Advances in Multiword Expression Identification for the Italian language: The PARSEME shared task edition 1.1

Johanna Monti[1], Silvio Ricardo Cordeiro[2], Carlos Ramisch[2]
Federico Sangati[1], Agata Savary[3], Veronika Vincze[4]
[1]University L'Orientale, Naples, Italy [2]Aix Marseille Univ, CNRS, LIS, Marseille, France
[3]University of Tours, France [4]MTA-SZTE Research Group on Artificial Intelligence, Hungary
jmonti@unior.it,silvioricardoc@gmail.com
carlos.ramisch@lis-lab.fr, fsangati@unior.it
agata.savary@univ-tours.fr, vinczev@inf.u-szeged.hu

Abstract

English. This contribution describes the results of the second edition of the shared task on automatic identification of verbal multiword expressions, organized as part of the *LAW-MWE-CxG 2018 workshop*, co-located with COLING 2018, concerning both the PARSEME-IT corpus and the systems that took part in the task for the Italian language. The paper will focus on the main advances in comparison to the first edition of the task.

Italiano. *Il presente contributo descrive i risultati della seconda edizione dello 'Shared task on automatic identification of verbal multiword expressions' organizzato nell'ambito del LAW-MWE-CxG 2018 workshop realizzato durante il COLING 2018 riguardo sia il corpus PARSEME-IT e i sistemi che hanno preso parte nel task per quel che riguarda l'italiano. L'articolo tratta i principali progressi ottenuti a confronto con la prima edizione del task.*

1 Introduction

Multiword expressions (MWEs) are a particularly challenging linguistic phenomenon to be handled by NLP tools. In recent years, there has been a growing interest in MWEs since the possible improvements of their computational treatment may help overcome one of the main shortcomings of many NLP applications, from Text Analytics to Machine Translation. Recent contributions to this topic, such as Mitkov et al. (2018) and Constant et al. (2017) have highlighted the difficulties that this complex phenomenon, halfway between lexicon and syntax, characterized by idiosyncrasy on various levels, poses to NLP tasks.

This contribution will focus on the advances in the identification of verbal multiword expressions (VMWEs) for the Italian language. In Section 2 we discuss related work. In Section 3 we give an overview of the PARSEME shared task. In Section 4 we present the resources developed for the Italian language, namely the guidelines and the corpus. Section 5 is devoted to the annotation process and the inter-annotator agreement. Section 6 briefly describes the thirteen systems that took part in the shared task and the results obtained. Finally, we discuss conclusions and future work (Section 7).

2 Related work

MWEs have been the focus of the PARSEME COST Action, which enabled the organization of an international and highly multilingual research community (Savary et al., 2015). This community launched in 2017 the first edition of the PARSEME shared task on automatic identification of verbal MWEs, aimed at developing universal terminologies, guidelines and methodologies for 18 languages, including the Italian language (Savary et al., 2017). The task was co-located with the 13th Workshop on Multiword Expressions (MWE 2017), which took place during the European Chapter of the Association for Computational Linguistics (EACL 2017). The main outcomes for the Italian language were the **PARSEME-IT Corpus**, a 427-thousand-word annotated corpus of verbal MWEs in Italian (Monti et al., 2017) and the participation of four systems[1], namely TRANSITION, a transition-based dependency parsing system (Al Saied et al., 2017), SZEGED based on the POS and dependency modules of the Bohnet parser (Simkó et al., 2017), ADAPT (Maldonado et al., 2017) and RACAI (Boroş et al., 2017), both based on sequence la-

[1]http://multiword.sourceforge.net/sharedtaskresults2017

beling with CRFs. Concerning the identification of verbal MWEs some further recent contributions specifically focusing on the Italian language are:

- A supervised token-based identification approach to Italian Verb+Noun expressions that belong to the category of complex predicates (Taslimipoor et al., 2017). The approach investigates the inclusion of concordance as part of the feature set used in supervised classification of MWEs in detecting literal and idiomatic usages of expressions. All concordances of the verbs *fare* ('to do/ to make'), *dare* ('to give'), *prendere* ('to take') and *trovare* ('to find') followed by any noun, taken from the itWaC corpus (Baroni and Kilgarriff, 2006) using SketchEngine (Kilgarriff et al., 2004) are considered.

- A neural network trained to classify and rank idiomatic expressions under constraints of data scarcity (Bizzoni et al., 2017).

With reference to corpora annotated with VMWEs for the Italian language and in comparison with the state of the art described in Monti et al. (2017), there are no further resources available so far. At the time of writing, therefore, the PARSEME-IT VMWE corpus still represents the first sample of a corpus which includes several types of VMWEs, specifically developed to foster NLP applications. The corpus is freely available, with the latest version (1.1) representing an enhanced corpus with some substantial changes in comparison with version 1.0 (cf. Section 4).

3 The PARSEME shared task

The second edition of the PARSEME shared task on automatic identification of verbal multiword expressions (VMWEs) was organized as part of the LAW-MWE-CxG 2018 workshop co-located with COLING 2018 (Santa Fe, USA)[2] and aimed at identifying verbal MWEs in running texts. According to the rules set forth in the shared task, system results could be submitted in two tracks:

- CLOSED TRACK: Systems using only the provided training/development data - VMWE annotations + morpho-syntactic data (if any) - to learn VMWE identification models and/or rules.

- OPEN TRACK: Systems using or not the provided training/development data, plus any additional resources deemed useful (MWE lexicons, symbolic grammars, wordnets, raw corpora, word embeddings, language models trained on external data, etc.). This track includes notably purely symbolic and rule-based systems.

The PARSEME members elaborated for each language i) annotation guidelines based on annotation experiments ii) corpora in which VMWEs are annotated according to the guidelines. Corpora were split in training, development and tests corpora for each language. Manually annotated training and development corpora were made available to the participants in advance, in order to allow them to train their systems and to tune/optimize the systems' parameters. Raw (unannotated) test corpora were used as input to the systems during the evaluation phase. The contribution of the PARSEME-IT research group[3] to the shared task is described in the next section.

4 Italian resources for the shared task

The PARSEME-IT research group contributed to the edition 1.1 of the shared task with the development of specific guidelines for the Italian language and with the annotation of the Italian corpus with over 3,700 VMWEs.

4.1 The shared task guidelines

The 2018 edition of the shared task relied on enhanced and revised guidelines (Ramisch et al., 2018). The guidelines[4] are provided with Italian examples for each category of VMWE.

The guidelines include two universal categories, i.e. valid for all languages participating in the task:

- **Light-verb constructions (LVCs)** with two subcategories: LVCs in which the verb is semantically totally bleached (LVC.full) like in *fare un discorso* ('to give a speech'), and LVCs in which the verb adds a causative meaning to the noun (LVC.cause) like in *dare il mal di testa* ('to give a headache');

- **Verbal idioms (VIDs)** like *gettare le perle ai porci* ('to throw pearls before swine').

[2] https:http://multiword.sourceforge.net/lawmwecxg2018

[3] https://sites.google.com/view/parseme-it/home

[4] http://parsemefr.lif.univ-mrs.fr/parseme-st-guidelines/1.1/

Three quasi-universal categories, valid for some language groups or languages but non-existent or very exceptional in others are:

- **Inherently reflexive verbs (IRV)** which are those reflexive verbal constructions which (a) never occur without the clitic e.g. *suicidarsi* ('to suicide'), or when (b) the IRV and non-reflexive versions have clearly different senses or subcategorization frames e.g. *riferirsi* ('to refer') opposed to *riferire* ('to report / to tell');

- **Verb-particle constructions (VPC)** with two subcategories: fully non-compositional VPCs (VPC.full), in which the particle totally changes the meaning of the verb, like *buttare giù* ('to swallow') and semi non-compositional VPCs (VPC.semi), in which the particle adds a partly predictable but non-spatial meaning to the verb like in *andare avanti* ('to proceed');

- **Multi-verb constructions (MVC)** composed by a sequence of two adjacent verbs like in *lasciar perdere* ('to give up').

An optional experimental category (if admitted by the given language, as is the case for Italian) is considered in a post-annotation step:

- **Inherently adpositional verbs (IAVs)**, which consist of a verb or VMWE and an idiomatic selected preposition or postposition that is either always required or, if absent, changes the meaning of the verb significantly, like in *confidare su* ('to trust on').

Finally, a language-specific category was introduced for the Italian language:

- **Inherently clitic verbs (LS.ICV)** formed by a full verb combined with one or more non-reflexive clitics that represent the pronominalization of one or more complements (CLI). LS.ICV is annotated when (a) the verb never occurs without one non-reflexive clitic, like in *entrarci* ('to be relevant to something'), or (b) when the LS.ICV and the non-clitic versions have clearly different senses or subcategorization frames like in *prenderle* ('to be beaten') vs *prendere* ('to take').

4.2 The PARSEME-IT corpus

The PARSEME-IT VMWE corpus version 1.1 is an updated version of the corpus used for edition 1.0 of the shared task. It is based on a selection of texts from the PAISÀ corpus of web texts (Lyding et al., 2014), including Wikibooks, Wikinews, Wikiversity, and blog services. The PARSEME-IT VMWE corpus was updated in edition 1.1 according to the new guidelines described in the previous section. Table 4.2 summarizes the size of the corpus developed for the Italian language and presents the distribution of the annotated VMWEs per category.

The training, development and test data are available in the LINDAT/Clarin repository[5], and all VMWE annotations are available under Creative Commons licenses (see README.md files for details). The released corpus' format is based on an extension of the widely-used CoNLL-U file format.[6]

5 Annotation process

The annotation was manually performed in running texts using the FoLiA linguistic annotation tool[7] (van Gompel and Reynaert, 2013) by six Italian native speakers with a background in linguistics, using a specific decision tree for the Italian language for joint VMWE identification and classification.[8]

In order to allow the annotation of IAVs, a new pre-processing step was introduced to split compound prepositions such as *della* ('of the') into two tokens. This step was necessary to annotate only lexicalised components of the IAV, as in *portare alla disperazione*, where only the verb and the preposition *a* should be annotated, without the article *la*.

Once the annotation was completed, in order to reduce noise and to increase the consistency of the annotations, we applied the consistency checking tool developed for edition 1.0 (Savary et al., forthcoming). The tool groups all annotations of the same VMWE, making it possible to spot annotation inconsistencies very easily.

[5] http://hdl.handle.net/11372/LRT-2842

[6] http://multiword.sf.net/cupt-format

[7] http://mwe.phil.hhu.de/

[8] http://parsemefr.lif.univ-mrs.fr/parseme-st-guidelines/1.1/?page=it-dectree

	sent.	tokens	VMWEs	IAV	IRV	LS.ICV	LVC.cause/full	MVC	VID	VPC.full/semi
IT-dev	917	32613	500	44	106	9	19/100	6	197	17/2
IT-train	13555	360883	3254	414	942	20	147/544	23	1098	66/0
IT-test	1256	37293	503	41	96	8	25/104	5	201	23/0
IT-Total	15728	430789	4257	499	7641	37	191/748	35	1496	106/2

Table 1: Statistics of the PARSEME-IT corpus version 1.1.

	#S	#A_1	#A_2	F_{span}	κ_{span}	κ_{cat}
PARSEME-IT-2017	2000	336	316	0.417	0.331	0.78
PARSEME-IT-2018	1000	341	379	0.586	0.550	0.882

Table 2: IAA scores for the PARSEME-IT corpus in versions from 2017 and 2018: #S is the number of sentences in the double-annotated corpus used for measuring the IAA. #A_1 and #A_2 refer to the number of VMWE instances annotated by each of the annotators. F_{span} is the F-measure for identifying the span of a VMWE, when considering that one of the annotators tries to predict the other's annotations (VMWE categories are ignored). κ_{span} and κ_{cat} are the values of Cohen's κ for span identification and categorization, respectively.

5.1 Inter-annotator agreement

A small portion of the corpus consisting in 1,000 sentences was double-annotated. In comparison with the previous edition, the inter-annotator agreement shown in Table 2 increased, although it is still not optimal.[9] The improvement is probably due to the fact that, this time, the group was based in one place with the exception of one annotator, and several meetings took place prior to the annotation phase in order to discuss the new guidelines.

The two annotators involved in the IAA task annotated 191 VMWEs with no disagreement, but there were several problems, which led to 44 cases of partial disagreement and 250 cases of total disagreement:

- PARTIAL MATCHES LABELED, (25 cases) in which there is at least one token of the VMWE in common between two annotators and the labels assigned are the same. The disagreement mainly concerns the lexicalized elements as part of the VMWE, as in the case of the VID *porre in cattiva luce* ('make look bad'). Annotators disagreed, indeed, about considering the adjective *cattiva* ('bad') as part of the VID.
- EXACT MATCHES UNLABELED, (18 cases) in which the annotators agreed on the lexicalized components of the VMWE to be annotated but not the label. This type of disagreement is mainly related to fine-grained categories such as **LVC.cause** and **LVC.full** as in the case of *dare ... segnale* (to give ... a signal) or **VPC.full** and **VPC.semi** as for *mettere insieme* ('to put together')
- PARTIAL MATCHES UNLABELED, (1 case) in which there is at least one token of the VMWE in common between two annotators but the labels assigned are different, such as in *buttar-si in la calca* ('to join the crowd') classified as VID by the first annotator and *buttar-si* ('to throw oneself') classified as IRV by the second one in the following sentence: *[...] attendendo il venerdì sera per buttarsi nella calca del divertimento [...].* ('waiting for the Friday evening to join the crowd for entertainment')
- ANNOTATIONS CARRIED OUT ONLY BY ONE OF THE ANNOTATORS: This is the category which collects the most numerous examples of disagremeent between annotators: 106 VMWE were annotated only by annotator 1 and 144 by annotator 2.

6 The systems and the results of the shared task for the Italian language

Whereas only four systems took part in edition 1.0 of the shared task for the Italian language, in edition 1.1, fourteen systems took on this challenge. The system that took part in the PARSEME shared task are listed in Table 3: 12 took part in the closed track and two in the open one. The two systems that took part in the open track reported the resources that were used, namely SHOMA used pretrained wikipedia word embeddings (Taslimipoor and Rohanian, 2018), while Deep-BGT (Berk et al., 2018) relied on the BIO tagging scheme

[9] As mentioned in Ramisch et al. (2018), the estimation of chance agreement in κ_{span} and κ_{cat} is slightly different between 2017 and 2018, therefore these results are not directly comparable.

and its variants (Schneider et al., 2014) to introduce additional tags to encode gappy (discontinuous) VMWEs. A distinctive characteristic of the systems of edition 1.1 is that most of them (GBD-NER-resplit and GBD-NER-standard, TRAPACC, and TRAPACC-S, SHOMA, Deep-BGT) use neural networks, while the rest of the systems adopt other approaches: CRF-DepTree-categs and CRF-Seq-nocategs are based on a tree-structured CRF, MWETreeC and TRAVERSAL on syntactic trees and parsing methods, Polirem-basic and Polirem-rich on statistical methods and association measures, and finally varIDE uses a Naive Bayes classifier. The systems were ranked according two types of evaluation measures (Ramisch et al., 2018): a strict per-VMWE score (in which each VMWE in gold is either deemed predicted or not, in a binary fashion) and a fuzzy per-token score (which takes partial matches into account). For each of these two, precision (P), recall (R) and F1-scores (F) were calculated. Table 3 shows the ranking of the systems which participated in the shared task for the Italian language. The systems with highest MWE-based Rank for Italian have F1 scores that are mostly comparable to the scores obtained in the General ranking of all languages (e.g. TRAVERSAL had a General F1 of 54.0 vs Italian F1 of 49.2, being ranked first in both cases). Nevertheless, the Italian scores are consistently lower than the ones in the General ranking, even if only by a moderate margin, suggesting that Italian VMWEs in this specific corpus might be particularly harder to identify. One of the outliers in the table is MWETreeC, which predicts much fewer VMWEs than in the annotated corpora. This turned out to be true for other languages as well. The few VMWEs that were predicted only obtained partial matches, which explains why its MWE-based score was 0. Another clear outlier is Polirem-basic. Both Polirem-basic and Polirem-rich had predictions for Italian, French and Portuguese. Their scores are somewhat comparable in the three languages, suggesting that the lower scores are a characteristic of the system and not some artifact of the Italian corpus.

TRASVERSAL (Waszczuk, 2018) was the best performing system in the closed track, while SHOMA (Taslimipoor and Rohanian, 2018) performed best in the open one. As shown in Figure 1, comparing the MWE-based F1 scores for each label for the two best performing systems,

System	Track	MWE-based				Token-based			
		P	R	F1	Rank	P	R	F1	Rank
TRAVERSAL	closed	63.09	40.32	**49.2**	1	74.42	42.11	**53.78**	1
TRAPACC	closed	52.43	30.44	**38.52**	2	61.54	30.34	**40.64**	4
CRF-Seq-nocategs	closed	55.14	27.02	**36.27**	3	78.49	33.05	**46.51**	2
TRAPACC_S	closed	55.66	23.79	**33.33**	4	65.42	22.99	**34.02**	8
CRF-DepTree-categs	closed	44.76	25.81	**32.74**	5	58.78	29.8	**39.55**	5
varIDE	closed	31.07	34.07	**32.5**	6	39.22	35.06	**37.02**	6
Veyn	closed	34.01	30.44	**32.13**	7	58.41	38.16	**46.16**	3
Polirem-rich	closed	72.36	17.94	**28.76**	8	86.54	21.9	**34.96**	7
GBD-NER-standard	closed	15.45	29.84	**20.36**	9	22.68	35.45	**27.67**	9
GBD-NER-resplit	closed	10.69	28.83	**15.59**	10	16.63	38.31	**23.19**	10
Polirem-basic	closed	83.33	4.03	**7.69**	11	81.82	3.48	**6.68**	11
MWETreeC	closed	0	0	**0**	n/a	1.45	6.58	**2.38**	12
SHOMA	open	50.37	41.33	**45.4**	1	67.49	46.59	**55.13**	1
Deep-BGT	open	45.52	25.6	**32.77**	2	70	27.63	**39.62**	2

Table 3: Results for the Italian language

TRASVERSAL obtained overall better results for almost all VMWEs categories with the exception of VID and MVC, for which SHOMA showed a better performance.

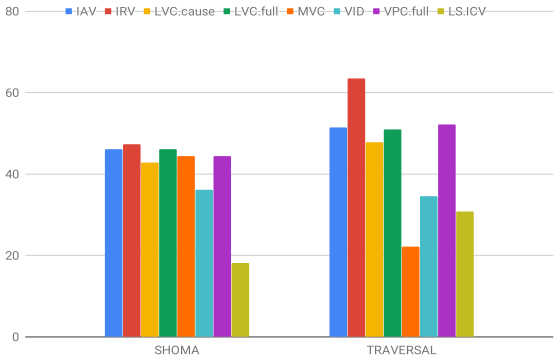

Figure 1: Chart comparing the MWE-based F1 scores for each label of the two best performing systems.

7 Conclusions and future work

Having presented the results of the PARSEME shared task edition 1.1, the paper described the advances achieved in this last edition in comparison with the previous one, but also highlighted that there is room for further improvements. We are working on some critical areas which emerged during the annotation task in particular with reference to some borderline cases and the refinement of the guidelines. Future work will focus on maintaining and increasing the quality and the size of the corpus but also on extending the shared task to other MWE categories, such as nominal MWEs.

Acknowledgments

Our thanks go to the Italian annotators Valeria Caruso, Maria Pia di Buono, Antonio Pascucci, Annalisa Raffone, Anna Riccio for their contributions.

References

Hazem Al Saied, Matthieu Constant, and Marie Candito. 2017. The ATILF-LLF system for PARSEME shared task: a transition-based verbal multiword expression tagger. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 127–132.

Marco Baroni and Adam Kilgarriff. 2006. Large linguistically-processed web corpora for multiple languages. In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*, pages 87–90. Association for Computational Linguistics.

Gözde Berk, Berna Erden, and Tunga Güngör. 2018. Deep-bgt at parseme shared task 2018: Bidirectional lstm-crf model for verbal multiword expression identification. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 248–253.

Yuri Bizzoni, Marco S. G. Senaldi, and Alessandro Lenci. 2017. Deep-learning the Ropes: Modeling Idiomaticity with Neural Networks. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017), Rome, Italy, December 11-13, 2017.*

Tiberiu Boroş, Sonia Pipa, Verginica Barbu Mititelu, and Dan Tufiş. 2017. A data-driven approach to verbal multiword expression detection. PARSEME Shared Task system description paper. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 121–126.

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. URL `https://doi.org/10.1162/COLI_a_00302`.

Adam Kilgarriff, Pavel Rychly, Pavel Smrz, and David Tugwell. 2004. Itri-04-08 the sketch engine. *Information Technology*, 105:116.

Verena Lyding, Egon Stemle, Claudia Borghetti, Marco Brunello, Sara Castagnoli, Felice Dell'Orletta, Henrik Dittmann, Alessandro Lenci, and Vito Pirrelli. 2014. The PAISÀ Corpus of Italian Web Texts. In *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, pages 36–43. Association for Computational Linguistics, Gothenburg, Sweden. URL `http://www.aclweb.org/anthology/W14-0406`.

Alfredo Maldonado, Lifeng Han, Erwan Moreau, Ashjan Alsulaimani, Koel Chowdhury, Carl Vogel, and Qun Liu. 2017. Detection of Verbal Multi-Word 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. Association for Computational Linguistics.

Ruslan Mitkov, Johanna Monti, Gloria Corpas Pastor, and Violeta Seretan. 2018. *Multiword units in machine translation and translation technology*, volume 341. John Benjamins Publishing Company.

Johanna Monti, Maria Pia di Buono, and Federico Sangati. 2017. PARSEME-IT Corpus. An annotated Corpus of Verbal Multiword Expressions in Italian. In *Fourth Italian Conference on Computational Linguistics-CLiC-it 2017*, pages 228–233. Accademia University Press.

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, and Abigail Walsh. 2018. Edition 1.1 of the PARSEME Shared Task
on Automatic Identification of Verbal Multiword Expressions. In *the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 222–240.

Agata Savary, Marie Candito, Verginica Barbu Mi-

titelu, Eduard Bejček, Fabienne Cap, Slavomír Čéplö, Silvio Ricardo Cordeiro, Gülşen Eryiğit, Voula Giouli, Maarten van Gompel, Yaakov HaCohen-Kerner, Jolanta Kovalevskaitė, Simon Krek, Chaya Liebes kind, Johanna Monti, Carla Parra Escartín, Lonneke van der Plas, Behrang QasemiZadeh, Carlos Ramisch, Federico Sangati, Ivelina Stoyanova, and Veronika Vincze. forthcoming. PARSEME multilingual 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*. 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. Association for Computational Linguistics, Valencia, Spain. URL `http://www.aclweb.org/anthology/W17-1704`.

Agata Savary, Manfred Sailer, Yannick Parmentier, Michael Rosner, Victoria Rosén, Adam Przepiórkowski, Cvetana Krstev, Veronika Vincze, Beata Wójtowicz, Gyri Smørdal Losnegaard, Carla Parra Escartín, Jakub Waszczuk, Mathieu Constant, Petya Osenova, and Federico Sangati. 2015. PARSEME – PARSing and Multiword Expressions within a European multilingual network. In *7th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC 2015)*. Poznań, Poland. URL `https://hal.archives-ouvertes.fr/hal-01223349`.

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.

Katalin Ilona Simkó, Viktória Kovács, and Veronika Vincze. 2017. USzeged: Identifying Verbal Multiword Expressions with POS Tagging and Parsing Techniques. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 48–53.

Shiva Taslimipoor and Omid Rohanian. 2018. Shoma at parseme shared task on automatic identification of vmwes: Neural multiword expression tagging with high generalisation. *arXiv preprint arXiv:1809.03056*.

Shiva Taslimipoor, Omid Rohanian, Ruslan Mitkov, and Afsaneh Fazly. 2017. Investigating the Opacity of Verb-Noun Multiword Expression Usages in Context. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 133–138. Association for Computational Linguistics. URL `http://aclweb.org/anthology/W17-1718`.

Maarten van Gompel and Martin Reynaert. 2013. FoLiA: A practical XML Format for Linguistic Annotation-a descriptive and comparative study. *Computational Linguistics in the Netherlands Journal*, 3:63–81.

Jakub Waszczuk. 2018. Traversal at parseme shared task 2018: Identification of verbal multiword expressions using a discriminative tree-structured model. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 275–282.