Classification and Generation of Derivational Morpho-semantic Relations for Polish Language

Wiktor Walentynowicz, Maciej Piasecki, Mateusz Gniewkowski

Faculty of Information and Communication Technology, Wrocław University of Science and Technology, Wrocław, Poland
wiktor.walentynowicz@pwr.edu.pl

Abstract. In this article, we take a new look on automated analysis and recognition of morpho-semantic relations in Polish. We present a combination of two methods for join exploration on word-form information – generating new forms and classifying pairs of words in derivational relations. As a method of generation, we used the Transformer architecture in the seq-2-seq task. Classification is performed using a neural network and using the fastText representation method. At the very end, we discussed the results obtained in the experiments.

Keywords: derivational morphology · morpho-semantic relations · word formation · relation classification

1 Introduction

Word formation processes can be observed in the vast majority of natural languages: *derivatives* are formed from *derivational bases* by means of language specific derivational mechanisms, e.g. *a teacher* from *to teach*, *a duchess* from *a duke* or Polish *domeczek* ≈'a nice, little house' from *dom* 'a house', *białość* ≈'a state of being white' from *biały* ≈'white'. For some natural languages, especially for the inflectional Slavic languages, such mechanisms constitute a very productive system. That is why native speakers can recognise a new derivative as a language unit and identify its derivational base with high precision. What is more derivational relations, in contrast to morpho-syntactic word formation processes signal a meaning change between a basis and the derivative, also predictive to a very large extent, e.g. *palarnia* ≈ 'a place for smoking' derived from *palić* 'to smoke'. Due to this, they are called *morphosemantic relations* [5].

Morphosemantic relations combine two transformations: between word forms and, in parallel, between lexical meanings, that are tightly coupled: different types of word form transformations are characteristic for semantic derivations, e.g. *kierowniczka* ≈ 'a female head or manager' derived from *kierownik* 'a head or manager' primarily by the suffix *-ka*. Derivation rules can be described to some extent by a combination of suffixes, prefixes and inside stem alternations. However such word form level rules are semantically, e.g., the suffix *-ka* mostly signals: `+Male` → `+Female`, but it appears in tool name derivation, too: *wiercić* 'to drill' → *wiertarka* 'a driller', and can be also misleading: *pierwiastka* 'a woman

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_33

giving birth for the first time' is not a female form of *pierwiastek* 'root', in spite of '*ka*'. Thus proper recognition and interpretation of derivational requires taking into account both types of transformations: morphological and semantic.

The goal of our work is to developed a mechanism for recognition and interpretation of derivatives in a way combining morphological and lexico-semantic level. For a given word, a potential derivative, we want to recognise a set of words with which it is in a certain lexico-semantic relation, and also its derivational basis. We propose machine learning means for both levels: word form and semantic. The unique feature of our approach is a combination of transformer-based neural architecture for modelling derivational patterns tightly coupled with recognition of lexico-semantic relations based on non-contextual word embeddings as semantic representation. We focus on the Polish language for which a large and rich model of morphosemantic relations is included in plWordNet [3]. Contrary to many other wordnets and derivational dictionaries, the plWordNet morphosemantic relations link particular senses of two words, not the word forms. In addition, these relations are always directed according to the derivational processes in Polish: from a derivational basis to the derivative.

2 Related Work

Derivational relations are often described in morphological dictionaries as links between lemmas, e.g. [6], [19] or a very large morphological and derivational network DeriNet [16], only later automatically classified to 5 very coarse-grained semantic classes [18]. In [18] the training data were pairs of words (not senses) and classification was based on morphological features of word forms. Semantic annotation of word pairs was adopted for wordnets (lexico-semantic networks), e.g. RoWordNet [10], BulNet [10, 2] or CroWN [20]. However, in wordnets, links between lemmas are additionally labelled with semantic relations, i.e. mapped onto morphosemantic relations. plWordNet [3] showed that such an approach is simplification and prone to errors, as different morphosemantic relations may be valid only for selected senses of lemmas. Thus, we focus on morphosemantic relations as linking senses, but signalled by derivational associations.

In [14] two character-level transducers extracted from training data and combined with internal stem alternations were proposed. Relations suggested by transducers were next filtered by grammatical patterns, corpus frequency and semantic classifiers for word pairs. trained a combination of features describing word distributions in a large corpus. The best results were reported for the set of 9 most populated relations: 36.84 (the young being relation) up to 97.19 (femininity) of F1. However, it should be emphasised that in this case wordnet-internal knowledge about assignment of lemmas to WordNet domains [4] was utilised. We do not use such knowledge in our approach. In a similar approach [8], but much more supported by hand-crafted knowledge F1=0.682 was achieved for verb and noun synset pairs in BulNet. A sequential pattern mining technique based on regular expressions as features for ML was proposed in [9] and tested on Polish and Spanish. It was trained on "1500 pairs of base words with their

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_33

derivatives". However, the annotation guidelines are unknown, semantics of the links was not taken into account, as well as the direction of derivation. Finally, the accuracy of 82.33% was achieved with "53.5 thousand links in the network".

Word embeddings (word2vec and neural language models) were investigated in [11] for the Czech coarse-grained derivational relations. Neural character encoder–decoder was applied to predict a derivative from a derivational base in [17]. It used occurrence context too, but was limited to deverbal nouns.

Our main contribution is a combined method for transforming a word form into its derivational basis, both perceived as a lexical units, and recognising the morphosemantic relation. For transformation, we applied Transformer [15] to characters. Information about the relation linking the input and output words is delivered to the decoder, so the encoder is relation-independent, and can be used as a source of embeddings vector for characters containing derivational information. The semantic aspect is represented by a classifier filtering the transformer results as lexical units from the morphosemantic relation perspective. The combination of these two allows for detection and suggestion of derivational relations between lexical units in any wordnet as our method does not require any language-specific knowledge resource, except a training set of relation instances.

3 Data

Coarse-grained	Fine-grained	Cardinality	Coarse-grained	Fine-grained	Cardinality
aspectuality	pure aspectuality	31030	role ADJ-V	agent	1694
	secondary aspectuality	7457		time	167
characteristic	characteristic	5366		location	937
markedness	diminutives	4184		instrument	322
	augmentatives	886		patient	306
	young being	83		product	85
markedness-intensity	markedness-intensity	996		cause	427
state/feature bearer	state/feature bearer	1410	role material	material	1315
similarity	similarity	2171	state/feature	state/feature	1410
predisposition	habituality	120	*CCS*	ADJ-N	4507
	quantification	15		ADV-ADJ	11355
	appreciation	21		N-ADJ	4506
	potential	334		N-V	30262
role	agent	153		V-N	30262
	time	36		for relational	17069
	location	25	role inclusion	agent inclusion	124
	instrument	299		time inclusion	38
	patient	1039		location inclusion	46
	product	1521		instrument inclusion	515
	agent of *HP*	10		patient inclusion	234
	location of *HP*	250		product inclusion	786
	product of *HP*	3762	femininity	femininity	3789

Table 1. Relationships found in plWordNet at different granularities. *HP* - Hidden Predicate, *CCS* - Cross-Categorial Synonymy

The applied dataset from plWordNet (the 25.02.2020 dump) consists of triples: a derivational base, a derivational relation and a derivative; 111,955 triples, of which 19,441 triples with multi-words. The data were divided into training and test: 9 to 1. The division was done in two ways. The first one is a split balanced in terms of the number of relations. The second is a split in which the same derivational bases do not occur in both sets. For the relation classification

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_33

task, we considered only relations with at least 300 instances. For our classification experiments, we selected relation instances consisting of only single words. plWordNet derivational relations can be divided according to two levels of granularity – sparse (coarse grained) and dense (fine grained), see Table 1 and [14].

4 Proposed approach & Experiments

The tasks is: for a triple (a derivational base lemma, a derivative lemma and the derivational relation), given two of these three pieces of information, we want to obtain the value of the unknown information. This results in three tasks: generation of a derivational base lemma (*derivational analysis*), generation of a derivative lemma (*derivational generation*), a *classification of a derivational relation*. The first two require transforming a character sequence into another character sequence, i.e., *sequence-to-sequence method*. The third task is classification. The *Derivator* system proposes a combination of these two types.

The derivational form generation is done by a sequence-to-sequence neural network based on Transformer [15]. Our Transformer model has a character-based input and is not pre-trained. In our approach, the decoder receives as an additional, special first token – relation tag for which it performs the transformation. The Transformer decoder, performs non-autoregressive decoding, so for our purpose it was enhanced with an autoregressive decoding module. This was necessary, because the length of the generated sequence is not known beforehand. Token selection, in each decoding step, is done using a greedy step-by-step method. In addition, we prepared a derivational analyser: a derivative to its base lemma, also taking into account the derivational relation. The analyser has the same architecture as the generator, the only difference is the transformation direction to be learned. The implementation is based on *PyTorch* [12].

For the derivation generation task, we performed experiments on the two defined splits of the datasets with and without multi-word lexical units (MWE): 75.35% (stratified, without MWE), 73.88% (grouped, without MWE), 76.07% (stratified, with MWE) and 77.14% (grouped, with MWE). In the derivation analysis task we obtained: 82.65% (stratified, without MWE), 74.42% (grouped, without MWE), 83.20% (stratified, with MWE) and 77.20% (grouped, with MWE). The given measure is the sequence identity accuracy.

The derivation relation classifier was based on Multilayer Perceptrons trained as one-vs-rest, that triggered a problem arises of selecting negative examples. We tested different approaches. The first was to use examples from the other relations as negative ones. However, it resulted in a highly unbalanced training set. For training our classifiers, we also chose examples from other relations as negative examples only that include lexical units of the same Part-of-Speech (PoS) pairs. This is well suited, given that our method of filtering candidate derivation pairs uses PoS information, so a classifier to classify a given relation, will never encounter a pair that does not match. Finally, we also tried samples for which the correspondence occurred at the PoS level of the word base as negative examples. The experiments were performed with *scikit-learn* [13].

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_33

We divided the classification problem into three experiments according to the input representation. In the first, we used difference of the derivative and base fastText vectors [1] from KGR10 [7]. In the second the difference vector was concatenated with the base and derivative vectors. In the last the difference vector was expanded with the difference of the vectors of pseudoaffixes. Pseudoaffixes were obtained by finding the longest common sequence between the lemmas of a base and a derivative, next separating prefixes and suffixes and building fastText vectors for them. In the presentation, we use the following notations: 1D – word vector difference; 1DNG – word vector difference vector only using fastText n-grams alone; 1DAF – concatenation of a word difference vector and pseudoaffixes difference vectors; 2DAF – concatenation of pseudoaffixes difference vectors; 3D – concatenation of word difference vector and word vectors. Table 2 presents classification results per relation. For the other two experiments, the differences in results were about 1-2 percentage points more for the weighted average. The implementation and experimental results are available in a repository at: `https://gitlab.clarin-pl.eu/morphology/derywator`

	General relation	Detailed relation	1D	1DNG	1DAF	2DAF	3D	Support
one-vs-rest classifiers	aspectuality	pure aspectuality	0,92	0,91	0,95	0,94	0,93	939
	aspectuality	secondary aspectuality	0,50	0,00	0,67	0,68	0,61	170
	characteristic	characteristic	0,65	0,66	0,70	0,66	0,62	411
	cross-categorial synonymy	ADJ-N	0,93	0,93	0,94	0,94	0,92	328
	cross-categorial synonymy	ADV-ADJ	0,99	0,98	0,99	0,99	0,98	578
	cross-categorial synonymy	for relational	0,84	0,83	0,86	0,85	0,86	1386
	cross-categorial synonymy	N-ADJ	0,95	0,95	0,96	0,94	0,96	328
	cross-categorial synonymy	N-V	0,98	0,98	0,98	0,98	0,98	1126
	cross-categorial synonymy	V-N	1,00	1,00	1,00	1,00	1,00	1126
	femininity	femininity	0,97	0,97	0,97	0,90	0,96	324
	markedness	augmentatives	0,00	0,00	0,83	0,78	0,64	74
	markedness	diminutives	0,86	0,86	0,94	0,86	0,93	306
	role	agent	0,79	0,80	0,81	0,77	0,81	122
	role	agent of hidden predicate	0,92	0,92	0,94	0,91	0,92	325
	role	instrument	0,66	0,64	0,70	0,38	0,64	78
	role	location of hidden predicate	0,86	0,76	0,87	0,78	0,84	39
	role	product	0,67	0,64	0,68	0,64	0,67	125
	role ADJ-V	agent	0,88	0,92	0,91	0,95	0,91	66
	role inclusion	instrument inclusion	0,65	0,26	0,71	0,47	0,70	44
	role inclusion	product inclusion	0,00	0,04	0,70	0,27	0,73	55
	role material	material	0,05	0,00	0,08	0,00	0,20	117
	similarity	similarity	0,34	0,35	0,40	0,31	0,38	184
	state/feature	state/feature	0,47	0,46	0,55	0,43	0,57	125
	state/feature bearer	state/feature bearer	0,60	0,63	0,69	0,61	0,61	125
	Mean	*uniform*	*0,69*	*0,65*	*0,78*	*0,71*	*0,76*	*8501*
	Mean	*weighted*	*0,85*	*0,83*	*0,88*	*0,86*	*0,87*	*8501*

Table 2. F1-Score measure values for all relations classifier.

5 Discussion

The character Transformer model handles the derivative analysis task more efficiently than the generation one, which is expected because the analysis task has a smaller target domain. The addition of multi-word lemmas to the training and evaluation data had positive effect. We believe that data that increases the variation of the training samples, even if they are more complex samples – increasing the sequence fragments that do not change – improves generalisation of the model. Considering the averages, the best classifier uses the combined information of the difference vectors for: words, prefixes and endings. It achieved

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_33

the highest weighted average values in all types of environment – including all relationships, filtering based on base PoS and filtering based on whole relationship PoS. The fact that the classifier is not based on vectors of specific words – like the 3D classifier – is also a good signal for generalisation. Comparing the results of the 1DAF classifier with those from [14] for selected relations we observe improvement 6 to 51 percentage points.

Explainability is important for a classification task, as it provides additional information on the quality and usability. The features that are most important to the classifier do not always make sense. In order to verify our model predictions, we try to estimate an importance of a word's n-grams. We do that by recalculating prediction probability of a previously trained model for every missing n-gram. If the probability of a sample not being in a given relation grows significantly, then the feature is relevant. The importance score of an n-gram is a difference between prediction probabilities (of a negative class) of the sample and its neighbour. As fastText generates a word embedding as the average of all the n-grams and the word representation, we remove each subsequent components of the sum. We do that only for derivatives. The input vector for a classifier is then calculated as described before – by subtracting an unchanged base vector from a newly created derivative embedding. We observed that for most relations, the key element in the classification is the ending, which agrees with linguistic research. It is also common for word endings to be considered as elements 2-3 characters long. For significant n-grams, the vast majority are 5-6 character n-grams.

6 Conclusion and Future Work

In this paper, we present a combined method for the generation of derivational word forms taking into account a potential morphosemantic relation, using a sequence-to-sequence technique based on Transformer and merged with semantic relation classifiers. We identified important features for decision-making by classifiers and compared them with linguistic research. Our system can be used to detect potential relations by generating potential derivational bases (derivatives) for different relations using Transformer and next verifying the results with the semantic classifier. The method can also be used to generate derivatives from the base lemmas and next to increase the coverage of a lexico-semantic network. The obtained results are better or at least comparable to those of the previous methods, but with more fine grained semantic classification. Our method works on lexical units, contrary to the previous methods focused on words alone. Definitely, there is still plenty of space for improvement. One of the future directions is tighter integration of relation classification with generation of new words.

Acknowledgements: Financed by the European Regional Development Fund as a part of the 2014-2020 Smart Growth Operational Programme, CLARIN - Common Language Resources and Technology Infrastructure, project no. POIR.04.02.00-00C002/19.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_33

References

1. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606 (2016)
2. Dimitrova, T., Tarpomanova, E., Rizov, B.: Coping with derivation in the bulgarian wordnet. In: Proceedings of the Seventh Global Wordnet Conference (2014)
3. Dziob, A., Piasecki, M., Rudnicka, E.: plWordNet 4.1 - a linguistically motivated, corpus-based bilingual resource. In: Proceedings of the 10th Global Wordnet Conference (2019)
4. Fellbaum, C. (ed.): WordNet – An Electronic Lexical Database. The MIT Press (1998)
5. Fellbaum, C., Osherson, A., Clark, P.E.: Putting semantics into wordnet's "morphosemantic" links. In: Human Language Technology. Challenges of the Information Society, Third Language and Technology Conference, LTC 2007 (2007)
6. Kanuparthi, N., Inumella, A., Sharma, D.M.: Hindi derivational morphological analyzer. In: Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology (2012)
7. Kocoń, J., Gawor, M.: Evaluating KGR10 Polish word embeddings in the recognition of temporal expressions using BiLSTM-CRF. Schedae Informaticae **27** (2018)
8. Koeva, S., Leseva, S., Stoyanova, I., Dimitrova, T., Todorova, M.: Automatic prediction of morphosemantic relations. In: Proceedings of the 8th Global WordNet Conference (GWC) (2016)
9. Lango, M., Ševčíková, M., Žabokrtský, Z.: Semi-automatic construction of word-formation networks (for polish and spanish). In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (2018)
10. Mititelu, V.B.: Adding morpho-semantic relations to the romanian wordnet. In: Proc. of the Eighth Inter. Conf. LREC'12 (2012)
11. Musil, T., Vidra, J., Mareček, D.: Derivational morphological relations in word embeddings. In: Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP (2019)
12. Paszke, A., et al.: Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32 (2019)
13. Pedregosa, F., et al.: Scikit-learn: Machine learning in Python (2011)
14. Piasecki, M., Ramocki, R., Maziarz, M.: Recognition of polish derivational relations based on supervised learning scheme. In: Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12) (2012)
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in neural information processing systems (2017)
16. Vidra, J., Žabokrtský, Z., Ševčíková, M., Kyjánek, L.: DeriNet 2.0: Towards an all-in-one word-formation resource. In: Proceedings of the Second International Workshop on Resources and Tools for Derivational Morphology (2019)
17. Vylomova, E., Cotterell, R., Baldwin, T., Cohn, T.: Context-aware prediction of derivational word-forms. In: Proceedings of the 15th Conference of the European Chapter of the ACL (2017)
18. Ševčíková, M., Kyjánek, L.: Introducing semantic labels into the derinet network. Journal of Linguistics/Jazykovedný casopis **70**(2) (2019)
19. Šnajder, J.: Derivbase.hr: A high-coverage derivational morphology resource for croatian. In: Proc. of the Ninth Inter. Conf. LREC'14 (2014)
20. Šojat, K., Srebačić, M.: Morphosemantic relations between verbs in croatian wordnet. In: Proceedings of the Seventh Global Wordnet Conference (2014)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_33