BBPOS: BERT-based Part-of-Speech Tagging for Uzbek

Latofat Bobojonova[1] **Arofat Akhundjanova**[2] **Phil Ostheimer**[1] **Sophie Fellenz**[1]
[1]RPTU Kaiserslautern-Landau [2]Saarland University
bobojono@rptu.de arak00001@stud.uni-saarland.de
{ostheimer, fellenz}@cs.uni-kl.de

Abstract

This paper advances NLP research for the low-resource Uzbek language by evaluating two previously untested monolingual Uzbek BERT models on the part-of-speech (POS) tagging task and introducing the first publicly available UPOS-tagged benchmark dataset for Uzbek. Our fine-tuned models achieve 91% average accuracy, outperforming the baseline multilingual BERT as well as the rule-based tagger. Notably, these models capture intermediate POS changes through affixes and demonstrate context sensitivity, unlike existing rule-based taggers.

1 Introduction

Uzbek (a.k.a Northern Uzbek) is the second most-spoken language among all Turkic languages after Turkish (Johanson and Csató, 2015). It has approximately 40 million native speakers and is the official language of the Republic of Uzbekistan. Although the official script for Uzbek is Latin, for historical reasons, it still heavily relies on Cyrillic script, both unofficially and officially. Uzbek is a morphologically rich language (MLR) and ranks as one of the most agglutinative languages in the world.

Although Uzbek is a low-resource language, several language models, particularly BERT-based models, have been pre-trained for Uzbek in recent years (e.g. Mansurov and Mansurov, 2021; Mamasaidov and Shopulatov, 2023; Davronov and Adilova, 2024; Kuriyozov et al., 2024). These models vary in size, quality, and the script of the data on which they have been pre-trained. While some are community projects rather than formal academic publications and lack comprehensive evaluation, others have been assessed only in terms of Masked Language Modeling (MLM) accuracy, with comparisons to multilingual mBERT (Devlin et al., 2019). This limitation stems from the lack of publicly available benchmark datasets for Uzbek (Mansurov and Mansurov, 2021). The main goal of this paper is to fill this gap by creating a new dataset for a downstream task and evaluating models based on this benchmark.

One such downstream task is POS tagging, which lacks publicly available annotated datasets or pre-trained models for Uzbek. POS tagging, specifically with neural models, has the potential to impact linguistic analysis, corpus linguistics, and computational efficiency (Allaberganova and Kuriyozov, 2023). Existing rule-based solutions lack context sensitivity, a limitation that a BERT model can address effectively through its attention mechanism (Murat and Ali, 2024). Finally, the fine-tuning approach using pre-trained language models may be the most effective solution for low-resource languages, helping to bridge both the resource and accuracy gap.

In this work, we introduce the first BERT-based POS tagging models (BBPOS) for Uzbek, available for two actively used scripts, Latin and Cyrillic, together with a newly POS-tagged dataset of 500 sentences. Our models show an average accuracy of 91% based on 5-fold cross-validation.

2 Related Work

Rule-Based POS Taggers: Sharipov et al. (2023) present UzbekTagger — a rule-based POS tagger tool that tags a word by looking up its root form from the dictionary. When it fails to find it, the tagger refers to the neighbouring words to make a decision using six custom grammatical rules. However, the tool only considers the immediate context, making it inferior to neural models (see Section 4).

Statistical POS Taggers: Elov et al. (2023) demonstrate the application of Hidden Markov Models (HMMs) on Uzbek by manually tagging a small set of sentences, without developing a full model or dataset.

Proceedings of the First Workshop on Language Models for Low-Resource Languages (LoResLM 2025), pages 313–319
January 20, 2025.

Neural POS Taggers: Murat and Ali (2024) present the only work on neural POS tagging in Uzbek, alongside two other MRLs: Uyghur and Kyrgyz. The authors propose a new POS tagging method for MRLs using a deeper representation through affix embeddings. They also employ a multi-head attention mechanism to the baseline models and capture dependencies between words regardless of their distance, thereby addressing POS tag ambiguity. This approach achieved an overall accuracy of 79.74% for Uzbek, representing an increase of up to 4.13% over other models that utilize only BiLSTMs, CNNs, and CRFs. Unfortunately, their trained models are not publicly available.

Dataset & Tagset: Initial work on the Uzbek morphological tagset identified 12 POS tags that correspond to word classes in traditional Uzbek grammar (Abjalova and Iskandarov, 2021). Sharipov et al. (2023) applied this tagset, though their annotated dataset has not been made publicly available. Murat and Ali (2024) used a distinct set of 12 POS labels in their dataset designed to be suitable for Uzbek, Uyghur and Kyrgyz. Although the dataset is relatively large, with 20k sentences in the training set and over 23k distinct stems in the Uzbek corpus, it is not publicly available. More morphologically comprehensive tagsets with over 100 tags were also proposed by Sharipov et al. (2022) and Abdullayeva et al. (2022), but no tagged datasets based on these frameworks currently exist.

3 Experiments

3.1 Methods

Due to the lack of a public dataset for POS tagging, we created our own dataset[1] (see Section 3.2). We chose one pre-trained model for each script (see Section 3.3) and fine-tuned them[2] with our dataset for the POS tagging task. As a baseline, we fine-tuned a multi-lingual mBERT model (Devlin et al., 2019). Each type of model was individually evaluated using a 5-fold cross-validation with a 80% - 20% train-test split. All BERT models were fine-tuned with the same hyperparameters (see Appendix A).

[1]The dataset is publicly available at https://huggingface.co/datasets/latofat/uzbekpos

[2]One fine-tuned model per script is available at: https://huggingface.co/latofat

Index	POS tag	# of words	# of unique words
0	ADJ	454	356
1	ADP	189	48
2	ADV	152	102
3	AUX	96	27
4	CCONJ	85	7
5	DET	**16**	**14**
6	INTJ	**11**	**6**
7	NOUN	2141	1751
8	NUM	217	94
9	PART	67	14
10	PRON	273	112
11	PROPN	300	261
12	PUNCT	810	19
13	SCONJ	**9**	**3**
14	SYM	**1**	**1**
15	VERB	1001	721
16	X	**9**	**3**
	Total	5831	3488

Table 1: Overview of the distribution of tags in the dataset. Bold numbers highlight relatively underrepresented tags.

3.2 Data

Tagset Selection: We used the Universal Part-of-Speech (UPOS) (Nivre et al., 2016), as it is a multilingual tagset that aims to cover similar linguistic features consistently across languages. Currently, it has been the foundation for 283 treebanks in 116 languages[3] and our dataset is the first work to employ UPOS for Uzbek. There are 17 tags in the UPOS as shown in Table 1, and Uzbek can use all of them. Furthermore, it is easy to map UPOS to 12 word classes identified in traditional Uzbek grammar (see Appendix B).

Dataset Development: We collected 500 sentences (5,831 words), 250 sourced from news articles and 250 from fictional books. We manually annotated the data written in Latin script with UPOS tags. Then it was transliterated into a Cyrillic script to fine-tune the Cyrillic model. Table 1 shows the distribution of tags in the dataset and the number of unique words per POS (more details in Appendix C). As the sentences are ordered according to their genre, i.e., fiction and news, the datasets for each script were shuffled with the same seed before a 5-fold split for training and testing.

[3]https://universaldependencies.org/

	UzbekTagger	mBERT		TahrirchiBERT	UzBERT
	rule-based	latin	cyrillic	(latin)	(cyrillic)
Accuracy	75.6 ± 1.6	$\mathbf{86.0 \pm 1.0}$	80.2 ± 1.0	90.9 ± 0.9	$\mathbf{91.6 \pm 0.4}$
F1	57.4 ± 2.3	$\mathbf{77.5 \pm 0.9}$	68.5 ± 1.9	85.2 ± 1.3	$\mathbf{86.4 \pm 0.6}$

Table 2: Accuracy and F1-score for different POS taggers measured in Mean $\pm$ Standard Deviation (%).

3.3 Models

Latin BERT: We chose the open source TahrirchiBERT (Mamasaidov and Shopulatov, 2023), a monolingual RoBERTa (Liu et al., 2019) model pre-trained on Uzbek Latin script. It is trained on large text data extracted from online blogs and scanned books (equivalent to 5B tokens $\approx$ 18.5GB). The dataset is fairly noisy due to the errors introduced by poor OCR applied to the books. Additionally, TahrirchiBERT does not handle the required pretokenization rules for the Latin script of Uzbek. Specifically, the modifier letters[4] used in oʻ and gʻ letters and the glottal stop sign ʼ are treated as delimiter signs that cause incorrect word splits. The authors introduced a normalization specific to Uzbek Latin script, preventing some common spelling errors.

Cyrillic BERT: We fine-tuned UzBERT (Mansurov and Mansurov, 2021), a monolingual BERT model (Devlin et al., 2019) pre-trained solely on Cyrillic scripted Uzbek text. According to the authors, the model is trained on high-quality Cyrillic text data with 142M words ($\approx$ 1.9GB) and has not been evaluated on any downstream tasks due to the lack of public datasets. There are no Uzbek Cyrillic script-specific rules to be applied during the normalization and pretokenization stages, as each letter in the Uzbek Cyrillic alphabet is represented by a single alphabetic character.

4 Results

Table 2 shows accuracy and F1-score for all trained models together with the results obtained from the rule-based UzbekTagger on the POS-converted dataset (see Appendix D). It presents the mean and standard deviation for accuracy and F1-score of 5-fold cross-validation. The rule-based POS tagger with an average accuracy of 75% falls behind all BERT models. Both monolingual models outperform mBERT by a good margin overall. Table 2, column three shows Latin Uzbek is better represented in mBERT than Cyrillic Uzbek.

[4]A modifier letter functions like diacritics, changing the sound-values of the letter it proceeds. Unlike diacritics, they do not combine with the letter.

POS	tah-rir-chi (lat)	uz-bert (cyr)	m-bert (lat)	m-bert (cyr)	rel. freq.
ADJ	77.0	79.3	54.7	23.3	8.9
ADP	92.8	88.6	88.5	63.0	3.6
ADV	64.3	75.4	11.1	5.7	3.6
AUX	88.2	83.3	90.9	55.2	1.9
CCONJ	84.8	94.1	90.9	90.9	1.9
DET	**0.0**	**0.0**	**0.0**	**0.0**	**0.3**
INTJ	**0.0**	**0.0**	**0.0**	**0.0**	**0.4**
NOUN	86.1	81.6	72.2	50.9	28.2
NUM	88.2	93.0	80.0	83.1	3.8
PART	92.3	85.7	92.3	64.5	1.5
PRON	76.8	84.7	77.2	70.6	5.8
PROPN	90.7	87.1	77.6	53.5	4.4
PUNCT	98.9	100.0	98.3	99.4	18.6
SCONJ	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
SYM	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
VERB	89.8	90.7	84.3	76.4	17.2
X	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**

Table 3: F1-scores of BBPOS models for each POS tag, including the tags' relative frequency in the evaluation set. Bold entries indicate tags that models failed to learn.

UzBERT vs TahrirchiBERT: Monolingual BERT models, regardless of script, show similarly high accuracies of at least 90% and F1-scores of at least 84%. Having been trained on ten times less data, UzBERT has outperformed TahrirchiBERT by a slight margin in both metrics. We hypothesize that this might be due to the data quality used for pre-training and incorrect pretokenization used for Latin scripted text. Especially during inference, when a sentence has to be pretokenized, TahrirchiBERT fails in successfully tagging words written with one of the modifier letters.

Learning per Tag: We randomly chose an evaluation fold to evaluate which tags are learned well by the BERT models. In Table 3, we present the

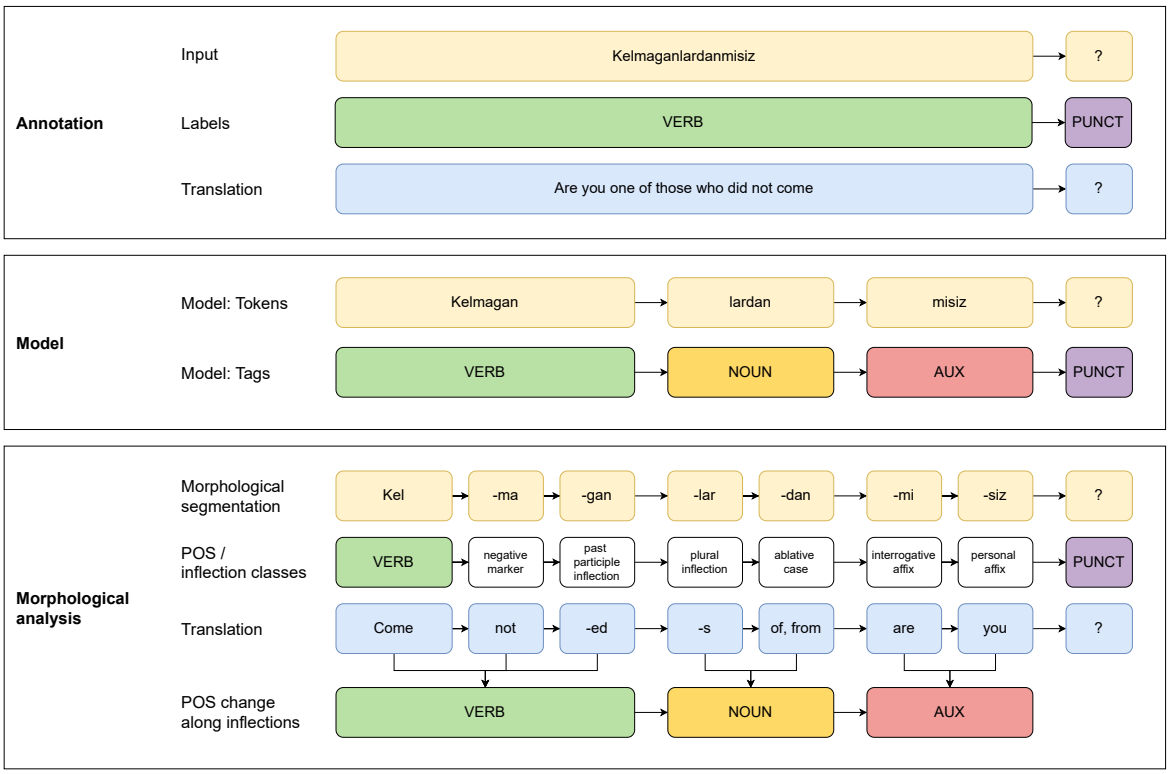

Figure 1: Analysis of one sentence-word in Uzbek: manual annotation according to UPOS guidelines (top); how BBPOS tags it (middle); comprehensive morphological analysis of the word (bottom).

relative frequency of POS tags in the chosen evaluation fold, together with the F1-scores obtained by the corresponding BERT models that were not trained on it. All models could not learn the same five tags, most likely due to the low representation in the overall dataset (see Table 1).

Context-sensitivity: We assess the rule-based and neural models for context sensitivity, running a couple of sentences containing homonyms. The sentence *Tortmani tortma* 'Don't pull the drawer' should be tagged as `[NOUN, VERB]`. The rule-based UzbekTagger will naturally tag it as `[NOUN, NOUN]` (treating it as 'The drawer drawer'). Similarly, mBERT fails at tagging this same sentence in both Latin and Cyrillic. However, TahrirchiBERT and UzBERT tag it correctly as `[NOUN, VERB]`.

5 Discussion

An interesting aspect of our experiments was how our models handled highly inflected words. They learned morphological features by detecting intermediate POS changes through affixes. For instance, in Figure 1 you can see how the word *Kelmaganlardanmisiz?* which corresponds to a whole sentence in English ('Are you one of those who did not come?') is tagged by our models. It also shows manual POS and morphological annotation for it. As you can see, our model's result resembles the morphological analysis rather than the simple POS labeling with which it was trained. In fact, according to Universal Dependencies (UD) guidelines, the word's POS relies solely on its lemma's POS.

Our work on POS tagging has the potential for extension to data generation in morphological analysis, specifically in morpheme classification. However, this requires BERT models to be pre-trained using morphological or morphologically informed tokenizers rather than relying on subword tokenization methods like BPE and WordPiece which are statistical algorithms. Additionally, the success of neural models in learning aspects of Uzbek morphology could inspire the linguistic community to develop a unified and comprehensive POS tagset for Uzbek, one that considers how morphemes influence word-level POS shifts. Previous work on Turkish (Çöltekin, 2016) also discusses the guidelines for this.

The inconsistent representation of the letters `o‘`, `g‘` and `’` in texts, caused by the use of varying forms of apostrophes, poses a significant challenge for Latin Uzbek. This issue, as evidenced by the

pre-tokenization problem detected in TahrirchiBERT, underscores the importance of pre-training language models for Latin-scripted Uzbek on data that adheres to consistent alphabet standards. Alternatively, we can focus on pre-training monolingual Uzbek models that apply normalization rules to standardize the singular form of the above letters across diverse Uzbek text data.

6 Conclusion

In this work, we introduced a new dataset for the low-resource Uzbek language tagged with the UPOS tagset and trained the first BERT-based POS taggers on it. We evaluated two monolingual Uzbek BERT models on the POS tagging downstream task, identifying potential improvements to pre-train Uzbek language models in the future. Our BBPOS models reached an average accuracy of 91% on 5-fold cross-validation, outperforming the baseline mBERT and the existing rule-based solution by far both in accuracy and F1-score. They show context sensitivity in handling ambiguous sentences with homonyms. They learned parts of speech for POS changing morphemes, generating enriched annotations with more linguistic information.

Limitations

We acknowledge the following limitations of the fine-tuned models:

- Even though our fine-tuned models performed better than the rule-based tagger on the evaluation sets, we acknowledge that our models fail to tag overly inflected words as single tokens due to the subword tokenization used in them. The models can be used for synthetic data generation although with heavy human supervision to ensure quality and accuracy.

- Additionally, due to the poor pretokenization of TahrirchiBERT, the Latin models fail at words containing the letters `o‘`, `g‘`, `’`, as they incorrectly split them into words treating the modifier letters as delimiters. This error is not evident during the validation and training stages of the token classification task as it is during inference.

We also acknowledge the following limitations of our benchmark dataset:

- Our models failed to learn five out of seventeen POS tags due to the small representation in the initial dataset. Our benchmark needs to be enriched on those POS tags.

- While not of major importance, our dataset is relatively small. The dataset is insufficient for training POS tagging models from scratch, such as HMM, CRF, RNN, or LSTM. While we trained an HMM model, its poor performance, achieving an accuracy of (40.7 ± 1) and an F1-score of (8.9 ± 1.8), proved it to be an inadequate baseline and therefore it is not included in the results.

References

Oqila Abdullayeva, Mokhiyakhon Uzokova, and Botir Elov. 2022. O‘zbek tilida POS tegging masalasi: muammo va takliflar. *Uzbekistan: Language and Culture*, 2(5):51–68.

Manzura Abjalova and Otabek Iskandarov. 2021. Methods of tagging part of speech of Uzbek language. In *2021 6th International Conference on Computer Science and Engineering (UBMK)*, pages 82–85. ISSN: 2521-1641.

Guliston Allaberganova and Elmurod Kuriyozov. 2023. Neural network-based models for Uzbek part-of-speech tagging, analysis and usage. In *RESPUBLIKA ILMIY-TEXNIK ANJUMANINING MA’RUZALAR TO’PLAMI*.

Rifkat Davronov and Fatima Adilova. 2024. UzRoberta: A pre-trained language model for Uzbek. *AIP Conference Proceedings*, 3004(1):050001.

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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Botir Elov, Shuhrat Sirojiddinov, Khamroeva Shahlo Mirdjonovna, Eşref Adalı, and Xusainova Zilola Yuldashevna. 2023. POS taging of Uzbek text using hidden markov model. In *2023 8th International Conference on Computer Science and Engineering (UBMK)*, pages 63–68. ISSN: 2521-1641.

Lars Johanson and Éva Ágnes Csató. 2015. *The Turkic Languages*. Routledge.

Elmurod Kuriyozov, David Vilares, and Carlos Gómez-Rodríguez. 2024. BERTbek: A pretrained language model for Uzbek. In *Proceedings of the 3rd Annual Meeting of the Special Interest Group on Underresourced Languages @ LREC-COLING 2024*, pages 33–44, Torino, Italia. ELRA and ICCL.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: a robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Mukhammadsaid Mamasaidov and Abror Shopulatov. 2023. TahrirchiBERT base. Accessed: 2024-06-14.

B. Mansurov and A. Mansurov. 2021. UzBERT: pretraining a BERT model for Uzbek. *CoRR*, abs/2108.09814.

Alim Murat and Samat Ali. 2024. Low-resource POS tagging with deep affix representation and multi-head attention. *IEEE Access*, 12:66495–66504.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan T. McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*. European Language Resources Association (ELRA).

Shavkat Rahmatullayev. 2006. *Hozirgi Adabiy O'zbek Tili [Contemporary Literary Uzbek Language (textbook)]*. Universitet.

Maksud Sharipov, Elmurod Kuriyozov, Ollabergan Yuldashov, and Ogabek Sobirov. 2023. Uzbektagger: The rule-based POS tagger for Uzbek language. *CoRR*, abs/2301.12711.

Maksud Sharipov, Jamolbek Mattiev, Jasur Sobirov, and Rustam Baltayev. 2022. Creating a morphological and syntactic tagged corpus for the Uzbek language. In *Proceedings of the ALTNLP The International Conference and workshop on Agglutanative Language Technologies as a challenge of Natural Language Processing, Virtual Event, Koper, Slovenia, June, 7th and 8th, 2022*, volume 3315 of *CEUR Workshop Proceedings*, pages 93–98. CEUR-WS.org.

Çağrı Çöltekin. 2016. (When) do we need inflectional groups? In *Proceedings of The First International Conference on Turkic Computational Linguistics*.

A Hyperparameter Settings

Table 4 shows hyperparameters and their values used in the fine-tuning of BERT models using `transformers`[5] library.

For all conducted evaluations we used the sequence labeling evaluation metric – `seqeval` – from the `evaluate`[6] package.

learning_rate	2e-5
per_device_train_batch_size	16
per_device_eval_batch_size	16
num_train_epochs	5
weight_decay	0.01

Table 4: Hyperparameters used for fine-tuning the BERT models

B Tagset Conversion

UPOS	Uzbek POS	Comment
NOUN PROPN	NOUN	
PRON DET	PRON	
CCONJ SCONJ	CONJ	
AUX	VERB	
ADJ	ADJ	
ADP	AUX	
INTJ	INTJ	
NUM	NUM	
PART	PART	
ADV	MOD ADV	There are finite modal words
VERB	IMIT VERB	There are finite immitation words
PUNCT SYM X	irrelevant	There is no specific POS tag for these gorup of tokens in the Uzbek grammar

Table 5: UPOS → traditional Uzbek POS

To align UPOS with the traditional Uzbek POS tagset and bridge prior research, we developed a conversion script that maps UPOS tags to Uzbek POS categories. Table 5 shows how individual tags are handled, grouped, or reclassified according to Uzbek linguistic rules (Abjalova and Iskandarov, 2021). While tags like `ADJ`, `INTJ`, `NUM` and `PART` are aligned directly, some are merged into broader word classes (e.g. `PROPN` ∪ `NOUN` = `NOUN`).

The most complex part of this conversion is `ADV` and `VERB` tags. Uzbek grammar splits adverbs into

[5] https://huggingface.co/docs/transformers

[6] https://huggingface.co/docs/evaluate

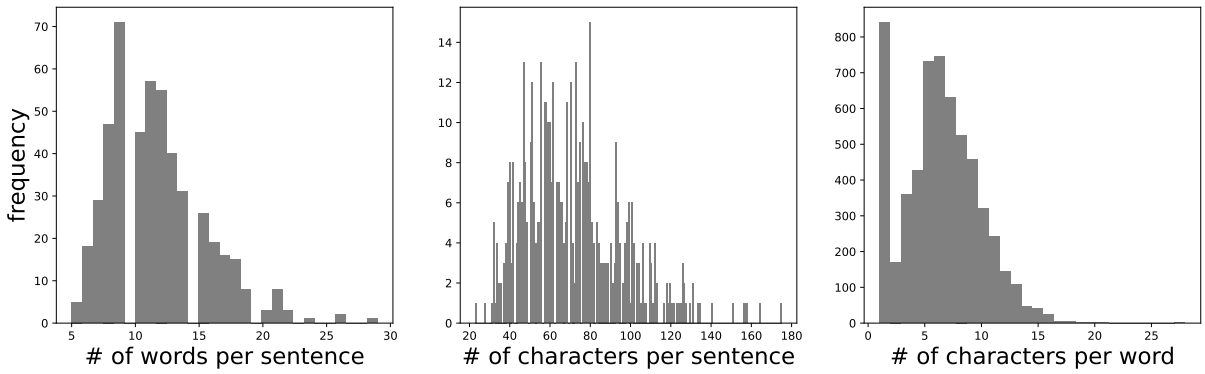

Figure 2: Words per sentence and characters per word/sentence in the dataset.

action-related (ADV) and attitude-related (MOD). Similarly, VERB is split into true verbs (VERB) and imitative words (IMIT). We made this distinction by simple set membership check, as words that belong to MOD and IMIT classes are finite and uninflected. Moreover, the Uzbek grammar does not specify POS tags for punctuation (PUNCT), symbols (SYM), and miscellaneous categories (X), so we excluded them from the mapping.

C Data Statement

We chose news and fiction genres to ensure broad domain coverage while preserving diversity in length, formality, and literary quality. All sentences were handpicked to ensure the quality of the data. News texts of the dataset were collected from the major news sites[7]. They cover various topics and reflect contemporary Uzbek language use. Fiction texts were chosen from the publicly available Uzbek works on the internet, including: "Og'riq Tishlar" and "Dahshat" by Abdulla Qahhor, "Shum Bola" and "Yodgor" by G'afur G'ulom, "Sofiya", "Hazrati Hizr Izidan", "Bibi Salima va Boqiy Darbadar", "Olisdagi Urushning Aks-Sadosi" and "Genetik" by Isajon Sulton, "Buxoro, Buxoro, Buxoro…", "Ozodlik" and "Lobarim Mening…" by Javlon Jovliyev, "Ko'k Tog'", "Insonga Qulluq Qiladurmen", "Fano va Baqo" and "Chodirxayol" by Asqar Muxtor, "Ajinasi Bor Yo'llar" by Anvar Obidjon, "Kecha va Kunduz" and "Qor Qo'ynida Lola" by Cho'lpon.

Figure 2 shows the number of words per sentence and the number of characters per word/sentence. The number of words per sentence ranges from 5 to 29, with an average of 11–12, likely reflecting natural linguistic patterns in Uzbek. This trend is further illustrated by the average number of characters per sentence (72) and per word (6).

Annotation was performed manually by one of the native Uzbek-speaking authors who is MSc in Computational Linguistics with a background in Uzbek linguistics, applying each UPOS tag according to the Universal Dependencies (UD) guidelines[8]. In addition to UD POS tagging guidelines, UD treebanks of other Turkic languages and Uzbek grammar rules (Rahmatullayev, 2006) were also used as a point of reference. Ambiguous cases such as the annotation of multiword expressions (MWEs) in compound verbs were solved through extensive discussions with other linguists and UD experts.

The Latin-scripted dataset was subsequently turned into a morpho-syntactically annotated UD treebank, released as part of UD version 2.15.

The transliteration was performed using an online transliterator tool[9].

D Comparison with UzbekTagger

To compare BBPOS models with the rule-based POS tagger tool, we relabeled our golden dataset with 12 conventional Uzbek POS tagset using the conversion script we developed (see Section B). The token families that are excluded by the logic of UzbekTagger, such as punctuations, symbols and other (i.e. PUNCT, SYM, X) were eliminated from the dataset to the favor of UzbekTagger results. We then ran the untagged 5 evaluation folds, each containing 100 sentences, through UzbekTagger and compared the results against the relabeled golden dataset.

[7] https://kun.uz/ and https://daryo.uz/

[8] https://universaldependencies.org/u/pos/

[9] https://tahrirchi.uz/uz/editor