Wordnet-oriented Recognition of Derivational Relations

Wiktor Walentynowicz, Maciej Piasecki
Wrocław University of Science and Technology
{wiktor.walentynowicz|maciej.piasecki}@pwr.edu.pl

Abstract

Derivational relations are an important element in defining meanings, as they help to explore word-formation schemes and predict senses of derivates (derived words). In this work, we analyse different methods of representing derivational forms obtained from WordNet – from quantitative vectors to contextual learned embedding methods – and compare ways of classifying the derivational relations occurring between them. Our research focuses on the explainability of the obtained representations and results. The data source for our research is plWordNet, which is the wordnet of the Polish language and includes a rich set of derivation examples.

1 Introduction

Word formation processes can be observed in many, if not all 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, from the Polish language, *domeczek* ≈'a nice, little house' from *dom* 'a house', *białość* ≈'a state of being white' from *biały* ≈'white'. In some natural languages, especially in the case of inflectional ones, e.g. Slavic languages, such mechanisms constitute a very productive system. That is why native speakers can recognise a new derived word forms (derivatives) as a language unit and identify their derivational bases with high precision. What is more, derivational relations, in contrast to morpho-syntactic word formation processes (e.g. different forms of nouns related to the grammatical cases or verb forms representing persons), signal a meaning change between a basis and the derivative. Such lexical meaning transformations are also predictive to a very large extent, e.g. *palarnia* ≈ 'a place for smoking' derived from *palić* 'to smoke'. Due to this property, such a class of derivational relations, described in lexico-semantic networks, is called *morphosemantic relations* (Fellbaum et al., 2007).

It is worth to notice that morphosemantic relations combine two transformations: one between word forms and, the second, in parallel, between lexical meanings, that are tightly coupled: different types of word form transformations are characteristic for some types of 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, ambiguous with respect to the meaning derivation. e.g., the suffix *-ka* mostly signals: a transformation from `+Male` → `+Female`, but it appears in tool name derivation, too: *wiercić* 'to drill' → *wiertarka* 'a driller', and can be also misleading: *pierwiastka* 'a woman 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 general objective 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 not only a set of words with which it is in a certain lexico-semantic relation, and also a word from which it has been morphologically derived – its derivational basis. We study machine learning means taking into account 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 (Dziob et al., 2019). Contrary to many other wordnets and deriva-

tional 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.

Derivational relations are often described in morphological dictionaries as links between lemmas[1] e.g. (Kanuparthi et al., 2012), (Šnajder, 2014) or a very large morphological and derivational network DeriNet (Vidra et al., 2019), only later automatically classified to 5 very coarse-grained semantic classes (Ševčíková and Kyjánek, 2019). In (Ševčíková and Kyjánek, 2019) 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 (Mititelu, 2012), BulNet (Mititelu, 2012; Dimitrova et al., 2014) or CroWN (Šojat and Srebačić, 2014). However, in wordnets, links between lemmas are additionally labelled with semantic relations, i.e. mapped onto morphosemantic relations. plWordNet (Dziob et al., 2019) 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 (Piasecki et al., 2012) two character-level transducers were built on the basis from training data (with post-pruning generalisation) and combined with internal stem alternations. 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 (Fellbaum, 1998) was utilised. We do not use such knowledge in our approach. In a similar approach (Koeva et al., 2016), 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 (Lango et al., 2018) and tested on Polish and Spanish. It was trained on "1500 pairs of base words with their 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 (Musil et al., 2019) for the Czech coarse-grained derivational relations. Neural character encoder–decoder was applied to predict a derivative from a derivational base in (Vylomova et al., 2017). It used occurrence context too, but was limited to deverbal nouns.

1.1 Contribution

Our main contribution is a method for recognition of morphosemantic relations and a comparison of several different representations of word forms in this task. The analysed method allows for detecting derivational relations between lexical units (word senses) in any wordnet as our method does not depend on any language-specific knowledge resource, except a training set of relation instances.

1.2 Data & Features

The data used in the experiments comes from the plWordNet[2] (Dziob et al., 2019) – precisely from the database dump from version 4.2. The dataset consists of samples represented as triples: a derivational base, a derivational relation and a derivative. Each triple originate from a morphosemantic, derivational, relation linking concrete lexical units (word senses), not lemmas, that have been manually edited and recently carefully manually verified by a separate team of lexicographers.

Statistics of the morphosemantic relations in plWordNet with respect to coarse and fine grained levels of classification is presented in Table 1. The acquired dataset consists of 134,201 triples, of which 77,122 are triples containing a single word lexical unit. The data has been divided into 5 equal numbered split folds. On the basis of the division into folds, five pairs of training and test sets were created. The training and test sets are lexically separable, what means in this case that the same derivational bases do not occur in both sets simultaneously. For the relation classification task, we

[1] Basic morphological word forms selected to represents sets of word forms that differ in the values of grammatical categories, but not meaning.

[2] http://plwordnet.pwr.edu.pl

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

Table 1: Relationships found in plWordNet at different granularities.

restricted the list of relations to those with a minimum of 150 examples in the dataset.

2 Embedding methods

In our experiments, we wanted to compare different methods for representing words (in fact lemmas) by vector spaces for the needs of recognition of semantic relations linking them, where all relations of interest are associated also with some relation between the word forms. First of all we used word embedding vectors, i.e. representation of words in dense spaces of real number vectors. Word embeddings were often used in recognition of lexical semantic relations. We conducted experiments with both context-free methods and those that use word context information (acquired during the learning process). We also tried to model words using vectors representing their character structure.

Concerning the latter, we call such a representation *Bag of Characters* (henceforth BoC). The vector for a word is simply constructed by counting the occurrences of different characters from the dictionary – i.e. simply letters of the Polish alphabet. Such a representation is an analogue of Bag of Words model used in Information Retrieval. It is relatively simple, but looses a lot of information related to object structures: documents and words in our case. It is known to be inferior in comparison to representations based on embeddings, so we expected it to be a kind of informative baseline.

A Bag of Characters vector of is easily interpretable in terms of its values, but unfortunately it is insensitive to the order of occurrence of the elements, i.e. character sequences that are very important in expressing derivational changes and morphems. Nevertheless, we wanted to check to what extent such an simplified representation is sufficient in representing derivational relations, which are characterised by relatively regular exchanges of characters in words. An example of such a vector is presented in Figure 1.

fastText (Bojanowski et al., 2017) is a word vectorisation model similar to *word2vec* (Mikolov et al., 2013), a kind of non-contextual word embedding model. The main difference is the use of orthographic representation in the vector creation process. The method learns the representation of character n-grams in text contexts and then constructs a vector of a given word as average of

	BoG Diff DT		BoG Diff RF		BoG Diff MLP		BoG 3-way DT		BoG 3-way RF		BoG 3-way MLP	
	Macro	**Weighted**	**Macro**	**Weighted**	**Macro**	**Weighted**	**Macro**	**Weighted**	**Macro**	**Weighted**	**Macro**	**Weighted**
Fold 0	0,60	0,83	0,60	0,83	0,58	0,83	0,56	0,80	0,57	0,82	0,59	0,83
Fold 1	0,61	0,83	0,61	0,83	0,61	0,83	0,56	0,80	0,59	0,82	0,59	0,82
Fold 2	0,60	0,82	0,60	0,83	0,60	0,83	0,56	0,79	0,59	0,82	0,57	0,82
Fold 3	0,60	0,82	0,61	0,82	0,60	0,82	0,55	0,79	0,57	0,81	0,58	0,82
Fold 4	0,60	0,83	0,61	0,83	0,59	0,82	0,56	0,79	0,59	0,82	0,59	0,82
Avg	***0,602***	***0,826***	***0,606***	***0,828***	***0,596***	***0,826***	***0,558***	***0,794***	***0,582***	***0,818***	***0,584***	***0,822***
St. dev	***0,004***	***0,005***	***0,005***	***0,004***	***0,011***	***0,005***	***0,004***	***0,005***	***0,011***	***0,004***	***0,009***	***0,004***
	FT 100 Diff		**FT 100 3-way**		**FT 300 Diff**		**FT 300 3-way**		**COMB 100**		**COMB 300**	
	Macro	**Weighted**	**Macro**	**Weighted**	**Macro**	**Weighted**	**Macro**	**Weighted**	**Macro**	**Weighted**	**Macro**	**Weighted**
Fold 0	0,58	0,82	0,61	0,83	0,60	0,83	0,63	0,85	0,61	0,83	0,62	0,84
Fold 1	0,57	0,81	0,61	0,83	0,60	0,83	0,63	0,84	0,60	0,83	0,64	0,85
Fold 2	0,57	0,81	0,61	0,83	0,61	0,83	0,62	0,84	0,60	0,83	0,64	0,84
Fold 3	0,59	0,82	0,61	0,83	0,59	0,82	0,64	0,84	0,60	0,83	0,63	0,84
Fold 4	0,57	0,82	0,61	0,83	0,61	0,83	0,61	0,84	0,61	0,83	0,62	0,84
Avg	***0,576***	***0,816***	***0,610***	***0,830***	***0,602***	***0,828***	***0,626***	***0,842***	***0,604***	***0,830***	***0,630***	***0,842***
St. dev	***0,009***	***0,005***	***0,000***	***0,000***	***0,008***	***0,004***	***0,011***	***0,004***	***0,005***	***0,000***	***0,010***	***0,004***

Table 2: Experimental results on the classifier. F-1 score measure. *DT – Decision Tree; RF – Random Forest; MLP – Multi Layer Perceptron; FT – fastText; COMB – Combination of FT and BoG vectors*

Word form: kotek *(ang. little cat)*

a	...	e	...	k	l	m	n	o	...	t	...	z
0	...	1	...	2	0	0	0	1	...	1	...	0

Figure 1: Example of bag-of-character vector for word "kotek".

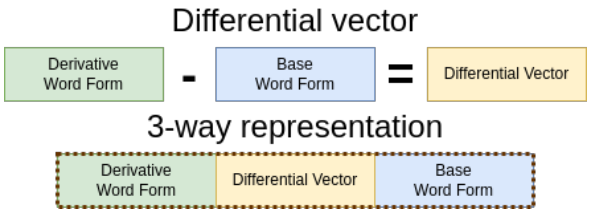

Figure 2: (Top) The way the differential vector is formed. (Bottom) The vector in the second phase of the experiments is formed by concatenating three basis vectors.

representations of the n-grams that constitute it. This process of building vectors goes around the problem of out-of-vocabulary words. The *fastText* based representation showed improvement in several NLP tasks in relation to inflectional languages, e.g. syntactic tasks relative to traditional *word2vec*, but also text classification and recognition of semantic relations.

3 Classification experiments

In order to compare the effectiveness of using different vector representations for the task of classifying derivational relations, we first used all vector versions to train a multi-class classifier based on an MLP neural network, as a classification model that seem to be in good balance between expressiveness and requirements for the size of a data set that is limited in our case (e.g. especially coverage for different relation types). We used the package default settings during learning the classifier, because our main focus was on different vector representations of examples.

Since Bag of Characters vectors are discrete in nature and their singular values are interpretable, we also decided to train classifiers using directly this representation, i.e. Decision Trees, both a single tree method and a Random Forest approach. In our experiments, we followed a multi-class classifier scheme. Each example in the training and test data subsets is an instance of a derivational relation (i.e. a pair of lemmas: a derivational basis and a derivative) so in the experiments we did not assume the possibility of labelling a pair with the label 'no relation'.

We examined each prepared vector representation in the following configurations:

1. differential vector of the derivation form and the base form;

2. concatenated vectors of a derivational form, a base form and a differential vector.

We called this vector a 3-way vector. This is shown in Figure 2. The 3-way representation was shown to be effective in recognition of wordnet relations, especially in combination with fastText representation, e.g. (Czachor et al., 2018). It is meant to represent semantic characteristic of both elements, but also to emphasise differences between them, together with the directions of the differences. The

directions are potentially important for plWordNet morphosemantic relations, as they are all defined and edited in the direction from a derivational basis to the derivate (the derived word).

For the final experiment, we also analysed combination of the two different representations. Whole words were embedded using *fastText* vectors and concatenated together with a difference vector obtained using the Bag-of-Characters technique. The aim of this experiment was to test whether combining a semantic representation based on word vectors and a discrete representation associated with an orthographic form would result in an improvement in the classification task.

We implemented the classifier models for all experiments using the *scikit-learn* library (Pedregosa et al., 2011).

3.1 Results

The obtained results are shown in Table 2. All experiments yielded approximately the same results – the differences are statistically non-significant – regardless of the representation method applied. These results are quite surprising in two aspects: lack of superiority of semantically-informed representation based on fastText and no preference for MLP representation.

Classifiers from the tree family, did not differ much in their results with respect to the neural network classifier, which may also suggest saturation of the problem rather than a specific classification method. Only increasing the size of the fastText vector improved the measure by ~1.5 percentage points in 3-way representation case. This can be also an effect of learning the association of some relation types with specific semantic dimensions. However, it is worth to emphasise that we applied a technique of lexical split in selecting folds, i.e. the same words were not selected for both the training and test subsets (needless to say that relations instances are obviously not repeating between both subsets). Such a split is known to prevent a classifier for memorising prototypes for relation instances. Such conformity of the classifier may indicate that a limit with respect to the efficiency of the method has been reached, which will not be exceeded without changing the assumptions of the problem.

A major limiting factor for further progress, we suggest, is the scheme in which the classification is performed out of use context. In tasks where semantics matter (for example WSD, NER) context is a strong stimulus for classification methods. Moreover, most of the lemmas we are working here with – relations link lexical units (word senses), but representations are built for lemmas – are polysemous. What is worse, in some number of cases a given morphosemantic relations links only selected lexical units from lemmas, depending on the meaning of these lexical units. It is also worth to notice that a representation based on word embeddings is a not only a mixture of several lexical meaning per a word, but also only more salient meanings dominates in it and less frequent meanings are often hard to trace in a vector. Thus, when we work with ambiguous, lemma-based representations that make the picture very blurred from the point of view of classifiers. In this task of recognition of morphosemantic relations, we need a shift in paradigm from context-less into analysing representations of lexical units in their use contexts, in order to make further progress. The task must be somehow combined with Word Sense Disambiguation and Word Sense Induction.

4 Conclusions

Our research has shown that the limit of context-free classification of derivational relations lies not in the representation of examples, but in the absence of any other source of information for the classifier. In the final version of the system for context-free classification of derivational relations, we decided to stay with Bag of Characters vectors, due to their simple human interpretability. We want to direct our further research to the study of derivation in the context of – both the preparation of datasets (such as a corpus) and methods for detecting and classifying relations.

References

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Gabriela Czachor, Maciej Piasecki, and Arkadiusz Janz. 2018. Recognition of lexico-semantic relations in word embeddings for polish. In *Proceedings of the 9th Global Wordnet Conference, Singapore, 8-12 January 2018*. Global WordNet Association.

Tsvetana Dimitrova, Ekaterina Tarpomanova, and Borislav Rizov. 2014. Coping with derivation in the bulgarian wordnet. In *Proceedings of the Seventh*

Global Wordnet Conference, pages 109–117. University of Tartu Press.

Agnieszka Dziob, Maciej Piasecki, and Ewa Rudnicka. 2019. plWordNet 4.1 - a linguistically motivated, corpus-based bilingual resource. In *Proceedings of the 10th Global Wordnet Conference*, pages 353–362, Wroclaw, Poland. Global Wordnet Association.

Christiane Fellbaum, editor. 1998. *WordNet – An Electronic Lexical Database*. The MIT Press.

Christiane Fellbaum, Anne Osherson, and Peter E. Clark. 2007. Putting semantics into wordnet's "morphosemantic" links. In *Human Language Technology. Challenges of the Information Society, Third Language and Technology Conference, LTC 2007, Poznan, Poland, October 5-7, 2007, Revised Selected Papers*, volume 5603 of *Lecture Notes in Computer Science*, pages 350–358. Springer.

Nikhil Kanuparthi, Abhilash Inumella, and Dipti Misra Sharma. 2012. Hindi derivational morphological analyzer. In *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology*, pages 10–16. Association for Computational Linguistics.

Svetla Koeva, Svetlozara Leseva, Ivelina Stoyanova, Tsvetana Dimitrova, and Maria Todorova. 2016. Automatic prediction of morphosemantic relations. In *Proceedings of the 8th Global WordNet Conference (GWC)*, pages 169–177. Global Wordnet Association.

Mateusz Lango, Magda Ševčíková, and Zdeněk Žabokrtský. 2018. 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)*. European Language Resources Association (ELRA).

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Verginica Barbu Mititelu. 2012. Adding morpho-semantic relations to the romanian wordnet. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2596–2601. European Language Resources Association (ELRA).

Tomáš Musil, Jonáš Vidra, and David Mareček. 2019. Derivational morphological relations in word embeddings. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 173–180. Association for Computional Linguistics.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Maciej Piasecki, Radosław 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. European Language Resources Association (ELRA).

Jonáš Vidra, Zdeněk Žabokrtský, Magda Ševčíková, and Lukáš Kyjánek. 2019. 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*, pages 81–89, Prague, Czechia. Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics.

Ekaterina Vylomova, Ryan Cotterell, Timothy Baldwin, and Trevor Cohn. 2017. Context-aware prediction of derivational word-forms. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, volume 2, pages 118–124. Association for Computational Linguistics.

Magda Ševčíková and Lukáš Kyjánek. 2019. Introducing semantic labels into the derinet network. *Journal of Linguistics/Jazykovedný casopis*, 70(2):412–423.

Jan Šnajder. 2014. Derivbase.hr: A high-coverage derivational morphology resource for croatian. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3371–3377. European Language Resources Association (ELRA).

Krešimir Šojat and Matea Srebačić. 2014. Morphosemantic relations between verbs in croatian wordnet. In *Proceedings of the Seventh Global Wordnet Conference*, pages 262–267. University of Tartu Press.