Article

The Phenomenon of Functional Shift Between Parts of Speech in Uzbek POS Tagging

Xudayberganov Nizomaddin Uktambay ugli*[1]

1. Lecturer at the Department of Computational Linguistics and Digital Technologies, Tashkent State University of Uzbek Language and Literature (TSUULL)
* Correspondence: nizomaddin@navoiy-uni.uz

Abstract: This article analyzes one of the key linguistic problems arising in the process of automatic Part-of-Speech (POS) tagging of Uzbek language units, namely the phenomenon of functional shift between parts of speech. The agglutinative nature of the Uzbek language, the multifunctionality of verb forms, and their strong dependence on context are shown – through scientific examples, tables, and diagrams – to have a direct impact on POS tagging accuracy. The article discusses the linguistic causes of functional shift and explores ways of accounting for this phenomenon in automatic POS tagging models.

1. Introduction

PPOS tagging (Part-of-Speech tagging) is one of the basic tasks of NLP. It is the process of categorizing each word form in a text as a particular grammatical form (e.g. noun, verb, adjective, numeral, adverb, pronoun). For Uzbek, POS tagging is a morphologically difficult problem because of its agglutinative linguistic nature where high affixation and the process of functional shift may occur between the POS.

Functional shift is characterized by the use of the same lexical unit as different parts of speech depending on the context. Since this phenomenon can lead to errors during the POS tagging process, its in-depth analysis from both linguistic and computational perspectives is of significant importance.

2. Methodology

This study applies a combined linguistic–computational methodology to analyze the phenomenon of functional shift and its impact on automatic POS tagging in the Uzbek language. The methodology consists of five main stages: corpus preparation, linguistic annotation, model development, experimental evaluation, and error analysis.

The research follows a comparative experimental design, in which different POS tagging approaches (rule-based, statistical, and neural) are tested and compared with respect to their ability to correctly identify functionally shifted forms. The study integrates principles of corpus linguistics and Natural Language Processing, drawing conceptually on the frameworks proposed in established NLP research such as Abney's statistical

approaches to linguistic knowledge and the computational methodologies described in Bird et al. and Jurafsky and Martin.

The empirical material consists of annotated Uzbek corpus texts selected from:

Educational and journalistic texts

Literary prose

Academic publications

Where available, the annotation scheme was aligned with the Universal Dependencies (UD) framework as described in Nivre et al. and Zeman et al.

The corpus preparation process included:

Tokenization – segmentation of text into word forms.

Morphological segmentation – identification of stems and affixes.

Manual annotation – assignment of:

Universal POS (UPOS) tags

Morphological features (FEATS)

Dependency relations (where necessary)

Special attention was given to lexical units demonstrating functional shift (e.g., participles, verbal nouns, adverbialized adjectives).

3. Identification of Functional Shift Instances

Functional shift cases were extracted using a two-stage filtering procedure:

Stage 1: Morphological ambiguity detection

Word forms capable of belonging to more than one grammatical category (e.g., participles, infinitives, adjectives used adverbially) were automatically flagged.

Stage 2: Contextual verification

Linguistic analysis determined the syntactic and semantic function of each flagged unit within the sentence.

Annotation decisions were guided by formal syntactic criteria as described in Sag et al. and by morphological descriptions of Uzbek.

By combining structural linguistic analysis with modern NLP techniques, the study systematically evaluates how functional shift affects POS tagging accuracy in Uzbek and proposes technically grounded solutions for improving automatic tagging systems..

3. Results and Discussion

Functional shift (conversion) refers to a word belonging to different parts of speech based on its syntactic and semantic function in a sentence, without changing its external morphological form. In the Uzbek language, this phenomenon is particularly prevalent within the verb category (Table 1).

Table 1. Examples of functional shift in the Uzbek language

№	Word Form	Usage in Context	Assigned Part of Speech
1.	yaxshi	Yaxshi natija. (Good result)	Adjective
2.	yaxshi	U yaxshi ishladi. *(He/She worked well)*	Adverb
3.	oʻqish	Oʻqish muhim. *(Studying/Education is important)*	Noun
4.	oʻqish	U oʻqish bilan band. *(He/She is busy with studying)*	Verbal Noun (Infinitve)
5.	kelgan	Kelgan mehmonlar. *(The guests who arrived)*	Adjective (Participle)
6.	kelgan	U kelgan. *(He/She has come)*	Verb

As seen from the table, in the process of POS tagging, the contextual function of a word takes precedence over its formal morphological structure.

Linguistic Causes of Functional Shift

The functional shift phenomenon in Uzbek is based by the following linguistic reasons:

1. Polyfunctionality of Verbs: The relation between verb stems and corresponding derived forms across multiple grammatical categories.

2. The scope of meaning: The range of meanings and functions which can be expressed by a given morpheme in different contexts.

3. Smothered Under Position : It is the position in a sentence that determines the class to which it belongs.

4. Active Use of Analytical Forms: The prevalence of compound structures where auxiliary components influence the categorical affiliation of the lexical unit.
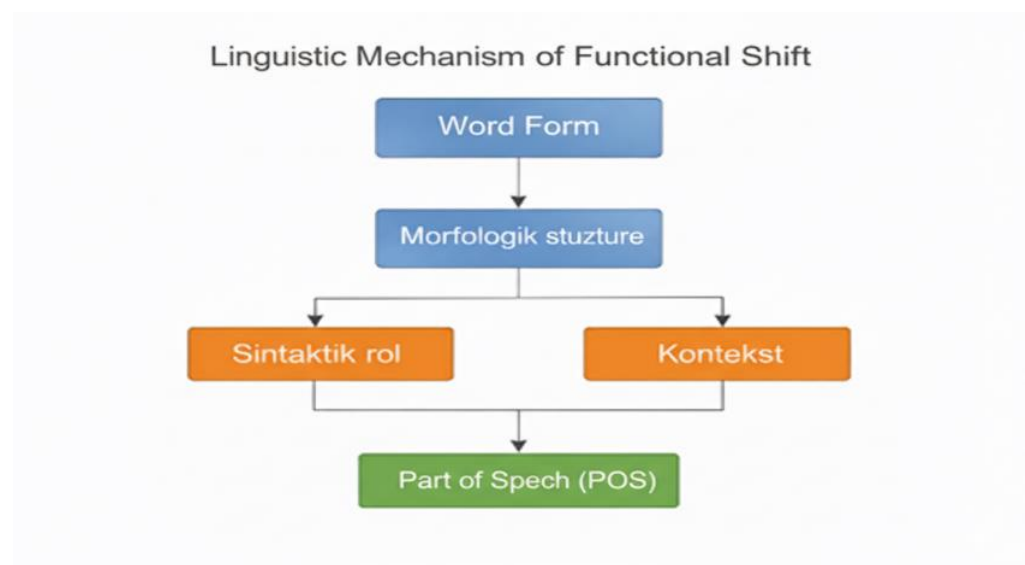

Figure 1. Linguistic mechanism of functional shift

This diagram illustrates that POS tagging is based not only on morphological form but also on the interaction of syntactic and semantic factors (Figure 1).

The Impact of Functional Shift on the POS Tagging Process

- **Rule-based POS tagging:** Rule-based POS tagging primarily relies on affixes and rigid grammatical rules. However, this approach is insufficient in cases of functional shift. For example:

  Kelgan talabalar zalga kirdi. (The students who arrived entered the hall.)

  In this sentence, although the unit *kelgan* functions as an adjective (participle), rule-based systems may mistakenly tag it as a verb.

- **Statistical and Neural Approaches:** While statistical and neural models take context into account, the limited size of annotated corpora for the Uzbek language hinders the comprehensive study and accurate processing of functional shifts (Table 2).

Table 2. The impact of functional shift on POS tagging models

№	Approach Type	Source of the Problem	Result
1.	Rule-based	Limited context	Incorrect POS
2.	Statistical	Small corpus size	Low accuracy

3.	Neural	Irregular annotation	Difficult to generalize

Approaches to Accounting for Functional Shift

To increase the accuracy of **POS tagging**, it is advisable to model functional shift through the following approaches:

1. Integrating Morphological and Syntactic Analysis: Combining word-form structure with sentence-level dependencies to resolve ambiguities.

2. Multi-stage POS Tagging Model: Implementing a layered processing system where initial tags are refined through subsequent contextual filters.

3. Separate Labeling of Universal POS (UPOS) Tags and Morphological Features (FEATS): Utilizing the Universal Dependencies framework to distinguish the core part of speech from its specific grammatical properties Figure 2.
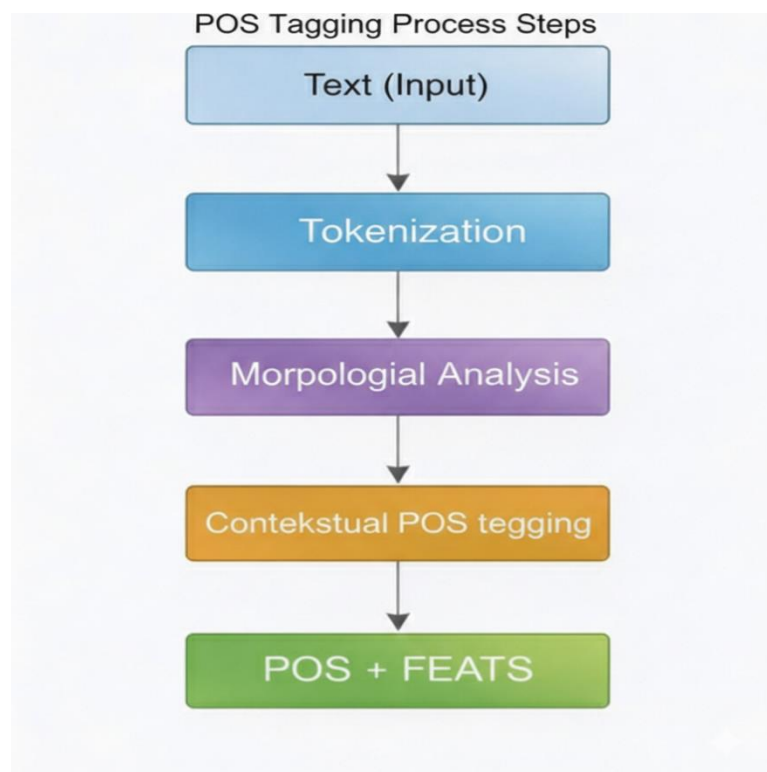

Figure 2. Multi-stage POS tagging model

4. Conclusion

The functional shift between POS of words is a widespread linguistic problem in POS tagging for Uzbek. This effect depends on the intrinsic grammatical properties of the language itself; its disregard leads to a drastic decrease in the performance of automatic tagging mechanisms. Hence, in constructing POS Tagging systems for the Uzbek language, context-sensitive linguistically motivated methods must be employed.

REFERENCES

[1] S. Abney, "Statistical methods and linguistic knowledge in corpus linguistics," Computational Linguistics, vol. 22, no. 1, pp. 1–15, 1996.

[2] B. Elov and N. Xudayberganov, "Methods of POS tagging for Uzbek language corpus texts," Computer Linguistics: Problems, Solutions, Prospects, vol. 1, no. 1, 2024.

[3] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. Sebastopol, CA, USA: O'Reilly Media, 2009.

[4] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed., draft version. Stanford, CA, USA: Stanford University, 2023.

[5] J. Nivre, M.-C. de Marneffe, F. Ginter et al., "Universal Dependencies v1: A multilingual treebank collection," in Proc. LREC 2016, 2016, pp. 1659–1666.

[6] I. A. Sag, T. Wasow, and E. M. Bender, Syntactic Theory: A Formal Introduction. Stanford, CA, USA: CSLI Publications, 2003.

[7] S. Usmonov, "Morphological features of the Uzbek language and issues of automatic analysis," Issues of Uzbek Linguistics, no. 4, pp. 45–58, 2020.

[8] D. Zeman, J. Nivre, M. Abrams et al., Universal Dependencies 2.13. LINDAT/CLARIAH-CZ, 2023.

[9] H. N. Qahramonovich, "Worthy descendants of the Samarkandians: enlighteners jadids," World Bulletin of Social Sciences, vol. 13, pp. 37–40, 2022.

[10] R. T. Usanov and N. K. Khakkulov, "Iz istorii izucheniya problemy 'Nenasilie'," Ekonomika i Sotsium, no. 6-2 (121), pp. 1406–1413, 2024.

[11] T. Shaymardonov, "Siddiqiy-Ajziy hayoti va ijodining o'rganilishi," Uzbekistan: Yazyk i Kultura, vol. 1, no. 1, pp. 27–39, 2022.

[12] T. A. Shaymardonov, "Sayidahmadxo'ja Siddiqiy-Ajziyning qofiya qo'llash mahorati," Yosh Olimlar Axborotnomasi – Vestnik Molodykh Uchenykh, vol. 1, no. 1, pp. 147–152, 2024.

[13] T. A. Shaymardonov, "The role of Siddiqi-Ajzi in Uzbek jadid literature," American Journal of Social and Humanitarian Research, vol. 1, no. 4, pp. 91–100, 2020.

[14] S. M. Ruzimurodov, "Ethnomadanian and ethnoesthetic aspects of the formation of a healthy lifestyle in society," The American Journal of Social Science and Education Innovations, vol. 3, no. 5, pp. 188–194, 2021.

[15] A. A. Azizkulov and S. M. Ruzimurodov, "Gazali on the significance of intellect in human activity," Paradigmata Poznani, no. 2, pp. 43–46, 2014.