tr> <td>Funkcja:</td> <td>subj</td> <td>obj</td> </tr> <tr> <td>Typy fraz:</td> <td>np(str)</td> <td>lex(np(str),_, 'calus', natr)</td> </tr> </table>		Schemat:	pewny [33629]		Funkcja:	subj	obj	Typy fraz:	np(str)	lex(np(str),_, 'calus', natr)
Rama:	metaforyczna [68941]	Lemma																			
Rola:	Recipient	Initiator																			
Preferencje selekcyjne:	istota ludzka-1	istota ludzka-1																			
Schemat:	pewny [33629]																				
Funkcja:	subj	obj																			
Typy fraz:	np(str)	lex(np(str),_, 'calus', natr)																			
		<p><i>kradzież (__,):</i></p> <table border="1"> <tr> <td>Schemat:</td> <td colspan="2">pewny [71236]</td> </tr> <tr> <td>Funkcja:</td> <td></td> <td></td> </tr> <tr> <td>Typy fraz:</td> <td>np(dat)</td> <td>lex(np(gen),_, 'calus', atr)</td> </tr> </table>		Schemat:	pewny [71236]		Funkcja:			Typy fraz:	np(dat)	lex(np(gen),_, 'calus', atr)									
Schemat:	pewny [71236]																				
Funkcja:																					
Typy fraz:	np(dat)	lex(np(gen),_, 'calus', atr)																			

Figure 8: A frame representing idiom *kraść calusa*(a) from the verb perspective (b) schema of the noun

(a)		(b)																						
<p><i>placz nad rozlanym mlekiem-A, płacz-2, płakać nad rozlanym mlekiem-A, płakać-3</i></p> <table border="1"> <tr> <td>Rama:</td> <td>metaforyczna [156725]</td> <td>Lemma</td> </tr> <tr> <td>Rola:</td> <td>Stimulus</td> <td>Experiencer</td> </tr> <tr> <td>Preferencje selekcyjne:</td> <td>ISTOTY</td> <td>LUDZIE</td> </tr> <tr> <td></td> <td>SYTUACJA</td> <td></td> </tr> </table>		Rama:	metaforyczna [156725]	Lemma	Rola:	Stimulus	Experiencer	Preferencje selekcyjne:	ISTOTY	LUDZIE		SYTUACJA		<p><i>płakać (__,imperf):</i></p> <table border="1"> <tr> <td>Schemat:</td> <td colspan="2">pewny [28241]</td> </tr> <tr> <td>Funkcja:</td> <td>subj</td> <td></td> </tr> <tr> <td>Typy fraz:</td> <td>np(str)</td> <td>lex(preppn(nad, inst), sg, 'mleko', ratr1({lex(ppasp(agr), agr, agr, aff, 'rozlać', natr)}))</td> </tr> </table>		Schemat:	pewny [28241]		Funkcja:	subj		Typy fraz:	np(str)	lex(preppn(nad, inst), sg, 'mleko', ratr1({lex(ppasp(agr), agr, agr, aff, 'rozlać', natr)}))
Rama:	metaforyczna [156725]	Lemma																						
Rola:	Stimulus	Experiencer																						
Preferencje selekcyjne:	ISTOTY	LUDZIE																						
	SYTUACJA																							
Schemat:	pewny [28241]																							
Funkcja:	subj																							
Typy fraz:	np(str)	lex(preppn(nad, inst), sg, 'mleko', ratr1({lex(ppasp(agr), agr, agr, aff, 'rozlać', natr)}))																						

Figure 9: A frame representing idiom *płakać nad rozlanym mlekiem*

Acknowledgements Work financed as part of the investment in the CLARIN-PL research infrastructure funded by the Polish Ministry of Science and Higher Education

References

Mirosław Bańko, editor. 2000. *Inny słownik języka polskiego*. Wydawnictwo Naukowe PWN, Warsaw, Poland.

Diego De Cao, Danilo Croce, and Roberto Basili. 2010. Extensive evaluation of a framenet-wordnet mapping resource. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC-2010)*, pages 2752–2757, Valetta, Malta. ELRA.

Noam Chomsky. 1970. Remarks on nominalization. In Roderic A. Jacobs and Peter S. Rosenbaum, editors, *Readings in English transformational grammar*, pages 184–221. Ginn and Company, Waltham, MA.

Hoa Trang Dang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. 1998. Investigating regular sense extensions based on intersective Levin classes. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics COLING-ACL'98*, pages 293–299, Montreal, Canada.

Christiane Fellbaum, editor. 1998. *WordNet — An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Charles J. Fillmore and Colin F. Baker. 2001. Frame semantics for text understanding. In *Proceedings of the WordNet and Other Lexical Resources Workshop*, Pittsburgh. NAACL.

Charles J. Fillmore, Christopher R. Johnson, and Miriam R. L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16(3):235–250.

Charles J. Fillmore. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, volume 280, pages 20–32. John Wiley.

Jan Hajič, Jarmila Panevová, Zdeňka Urešová, Alevtina Bémová, Veronika Kolářová, and Petr Pajas. 2003. PDT-VALLEX: Creating a large-coverage valency lexicon for treebank annotation. *Mathematical Modelling in Physics, Engineering and Cognitive Science*, 9:57–68.

Elżbieta Hajnicz, Anna Andrzejczuk, and Tomasz Bartosiak. 2016a. Semantic layer of the valence dictionary of Polish *Walenty*. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC-2016)*, pages 2625–2632, Portorož, Slovenia. ELRA.

Elżbieta Hajnicz, Agnieszka Patejuk, Adam Przepiórkowski, and Marcin Woliński. 2016b. *Walenty: słownik walencyjny języka polskiego z bogatym komponentem frazeologicznym*. In Karolina Skwarska and Elżbieta Kaczmarek, editors, *Výzkum slovesné valence ve slovanských zemích*, pages 71–102. Slovanský ústav Akademie věd ČR, Prague, Czech Republic.

Dana Hlaváčková and Aleš Horák. 2006. VerbaLex — new comprehensive lexicon of verb valences for Czech. In *Proceedings of the Third International Seminar on Computer Treatment of Slavic and East European Languages*, pages 107–115, Bratislava, Slovakia.

- J. J. Katz and J. A. Fodor. 1964. The structure of a semantic theory. In J. A. Fodor and J. J. Katz, editors, *The Structure of Language*, pages 479–518. Prentice Hall.
- Karin Kipper, Hoa Trang Dang, and Martha Palmer. 2000. Class-based construction of a verb lexicon. In *Proceedings of the 17th National Conference on Artificial Intelligence*, pages 691–696, Austin, TX. AAAI Press.
- Karin Kipper-Schuler. 2005. *VerbNet: A broad coverage, comprehensive verb lexicon*. PhD thesis, Computer and Information Science Department, University of Pennsylvania.
- Beth Levin. 1993. *English verb classes and alternation: a preliminary investigation*. University of Chicago Press, Chicago, IL.
- Markéta Lopatková, Zdeňek Žabokrtský, Karolina Skwarska, and Václava Benešová. 2003. VALLEX 1.0 valency lexicon of Czech verbs. Technical Report TR-2003-18, ÚFAL/CKL MFF UK, Prague, Czech Republic.
- George A. Miller and Christiane Fellbaum. 2007. WordNet then and now. *Language Resources and Evaluation*, 41:209–214.
- Karel Pala and Pavel Smrž. 2004. Building the Czech WordNet. *Romanian Journal of Information Science and Technology*, 7(2–3):79–88.
- Agnieszka Patejuk and Adam Przepiórkowski. 2012. Towards an LFG parser for Polish. an exercise in parasitic grammar development. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC-2012)*, pages 3849–3852, Istanbul, Turkey. ELRA.
- Maciej Piasecki, Stanisław Szpakowicz, and Bartosz Broda. 2009. *A Wordnet from the Ground Up*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, Poland.
- Maciej Piasecki, Stan Szpakowicz, Marek Maziarz, and Ewa Rudnicka. 2016. PIWordNet 3.0 – Almost There. In *Proceedings of the 8th International WordNet Conference (GWC 2016)*, pages 290–299, Bucharest, Romania. Global Wordnet Association.
- Kazimierz Polański, editor. 1980–1992. *Słownik syntaktyczno-generatywny czasowników polskich*, volume I–V. Zakład Narodowy imienia Ossolińskich, Wrocław · Warszawa · Kraków · Gdańsk, Poland.
- Adam Przepiórkowski, Elżbieta Hajnicz, Agnieszka Patejuk, Filip Skwarski, Marcin Woliński, and Marek Świdziński. 2014a. Walenty: Towards a comprehensive valence dictionary of Polish. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC-2014)*, pages 2785–2792, Reykjavík, Iceland. ELRA.
- Adam Przepiórkowski, Elżbieta Hajnicz, Agnieszka Patejuk, and Marcin Woliński. 2014b. Extended phraseological information in a valence dictionary for NLP applications. In *Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014)*, pages 83–91, Dublin, Ireland.
- Adam Przepiórkowski, Filip Skwarski, Elżbieta Hajnicz, Agnieszka Patejuk, Marek Świdziński, and Marcin Woliński. 2014c. Modelowanie własności składniowych czasowników w nowym słowniku walencyjnym języka polskiego. *Polonica*, XXXIII:159–178.
- Philip Resnik. 1993. *Selection and Information: A Class-Based Approach to Lexical Relationships*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA, December.
- Adam Rmbosek, Karel Pala, and Sandra Tukačová. 2017. Overview and future of Czech WordNet. In *LDK Workshops: OntoLex, TIAD and Challenges for Wordnets*, pages 146–151, Galway, Ireland.
- Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, and Christopher R. Johnson. 2006. *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, CA.
- Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. D. Reidel, Dordrecht, Holland.
- Maria Szupryczyńska. 1996. Problem pozycji składniowej. In Krystyna Kallas, editor, *Polonistyka Toruńska Uniwersytetowi w 50. Rocznice Utworzenia UMK*, Językoznawstwo, pages 135–144. Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń, Poland.
- Marek Świdziński. 1994. *Syntactic Dictionary of Polish Verbs*. Uniwersytet Warszawski / Universiteit van Amsterdam.
- Zdeňka Urešová. 2009. Building the PDT-Vallex valency lexicon. In *Proceedings of the 5th Corpus Linguistics Conference*. University of Liverpool.
- Zygmunt Vetulani and Bartłomiej Kochanowski. 2014. “PolNet – polish wordnet” project: PolNet 2.0 – a short description of the release. In *Proceedings of the 7th International WordNet Conference (GWC 2014)*, pages 400–404, Tartu, Estonia. University of Tartu.
- Zygmunt Vetulani, Justyna Walkowska, Tomasz Obrębski, Jacek Marciniak, Paweł Konieczka, and Przemysław Rzepecki. 2009. An algorithm for building lexical semantic network and its application to PolNet — Polish WordNet project. In Zygmunt Vetulani and Hans Uszkoreit, editors, *Human Language Technology. Challenges of the Information Society. 3rd Language & Technology Conference*, volume 5603 of *LNAI*, pages 369–381. Springer-Verlag. Revised Selected Papers.
- Zygmunt Vetulani. 2014. PolNet – Polish WordNet. In Zygmunt Vetulani and Joseph Mariani, editors, *Human Language Technology Challenges for Computer Science and Linguistics. LTC 2011*, volume 8387 of *LNAI*, pages 408–416. Springer-Verlag. Revised Selected Papers.

- Piek Vossen, editor. 1998. *EuroWordNet: a multilingual database with lexical semantic network*. Kluwer Academic Publishers, Dordrecht, Holland.
- Marcin Woliński. 2004. *Komputerowa weryfikacja gramatyki Świdzińskiego*. PhD thesis, Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland.
- Zdeňek Žabokrtský and Markéta Lopatková. 2007. Valency information in VALLEX 2.0: Logical structure of the lexicon. *The Prague Bulletin of Mathematical Linguistics*, 87:41–60.