

Towards Emotive Annotation in plWordNet 4.0

Monika Zaśko-Zielińska¹, Maciej Piasecki²

¹,University of Wrocław, Wrocław, Poland

²G4.19 Research Group, Department of Computational Intelligence

Wrocław University of Technology, Wrocław, Poland

maciej.piasecki@pwr.edu.pl, monik@uni.wroc.pl

Abstract

The paper presents an approach to building a very large emotive lexicon for Polish based on plWordNet. An expanded annotation model is discussed, in which lexical units (word senses) are annotated with basic emotions, fundamental human values and sentiment polarisation. The annotation process is performed manually in the 2+1 scheme by pairs of linguists and psychologists. Guidelines referring to the usage in corpora, substitution tests as well linguistic properties of lexical units (e.g. derivational associations) are discussed. Application of the model in a substantial extension of the emotive annotation of plWordNet is presented. The achieved high inter-annotator agreement shows that with relatively small workload a promising emotive resource can be created.

1 Introduction

Since plWordNet (Maziarz et al., 2016) achieved good coverage with the version 2.0 the number of its users and applications has been quickly growing. Many users declared sentiment analysis, as their intended use of plWordNet, contrary to the lack of sentiment information in it. In response, a pilot project on emotive annotation of a selected subset of senses in plWordNet was conducted (Zaśko-Zielińska et al., 2015) which next resulted with plWordNet 2.3 emo including *emotive annotation* for more than 31,000 lexical units (word senses). This prototype emotive annotation showed its usefulness in lexicon-based sentiment analysis, but its coverage was limited and selective (i.e. around 10% of noun senses and 25% of adjective senses of plWordNet 3.0 emo).

Our goal is to develop an improved and expanded model of emotive annotation for a word-

net, and also an expanded version of the manual annotation procedures. In addition we will also present application of the model in a substantial extension of the emotive annotation of plWordNet.

2 Emotions in Wordnets

Several sentiment lexicons are available for English, but hardly any for most other languages. Chen and Skiena (2014) found 12 publicly available sentiment lexicons for 5 languages; there were none for Polish. Some sentiment lexicons were built upon Princeton WordNet (PWN), 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.¹ In the end, SentiWordNet 3.0 added automatic sentiment annotation to all of PWN 3.0.

¹good, nice, excellent, positive, fortunate, correct, superior; bad, nasty, poor, negative, unfortunate, wrong, inferior (Turney and Littman, 2003)

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, 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 Annotation Model

For the sake of compatibility with plWordNet 2.3 emo, the main assumptions and annotation scheme have been preserved without significant changes, see Sec. 3.1,3.2. However, we plan to encompass by emotive annotation all PoS in plWordNet (i.e. nouns, adjectives, verbs and adverbs) and expand it very substantially by 100,000 annotated lexical units. It goes beyond typical sentiment polarity encoding and includes: sentiment polarity, basic emotions and fundamental values. On the basis of the analysis of the results of (Zaśko-Zielińska et al., 2015) we modified the annotation guidelines for nouns and adjectives to improve annotation quality, see Sec. 4.

3.1 Main Assumptions

plWordNet has been constructed on the basis of the corpus-based wordnet development method (Piasecki et al., 2009), according to which *lexical units* (henceforth LUs) are basic building blocks of the wordnet, *e.g.* use examples for LUs can be collected and analysed in corpora, but not for synsets, linguistic lexico-semantic relations are defined for LUs, and linguistic substitution tests can be applied to LUs. Synsets and their relations are derived in plWordNet from the structure of relations linking LUs, cf (Maziarz et al., 2013). Thus, emotive annotation is naturally done on level of LUs and includes LU use examples.

The analysis of the results of (Zaśko-Zielińska et al., 2015), i.e. the model, annotated LUs and the first feedback from the applications, has brought us to the revision of that model. However, first we agree with (Zaśko-Zielińska et al., 2015), that emotive annotation is focused on those emotive properties of LUs that are revealed in situation in which the given LU is maximally detached from the interpretation context, or, from the other point

of view, the description requires as little knowledge about the context as possible. This assumption coincides with the idea of dictionary and plWordNet is undoubtedly one.

As in (Zaśko-Zielińska et al., 2015), context-independent emotive characterisation of an LU is obtained by comparing its authentic uses found in the text corpora. During the annotation process features that are common to the LU usages are isolated, while the occasional ones discarded. Validating the obtained results we search for polarisation stability that should be repeated in the collocations of the given LU. However, contrary to (Zaśko-Zielińska et al., 2015), we claim that LU's emotive polarisation determined in this way does not provide information about emotive attitudes of the speaker. We can only read what is expressed in the examples. This is a difference similar to the one between the intent and the statement function, cf (Bartmiński and Niebrzegowska-Bartmińska, 2009). Thus, while still preserving the fundamental premise of aiming at the detection of the LU characteristics outside of the context, we assume that it is not the knowledge of the speaker's emotive attitude that is being described in annotation, but the emotive characteristics that is common to the analysed expressions and salient to the recipient, i.e. an annotator. The process of averaging across different LU use examples in search for emotive features independent of the contexts, or common to different contexts, is amplified by searching for agreement of the annotators applying independently the same annotation procedure.

In (Zaśko-Zielińska et al., 2015) fundamental human values (Puzynina, 1992), see Sec. 3.2, have been also included into the emotive annotation. This is a unique solution in comparison to other wordnet-based emotive annotations. There are also important reasons to follow and expand it in our work. While emotional assessment is always associated with a kind of evaluation in the meaning of LUs, it is very often neglected that some LUs lack emotional aspect, but still are associated with a form of evaluation. Annotating of LUs with fundamental human values expressing evaluation is particularly important for the analysis of product reviews or brands (opinion mining) It helps to extend sentiment polarisation also to multiword LUs that are quite numerous in plWordNet (>54k) and many of them belong to terminology. This is especially valuable because general dictionar-

ies usually omit this type of LUs. They are often treated as a group of vocabulary without polarisation. However, it is worth to notice that in works on opinion mining in Polish texts from the economics point of view, speaker's attitude is an important factor in the analysis of product reviews This is partially possible, but does not take into account the impact of the speaker's error on the quality of the message or the beliefs of the recipient, which, as contextual information, is inherent in receiving the message (Lula et al., 2016).

3.2 Annotation Scheme

Following (Zaśko-Zielińska et al., 2015) the main distinction is between *neutrality vs polarity* of LUs. Polarised LUs are assigned the *intensity* of the sentiment polarisation, *basic emotions* and *fundamental human values*. The latter two help to determine the sentiment polarity and its intensity expressed in the 5 grade scale: *strong* or *weak vs negative* and *positive*. Annotator decisions are supported by use examples that must be included in the annotations.

Due to the compatibility (Zaśko-Zielińska et al., 2015) with other wordnet-based annotations, we use the set of eight basic emotions recognised by Plutchik (Plutchik, 1980) whose wheel shows how basic emotions can contribute to secondary emotions. It contains Ekman's six basic emotions (Ekman, 1992): *joy*, *fear*, *surprise*, *sadness*, *disgust*, *anger*, complemented by Plutchik's *trust* and *anticipation*. As a result, negative emotions do not prevail in the set, cf (Mohammad and Turney, 2013). One LU can be assigned more than one emotion and, as a result, complex emotions can be represented by using the same eight-element set. Plutchik states (Plutchik, 1980) that his basic emotions are primary, that is, they appear first in ontogenesis and phylogenesis. So we assume that they are repetitive for all language users regardless of their age and development. Ekman, on the other hand, refers not to evolution but to intercultural nature and claims that facial expressions and underlying emotions are common to different cultures (Ekman and Friesen, 1971).

As in (Zaśko-Zielińska et al., 2015), we use the set of fundamental human values postulated by Puzynina (Puzynina, 1992), later followed in many works on lexicology and derivation. Thus we assume that the emotive state of the speaker is linked to the *evaluative attitude*.

Although, the evaluation can also be independent of emotions (Waszakowa, 1991). The set of the fundamental human values encompasses: *użyteczność* ‘utility’, *dobro drugiego 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).

3.3 Examples of Annotation

Below we present examples of complete emotive annotations for three LUs (where A1 and A2 means, respectively the first and the second annotation added, **BE** – basic emotions, **FHV** – fundamental human values, **SP** – sentiment polarity, and **Exam** – usage example):

dziad 1 gloss: “stary mężczyzna” ‘an old man’
 (**Annot.:**A1, **BE:** {*złość* ‘anger’, *wstręt* ‘disgust’}, **FHV:**{*nieużyteczność* ‘futility’, *niewiedza* ‘ignorance’}, **SP:**–s

Exam: “Stary dziad nie powinien podrywać młodych dziewczyn.”

‘An old geezer should not pick up young girls.’
 (**Annot.:**A2, **BE:** {*wstręt* ‘disgust’}, **FHV:**{*nieużyteczność* ‘futility’, *brzydota* ‘ugliness’}, **SP:**–w

Exam: “Jakiś dziad się dosiadł do naszego przedziału i wyciągnął śmierdzące kanapki z jajkiem.” ‘An old geezer joined our compartment and took out stinky egg sandwiches.’

(**Annot.:**A3, **BE:** {*wstręt* ‘disgust’}, **FHV:**{*nieużyteczność* ‘futility’, *brzydota* ‘ugliness’}, **SP:**–s

Exam: “Kilkanaście lat minęło i zrobił się z niego stary dziad.”

‘Several years have passed and he has become an old geezer’)

szalbierski 2 ‘deceitful’

(**Annot.:**A1, **BE:** {*smutek* ‘sadness’, *złość* ‘anger’}, **FHV:** {*krzywda* ‘harm’, *błąd* ‘error’}, **SP:**–s,

Exam: “Nie chciałam brać udziału w tym szalbierskim planie, którego pomyślność zależała od stopnia naiwności nieświadomych klientów.”

‘I did not want to take part in this deceitful plan, whose success depended on the level of naiveness of the unaware clients.’)

(**A2**, **BE:** {*smutek* ‘sadness’, *złość* ‘anger’}, **FHV:** {*krzywda* ‘harm’, *błąd* ‘error’}, **SP:**–s,

Exam: “Mam szalbierski pomysł, który pomoże nam naciągnąć paru idiotów.”

‘I have a deceitful idea which might help us to con a couple of idiots.’)

wytrzymały 2 ‘enduring’

(**Annot.:**A1, **BE:**{*zaufanie* ‘trust’}, **FHV:**{*użyteczność* ‘utility’}, **SP:**+w,

Exam: “Wykonaliśmy podłogę z wytrzymałych paneli, dzięki temu od lat prezentuje się wspaniale.”

‘We made the floor from enduring panels, that is why it has been looking splendid for years’)

(**Annot.:**A2: **BE:**{*zaufanie* ‘trust’}, **FHV:**{*użyteczność* ‘utility’}, **SP:**+w

Exam: “Postanowiłem nie oszczędzać i kupić plecak z wytrzymałego materiału — przynajmniej wiem, że nie rozleci mi się po roku.”

‘I decided to not economize and to buy a backpack made of enduring material — at least I know that it will not tear apart after one year.’)

4 Annotation Procedure

The annotation is performed² by: linguists and psychologists, where each LUs is annotated by a mixed pair: one psychologist and one linguist. The annotators must follow guidelines that consist of a core common to all PoSs and detailed guidelines dedicated to each PoS. The work of annotators is coordinated and verified by a supervisor, who can access all annotations and solve disagreements³ by adding the final annotation.

The common core is similar to the procedure in (Zaśko-Zielińska et al., 2015):

Step 1 identification of LUs with *neutral* and *non-neutral* sentiment polarity;

Step 2 assignment of the basic emotions and fundamental human values;

Step 3 recognition of the LU polarity direction: negative or positive, but also *ambiguous*, if the collected use examples show both behaviours;

Step 4 assignment of sentiment polarity intensity;

Step 5 illustration of the assigned annotation by sentences representing use examples: at least

² Six persons were working on the results reported here: four linguists and two psychologists.

³ As it is presented in Sec. 5 disagreements in sentiment polarity are quite infrequent.

one sentence in the case of positive and negative LUs, and at least two example sentences for ambiguous LUs.

Each step is associated with several linguistic tests, including substitution tests and requires consulting corpus data. The detailed specification of the subsequent steps is dependant on a particular PoS. In the case of nouns see (Zaśko-Zielińska et al., 2015), the specification for adjectives proposed by us is presented in Sec. 4.2.

Annotators can return from the later steps to the previous ones. We could observe that, e.g., assignment of fundamental human values or basic emotions can be helpful in verifying the polarity of the given LU.

For the annotation process, we use Wordnet-Loom – a wordnet editing system (Piasecki et al., 2013) – which has been extended by additional windows and database tables (to eliminate errors in the annotation representation), as well as a mechanism that separates work of individual annotators. They do not see annotation decisions of other annotators and they do not know who is the second annotator of the given LU. Moreover, annotators are rotated in the pairs in order to minimise a potential bias. This strict separation of annotators is a significant difference in relation to (Zaśko-Zielińska et al., 2015), where the second annotator was told not to take a look into the decision of the first annotator before having made his own one, but he could see it and could change his own one later. The second could report a possible error of the first one in the pilot project, but we decided to resign from this possibility and to separate them strictly. The inter-annotator agreement is on a high level, but inevitably lower than reported in (Zaśko-Zielińska et al., 2015), see Sec.5. However, we sometimes observed a tendency to too prompt classification of a LU as a neutral one. If such a decision is taken without a detailed analysis, then the annotation process is actually discontinued after the first step and any change of mind of the given annotator later along the process is impossible. To amend this potential problem we paid more attention to the detailed guidelines for Step 1, as well as to the training of annotators and verification of their work.

4.1 Nouns

As annotation of nouns was not completed in the pilot project, we also started with nouns. We used

guidelines from the pilot project. Only minor details were fine-tuned, e.g. we added a test for distinguishing diminutive formant function (Siudzinska, 2016). Formants appropriate for diminutives are not always connected with sentiment polarity. The test involves attaching three groups of adjuncts to the nouns:

- A adjuncts indicating size (e.g. expressing senses: *small, fine, young, ...*);
- B adjuncts showing positive emotions towards the person represented by the derivative or emotional bond with a person (e.g., senses: *my, our, good, loved, nice, sympathetic, unusual, modest, poor, tiny, thin, mischievous, miserable, etc.*);
- C adjuncts indicating negative emotions (e.g., *clumsy, unfulfilled, stupid, backward, lying, poor*); in this way, the sender may indicate the immaturity, helplessness of the person called by the derivative, and also show pity, irony, disregard and contempt.

Test A covers LUs like: *minutka* ‘≈a small minute’, *chwilunia* ‘≈a tiny moment’, that are related to size.

4.2 Adjectives

Annotation of adjectives started at the end of the pilot project on limited material, so the guidelines for adjectives required more substantial changes.

First annotators are reminded that adjective LUs in plWordNet have mostly more fine grained meanings than those in Polish dictionaries. Thus, all the time the annotator has to check whether he is working on the same and appropriate LU, not, e.g. deviating accidentally to another sense of the LU lemma. For this purpose annotators should check and use collocations as a tool for prompting a particular meaning. For instance *ciężki* ‘≈heavy’ corresponds to 23 LUs, that can be distinguished (mentally or in the corpus) by different collocations, e.g.: *heavy 1* – ‘weighs a lot’ (heavy bag), *heavy 2* – ‘sluggish, slow’ (heavy steps); *heavy 8* – ‘bulky, overwhelming’ (heavy curtains), *heavy 9* – ‘thick, not transparent’ (heavy air), *heavy 12* – ‘sad’ (heavy film), *heavy 14* – ‘difficult to bear’ (heavy silence), *heavy 15* – ‘heavy with fatigue’ (heavy eyelids), *heavy 18* – ‘intense, expressive’ (heavy wine), *heavy 19* – ‘ponderous’ (feels heavy); *heavy 22* – ‘with great power

(heavy artillery), or *heavy* 23 – ‘strong, aggressive’ (heavy sound).

Step 1 Neutrality test for adjectives is related to the wordnet structure of derivational relations for adjectives, non-derived adjectives are analysed according to the noun procedure. Adjectives derived from adjectives can be skipped in **Step 1**. The rest of derived adjectives are recognised as non-neutral:

- adjectives from polarised nouns: *domowa atmosfera* ‘home atmosphere’, derived from *dom* ‘from home (as a group of people)’ in opposition to the neutral *domowy strój* ‘a casual outfit’ where *domowy* is derived from ‘home (place), ≈ ‘somebody’s flat’;
- adjectives derived from verbs, called *dispositional*, including subtypes: *potential* – expressing potentia to do something, e.g. *powtarzalny* ‘repeatable’, *habitual* emphasising that something is permanent and in large amounts, e.g. *krzykliwy* ‘≈ noisy, vociferous’ in *krzykliwe dziecko* ‘a noisy child’, *quantificational* signalling large amount or quantity, e.g. *wytrzymały* ‘hardy, inured, hardened’ in *wytrzymały człowiek* ‘a hardened man’, and positively *evaluating*, e.g. *bitny* ‘valiant’ in *bitny żołnierz* ‘a valiant soldier’.

Step 2 Assignment of emotions and values: adjectives derived from verbs by the suffix *-alny* (meaning ‘to be able to’, ‘it is possible to’) form a very characteristic group of LUs. They are not connected with emotions, but they are related to the fundamental values: utility, futility, e.g. *zmywalny* ‘such that, can be removed by washing’ in *tataż zmywalny* ‘a tattoo that can be washed out’, *egzekwowny* ‘such that can be enforced’.

Step 3 Marking LUs as negative, positive or ambiguous: this step requires especially careful identification of meanings. In order to recognise polarity we perform tests: a *congruence test*, a *discord test*, a *test of collocation* and a *test of dictionary definitions*. The way they are formulated and applied is similar to the corresponding test for nouns, see (Zaśko-Zielińska et al., 2015). However, more attention should be sometimes paid to affixes, whose semantic transparency in adjective derivatives seems to be weaker.

The congruence test not only allows to detect the LU polarity, but also helps in creating exam-

ple sentences in **Step 5** that confirm the polarity recognised earlier, e.g. for *tęskny* ‘wistful’:

positive: *Upajaliśmy się tym tęsknym, nastrojowym widokiem.*

‘We were intoxicated by this **wistful** and romantic view.’

negative: *Nie mogłam już dłużej wytrzymać tego zawodzenia i jego tęsknych pieśni.*

‘I could not bear this crooning and his **wistful** songs any longer.’

The presence of the same LU in the two opposing contexts reveals its ambiguous emotive character. The occurrence of suffixes: *-usieńki*, *-uteńki*, *-eńki* does not determine the polarisation of LUs, because it also depends on the derivation basis. Although these suffixes express a positive polarisation (Grzegorzczkowska et al., 1998), the combination with the derivation basis, which can be polarised negatively, only weakens the marking, for example: *chudzieńki* ‘≈ very thin and weak’, *pijaniusieńki* ‘≈ completely drunk, not controlling himself’.

The discord test is used to correct linguistic awareness, which is primarily focused on negative polarisation: only antonyms, e.g. *clean – dirty* shows that both elements are polarised in this pair. Often only the collocation test allows you to capture the ambiguity of the polarisation for example: for *pedantic a pedantic order vs morbidly pedantic*.

Step 4 Assignment intensity of sentiment polarity: annotators are reminded that grade forms of adjectives do not inform about the sentiment polarity intensity of the derivational basis, but they show comparison between objects or phenomena; e.g., the suffix derivative *-utki* which expresses that the described feature is not at its maximum, in the lower part of a scale, and there may be something that is even smaller than *malutki* ‘≈very small’. In comparison to it, LUs with *-uteńki* ‘≈tiny’, *-usieńki* ‘≈very tiny’ may be a cause of doubt, as their suffixes signals that some feature value is even smaller. In resolving this problem one has to remember two aspects of such a derivation process: semantic and pragmatic. Although LUs *mokrzuteńki* ‘≈completely wet’, *mokrzusieńki* ‘≈completely wet’ can be interpreted as representing some extreme values of the feature, this is rather semantic information, and the emotive aspect of this LUs is be maximised. (Bogusławski, 1991) argues that the func-

PoS	# Comp	# Sing	-s	-w	n	+w	+s	amb
N	25,919	18,574	16.62	14.64	51.59	6.05	4.23	6.87
Adj	14,817	5,392	14.87	22.59	31.39	15.03	7.50	8.62
All	40,773	24,002	15.89	17.95	43.18	9.79	5.59	7.60

Table 1: Sentiment polarity annotation of plWordNet 4.0 in progress (Comp – completed, Sing – one annotator only so far); -s, -w, n, +w, +s, amb (negative strong/weak, neutral, positive weak/strong, ambiguous) are shown in percentage points.

tion of these affixes is similar to inflection, i.e., it corresponds to grade of adjectives.

5 Intermediate Results

During the pilot project more than 31,000 LUs (19,625 noun LUs and 11,573 adjective LUs) were described in plWordNet 3.0 emo by emotive annotation (Zaśko-Zielińska et al., 2015). From that point we started the annotation process aiming at its expansion by complete emotive annotations (2+1) for around 100k more LUs. Annotations done in the pilot project including decisions of only one annotator had to be completed.

We started adding emotive annotation from noun LUs with focus on hypernymic branches that are likely to include LUs with polarised sentiment. In addition we try to distribute manual annotations across the network of synsets in such a way that it will be possible to apply an algorithm for automated spreading annotations to the rest of LUs.

The statistics describing the current state of the work are presented in Tab. 1. Only LUs annotated by two annotators are counted as completed. This number includes also completed annotations for LUs processed during the pilot project. As annotators are mixed in pairs and subsets of LUs are assigned to them in diversified ways, a large number of LUs have received so far only one annotation. As it was also the case in the pilot project, more than half of the noun LUs are annotated as neutral. However, only $\approx 30\%$ of adjective LUs are neutral contrary to almost 60% in plWordNet 3.0 emo. This difference can be caused by a much broader coverage of noun LUs, while adjective LUs were selected by (Zaśko-Zielińska et al., 2015) in a slightly accidental way (there was an ongoing plWordNet expansion work on that time).

As our annotators work completely independently, we could measure the inter-annotator agreement (IAA) with respect to the sentiment polarity using the Cohen’s Kappa measure (Cohen, 1960), see Tab. 2. Due to the large number of annotators, and simplifying a little bit, we present the

PoS	All	-s	-w	n	+w	+s	amb
All	0.78	0.77	0.78	0.82	0.74	0.73	0.65
Mrk.	0.84	0.80	0.84	–	0.89	0.80	0.86

Table 2: Inter-annotator agreement (IAA), measured in Cohen’s κ , for different types of sentiment polarity: -s, -w, n, +w, +s, amb (negative strong/weak, neutral, positive weak/strong, ambiguous). *All* describes agreement for all decisions, *Mrk* presents estimated IAA value for marked LUs only.

agreement between the first and the second decision registered in the system for LUs. LUs with at least one annotation from the pilot project were excluded from this analysis. The observed IAA values, both, 0.78 for all decisions and around 0.75 for different sentiment polarity values, are very high. The value for the neutral polarity is a value for the decision: polarised vs non-polarised in fact. It can show that the annotators are quite confident about the neutrality of the LUs, but also it can be biased by the fact that describing a LU as a neutral can be easier than by other values. This issue needs further investigation.

As the neutral annotations dominate (almost half of all decisions), we have calculated an estimated IAA value for the marked LUs only by simply taking into account LUs for which any annotator did not proposed the neutral value. The obtained values are much higher than for all decisions, so we can conclude that neutral values do not increase artificially the general IAA.

Negative sentiment polarity values dominate in annotation: 33.84% vs 15.38% in Tab. 2. This correlates with the dominance of the negative basic emotions that can be observed in the statistics presented in Tab. 3, i.e. 76.48% emotions associated with noun LUs and 70.13% with adjective LUs are negative. A similar dominance of words marked negatively could be also observed in the dictionary of the colloquial Polish language (Anusiewicz and Skawiński, 1996). For instance, if we compare two thematic fields of this dictionary, namely: act-

PoS	joy	trust	antic.	surprise	fear	disgust	sadness	anger				
N	15.17	6.74	0.96	0.65	7.66	21.78	16.77	30.27	–	–	–	–
Adj	20.95	8.01	0.54	0.37	5.31	18.56	21.56	24.71	–	–	–	–
	util.	good	truth	know.	beauty	happ.	futility	harm	ignor.	error	uglin.	misfor.
N	18.89	3.06	0.76	4.76	2.17	14.98	13.93	12.69	3.07	13.40	2.71	9.58
Adj	23.88	3.62	1.01	2.53	4.03	14.37	15.29	8.85	1.18	14.30	3.56	7.40

Table 3: Basic emotions (see Sec. 3.2) and fundamental human values (see Sec. 3.2) annotation of plWordNet 4.0 (in progress) are shown in percentage points.

ing towards somebody’s harm – enforcing some particular behaviours (id:2.3.2) and acting towards somebody’s profit (id.: 2.3.3), we can notice that the former includes 324 entries while the latter only 20. In plWordNet emo it is also characteristic that almost all emotions except *fear* are approximately frequent while *joy* is a single dominating positive emotion. This bias can be a result of assigning *joy* not only as a simple emotions, but also as a basic component of the complex emotions.

Contrary to the basic emotions, the fundamental human values are evenly distributed between the positive and negative ones, see Tab. 2: 55.38% negative values assigned to nouns and 50.57% to adjectives. There are no single fundamental human values that are substantially more frequent across the annotations, but only some of them, e.g. *prawda* ‘truth’ are significantly less frequent. Language users mostly perform evaluations of an emotional or utility (advantageous vs non-advantageous) character. They relatively infrequently assess phenomena from the rational perspective. Emotively marked LUs are more frequent in colloquial or informal communication where emotions and advantages are more important than rational thinking.

We checked also combinations of sentiment polarity values inside synsets. Almost all synsets are consistent with respect to the sentiment polarity, i.e. only ≈ 20 synsets from many thousands analysed included LUs of both positive and negative polarity, and most of them result from errors in plWordNet, e.g. too broad synsets. Synsets including marked LUs and neutral or ambiguous ones are more frequent, but perfectly compatible with the annotation guidelines. LU linked by hypernymy (via synset hypernymy) are in the vast majority of cases in the same polarity. We found only less than 700 hundred LUs linked by hypernymy per more than 70,000 analysed pairs in which both LUs were in the different polarity, among which we found only 32 $\langle -s, +s \rangle$ pairs.

6 Conclusions and Further Works

A large emotive lexicon can be an indispensable language resources for sentiment analysis and opinion mining, if it is of good coverage and quality, especially if the lexicon-based method is expanded with domain adaptation on the basis of machine learning. At least the use of the lexicon can help to improve the domain independent aspect of the method. The pilot project (Zaško-Zielińska et al., 2015) showed that with relatively small workload a promising emotive resource was be created. We presented an annotation process following this project and aiming at building a very large emotive lexicon of Polish of more than 130k manually annotated lexical units from plWordNet, i.e. on a scale incomparable to the majority of existing resources. The intended size is meant to suit the envisaged applications. A slightly modified general model and annotation guidelines were presented, together with improved specific guidelines for adjectives. Both the lexicon as well guidelines utilise the rich relation structure of plWordNet. The observed high values of the inter-annotator agreement (measured on a large sample of data according to an objective procedure) is very promising for the future applications and is a strong argument in favour of the assumed model and annotation procedure. The presented first results for nouns and adjectives, but for quite large sample, allows for collecting interesting observation that are in line with qualitative analysis in literature. We plan to complete annotation (>130k lexical units in total) of all Parts of Speech in plWordNet by the July 2018. The results will be completely open. The annotation will be extended to the rest of plWordNet by automated method (e.g. based on activation propagation or machine learning.) We plan also to compare our annotation with annotation built for English using the mapping of plWordNet onto Princeton WordNet.

Acknowledgment

Work supported by the Polish Ministry of Education and Science, Project CLARIN-PL.

References

- [Anusiewicz and Skawiński1996] Janusz Anusiewicz and Jacek Skawiński. 1996. *Słownik polszczyzny potocznej*. Wrocław.
- [Baccianella et al.2010] 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.
- [Bartmiński and Niebrzegowska-Bartmińska2009] Jerzy Bartmiński and Stanisława Niebrzegowska-Bartmińska. 2009. *Tekstologia*. Wydawnictwo PWN, Warszawa.
- [Bogusławski1991] Andrzej Bogusławski. 1991. Polski sufiks –utki,. *Poradnik Językowy*, 5–6:174–179.
- [Carrillo de Albornoz et al.2012] 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.
- [Chen and Skiena2014] 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.
- [Cohen1960] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- [Das and Bandyopadhyay2010] Amitava Das and Sivaji Bandyopadhyay. 2010. SentiWordNet for Indian Languages. In *Proceedings of the Eighth Workshop on Asian Language Resources*, pages 56–63.
- [Ekman and Friesen1971] P. Ekman and WV. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124–129, Feb.
- [Ekman1992] Paul Ekman. 1992. An argument for basic emotions. *Cognition & Emotion*, 6(3):169–200.
- [Esuli and Sebastiani2006] 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.
- [Grzegorzczkova et al.1998] R. Grzegorzczkova, R. Laskowski, and H. Wróbel, editors. 1998. *Morfologia. Gramatyka współczesnego języka polskiego*, volume T. 2. PWN.
- [Lula et al.2016] Paweł Lula, Katarzyna Wójcik, and Janusz Tuchowski. 2016. Analiza wydźwięku polskojęzycznych opinii konsumenckich ukierunkowanych na cechy produktu [feature-based sentiment analysis of opinions in polish]., *Prace Naukowe Uniwersytetu Ekonomicznego We Wrocławiu [Research Papers of Wrocław University of Economics]*, 207:p.155.
- [Maziarz et al.2013] 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.
- [Maziarz et al.2016] Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stan Szpakowicz, and Paweł Kędzia. 2016. 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.
- [Mohammad and Turney2013] Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29(3):436–465.
- [Piasecki et al.2009] Maciej Piasecki, Stanisław Szpakowicz, and Bartosz Broda. 2009. *A Wordnet from the Ground Up*. Wrocław University of Technology Press.
- [Piasecki et al.2013] Maciej Piasecki, Michał Marcińczuk, Radosław Ramocki, and Marek Maziarz. 2013. WordnetLoom: a wordnet development system integrating form-based and graph-based perspectives. *International Journal of Data Mining, Modelling and Management*, 5(3):210–232.
- [Plutchik1980] Robert Plutchik. 1980. *EMOTION: A Psychoevolutionary Synthesis*. Harper & Row.
- [Poria et al.2012] S. Poria, A. Gelbukh, E. Cambria, PeiPei Yang, A. Hussain, and T. 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.
- [Puzynina1992] Jadwiga Puzynina. 1992. *Język wartości [The language of values]*. Scientific Publishers PWN.
- [Siudzińska2016] Natalia Siudzińska. 2016. *Formacje ekspresywne we współczesnym języku polskim (na przykładzie wybranych pospolitych nazw osobowych)*. Warszawa.

- [Strapparava and Valitutti2004] 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.
- [Torii et al.2011] 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.
- [Turney and Littman2003] 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.
- [Waszakowa1991] Krystyna Waszakowa. 1991. O wartościowaniu w słowotwórstwie. *Poradnik Językowy*, 5-6:180–186.
- [Xu et al.2013] Jun Xu, Ruifeng Xu, Yanzhen Zheng, Qin Lu, Kam-Fai Wong, and Xiaolong Wang. 2013. Chinese Emotion Lexicon Developing via Multilingual Lexical Resources Integration. In *Proceedings of 14th International Conference on Intelligent Text Processing and Computational Linguistics CI-Cling 2013*, pages 174–182.
- [Zaśko-Zielińska et al.2015] Monika Zaśko-Zielińska, Maciej Piasecki, and Stan Szpakowicz. 2015. A large wordnet-based sentiment lexicon for Polish. In Ruslan Mitkov, Galia Angelova, and Kalina Boncheva, editors, *Proceedings of the International Conference Recent Advances in Natural Language Processing – RANLP’2015*, pages 721–730, Hissar, Bulgaria. INCOMA Ltd. Shoumen, BULGARIA.