Exploring the Use of Foundation Models for Named Entity Recognition and Lemmatization Tasks in Slavic Languages

Gabriela Pałka and **Artur Nowakowski***
Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznań, Poland
{gabriela.palka,artur.nowakowski}@amu.edu.pl

Abstract

This paper describes Adam Mickiewicz University's (AMU) solution for the 4th Shared Task on SlavNER. The task involves the identification, categorization, and lemmatization of named entities in Slavic languages. Our approach involved exploring the use of foundation models for these tasks. In particular, we used models based on the popular BERT and T5 model architectures. Additionally, we used external datasets to further improve the quality of our models. Our solution obtained promising results, achieving high metrics scores in both tasks. We describe our approach and the results of our experiments in detail, showing that the method is effective for NER and lemmatization in Slavic languages. Additionally, our models for lemmatization will be available at: https://huggingface.co/amu-cai.

1 Introduction

Named entity recognition and lemmatization are important tasks in natural language processing. Fine-tuning pre-trained neural language models has become a popular approach to achieve the best results in these tasks. However, the performance of this method can vary across languages and language families. In this paper, we investigate the performance of fine-tuned, language-specific neural language models in named entity recognition and lemmatization in a set of Slavic languages and compare them with multilingual solutions.

We describe Adam Mickiewicz University's (AMU) solution for the 4th Shared Task on SlavNER, which is a part of The 9th Workshop on Slavic Natural Language Processing (Slavic NLP 2023). Our solution is based on foundation models (Bommasani et al., 2021). In particular, we used models based on the popular BERT and T5 model architectures. To increase the effectiveness of our approach, we conducted experiments with different versions of monolingual and multilingual models, investigating the potential benefits of each model variant for specific tasks. The data provided by the organizers and external resources used for named entity recognition and lemmatization were processed and prepared as described in section 2. Specific details regarding the approach are further discussed in section 3.

In order to evaluate the effectiveness of our method, we performed several experiments on the previous Shared Task edition test set. This particular set was chosen because it is a well-known benchmark for named entity recognition and lemmatization in Slavic languages. The results of our experiments are described in section 4.

2 Data

This section provides a brief description of the datasets used in our solution. In addition to the data released by the organizers, we also used external datasets for named entity recognition and lemmatization. All training and validation samples containing named entities were converted to a CoNLL-2003 dataset format (Tjong Kim Sang and De Meulder, 2003).

2.1 Shared Task Dataset

The 4th Shared Task on SlavNER focuses on recognition, lemmatization, and cross-lingual linking of named entities in Polish, Czech and Russian languages. The training and validation data provided by the organizers come from the previous editions of the Shared Task and consist of news articles related to a single entity or event such as Asia Bibi, Brexit, Ryanair, Nord Stream, COVID-19 pandemic and USA 2020 Elections. The documents contain annotations of the following named entities: person (PER), location (LOC), organization (ORG), event (EVT) and product (PRO) (Piskorski et al., 2021).

*Artur Nowakowski is a scholarship recipient of the Adam Mickiewicz University Foundation for the 2022/2023 academic year.

Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023), pages 165–171

To obtain NER training and validation samples in the CoNLL-2003 format, we processed the data using the code provided by the Tilde team (Vīksna and Skadina, 2021)[1].

2.2 External NER Datasets

One way to improve the performance of NER models is to use external NER datasets to increase the volume of the training data. These datasets contain pre-labeled documents that have been annotated with named entities, and can be used to fine-tune existing models. This technique allows the model to learn from the additional data, which can provide a more comprehensive understanding of the context and complexities of the named entities.

2.2.1 Collection3

The *Collection3* dataset (Mozharova and Loukachevitch, 2016) is based on *Persons-1000*, a publicly available Russian document collection consisting of 1,000 news articles. Currently, the dataset contains 26,000 annotated named entities (11,000 persons, 7,000 locations and 8,000 organizations).

2.2.2 MultiNERD

The *MultiNERD* dataset (Tedeschi and Navigli, 2022) covers 10 languages, including Polish and Russian, and contains annotations of multiple NER categories, from which we extracted categories present in the Shared Task. The labels were obtained by processing the Wikipedia and Wikinews articles. In addition, the sentences were tagged automatically, in a way that can also be adapted to the Czech language.

2.2.3 Polyglot-NER

A *Polyglot-NER* dataset (Al-Rfou et al., 2015) covers 40 languages, including Polish, Czech and Russian. The annotations were automatically generated from Wikipedia and Freebase. The obtained entity categories are: person, location and organization.

2.2.4 WikiNEuRal

The *WikiNEuRal* dataset (Tedeschi et al., 2021) consists of named entities in the following categories: person, location, organization and miscellaneous. Wikipedia was used as the source for the labels, which were automatically obtained using a combination of knowledge-based approaches and neural models. The datasets cover 9 languages, including Polish and Russian.

[1] https://github.com/tilde-nlp/BSNLP_2021

2.3 External Lemmatization Datasets

Lemmatization, the process of reducing a word or phrase to its base form, is an essential component, especially for tasks such as information retrieval and text mining. External lemmatization datasets can improve the quality of lemmatization models by providing additional training samples that contain more inflectional variants of phrases. Such data consists of inflected words, collocations or phrases with corresponding lemmatized forms.

2.3.1 SEJF

SEJF (Czerepowicka and Savary, 2018) is a linguistic resource consisting of a grammatical lexicon of Polish multi-word expressions. It contains two modules: an intensional module, which consists of 4,700 multiword lemmas assigned to 100 inflection graphs, and an extensional module, which contains 88,000 automatically generated inflected forms annotated with grammatical tags.

2.3.2 SEJFEK

SEJFEK (Savary et al., 2012) refers to a lexical and grammatical resource related to Polish economic terms. It contains a grammatical lexicon module with over 11,000 terminological multi-word units and a fully lexicalized shallow grammar with over 146,000 inflected forms, which was produced by an automatic conversion of the lexicon.

2.3.3 PolEval 2019: Task 2

PolEval 2019: Task 2 (Marcińczuk and Bernaś, 2019) is a part of a workshop focusing on natural language processing in the Polish language. The main goal of this task was to lemmatize proper names and multi-word phrases. The train set consists of over 24,000 annotated and lemmatized phrases. The validation set and the test set contain 200 and 1,997 phrases, respectively.

2.3.4 Machine Translation of External Datasets

Due to the lack of external Czech and Russian datasets dedicated to lemmatization tasks, we decided to use OPUS-MT (Tiedemann and Thottingal, 2020), which is a resource containing open-source machine translation models. We machine translated all the samples prepared from the three aforementioned datasets.

3 Approach

We participated in the two subtasks of the Multilingual Named Entity Recognition Task - *Named Entity Mention Detection and Classification* and *Named Entity Lemmatization*. The solution involved fine-tuning the foundation models using task-specific modifications and additional training data. All models used in the experiments can be found on the Hugging Face Hub[2].

3.1 Named Entity Recognition

Recently, the BERT (Devlin et al., 2019) model architecture has been adapted to address Slavic languages such as Polish, Czech and Russian, among others. These languages present unique challenges because of their complex grammatical structures, declensions and inflections, making NLP tasks even more difficult. However, the application of BERT to these languages has resulted in significant improvements in language processing and understanding.

In our solution, we used several monolingual BERT models to better handle the specific linguistic nuances of individual Slavic languages. In particular, we employed of the following models: HerBERT (Mroczkowski et al., 2021) for Polish, Czert (Sido et al., 2021) for Czech and RuBERT (Kuratov and Arkhipov, 2019) for Russian. For comparison, we also used multilingual BERT models that can handle multiple languages, including Slavic BERT (Arkhipov et al., 2019) and XLM-RoBERTa (Conneau et al., 2020).

In the experiments, we also added a Conditional Random Fields (CRF) layer on the top of each BERT model. A similar approach of combining CRF with neural networks has been used previously (Lample et al., 2016), as the CRF layer can capture the dependencies between neighboring tokens and provide a smoother transition between different entity types.

3.2 Lemmatization

Models based on the T5 (Raffel et al., 2020) model architecture have achieved state-of-the-art results in various natural language processing challenges and can be fine-tuned for specific tasks. One of the applications of T5 can be lemmatization, the process of reducing a word or phrase to its basic form (lemma). In Slavic languages such as Polish, Czech and Russian, lemmatization is particularly important due to the complex inflection of these languages.

We approached the lemmatization task as a text-to-text problem. The input to the model is an inflected phrase or named entity, which can consist of several word forms. For example, it can consist of nouns in singular or plural form, or verbs in different tenses. The output of the model is the base, normalized form of the phrase or named entity.

To address the lack of dedicated models for Czech and Russian, we used one monolingual and a multilingual T5 model. Specifically, we chose plT5 (Chrabrowa et al., 2022) for Polish and mT5 (Xue et al., 2021) for multilingual experiments. For comparison purposes, we also conducted our experiments on the small, base and large sizes of the above models.

In the multilingual experiments, we included a language token (»pl«, »cs«, »ru«) as the first token of the source phrases, depending on the language of the phrase. Our preliminary experiments have shown that incorporating the language token improves the results, increasing the exact match by approximately 2 points in each language. We noticed that the model sometimes tends to change the grammatical number from plural to singular - possibly due to the fact that singular named entities occur more often in the training data.

4 Results

4.1 Named Entity Recognition Results

The results of our named entity recognition experiments are presented in table 1. We evaluated our models with a case-sensitive F1 score, which is a standard span-level metric calculated on the ConLL-2003 dataset format. As test sets, we choose COVID-19 and USA 2020 Elections subsets of the 3rd Shared Task on SlavNER.

We tested our solution in two approaches: monolingual and multilingual. For Polish and Czech, we found that monolingual models perform better for language-specific data. In the case of Russian, multilingual models strongly outperform language-specific solutions. We assume that this is due to the lack of sufficient data for this language. In addition, multilingual models can learn common rules in Slavic languages to overcome weaknesses related to insufficient data.

We also found that adding a CRF layer significantly improves the quality of the models in most cases. However, including external datasets wors-

[2] https://huggingface.co/models

Model	original data						+ external datasets					
	COVID-19			USA 2020 Elections			COVID-19			USA 2020 Elections		
	pl	cs	ru	pl	cs	ru	pl	cs	ru	pl	cs	ru
HerBERT$_{\text{BASE}}$	79.50	-	-	89.27	-	-	78.70	-	-	84.63	-	-
HerBERT$_{\text{BASE}}$ + CRF	80.11	-	-	90.16	-	-	80.86	-	-	87.43	-	-
HerBERT$_{\text{LARGE}}$	81.18	-	-	91.71	-	-	81.29	-	-	89.83	-	-
HerBERT$_{\text{LARGE}}$ + CRF	81.75	-	-	**92.13**	-	-	**82.33**	-	-	89.20	-	-
Czert	-	84.10	-	-	88.82	-	-	73.05	-	-	84.06	-
Czert + CRF	-	**84.22**	-	-	**90.29**	-	-	71.36	-	-	83.70	-
RuBERT	-	-	**62.06**	-	-	76.97	-	-	58.51	-	-	77.63
RuBERT + CRF	-	-	61.80	-	-	**77.69**	-	-	59.55	-	-	76.72
Slavic-BERT	79.06	78.67	61.42	89.07	90.31	78.21	73.73	68.22	59.32	83.72	78.16	77.29
Slavic-BERT + CRF	78.15	80.68	63.08	89.97	90.13	78.72	77.76	69.12	58.08	86.76	80.51	77.05
XLM-RoBERTa$_{\text{BASE}}$	79.53	77.89	62.12	88.30	89.51	77.56	76.92	68.46	60.45	83.25	80.89	77.21
XLM-RoBERTa$_{\text{BASE}}$ + CRF	81.10	78.80	65.94	88.48	90.88	77.58	79.45	73.42	58.86	87.02	84.20	76.87
XLM-RoBERTa$_{\text{LARGE}}$	81.43	80.58	**66.26**	**90.36**	**91.62**	**80.22**	81.12	75.35	61.95	87.46	86.96	77.60
XLM-RoBERTa$_{\text{LARGE}}$ + CRF	**81.81**	**81.20**	64.95	89.37	91.53	79.93	80.72	75.01	61.80	86.78	87.66	77.73

Table 1: Results of case-sensitive F1 score for named entity recognition on the COVID-19 and USA 2020 Elections test sets from the 3rd Shared Task on SlavNER. For each language in a given test set, the best score for the monolingual and multilingual solution is shown in bold. In addition, the best score for each language in a given test set is underlined.

		original data			+ PolEval 2019			+ Lexicon		
		pl	cs	ru	pl	cs	ru	pl	cs	ru
	COVID-19									
Model	*Size*									
plT5	small	86.36	-	-	91.15	-	-	92.02	-	-
	base	89.99	-	-	93.03	-	-	80.70	-	-
	large	94.05	-	-	94.78	-	-	**95.36**	-	-
mT5	small	74.46	73.75	70.17	86.80	80.98	73.83	81.13	75.45	71.84
	base	87.66	85.44	76.96	91.00	86.29	76.10	90.42	83.32	75.30
	large	90.57	88.84	**79.09**	93.76	**89.80**	77.30	93.03	89.27	77.16
USA 2020 Elections										
Model	*Size*									
plT5	small	83.37	-	-	87.47	-	-	86.65	-	-
	base	85.22	-	-	87.89	-	-	76.80	-	-
	large	90.97	-	-	90.76	-	-	**91.38**	-	-
mT5	small	71.46	70.03	72.18	78.85	75.86	76.18	74.54	69.76	68.92
	base	83.98	80.37	80.51	84.19	81.97	80.27	85.63	78.78	78.25
	large	88.71	**88.33**	**82.86**	89.12	87.27	82.50	89.94	86.74	81.76

Table 2: Results of the case-insensitive exact match for lemmatization on the COVID-19 and USA 2020 Elections test sets from the 3rd Shared Task on SlavNER. For each test set, the best score in a given language is shown in bold and underlined.

ens the results in almost all cases. We suspect that this is due to the specific domain of the test sets, which are news articles. In addition, some annotation errors can be found in all datasets presented in the 2.2 section.

4.2 Lemmatization Results

The results of our lemmatization experiments are presented in the table 2. We evaluated our models with a case-insensitive exact match on the same test sets as for named entity recognition, but only on the data specific to this task.

We tested our solution based on two models: a monolingual plT5 (only for the Polish language), and a multilingual mT5 model. We observed that the addition of each external dataset significantly improves the quality of the Polish language-specific model. Moreover, the addition of the data from PolEval 2019 also improves the results for the multilingual model. Unfortunately, the addition of data from the lexicon generated by machine translation of the SEJF and SEJFEK datasets causes a decrease

Submission	Recognition			Normalization		
	pl	cs	ru	pl	cs	ru
System 1	83.33	88.08	84.30	80.27	76.62	79.32
System 2	**85.37**	**89.70**	**86.16**	**82.37**	**76.89**	81.27
System 3	83.40	85.19	82.77	80.32	73.06	**81.47**
System 4	83.33	81.70	79.20	80.27	71.11	76.84

Table 3: Results of our systems on the released test set for named entity recognition and normalization (lemmatization). The scores are computed as case-insensitive strict matching for recognition and case-insensitive F1 score for normalization. All scores were received from the organizers.

in the model performance for the Czech and Russian languages. We assume that this is due to the quality of the translation of the phrases into these languages.

We also noticed that the quality of the lemmatization improves as the size of the model increases in almost all cases. However, for Polish, the small model trained on all available data is better than the base model. Furthermore, it is only 3 points worse than the large model, so it can be used efficiently considering the hardware limitations.

4.3 The 4th Shared Task on SlavNER Results

The current edition of the shared task features news articles about the Russian-Ukrainian war, and the test set includes raw texts in Polish, Czech and Russian languages.

As a solution, we submitted four systems consisting of the following fine-tuned models with an additional CRF layer for named entity recognition:

- System 1: HerBERT$_{\text{LARGE}}$ for Polish trained on all available data, Czert for Czech and RuBERT for Russian trained only on the data provided by the organizers,
- System 2: XLM-RoBERTa$_{\text{LARGE}}$ for all languages trained only on the data provided by the organizers,
- System 3: XLM-RoBERTa$_{\text{LARGE}}$ for all languages trained on all available data,
- System 4: HerBERT$_{\text{LARGE}}$ for Polish, Czert for Czech and RuBERT for Russian trained on all available data.

In all the systems mentioned above, we used the following lemmatization models: plT5$_{\text{LARGE}}$ for Polish (trained on all available data) and mT5$_{\text{LARGE}}$ for Czech and Russian (trained on the data provided by the organizers and the data from PolEval 2019 Task 2).

The best solution for recognizing and categorizing named entities turned out to be System 2, which also achieved the best results for normalization (lemmatization). In addition, the normalization scores are highly dependent on the NER results, since only recognized entities are normalized.

5 Conclusions

We described the Adam Mickiewicz University's (AMU) participation in the 4th Shared Task on SlavNER for named entity recognition and lemmatization tasks. Our experiments encompassed various foundation models, including monolingual and multilingual BERT and T5 models. We found that incorporating a CRF layer enhanced the quality of our named entity recognition models. Additionally, our results indicate that the use of T5 models for lemmatization yields high-quality lemmatization of named entities. We will release the lemmatization models to the community and make them available at: https://huggingface.co/amu-cai.

References

Rami Al-Rfou, Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. 2015. Polyglot-NER: Massive multilingual named entity recognition. *Proceedings of the 2015 SIAM International Conference on Data Mining, Vancouver, British Columbia, Canada, April 30- May 2, 2015*.

Mikhail Arkhipov, Maria Trofimova, Yuri Kuratov, and Alexey Sorokin. 2019. Tuning multilingual transformers for language-specific named entity recognition. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 89–93, Florence, Italy. Association for Computational Linguistics.

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S.

Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, S. Buch, Dallas Card, Rodrigo Castellon, Niladri S. Chatterji, Annie S. Chen, Kathleen A. Creel, Jared Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren E. Gillespie, Karan Goel, Noah D. Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas F. Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, O. Khattab, Pang Wei Koh, Mark S. Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Benjamin Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, J. F. Nyarko, Giray Ogut, Laurel J. Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Robert Reich, Hongyu Ren, Frieda Rong, Yusuf H. Roohani, Camilo Ruiz, Jack Ryan, Christopher R'e, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishna Parasuram Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei A. Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2021. On the opportunities and risks of foundation models. *ArXiv*, abs/2108.07258.

Aleksandra Chrabrowa, Łukasz Dragan, Karol Grzegorczyk, Dariusz Kajtoch, Mikołaj Koszowski, Robert Mroczkowski, and Piotr Rybak. 2022. Evaluation of transfer learning for Polish with a text-to-text model. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4374–4394, Marseille, France. European Language Resources Association.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard 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, Online. Association for Computational Linguistics.

Monika Czerepowicka and Agata Savary. 2018. Sejf - a grammatical lexicon of polish multiword expressions. In *Human Language Technology. Challenges for Computer Science and Linguistics*, pages 59–73, Cham. Springer International Publishing.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Yuri Kuratov and Mikhail Arkhipov. 2019. Adaptation of deep bidirectional multilingual transformers for russian language. *ArXiv*, abs/1905.07213.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California. Association for Computational Linguistics.

Michał Marcińczuk and Tomasz Bernaś. 2019. Results of the poleval 2019 task 2: Lemmatization of proper names and multi-word phrases.

Valerie Mozharova and Natalia Loukachevitch. 2016. Two-stage approach in russian named entity recognition. In *2016 International FRUCT Conference on Intelligence, Social Media and Web (ISMW FRUCT)*, pages 1–6.

Robert Mroczkowski, Piotr Rybak, Alina Wróblewska, and Ireneusz Gawlik. 2021. HerBERT: Efficiently pretrained transformer-based language model for Polish. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 1–10, Kyiv, Ukraine. Association for Computational Linguistics.

Jakub Piskorski, Bogdan Babych, Zara Kancheva, Olga Kanishcheva, Maria Lebedeva, Michał Marcińczuk, Preslav Nakov, Petya Osenova, Lidia Pivovarova, Senja Pollak, Pavel Přibáň, Ivaylo Radev, Marko Robnik-Sikonja, Vasyl Starko, Josef Steinberger, and Roman Yangarber. 2021. Slav-NER: the 3rd cross-lingual challenge on recognition, normalization, classification, and linking of named entities across Slavic languages. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 122–133, Kyiv, Ukraine. Association for Computational Linguistics.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Agata Savary, Bartosz Zaborowski, Aleksandra Krawczyk-Wieczorek, and Filip Makowiecki. 2012. SEJFEK - a lexicon and a shallow grammar of Polish economic multi-word units. In *Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon*, pages 195–214, Mumbai, India. The COLING 2012 Organizing Committee.

Jakub Sido, Ondřej Pražák, Pavel Přibáň, Jan Pašek, Michal Seják, and Miloslav Konopík. 2021. Czert – Czech BERT-like model for language representation. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1326–1338, Held Online. INCOMA Ltd.

Simone Tedeschi, Valentino Maiorca, Niccolò Campolungo, Francesco Cecconi, and Roberto Navigli. 2021. 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 and Roberto Navigli. 2022. MultiNERD: A multilingual, multi-genre and fine-grained dataset for named entity recognition (and disambiguation). In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 801–812, Seattle, United States. Association for Computational Linguistics.

Jörg Tiedemann and Santhosh Thottingal. 2020. OPUS-MT – building open translation services for the world. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 479–480, Lisboa, Portugal. European Association for Machine Translation.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.

Rinalds Vīksna and Inguna Skadina. 2021. Multilingual Slavic named entity recognition. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 93–97, Kyiv, Ukraine. Association for Computational Linguistics.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.