Adverbs in plWordNet: Theory and Implementation

Marek Maziarz[A]**, Stan Szpakowicz**[B]**, Michał Kaliński**[A]

[A] Department of Computational Intelligence
Wrocław University of Technology, Wrocław, Poland
[B] School of Electrical Engineering and Computer Science
University of Ottawa, Ottawa, Ontario, Canada

[A] mawroc@gmail.com, m.kalinski.9@gmail.com
[B] szpak@eecs.uottawa.ca

Abstract

Adverbs are seldom well represented in wordnets. Princeton WordNet, for example, derives from adjectives practically all its adverbs and whatever involvement they have. GermaNet stays away from this part of speech. Adverbs in plWordNet will be emphatically present in all their semantic and syntactic distinctness. We briefly discuss the linguistic background of the lexical system of Polish adverbs. We describe an automated generator of accurate candidate adverbs, and introduce the lexicographic procedures which will ensure high consistency of wordnet editors' decisions about adverbs.

1 Adverbs in wordnets and monographs

Adverbs have yet to receive their due in wordnets.

There are only few adverbs in WordNet (hardly, mostly, really, etc.) as the majority of English adverbs are straightforwardly derived from adjectives via morphological affixation (surprisingly, strangely, etc.)[1]

GermaNet shares the basic division of the database into the four lexical categories noun, adjective, verb, and adverb with WordNet®, although it is not planned to implement adverbs in the current work phase.[2]

Curiously, English monographs on lexical semantics (Cruse, 1997; Geeraerts, 2010) give the adverb a short shrift. The term does not even appear in the index of either book!

Yes, most adverbs do derive from adjectives.[3] And yet, the adverb is a *bona fide* open-class part of speech. Its distinctness and its peculiarities cannot be "swept under the carpet" by making it merely an inflected adjective.

Polish morphology acknowledges the adverb grudgingly, but at least it is present in several monographs, notably in (Grzegorczykowa, 1975).

The paper presents a definition of adverbs in plWordNet (section 2), a procedure to generate candidate adverbs (section 3), a manual verification (section 4) and a few conclusions (section 5).

2 Adverbs in plWordNet

The designers of plWordNet established a spectrum of relations for nouns, verbs and adjectives (Maziarz et al., 2011a; Maziarz et al., 2011b; Maziarz et al., 2012). Table 1 lists the relations for adverbs, with examples.[4] The list is based on the adjective model (Maziarz et al., 2012); we have assumed that those relations will fit adverbs, given that most adverbs are transposition derivatives from adjectives.

Every relation type has its own test expressions. (The substitution of lexical units for variables yields correct expressions in Polish.) Language forces the tests to be polymorphic. That is because an adverb can modify a verb, an adjective or an adverb, and it can appear in a predicative position (*jest* 'to be$_{3^{rd} \ person}$' + adverb).

---

[1] https://wordnet.princeton.edu/

[2] http://www.sfs.uni-tuebingen.de/lsd/ germanet_structure.shtml – dated 2009

[3] Calculations on dictionary material show that only 1% of all adverbs is not derived from adjectives (Grzegorczykowa, 1998, p. 524).

[4] See http://tinyurl.com/okdc5w7 for all relations and wordnet editors' instructions (in Polish).

Relation type	definition
Synset relations	
hyponymy	gorączkowo$_1$ 'frantically' $\to$ nerwowo$_1$ 'anxiously'
value of the attribute	intensywnie$_2$ 'intensively' $\to$ intensywność$_1$ 'intensity'
gradation	brązowawo$_1$ 'in brownish colour' $\to$ brązowo$_2$ 'in brown colour'
fuzzynymy	weselnie$_1$ 'in a wedding mood' $\to$ wódka$_1$ 'vodka'
inter-register synonymy	dziwnie$_1$ 'strangely' $\to$ dziwno$_1$ 'strangely (obsolete)'
Lexical unit relations	
antonymy	apriorycznie$_1$ 'a priori' $\leftrightarrow$ aposteriorycznie$_1$ 'a posteriori'
converseness	lepiej$_1$ 'better' $\leftrightarrow$ gorzej$_1$ 'worse'
XPOS synonymy	gorączkowo$_1$ adv. 'frantically' $\to$ gorączkowy$_1$ adj. 'frantic'
degree	lepiej$_1$ 'better' $\to$ dobrze$_1$ 'well'
derivation	intonacyjnie$_1$ 'with regard to intonation' $\to$ intonacja$_3$ 'intonation'

Table 1: Relations in plWordNet with examples.

2.1 Synset relations

Synset relations are short-cuts for a bundle of links between lexical units belonging to two different synsets (Maziarz et al., 2013, pp. 774-775). Our test expression, then, admit pairs of lexical units belonging to synsets which are supposed to be linked by a synset relation.

We present four such tests for *hyponymy*.[5] Symbols x, y denote adverb lexical units. The awkward phrase 'does it x' is meant as "does it in a manner x", *etc.*

When we insert actual words into these tests, we can decide whether the resulting assertion is true. For example, let x and y in Listing 1 be *gorączkowo$_1$* 'frantically' and *nerwowo$_1$* 'anxiously.'

- Jeżeli ktoś robi coś *gorączkowo$_1$*, to robi to *nerwowo$_1$*. 'If someone does something frantically, he does it anxiously.'
- Jeżeli ktoś/coś robi coś *nerwowo$_1$*, to niekoniecznie robi to *gorączkowo$_1$*. 'If someone does something anxiously, he does not necessarily do it frantically.'

Both these statements hold for Polish: the re-

---

[5] We give separate tests for the adjective modifier, the predicative position, and the modifiers of intentional and unintentional verbs; Laskowski (1998) gives an exact definition.

lation **hypo**(*gorączkowo$_1$*, *nerwowo$_1$*), then, is an instance of hyponymy in plWordNet.

Listing 1: Hyponymy. Modifier of intentional verbs.

```
Jeżeli ktoś/coś robi coś x, to robi to y.
Jeżeli ktoś/coś robi coś y, to niekoniecznie robi to x.

'If someone/something does something x,
they do it y.'
'If someone/something does something y,
they do not necessarily do it x.'
```

Listing 2: Hyponymy. Modifier of unintentional verbs.

```
Jeżeli coś dzieje się x, to dzieje się y.
Jeżeli coś dzieje się y, to niekoniecznie dzieje się x.

'If something happens x, it happens y.'
'If something happens y,
it does not necessarily happen x.'
```

Listing 3: Hyponymy. Adjective modifier.

```
Jeżeli ktoś/coś jest x jakiś, to jest też y jakiś.
Jeżeli ktoś/coś jest y jakiś, to niekoniecznie jest x jakiś.

'If someone/something is x so,
they are also y so.'
'If someone/something is y so,
they are not necessarily x so.'
```

Listing 4: Hyponymy. Predicative adverb.

```
Jeżeli jest x, to jest też y.
Jeżeli jest y, to niekoniecznie jest x.

'If it is x, it is also y.'
'If it is y, it is not necessarily x.'
```

Let us now put the hyponymous pair *fiołkowo$_1$* '± like a violet' and *słodko$_2$* 'sweetly' in Listing 2, and replace the generic non-volitional *dzieje się* 'it happens' with its hyponym pachnie 'it smells':

- Jeżeli coś pachnie *fiołkowo$_2$*, to pachnie *słodko$_3$*. 'If something smells like a violet, it smells sweetly.'
- Jeżeli coś pachnie *słodko$_3$*, to niekoniecznie pachnie *fiołkowo$_2$*. 'If something smells sweetly, it does not necessarily smell like a violet.'

In Listing 3, we put the hyponymous pair *bordowo$_1$* 'maroon$_{adv}$' and *ciemnoczerwono$_1$* 'dark-red$_{adv}$' and a specific passive participle *zabarwiony* '*-hued' to replace the generic "so".

- Jeżeli coś jest *bordowo*$_1$ <u>zabarwione</u>, to jest też *ciemnoczerwono*$_1$ <u>zabarwione</u>. 'If something is maroon-hued, it is also dark-red-hued.'
- Jeżeli coś jest *ciemnoczerwono*$_1$ <u>zabarwione</u>, to niekoniecznie jest *bordowo*$_1$ <u>zabarwione</u>. 'If something is dark-red-hued, it is not necessarily maroon-hued.'

Finally, two hyponymous adverbs in a predicative context (to be$_{3rd person}$ + adverb).[6]

- Jeżeli jest *słonecznie*$_6$, to jest też *bezchmurnie*$_4$. 'If it is sunny$_{adv}$, it is also cloudless$_{adv}$'.
- Jeżeli jest *bezchmurnie*$_4$, to niekoniecznie jest *słonecznie*$_6$. 'If it is cloudless$_{adv}$, it is not necessarily sunny$_{adv}$'.

If any of these four tests admits a given pair of lexical units, we will say they are a hyponymy pair.

The relation **value of the attribute** resembles hyponymy. It holds between an adverb, treated as a feature value and a noun, which represents certain category (attribute). For example, the attribute *intensywność*$_1$ 'intensity', has several values, among them the adverbs $intensywnie_2$ 'intesively', $fanatycznie_1$ 'fanatically' and $wydajnie_3$ 'about cough in medicine: efficiently'. Actual hyponymy and value of the attribute together form the backbone of plWordNet's adverb structure.

The **gradation** relation is applied when a series of adverbs can be arranged into a sequence according to some scale. The adverbs *brązowawo*$_1$ 'in brownish colour' and *brązowo*$_2$ 'in brown colour' represent the same attribute *hue* and could be ordered according to that attribute. Adverb sequences can be quite long. Consider adverbs of temperature: *lodowato*$_1$ 'icily', *zimno*$_5$ 'coldly', *zimnawo*$_1$ 'coldishly', *chłodno*$_6$ 'coolly', *chłodnawo*$_1$ 'coolishly', *letnio*$_1$ 'lukewarmly', *ciepło*$_1$ 'warmly', *gorąco*$_1$ 'hotly'.

Inter-register synonymy links adverbs which would be synonymous if not for minor differences in register (in usage). For example, the adverbs *dziwnie*$_1$ and *dziwno*$_1$ occupy nearly the same place in plWordNet's lexico-semantic relation net. They are related to the same lexical units, except for hyponymy (see Figure 1 at the end of section 3). They cannot be in the same synset: *dziwno*$_1$ is obsolete, so is a poor hypernym choice for

contemporary vocabulary, while *dziwnie*$_1$ belongs to the general register.

2.2 Lexical unit relations

The most prominent relation among lexical units is **cross-categorial synonymy**, which we refer to as XPOS synonymy. It binds the adjectival net with the adverbial net. Almost all plWordNet adverbs are related to their derivative bases.[3] An adverb x and its adjective base a are XPOS-synonymous if they can be replaced in the nominalisation process – see (Nagórko, 1987, p. 140) and (Jędrzejko, 1993, p. 61). Two transpositions are possible from a verb context to a nominalised phrase (denoted by the symbol $\Rightarrow$):

- krzątał się gorączkowo 'he bustled frantically' $\Rightarrow$ gorączkowa krzątanina 'frantic bustle',
- jest zimno na ulicy 'it is cold in the street' $\Rightarrow$ zimna ulica 'cold street'.

The test expressions make use of these transpositions. Let us present a test for a modifier of *intentional* verbs (Listing 5; x is an adverb, a is an adjective).

Listing 5: XPOS synonymy. Modifier of intentional verbs.

```
Jeżeli ktoś/coś robi coś x,
to jest to a robienie czegoś przez kogoś/coś.
Jeżeli to jest a robienie czegoś przez kogoś/coś,
to ktoś/coś robi to x.

'If someone/something does something x,
then it is a doing it by someone/
something.'
'If it is a doing something by someone/
something, then someone/something does
not necessarily do it x.'
```

For *gorączkowo*$_1$ and *gorączkowy*$_1$, we get the following test expressions:

- Jeżeli ktoś/coś robi coś *gorączkowo*$_1$, to jest to *gorączkowe*$_1$ robienie czegoś przez kogoś/coś. 'If someone/something does something frantically, then it is frantic doing something by someone/something.'
- Jeżeli jest to *gorączkowe*$_1$ robienie czegoś przez kogoś/coś, to ktoś/coś robi coś *gorączkowo*$_1$. 'If it is frantic doing something by someone/something, then someone/something does something frantically.'

The tests check the truth of two hyponymy-like implications which go in opposite directions.

---

[6] Unlike English, Polish allows both adjectives and adverbs in this position.

Since synonymy can bee seen as bi-directional hyponymy, the tests effectively investigate synonymy conditions for the two parts of speech.

Apart from XPOS-synonymy, the adverbial plWordNet has two more derivationally motivated relations: **degree** and **derivation**. The former caters for synthetic comparatives and superlatives.[7] The latter is a catch-all for other derivational relations.

Antonymy links two adverb lexical units if they satisfy the conditions in Listing 6.

Listing 6: Antonymy. Predicative context.

```
– Jest x?    – Wręcz przeciwnie: jest y .
Jeżeli jest x, to nie jest y.
Jeżeli nie jest x, to niekoniecznie jest y.

– Is it x? – On the contrary:  it is y.
'If it is x, then it is not y.'
'If it is not x, then it is
    not necessarily y.'
```

Semantic opposition was introduced into this test with a short dialogue, with the key word *przeciwnie* 'on the contrary, conversely' (note the predicative context):[8]

- – Jest **x**? '– Is it **x**?'
- – Wręcz przeciwnie: jest **y**. 'On the contrary: it is **y**.'

Consider the pair *słonecznie$_6$* 'sunny$_{adv}$' and *deszczowo$_1$* 'rainy$_{adv}$':

- – Jest *słonecznie$_6$*? – Nie, wręcz przeciwnie: jest *deszczowo$_1$*. '– Is it sunny? – On the contrary: it is rainy.'
- Jeżeli jest *słonecznie$_6$*, to nie jest *deszczowo$_1$*. 'If it is sunny, then it is not rainy.'
- Jeżeli nie jest *słonecznie$_6$*, to niekoniecznie jest *deszczowo$_1$*. 'If it is not sunny, then it is not necessarily rainy.'

---

[7]Degree in Polish adverbs is either synthetic (affix-*ej* for comparatives and *naj-...-ej* for superlatives) or analytic (precede with the adverb *bardzo* 'more' or *najbardziej* 'most', respectively) (Grzegorczykowa, 1998, pp. 533-534).

[8]We follow here a very interesting synonymy test (Cruse, 1997, pp. 257-258): "[N]ot all lexical items are felt to have opposites. Ask someone for the opposite of *table*, or *gold*, or *triangle*, and he will be unable to oblige. Some lexical items, it seems, are inherently non-opposable." The dialogue from our test suggests a language-game in oppositions ("[a]sk someone for the opposite of..."). This helps us throw out those lexical unit pairs which only satisfy the main condition of antonymy, *i.e.*, the incompatibility implication $x \Rightarrow \sim y$ (Lyons, 1981, 154-155).

According to Lyons (1981), converseness is quite frequent among adverbs in the comparative degree whose positive degree is involved in antonymy. We found many such pairs. Listing 7 shows tests for an adjective modifier.

Listing 7: Converseness. Predicative context.

```
Jeżeli p robi coś x niż q, to q robi to y niż p.

'If p does something x than q,
  then q does it y than p.'
```

For example, the lexical units *wolno$_6$* 'slowly' and *szybko$_3$* 'quickly' have the comparatives *wolniej$_1$* 'more slowly' and *szybciej$_1$* 'more quickly'. The test becomes:

- Jeżeli p robi coś wolniej niż q, to q robi to szybciej niż p. 'If p does something more slowly than q, then q does it more quickly than p.'

3 Automatic generation of candidate adverbs

We followed six steps in the generation of new adverbs from their adjective bases. We worked all along with a copy of plWordNet, which we denote plWordNet$_c$.

1. **Derivator**. Consider every existing adjective lemma X within the domain `qualitative` in plWordNet$_c$. Using the Derivator tool (Piasecki et al., 2012) create all possible adverbial derivatives A of adjectives X housed in plWordNet$_c$. The resulting lexicon L contains adverb-adjective pairs of lemmas (A, X).

Table 2 presents the statistics of the derivation process. Since mainly qualitative adjectives form their adverbs, it is interesting that more than one-third of them have their derivatives. For example, for the adjective *czyściutki* 'pleasantly clean, clear, pure' the Derivator created its adverb derivative *czyściutko* '≈cleanly, neatly; purely', whereas for the adjective *poszkodowany* 'injured, damaged' no adverb was derived.

2. **Adverbial lexical units.** For every given qualitative adjective lexical unit x in plWordNet$_c$ representing lemma X which is present in L, create its counterpart lexical unit a representing lemma A. Omit the lexical units housed in artificial (non-lexical) synsets (Piasecki et al., 2009, p. 30). Equip every thus created adverb lexical unit with register labels and glosses copied from the corresponding adjective unit.

Lemma type	Freq.	[%]
Adj. lemmas	27,042	100.0
Qualitative Adj. lemmas	17,045	63.0
Adv. derivative lemmas, $|L|$	6,321	23.4

Table 2: Statistics for automatic adverb derivation by the Derivator and plWordNet$_c$. Abbreviations: Adj. – adjective, Adv. – adverb, $|L|$ – cardinality of the set L.

The rule states that whenever an adjective lexical unit x from the domain `qualitative` has an entry (A, X) in the dictionary L, we create for it its counterpart lexical unit a. For example, lemma *czyściutki* has 5 senses in plWordNet$_c$ in the domain `qualitative`, so the lemma *czyściutko* would have also 5 senses (as).

3. **Filtering rules.** Having created counterparts as for senses xs, we perform filtering based on six rules. Two of them are shown in Listings 8-9. If a rule's premise holds, we remove from plWordNet$_c$ the considered sense a_0 of a given adverb lemma A.

Listing 8: Illustration for rule #1.

```
mod(x_0, istota_1) ∨
∃y [mod(x_0,y) ∧ hypo′(y,istota_1)] ∨
∃y [hypo′(x_0,y) ∧ mod(y,istota_1)] ∨
∃y,n [hypo′(x_0,y) ∧ mod(y,n) ∧ hypo′(n,istota_1)]
```

Symbols x_0, y, z in Listing 8 are lexical units, x and y are adjectives, a_0 is an adverb counterpart of adjective x_0, n is a noun. The noun $istota_1$ means 'being, causal agent, human being, spirit or animal'; $\textbf{hypo}^{'}(x,y)$ holds if y is a direct or indirect hypernym of x; $\textbf{mod}(x,n)$ holds if x is a modifier of n; $\textbf{val}(x,n)$ holds if x is a value of the attribute n.

Listing 9: Illustration for rule #4.

```
val(x_0, zachowanie_7) ∨
∃y [hypo′(x_0,y) ∧ val(y, zachowanie_7)]
```

Symbols in Listing 9 – see Listing 8. The noun $zachowanie_7$ means 'behaviour, manner of acting or controlling oneself'.

Rules #2 and #3 are derived from rule #1 by replacing $istota_1$ with $organizm_1$ 'living organism' and $grupa_5$ 'group of people', respectively. Rules #5 and #6 arise from rule #4 by replacing $zachowanie_7$ by *cecha osobowości*$_1$ 'character trait' and $pochodzenie_5$ 'origin, source of someone/something', respectively. The rules are based on a simple random sample of 69 *adjective* lexical units from plWordNet$_c$ (more in Section 4).

4. **Synsets.** Group all adverbial lexical units into synsets, mirroring their counterpart adjective synsets: two adverb units a_1, a_2 are in the same synset **iff** the corresponding adjective lemmas x_1, x_2 are in the same synset. An adjective lemma can also correspond to two or more adverb lemmas (each with perhaps a slightly different meaning). In such cases, all adverb lexical units $a_1, a_2, \ldots$ are considered counterparts of the same adjective lexical unit x; the register `obsolete` (Maziarz et al., 2014; Maziarz et al., 2015) is assigned to all a_k except the unit of the most frequent adverb lemma.

For example, the lemma *żmudny* 'arduous; laborious' has only one meaning in plWordNet, but two adverbial derivatives in the lexicon L: *żmudnie*, *żmudno* 'arduously; laboriously' (of which the second one is almost absent in modern Polish texts). It has also one synonym *mozolny*. Since *mozolny* has its own adverb derivative *mozolnie*, finally, we get a 3-element synset: {żmudnie$_1$, żmudno$_1$ (`obsolete`), mozolnie$_1$}.

5. **XPOS synonymy.** Add the cross-categorial (XPOS) synonymy between adverb lexical units a and the corresponding adjective lexical units x.

For the adverbs described above, the XPOS synonymy relation instances are the following:

$$\textit{żmudnie} \rightarrow \textit{żmudny},$$
$$\textit{żmudno} \rightarrow \textit{żmudny},$$
$$\textit{mozolnie} \rightarrow \textit{mozolny}.$$

The last step is to copy relations from the adjective part of plWordNet$_c$.

6. **Copying relations.** Copy relations from the adjective part of plWordNet$_c$ onto the adverbial part. This step is split in two sub-steps, one for copying hyponymy chains, and another for copying various other relations.

(a) **Hypernymy/value.** If there is hyponymy between adjectives x and y, their counterpart adverbs a and b are also connected by hyponymy. There also may be "holes" in hyponymy chains, created by adjective synsets which do not have any corresponding adverb synsets (either not generated or filtered

out). Such "holes" are stepped over; see Listing 10.[9] For example, given an adjective chain $x_1 \rightarrow x_2 \rightarrow x_3$ such that only the adverbs a_1 and a_3 exist, the link $a_1 \rightarrow a_3$ is created. The relation "value of the attribute" is treated specially here; it may connect a top adjective hypernym in a chain to a noun. When copying this relation, a top adverb in a hypernymy chain will be linked to that noun if there is a hypernymy + value-of-the-attribute path from its counterpart to the noun; see Listing 11. Figure 1 is a descriptive example of this process.

(b) **Other relations.** Four other adjective-linking relations are copied onto their counterpart adverbs: gradation, inter-register synonymy, antonymy, and converseness. So, if one of these relations connects adjectives x_1, x_2, their counterparts a_1, a_2 will also be connected. Since these relations do not form chains, only immediate neighbours are considered; if one of the connected adjectives has no adverb counterpart, the relation will not be copied.

Listing 10: Illustration for hyponymy chain copying conditions.

$$\forall a, b \;\; \exists x, y \;\; \mathbf{hypo'}(a, b) \Leftarrow$$
$$\mathbf{hypo'}(x, y) \;\wedge\; \mathbf{xpos}(a, x) \;\wedge\; \mathbf{xpos}(b, y)$$

Listing 11: Illustration for value-of-the-attribute relation repair conditions.

$$\forall a, b \;\; \exists x, y, n \;\; \mathbf{val}(a, n) \Leftarrow$$
$$\mathbf{val}(x, n) \;\wedge\; \mathbf{xpos}(a, x) \;\vee$$
$$\mathbf{hypo'}(x, y) \;\wedge\; \mathbf{xpos}(a, x) \;\wedge\; \mathbf{val}(y, n)$$

Symbols x, y, a, b, n in Listings 10-11 are lexical units: x, y are adjectives, a and b are adverbs, n is a noun; $\mathbf{hypo'}(x, y)$ holds if y is a direct or indirect hypernym of x; $\mathbf{val}(x, n)$ holds if x is a value of the attribute n; $\mathbf{xpos}(a, x)$ holds if a is a cross-categorial synonym of x.

Figure 1 illustrates the rule with the hyponymy chain of the synset {postrzelony$_2$} 'crazy'. There are 6 elements in the adjective path (on the left), including the value of the attribute relation. The Derivator did not create some derivatives, so the adverb structure (on the right) is not an exact copy of the adjective part. Luckily, in this case only derivatives forbidden in Polish (marked with "X"

---

[9]$\mathbf{hypo'}$ (•, •) stands for direct or <u>indirect</u> hyponymy.
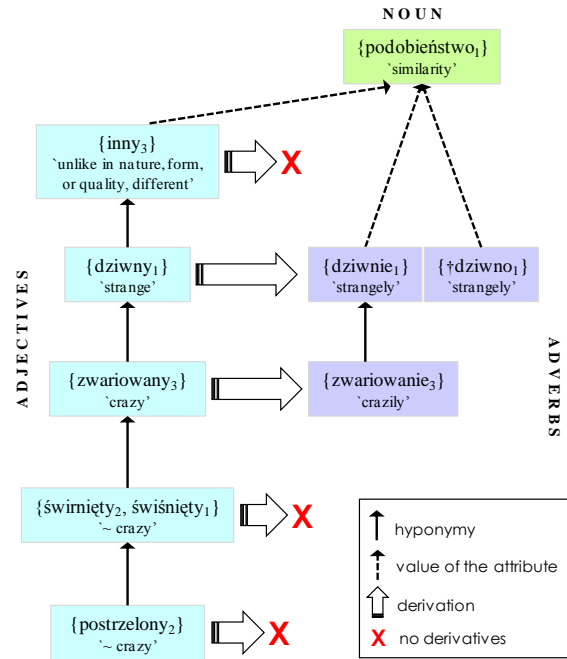

Figure 1: The hyponymy path for *postrzelony* 'crazy'. "X" marks synsets left empty by the algorithm in plWordNet$_c$.

in the Figure) were omitted. Instances lacking relation were stepped over by pointing to the closest synset possible (*dziwnie – podobieństwo*).

4 Manual verification

We evaluate the procedure from section 3 in three experiments, two before copying plWordNet$_c$ onto plWordNet (S_L, S_T), and one afterwards (S_V). The former two were based on simple random samples of 69 (S_L) and 70 (S_T) *adjective* lexical units from plWordNet. The development set S_L helped write and check the filtering rules in Section 3. As a baseline B_L we chose the procedure's performance, without filtering, on the first set of 69 adjectives. The test set S_T was used to reassess the measures of efficiency. The randomly drawn adjectives were checked manually by plWordNet editors (all of them linguists with a university degree) for correspondence with adverbial lexical units .

In the S_L sample (Table 3).[10], two of 27 adverbs in plWordNet$_c$ are our procedure's "creation", and

---

[10]In Tables 3-5, $A+$ / $A-$ denote lexical units which are / are not proper Polish adverbs. $W+$ / $W-$ denote lexical units present / not present in plWordNet$_c$, because either the Derivator did not create them, or they were filtered by rules #1-#6 from step 3 in section 3. $P(W+)$ and $R(A+)$ are precision and recall of recognising real adverb lexical units. CI is the confidence interval.

25 of 36 existing adverbs were introduced into plWordNet$_c$. Let us calculate the precision of introducing adverbs into plWordNet $P(W_+)$ and recall of automatic recognition of adverbial lexical units $R(A_+)$, the most important measures of reliability in this case ($N(\bullet)$ is set cardinality):

$$P(W_+) = N(W_+ \cap A_+)/N(W_+) = 93\% \quad (1)$$

$$R(A_+) = N(W_+ \cap A_+)/N(A_+) = 69\% \quad (2)$$

The set $W_+ \cap A_-$ contains false positives: adverbs which do not exist in reality but were introduced by the algorithm. The set $W_- \cap A_+$ contains false negatives: adverbs which do exist in language but were omitted by the algorithm. For illustration, we present their elements.

- $W_+ \cap A_-$ =
  {kurczliwy$_1$ 'contractible', żeński$_3$ 'female'}

- $W_- \cap A_+$ =
  {redukowalny$_1$ 'reducible', jednosetowy$_1$ 'one-set [*e.g.*, in tennis]', polarny$_1$ 'arctic or antarctic', ropuchowaty$_1$ 'toadlike', włókienkowaty$_1$ 'fibrillose', brutalny$_2$ 'brutal', warzywny$_3$ 'vegetable$_{Adj}$', jednopasmowy$_1$ 'single-lane', równobrzmiący$_1$ 'consonant', pilśniowaty$_1$ 'felt-like', dwupolowy$_2$ 'bi-polar'}

Precision and recall answer two questions:

- How many automatically generated lexical units are real adverb lexical units?

- How many adverb lexical units that could be generated from copying structure from adjective part of plWordNetwere indeed created?

Our procedure performed better on the S_L sample, with a statistically significant increase of precision (from 70% to 93%), and a small, not significant, decrease of recall (from 72% to 69%). The size of the adverbial base in plWordNet$_c$ was only 10% smaller after filtering the original base (see the row M in Table 3).

The results were promising, so we drew yet another sample S_T. Now precision was still high, but recall was lower, however – since we ran the very same algorithm as in S_L – the size M of adverb plWordNet$_c$ (in lexical units) did not change.

With high precision and a reasonably slight "leakage" of lexical units (reasonably high M), we finally decided to copy plWordNet$_c$ onto the live base plWordNet. The plWordNet$_c$ set consisted of

	$\mathbf{B}_L$ ($n = 69$)		$\mathbf{S}_L$ ($n = 69$)	
	W_-	W_+	W_-	W_+
A_-	22	11	31	2
A_+	10	26	11	25
M	11,402		10,190	
$P(W_+)$	70%*		93%*	
95% CI	[53÷84%]		[76÷99%]	
$R(A_+)$	72%		69%	
95% CI	[55÷86%]		[52÷84%]	

Table 3: The confusion matrix for our automatic procedure on the development set. B_L – baseline, the procedure without filtering; S_L – the development set; M is plWordNet$_c$ size, n is sample size, both in lexical units. The asterisks mark statistically significant differences between $\mathbf{B}_L$ and $\mathbf{S}_L$ at the confidence level 95%.

	$\mathbf{S}_T$ ($n = 70$)	
	W_-	W_+
A_-	20	4
A_+	24	22
M	10,190	
$P(W_+)$	85%	
95% CI	[65÷96%]	
$R(A_+)$	45%	
95% CI	[33÷63%]	

Table 4: The confusion matrix for our automatic procedure on the test set. M is plWordNet$_c$ size, n is sample size, both in lexical units.

	$\mathbf{S}_V$ ($n = 517$)	
	W_-	W_+
A_-	NA	86
A_+	100	331
Z	241	
$P(W_+)$	79%	
95% CI	[75÷83%]	
$R(A_+)$	78%	
95% CI	[72÷81%]	

Table 5: The confusion matrix for our automatic procedure on the validation set. S_V – the validation set; Z – the number of adverb lemmas in S_V, and n – sample size in lexical units. Note that the cell $W_- \cap A'_-$ is empty because we changed the interpretation of recall.

10,190 lexical units. We gave the resulting "adverbial" plWordNet to a team of 10 editors, asking them to build upon this automatically generated
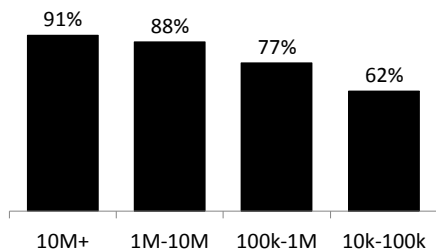
Figure 2: Coverage of lexicon built from plWord-Net Corpus with regard to different frequency bins.

wordnet. Table 5 presents the results of manual verification of part of the automatically generated adverb wordnet; that is the validation set S_V. The conditions of the validation were different than in two earlier experiments S_L and S_T, in which the starting point were adjective lexical units. S_V contained only the *adverb* lemmas generated by the procedure and worked upon by the editors. In S_V, we were not interested in recall of adverbs derivable from the existing adjectives. We changed the interpretation:

- How many adverb lexical units which could have been introduced into plWordNet from generated adverb lemmas were indeed created?

Around one of four-five lexical units is not an appropriate adverb lexical unit; one of four-five existing senses of a given lemma is missing.[11]

5 Whither adverbs in plWordNet?

We have so far only considered adverbs which can be generated from adjectives in plWordNet. It stands to reason that coverage could increase if we worked instead with corpus-based frequency lists. Figure 2 presents coverage of a lexicon built from the plWordNet corpus.[12] The more frequent an adverb is, the more likely it is to appear plWordNet. Even for the least frequent adverbs, the coverage is still a high 62%.

Table 6 shows that our procedure does not miss much. For example (row 3), it only omitted 1418 adverbs with frequency above 10.

	Adverb class	lemmas	%
1	in plWN, $f > 10$	3,720	42.8
2	in plWN, $f <= 10$	2,601	29.9
3	not in plWN, $f > 10$	1,418	16.3
4	multi-word adverbs, *po polsku* type, $f > 10$	958	11.0
	Total (with multi-word adverbs, a guess)	8,697 (≈9,000÷10,000)	100.0

Table 6: The estimated size of plWordNet's adverb list, based of frequencies (f) in the plWordNet corpus.

Row 4 in Table 6 refers to a productive class of multi-word adverbs such as *(mówić) po polsku, po angielsku* '(speak) Polish, English'. There also are other productive patterns, *e.g.*, *(ubierać się) z polska, z niemiecka* '(dress) Polish-style, German-style', as well as non-compositional constructions, *e.g.*, *z dobroci serca* 'out of the goodness of one's heart'. All such adverbial expressions must be added to plWordNet. The "po polsku" type is much more frequent than other types; we found almost 1,000 such word combinations in the corpus. Thus we estimate the number of all other multi-word adverb lexical units at yet another 1,000. We expect, all told, 9 to 10 thousand lemmas.

Clearly, the adding of adverbs to plWordNet is work in progress. Detailed instructions for the editors,[4] in keeping with our practice over the years, are meant to ensure the consistency of editorial decisions. Editors now verify, add to and complete the list of adverb lexical units, automatically generated from plWordNet's adjectives. Next, we plan to add multi-word lexical units of the *po polsku* type and of other types.

Acknowledgments

References

[Broda et al.2012] Bartosz Broda, Michał Marcińczuk, Marek Maziarz, Adam Radziszewski, and Adam Wardyński. 2012. KPWr: Towards a Free Corpus of Polish.

[Cruse1997] Alan Cruse. 1997. *Lexical semantics*. Cambridge University Press.

[Geeraerts2010] Dirk Geeraerts. 2010. *Theories of Lexical Semantics*. Oxford University Press.

[Grzegorczykowa1975] Renata Grzegorczykowa. 1975. *Funkcje semantyczne i składniowe polskich przysłówków [The semantic and syntactic function of Polish adverbs]*. Ossolineum, Wrocław.

[Grzegorczykowa1998] Renata Grzegorczykowa. 1998. IV. Słowotwórstwo: Przysłówek [IV. Derivation: The adverb]. In Renata Grzegorczykowa, Roman Laskowski, and Henryk Wróbel, editors, *Gramatyka współczesnego języka polskiego [Grammar of contemporary Polish]*, volume 2 of *Morfologia [Morphology]*, pages 524–535. PWN, 2 edition.

[Jędrzejko1993] Ewa Jędrzejko. 1993. *Nominalizacje w systemie i w tekstach współczesnej polszczyzny [Nominalisations in language system and in texts of contemporary Polish]*. University of Silesia Press, Katowice.

[Laskowski1998] Roman Laskowski. 1998. Kategorie morfologiczne języka polskiego – charakterystyka funkcjonalna [The morphological categories of Polish – a functional characterisation]. In Renata Grzegorczykowa, Roman Laskowski, and Henryk Wróbel, editors, *Gramatyka współczesnego języka polskiego [Grammar of Contemporary Polish]*, volume 1 of *Morfologia [Morphology]*, pages 151–224. PWN, 2 edition.

[Lyons1981] John Lyons. 1981. *Language and Linguistics: An Introduction*. Cambridge University Press.

[Maziarz et al.2011a] Marek Maziarz, Maciej Piasecki, Stan Szpakowicz, and Joanna Rabiega-Wiśniewska. 2011a. Semantic Relations among Nouns in Polish WordNet Grounded in Lexicographic and Semantic Tradition. *Cognitive Studies / Études Cognitives*, 11:161–181.

[Maziarz et al.2011b] Marek Maziarz, Maciej Piasecki, Stan Szpakowicz, Joanna Rabiega-Wiśniewska, and Bożena Hojka. 2011b. Semantic Relations Between Verbs in Polish Wordnet. *Cognitive Studies / Études Cognitives*, 11:183–200.

[Maziarz et al.2012] Marek Maziarz, Stan Szpakowicz, and Maciej Piasecki. 2012. Semantic Relations among Adjectives in Polish WordNet 2.0: A New Relation Set, Discussion and Evaluation. *Cognitive Studies / Études Cognitives*, 12:149–179.

[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. DOI 10.1007/s10579-012-9209-9.

[Maziarz et al.2014] Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, and Stan Szpakowicz. 2014. Registers in the System of Semantic Relations in plWordNet. In *Proc. 7th International Global Wordnet Conference*, pages 330–337.

[Maziarz et al.2015] Marek Maziarz, Maciej Piasecki, and Stan Szpakowicz. 2015. The system of register labels in plWordNet. *Cognitive Studies / Études Cognitives*, 15:in print.

[Nagórko1987] Alicja Nagórko. 1987. *Zagadnienia derywacji przymiotników [Issues on adjective derivation]*. Warsaw University Press.

[Piasecki et al.2009] Maciej Piasecki, Stanisław Szpakowicz, and Bartosz Broda. 2009. *A Wordnet from the Ground Up*. Wrocław University of Technology Press. http://www.eecs.uottawa.ca/ szpak/pub/A_Wordnet_from_the_Ground_Up.zip.

[Piasecki et al.2012] Maciej Piasecki, Radoslaw Ramocki, and Marek Maziarz. 2012. Recognition of Polish Derivational Relations Based on Supervised Learning Scheme. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 916–922, Istanbul, Turkey. European Language Resources Association (ELRA).

[Przepiórkowski2004] Adam Przepiórkowski. 2004. *The IPI PAN Corpus, Preliminary Version*. Institute of Computer Science PAS.

[Weiss2008] Dawid Weiss. 2008. Korpus Rzeczpospolitej. http://www.cs.put.poznan.pl/dweiss/rzeczpospolita. Corpus of text from the online edtion of Rzeczypospolita.

[Wikipedia2010] Wikipedia. 2010. Polish Wikipedia. https://pl.wikipedia.org, accessed in 2010.