

The Impact of Word Representations on Sequential Neural MWE Identification

Nicolas Zampieri

Aix Marseille Univ,
Université de Toulon, CNRS,
LIS, Marseille, France
first.last@etu.univ-amu.fr

Carlos Ramisch

Aix Marseille Univ,
Université de Toulon, CNRS,
LIS, Marseille, France
first.last@lis-lab.fr

Géraldine Damnati

Orange Labs
Lannion, France
first.last@orange.com

Abstract

Recent initiatives such as the PARSEME shared task have allowed the rapid development of MWE identification systems. Many of those are based on recent NLP advances, using neural sequence models that take continuous word representations as input. We study two related questions in neural verbal MWE identification: (a) the use of lemmas and/or surface forms as input features, and (b) the use of word-based or character-based embeddings to represent them. Our experiments on Basque, French, and Polish show that character-based representations yield systematically better results than word-based ones. In some cases, character-based representations of surface forms can be used as a proxy for lemmas, depending on the morphological complexity of the language.

1 Introduction

MWE identification consists in finding multiword expressions (MWEs) in running text (Constant et al., 2017). For many years, MWE identification was considered unrealistic, with most MWE research focusing on out-of-context MWE discovery (Ramisch et al., 2013). Indeed, the availability of MWE-annotated corpora was limited to some treebanks with partial annotations, often a by-product of syntax trees (Green et al., 2013; Constant et al., 2013). This prevented the widespread development and evaluation of MWE identification systems, as compared to other tasks such as POS tagging and named entity recognition.

This landscape has drastically changed in the last few years, thanks to shared tasks such as DiMSUM (Schneider et al., 2016) and PARSEME 1.0 and 1.1 (Savary et al., 2017; Ramisch et al., 2018) and to the release of open corpora annotated for MWEs in ~ 20 languages. These initiatives provide a unified framework for MWE identifica-

tion, including training/test corpus splits, evaluation metrics, benchmark results, and analysis tools. As a consequence, it is now possible to study some classical text processing problems and their impact on MWE identification systems.

One of these problems is the relation between a language’s morphology, lemmatisation, input feature representations, out-of-vocabulary (OOV) words, and the performance of the system. For instance, an MWE identification system based on (inflected) surface forms will likely encounter more OOV words than a system based on lemmas, especially for morphologically-rich languages in which a single lemma may correspond to dozens of surface forms (Seddah et al., 2013). This problem is particularly relevant for verbal MWEs, which present high morphological and syntactic variability (Savary et al., 2018).

Our goal is to study the impact of word representations on verbal MWE (VMWE) identification, comparing lemmas, surface forms, traditional word embeddings and subword representations. We compare the performance of an off-the-shelf MWE identification system based on neural sequence tagging (Zampieri et al., 2018) using lemmas and surface forms as input features, encoded in the form of classical pre-initialised word2vec embeddings (Mikolov et al., 2013) or, alternatively, using new-generation FastText embeddings built from character n-grams (Bojanowski et al., 2017). Our main hypothesis is that the latter can model morphological variability, representing an alternative for lemmatisation. We carry out experiments in 3 languages with varying morphological complexity: French, Polish and Basque.

2 Related Work

Rule-based matching, supervised classification, sequence tagging, and parsing are among the most

popular models for MWE identification (Constant et al., 2017). Parsing-based methods take the (recursive) structure of language into account, trying to identify MWEs as a by-product of parsing (Green et al., 2013; Constant et al., 2013), or jointly (Constant and Nivre, 2016). Sequence tagging models, on the other hand, consider only linear context, using models such as CRFs (Vincze et al., 2011; Shigeto et al., 2013; Riedl and Biemann, 2016) and averaged perceptron (Schneider et al., 2014) combined with some variant of begin-inside-outside (BIO) encoding (Ramshaw and Marcus, 1995).

Recurrent neural networks can be used for sequence tagging, being able to handle continuous word representations and unlimited context. The first neural identification system was MUMULS, submitted to the PARSEME shared task 1.0 (Klyueva et al., 2017). Although it did not obtain the best results, MUMULS influenced the development of more advanced models (Gharbieh et al., 2017) which ultimately led to the popularisation of the approach. As a consequence, and inspired by the success of neural models in NLP, nine out of the 17 systems submitted to the PARSEME shared task 1.1 used neural networks (Ramisch et al., 2018). Recently, improvements have been proposed, e.g. to deal with discontinuous MWEs (Rohanian et al., 2019).

Previous work studied the impact of external lexicons (Riedl and Biemann, 2016) and of several feature sets (Maldonado et al., 2017) on CRFs for MWE identification. Character-based embeddings have been shown useful to predict MWE compositionality out of context (Hakimi Parizi and Cook, 2018). In other tasks such as named entity recognition, character convolution layers have been successfully applied (Ma and Hovy, 2016). The use of pre-trained vs. randomly initialised embeddings has been analysed in some PARSEME shared task papers (Ehren et al., 2018; Zampieri et al., 2018). The closest works to ours are the Veyn (Zampieri et al., 2018) and SHOMA (Taslimipoor and Rohanian, 2018) systems, submitted to the PARSEME shared task 1.1. Veyn is used as our off-the-shelf base system, so most of its architecture is identical to ours. Similarly to us, SHOMA employs FastText embeddings, a recurrent layer and a CRF output layer. To our knowledge, however, this is the first study to compare input representations for neural MWE identification.

3 Experimental Setup

Corpora The PARSEME shared task 1.1 released freely available VMWE-annotated corpora in 20 languages.¹ Each language’s corpus is split into training, development and test parts. To choose our target languages, we analysed the PARSEME corpora, choosing 3 languages with varying morphological richness: Basque (EU), French (FR) and Polish (PL), shown in Table 1.² The FR training corpus has more than 420K tokens, whereas the PL and EU training corpora have around 220K and 117K tokens. EU contains less annotated VMWE occurrences than both FR and PL. The average length of annotated VMWE occurrences is similar in the three languages (2.02/2.29/2.13 in EU/FR/PL). The proportion of discontinuous VMWEs is highest in FR (42.12%), whereas in Polish (29.76%) and in Basque (19.28%) they are less frequent. These languages do have not the same morphological richness, as measured by the average number of surface forms per lemma in the vocabulary (‘Morph’ column). For instance, the EU training corpus (2.32) has a higher morphological richness than PL (2.21) and FR (1.33). The rate of OOVs, that is, of words that appear in the dev or test corpus vocabularies, but not in the training corpus, is higher for surface forms than for lemmas, with a potential negative impact on VMWE identification systems based on surface forms only. As expected, the OOV rate for surface forms is lowest in FR (20-26%), which also has the lowest morphological richness, and highest for EU (43%). These differences are less visible for lemmas, which abstract away from morphology.³ An interesting figure is the OOV rate focusing on verbs only.⁴ Here, PL presents more OOV verb forms (42-44%) than EU (32%), but again this difference disappears for lemmas. This is relevant because our experimen-

¹<http://hdl.handle.net/11372/LRT-2842>

²Other languages have similar characteristics but were not selected due to the size of the corpora or to incomplete information (e.g. Turkish has missing surface forms for some verbs, preventing us from training a system based on surface forms only).

³The official PARSEME French test corpus presents 11,632 missing lemmas. We have lemmatised it using UD-Pipe (<http://ufal.mff.cuni.cz/udpipe>) with default parameters, trained on the PARSEME shared task training corpus, to remain in the “closed track” conditions.

⁴For EU, we consider the POS tags VERB, ADI and ADT according to the conversion table <https://universaldependencies.org/tagset-conversion/eu-conll-uposf.html>

	Tokens	VMWEs	Vocabulary		Morph	OOVs-Vocabulary		OOVs-Verbs	
			forms	lemmas		forms	lemmas	forms	lemmas
EU-train	117,165	2,823	26,912	11,602	2.32	—	—	—	—
EU-dev	21,604	500	7,766	4,178	1.86	43% (3,365)	29% (1,225)	32% (454)	18% (93)
EU-test	19,038	500	7,226	3,902	1.85	43% (3,085)	28% (1,080)	32% (448)	15% (73)
FR-train	420,762	4,550	45,166	33,928	1.33	—	—	—	—
FR-dev	54,685	629	11,593	8,814	1.32	26% (3,032)	27% (2,383)	23% (550)	12% (126)
FR-test	38,402	498	8,160	6,052	1.35	20% (1,666)	19% (1,172)	23% (441)	16% (144)
PL-train	220,352	4,122	48,211	21,795	2.21	—	—	—	—
PL-dev	26,014	515	10,007	5,955	1.68	34% (3,452)	19% (1,136)	42% (1,047)	14% (180)
PL-test	27,661	515	10,285	6,408	1.61	40% (4,145)	25% (1,605)	44% (825)	16% (177)

Table 1: Description of the training (train), development (dev), and test corpora for Basque (EU), French (FR), and Polish (PL). The number of tokens excludes ranges (multiword tokens). ‘VMWEs’ denotes the number of verbal MWEs. The ‘Vocabulary’ column shows the number of types in the vocabulary of surface forms and lemmas. The ‘Morph’ column indicates the morphological richness: ratio between the number of forms and lemmas in the vocabulary. The rate and number (in parentheses) of OOVs is given for the whole vocabulary, and for verbs only.

tal setup implies that it is difficult for a system to predict a VMWE without a reliable representation for a verb, learned from the training data.

MWE Identification System We use our in-house MWE identification system Veyn (Zampieri et al., 2018), based on sequence tagging using recurrent neural networks.⁵ The system takes as input the concatenation of the embeddings of the words’ features (e.g. lemmas and POS). It uses a CRF output layer (conditional random fields) to predict valid label sequences, with VMWEs encoded using the ‘BIOG+cat’ format. Each token is tagged ‘B’ if it is at the beginning of a VMWE, ‘I’ if it is inside a VMWE, ‘O’ if it does not belong to a VMWE, and ‘G’, if it does not belong to a VMWE but it is in the gap between two words that are part of a VMWE. The tags ‘B’ and ‘I’ are concatenated with the VMWE categories (VID, LVC.full, etc.) present in the corpus. The system is trained on the shared task training corpora, so that the results are comparable with the systems submitted to the closed track.⁶ We use the dev corpus as validation data, training for 25 epochs which 3 epochs of patience for early stopping. We configure it to use two layers of bidirectional gated recurrent units (GRU) of dimension 128, with all other parameters taking the default values suggested in the Veyn documentation.

Word Representations We use two types of word embeddings to represent input surface forms

⁵<https://github.com/zamp13/Veyn>

⁶<http://multiword.sourceforge.net/sharedtaskresults2018>

and lemmas: word2vec and FastText. Word2vec is a prediction-based distributional model in which a word representation is obtained from a neural network trying to predict a word from its context or vice-versa (Mikolov et al., 2013). FastText is an adaptation which also takes into account character n-grams, being able to build vectors for OOVs from its character n-grams (Bojanowski et al., 2017). For each representation, we used the gensim library⁷ to train 256-dimensional vectors for both forms and lemmas on the training corpus of the shared task for 10 epochs. Furthermore, all embeddings use the CBOW algorithm with the same hyper-parameter values of 5 for the window size (left/right context of words) and 1 for min-count (minimum number of occurrences of words). For FastText, we set the size of character n-grams to 1 to combine the whole word’s embedding with the embeddings of its characters. We did not use contextual representations, like BERT, Elmo or Flair (Devlin et al., 2018; Peters et al., 2018; Akbik et al., 2018), because they have to be pre-trained on large corpora and we wanted to have an experimental setup compatible with the closed track of the PARSEME shared task.

Evaluation Measures We adopt the metrics proposed in the PARSEME shared tasks (Savary et al., 2017). The MWE-based measure (F-MWE) is the F1 score for fully predicted VMWEs, whereas the token-based measure (F-TOK) is the F1 score for tokens belonging to a VMWE.

⁷<https://radimrehurek.com/gensim/>

Features	Embeddings	EU		FR		PL	
		F-MWE	F-Tok	F-MWE	F-Tok	F-MWE	F-Tok
form	word2vec	60.37	70.93	47.41	56.64	42.27	58.23
form	FastText	66.52	72.36	52.60	63.47	47.24	56.08
lemma	word2vec	53.36	65.37	53.28	63.76	57.82	65.48
lemma	FastText	62.86	68.79	59.35	68.60	61.49	63.98
form-lemma	word2vec	60.56	73.07	56.11	66.31	56.80	67.16
form-lemma	FastText	69.24	74.01	60.41	68.39	57.39	64.63

Table 2: MWE-based F-measure (F-MWE) and token-based F-measures (F-TOK) of the models on the test corpus, using word2vec and FastText word representations for different feature sets: lemmas, surface forms, and both.

4 Results

We train Veyn using UPOS tags as input features, combined with word2vec and FastText embeddings for lemmas, surface forms, or both.⁸ Performances are given on the PARSEME test corpus for Basque (EU), French (FR) and Polish (PL). On one hand, we compare performances with FastText and word2vec representations, and on the other hand, we compare performances with various input feature sets (Table 2).

Impact of Word Vector Representation For French and Basque, the use of FastText outperforms word2vec by a large margin on both F-measures with any input feature set. For Basque, the best input features and subword embeddings get a score of 69.24 (74.01) in F-MWE (F-TOK), while they get a score of 60.56 (73.07) in F-MWE (F-TOK) with word2vec representations. Similar results are obtained for French, with FastText corresponding to the best choice for both metrics. Results for PL are more contrasted: word2vec representation yields the best results for the F-TOK metric (67.16 against 64.63 for FastText) but results are better with FastText in terms of F-MWE scores (61.29 against 57.82 with word representation). This suggests that the word2vec model has difficulties in predicting MWE boundaries, but predicts correct parts of VMWEs more often than with FastText. Looking into the details of the system’s output, we observed that the system with word2vec predicts more MWEs (540 predictions against 365 predictions with FastText). These predictions include a large amount of single-token VMWEs (22% with word2vec against 5% with FastText), but the training and development corpora have no single-token VMWE in Polish. For example, for the verbal idiom expression *będzie*

się w stanie, the system with FastText makes no prediction whereas with word2vec the prediction is *będzie się w stanie* where the reflexive clitic *się* is tagged as being a single-token inherently reflexive verb and *w stanie* is predicted as a verbal idiom. More single token predictions increase the recall of F-TOK, but decrease the precision of F-MWE. Further investigation will be made to understand this phenomenon, which could be compensated by simple post-processing, e.g. grouping single-token predictions with adjacent ones. We hypothesise that the system with subword representation is able to take the morphological inflection into account. For example, the French expression *faire référence* ‘to make reference’ is seen in this form in the training corpus, but the test corpus contains a different inflection of the verb *fait référence* ‘makes reference’. For this example, with FastText representation the system is able to find the expression, but with word2vec representation the system can not find it if we use surface form and lemma at input.

Impact of Input Pre-processing For Basque, which has a high morphological richness, the model with the richest information provides the best results. Performances are maximised with the form-lemma model, providing an F-MWE score of 69.24, while the form model yields a 66.52 score and the lemma model gives 62.86, suggesting that relevant information for VMWE identification is lost in lemmatisation. For Polish, similar results are obtained in terms of F-Tok while F-MWE is maximised for the lemma configuration with FastText. This is also a consequence of the phenomenon described in the previous subsection where single-token expressions are predicted for Polish. The lemma configuration is less affected by this phenomenon (F-TOK is lower) and thus full-expression identification is more effec-

⁸Other features (e.g. morphology and syntax) are ignored.

tive (higher F-MWE of 61.49). Results on French corroborate this trend: although French has simpler morphology, lemmas are still important to obtain best results. As opposed to highly morphological languages like Basque, the combination of lemmas and forms for French does not yield as much improvement. Performances in terms of F-TOK are equivalent for lemma and form-lemma and are slightly better in terms of F-MWE.

For the three languages under consideration, our best models would have ranked in the top-3 in the closed track of the official shared task results.

5 Conclusions and Future Work

We have studied the impact of word representations on VMWE identification for Basque, French and Polish, comparing lemmas and surface forms as input features and comparing traditional word embeddings (word2vec) and subword representations (FastText). Regarding the latter, subword representations proved to be efficient for our task. For the former, we have highlighted that the use of lemmas always have a positive impact. For languages with high morphological richness, the combination of lemmas and forms has an even higher impact, especially for Basque. Considering the high Out-of-Vocabulary rate, including for verbs, we intend to improve OOV handling in the future. The use of recent embeddings such as BERT, Elmo and Flair, trained on large external corpora, could help with OOVs.

References

- Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. [Contextual string embeddings for sequence labeling](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1638–1649, Santa Fe, NM, USA. Association for Computational Linguistics.
- 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.
- Mathieu Constant, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner, and Amalia Todirascu. 2017. [Multiword expression processing: A survey](#). *Computational Linguistics*, 43(4):837–892.
- Mathieu Constant, Joseph Le Roux, and Anthony Sigogne. 2013. [Combining compound recognition and PCFG-LA parsing with word lattices and conditional random fields](#). *ACM TSLP Special Issue on MWEs*, 10(3):1–24.
- Mathieu Constant and Joakim Nivre. 2016. [A transition-based system for joint lexical and syntactic analysis](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 161–171, Berlin, Germany. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. [Bert: Pre-training of deep bidirectional transformers for language understanding](#). *arXiv preprint arXiv:1810.04805*.
- Rafael Ehren, Timm Lichte, and Younes Samih. 2018. [Mumpitz at PARSEME shared task 2018: A bidirectional LSTM for the identification of verbal multiword expressions](#). In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 261–267, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Waseem Gharbieh, Virendrakumar Bhavsar, and Paul Cook. 2017. [Deep learning models for multiword expression identification](#). In *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (*SEM 2017)*, pages 54–64, Vancouver, Canada. Association for Computational Linguistics.
- Spence Green, Marie-Catherine de Marneffe, and Christopher D. Manning. 2013. [Parsing models for identifying multiword expressions](#). *Computational Linguistics*, 39(1):195–227.
- Ali Hakimi Parizi and Paul Cook. 2018. [Do character-level neural network language models capture knowledge of multiword expression compositionality?](#) In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 185–192, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Natalia Klyueva, Antoine Doucet, and Milan Straka. 2017. [Neural networks for multi-word expression detection](#). In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 60–65, Valencia, Spain. Association for Computational Linguistics.
- Xuezhe Ma and Eduard Hovy. 2016. [End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1064–1074, Berlin, Germany. Association for Computational Linguistics.
- Alfredo Maldonado, Lifeng Han, Erwan Moreau, Ashjan Alsulaimani, Koel Dutta Chowdhury, Carl Vogel, and Qun Liu. 2017. [Detection of verbal multiword expressions via conditional random fields](#)

- with syntactic dependency features and semantic re-ranking. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 114–120, Valencia, Spain. Association for Computational Linguistics.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. [Efficient estimation of word representations in vector space](#). In *ICLR Workshop Papers*.
- Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep contextualized word representations](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, LA, USA. Association for Computational Linguistics.
- Carlos Ramisch, Silvio Ricardo Cordeiro, Agata Savary, Veronika Vincze, Verginica Barbu Mititelu, Archana Bhatia, Maja Buljan, Marie Candito, Polona Gantar, Voula Giouli, Tunga Güngör, Abdelati Hawwari, Uxoá Iñurrieta, Jolanta Kovalevskaitė, Simon Krek, Timm Lichte, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Behrang QasemiZadeh, Renata Ramisch, Nathan Schneider, Ivelina Stoyanova, Ashwini Vaidya, and Abigail Walsh. 2018. [Edition 1.1 of the parseme shared task on automatic identification of verbal multiword expressions](#). In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 222–240, Santa Fe, NM, USA. Association for Computational Linguistics.
- Carlos Ramisch, Aline Villavicencio, and Valia Kordoni, editors. 2013. *ACM TSLP Special Issue on MWEs*, volume 10. ACM, New York, NY, USA.
- Lance Ramshaw and Mitch Marcus. 1995. [Text chunking using transformation-based learning](#). In *Proceedings of the Third Workshop on Very Large Corpora*, pages 82–94, Cambridge, MA, USA.
- Martin Riedl and Chris Biemann. 2016. [Impact of MWE resources on multiword recognition](#). In *Proc. of the ACL 2016 Workshop on MWEs*, pages 107–111, Berlin, Germany.
- Omid Rohanian, Shiva Taslimipoor, Samaneh Kouchaki, Le An Ha, and Ruslan Mitkov. 2019. [Bridging the gap: Attending to discontinuity in identification of multiword expressions](#). *CoRR*, abs/1902.10667.
- Agata Savary, Marie Candito, Verginica Barbu Mititelu, Eduard Bejček, Fabienne Cap, Slavomír Čéplö, Silvio Ricardo Cordeiro, Gülşen Eryiğit, Voula Giouli, Maarten van Gompel, Yaakov HaCohen-Kerner, Jolanta Kovalevskaitė, Simon Krek, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Lonneke van der Plas, Behrang QasemiZadeh, Carlos Ramisch, Federico Sangati, Ivelina Stoyanova, and Veronika Vincze. 2018. [PARSEME multi-lingual corpus of verbal multiword expressions](#). In Stella Markantonatou, Carlos Ramisch, Agata Savary, and Veronika Vincze, editors, *Multiword expressions at length and in depth. Extended papers from the MWE 2017 workshop*, pages 87–147. Language Science Press, Berlin, Germany.
- Agata Savary, Carlos Ramisch, Silvio Cordeiro, Federico Sangati, Veronika Vincze, Behrang QasemiZadeh, Marie Candito, Fabienne Cap, Voula Giouli, Ivelina Stoyanova, and Antoine Doucet. 2017. [The PARSEME shared task on automatic identification of verbal multiword expressions](#). In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 31–47, Valencia, Spain. Association for Computational Linguistics.
- Nathan Schneider, Emily Danchik, Chris Dyer, and Noah A. Smith. 2014. [Discriminative lexical semantic segmentation with gaps: running the MWE gamut](#). *Transactions of the Association for Computational Linguistics*, 2:193–206.
- Nathan Schneider, Dirk Hovy, Anders Johannsen, and Marine Carpuat. 2016. [SemEval-2016 task 10: Detecting minimal semantic units and their meanings \(DiMSUM\)](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 546–559, San Diego, CA, USA. Association for Computational Linguistics.
- Djamé Seddah, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho Choi, Richárd Farkas, Jennifer Foster, Iakes Goenaga, Koldo Gojenola, Yoav Goldberg, Spence Green, Nizar Habash, Marco Kuhlmann, Wolfgang Maier, Joakim Nivre, Adam Przepiorkowski, Ryan Roth, Wolfgang Seeker, Yannick Versley, Veronika Vincze, Marcin Woliński, Alina Wróblewska, and Eric Villemonte de la Clergerie. 2013. [Overview of the SPMRL 2013 shared task: A cross-framework evaluation of parsing morphologically rich languages](#). In *Proc. of SPRML 2013*, pages 146–182, Seattle, WA, USA.
- Yutaro Shigeto, Ai Azuma, Sorami Hisamoto, Shuhei Kondo, Tomoya Kose, Keisuke Sakaguchi, Akifumi Yoshimoto, Frances Yung, and Yuji Matsumoto. 2013. [Construction of English MWE dictionary and its application to POS tagging](#). In *Proceedings of the 9th Workshop on Multiword Expressions*, pages 139–144, Atlanta, Georgia, USA. Association for Computational Linguistics.
- Shiva Taslimipoor and Omid Rohanian. 2018. [SHOMA at parseme shared task on automatic identification of vmwes: Neural multiword expression tagging with high generalisation](#). *CoRR*, abs/1809.03056.
- Veronika Vincze, István Nagy T., and Gábor Berend. 2011. [Detecting noun compounds and light verb constructions: a contrastive study](#). In *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, pages

116–121, Portland, OR, USA. Association for Computational Linguistics.

Nicolas Zampieri, Manon Scholivet, Carlos Ramisch, and Benoit Favre. 2018. [Veyn at PARSEME shared task 2018: Recurrent neural networks for VMWE identification](#). In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 290–296, Santa Fe, NM, USA. ACL.