

NER4ID at SemEval-2022 Task 2: Named Entity Recognition for Idiomaticity Detection

Simone Tedeschi

Babelscape & Sapienza University of Rome
tedeschi@babelscape.com

Roberto Navigli

Sapienza University of Rome
navigli@diag.uniroma1.it

Abstract

Idioms are lexically-complex phrases whose meaning cannot be derived by compositionally interpreting their components. Although the automatic identification and understanding of idioms is essential for a wide range of Natural Language Understanding tasks, they are still largely under-investigated. This motivated the organization of the SemEval-2022 Task 2, which is divided into two multilingual subtasks: one about idiomaticity detection, and the other about sentence embeddings. In this work, we focus on the first subtask and propose a Transformer-based dual-encoder architecture to compute the semantic similarity between a potentially-idiomatic expression and its context and, based on this, predict idiomaticity. Then, we show how and to what extent Named Entity Recognition can be exploited to reduce the degree of confusion of idiom identification systems and, therefore, improve performance. Our model achieves 92.1 F_1 in the one-shot setting and shows strong robustness towards unseen idioms achieving 77.4 F_1 in the zero-shot setting. We release our code at <https://github.com/Babelscape/ner4id>.

1 Introduction

One of the main challenges in Natural Language Processing (NLP) is to embed the meaning of a piece of raw text (e.g. a word or a sentence) in a low-dimensional dense vector. With the advent of pretrained language models, which exploit contextual information and assume compositionality of word representations, significant improvements have been made in this direction (Peters et al., 2018; Devlin et al., 2019). On the other hand, very little attention has been paid to idiomatic expressions, i.e. multi-word expressions (MWEs) with an established meaning unrelated to the meanings of the individual constituents. However, since idiomaticity is a frequent phenomenon that can be observed in all languages, idiomatic expressions should play

an important role in NLP. Indeed, their identification and understanding is crucial not only for Natural Language Understanding tasks such as Word Sense Disambiguation (Bevilacqua et al., 2021b), Semantic Role Labeling (Conia et al., 2021) and Semantic Parsing (Bevilacqua et al., 2021a), but also for Machine Translation (Edunov et al., 2018; Liu et al., 2020), Question Answering (Mishra and Jain, 2016) and Text Summarization (Chu and Wang, 2018), *inter alia*.

In the SemEval-2022 Task 2: *Multilingual Idiomaticity Detection and Sentence Embedding* (Tayyar Madabushi et al., 2022), research on idioms has been promoted by adapting datasets and tasks from the work carried out by Tayyar Madabushi et al. (2021). Specifically, the organizers propose two subtasks:

- **Subtask A:** a binary classification task in which potentially-idiomatic expressions (PIEs) must be labeled as either "Idiomatic" or "Literal", based on the context they appear in. To better test models' generalization capabilities, two different settings are provided: *zero-shot* and *one-shot*;
- **Subtask B:** requires models to output the correct Semantic Text Similarity (STS) scores between sentence pairs based on whether or not each sentence contains an idiomatic expression. Subtask B is also available in two settings: *pre-train* and *fine-tune*.

Both subtasks cover three languages: English, Portuguese and Galician¹. In addition, the organizers provide strong baseline systems to compare with.

In this paper, we present the NER4ID submission to the SemEval-2022 Task 2 which focuses on Subtask A. Specifically, we successfully tackle

¹Galician is included only in the test sets to test transfer-learning abilities of the models.

the idiom identification task by introducing a two-step system that: i) uses Named Entity Recognition (NER) to pre-identify non-idiomatic expressions, and ii) exploits a novel Transformer-based dual-encoder architecture to compute the semantic similarities between the remaining potentially-idiomatic expressions and their contexts and, based on these, predict idiomaticity. Finally, we extensively evaluate our system on both one-shot and zero-shot settings. We release our code at <https://github.com/Babelscape/ner4id>.

2 Related Work

Approaches to idiom identification were initially built on the notion that idiomatic expressions, like other MWEs, are less syntactically and lexically flexible than non-idiomatic and compositional ones. Indeed, initial studies focused on specific syntactic constructions. [Fazly and Stevenson \(2006\)](#) focused on verb/noun idioms, e.g. *shoot the breeze*, and used the Pointwise Mutual Information (PMI, [Church et al., 1991](#)) measure to quantify the degree of lexical, syntactic, and overall fixedness of a given verb+noun combination. [Cook et al. \(2007\)](#) and [Diab and Bhutada \(2009\)](#) also focused on verb/noun idioms using similar strategies. Other studies, instead, focused on verb/particle idioms, e.g. *call off* ([Ramisch et al., 2008](#)), or on idioms satisfying specific restrictions, i.e. subject/verb, such as *tension mounted*, and verb/direct-object, e.g. *break the ice* ([Shutova et al., 2010](#)).

The following generation of approaches exploited semantic idiosyncrasy, i.e. the linguistic property in which the meaning of an idiomatic expression cannot be completely derived from the meaning of its individual constituents. This property causes idioms to appear in contexts typically unrelated to the meaning of their individual components, hence it provides a key aspect to be exploited in an automatic approach. In particular, [Muzny and Zettlemoyer \(2013\)](#) introduced new lexical and graph-based features that use WordNet² and Wiktionary³, and proposed a simple yet efficient binary Perceptron classifier to distinguish idiomatic and literal expressions by exploiting their components and dictionary definitions. A similar, but unsupervised approach that relied on the dictionary definitions of each component of a given idiom was adopted by [Verma and Vuppuluri \(2015\)](#).

²<https://wordnet.princeton.edu/>

³<https://www.wiktionary.org/>

Finally, these latter methods have been superseded by approaches making use of distributional similarity in the form of both static and contextualized word embeddings ([Gharbieh et al., 2016](#); [Ehren, 2017](#); [Senaldi et al., 2019](#); [Liu and Hwa, 2019](#); [Hashempour and Villavicencio, 2020](#); [Kurfalı and Östling, 2020](#); [Fakharian, 2021](#); [Garcia et al., 2021](#); [Nedumpozhimana and Kelleher, 2021](#)), while keeping the underlying assumption unchanged, that is, the vector representation of the component words should be distant from the vector representation of the context, or of the expression as a whole.

Although efforts have been made in this direction, most of the studies to date have focused on the English language. Additionally, the low performance of current idiomaticity detection systems makes them not very reliable, and therefore such systems tend not to be included in downstream applications. In this work, instead, we propose a high-performance multilingual system for idiomaticity identification.

3 NER4ID

We first describe our architecture for idiomaticity detection (Section 3.1), and then we show how Named Entity Recognition can be included to obtain a more robust idiom identification system (Section 3.2). Figure 1 provides a graphical representation of the overall idiom identification system.

3.1 Dual-Encoder Architecture

In order to distinguish between compositional and idiomatic phrases, we exploit the semantic idiosyncrasy property of idiomatic expressions. This property often implies that when a MWE occurs with its idiomatic meaning, then the meaning of its individual components is unrelated to the surrounding context. On the other hand, when the expression has a compositional meaning, individual words are related to the context. To better explain, consider the following two alternatives in which the potentially idiomatic expression *piece of cake* occurs:

- a) Decryption is a *piece of cake* if you know the override codes;
- b) Tom ate the last *piece of cake*, but if you want, I'm making another dessert.

In the first case, where *piece of cake* has an idiomatic meaning (i.e. it means *straightforward*),

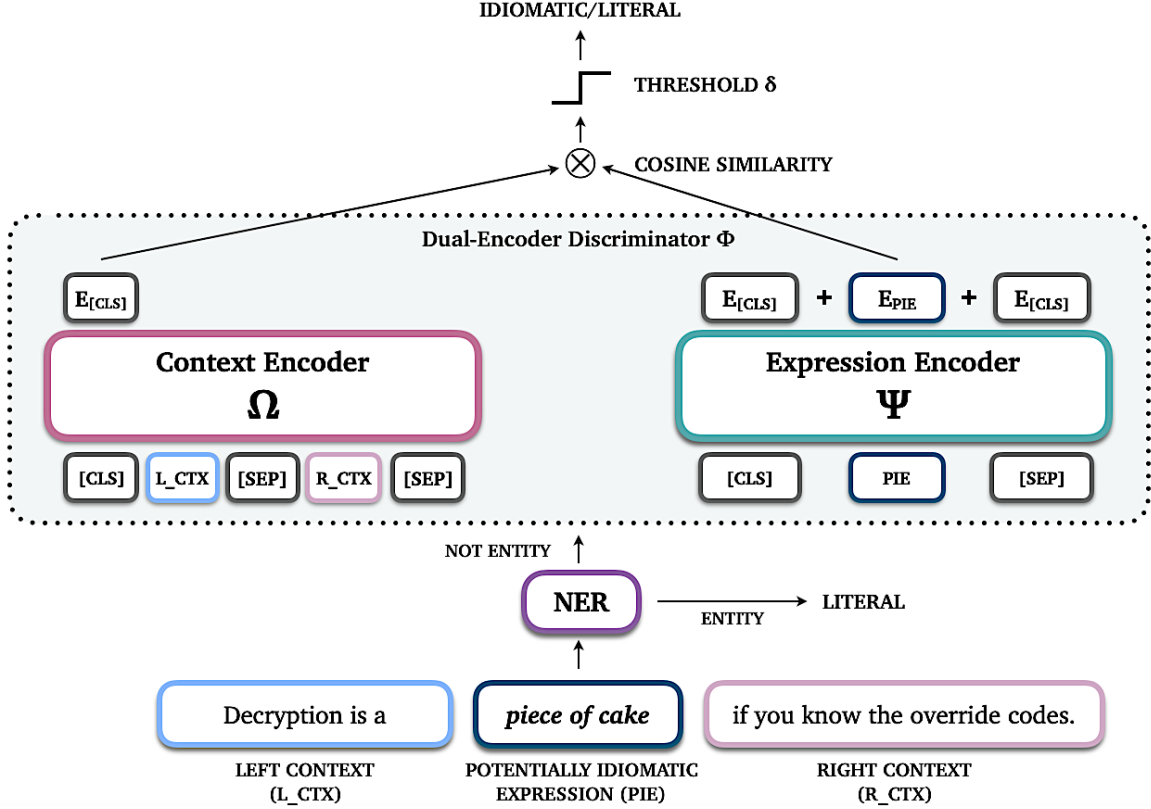


Figure 1: Graphical representation of our architecture for idiomaticity detection. “E” stands for Embedding. A potentially idiomatic expression e is labeled as idiomatic when: i) e is not an entity, and ii) the cosine similarity score between the representations $\Omega(c)$ and $\Psi(e)$, where c is the surrounding context, is lower than the threshold δ .

the word *cake* has nothing to do with the surrounding context. In the second case, instead, we find multiple words whose meaning is related to the meaning of *cake*, i.e. *ate* and *dessert*.

Following the above described intuition, and taking inspiration from recent advances in the main disambiguation tasks (Blevins and Zettlemoyer, 2020; Botha et al., 2020; Tedeschi et al., 2021a), we design a dual-encoder architecture to produce a vector representation for both the expression and its context, and then, based on their cosine similarity, we label the expression as either idiomatic or literal. More formally, let us define an expression encoder Ψ and a context encoder Ω . Then, given an expression-context pair $\langle e, c \rangle$, the output of the dual-encoder architecture Φ is defined as follows:

$$\Phi(e, c) = \begin{cases} 0, & \text{if } \frac{\Psi(e)^T \Omega(c)}{\|\Psi(e)\| \|\Omega(c)\|} \leq \delta \\ 1, & \text{otherwise} \end{cases} \quad (1)$$

where $\Phi(e, c) = 0$ means that e is idiomatic in c , while $\Phi(e, c) = 1$ if e has a literal meaning in c . δ is a manually-tuned threshold. Both encoders are BERT-based architectures that take as input the

tokenized versions of expressions and their contexts, respectively, surrounded by the special tokens [CLS] and [SEP]. To encode an expression, we take the sum of the individual representations of all its subwords. Instead, for the context we take the representation of the [CLS] token.

3.2 Entity or Idiom?

As we discussed in the previous Section, semantic idiosyncrasy is essential for discriminating between idiomatic and literal expressions. However, there are cases in which the individual constituents of a potentially idiomatic expression are unrelated to the context, but the expression as used in that particular context is not idiomatic. Many of these cases correspond to named entities. Table 1 provides a selection of examples – extracted from the Sub-task A datasets – in which PIEs are named entities. For instance, in the first example, *Blood Bath* is a movie and, therefore, it does not have an idiomatic meaning. Nevertheless, its constituents (i.e. *blood* and *bath*) are unrelated to the context, hence misleading our dual-encoder architecture (Section 3.1) to classify it as idiomatic.

PIE	Context
blood bath	Deborah Loomis is an actress, known for Hercules in New York (1970), Foreplay (1975) and <i>Blood Bath</i> (1976).
fine line	<i>Fine Line</i> received generally positive reviews from music critics, particularly towards its production and stylistic influences.
monkey business	<i>Monkey Business</i> is an Action, Adventure, Comedy, Crime movie that was released in 1998 and has a run time of 1 hr 29 min.
rocket science	After finishing "Confrontation", the band shifted to " <i>Rocket Science</i> ".
night owl	Andrew Gonzalez, owner, <i>Night Owl</i> Cookies: "Nobody believed in me except for Deco Drive.", "They got me on air very quickly!"
silver spoon	Not only is it endorsed by the UK's biggest food brands – Weetabix, Shredded Wheat, <i>Silver Spoon</i> , Carling lager, Marriage's flour – but being Red Tractor also means you can supply different retailers without lots of different requirements.

Table 1: Examples of sentences where potentially idiomatic expressions (PIEs) are named entities.

In order to cope with this issue, we exploit Named Entity Recognition, i.e. the task of identifying specific words as belonging to predefined semantic types, such as Person, Location and Organization (Nadeau and Sekine, 2007). Specifically, we introduce an auxiliary NER module in our classification pipeline that, given as input a raw text sequence of n tokens $X = x_1, \dots, x_n$ containing a potentially idiomatic expression p , predicts all the entities $E = e_1, \dots, e_m$ in X . Then, if $p \in E$, p is labeled as literal, otherwise p is provided to the dual encoder, together with its context. To detect further entities, we also exploit capitalization.

4 Experiments

In this Section, we describe our experimental setup (Section 4.1), the datasets we use to train and evaluate our idiom identification system (Section 4.2), and the obtained results (Section 4.3).

4.1 Experimental Setup

We implement our dual-encoder architecture (Section 3.1) with PyTorch (Paszke et al., 2019), using the Transformers library (Wolf et al., 2019) to load the weights of BERT-base-cased for English and of BERT-base-portuguese-cased for Portuguese and Galician. We fine-tune our idiom identification system for 100 epochs with a Mean-Squared Error loss criterion, adopting an early stopping strategy with a patience value of 20, Adam

(Kingma and Ba, 2015) optimizer, and a learning rate of 10^{-5} . Additionally, we set $\delta = 0^4$, and use 32 as batch size, with 4 steps of gradient accumulation. To identify entities, instead, we employ wikineural-multilingual-ner⁵, a Multilingual BERT (mBERT) model fine-tuned on the WikiNEuRal dataset (Tedeschi et al., 2021b). We compare systems by means of their Macro F_1 scores, as specified by the competition rules. Our final scores are obtained by ensembling the predictions of $N = 9$ model checkpoints⁶ and taking the class with the highest number of votes.

Model training was carried out on a NVIDIA GeForce RTX 3090. Each training (i.e. for each model configuration) required ~ 1 min/epoch on average, for a mean of ~ 30 epochs.

4.2 Training, Validation and Test Data

The training, validation and test sets we use in our experiments are those provided for SubTask A⁷. Data statistics are provided in Table 2.

⁴We train our system to produce a cosine similarity score s between a MWE e and its context c , which is $s = -1$ when e is idiomatic in c , or $s = 1$ otherwise. Therefore, in Eq. 1, $\delta = 0$ means that negative similarity scores are mapped to 0 (*Idiomatic*), while positive scores are mapped to 1 (*Literal*).

⁵<https://huggingface.co/Babelscape/wikineural-multilingual-ner>

⁶We use the dev set to search for the optimal value of N by choosing from $N = \{1, 3, 5, 7, 9, 11, 13\}$.

⁷https://github.com/H-TayyarMadabushi/SemEval_2022_Task2-idiomaticity

Split	EN	PT	GL	Total
train-one-shot	73	73	63	209
train-zero-shot	3327	1164	0	4491
dev	466	273	0	739
test	916	713	713	2342

Table 2: Number of examples in the training, validation and test sets for each of the covered languages: English (EN), Portuguese (PT) and Galician (GL).

	System	EN	PT	GL	ALL
zero-shot	Baseline	70.1	68.0	50.7	65.4
	Our System w/o NER	76.4	63.5	59.7	69.9
	Our System	86.8	70.4	65.5	77.4
one-shot	Baseline	88.6	86.4	81.6	86.5
	Our System w/o NER	91.0	86.8	83.9	88.8
	Our System	95.8	88.9	87.4	92.1

Table 3: Results of our system with and without the inclusion of the NER module on English (EN), Portuguese (PT) and Galician (GL) languages using the Macro-F₁ score metric. The ALL column reports the overall results. The baselines are provided by task organizers.

In the *zero-shot* setting, potentially idiomatic expressions in the training set are completely disjoint from those in the validation and test sets. In the *one-shot* setting, instead, one positive and one negative example are included for each MWE in the test and validation sets. Finally, note that the zero-shot training set and the validation set cover only English and Portuguese languages, while the test set also contains the Galician language, hence further increasing the difficulty of the zero-shot setting.

4.3 Results

In preliminary experiments, we measure the impact that context inclusion (i.e., the sentences preceding and following the one containing the PIEs) has on our system’s performance. Similar to [Tayyar Madabushi et al. \(2021\)](#), we observe a slight drop in performance (i.e., -0.3 F₁ points, on average on the zero-shot and one-shot settings) and longer training times, hence we do not include context in our experiments. Then, in order to show the effectiveness of our dual-encoder architecture (Section 3.1) and of our entire idiomaticity detection system that includes the NER module (Section 3.2), we compare them with the strong mBERT-based baselines provided by the task organizers ([Tayyar Madabushi et al., 2022](#)): for the zero-shot setting, their model takes as input the context, while for the one-shot setting, they exclude the context and provide as

input only the sentence containing the PIE, where the latter is separated from the rest of the input by using the “[SEP]” special token.

In both zero-shot and one-shot settings, our system far exceeds the performance of the competitive baselines. Specifically, in the zero-shot setting we observe an average improvement of 12 F₁ points for the complete system (Figure 1), and of 4.5 F₁ points using only the dual-encoder architecture (Section 3.1). Likewise, in the one-shot setting we point out an average improvement of 5.6 F₁ points for the overall architecture, and of 2.3 F₁ points for the dual encoder. Therefore, the findings are twofold: i) dual encoders that exploit semantic idiosyncrasy discriminate well between idiomatic and literal expressions, and ii) an idiomaticity detection system can greatly benefit from the inclusion of a NER module in the classification pipeline to manage such ambiguous cases (cf. the examples in Table 1, extracted by using our NER classifier).

5 Conclusions

In this paper, we presented our NER4ID submission to SemEval-2022 Task 2 focusing on the Multilingual Idiomaticity Detection subtask. We started by exploiting the semantic idiosyncrasy property of idiomatic expressions and introduced a novel dual-encoder Transformer-based architecture that encodes both the potentially idiomatic expression (PIE) and its context, and based on their similarity predicts idiomaticity. Further, by manually inspecting our system’s errors we discovered critical cases in which, although the individual constituents of a PIE were unrelated to the context, the expressions were not idiomatic in that particular context in which they were used. In most of these cases, the PIEs were part of a named entity. Hence, our second main contribution was devoted to the inclusion of an auxiliary NER module in the idiomaticity detection pipeline in order to avoid these errors. Our experiments showed that: i) our dual-encoder architecture was able to successfully solve the idiom identification task by consistently outperforming the strong baselines provided by the task organizers, and ii) the inclusion of NER in the pipeline provided further improvements of up to 7.5 F₁ points.

As future work, we plan to follow the research line proposed by [Tedeschi et al. \(2022\)](#), and explore the identification of idioms directly on raw texts, i.e., without pre-identified potentially idiomatic expressions, and study a broader set of languages.

Acknowledgments

The authors gratefully acknowledge the support of the ERC Consolidator Grant MOUSSE No. 726487 under the European Union’s Horizon 2020 research and innovation programme.



References

- Michele Bevilacqua, Rexhina Blloshmi, and Roberto Navigli. 2021a. [One SPRING to rule them both: Symmetric AMR semantic parsing and generation without a complex pipeline](#). In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 12564–12573. AAAI Press.
- Michele Bevilacqua, Tommaso Pasini, Alessandro Raganato, and Roberto Navigli. 2021b. [Recent trends in word sense disambiguation: A survey](#). In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 4330–4338. International Joint Conferences on Artificial Intelligence Organization. Survey Track.
- Terra Blevins and Luke Zettlemoyer. 2020. [Moving down the long tail of word sense disambiguation with gloss informed bi-encoders](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1006–1017, Online. Association for Computational Linguistics.
- Jan A. Botha, Zifei Shan, and Daniel Gillick. 2020. [Entity Linking in 100 Languages](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7833–7845, Online. Association for Computational Linguistics.
- Chenhui Chu and Rui Wang. 2018. [A survey of domain adaptation for neural machine translation](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1304–1319, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Kenneth Church, W Gale, P Hanks, D Hindle, and Uri Zernik. 1991. [Lexical acquisition: Exploiting on-line resources to build a lexicon](#). *U. Zernik (Ed.)*, pages 115–164.
- Simone Conia, Andrea Bacciu, and Roberto Navigli. 2021. [Unifying cross-lingual semantic role labeling with heterogeneous linguistic resources](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 338–351, Online. Association for Computational Linguistics.
- Paul Cook, Afsaneh Fazly, and Suzanne Stevenson. 2007. [Pulling their weight: Exploiting syntactic forms for the automatic identification of idiomatic expressions in context](#). In *Proceedings of the Workshop on A Broader Perspective on Multiword Expressions*, pages 41–48, Prague, Czech Republic. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [Bert: Pre-training of deep bidirectional transformers for language understanding](#). In *NAACL-HLT (1)*.
- Mona Diab and Pravin Bhutada. 2009. [Verb noun construction mwe token classification](#). In *Proceedings of the Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications (MWE 2009)*, pages 17–22.
- Sergey Edunov, Myle Ott, Michael Auli, and David Grangier. 2018. [Understanding back-translation at scale](#). *arXiv preprint arXiv:1808.09381*.
- Rafael Ehren. 2017. [Literal or idiomatic? identifying the reading of single occurrences of German multiword expressions using word embeddings](#). In *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 103–112, Valencia, Spain. Association for Computational Linguistics.
- Samin Fakharian. 2021. [Contextualized embeddings encode knowledge of English verb-noun combination idiomaticity](#). Ph.D. thesis, University of New Brunswick.
- Afsaneh Fazly and Suzanne Stevenson. 2006. [Automatically constructing a lexicon of verb phrase idiomatic combinations](#). In *11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy. Association for Computational Linguistics.
- Marcos Garcia, Tiago Kramer Vieira, Carolina Scarton, Marco Idiart, and Aline Villavicencio. 2021. [Probing for idiomaticity in vector space models](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3551–3564, Online. Association for Computational Linguistics.
- Waseem Gharbieh, Virendra Bhavsar, and Paul Cook. 2016. [A word embedding approach to identifying verb-noun idiomatic combinations](#). In *Proceedings of the 12th Workshop on Multiword Expressions*, pages 112–118, Berlin, Germany. Association for Computational Linguistics.
- Reyhaneh Hashempour and Aline Villavicencio. 2020. [Leveraging contextual embeddings and idiom principle for detecting idiomaticity in potentially idiomatic expressions](#). In *Proceedings of the Workshop on the Cognitive Aspects of the Lexicon*, pages 72–80.

- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Murathan Kurfalı and Robert Östling. 2020. [Disambiguation of potentially idiomatic expressions with contextual embeddings](#). In *Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons*, pages 85–94, online. Association for Computational Linguistics.
- Changsheng Liu and Rebecca Hwa. 2019. [A generalized idiom usage recognition model based on semantic compatibility](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6738–6745.
- Xiaodong Liu, Kevin Duh, Liyuan Liu, and Jianfeng Gao. 2020. [Very deep transformers for neural machine translation](#).
- Amit Mishra and Sanjay Kumar Jain. 2016. [A survey on question answering systems with classification](#). *Journal of King Saud University-Computer and Information Sciences*, 28(3):345–361.
- Grace Muzny and Luke Zettlemoyer. 2013. [Automatic idiom identification in Wiktionary](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1417–1421, Seattle, Washington, USA. Association for Computational Linguistics.
- David Nadeau and Satoshi Sekine. 2007. [A survey of named entity recognition and classification](#). *Linguisticae Investigationes*, 30(1):3–26.
- Vasudevan Nedumpozhimana and John Kelleher. 2021. [Finding bert’s idiomatic key](#). In *Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021)*, pages 57–62.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. [Pytorch: An imperative style, high-performance deep learning library](#). *Advances in neural information processing systems*, 32.
- Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep contextualized word representations](#). *arXiv preprint arXiv:1802.05365*.
- Carlos Ramisch, Aline Villavicencio, Leonardo Moura, and Marco Idiart. 2008. [Picking them up and figuring them out: Verb-particle constructions, noise and idiomaticity](#). In *CoNLL 2008: Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 49–56, Manchester, England. Coling 2008 Organizing Committee.
- Marco Silvio Giuseppe Senaldi, Yuri Bizzoni, and Alessandro Lenci. 2019. [What do neural networks actually learn, when they learn to identify idioms?](#) *Proceedings of the Society for Computation in Linguistics*, 2(1):310–313.
- Ekaterina Shutova, Lin Sun, and Anna Korhonen. 2010. [Metaphor identification using verb and noun clustering](#). In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1002–1010, Beijing, China. Coling 2010 Organizing Committee.
- Harish Tayyar Madabushi, Edward Gow-Smith, Marcos Garcia, Carolina Scarton, Marco Idiart, and Aline Villavicencio. 2022. [SemEval-2022 Task 2: Multilingual Idiomaticity Detection and Sentence Embedding](#). In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Association for Computational Linguistics.
- Harish Tayyar Madabushi, Edward Gow-Smith, Carolina Scarton, and Aline Villavicencio. 2021. [AStitchInLanguageModels: Dataset and methods for the exploration of idiomaticity in pre-trained language models](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3464–3477, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Simone Tedeschi, Simone Conia, Francesco Cecconi, and Roberto Navigli. 2021a. [Named Entity Recognition for Entity Linking: What works and what’s next](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2584–2596, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Simone Tedeschi, Valentino Maiorca, Niccolò Campolungo, Francesco Cecconi, and Roberto Navigli. 2021b. [WikiNEuRal: Combined neural and knowledge-based silver data creation for multilingual NER](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2521–2533, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Simone Tedeschi, Federico Martelli, and Roberto Navigli. 2022. [ID10M: Idiom Identification in 10 Languages](#). In *Findings of the Association for Computational Linguistics: NAACL 2022*, Seattle, Washington. Association for Computational Linguistics.
- Rakesh Verma and Vasanthi Vuppuluri. 2015. [A new approach for idiom identification using meanings and the web](#). In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 681–687, Hissar, Bulgaria. INCOMA Ltd. Shoumen, BULGARIA.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. [Huggingface’s transformers: State-of-the-art natural language processing](#). *arXiv preprint arXiv:1910.03771*.