A Large Wordnet-based Sentiment Lexicon for Polish

Monika Zaśko-Zielińska
University of Wrocław
Wrocław, Poland
monika.zasko-zielinska@uwr.edu.pl

Maciej Piasecki
Wrocław University of Technology
Wrocław, Poland
maciej.piasecki@pwr.edu.pl

Stan Szpakowicz
University of Ottawa
Ottawa, Ontario, Canada
szpak@eecs.uottawa.ca

Abstract

The applications of plWordNet, a very large wordnet for Polish, do not yet include work on sentiment and emotions. We present a pilot project to annotate plWordNet manually with sentiment polarity values and basic emotion values. We work with lexical units, plWordNet's basic building blocks.[1] So far, we have annotated about 30,000 nominal and adjectival LUs. The resulting lexicon is already one of the largest sentiment and emotion resources, in particular among those based on wordnets. We opted for manual annotation to ensure high accuracy, and to provide a reliable starting point for future semi-automated expansion. The paper lists the principal assumptions, outlines the annotation process, and introduces the resulting resource, plWordNet-emo. We discuss the selection of the material for the pilot study, show the distribution of annotations across the wordnet, and consider the statistics, including inter-annotator agreement and the resolution of disagreement.

1 Introduction

The Polish wordnet, plWordNet (Piasecki et al., 2009; Maziarz et al., 2013), is very large and comprehensive, with well over 150,000 synsets and 200,000 LUs at the time of writing. It has many applications, *e.g.,* text similarity (Siemiński, 2012), terminology extraction and clustering (Mykowiecka and Marciniak, 2012), extraction of opinion attributes from product descriptions (Wawer and Gołuchowski, 2012), addition of features for text mining (Maciołek and Dobrowolski, 2013), or a mapping between a lexicon and an ontology (Wróblewska et al., 2013). It is fast becoming a go-to resource in Polish lexical semantics. So far, however, it has not supported applications in the crucially important area of sentiment analysis and opinion mining. That area requires annotation: a word or word sense either does or does not carry sentiment, emotion or affect. That is why we have recently set out to annotate plWordNet with sentiment polarity and basic emotions.

Automatic annotation of lexical material is not a viable option. Wordnets are reference resources, relied upon for the absence of lexical errors. In fact, all widely published sentiment-marked and emotion-marked lists of lexical items have been created manually, sometimes by crowdsourcing. Now, plWordNet is much too large for complete, *affordable* manual annotation, but a reliable core of as little as 10% of the wordnet annotated makes it entirely possible to continue with semi-automatic expansion. Our pilot project manually annotated around 30,000 LUs (15% of plWordNet)[2] with sentiment and basic emotions, so we have ample material to also compare fully manual and semi-automatic annotation.

[1] The term *lexical unit* will be abbreviated to *LU* throughout this paper.

[2] This annotation is already on a scale several times larger than SentiWordNet (Esuli and Sebastiani, 2006).

Proceedings of Recent Advances in Natural Language Processing, pages 721–730,
Hissar, Bulgaria, Sep 7–9 2015.

2 Sentiment and Affect Annotations in Wordnets

Several sentiment lexicons are available for English, but hardly any for most other languages. Chen and Skiena (2014) have found 12 publicly available sentiment lexicons for 5 languages; there are none for Polish. Some sentiment lexicons have been built upon Princeton WordNet,[3] a natural starting point because of its comprehensive coverage and its numerous applications. The lexicons not based on PWN consider lemmas rather than lexical meanings or concepts.

WordNet-Affect is a selection of synsets very likely to represent "affective concepts" (Strapparava and Valitutti, 2004). A small core of 1903 lemmas was selected and described manually with "affective labels". Next, a set of rules based on wordnet relation semantics drove the transfer of the sentiment description onto the synsets connected to the core by wordnet relations. This produced 2874 synsets and 4787 lemmas.

SentiWordNet (Esuli and Sebastiani, 2006) annotates a synset with three values from the interval $\langle 0, 1 \rangle$. They describe "how objective, positive, and negative the terms contained in the synset are". About 10% of the adjectives were manually annotated, each by 3-5 annotators (Baccianella et al., 2010). In SentiWordNet 3.0, the automated annotation process starts with all the synsets which include 7 "paradigmatically positive" and 7 "paradigmatically negative" lemmas.[4] In the end, SentiWordNet 3.0 added automatic sentiment annotation to all of PWN 3.0.

SentiSense (Carrillo de Albornoz et al., 2012) is also a concept-based affective lexicon, with emotion categories assigned to PWN synsets. The initial list of 20 categories, a sum of several sets including WordNet-Affect, was reduced to 14 after some work with annotators. The authors write: "the manual labelling techniques generate resources with very low coverage but very high precision", but note that such precision can be only achieved for specific domains. The construction of SentiSense began with a manual annotation of only 1200 synsets with 14 emotions. Annotation was transferred onto other synsets using wordnet relations. The authors' visualisation and editing tools, designed to allow relatively easy expansion and adaptation, did not add much to the resource, so every user must enlarge it further to make it really applicable.

To sum up, a wordnet may be a good starting point for the construction of a sentiment lexicon: annotation can be done at the level of lexical meanings (concepts) or lemmas. PWN appears to be a good choice due to its sense-based model and large coverage. All large wordnet-based sentiment lexicons have been built by giving very limited manual annotation to algorithms for automated expansion onto other synsets. This, however, seems to have to result in lower precision, as noted, *e.g.,* by Poria et al. (2012): "Currently available lexical resources for opinion polarity and affect recognition such as SentiWordNet (Esuli and Sebastiani, 2006) or WordNet-Affect are known to be rather noisy and limited."

No large wordnets are available for most languages other than English. Many sentiment lexicons were created by translating sentiment-annotated PWN, *e.g.,* Bengali WordNet-Affect (Das and Bandyopadhyay, 2010), Japanese WordNet-Affect (Torii et al., 2011) and Chinese Emotion Lexicon (Xu et al., 2013). It is not clear how well annotations of that kind can be transferred across the language barrier. Moreover, as we discuss it in section 3.1, plWordNet's model differs slightly from that of PWN.

Crowdsourcing has also been used to develop sentiment lexicons (Mohammad and Turney, 2013). It *can* outdo automated annotation (or automatic expansion of a manually annotated part), but the consistency of the result is low compared to manual description by trained annotators.

Unlike most of the existing methods, our aim is a manual annotation of a substantial part of plWordNet by a team of linguists and psychologists. The manually annotated part – several times larger than other known manually created sentiment lexicons – can be an important resource on its own. It can also be a solid basis for the development of automated sentiment annotation methods for more lexical material in a wordnet. We have adopted a rich annotation model in which sentiment polarity description is combined with emotion categories.

[3] *Princeton WordNet* will be abbreviated to PWN throughout this paper.

[4] good, nice, excellent, positive, fortunate, correct, superior; bad, nasty, poor, negative, unfortunate, wrong, inferior (Turney and Littman, 2003)

3 An Annotation Model for plWordNet

3.1 The Principles

In contrast with most wordnet-building projects, plWordNet is not based on PWN. It also has a slightly, but significantly different model of word sense description. Its main building block is an LU understood as a pair: lemma plus sense number. LUs are grouped into a synset when they share constitutive lexico-semantic relations (hyponymy/hypernymy, meronymy/holonymy etc.) (Maziarz et al., 2013) Synsets are a notational shorthand for LUs which share their relations, so that all plWordNet relations recorded at the level of synsets can also be expressed at the level of LUs. More than half of relation instances in plWordNet are defined for LUs, because they are LU-specific, among them antonymy and relations signalled derivationally. Glosses and use examples in plWordNet are also assigned to LUs. LUs, then, seem to be a natural place to represent information related to sentiment polarity and emotions.

Sentiment polarity of an utterance is the result of a complex process influenced by word sense, language structure, communication and interpretation. It is difficult to describe sentiment polarity of a word sense in isolation from the context, but a "context-agnostic" sentiment lexicon can be a useful approximation for many applications. Too many factors govern sentiment perception from the point of view of the hearer (receiver) of the utterance. That is why we have assumed that the description from the point of view of the speaker would let us concentrate on the word sense typically intended by the speaker and its sentiment polarity included in that sense. We wanted to abstract away any further interpretation process and concentrate on the core of a word sense, which can be understood with no information about the context of interpretation.

Sentiment polarity appears to be associated with emotions which typify the source of the polarity in question. It can also be characterised by the fundamental human values associated with a given type of polarity (Puzynina, 1992) – more on that in section 3.3, step 2.

All in all, we have annotated LUs, plWordNet's basic building blocks, as completely as possible. We encode the sign of polarity (positive, negative, ambiguous), its intensity (strong, weak), as well as emotions and fundamental values.

3.2 The Pilot Project

The pilot sentiment annotation has been designed to add annotations to plWordNet manually. This is not what other wordnet annotation projects did – see section 2. Manual annotation on a larger scale does not only allow a broader vocabulary annotated with higher accuracy, often negotiated between annotators, but also becomes a much more reliable basis for *semi*-automatic expansion. We also wanted to test on a suitable scale the annotation guidelines we had adopted. Finally, we wanted to investigate how sentiment values and other related values are distributed over the various plWordNet relations and over synsets. It was not clear if LUs in a synsets must all have the same sentiment description. To avoid any bias, all that work was entrusted to a new group of linguists, separate from the main plWordNet team. A fresh look was also to be an independent diagnostic test for a sizeable part of the contents of plWordNet.

A manual analysis of the first sample of plWordNet LUs showed that even synsets with no positively or no negatively marked LUs can include LUs neutral in relation to sentiment, *e.g.,* {*mańkut* 1 '*coll.* left-hander' -weak, *leworęczny* 1 'left-handed' neutral, *szmaja* 1 '≈southpaw' -weak} or {*bliźni* 1 'neighbour [biblical]' +weak, *brat* 2 'brother' +weak, *drugi* 2 'the other' neutral}. Mixed-sentiment synsets rarely include positive, negative and ambiguous LUs, but they do occur, *e.g.,* {*pożądanie* 3 'desire' +strong, *pociąg fizyczny* 1 'physical attraction' +strong, *chuć* 1 '*coll.* sexual attraction, lust' -strong, *pożądliwość* 1 'lust' ambiguous}.

Notwithstanding, such synsets are well formed according to the general plWordNet guidelines. We also noted that LUs which share a derivational basis do not necessarily share their sentiment marking. There are marked bases with neutral derivatives, *e.g., gadać* 'to chatter' → *pogadanka* 'a chat', or *łazić* 'to tramp' → *łazik* 'a jeep' (Burkacka, 2003, p. 127). Derivational semantic relations, then, cannot be treated as copying the sentiment values to the derivatives.[5]

The sentiment of an LU x was determined in five main steps.

1. Decide if x is marked with respect to senti-

[5]Derivational semantic relations originate from a formal derivation relation and are mostly represented by a derivational link, but they are semantic in nature. They are suitably defined in the plWordNet model (Maziarz et al., 2013).

ment polarity, or neutral; if x is neutral, skip the remaining steps.

2. Assign the basic emotions and fundamental human values which appear to be associated with x.
3. Mark x as negative, positive or ambiguous.
4. Evaluate the intensity of x's sentiment polarity: strong or weak.
5. Give example sentences: one for x with a positive or negative polarity, two for an ambiguous x.

3.3 The Steps

Step 1 identifies noun LUs marked by non-neutral sentiment polarity. We have adopted two linguistic test procedures.

The first procedure is based on the method introduced by Markowski (1992) for the recognition of the lexis common to different genres, *i.e.,* nouns which are unmarked, non-erudite, and not terminological. A marked LU's expressivity can be implicit (*e.g.,* names of emotional states) or explicit (motivated by form or meaning) (Grabias, 1981, p. 40). The former are relatively easily spotted: they are established in language and occur in all genres (Zaśko-Zielińska and Piasecki, 2015), and their emotional markedness can be recognised without referring to context. The latter require the language user to check how she or other language users deploy it. For example, *troll* is either a Norse mythical creature or a person whose sole purpose in life is to seek out people to argue with on the internet over extremely trivial issues.[6]

For each LU analysed, we tested corpora for its occurrences together with deictic and possessive pronouns and operators which specify markedness.[7] Consider examples of the form *proszę pomyśleć o...* 'please think of...': *krześle* 'a chair' – acceptable; *tym krześle* 'this chair' – acceptable; *starociu* 'a relic' – unacceptable (this cannot be left unspecific); *tym naszym starociu* 'this relic of ours' – acceptable.

This method was applied earlier in research on Polish expressive lexis: expressivity is confirmed in context, and signalled (among others) by concretisation due to the use of pronouns (Rejter, 2006, pp. 88-90). For the recognition of marked LUs, we also used a concreteness test (Markowski, 1992): whether the LU can be modified by the pronouns *ten* 'this, the', *taki* 'such, such as', *twój* 'your$_{possessive}$' and *jakiś* 'some, a$_{referential}$, one'. The verdict was based on corpus search and the linguist's intuition.

We had to distinguish between neutral and marked adjectives. As in the analysis of nouns, we took into account such interrelated factors as meaning, word formation and context. Adjectives participate in the construction of expressive contexts in a sentence. Alongside such language mechanisms as the already noted deictic and possessive pronouns, adjectives are responsible for the semantic consistency of an utterance (Rejter, 2006, p. 76). That is why we placed a strong emphasis on the analysis of contexts in which adjectives occur.

The second test procedure in step 1 is based on checking the presence of pragmatic elements in the wordnet glosses for the analysed LUs and in their definitions in various dictionaries. We also tested the presence of qualifiers for genres—posp. (*pospolity* 'common'), pot. (*potoczny* 'colloquial'), wulg. (*wulgarny* 'vulgar') and książk. (*książkowy* 'bookish, literary')[8]—in the wordnet glosses of the analysed LUs.

The recognition of marked words is aimed not only at determining which LUs go through the subsequent steps of emotion analysis, but also at collecting neutral LUs (those not carrying polarity or emotion). Such LUs can play a role in automatic methods of emotional markedness recognition, see, *e.g.,* (Koppel and Schler, 2006).

Step 2 assigns emotions and values to LUs. We initially intended only to use the set of basic emotions which Plutchik (1980) identified in his Wheel of Emotions: joy, trust, fear, surprise, sadness, disgust, anger, anticipation. This set had figured in many later publications, *e.g.,* in (Ekman, 1992), and a number of resources and projects, including the NRC emotion lexicon (Mohammad and Turney, 2013) and the SentiSense Affective Lexicon (Carrillo de Albornoz et al., 2012).

In the Polish linguistic tradition, however, the description of the basic emotions is often associated with references to the fundamental values, like *użyteczność* 'utility', *dobro drugiego*

[6] http://www.urbandictionary.com/

[7] The corpora and other sources include:
http://tinyurl.com/kpwr1
http://www.nkjp.uni.lodz.pl/
http://www.nowewyrazy.uw.edu.pl/
http://www.miejski.pl/

[8] The term "książkowy" suggests *podniosły/uroczysty* 'solemn' as well as 'formal'.

człowieka ‘another’s good’, *prawda* ‘truth’, *wiedza* ‘knowledge’, *piękno* ‘beauty’, *szczęście* ‘happiness’ (all of them positive), *nieużyteczność* ‘’futility’, *krzywda* ‘harm’, *niewiedza* ‘ignorance’, *błąd* ‘error’, *brzydota* ‘ugliness’, *nieszczęście* ‘misfortune’ (all negative) (Puzynina, 1992). This set of fundamental values was proposed as a tool of linguistic analysis in the research on the language of values. We used it in our annotations. Kaproń-Charzyńska (2014, pp. 134-137) argues that expressions of emotions and values are usually associated in language expressions, and that it is difficult to separate them.

Evidence from psychological research, *e.g.,* (Barrett, 2006), and from linguistic research, *e.g.,* (Fries, 1992), shows that evaluation in terms of values is tightly connected with the feeling of emotions. Values can have different status in the description of lexical meaning: from included in the central aspects to peripheral. That is compatible with the semantics of prototypes, *e.g.,* (Mikołajczuk, 2000, p. 120).

To account for fundamental human values, then, the annotators could select labels from a predefined list, but they also could omit this sub-step.

The assignment of the emotion value helps annotators decide on the sentiment polarity of an LU. If the annotator selects, *e.g., wiedza* ‘knowledge’ and *piękno* ‘beauty’, *szczęście*, then we can assume that the given LU has a positive sentiment. If there are only negative emotions in the assigned set, *i.e.,* fear, surprise, sadness, anger, and disgust, and the values are only negative, then we can be sure that the LU has a negative sentiment. The presence of positive and negative emotions or values in the annotation of the given LU is a strong signal in favour of its ambiguity in relation to sentiment polarity.

We initially assumed that in some cases only emotions or values can be assigned to an LU. We observed, however, that only rarely did the annotator refrain from an assignment. Here is a likely reason: the annotators, while using combinations of basic emotions, tried to express complex emotions for which association with fundamental human values was much less straightforward. A mechanism for constructing complex emotions from basic ones (*e.g.,* disgust + anger = hostility) has been already described by Plutchik (2001, p. 349). That is why LUs marked by sentiment polarity and given some fundamental value had at least two or three basic emotions assigned.

The annotation with emotions and fundamental values was treated as supplementary to the primary annotation with sentiment polarity. We did not require perfect agreement in the assignment of basic emotions and fundamental values. High inter-annotator agreement was expected in the case of sentiment polarity, where the third annotator, the supervisor, arbitrated any disagreement. (See section 4 for more on team organisation.) The practice has shown, hovewer, that there is very high overlap between the sets assigned by two annotators. One set is mostly a subset of the other, which adds only one or two emotions or values. Consider *antytalent* 2 ‘a person who exhibits lack of skill in some area’:
A1: {*smutek* ‘sadness’, *wstręt* ‘disgust’}; {*nieużyteczność* ‘futility’, *niewiedza* ‘ignorance’}
A2: {*smutek* ‘sadness’, *złość* ‘anger’, *wstręt* ‘disgust’}; {*nieużyteczność* ‘futility’, *niewiedza* ‘ignorance’}

The evaluation of the sentiment polarity in **step 3** was based on several tests applied in parallel:

- a congruence test,
- a discord test,
- a test of collocations,
- a test of dictionary definitions.

The ongruence test requires all occurrences of the given LU x (not a lemma/word) in the usage examples to have the same sentiment polarity as that considered for x. The co-occurring adjectives, nouns and verbs do not change the polarity value, but support the polarity value considered for the given LU. For example:

- ***Przyjaźń*** *to lojalność, wierność i bezgraniczne oddanie.* ‘Friendship is loyalty, faithfulness and all-embracing devotion.’
  This supports the positive sentiment polarity for *przyjaźń* 1 ‘friendship’.
- *I że dolega mu jakiś* ***niepokój****, gorycz lub zgoła rozczarowanie.* ‘And that he feels some restlessness, bitterness or even disappointment.’
  This supports the negative polarity for *niepokój* 1 ‘restlessness’.

The congruence test can be also applied to LUs suspected of having ambiguous sentiment polarity. In such cases, we expect to find diverse usage

examples supported by the sentiment polarity of words co-occurring with the LU under analysis.

The discord test refers to plWordNet (or a wordnet in a more general setting). It checks the presence of the proper antonymy link between the LU considered and some other LUs with clear sentiment polarity. We assume that proper antonyms have opposite sentiment polarity values, *e.g.,* the relation *skłonność* 'inclination' – *niechęć* 'aversion' [negative] suggests the positive value for *skłonność* , and *nadzieja* 'hope' [positive] – *rozczarowanie* 'disappointment' suggests the negative value for *rozczarowanie*.

In the collocation test, words included in collocations for the given LUs are examined with respect to their sentiment polarity. In the ideal case, a positive LU is associated only with the positive words, and a negative one with the negative words. Such perfect association happens rarely, but the strength of the observed tendency supplies evidence for the annotator's decision about x.

Finally, annotators search through dictionary definitions for the given LU in order to check if all components of the definition (definition parts) are clearly positive, negative or mixed. Examples:

1. *szatan* 'devil' – *z* ***podziwem*** *o człowieku bardzo* ***zdolnym****, sprytnym,* ***odważnym*** '**admiringly** about someone very **capable**, canny, **courageous**' [plWordNet gloss]. This suggests positive polarity.
2. *bubek* 'a kind of ass and upstart' – *z niechęcią o mężczyźnie* ***mało wartym****, ale mającym* ***wygórowane*** *mniemanie o sobie* 'with dislike about a man **worth little** but with an **excessively high** opinion of himself'. This suggests negative polarity.
3. *zlewka* 3 '*coll.* ≈ funny situation' – *ubaw, dużo śmiechu, śmieszna sytuacja, ale bardziej w znaczeniu wyśmiewania się z kogoś* 'hilarity, much laughter, an amusing situation, but more in the sense of mocking somone'. This suggests both positive and negative polarity. Both annotators assigned contradictory annotations: +weak and - weak. The coordinator described the LU as ambiguous, with examples for either polarity.

We have developed several heuristics for **step 4** to evaluate the strength of polarity.

1. Given the basic emotions and fundamental values assigned to an LU, we can examine how close it is to them on some intensity scale, such as strong versus weak polarity. If, *e.g., smutek* 'saddness' and *złość* 'pique' are assigned to the LU *niezadowolenie* 'dissatisfaction', then we can consider whether they fully describe the state of dissatisfaction.
2. We can compare an LU with another, similar in meaning. If that LU is evidently more marked, the given one gets weak polarity.
3. If the given LU seems to have negative polarity but it is used to characterise a child humorously, we assign it weak polarity.[9]

It must also be noted that, for the common genre of Polish, the expressiveness and strength of markedness (including polarity) decreases in time. Very often, then, new marked words replace older less marked ones. For the native speakers today, old words do not have so clear a character and do not have the full strength of polarity, In the pilot project, we try to evaluate only the contemporary state and the contemporary polarity of LUs.

Examples added in **step 5** play a double role: they illustrate the annotations and the related aspects of the LU's meaning, and they verify the earlier decisions. Concerning the first role, it is especially important for the LUs considered ambiguous with respect to sentiment polarity.

The selection or creation of an example by the annotator is also the moment of the verification of the annotation decisions made so far. The example sentence should include frequent collocations of the LU under consideration. The sentence should show that the selected sentiment polarity does not result from the annotator's individual experience, but is also supported by the observed connectivity of the LU. So, all examples which the annotators create contain collocations found in corpora or other sources.

The language material stored in the examples is very interesting from the linguistic point of view. It often shows language use in unofficial situations. Examples also include also samples of transcribed speech. Such illustrations are not frequent in dictionaries. The corpus-based material needed careful selection and finding examples to match the given LU and its meaning, as well as illustrating the polarity value.

[9]For example, *ty draniu* 'you son of a gun' directed to a child is neither offensive nor angry. Related words *łobuziak, psotnik, urwipołeć* 'scamp, prankster, rascal' in the same usage serve to point out improper, but not harmful, behaviour.

4 The Annotation Process

The project team consisted of six annotators, coordinated by a "super-annotator". We had to find a balance between the available funds and the future practical value of the resource. We decided to aim at two annotations per LU. Everyone worked half of their time as the first annotator, *i.e.,* the one who assigns basic emotions, fundamental values, sentiment polarity values and examples.[10] The second annotator processed the same LU independently but, right after having recorded the result in plWordNet, could see what the first annotator did and then perhaps adjust the decision.

If the second annotator disagreed, a report went to the coordinator. Also, if the coordinator found an annotator's error, a re-analysis was requested. Practically the only cause was a wrong interpretation of the LU's meaning description in plWordNet.[11] Annotators occasionally discovered likely errors in plWordNet's structure. In such cases, the analysis was postponed until the main plWordNet team has intervened.

We selected several areas of plWordNet for the annotation project. In the first phase, we worked only with nouns, in the second phase – also with adjectives. Proper names were omitted in both phases. To start with nouns may be uncommon: the WordNet-Affect project, *e.g.,* started from adjectives (Strapparava and Valitutti, 2004). We had a good practical reason. The adjectival part of plWordNet was undergoing major expansion, but the annotation project had to go ahead, not to mention the fact that the main team could inadvertently undo annotation decisions.

There also was a serious reason. Annotation turned out to be simpler for nouns, so we gained experience before taking upon the more difficult area: adjectives. To assign sentiment polarity and other elements of the annotation is not harder. The main difference is in the proper interpretation of the description of an LU's meaning – in the linking of sentiment polarity evaluation with particular meanings of individual nominal and adjectival LUs. The work with use examples requires permanent word sense disambiguation – see (Mohammad and Turney, 2013). The adjectival meaning is often revealed in combination with nouns, so practice with nouns was very helpful for annotators.

We record in plWordNet fine-grained lexical meanings, linguistically well motivated. Nouns are described by the hypernymy hierarchy. Adjectives have a much shallower hierarchy and a lower density of relations (per one LU). So, there is more effort in understanding the meaning of an adjectival LU. Adjective lemmas are also on average more semantically ambiguous, *e.g.,* the average polysemy rate per lemma is higher for adjectives.[12] We started on adjectives when the adjective database reorganisation was already well advanced, so we effectively "played catch-up". An added advantage was the possibility of a close cooperation with the main plWordNet team.

In the case of nouns, we selected several domains, represented by hypernymy subgraphs, as more significant for sentiment polarity:

- the hypernymy sub-hierarchies for affect, feelings and emotions – the domain 'czuj' in plWordNet;
- noun sub-hierarchies describing people, *e.g.,* those dominated by non-lexical ("artificial") LUs *a person characterised by personality – age – physical properties – financial status – qualifications – positivity – negativity*;
- features of people and animals ('cech'),
- events ('zdarz'), *e.g.,* the sub-hierarchy of the artificial LU *events rated negatively, evaluated as negative* and the sub-hierarchy of *entertainment*.

5 The End Product: plWordNet-emo

Table 1 shows the number of LUs eventually annotated in the pilot project. The numbers refer to LUs which received the same sentiment polarity and strength from two annotators or whose sentiment label was decided by the coordinator. The project has annotated over 27% of adjectival LUs, but only around 12% of noun LUs from plWordNet 2.3. 12% is not high, but the processed portion covers the domains most likely to include LUs with non-neutral sentiment polarity. The manual annotations should be of high quality, and thus facilitate automated propagation of sentiment polarity to the remaining parts of plWordNet 2.3.

As noted in section 4, the second annotator did not look at the first annotator's decision before

[10] The pairing of annotators, and their first/second status changed regularly.

[11] Wordnets describe lexical meaning in terms of networks of relations. Not all LUs in plWordNet have glosses.

[12] `plwordnet.pwr.wroc.pl/wordnet/stats`

PoS	#	-s	-w	n	+w	+s	amb
N	19,625	11.29	8.78	69.06	3.24	2.88	4.74
Adj	11,573	9.89	11.22	58.85	9.21	5.60	5.24
Both	31,198	10.77	9.69	65.27	5.46	3.89	4.92

Table 1: Experimental sentiment annotation of plWordNet 2.3 in numbers; -s, -w, n, +w, +s, amb (negative strong/weak, neutral, positive weak/strong, ambiguous) are shown in percentage points.

PoS	#	-s	-w	n	+w	+s	amb
N	19,625	0.961	0.915	0.976	0.864	0.930	0.868
Adj	11,573	0.958	0.935	0.960	0.919	0.976	0.935

Table 2: Inter-annotator agreement, measured in Fleiss' κ, for different types of sentiment polarity: -s, -w, n, +w, +s, amb (negative strong/weak, neutral, positive weak/strong, ambiguous).

having made her own. Only in the case of evident errors did the coordinator ask the annotators to analyse the meaning of the given LU and to rethink the decision. We store all final decisions of the two annotators for every LU, so it is natural to measure inter-annotator agreement.

For nouns, the value of Fleiss' κ (Fleiss, 1971) – calculated for the two annotators and all decisions – is 0.943: very high agreement, even if we allow that the second annotator could sometimes change the decision after seeing the work of the first annotator. A very similar Fleiss' κ value of 0.95 was calculated for all annotators' decisions on adjectives. A detailed picture of inter-annotator agreement for all types of polarity appears in Table 2.[13]

A little surprisingly, the agreement for adjectives is higher than for nouns, and it is relatively equal across different types of polarity. A possible explanation: it is harder to read the meaning of adjectival LUs from plWordNet, and the annotators were more careful in reading the wordnet structures exactly.

6 Conclusions

The resource we have constructed is a first, important step towards sentiment annotation of the whole plWordNet. That is because the achieved size is very high in comparison to other manual annotation projects. We plan to expand the annotation to other LUs by means of algorithms based on sentiment polarity propagation along the wordnet graph.

The development of plWordNet has been independent of PWN, and the amount of sentiment annotation in our pilot project exceeds that in SentiWordNet and WordNet-Affect. It might therefore be interesting to compare our annotation with the automatic annotation in those wordnets, using the manual mapping of plWordNet onto PWN (Rudnicka et al., 2012).

Acknowledgments

Work financed by the Polish Ministry of Science and Higher Education, a program in support of scientific units involved in the development of a European research infrastructure for the humanities and social sciences in the scope of the consortia CLARIN ERIC and ESS-ERIC, 2015-2016.

Heartfelt thanks to our annotators: Ada Zajączkowska, Anna Nizińska, Kamil Wabnic, Monika Górka, Amelia Kiełbawska and Joanna Wieczorek.

[13] The κ values would have probably decreased a little if we calculated them for the second annotator's initial answer, before "reconciliation" with the first annotator's verdict. There are low-level technical reasons why we did not record that initial answer: the interface had been designed to streamline the annotators' task, and we decided to leave out clerical steps deemed *a priori* to be inessential.

References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7th Conference on Language Resources and Evaluation (LREC 2010), Valletta, MT,*, pages 2200–2204. ELRA.

Lisa Feldman Barrett. 2006. Valence is a basic building block of emotional life. *Journal of Research in Personality*, 40:35–55.

Iwona Burkacka. 2003. Aspekt stylistyczny opisu gniazdowego [The stylistic aspect of nests description]. In Mirosław Skarżyński, editor, *Słowotwórstwo gniazdowe. Historia, metoda, zastosowania [Nests derivation. History, method, applications]*, pages 114–136. Księgarnia Akademicka, Kraków.

Jorge Carrillo de Albornoz, Laura Plaza, and Pablo Gervás. 2012. SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*. `http://www.lrec-conf.org/proceedings/lrec2012/pdf/236_Paper.pdf`.

Yanqing Chen and Steven Skiena. 2014. Building sentiment lexicons for all major languages. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)*, pages 383–389, Baltimore, Maryland, USA, June 23-25 2014. ACL.

Amitava Das and Sivaji Bandyopadhyay. 2010. SentiWordNet for Indian Languages. In *Proceedings of the Eighth Workshop on Asian Language Resources*, pages 56–63.

Paul Ekman. 1992. An argument for basic emotions. *Cognition & Emotion*, 6(3):169–200.

Andrea Esuli and Fabrizio Sebastiani. 2006. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of 5th Conference on Language Resources and Evaluation LREC 2006*, pages 417–422.

Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 75(5):378–382.

Norbert Fries. 1992. Wartościowanie. Aspekty językowe i pojęciowe [Valuation. Linguistic and conceptual aspects]. In Gabriel Falkenberg, Norbert Fries, and Jadwiga Puzynina, editors, *Sprachliche Bewertung (polnisch und deutsch) [Valuation in language and text (on Polish and German material)]*, pages 25–44. University of Warsaw Press, Warszawa.

Stanisław Grabias. 1981. *O ekspresywności języka: ekspresja a słowotwórstwo [On the expressiveness of language: expression versus word formation]*. Wydawnictwo Lubelskie, Lublin.

Iwona Kaproń-Charzyńska. 2014. *Pragmatyczne aspekty słowotwórstwa. Funkcja ekspresywna i poetycka [The pragmatic aspects of word formation. The expressive and poetic function]*. Mikołaj Kopernik University Press, Toruń.

Moshe Koppel and Jonathan Schler. 2006. The Importance of Neutral Examples in Learning Sentiment. *Computational Intelligence*, 22(2):100–109.

Przemysław Maciołek and Grzegorz Dobrowolski. 2013. CLUO: Web-Scale Text Mining System for Open Source Intelligence Purposes. *Computer Science*, 14(1):45–62.

Andrzej Markowski. 1992. *Leksyka wspólna różnym odmianom polszczyzny, vol. 1-2 [Lexicon common to variations of the Polish language]*. Wydawnictwo "Wiedza o Kulturze", Wrocław.

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

Agnieszka Mikołajczuk. 2000. Problem ocen w analizie wybranych polskich nazw uczuć z klasy semantycznej GNIEWU [The problem of evaluation in the analysis of selected Polish emotion names from the class ANGER]. *Język a Kultura [Language and Culture]*, 14:117–134.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29(3):436–465.

Agnieszka Mykowiecka and Małgorzata Marciniak. 2012. Combining Wordnet and Morphosyntactic Information in Terminology Clustering. In *Proceedings of COLING 2012: Technical Papers*, pages 1951–1962.

Maciej Piasecki, Stanisław Szpakowicz, and Bartosz Broda. 2009. *A Wordnet from the Ground Up*. Wrocław University of Technology Press.

Robert Plutchik. 1980. *EMOTION: A Psychoevolutionary Synthesis*. Harper & Row.

Robert Plutchik. 2001. The Nature of Emotions. *American Scientist*, 89(4):344.

Soujanya Poria, Alexander Gelbukh, Erik Cambria, Peipei Yang, Amir Hussain, and Tariq Durrani. 2012. Merging SenticNet and WordNet-Affect emotion lists for sentiment analysis. In *IEEE 11th International Conference on Signal Processing (ICSP), 2012*, volume 2, pages 1251–1255, Beijing.

Jadwiga Puzynina. 1992. *Język wartości [The language of values]*. Scientific Publishers PWN.

Artur Rejter. 2006. *Leksyka ekspresywna w historii języka polskiego. Kulturowo-komunikacyjne konteksty potoczności [Expressive lexicon in the history of*

Polish language. Cultural and communicative contexts of colloquialism]. University of Silesia Press, Katowice.

Ewa Rudnicka, Marek Maziarz, Maciej Piasecki, and Stan Szpakowicz. 2012. A strategy of mapping Polish WordNet onto Princeton WordNet. In *Proceedings of the 24th International Conference on Computational Linguistics COLING*, pages 1039–1048.

Andrzej Siemiński. 2012. Fast algorithm for assessing semantic similarity of texts. *International Journal of Intelligent Information and Database Systems*, 6(5):495–512.

Carlo Strapparava and Alessandro Valitutti. 2004. WordNet-Affect: An affective extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1083–1086.

Yoshimitsu Torii, Dipankar Das, Sivaji Bandyopadhyay, and Manabu Okumura. 2011. A Developing Japanese WordNet Affect for Analyzing Emotions. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2011), 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011)*, pages 80–86.

Peter D. Turney and Michael L. Littman. 2003. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. *ACM Transactions on Information Systems*, 21(4):315–346.

Aleksander Wawer and Konrad Gołuchowski. 2012. Expanding Opinion Attribute Lexicons. In *Proceedings of Text, Speech and Dialogue, TSD 2012*, volume 7499 of *Lecture Notes in Computer Science*, pages 72–80. Springer.

Anna Wróblewska, Grzegorz Protaziuk, Robert Bembenik, and Teresa Podsiadły-Marczykowska. 2013. Associations between Texts and Ontology. In *Intelligent Tools for Building a Scientific Information Platform*, volume 467 of *Studies in Computational Intelligence*, pages 305–321. Springer.

Jun Xu, Ruifeng Xu, Yanzhen Zheng, Qin Lu, Kam-Fai Wong, and Xiaolong Wang. 2013. Chinese Emotion Lexicon Developing via Multi-lingual Lexical Resources Integration. In *Proceedings of 14th International Conference on Intelligent Text Processing and Computational Linguistics CICLing 2013*, pages 174–182.

Monika Zaśko-Zielińska and Maciej Piasecki. 2015. Lexical Means in Communicating Emotion in Suicide Notes – On the Basis of the Polish Corpus of Suicide Notes. *Cognitive Studies | Études Cognitives*, *to appear*.