A (Non)-Perfect Match: Mapping plWordNet onto Princeton WordNet

Ewa Rudnicka, Wojciech Witkowski, Maciej Piasecki
G4.19 Research Group, Department of Computational Intelligence
Wrocław University of Technology, Wrocław, Poland
University of Wroclaw, Wroclaw, Poland
{ewa.rudnicka,maciej.piasecki}@pwr.edu.pl;wojciech.witkowski@uwr.edu.pl

Abstract

The paper reports on the methodology and final results of a large-scale synset mapping between plWordNet and Princeton WordNet. Dedicated manual and semi-automatic mapping procedures as well as interlingual relation types for nouns, verbs, adjectives and adverbs are described. The statistics of all types of interlingual relations are also provided.

1 Introduction

The goal of this paper is to present the machinery and guiding ideas behind a large-scale synset mapping between the Polish plWordNet (henceforth, plWN) and the English Princeton WordNet (henceforth, PWN). The resulting mapping is unique in terms of its character, scale and methodology. First, it is probably the only mapping built between the two very large wordnets constructed completely independently of each other (Fellbaum, 1998), (Piasecki et al., 2009), that is with no form of PWN content (synset) translation or structure mapping (in plWN construction)[1]. Second, it employs and further extends the whole array of inter-lingual inter-wordnet relations proposed for EuroWordNet by Vossen (2002), yet never fully implemented. Third, it is a manual mapping partially enhanced by automatic prompt systems yet not relying on them (Kędzia et al., 2013), (Rudnicka et al., 2015).

As such, the mapping had a lot of challenges: (partially) different wordnet construction methodologies of plWN and PWN; profound cross-linguistic differences between the synthetic Polish and the analytic English language; numerous cultural, sociological and historical differences between the two language communities affecting their lexicons to a large extent. Still, it seems the above challenges have been successfully met, proven by a number of cross-linguistic applications of the bilingual Polish-English wordnet.

In the paper we describe the systems of interlingual relations proposed and implemented for noun, adjective, adverb and verb synsets, presented in a chronological order motivated by the increasing difficulty of the milestones of the project. We close with the current statistics of interlingual synset relations.

2 Related works

WordNet started off as a lexico-semantic network for English (Fellbaum, 1998). With a quickly manifested potential both for linguistics and NLP research, it soon found its followers for other languages, e.g. GermaNet for German, (Hamp and Feldweg, 1997). Since cross-linguistic applications are always welcome, the idea of linking monolingual wordnets into a multilingual network naturally arose. It was put into practice in the EuroWordNet project (Vossen, 1998, 2002) and more recently in the OpenMultilingualWordnet project (Bond and Foster, 2013). Even before the start of EuroWordNet project, it was clear that constructing a wordnet from scratch and later linking it to similar resources is time and money-consuming. Few teams could afford it. On the other hand, taking Princeton WordNet as a basic template and expanding on it proved much more economical in terms of the investment needed. The two approaches were called merge and expand, respectively. Yet, science does not operate on a simple win-lose model. The expand approach rests on a

[1] The very model of plWordNet, clearly wordnet-like, is unique in several aspects in comparison to Princeton WordNet, cf Maziarz et al. (2013a)

long-abandoned assumption of the universality of mental lexicon (von Fintel and Matthewson, 2008). Thus, expand wordnets are useful language resources, but the accuracy of the specific language structure of the internal relation network may be arguable. Examples of expand wordnets are MultiWordNet (Bentivogli and Pianta, 2004), AsianWordNet (Robkop et al., 2009), IndoWordNet (Sinha et al., 2006), Open Dutch WordNet (Postma et al., 2016), and sloWNet for Slovenian (Fišer and Sagot, 2015). The few wordnets built through the merge approach are GermaNet (Hamp and Feldweg, 1997), DanNet, built on the basis of a Danish dictionary (Pedersen et al., 2009), and RuWordNet, constructed by a semi-automatic transformation of a Russian thesaurus (Loukachevitch et al., 2016).

plWordNet was constructed from scratch with a method exceptional in the wordnet world. It relied on the extraction of information about lexico-semantic relations from large text corpora (Piasecki et al., 2009). Automatically extracted relations and structures were presented as prompts to trained and supervised lexicographers who verified them with the help of reliable language resources for Polish (such as dictionaries, encyclopedia etc.) Thus, the method can be classified as a variant of the merge approach. As a consequence, the link to PWN had to be provided independently of plWN construction.

3 Nouns – bottom up

The prerequisite for mapping plWordNet onto Princeton WordNet was a reasonably advanced state of development of the former. Thus, we started with plWN 2.0 (Maziarz et al., 2013a) and PWN 3.1, see Table 1. Despite the bigger number of lemmas and lexical units in PWN, the number of synsets was comparable. It was especially visible for nouns, see Table 2.

The first challenge of mapping were partly different philosophies behind the construction of plWN and PWN (Maziarz et al., 2013a). In plWN, synset is rigidly defined as a set of lexical units sharing a set of constitutive relations (hypo/hypernymy, mero/holonymy, antonymy) (Maziarz et al., 2013b). This explains more fine-grained sense differentiation in plWN than in PWN, where synset membership is more arbitrary (Fellbaum, 1998). The second challenge of mapping were substantial cross-linguistic differences between English and Polish that also found its reflection in relation structures of the two wordnets. To explore the synthetic character of Polish, new types of lexical unit relations (PWN morphosemantic links) were added to plWN, such as, for instance, *diminutivity* (e.g. *pies* 2 - 'a dog' - *piesek* 1 - 'a small or young dog'), or *cross-categorial synonymy* (e.g. *piesek* 1 - *pieskowy* 1 - '[ADJ] related to a small dog'. The latter relation is established between a base and its derivative of a different POS when they relate to the same concept. In addition, PWN provided short definitions called *glosses*, sometimes followed by examples, for every synset, while plWN started adding glosses for lexical units (not for synsets) only at a later stage[2] of its development around the 3.0 version.

With the above challenges in mind, the guiding idea of the mapping was to link nodes of wordnet graphs that would mainly correspond in terms of relation structures (and possibly also with respect to glosses). This turned out a non-trivial task. Often, even for closely related concepts, their relation structures were partially or wholly different (Rudnicka et al., 2012). This explains the use of different types of interlingual relations (I-relations) going far beyond interlingual synonymy defined as *Simple Equivalence* by (Vossen, 2002). Most of Vossen's *Complex Equivalence* relations were adopted (e.g. I-hypo/hypernymy, I-near-synonymy, or I-mero/holonymy). Some new ones were added too, such as, for instance, *interlingual inter-register synonymy.*

Of the four parts of speech described by plWN and PWN, nouns share the most in terms of the fundamentals of their internal synset relation structure (Maziarz et al., 2013a). The basic relation is hypo/hypernymy, followed by mero/holonymy and near-synonymy. Therefore, the set of I-relations and the mapping procedure were first defined for and applied to noun synsets. It consisted of I-synonymy, I-partial-synonymy,

[2] However, even at the earlier stages of plWN development editors could add comments that were visible to other editors and facilitated the identification of the intended meaning of lexical units. Such comments were later transformed into the first version of glosses.

Elements	plWordNet 2.0	Princeton WordNet 3.1
Synsets	116 323	117 659
Lexical units	160 100	206 978
Lemmas	106 438	155 593

Table 1: Basic statistics for plWordNet 2.0 and Princeton WordNet 3.1

Elements	plWordNet 2.0	Princeton WordNet 3.1
Synsets	80 037	82 115
Lexical units	109 967	146 347
Lemmas	77 662	117 798

Table 2: Basic statistics for nouns in plWordNet 2.0 and Princeton WordNet 3.1

I-inter-register synonymy, I-hypo/hypernymy and I-mero/holonymy (Rudnicka et al., 2012). Later, it was supplemented by *I-type/instance* to link proper names, especially from PWN (Dziob et al., 2019).

For the first stage of mapping, we chose nouns from semantic domains (PWN lexicographer files) such as person, artefact, food, place, time, and names connected with thinking and communication, so as to start with concrete nouns, more likely to have unique referents regardless of a language, and then move to abstract nouns, for which the reference is often more culturally and socially dependent. We decided to go 'bottom up' in a wordnet graph. Such a move was motivated by the idea to start with the most specific, possibly unambiguous part of plWN which would form a basis for the further mapping. Also, we were trying to cover the whole branches of the hyponymy tree. The mapping direction was plWN-PWN.

4 Adjectives and adverbs – from hierarchy to dumbbell and island

With the mapping of nouns in progress, we proceeded to adjectives. The biggest challenge of adjective mapping were very different models of their internal synset relation structure in plWordNet and Princeton WordNet, shown in Table 4. As for numbers, plWN had approximately twice as much adjective synsets as PWN at the start of mapping, see Table 3.

Adjectives in plWN follow the same hierarchical, hyponymy-based model as nouns[3]. The subsidiary relations are: *gradability*, *near-synonymy* and *modifier* (Maziarz et al., 2012). On the contrary, adjectives in PWN are organized around *Similar to* relation with central and peripheral adjectives in the so called dumbbell model (Miller, 1998, Sheinman et al., 2013). *Similar to* relation is rather vague, but can be re-interpreted as one level hyponymy with a central adjective functioning as a hypernym of its peripheral adjectives. A subsidiary relation is *Member of this domain*. These profound differences in synset relation structures made us look into lexical unit relations. The latter exhibit more correspondence: antonymy/antonym, cross-categorial synonymy/pertainym, derivativity/derivationally related form. Therefore, in the mapping process we decided to consider both types of internal relations as well as the already existing I-relations between noun synsets, because some of the adjective relations are cross-categorial relations to nouns (e.g. cross-categorial synonymy, derivativity, Member of this domain). We proposed the following set of I-relations for adjective synsets: I-synonymy, I-hypo/hypernymy, I-partial synonymy, I-inter-register synonymy, and I-cross-categorial synonymy (Rudnicka et al., 2015), (Rudnicka et al., 2016). The latter relation was always used together with I-hyponymy to keep the POS information, and, in the case of very general I-hyponyms, to give more specification to the meaning of a mapped adjective. The use of such a pair of relations also allowed us to make up for the difference in size between plWN and PWN.

Having mapped a substantial part of adjective synsets, we moved to adverbs. Again, adverbs are more numerous in plWordNet than in Princeton WordNet, with twice as much lemmas, almost three times more lexical units and almost four times more synsets, as illustrated

[3]The placement of adjectives within specific hyponymy trees is conditioned on substitution tests and verified in corpora.

Elements	plWordNet 2.2	Princeton WordNet 3.1
Synsets	38 868	18 185
Lexical units	45 514	30 072
Lemmas	26 961	21 808

Table 3: Basic statistics for adjectives in plWordNet 2.2 and Princeton WordNet 3.1

Relation	plWordNet 2.2	Princeton WordNet 3.1
Value (of the attribute)	9658	639
Modifier	2 108	—
Hyponymy	18 225	—
Gradability	991	—
Near-synonymy	1 308	—
Similar to	—	21 434
Member of this domain	—	1 418

Table 4: Counts for adjective synset relations in plWordNet 2.2 and Princeton WordNet 3.1

in Table 5. Adverbs are unique in Princeton WordNet in that they have no synset relations. On the other hand, in plWordNet many adverbs were systematically derived from adjectives (Maziarz et al., 2016), hence they have a similar set of synset relations as adjectives (apart from Modifier), see Table 6. Therefore, we decided to take advantage of a previously established network of I-relations between adjectives in plWN and PWN and constructed automatic prompts for adverb synsets on the basis of internal relations between adjectives and adverbs and I-relations between adjectives. Those automatic prompts were later manually verified by lexicographers, for details see Section 6. The I-relations used were I-synonymy, I-hypo/hypernymy, I-partial synonymy, and I-inter-register synonymy.

5 Verbs – from lexico-grammatical hierarchy to semantic fields

Verbs were the last category to map due to their most complex and divergent relation structures resulting from substantial differences in rendering aspect and other verbal categories in English and Polish. Polish lexicalises aspect and in plWN perfective and imperfective verb forms always land in separate synsets (even in the case of pure aspectual pairs e.g. *czytać* - 'read/be reading' vs *przeczytać* - 'have read'). This partly accounts for bigger number of verbal elements in plWordNet 3.1 in comparison to Princeton WordNet 3.1 (see Table 7).

Similarly to nouns, verb synsets are organised around hypo/hypernymy relation both in plWN and in PWN. Other PWN verb relations include Verb group, Member of this domain (mainly Topic), and, relatively sparse, Entailment and Cause. At the start of verb mapping, plWN 3.1 used Causation, Processuality, and less numerous Distributivity, Inchoativity and Iterativity (Table 8). plWN verbs are also grouped into verb classes. These are drawn from situation types (Aktionsart (Vendler, 1967)) the verbs denote and their grammatical aspect. Class assignment is based on verb's meaning as it is evoked by a clausal context. Vendlerian classification served as the basis for creating artificial synsets whose function is to provide systematic hierarchical organization of verb synsets in plWN. Vendlerian *Activities, Achievements, and Accomplishments* are subsumed under Dynamic verbs. These are further subdivided into: distributive, cumulative, perdurative and delimitive, based on the meaning of the prefixes that attach to verbal roots. Vendler's *States* correspond to Stative verbs in plWN. Additionally, plWN distinguishes Action verbs which include: perfective forms of non-distributive, non-cumulative, non-perdurative, and non-delimitive verbs; imperfective forms of distributive, cumulative, perdurative, and delimitive verbs, and imperfective verbs with causative, procesual, inchoative, and completive meanings.

Thus, verb mapping had to overcome (partly) non-congruent internal relation networks and specific linguistic differences. Therefore, we decided to use I-synonymy, I-inter register synonymy, and I-hypo/hypernymy and introduce new interlingual relations specific to verb mapping. They were based on (Wiland, 2011) and include: I-attenuativity (to V to a lesser

Elements	plWordNet 3.1	Princeton WordNet 3.1
Synsets	11 396	3 625
Lexical units	14 207	5 592
Lemmas	8 113	4 475

Table 5: Basic statistics for adverbs in plWordNet 3.1 and Princeton WordNet 3.1

Relation	plWordNet 3.1	Princeton WordNet 3.1
Value (of the attribute)	4302	—
Fuzzynymy	9280	—
Hyponymy	10 082	—
Gradability	690	—
Near-synonymy	647	—

Table 6: Counts for adverb synset relations in plWordNet 3.1 and Princeton WordNet 3.1.

extent), e.g. *podczytać* 1 'to read a little' - *read* 1 'to interpret something that is written or printed', I-iterativity (to V repeatedly), e.g. *czytywać* 1 - 'to read from time to time' - *read* 1, I-perdurativity (to V for a period of time), e.g. *zaczytać się* 1 'to be continuously engaged in reading' - *read* 1, I-delimitivity (to V to a point in time), e,g, *poczytać* 2 'to spend some time reading' - *read* 1, I-inchoativity (onset of an action or state), e.g. *zaczytać* 2 ' to start reading' - *read* 1, I-completivity (completion of an action), e.g. *doczytać* 1 'to read to the end' - *read* 1, I-cumulativity (to V to a satisfying extent), e.g. *naczytać się* 1 'to read a lot, so that one does not want to read anymore' - *read* 1, I-distributivity (to V among many recipients), I-excessivity (to V to an excessive extent), I-causativity (to cause V), I-processuality (to become V), I-terminativity (termination of an action), I-anticausativity (be in a state caused by V), I-stativity (be in a state denoted by V), I-ablativity (V from), and I-allativity (V to). In addition, we proposed three types of cross-categorial relations: I-cross-categorial processuality, I-cross-categorial stativity, and I-cross-categorial causativity, which are always coupled with I-hyponymy relation. The function of verb-specific I-relations is to render the meaning correspondence as accurate as possible.

6 Procedures and tools

The entire mapping has been performed manually by a team of trained bilingual lexicographers supervised by senior lexicographers (Rudnicka et al., 2012). The actual mapping has been taking place in a custom-designed wordnet editing system called WordNetLoom which allows to visualise wordnet graphs for different languages on the same screen, manipulate them, compare their fragments and establish relations (Piasecki et al., 2010, Naskręt et al., 2018). The fact that an editor can see the relation structures for both languages[4], interactively explore them in any direction and depth, and make changes, e.g. by adding I-relation links directly to the graphs, noticeably facilitated the mapping process. Moreover, lexicographers' work has been monitored via another custom-designed tool, namely the WordNet Tracker system (Naskręt et al., 2018) documenting every action of a lexicographer in real time. Due to the scale of the project and its financing conditions, we worked in 1+1 model (a lexicographer establishing I-links plus a supervisor checking their adequacy).

The manual mapping procedure was first designed for nouns, but its basics have been kept for other parts of speech as well (Rudnicka et al., 2012, 2016). It consists of three main stages: (1) recognising the sense of a source synset, (2) searching candidates for a target synset, and (3) choosing a target synset and a type of interlingual relation. In the first stage, we carefully analyse the source synset internal relation structure, gloss, examples as well as interlingual relations if there are any within the close nodes in its hyponymy tree. In the second stage, candidates for a target synset are nominated on the basis of a bilingual linguist's intuition and information found in Polish-English language resources. Next, candidates for a tar-

[4]Automatically generated suggestions for I-relation links are also presented on the same screen, but marked as different kinds of relations – 'generated'.

Elements	plWordNet 3.1	Princeton WordNet 3.1
Synsets	29 110	13 789
Lexical units	40 181	25 061
Lemmas	19 836	11 540
Monosemous lemmas	11 265	6 284
Polysemous lemmas	8 571	5 256

Table 7: Basic statistics for verbs in plWordNet 3.1 and Princeton WordNet 3.1

Relation	plWordNet 3.1	Princeton WordNet 3.1
Hyponymy	31 784	13 251
Hypernymy	31 784	13 251
Causation	3 427	—
Processuality	1 204	—
Distributivity	676	—
Inchoativity	519	—
Iterativity	148	—
Entailment	—	406
Cause	—	214

Table 8: Counts for verb synset relations in plWordNet 3.1 and Princeton WordNet 3.1

get synset are analysed in a similar fashion as it is done for a source synset. In the third stage, the target synset is chosen, and depending on the degree and type of correspondence between the source and target synset, an interlingual relation is chosen and the two synsets get linked.

The results of the first stage of mapping of nouns were used as the input to an automatic prompt system developed for further stages of noun mapping (Kędzia et al., 2013). The system was based on the Relaxation Labeling algorithm of (Daudé et al., 1999). It mirrors the manual mapping procedure to the extent that it compares parts of a wordnet graph and suggests the closely related fragments. Next, Polish-English synset pairs produced by the algorithm were filtered by the so called cascade Polish-English dictionary[5] and pairs of synsets whose lemmas appear as dictionary equivalents were given the status of automatic prompts and presented to lexicographers as special links in the WordNetLoom system.

Moving to the mapping of adjectives, we could not resort to the automatic prompt system developed for nouns, because Relaxation Labeling algorithm used there requires parallel hierarchical structures to operate. Such structures are missing in the case of PWN adjectives. That made us look for other solutions. Despite superficially divergent models in plWN and PWN, we searched for common points in the relation structure both at synset and lexical unit level. As a consequence, two rule-based algorithms were designed that produced automatic prompts for the first stage of adjective mapping (Rudnicka et al., 2015). The first one relied on synset relations exclusively, the second one on synset and lexical unit relations. Both also took advantage of I-synonymy relations between noun synsets (provided the latter were internally linked to adjectives). Pairs of Polish-English lemmas (from the pairs of adjective synsets generated by the algorithms) were automatically verified by the cascade dictionary. Those recorded in the dictionary were presented to lexicographers in the form of prompts for manual mapping, (Rudnicka et al., 2015).

The procedures and relations developed for adjectives also found its use in the mapping of adverbs (Maziarz et al., 2016). In plWN many adverbs were automatically derived from adjectives. That allowed to generate automatic prompts for adverb mapping on the basis of adjective mapping. It consisted in copying interlingual relations established for adjective synsets to adverb synsets provided that the latter were systematically derived from the former in plWN. Another necessary condition was that target PWN adverbs were also derived from PWN adjectives already linked by an interlingual relation to plWN adjectives. Next, the prompts were verified by lexicographers and manual links were established. At

[5] Bilingual Cascade Dictionary is a collection of dictionaries organised in a cascade with the top-most dictionaries having the highest priority in applications.

the same time, adjective links were critically evaluated and modified, when necessary. In the overwhelming majority of cases automatic prompts were valid and a manual relation was established. More work was required for cases of one-to-many mapping (e.g. one synset serving as a hypernym for several other wordnet's synsets), while the most difficult cases constituted adverbs that were not derived from plWN adjectives holding I-relations to PWN adjectives. These required independent search for target synsets.

As for verb mapping, no automatic prompt system was designed. Although both plWN and PWN verb relation networks are hierarchical, these hierarchies are based on non-congruent prerequisites. Moreover, verb synsets are also linked via non-hierarchical relations, different in both wordnets. These differences combined with linguistic differences between Polish and English aspect morphology enforced a fully manual mapping procedure. The main focus in the procedure was put on providing as close meaning correspondence as possible. This was achieved by finding the most suitable plWN – PWN synset pair and choosing the I-relation that most adequately captures the meaning relation. The selection of I-relations was hierarchical. I-synonymy and I-inter register synonymy were prioritized. For verbs, we have exceptionally allowed for double synonymy in the case of pure aspectual pairs of verbs in Polish linked to aspectually unmarked English verbs (creating 2 – to – 1 mapping). Verb-specific relations were selected if I-synonymy and I-inter register synonymy relations could not be established and the prefix of the plWN verb carried a relevant facet of meaning. I-hyponymy relation was treated a 'last resort' relation, as it provided the most general meaning correspondence. In the cases in which PWN lacked a relevant verb synset, but a noun or adjective synset which would be used in a copula-construction in English was present, I-cross-categorial relations coupled with I-hyponymy relation linking plWN and PWN verb synsets were selected.

7 Result: a bilingual network and its applications

The result of bidirectional mapping of plWordNet and Princeton WordNet is a large Polish-English wordnet with almost 300k of unique interlingual relations. The counts of all types of I-relations are shown in Table 9. We can see that despite the priority of I-synonymy in the mapping procedure it is strongly overruled by I-hyponymy for all parts of speech. This tendency has been observed since the beginning of mapping, e.g. (Rudnicka et al., 2012, Maziarz et al., 2016), and is caused by independent methodologies and times of the two wordnets' construction leading to partly different relation structures and vocabulary coverage. Moreover, plWordNet currently outgrows Princeton WordNet in the number of synsets (and other basic elements) for all parts of speech [6]. Another reason are profound lexico-grammatical differences between English and Polish (e.g. systematic lexicalisation of aspect, gender and other grammatical categories) and socio-cultural differences between the two language communities resulting in lexical gaps (such as names of meals, administrative divisions and posts, or related to history (e.g. the Communist period or the WWII)).

The quality of the created resource has been confirmed by a number of applications, fostered by its open wordnet licence [7]. These include language learning and teaching (e.g. the creation of didactic tools such as CloudNet Word Cloud Generator [8]), dictionary making and machine translation (a component for PONS, Glosbe, Kamusigold, Ling.pl, BabelNet, Open Multilingual Wordnet, Google Translate), semi-automatic mapping of a number of domain thesauri as well as SUMO ontology on plWN, bilingual word sense disambiguation (Sherlock Holmes corpus?), multilingual wordnet construction and contrastive studies (Open Multingual Wordnet, the Yiddish project[9],[10]).

[6] `http://plwordnet.pwr.wroc.pl/wordnet/stats`
[7] http://nlp.pwr.wroc.pl/plwordnet/license/
[8] www.cloud-net.pl
[9] http://polonjid-dictionary.clarin-pl.eu/
[10] https://polonjid.wn.uw.edu.pl/?lang=en

I-relation	V		N		Adv		Adj		Total	
I-relation	pl	en	pl	en	pl	en	pl	en	pl	en
I-synonymy	3933	3933	38056	38056	1002	1002	4156	4156	47147	47147
I-partial syn.	—	—	6602	6601	330	330	1373	1373	8305	8304
I-int.-reg. syn.	602	602	1983	1983	53	53	98	98	2736	2736
I-meronymy	—	—	10946	8109	—	—	—	—	10946	8109
I-hypernymy	264	11274	34651	83355	182	9910	383	44613	35480	149152
I-hyponymy	11277	264	83359	34656	9910	182	44613	383	149159	35485
I-holonymy	—	—	8106	10946	—	—	—	—	8106	10946
I-Type	—	—	7930	707	—	—	—	—	7930	707
I-Instance	—	—	707	7930	—	—	—	—	707	7930
I-allative	90	—	—	—	—	—	—	—	90	—
I-delimitive	461	—	—	—	—	—	—	—	461	—
I-excess	72	—	—	—	—	—	—	—	72	—
I-perdurative	24	—	—	—	—	—	—	—	24	—
I-anticausative	1717	—	—	—	—	—	—	—	1717	—
I-atenuative	233	—	—	—	—	—	—	—	233	—
I-cumulative	360	—	—	—	—	—	—	—	360	—
I-procesuality	16	—	—	—	—	—	—	—	16	—
I-completive	78	—	—	—	—	—	—	—	78	—
I-inchoative	215	—	—	—	—	—	—	—	215	—
I-distributive	840	—	—	—	—	—	—	—	840	—
I-iterative	82	—	—	—	—	—	—	—	82	—
I-terminative	12	—	—	—	—	—	—	—	12	—
I-ablative	42	—	—	—	—	—	—	—	42	—
I-causative	297	—	—	—	—	—	—	—	297	—
I-c-c-made-of	—	—	—	—	—	—	1059	—	1059	—
I-c-c-resembling	—	—	—	—	—	—	938	—	938	—
I-c-c-related-to	—	—	—	—	93	—	22694	—	22787	—
Total	20615	16074	192341	192343	11570	11477	75315	50623	299841	270571

Table 9: Interlingual relation counts

8 Conclusion and Further works

The created resource is unique not only because of its scale and method of construction, but mainly due to the fact that it uses a rich network of interlingual relations which had not been done before. Such approach has its pluses and minuses. It shows the complexity of a bilingual lexicon, yet it does not offer that many simple equivalents (often very much wanted by dictionary users). This is also partly due to the fact that wordnet mapping is synset mapping. However, we saw a significant potential for future development of the created bilingual resource.

Thus, we have started a project (Rudnicka et al., 2017) on converting the synset level mapping to an interlingual mapping between lexical units based on the concept of translational equivalence (Rudnicka et al., 2019). Three types of equivalence links were identified: strong, regular and weak, in addition to the lack of equivalence. The recognition of a type of equivalence was based on the manual verification of values of equivalence features, cf (Rudnicka et al., 2019). In a pilot study, equivalence links were manually described for ≈10 000 bilingual pairs of senses (lexical units) coming mostly from noun synsets linked by I-synonymy. On average, only 1-2 strong equivalence links were identified for a pair of synsets (Rudnicka and Naskręt, 2020). As a result, a precise bilingual sense-level dictionary that can be used in translation, but also in many bilingual wordnet application was developed. We plan to expand this mapping both to remaining noun pairs and to other parts of speech.

Acknowledgements

The work co-financed as part of the investment in the CLARIN-PL research infrastructure funded by the Polish Ministry of Science and Higher Education.

References

Luisa Bentivogli and Emanuele Pianta. Extending wordnet with syntagmatic information. In *Proceedings of the Second Global WordNet Conference*, pages 47–53, Brno, Czech Republic, January, 20th-23rd 2004.

Francis Bond and Ryan Foster. Linking and extending an open multilingual wordnet. In *Proc. 51st Annual Meeting of the Association for Computational Lin-*

guistics (Volume 1: Long Papers), pages 1352–1362, 2013.

Jordi Daudé, Lluís. Padró, and German Rigau. Mapping multilingual hierarchies using relaxation labeling. *ArXiv*, cs.CL/9906025, 1999.

Agnieszka Dziob, Maciej Piasecki, and Ewa K. Rudnicka. plwordnet 4.1 - a linguistically motivated, corpus-based bilingual resource. In *Proceedings of the Tenth Global Wordnet Conference : July 23-27, 2019, Wrocław (Poland)*, pages 353–362, 2019. URL `https://clarin-pl.eu/dspace/handle/11321/718`.

Christiane Fellbaum, editor. *WordNet – An Electronic Lexical Database*. The MIT Press, 1998.

Darja Fišer and Benoît Sagot. Constructing a poor man's wordnet in a resource-rich world. *Language Resources and Evaluation*, 49(3):601–635, 2015. ISSN 1574-0218. doi: 10.1007/s10579-015-9295-6. URL `http://dx.doi.org/10.1007/s10579-015-9295-6`.

Birgit Hamp and Helmut Feldweg. GermaNet – a Lexical-Semantic Net for German. In *Proc. ACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15. Madrid, 1997.

Paweł Kędzia, Maciej Piasecki, Ewa Rudnicka, and Konrad Przybycień. Automatic Prompt System in the Process of Mapping plWordNet on Princeton WordNet. *Cognitive Studies*, 2013.

Natalia V. Loukachevitch, German Lashevich, Anna. A. Gerasimova, Vladimir V. Ivanov, and Boris. Dobrov. Creating russian WordNet by conversion. In *Proceedings of Conference on Computational Linguistics and Intellectual Technologies Dialog-2016*, pages 405–415, 2016.

Marek Maziarz, Stanisław Szpakowicz, and Maciej Piasecki. Semantic relations among adjectives in polish wordnet 2.0: a new relation set, discussion and evaluation. *Cognitive Studies*, (12), 2012.

Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, and Stan Szpakowicz. Beyond the transfer-and-merge wordnet construction: plWordNet and a comparison with WordNet. In G. Angelova, K. Bontcheva, and R. Mitkov, editors, *Proc. International Conference Recent Advances in Natural Language Processing RANLP 2013*, pages 443–452. INCOMA Ltd. Shoumen, BULGARIA, 2013a.

Marek Maziarz, Maciej Piasecki, and Stanisław Szpakowicz. The chicken-and-egg problem in wordnet design: synonymy, synsets and constitutive relations. *Language Resources and Evaluation*, 47(3): 769–796, 2013b. doi: 10.1007/s10579-012-9209-9.

Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stan Szpakowicz, and Paweł Kędzia. plwordnet 3.0 – a comprehensive lexical-semantic resource. In Nicoletta Calzolari, Yuji Matsumoto, and Rashmi Prasad, editors, *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 2259–2268. ACL, ACL, 2016. URL `http://aclweb.org/anthology/C/C16/`.

George A. Miller. *A Semantic Network of English Verbs*, chapter 2, pages 47–68. In Fellbaum (1998), 1998.

Tomasz Naskręt, Agnieszka Dziob, Maciej Piasecki, Chakaveh Saedi, and António Branco. WordnetLoom–a Multilingual Wordnet Editing System Focused on Graph-based Presentation. In Francis Bond, Christiane Fellbaum, and Piek Vossen, editors, *Proceedings of the 9th Global Wordnet Conference, Singapore, 8-12 January 2018*. Global WordNet Association, 2018. URL `http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/GWC2018_paper_57.pdf`.

Bolette S. Pedersen, Sanni Nimb, Jørg Asmussen, Nicolai H. Sørensen, Lars Trap-Jensen, and Henrik Lorentzen. DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation*, 43: 269–299, 2009.

Maciej Piasecki, Stan Szpakowicz, and Bartosz Broda. *A Wordnet from the Ground Up*. Oficyna Wydawnicza Politechniki Wroclawskiej, Wroclaw, 2009. URL `http://www.dbc.wroc.pl/Content/4220/Piasecki_Wordnet.pdf`.

Maciej Piasecki, Michał Marcińczuk, Adam Musiał, Radosław Ramocki, and Marek Maziarz. WordnetLoom: a Graph-based Visual Wordnet Development Framework. In *Proceedings of the 2010 International Multiconference on Computer Science and Information Technology (IMCSIT)*, 2010. URL `http://ieeexplore.ieee.org/document/5679686/`.

Marten Postma, Emiel van Miltenburg, Roxane Segers, Anneleen Schoen, and Piek Vossen. Open Dutch WordNet. In *Proceedings of the Eight Global Wordnet Conference*, Bucharest, Romania, 2016.

Kergrit Robkop, Sareewan Thoongsup, T. Charoenporn, Virach Sornlertlamvanich, and H. Isahara. Wnms: Connecting the distributed wordnet in the case of asian wordnet. the 5th international conference of the global wordnet association (gwc-2010). 2009.

Ewa Rudnicka and Tomasz Naskręt. A dataset of translational equivalents built on the basis of plWordNet-Princeton WordNet synset mapping. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3260–3264, Marseille, France, May 2020. European Language Resources Association. URL `https://www.aclweb.org/anthology/2020.lrec-1.398`.

Ewa Rudnicka, Marek Maziarz, Maciej Piasecki, and Stan Szpakowicz. A Strategy of Mapping Polish WordNet onto Princeton WordNet. In *Proc. COLING 2012, posters*, pages 1039–1048, 2012.

Ewa Rudnicka, Wojciech Witkowski, and Michał Kaliński. A Semi-automatic Adjective Mapping Between plWordNet and Princeton WordNet. In *Text, Speech, and Dialogue*, pages 360–368. Springer, 2015. URL `https://link.springer.com/chapter/10.1007/978-3-319-24033-6_41`.

Ewa Rudnicka, Wojciech Witkowski, and Katarzyna Podlaska. Challenges of Adjective Mapping between plWordNet and Princeton WordNet. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation*

(LREC 2016). European Language Resources Association (ELRA), 2016. ISBN 978-2-9517408-9-1. URL http://www.lrec-conf.org/proceedings/lrec2016/pdf/481_Paper.pdf.

Ewa Rudnicka, Francis Bond, Łukasz Grabowski, Maciej Piasecki, and Tadeusz Piotrowski. Towards equivalence links between senses in plWordNet and princeton WordNet. *Lodz Papers in Pragmatics*, 13 (1):3–24, 2017.

Ewa K. Rudnicka, Maciej Piasecki, Francis Bond, Łukasz Grabowski, and Tadeusz Piotrowski. Sense equivalence in plwordnet to princeton wordnet mapping. *International Journal of Lexicography*, 32:296–325, 2019. doi: 10.1093/ijl/ecz004. URL https://academic.oup.com/ijl/article/32/3/296/5382106.

Vera Sheinman, Christiane Fellbaum, Isaac Julien, Peter F. Schulam, and Takenobu Tokunaga. Large, huge or gigantic? identifying and encoding intensity relations among adjectives in wordnet. *Language Resources and Evaluation*, 47:797–816, 2013.

Manish Sinha, Mahesh Reddy, and Pushpak Bhattacharyya. An approach towards construction and application of multilingual indo-wordnet. 3rd global wordnet conference (gwc 06). jeju island, korea. 2006. URL www.cse.iitb.ac.in/~pb/papers/gwc06_IITB_IndoWN.pdf.

Zeno Vendler. *Linguistics in Philosophy*. Ithaca: N.Y., Cornell University Press, 1967.

Kai von Fintel and Lisa Matthewson. Universals in semantics. *The Linguistic Review*, 25(1-2):139 – 201, 2008. doi: https://doi.org/10.1515/TLIR.2008.004. URL https://www.degruyter.com/view/journals/tlir/25/1-2/article-p139.xml.

Piek Vossen, editor. *EuroWordNet. A multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, 1998.

Piek Vossen. EuroWordNet General Document Version 3. Technical report, Univ. of Amsterdam, 2002.

Bartosz Wiland. Prefix stacking, syncretism and the syntactic hierarchy. In Marketa Zikova and Mojmir Docekal, editors, *Slavic Languages in Formal Grammar*, pages 307–324. Peter Lang, Bern, Switzerland, 2011.