Bulgarian ParlaMint 4.0 Corpus as a Testset for Part-of-Speech Tagging and Named Entity Recognition

Kiril Simov, Petya Osenova
Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
{kivs, petya}@bultreebank.org

Abstract

The paper discusses some fine-tuned models for the tasks of part-of-speech tagging and named entity recognition. The fine-tuning was performed on the basis of an existing BERT pre-trained model and two newly pre-trained BERT models for Bulgarian that are cross-tested on the domain of the Bulgarian part of the ParlaMint corpora as a new domain. In addition, a comparison has been made between the performance of the new fine-tuned BERT models and the available results from the Stanza-based model which the Bulgarian part of the ParlaMint corpora has been annotated with. The observations show the weaknesses in each model as well as the common challenges.

Keywords: BERT model, Bulgarian, parliamentary sessions, domain cross-validation, POS tagging, NER

1. Introduction

The Bulgarian Parliamentary Corpus is part of the ParlaMint 4.0 multilingual corpus of 29 national and regional parliaments in Europe (Erjavec et al., 2023). The data is publicly available for usage through the CLARIN.SI repository[1]. Our plans for the further development of the Bulgarian ParlaMint Corpus (BParlC) go in two main directions: an extension of the phenomena covered by the annotations of the corpus as well as an extension of BParlC with additional data. Concerning the latter direction, the additional data will include: (1) diachronically available editions of parliament activity to cover the period after the liberation of Bulgaria from Ottoman Empire in 1878; (2) additional documents such as records of debates in the Parliamentary Committees; (3) similar corpora related to Municipality Councils for some economically influential cities in Bulgaria; (4) documents of political parties; (5) linking to the Bulgaria-centric Knowledge Graph (BGKG); and others.

Concerning the former direction, our first goal is to perform linking of the named entities within the texts of BParlC with the Bulgaria-centric Knowledge Graph. However, in order to achieve this task in the best possible way, we decided to check the quality of the annotation already present in BParlC, and to implement some processing tools for Bulgarian which to improve the current annotations — especially the part-of-speech tagging (POS) (UPOS for tagging with Universal Universal POS tags[2] and the XPOS for tagging with BulTreeBank POS tags[3]) as well as the Named Entity Recognition (NER).

Thus, our immediate aim is to test the BERT pre-trained models for POS tagging and NER tasks in a cross-domain setting. We had at our disposal a BERT model, already pre-trained over newsmedia texts – BERT-WEB-BG[4] — see Marinova et al. (2023) – and two new BERT models BERT-NEWS-LIT-BG-1 and BERT-NEWS-LIT-BG-2 which were pre-trained especially for the purposes of these experiments on more and other newsmedia texts as well as on additions of fictional texts (original Bulgarian and Translated into Bulgarian Foreign Literature).

The fine-tuning was performed on the two datasets for POS tagging (the BulTreeBank Dataset and the CLaDA-BG-POS Dataset[5]) as well as on the two datasets for NER (the BulTreeBank dataset and the Bulgarian Balto-Slavic Dataset). We first fine-tuned the models for each task mentioned above, and evaluate them with respect to the test sets within the corresponding dataset. Additionally, we evaluate the fine-tuned models on a new genre of texts — parliament debates.

Our work somewhat relates to the task of Domain Adaptation (DA) in sequential labeling tasks. However, at this stage we did not use the strategy of adding some in-domain data either in the pre-training phase or in the fine-tuned model to check whether it would improve for the target domain. This

[1] https://www.clarin.eu/parlamint#parlamint-corpora

[2] https://universaldependencies.org/u/pos/all.html

[3] http://bultreebank.org/wp-content/uploads/2017/04/BTB-TR03.pdf

[4] https://huggingface.co/usmiva/bert-web-bg

[5] This dataset is under development within the Bulgarian Infrastructure project CLaDA-BG: https://clada-bg.eu/en/

ParlaCLARIN IV Workshop on Creating, Analysing, and Increasing Accessibility of Parliamentary Corpora, pages 30–35

setting remains for future work. In addition, a semi-automatic comparison through human checks has been made between the performance of our BERT model and the Stanza-based one.

The motivation behind such a task includes the following aspects: i) improving the quality and the coverage of the BERT models for the two above-mentioned tasks, and (ii) evaluating the applicability of the models with respect to a different domain.

At first sight it might seem that part-of-speech tagging and named entity recognition are already solved tasks to a great extent. And this is true already for many languages and domains since the SOTA results are beyond 90 % F-measure (even beyond 95 %). However, it would be useful to track the systemic and occasional errors in the remaining percentages of unrecognized or wrongly annotated tokens in the data.

The paper is structured as follows: in the next section a brief overview is given of related work. Section 3 outlines the experimental setting. Section 4 discusses the results and provides quality comparison between the two models. Section 5 presents the conclusions.

2. Related Work

For the POS tagging there are a number of works that evaluate taggers' F-measure out-of-domain. For example, Schnabel and Schütze (2013) show that there is no single representation and method that works equally well for all target domains. In the target domains that were considered in the paper there is politics or parliamentary data.

Then, Hansen and van der Goot (2023) evaluate the performance of two taggers for English in a domain different from the Wall Street Journal section of the Penn Treebank, namely – on video games related dataset. Authors conclude that the accuracy on unknown tokens decreases and that the main problems are with the proper nouns and inconsistent capitalization.

In (Kübler and Baucom, 2011) a fast method for adding in-domain training data has been proposed that uses three taggers trained on the source data and run on the target unannotated data. The source domain is the Wall Street Journal part of the Penn Treebank, and the target one consists of dialogues in a collaborative task. The authors add sentences to the training data only when the majority of the taggers agree on the POS tags.

For NER also there is a lot of work devoted to its handling in a cross- and/or out-of-domain setting. For example, Liu et al. (2020) introduce a cross-NER dataset that comprises five domains among which politics. The authors provide specific NE for each domain. For the politics they are: politician, person, organization, political party, event, election, country, location, miscellaneous. The authors find that this domain overlaps mostly with the Reuters domain, i.e. newsmedia (35.7%). With BERT on English they report an integrated F1 on token level of 68.83.

Later on Zheng et al. (2022) use a graph matching method that learns graph structure via matching label graphs from source to target domain, and improve these results on all domains among which the domain of politics. In contrast to Liu et al. (2020) we do not use parliament data in the training phase but only in the pre-training one, thus making the task slightly harder. On the other hand, we facilitate our work by applying the standard set of categories for NER used in the corresponding datasets: Person, Organization, Location, and Other for Bultreebank dataset and Person, Organization, Location, Event, and Product for Balto-Slavic dataset.

3. The Experimental Setting

In this section we present the characteristics of the models as well as the specifics of the datasets that were used in the experiments.

3.1. BERT Pre-trained Models

In the experiments we exploited one of the existing models released for free public usage — BERT-WEB-BG. This model has been pre-trained on 30 GB of text which we estimated to comprise 3 536 668 132 tokens. The domain was newsmedia data. However, in order to check the impact of the text types used in the pre-training, we decided to pre-train a new model with a size of 2 192 734 242 tokens, from which about 800 000 000 tokens are fiction and the rest are newsmedia data. The other parameters have not been changed, including the number of epochs. In Table 1 the characteristics of the pre-training datasets are given.

3.2. BERT Fine-tuning Datasets

BulTreeBank Datasets In our experiments the following datasets were used:

1. *BulTreeBank-UD (BTB-UD)*. The BulTreeBank in its Universal Dependency format comprises data of 156K in tokens. It contains annotation for POS tagging divided into two: UPOS - annotation with Universal parts-of-speech and XPOS - annotation with the original BulTreeBank tags. The dataset follows the division into training and test sets in the Universal Dependencies package[6].

2. *BulTreeBank-NER (BTB-NER)*. The BTB-NER used the original BulTreeBank resource that

[6] https://universaldependencies.org/

Dataset	Size in GB	Size in tokens	Loss	Accuracy
BERT-WEB-BG	30.0	3 536 668 132	1.451	0.6906
BERT-NEWS-LIT-BG-1	18.6	2 192 734 242	2.153	0.5593
BERT-NEWS-LIT-BG-2	18.6	2 192 734 242	1.414	0.6913

Table 1: Characteristics of the pre-trained models. The dataset for the training of BERT-WEB-BG is proprietary and for that reason the exact size in tokens is not available to us. Thus, we estimated it on the basis of the other datasets. The main differences between BERT-NEWS-LIT-BG-1 and BERT-NEWS-LIT-BG-2 are the following hyper parameters: the hidden size of the first model is the default 768, but it is 1024 in the second model; the number of the attention heads is 12 for the first model, and 16 for the second one; the intermediate size of the first model is 3072, and 4096 for the second one respectively. The sizes of the parameters in the models are as follows: 109 113 649 parameters for the first one, and 183 485 745 parameters for the second.

is constituency-based and dependency-aware. Thus, it used the data of 256K in tokens. It includes four kinds of Named Entities: PER (persons), LOC (locations), ORG (organization), and OTH (other names). The treebank consists of 40 sets of sentences. Some of these sets are just small segments of texts extracted from different sources like Bulgarian grammar books, random paragraphs from corpora. Other sets are whole articles or other genres like newspaper articles, chapters of books, Bulgarian constitution, etc. The division in training, development, and test sets was performed on the basis of the whole sets of the treebank in the proportion of 80 % training set, 10 % development set and 10 % test set.

The additional two datasets used for fine-tuning are the following:

3. *Bulgarian Balto-Slavic Dataset (BS-NER)*. The datasets are from years 2019 (Piskorski et al., 2019) and 2021 (Piskorski et al., 2021). This integrated dataset contains annotations of Named Entities in the following categories: PER (person), ORG (organization), LOC (location), EVN (event) and PRO (product). The dataset follows the division of training and test sets as described in (Hardalov et al., 2023).

4. *CLaDA-POS Dataset*. This is a newly annotated dataset created within CLaDA-BG research infrastructure. The texts are collected from different sources. They include all the definitions from BTB-WN: the BTB Bulgarian Wordnet — see (Simov and Osenova, 2023), all the examples related to meanings from BTB-WN, newsmedia documents, first paragraphs of about 1000 articles from the Bulgarian Wikipedia. The dataset is divided into training, development and test sets by us for the purposes of these experiments.

It can be seen that no parliamentary data was used in the fine-tuning step due to the lack of sufficient gold data. At the same time, only some of the characteristics that can be found in the parliament corpora, are already present albeit in small portions in the above-mentioned datasets. This means that there are politically oriented topics and named entities that refer to politicians, especially in the newsmedia data.

4. Results

In order to evaluate different models for POS tagging and NER over ParlaMint data we performed fine-tuning of the two pre-trained models on the above described datasets for these tasks. The results from the first experiments are given in Table 2 where some evaluation was performed within the same datasets.

It can be seen that the best F1 measure metrics for all the tasks was achieved by BERT-NEWS-LIT-BG-2 model. These results reflect the increase in both metrics - Precision and Recall. This means that the more parameters and the more context included, the better the results. Table 2 also shows that concerning Precision and Recall, the two new models outperform the previous one in all tasks with the exception of the results on the BS-NER dataset by BERT-NEWS-LIT-BG-1 model.

For the task of XPOS tagging we could compare our results with the state-of-the-art performance reported in (Georgiev et al., 2012). There the authors report a method based on Guided Learning with results 95.72 % Accuracy; Guided Learning + Lexicon 97.83 % Accuracy; and Guided Learning + Lexicon + Rules 97.98 %. The results are achieved by training on the constituent version of BulTreeBank, because at that time BulTreebank-UD has not existed yet. Thus, we consider our current models comparable to the state-of-the-art. Since the best results with Guided Learning were achieved by the inclusion of an inflectional lexicon, as a further step we plan to encode this lexicon in the pre-trained models.

With respect to the NER Task, our results are comparable with the results given in (Marinova

Pre-trained Models	Task	Classes	Dataset	Precision	Recall	F1
BERT-WEB-BG	NER	11	BS-NER	0.986718	0.991105	0.988907
BERT-WEB-BG	NER	11	BTB-NER	0.810180	0.813631	0.811902
BERT-WEB-BG	UPOS	16	BTB-UD	0.987725	0.987725	0.987725
BERT-WEB-BG	XPOS	546	BTB-UD	0.943907	0.943907	0.943907
BERT-WEB-BG	XPOS	674	CLaDA-POS	0.948318	0.948318	0.948318
BERT-NEWS-LIT-BG-1	NER	11	BS-NER	0.983014	0.988522	0.985760
BERT-NEWS-LIT-BG-1	NER	11	BTB-NER	0.837433	0.833865	0.835645
BERT-NEWS-LIT-BG-1	UPOS	16	BTB-UD	0.991668	0.991668	0.991668
BERT-NEWS-LIT-BG-1	XPOS	546	BTB-UD	0.953256	0.953256	0.953256
BERT-NEWS-LIT-BG-1	XPOS	674	CLaDA-POS	0.952327	0.952327	0.952327
BERT-NEWS-LIT-BG-2	NER	11	BS-NER	0.993962	0.996836	**0.995397**
BERT-NEWS-LIT-BG-2	NER	11	BTB-NER	0.869374	0.843450	**0.856216**
BERT-NEWS-LIT-BG-2	UPOS	16	BTB-UD	0.992877	0.992877	**0.992877**
BERT-NEWS-LIT-BG-2	XPOS	546	BTB-UD	0.977995	0.977995	**0.977995**
BERT-NEWS-LIT-BG-2	XPOS	674	CLaDA-POS	0.954940	0.954940	**0.954940**

Table 2: Fine-tuning tasks performance for the two pre-trained models. The number of epochs for the NER tasks is 7 and for the POS tasks – 10.

Task	BothTrue	CLTrueBTBFalse	CLFalseBTBTrue	BothFalse	CLASSLA	BERT-NEWS-LIT-BG-1
	#	#	#	#	Accuracy	Accuracy
NER	1079	55	38	60	**92,04 %**	90.66 %
UPOS	1162	21	28	21	96.02 %	**96.59 %**
XPOS	959	8	16	57	92.98 %	**94.51 %**

Table 3: Evaluation over the ParlaMint data of the BERT-NEWS-LIT-BG-1-based models.

et al., 2020) and (Marinova et al., 2023). This similarity is obvious since we also rely on their BERT pre-trained model. The main difference between the two models is that the division of the BS-NER data into training and test subsets is not the same. We were surprised to see that the model performance on the BTB-NER dataset was quite poor. The analysis shows that this is due to selection mainly literature tests for the test set. The category OTHER is problematic, because it practically covers a very diverse set of named entities like names of books, movies, and similar names that sometimes are long phrases or even full sentences. Later on, it was discovered that this type of names were also the largest problem with respect to the NER performance within the Bulgarian part of the ParlaMint corpora.

Evaluation over the ParlaMint data. At the moment we do not have a gold standard dataset for POS tagging and NER over the Bulgarian ParlaMint corpus that is significant in size. Thus, direct measurements of the performance of the train models are not possible. However, in order to perform some initial evaluation, the debates from three days (27/28/29.07.2022) were selected and then annotated automatically with the best one from the above fine-tuned models, based on BERT-WEB-BG and BERT-NEWS-LIT-BG-1 pre-trained models[7]. Then the annotations were manually checked for about 1000 occurrences of NEs and POS tags. More precisely — 1232 for named entities and for UPOS task, and 1040 for the XPOS task. Since the ParlaMint corpora of South Slavic languages were already annotated by Nikola Ljubešić with the CLASSLA models, we were able to compare the results from the models.

The evaluation was performed by our best annotator and the process was executed by the usage of the following categories:

- *BothTrue*: this label means that both CLASSLA and our model took the same decision.
- *CLTrueBTBFalse*: this label means that the decision of CLASSLA was correct and the decision of our model was wrong.
- *CLFalseBTBTrue*: this label means that the decision of CLASSLA was wrong and the decision of our model was correct.
- *BothFalse*: this label means that the decision of both – CLASSLA and our model – was wrong.

The following annotations were considered: UPOS Tagging, XPOS Tagging and NER. For the UPOS Tagging and XPOS Tagging tasks we used the fine-tuned model on BTB-UD dataset. The NER task used the fine-tuned model on BS-NER dataset. The results are given in Table 3 for the models that were fine-tuned on the BERT-NEWS-LIT-BG-1 pre-trained model, and in Table 4 for the models that

[7]We did not have the same evaluation based on the BERT-NEWS-LIT-BG-2 model.

Task	BothTrue	CLTrueBTBFalse	CLFalseBTBTrue	BothFalse	CLASSLA	BERT-WEB-BG
	#	#	#	#	Accuracy	Accuracy
NER	1055	79	33	65	**92,04 %**	89.12 %
UPOS	1168	15	30	19	96.02 %	**97,24 %**
XPOS	959	8	16	57	92.98 %	**93.75** %

Table 4: Evaluation over the ParlaMint data of the BERT-WEB-BG-based models.

were fine-tuned on the BERT-NEWS-LIT-BG-1 pre-trained model.

It can be seen that the biggest drop of the performance is on the NER tasks – with about 8-9 %. The manual check shows that the most problematic cases are the names of documents/regulations/laws that have been discussed during the debates in the Parliament. Here is an example of such a name: "Zakon za ratifitsirane na Memoranduma za razbiratelstvo otnosno podkrepa za proekti na Evropeiskiya sayuz mezhdu pravitelstvoto na Republika Bulgaria i Evropejskata investitsionna banka." (Law on the ratification of the Memorandum of Understanding regarding support for the European Union projects between the Government of the Republic of Bulgaria and the European Investment Bank.) Another type of problematic names are the names of some parties. For example, compare the name "Ima takav narod" (There is Such a People) which constitutes a complete sentence. The same holds for the party "Prodalzhavame promyanata" (We continue with the changes).

These above examples illustrate two issues: i) some names are very long and thus, the usual encoding in a BIO format is not appropriate for them, and ii) there is a big density of novel complex names where the recursive chain is very deep since one name can contain a number of other names. In our view, such newly generated and long names also require new approaches and strategies for the domain adaptation of the existing NER models.

As for the POS annotation, the BERT-NEWS-LIT-BG-1 model performs better on the UPOS tags than on the in-house XPOS ones. However, it must be noted that the UD tags are 16, while the number of XPOS classes are much higher.

To conclude the section, not surprisingly, the out-of-domain data are predominantly sensitive to the named entities and not so much to the POS tags. Despite this it can be seen that POS tagging also drops. This means that there are morphosyntactic specifics in the parliamentary domain that have to be addressed.

5. Conclusions

At this stage of our work no in-domain data was used either in pre-training or fine-tuning phases. The reason for not including data in pre-train stages was that we considered the available one not large enough.

The reason for not including data during the fine-tuning process is the fact that the semi-automatic morphosyntactic disambiguation and the NER checking on the parliamentary sessions has not been finished yet. Only some manual inspection was made on POS tags and NER labels from our fine-tuned BERT models and the CLASSLA Stanza-based model. However, these cross-checks were sufficient to give some main orientation to us about the sources of the drops in the respective results.

For future work we plan to specialize the NER labels towards the parliamentary data. Our idea is to explore the following places for getting domain information on named entities: i) within the specific structure of the sessions such as the interaction formulas; ii) through the referring to various legislative acts and iii) through the discussion over various topics where topic modeling experiments might be applied in advance.

6. Acknowledgements

The reported work has been partially supported by CLaDA-BG, *the Bulgarian National Interdisciplinary Research e-Infrastructure for Resources and Technologies in favor of the Bulgarian Language and Cultural Heritage, part of the EU infrastructures CLARIN and DARIAH*, funded by the Ministry of Education and Science of Bulgaria (support for the Bulgarian National Roadmap for Research Infrastructure). We would like to thank Nikolay Paev, Veneta Kireva and Silvia Petrova for their help with training and testing the models. We would like to thank also the reviewers for their valuable and insightful comments.

7. Bibliographical References

Tomaž Erjavec, Maciej Ogrodniczuk, Petya Osenova, Nikola Ljubešic, Kiril Simov, Andrej Pancur, Michał Rudolf, Matyáš Kopp, Starkaður Barkarson, Steinþór Steingrímsson, Çagrı Çöltekin, Jesse de Does, Katrien Depuydt, Tommaso Agnoloni, Giulia Venturi, María Calzada Pérez, Luciana D. de Macedo, Costanza Navarretta, Giancarlo Luxardo, Matthew Coole, Paul

Rayson, Vaidas Morkevicius, Tomas Krilavicius, Roberts Dargis, Orsolya Ring, Ruben van Heusden, Maarten Marx, and Darja Fiser. 2023. The parlamint corpora of parliamentary proceedings. *Lang Resources and Evaluation*, 57:415–448.

Georgi Georgiev, Valentin Zhikov, Kiril Simov, Petya Osenova, and Preslav Nakov. 2012. Feature-rich part-of-speech tagging for morphologically complex languages: Application to Bulgarian. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 492–502, Avignon, France. Association for Computational Linguistics.

Kia Kirstein Hansen and Rob van der Goot. 2023. Cross-domain evaluation of pos taggers: From wall street journal to fandom wiki.

Momchil Hardalov, Pepa Atanasova, Todor Mihaylov, Galia Angelova, Kiril Simov, Petya Osenova, Veselin Stoyanov, Ivan Koychev, Preslav Nakov, and Dragomir Radev. 2023. bgGLUE: A Bulgarian general language understanding evaluation benchmark. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8733–8759, Toronto, Canada. Association for Computational Linguistics.

Sandra Kübler and Eric Baucom. 2011. Fast domain adaptation for part of speech tagging for dialogues. In *Proceedings of the International Conference Recent Advances in Natural Language 2011*, pages 41–48, Hissar, Bulgaria.

Zihan Liu, Yan Xu, Tiezheng Yu, Wenliang Dai, Ziwei Ji, Samuel Cahyawijaya, Andrea Madotto, and Pascale Fung. 2020. Crossner: Evaluating cross-domain named entity recognition.

Nikola Ljubešić and Kaja Dobrovoljc. 2019. What does neural bring? analysing improvements in morphosyntactic annotation and lemmatisation of Slovenian, Croatian and Serbian. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 29–34, Florence, Italy. Association for Computational Linguistics.

Iva Marinova, Laska Laskova, Petya Osenova, Kiril Simov, and Alexander Popov. 2020. Reconstructing NER corpora: a case study on Bulgarian. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4647–4652. European Language Resources Association.

Iva Marinova, Kiril Simov, and Petya Osenova. 2023. Transformer-based language models for Bulgarian. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 712–720, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Petya Osenova, Kiril Simov, Iva Marinova, and Melania Berbatova. 2022. The Bulgarian event corpus: Overview and initial NER experiments. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3491–3499, Marseille, France. European Language Resources Association.

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, Kiyv, Ukraine. Association for Computational Linguistics.

Jakub Piskorski, Laska Laskova, Michał Marcińczuk, Lidia Pivovarova, Pavel Přibáň, Josef Steinberger, and Roman Yangarber. 2019. The second cross-lingual challenge on recognition, normalization, classification, and linking of named entities across Slavic languages. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 63–74, Florence, Italy. Association for Computational Linguistics.

Marko Prelevikj and Slavko Zitnik. 2021. Multilingual named entity recognition and matching using BERT and dedupe for Slavic languages. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 80–85, Kiyv, Ukraine. Association for Computational Linguistics.

Tobias Schnabel and Hinrich Schütze. 2013. Towards robust cross-domain domain adaptation for part-of-speech tagging. In *International Joint Conference on Natural Language Processing*.

Kiril Simov and Petya Osenova. 2023. Recent developments in BTB-WordNet. In *Proceedings of the 12th Global Wordnet Conference*, pages 220–227, University of the Basque Country, Donostia - San Sebastian, Basque Country. Global Wordnet Association.

Junhao Zheng, Haibin Chen, and Qianli Ma. 2022. Cross-domain named entity recognition via graph matching. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2670–2680, Dublin, Ireland. Association for Computational Linguistics.