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.).

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.

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. 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. 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 be3rd person + adverb).

Calculations on dictionary material show that only 1% of all adverbs is not derived from adjectives (Grzegorczykowa, 1998, p. 524). See [http://tinyurl.com/okdc5w7](http://tinyurl.com/okdc5w7) for all relations and wordnet editors’ instructions (in Polish).
relation hypo(goręczkowo₁, nerwowo₁), 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 jaki's, to jest też y jaki's.
Jeżeli ktoś/coś jest y jaki's, to niekoniecznie jest x jaki's.

'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 fiołkowo₁ ‘± like a violet’ and słodkowo₂ ‘sweetly’ in Listing 2 and replace the generic non-volitional dzieje się ‘it happens’ with its hyponym pachnie ‘it smells’:

• Jeżeli coś pachnie fiołkowo₂, to pachnie słodkowo₃. ‘If something smells like a violet, it smells sweetly.’
• Jeżeli coś pachnie słodkowo₃, to niekoniecznie pachnie fiołkowo₂. ‘If something smells sweetly, it does not necessarily smell like a violet.’

In Listing 3 we put the hyponymous pair bordowo₁ ‘maroon adv’ and ciemnoczerwono₁ ‘dark-red adv’ and a specific passive participle zabarwiony ‘*-hued’ to replace the generic “so”.

| Relation type | definition |
|---------------|------------|
| Synset relations | goręczkowo ‘frantically’ → nerwowo ‘anxiously’ |
| value of the attribute | intensywnie ‘intensively’ |
| gradation | brązowo ‘brownish colour’ |
| fuzzynymy | weselnie ‘in a wedding mood’ |
| inter-register | dzwieni ‘strangely’ |
| synonymy | → dziwno ‘strangely obsolete’ |

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[^1][^2] 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 goręczkowo₁ ‘frantically’ and nerwowo₁ ‘anxiously’.

- Jeżeli ktoś robi coś goręczkowo₁, to robi to nerwowo₁. ‘If someone does something frantically, he does it anxiously.’
- Jeżeli ktoś/coś robi coś nerwowo₁, to niekoniecznie robi to goręczkowo₁. ‘If someone does something anxiously, he does not necessarily do it frantically.’

Both these statements hold for Polish: the re-
Finally, two hyponymous adverbs in a predica
tive context (to be3y4person + adv)6

• Jeżeli coś jest bordowo1 zabarwione, to jest też
ciennoczerwono1 zabarwione. ‘If something
is maroon-hued, it is also dark-red-hued.’
• Jeżeli coś jest ciennoczerwono1 zabarwione,
to niekoniecznie jest bordowo1 zabarwione. ‘If
something is dark-red-hued, it is not necessarily
maroon-hued.’

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 intensywność1 ‘intensity’,
has several values, among them the adverbs
intensywnie2 ‘intensively’, fanatycznie1 ‘fanatic-
ally’ and wydajnie3 ‘about cough in medicine:
efficiently’. Actual hyponymy and value of the at-
tribute 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 accord-
ing to some scale. The adverbs brązowawo1 ‘in
brownish colour’ and brązowo2 ‘in brown colour’
represent the same attribute hue and could be
ordered according to that attribute. Adverb se-
quences can be quite long. Consider adverbs of
temperature:lodowato1 ‘icily’, zimno5 ‘coldly’,
zinnawo1 ‘coldishly’, chłodno6 ‘coolly’, chłod-
nawo1 ‘coolishly’, letnio1 ‘lukewarmly’, ciepło1
‘warmly’, gorąco1 ‘hotly’.

**Inter-register synonymy** links adverbs which
would be synonymous if not for minor differences
in register (in usage). For example, the
adverbs dziwnie1 and dziwno1 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). They cannot be in the same synset: dzi-
wno1 is obsolete, so is a poor hypernym choice for

6Unlike English, Polish allows both adjectives and ad-
verbs in this position.

### 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. An adverb
x and its adjective base a are XPOS-synonymous
if they can be replaced in the nominalisation pro-
cess – see (Nagórko, 1987, p. 140) and (Jędrejko,
1993, p. 61). Two transpositions are possible from
a verb context to a nominalised phrase (denoted by
the symbol ⇒):

- krzątał się gorączkowo ‘he hustled frantically’
⇒ gorączkowa krzątanina ‘frantic bustle’;
- jest zimno na ulicy ‘it is cold in the street’ ⇒
zimna ulica ‘cold street’.

The test expressions make use of these trans-
positions. 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 inten-
tional verbs.

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

‘If someone/something does something x,
then it is doing it by someone/ something.’

‘If it is doing something by someone/
something, then someone/something does
not necessarily do it x.’

For gorączkowo1 and gorączkowy1, we get the
following test expressions:

- Jeżeli ktoś/coś robi coś gorączkowo1, to jest
to gorączkowe1 robienie czegoś przez kogoś/coś. ‘If someone/something does something
frantically, then it is frantic doing something
by someone/something.’
- Jeżeli jest to gorączkowe1 robienie czegoś przez kogoś/coś, to ktoś/coś robi coś gorączkowo1, ‘If it is frantic doing
something by someone/something, then some-
one/something does something frantically.’

The tests check the truth of two hyponymy-
like implications which go in opposite directions.
Since synonymy can be 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. 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).

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

Consider the pair **słonecznie**6 ‘sunny+a’ and **deszczowo**7 ‘rainy+a’:

- – Jest **słonecznie**6? – Nie, wręcz przeciwnie: jest **deszczowo**7. ‘Is it sunny? – On the contrary: it is rainy.’
- Jeżeli jest **słonecznie**6, to nie jest **deszczowo**7. ‘If it is sunny, then it is not rainy.’
- Jeżeli nie jest **słonecznie**6, to niekoniecznie jest **deszczowo**7. ‘If it is not sunny, then it is not necessarily rainy.’

According to [Lyons (1981)](Lyons1981), 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 niż q, to q robi do 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**9 ‘quickly’ have the comparatives **wolniej**8 ‘more slowly’ and **szybciej**8 ‘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](Piasecki2012)) 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 **czyściuto** ‘≈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 plWordNetc, 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](Piasecki2009), p. 30). Equip every thus created adverb lexical unit with register labels and glosses copied from the corresponding adjective unit.

---

7In Polish adverbs is either synthetic (suffix-ef for comparatives and naj-...ej for superlatives) or analytic (precede with the adverb bardzo ‘more’ or najbardziej ‘most’, respectively) ([Chmiel, 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 ⇒ ¬ y ([Lyons, 1981] 154-155).
Table 2: Statistics for automatic adverb derivation by the Derivator and plWordNet,\(_c\). Abbreviations: Adj. – adjective, Adv. – adverb, \(|L|\) – cardinality of the set \(L\).

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

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 \(\text{czyściutko}\) has 5 senses in plWordNet, in the domain qualitative, so the lemma \(\text{czyśćciutko}\) would have also 5 senses (as).

3. Filtering rules. Having created counterparts as for senses \(x\), we perform filtering based on six rules. Two of them are shown in Listings 8 and 9. If a rule’s premise holds, we remove from plWordNet, the considered sense \(a_0\) of a given adverb lemma \(A\).

Listing 8: Illustration for rule #1.

| \(\text{mod}(x_0, \text{istota}_1)\) \& \(\exists y [\text{mod}(x_0, y) \land \text{hypo}'(y, \text{istota}_1)]\) \& \(\exists y [\text{hypo}'(x_0, y) \land \text{mod}(y, \text{istota}_1)]\) \& \(\exists y, n [\text{hypo}'(x_0, y) \land \text{mod}(y, n) \land \text{hypo}'(n, \text{istota}_1)]\) |

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 \(\text{istota}_1\) means ‘being, causal agent, human being, spirit or animal’; \(\text{hypo}'(x, y)\) holds if \(y\) is a direct or indirect hypernym of \(x\); \(\text{mod}(x, n)\) holds if \(x\) is a modifier of \(n\); \(\text{val}(x, n)\) holds if \(x\) is a value of the attribute \(n\).

Listing 9: Illustration for rule #4.

| \(\text{val}(x_0, \text{zachowanie}_7)\) \& \(\exists y [\text{hypo}'(x_0, y) \land \text{val}(y, \text{zachowanie}_7)]\) |

Symbols in Listing 9 – see Listing 8. The noun \(\text{zachowanie}_7\) means ‘behaviour, manner of acting or controlling oneself’.

Rules #2 and #3 are derived from rule #1 by replacing \(\text{istota}_1\) with \(\text{organizm}_1\) ‘living organism’ and \(\text{grupa}_5\) ‘group of people’, respectively. Rules #5 and #6 arise from rule #4 by replacing \(\text{zachowanie}_7\) by \(\text{cecha osobowości}_1\) ‘character trait’ and \(\text{pochodzenie}_5\) ‘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 \(\text{iff}\) 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 \(\text{źmudny}\) ‘arduous; laborious’ has only one meaning in plWordNet, but two adverbial derivatives in the lexicon \(L\): \(\text{źmudnie}, \text{źmudno}\) ‘arduously; laboriously’ (of which the second one is almost absent in modern Polish texts). It has also one synonym \(\text{mozolny}\). Since \(\text{mozolny}\) has its own adverb derivative \(\text{mozolnie}\), finally, we get a 3-element synset: \{\(\text{źmudnie}_1, \text{źmudno}_1\) (obsolete), mozolnie\}.

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:

\(\text{źmudnie} \rightarrow \text{źmudny}, \text{źmudno} \rightarrow \text{źmudny}, \text{mozolnie} \rightarrow \text{mozolny}\).

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) Hyponymy/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. For example, given an adjective chain \(x_1 \rightarrow x_2 \rightarrow x_3\) such that only the adverbs \(a_1\) and \(a_2\) 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.

\[
\forall a, b \exists x, y \ hypo(a, b) \iff \\
\hypo(x, y) \land xpos(a, x) \land xpos(b, y)
\]

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

\[
\forall a, b \exists x, y, n \ val(a, n) \iff \\
val(x, n) \land xpos(a, x) \lor \\
hypo(x, y) \land xpos(a, x) \land \val(y, n)
\]

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; \(\hypo(x, y)\) holds if \(y\) is a direct or indirect hypernym of \(x\); \(\val(x, n)\) holds if \(x\) is a value of the attribute \(n\); \(\xpos(a, x)\) holds if \(a\) is a cross-categorical synonym of \(x\).

Figure 1 illustrates the rule with the hyponymy chain of the synset \{postrzeleny2\} ‘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”

Figure 1: The hyponymy path for postrzeleny ‘crazy’. “X” marks synsets left empty by the algorithm in plWordNet.

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, 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), two of 27 adverbs in plWordNet, are our procedure’s “creation”, and

---

\(\hypo(\bullet, \bullet)\) stands for direct or indirect hyponymy.
25 of 36 existing adverbs were introduced into plWordNet. 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 (\(N(A)\) is set cardinality):

\[
P(W_+) = \frac{N(W_+ \cap A_+)}{N(W_+)} = 93\% \quad (1)
\]

\[
R(A_+) = \frac{N(W_+ \cap A_+)}{N(A_+)} = 69\% \quad (2)
\]

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\ ‘contractible’, żeński\ ‘female’\}

- \(W_- \cap A_+\) =

\{redukowalny\ ‘reducible’, jednosetowy\ ‘one-set [e.g., in tennis]’, polarny\ ‘arctic or antarctic’, ropuchowy\ ‘toadlike’, włókienkowaty\ ‘fibrillose’, brutalny\ ‘brutal’, warzywny\ ‘vegetable\(\text{Adj}\)’, jednopasmowy\ ‘single-lane’, równobrzmiący\ ‘consonant’, piśmienniowaty\ ‘felt-like’, dwupolowy\ ‘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 plWordNet were 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 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 \(S_V\). 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 (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, onto the live base plWordNet. The plWordNet set consisted of

| \(W_-\) (\(n = 69\)) | \(W_+\) (\(n = 69\)) | \(W_-\) (\(n = 69\)) | \(W_+\) (\(n = 69\)) |
|-----------------|-----------------|-----------------|-----------------|
| \(A_-\) | 22 | 11 | 31 | 2 |
| \(A_+\) | 10 | 26 | 11 | 25 |
| \(M\) | 11,402 | 10,190 | | |

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 size, \(n\) is sample size, both in lexical units. The asterisks mark statistically significant differences between \(B_L\) and \(S_L\) at the confidence level 95%.

| \(S_T\) (\(n = 70\)) | | | |
|-----------------|-----------------|-----------------|-----------------|
| \(W_-\) | 20 | 4 | | |
| \(A_+\) | 24 | 22 | | |
| \(M\) | 10,190 | | | |

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

| \(S_V\) (\(n = 517\)) | | | |
|-----------------|-----------------|-----------------|-----------------|
| \(W_-\) | NA | 86 | | |
| \(A_+\) | 100 | 331 | | |
| \(Z\) | 241 | | | |

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
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.  

### 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. 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, $f \leq 10$ | 2,601 | 29.9 |
| 3 not in plWN, $f > 10$ | 1,418 | 16.3 |
| 4 multi-word adverbs, $f > 10$ | 958 | 11.0 |
| Total (with multi-word adverbs, a guess) | $8,697$ | 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 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

Work supported by the Polish Ministry of Education and Science, Project CLARIN-PL, the European Innovative Economy Programme project POIG.01.01.02-14-013/09, and by the EU’s 7FP under grant agreement No. 316097 [ENGINE]. Thanks to Paweł Kędzia for help with the adverb generation algorithm. Thanks to Agnieszka Dziob and Justyna Wieczorek, the co-authors of the adverb guidelines, for the manual verification of the learning and test sets.
References

[Broda et al.2012] Bartosz Broda, Michał Marciniczuk, Marek Maziarz, Adam Radziiszewski, 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 semanticzne 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.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. Wroclaw 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/rzeczpospolit.html. Corpus of text from the online edition of Rzeczypospolita.

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