Romanian Multiword Expression Detection Using Multilingual Adversarial Training and Lateral Inhibition

Andrei-Marius Avram[1], **Verginica Barbu Mititelu**[2], **Dumitru-Clementin Cercel**[1]

[1]University Politehnica of Bucharest, Faculty of Automatic Control and Computers
[2]Romanian Academy Research Institute for Artificial Intelligence
andrei_marius.avram@stud.acs.upb.ro, vergi@racai.ro, dumitru.cercel@upb.ro

Abstract

Multiword expressions are a key ingredient for developing large-scale and linguistically sound natural language processing technology. This paper describes our improvements in automatically identifying Romanian multiword expressions on the corpus released for the PARSEME v1.2 shared task. Our approach assumes a multilingual perspective based on the recently introduced lateral inhibition layer and adversarial training to boost the performance of the employed multilingual language models. With the help of these two methods, we improve the F1-score of XLM-RoBERTa by approximately 2.7% on unseen multiword expressions, the main task of the PARSEME 1.2 edition. In addition, our results can be considered SOTA performance, as they outperform the previous results on Romanian obtained by the participants in this competition.

1 Introduction

The correct identification and handling of multi-word expressions (MWEs) are important for various natural language processing (NLP) applications, such as machine translation, text classification, or information retrieval. For example, in machine translation, if an MWE is not recognized as such and is literally translated rather than as an expression, the resulting translation either is confusing or has the wrong meaning (Zaninello and Birch, 2020). In text classification, MWEs recognition can provide important information about the topic or sentiment of a text (Catone et al., 2019), while in information retrieval, MWEs can clarify the meaning of a query and improve the accuracy of search results (Englmeier and Contreras, 2021).

The PARSEME COST Action[1] organized three editions Savary et al. (2017); Ramisch et al. (2018, 2020) of a shared task that aimed at improving the identification of verbal MWEs (VMWEs) in

text. This work improves the results obtained in PARSEME 1.2 (Ramisch et al., 2020) for the Romanian language. We investigate the advantages of using Romanian monolingual Transformer-based (Vaswani et al., 2017) language models together with merging all the datasets for each language presented at the competition in a single corpus and then fine-tuning several multilingual language models on it. Additionally, for the latter, we aim to enhance the overall system's performance by generating language-independent features, with the help of two techniques, namely the lateral inhibition layer (Păiș, 2022) on top of the language models and adversarial training (Lowd and Meek, 2005) between languages.

Our experiments show that by employing these two algorithms, the results of the cross-lingual robustly optimized BERT approach (XLM-RoBERTa) (Conneau et al., 2020) improve by 2.7% on unseen MWEs when trained on the combined dataset. Additionally, we report state-of-the-art (SOTA) results with the monolingual training of Romanian Bidirectional Encoder Representations from Transformer (RoBERT) (Dumitrescu et al., 2020) in comparison with the results obtained at the PARSEME 1.2 edition, achieving an F1-score of 60.46%, an improvement of over 20%.

2 Dataset

The PARSEME multilingual corpus was annotated with several types of VMWEs, to serve as training and testing material for the shared task. The quality of the manual annotation was further enhanced by a semi-automatic way of ensuring annotation consistency. For edition 1.2, the corpus contained 14 languages: Basque, Chinese, French, German, Hebrew, Hindi, Irish, Italian, Modern Greek, Polish, Portuguese, Romanian, Swedish, and Turkish.

The types of VMWEs (i.e., universal, quasi-universal, and language-specific types) annotated

---

[1]https://typo.uni-konstanz.de/parseme/.

therein are described in the annotation guidelines[2]. The types of VMWEs annotated for Romanian are as follows: VID (verbal idiom) like "fura somnul" (eng., "steal sleep-the", "fall asleep"), LVC.full (light verb construction with a semantically bleached verb) like "da citire" (eng., "give reading", "read"), LVC.cause (light verb construction in which the verb has a causative meaning) like "da foc" (eng., "give fire", "put on fire"), and IRV (inherently reflexive verb) like "se gândi" (eng., "Refl.Cl. think", "think").

The whole corpus version 1.2 contains 5.5 million tokens with 68k VMWEs annotations, split into train, dev, and test sets, on the one hand for controlling the distribution of unseen VMWEs both in dev with respect to test and in test with respect to train+dev, and on the other hand in ensuring a sufficient number of unseen VMWEs in the test set for each language. The Romanian training corpus contains 195k tokens in which 1,218 VMWEs are annotated. The Romanian dev set contains 134,340 tokens and 818 annotated VMWEs; the Romanian test set includes 685,566 tokens and 4,135 annotated VMWEs. The frequency of occurrence of VMWEs in Romanian ranges from 8% (for LVC.full) to 22% (for LVC.cause), with an average of 12%, thus being quite redundant (Barbu Mititelu et al., 2019).

3 System Description

3.1 Monolingual Training

We experiment with four BERT-based models (first two monolingual and last two multilingual) for MWE identification using only the Romanian part of the PARSEME 1.2 corpus, namely the RoBERT, the Distilled Romanian BERT (Distil-RoBERT) (Avram et al., 2022a), the multilingual BERT (M-BERT) (Kenton and Toutanova, 2019), and the XLM-RoBERTa (Conneau et al., 2020). We follow the standard sequence tagging procedure described in the original BERT model and fine-tune the embeddings produced by the last layer for the input tokens to predict the corresponding MWE labels using a feed-forward layer.

3.2 Multilingual Training

Our second and principal line of work here combines all the training sets of the corpora. Therefore, we train the two multilingual language models on

the resulting dataset and then evaluate the models on the Romanian test set of the PARSEME 1.2 shared task. In addition, we improve the performance of the system by forcing the embeddings of the respective language models to depend less on their source language and more on the semantic specificities of an MWE using a lateral inhibition layer and adversarial training.

The general architecture of our multilingual training methodology is depicted in Figure 1. It is divided into three major components: a multilingual BERT model that acts as a feature extractor F and produces the embeddings of the tokens, a classifier C whose role is to identify the MWEs in the given texts, and a language discriminator LG whose role is to recognize the language of the input. We employ the lateral inhibition layer before feeding the embeddings to C and adversarially train LG by reversing its gradient before backpropagating through F. Further details on these two methods are given below.

3.3 Lateral Inhibition

The neural inhibitory layer, modelled after the biological process of lateral inhibition in the brain, has been successfully used for the named entity recognition (NER) task in the past (Păiș, 2022; Avram et al., 2022b; Mitrofan and Păiș, 2022). We envisage that since the terms recognised by NER are just a subset of the MWEs identification, both being grounded in sequence tagging, introducing this layer into our model would also bring improvements in the final performance of our system. However, in the previous work, the neural inhibitory layer was mainly used to enhance the quality of the extracted named entities. In contrast, in this work, we employ it to achieve language-independent embeddings out of the multilingual transformer models.

The main idea behind the lateral inhibitory layer is quite simple. Given the embeddings X produced by a language model and a weight matrix W with a bias b, the output Y of this layer is described in the following formula:

$$Y = X * Diag(H(X * ZeroDiag(W) + b)) \quad (1)$$

where $Diag$ is a function that creates a matrix whose main diagonal is the vector given as input, $ZeroDiag$ is a function that sets a given matrix with the zero value on the main diagonal, and H is the Heaviside step function.
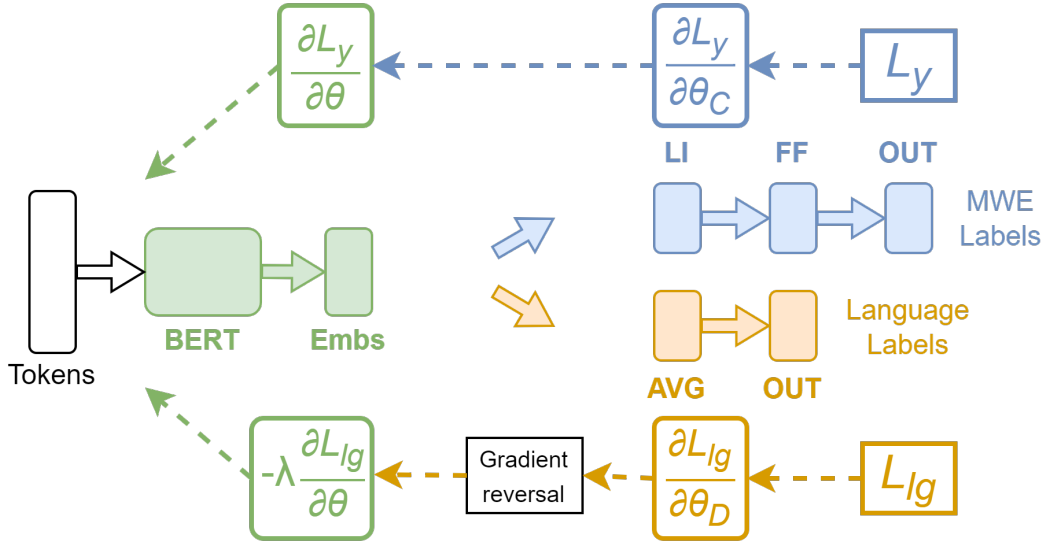
Figure 1: The multilingual training architecture. We use a multilingual BERT-based model to extract the embeddings from the input tokens (green). All these embeddings are fed into a classifier with a lateral inhibition layer to predict the MWE labels (blue) and into an adversarially trained language discriminator (orange). The block arrow depicts the forward pass, and the dotted arrow the backward pass.

Equation 1 works well for the forward pass. However, since the Heaviside step function is not differentiable, the lateral inhibition layer approximates the respective gradients with the gradients of the parameterized Sigmoid function (Wunderlich and Pehle, 2021), a technique known as surrogate gradient learning (Neftci et al., 2019).

3.4 Adversarial Training

Adversarial training of neural networks has been a highly influential area of research in recent years, particularly in fields such as computer vision with generative unsupervised models (Gui et al., 2021). Adversarial training has also been used to train predictive models (Zhao et al., 2022), and in recent research, both multilingual and cross-lingual adversarial neural networks were introduced (Hu et al., 2019; Guzman-Nateras et al., 2022). These networks are designed to learn discriminative representations that are invariant to language. In this study, we utilize the same methodology to learn task-specific representations in a multilingual setting, trying to improve the predictive capabilities of the employed multilingual transformer models.

Our methodology closely follows the domain adversarial neural network algorithm (DANN) (Ganin et al., 2016), the difference here being that instead of reversing the gradient to create domain-independent features, we reverse it to generate language-independent embeddings out of the multilingual transformer models. As is the case for our system, DANN has in its composition a feature extractor F, a label classifier C, and a domain classifier D that is replaced in our work with a language classifier LG. Thus, the gradient computation of each component can be formalized in the following equations:

$$
\begin{aligned}
\theta_C &= \theta_C - \alpha \frac{\partial L_y}{\partial \theta_C} \\
\theta_{LG} &= \theta_{LG} - \alpha \frac{\partial L_{lg}}{\partial \theta_{LG}} \\
\theta_F &= \theta_F - \alpha \left(\frac{\partial L_y}{\partial \theta_F} - \lambda \frac{\partial L_{lg}}{\partial \theta_F} \right)
\end{aligned}
\tag{2}
$$

where θ_C are the parameters of the label classifier, L_y is the loss obtained by the label classifier when predicting the class labels y, θ_{LG} are the parameters of the language classifier, L_{lg} is the loss obtained by the language classifier when predicting the language labels d, θ_F are the parameters of the feature extractor, λ is the hyperparameter used to reverse the gradients, and α is the learning rate.

4 Results

4.1 Monolingual Training

Table 1 shows the results of our monolingual training. We report both the overall scores (called global MWE) and the scores of the identified MWEs that do not appear in the training set (called unseen

Model	Global MWE			Unseen MWE		
	P	R	F1	P	R	F1
MTLB-STRUCT	89.88	91.05	90.46	28.84	41.47	34.02
TRAVIS-mono	**90.80**	91.39	91.09	33.05	51.51	40.26
RoBERT	90.73	**93.74**	**92.21**	**52.97**	**70.69**	**60.56**
Distil-RoBERT	87.56	90.40	88.96	41.06	62.77	49.65
M-BERT	90.39	90.11	90.25	46.82	51.09	48.86
XLM-RoBERTa	90.72	91.46	91.09	51.54	62.77	56.61

Table 1: The results of the models trained on the monolingual Romanian set.

Model	Global MWE			Unseen MWE		
	P	R	F1	P	R	F1
M-BERT	**91.34**	88.46	**89.88**	**49.90**	48.12	48.99
M-BERT + LI	90.78	88.85	89.81	45.06	45.15	45.10
M-BERT + Adv	89.14	**90.13**	89.63	46.27	**56.44**	**50.85**
M-BERT + LI + Adv	89.95	88.78	89.36	45.44	50.30	47.74
XLM-RoBERTa	**91.23**	92.53	**91.87**	52.92	**64.55**	58.16
XLM-RoBERTa + LI	91.12	92.02	91.02	52.11	61.19	56.28
XLM-RoBERTa + Adv	89.45	**92.87**	91.12	54.91	63.96	59.09
XLM-RoBERTa + Adv + LI	90.49	92.61	91.53	**55.01**	64.47	**59.36**

Table 2: The results of the multilingual models trained on the multilingual combined dataset and evaluated on the Romanian set. LI means lateral inhibition, and Adv means multilingual adversarial training.

MWE), as well as the results of the best overall system (MTLB-STRUCT) (Taslimipoor et al., 2020) and the results of the best system on Romanian (TRAVIS-mono) (Kurfalı, 2020). All our monolingual models outperform the MTLB-STRUCT and TRAVIS-mono systems by more than 8% on unseen MWE, with RoBERT achieving an improvement of more than 20%. We believe that this is due to the more intensive hyperparameter search that we performed and the text preprocessing which consisted of things like replacing the letters with diacritics in Romanian to the standard used in pretraining or making sure that the tokenizer produces cased subtokens[3].

Both the highest global MWE and unseen MWE performance were achieved by the monolingual RoBERT model, with F1-scores of 92.21% and 60.56%, respectively. The second highest performance was obtained by the XLM-RoBERTa model, although it is a multilingual model. Thus, XLM-RoBERTa outperformed the other monolin-

gual model, Distil-RoBERT, by 2.1% on global MWE and 7% on unseen MWE. This phenomenon has also been noticed by Conneau et al. (2020), showing the raw power of multilingual models pretrained on a large amount of textual data.

4.2 Multilingual Training

Table 2 shows the results for the multilingual training of both M-BERT and XLM-RoBERTa. As in the monolingual training case, XLM-RoBERTa achieves better performance, coming out on top with an F1-score of 58.16% in comparison with the 48.99% F1-score obtained by M-BERT. We also notice that the simple multilingual training (i.e., without lateral inhibition and adversarial training) improves the results of the two models when trained on the monolingual Romanian set.

The adversarial training improves the performance of both M-BERT and XLM-RoBERTa in multilingual training. At the same time, the lateral inhibition layer brought improvements only to the later when it was combined with adversarial training. Thus, by merging the two methodologies, we outperform the XLM-RoBERTa's results trained

---
[3]These text preprocessing techniques are suggested at https://github.com/dumitrescustefan/Romanian-Transformers.

on monolingual data (i.e., around 2.7% on unseen MWEs), which was the main target of the competition, being behind RoBERT with only 1.2%.

5 Conclusions

The detection and processing of MWEs play an important role in various areas of NLP. This paper made notable improvements in unseen Romanian MWE identification by employing a lateral inhibition layer and adversarial training to multilingual large language models like XLM-RoBERTa. This way, we were able to improve the results of XLM-RoBERTa. In addition, we achieved SOTA results on this task with a simple fine-tuning of RoBERT that involved a better hyperparameter search and text preprocessing pipeline, respectively.

Future work considers an analysis of the language-independent embeddings produced in the multilingual training, together with more experiments on other languages, to validate the generalization of this approach. In addition, we intend to add these results in LiRo - the public benchmark for Romanian NLP models (Dumitrescu et al., 2021).

References

Andrei-Marius Avram, Darius Catrina, Dumitru-Clementin Cercel, Mihai Dascalu, Traian Rebedea, Vasile Pais, and Dan Tufis. 2022a. Distilling the knowledge of Romanian BERTs using multiple teachers. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 374–384, Marseille, France. European Language Resources Association.

Andrei-Marius Avram, Vasile Păiș, and Maria Mitrofan. 2022b. Racai@ smm4h'22: Tweets disease mention detection using a neural lateral inhibitory mechanism. In *Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task*, pages 1–3.

Verginica Barbu Mititelu, Mihaela Cristescu, and Mihaela Onofrei. 2019. The Romanian corpus annotated with verbal multiword expressions. In *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*, pages 13–21, Florence, Italy. Association for Computational Linguistics.

Maria Carmela Catone, Mariacristina Falco, Alessandro Maisto, Serena Pelosi, and Alfonso Siano. 2019. Automatic text classification through point of cultural interest digital identifiers. In *International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, pages 211–220. Springer.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Édouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451.

Stefan Dumitrescu, Andrei-Marius Avram, and Sampo Pyysalo. 2020. The birth of romanian bert. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4324–4328.

Stefan Daniel Dumitrescu, Petru Rebeja, Beata Lorincz, Mihaela Gaman, Andrei Avram, Mihai Ilie, Andrei Pruteanu, Adriana Stan, Lorena Rosia, Cristina Iacobescu, et al. 2021. Liro: Benchmark and leaderboard for romanian language tasks. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*.

Kurt Englmeier and Pedro Contreras. 2021. Aspect fusion as design paradigm for legal information retrieval. In *International Conference on Intelligent Human Systems Integration*, pages 547–553. Springer.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030.

Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. 2021. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*.

Luis Guzman-Nateras, Minh Van Nguyen, and Thien Nguyen. 2022. Cross-lingual event detection via optimized adversarial training. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5588–5599.

Ke Hu, Hasim Sak, and Hank Liao. 2019. Adversarial training for multilingual acoustic modeling. *arXiv preprint arXiv:1906.07093*.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

Murathan Kurfalı. 2020. Travis at parseme shared task 2020: How good is (m) bert at seeing the unseen? In *Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons*, pages 136–141.

Daniel Lowd and Christopher Meek. 2005. Adversarial learning. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 641–647.

Maria Mitrofan and Vasile Păiș. 2022. Improving romanian bioner using a biologically inspired system. In *Proceedings of the 21st Workshop on Biomedical Language Processing*, pages 316–322.

Emre O Neftci, Hesham Mostafa, and Friedemann Zenke. 2019. Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 36(6):51–63.

Vasile Păiș. 2022. Racai at semeval-2022 task 11: Complex named entity recognition using a lateral inhibition mechanism. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 1562–1569.

Carlos Ramisch, Silvio Ricardo Cordeiro, Agata Savary, Veronika Vincze, Verginica Barbu Mititelu, Archna Bhatia, Maja Buljan, Marie Candito, Polona Gantar, Voula Giouli, Tunga Güngör, Abdelati Hawwari, Uxoa Iñurrieta, Jolanta Kovalevskaitė, Simon Krek, Timm Lichte, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Behrang QasemiZadeh, Renata Ramisch, Nathan Schneider, Ivelina Stoyanova, Ashwini Vaidya, Abigail Walsh, Cristina Aceta, Itziar Aduriz, Jean-Yves Antoine, Špela Arhar Holdt, Gözde Berk, Agnė Bielinskienė, Goranka Blagus, Loic Boizou, Claire Bonial, Valeria Caruso, Jaka Čibej, Matthieu Constant, Paul Cook, Mona Diab, Tsvetana Dimitrova, Rafael Ehren, Mohamed Elbadrashiny, Hevi Elyovich, Berna Erden, Ainara Estarrona, Aggeliki Fotopoulou, Vassiliki Foufi, Kristina Geeraert, Maarten van Gompel, Itziar Gonzalez, Antton Gurrutxaga, Yaakov Ha-Cohen Kerner, Rehab Ibrahim, Mihaela Ionescu, Kanishka Jain, Ivo-Pavao Jazbec, Teja Kavčič, Natalia Klyueva, Kristina Kocijan, Viktória Kovács, Taja Kuzman, Svetlozara Leseva, Nikola Ljubešić, Ruth Malka, Stella Markantonatou, Héctor Martínez Alonso, Ivana Matas, John McCrae, Helena de Medeiros Caseli, Mihaela Onofrei, Emilia Palka-Binkiewicz, Stella Papadelli, Yannick Parmentier, Antonio Pascucci, Caroline Pasquer, Maria Pia di Buono, Vandana Puri, Annalisa Raffone, Shraddha Ratori, Anna Riccio, Federico Sangati, Vishakha Shukla, Katalin Simkó, Jan Šnajder, Clarissa Somers, Shubham Srivastava, Valentina Stefanova, Shiva Taslimipoor, Natasa Theoxari, Maria Todorova, Ruben Urizar, Aline Villavicencio, and Leonardo Zilio. 2018. Annotated corpora and tools of the PARSEME shared task on automatic identification of verbal multiword expressions (edition 1.1). LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Carlos Ramisch, Bruno Guillaume, Agata Savary, Jakub Waszczuk, Marie Candito, Ashwini Vaidya, Verginica Barbu Mititelu, Archna Bhatia, Uxoa Iñurrieta, Voula Giouli, Tunga Güngör, Menghan Jiang, Timm Lichte, Chaya Liebeskind, Johanna Monti, Renata Ramisch, Sara Stymme, Abigail Walsh, Hongzhi Xu, Emilia Palka-Binkiewicz, Rafael Ehren, Sara Stymne, Matthieu Constant, Caroline Pasquer, Yannick Parmentier, Jean-Yves Antoine, Carola Carlino, Valeria Caruso, Maria Pia Di Buono, Antonio Pascucci, Annalisa Raffone, Anna Riccio, Federico Sangati, Giulia Speranza, Renata Ramisch, Silvio Ricardo Cordeiro, Helena de Medeiros Caseli, Isaac Miranda, Alexandre Rademaker, Oto Vale, Aline Villavicencio, Gabriela Wick Pedro, Rodrigo Wilkens, Leonardo Zilio, Monica-Mihaela Rizea, Mihaela Ionescu, Mihaela Onofrei, Jia Chen, Xiaomin Ge, Fangyuan Hu, Sha Hu, Minli Li, Siyuan Liu, Zhenzhen Qin, Ruilong Sun, Chenweng Wang, Huangyang Xiao, Peiyi Yan, Tsy Yih, Ke Yu, Songping Yu, Si Zeng, Yongchen Zhang, Yun Zhao, Vassiliki Foufi, Aggeliki Fotopoulou, Stella Markantonatou, Stella Papadelli, Sevasti Louizou, Itziar Aduriz, Ainara Estarrona, Itziar Gonzalez, Antton Gurrutxaga, Larraitz Uria, Ruben Urizar, Jennifer Foster, Teresa Lynn, Hevi Elyovitch, Yaakov Ha-Cohen Kerner, Ruth Malka, Kanishka Jain, Vandana Puri, Shraddha Ratori, Vishakha Shukla, Shubham Srivastava, Gozde Berk, Berna Erden, and Zeynep Yirmibeşoğlu. 2020. Annotated corpora and tools of the PARSEME shared task on semi-supervised identification of verbal multiword expressions (edition 1.2). LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Agata Savary, Carlos Ramisch, Silvio Ricardo Cordeiro, Federico Sangati, Veronika Vincze, Behrang QasemiZadeh, Marie Candito, Fabienne Cap, Voula Giouli, Ivelina Stoyanova, Antoine Doucet, Kübra Adalı, Verginica Barbu Mititelu, Eduard Bejček, Ismail El Maarouf, Gülşen Eryiğit, Luke Galea, Yaakov Ha-Cohen Kerner, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Jolanta Kovalevskaitė, Simon Krek, Lonneke van der Plas, Cristina Aceta, Itziar Aduriz, Jean-Yves Antoine, Greta Attard, Kirsty Azzopardi, Loic Boizou, Janice Bonnici, Mert Boz, Ieva Bumbulienė, Jael Busuttil, Valeria Caruso, Manuela Cherchi, Matthieu Constant, Monika Czerepowicka, Anna De Santis, Tsvetana Dimitrova, Tutkum Dinç, Hevi Elyovich, Ray Fabri, Alison Farrugia, Jamie Findlay, Aggeliki Fotopoulou, Vassiliki Foufi, Sara Anne Galea, Polona Gantar, Albert Gatt, Anabelle Gatt, Carlos Herrero, Uxoa Iñurrieta, Glorianna Jagfeld, Milena Hnátková, Mihaela Ionescu, Natalia Klyueva, Svetla Koeva, Viktória Kovács, Taja Kuzman, Svetlozara Leseva, Sevi Louisou, Teresa Lynn, Ruth Malka, Héctor Martínez Alonso, John McCrae, Helena de Medeiros Caseli, Ayşenur Miral, Amanda Muscat, Joakim Nivre, Michael Oakes, Mihaela Onofrei, Yannick Parmentier, Caroline Pasquer, Maria Pia di Buono, Belem Priego Sanchez, Annalisa Raffone, Renata Ramisch, Erika Rimkutė, Monica-Mihaela Rizea, Katalin Simkó, Michael Spagnol, Valentina Stefanova, Sara Stymne, Umut Sulubacak, Nicole Tabone, Marc Tanti, Maria Todorova, Zdenka Urešová, Aline Villavicencio, and Leonardo Zilio. 2017. Annotated corpora and tools of the PARSEME shared task on automatic identification of verbal multiword expressions (edition 1.0).

LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Shiva Taslimipoor, Sara Bahaadini, and Ekaterina Kochmar. 2020. Mtlb-struct@ parseme 2020: Capturing unseen multiword expressions using multi-task learning and pre-trained masked language models. In *Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons*, pages 142–148.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Timo C Wunderlich and Christian Pehle. 2021. Event-based backpropagation can compute exact gradients for spiking neural networks. *Scientific Reports*, 11(1):12829.

Andrea Zaninello and Alexandra Birch. 2020. Multiword expression aware neural machine translation. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3816–3825.

Weimin Zhao, Sanaa Alwidian, and Qusay H Mahmoud. 2022. Adversarial training methods for deep learning: A systematic review. *Algorithms*, 15(8):283.