

Problems of Automatic Recognition of Verb Analytical Forms in the Uzbek Language

Nurova Firuza Kamolovna,

Senior teacher of Bukhara State Pedagogical Institute, Uzbekistan

nuruvofiruza@buxdpi.uz

Abstract. This article addresses theoretical and practical challenges in the automatic recognition of analytical verb forms in the Uzbek language. It analyzes criteria for distinguishing the grammatical function of auxiliary verbs, as well as disambiguating them from phraseological units and compound verbs. Special focus is given to the contextual differentiation of “converb + verb” constructions. Issues of segmentation and disambiguation essential for morphological analyzers are examined. Based on A. Hojiyev’s theory, the study aims to provide a scientific and practical foundation for machine translation and natural language processing systems.

Keywords: language processing, morphological analyzer, analytical forms, machine translation systems, semantic differentiation, homonymy in verb combinations, machine translation, morphological analysis and automatic syntactic parsing, grammatical and semantic interpretation, semantic modification, automatic recognition, auxiliary verb.

Introduction

With the development of the natural language processing (NLP) direction in modern linguistic technologies, the creation of morphological analyzers for the Uzbek language is becoming increasingly relevant. In particular, the widely used **analytical verb forms** in the Uzbek language and their automatic recognition are one of the complex problems that currently need to be solved.

The auxiliary verbs existing in our language are considered **analytical forms** according to their grammatical function, structure and semantic load. This situation has been studied in depth in the research of Uzbek linguists, and it is this approach that is the impetus for the perfect modeling of Uzbek morphoanalyzers.

According to A. Hojiyev, in the Uzbek language there are verbs that give an **additional semantic tone to the action**, which, when combined with the main verb, form a very important syntactic unit in the language - the analytical verb form. He writes as follows: “These verbs, when combined with an independent verb, serve to give various additional meanings to the meaning of this verb” [A. Hojiyev, 1966, p. 3]. For example:

- *yozi**b** yurmoq* (to be writing) - the duration of the action
- *yozi**b** bermoq* (to write for smb) - the performance of the action for another
- *yozi**b** bo‘lmoq* (have write) - the complete performance of the action
- *yozi**b** ko‘rmoq* (to try to write) - the performance of the action with trial and knowledge

These are independent grammatical units, and although they are formally divided into two verbs, they form a single whole semantically. The main task of the morphological analyzer is to analyze these semantic single units as a single word. Analytical forms of the verb can form homonymy in our speech with phraseological units, adverb + verb constructions, and compound verbs

with verb + verb structure. In their research, scholars have cited a number of data on the homonymy of the auxiliary verb phrase with the adverb + verb construction and the verb + verb compound. [9]

The Main Part

In the Uzbek language, within the analytical forms of the verb, the constructional compounds “ravishdosh + verb” occur with the participation of an auxiliary verb. Such units sometimes cause cases of homonymy due to their external formal similarity with compound verbs. The compound “o‘qib brâdi” in the sentence is one of such types.

U bu kitobni bir kechada o‘qib tashladi (He read this book in one night).

In this context, the conjunction "*o‘qib tashladi*" ("read") indicates **the completeness, completion, and intensity of the action**. That is, the process of reading the book was carried out smoothly, completely, and quickly.

U xatni o‘qib tashladi. (He read the letter)

In this case, the meaning of "*o‘qib tashladi*" ("read") can have two different interpretations:

- **Interpretation 1:** as above – the letter was read, that is, the act of reading was completed.
- **Interpretation 2 (semantic expansion):** after reading the letter, he *tashlab yubordi* (**threw it away**) (representing a real action: physically throwing the letter away).

Therefore, this construction has a homonymous appearance, and both components — “read” and “threw” — have retained their independent lexical meaning, which indicates that they are not compound verbs, but word combinations.

Syntactic differential signs

• If the component "*o‘qib tashladi*" ("threw it away") is independent of the context and can be applied to another subject (for example: read [the letter] threw [into the trash can]), then it is evaluated as **two separate predicative acts**.

• On the contrary, if "*o‘qib tashladi*" ("threw it away") performs only semantic modification (intensifies intensity), it is considered as **an auxiliary verb** and forms a whole verb complex.[8]

Thus, the combination "*o‘qib tashladi*" ("**threw it away**") has two different grammatical and semantic interpretations, depending on its contextual use:

- **As an analytical verb form** – it means a single action in semantic terms (to finish reading quickly, completely);
- **As a homonym** – it means two independent actions (to read and to throw).

This shows that the differentiation of verb combinations based on context is an urgent problem in the processes of **machine translation, morphological analysis** and **automatic syntax separation**. [11]

Criteria for differentiating analytical forms from compound verbs in homonymy

1. Analysis from the point of view of grammatical structure

The main characteristics in analyzing the homonymous properties of verb compounds are **the independence of the components, the degree of syntactic connection, and the property of having a grammatical form**.

Characters	Analytical form (adverb + auxiliary verb)	Compound verb (verb + verb)
Semantic integrity	Semantically, it forms a unit	A new lexical unit creates a new meaning
Independence of components	The auxiliary verb does not have an independent meaning	Both verbs are separate independent lexical units
Intonational emphasis	It is not separated intonationally	The emphasis is equally distributed between the two components

Characters	Analytical form (adverb + auxiliary verb)	Compound verb (verb + verb)
Syntactic separability	It forms a unit, no word is inserted in the middle	Sometimes divisibility is possible (<i>also acquired</i>)
Dropout test	When the auxiliary verb is omitted, the meaning decreases	When reduced, the meaning changes radically

2. Semantic differentiation of homonym forms

The form "*o'qib tashladi*" ("threw it away") is one of them and can be differentiated in the following semantic contexts:

(a) **Analytic form** – He read the book (finality, speed):

- The action process is completed and emphasized.
- “Threw” as an auxiliary verb indicates the completion of the reading action.

(b) **Independent predicative units in the homonym case** – He read the letter (and then threw it):

- There are two actions: the first is reading, the second is throwing.
- In this case, “threw” is analyzed as an independent predicate.

3. Approach in computer linguistics

Morphological analyzers and machine translation systems, when determining this homonym:

- Enter a **list of auxiliary verbs** into the database;
- Use disambiguation algorithms (i.e., eliminating ambiguity) based on **contextual analysis**;
- Generates and automatically identifies semantic **lists of verb combinations**.

Example: "verb_combination": "read",

"type": "ambiguous",

"possible_analyses":

"structure": "analytic_form",

"meaning": "read completely/intensively"

"structure": "two_predicates",

"meaning": "read and then threw it away"

Conclusion

The homonymy between analytical forms and compound verbs in the Uzbek language requires complex linguistic analysis, especially in applied linguistics and computational linguistics. The correct identification of these constructions is important for:

- machine translation,
- automatic text analysis,
- semantic tagging,
- and the formation of lexical databases.

References

1. Abduraxmonova N. Korpus lingvistikasi (darslik). – Toshkent: Globeedit, 2023. –357 b.
2. Abduraxmonova N. O‘zbek tili elektron korpusining kompyuter modellari (monografiya) – Germaniya: Globeedit, 2021. –200 b.
3. Abjalova M. Korpus lingvistikasi (uslubiy qo‘llanma). Toshkent: Bookmany print, 2010.
4. 4.Абжалова М.А. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар таҳрири дастури учун): филол.фан.бўйича фалсафа док. (PhD)...дисс. – Фарғона, 2019. – 164 б.

5. Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn]: monografiya. – Toshkent, 2020. – 176 b
6. Abdurahmonova N. Mashina tarjimasining lingvistik ta'minoti (monografiya). – Toshkent: Muharrir, 2018. –165 b.
7. Elov, B. B., Khamroeva, S. M., & Xusainova, Z. Y. The pipeline processing of NLP. In *E3S Web of Conferences* (Vol. 413, p. 03011). EDP Sciences. (2023).
8. Khamrayeva, Sh.M. Specific and prevalent peculiarities of the authorship corpus // *IMPACT: International Journal of Research in Humanities, Arts and Literature*, 6(6), Jun 2018. – P. 431-438.
9. . Khamroeva, Sh. Morphotactic rules in the morphological analyzer of the uzbek language // *Middle European scientific bulletin*, 6, NOVEMBER 2020. ISSN 2694-9970 – P. 45-49.
10. Khamroeva, Sh.M., Gulyamova, Sh. K.. Electronic dictionaries – the product of applied linguistics. *ISJ Theoretical & Applied Science*, 07(87), 2020 – 463-466.
11. Менглиев Б., Шахабитдинова С., Хамроева С., Гулямова С., Ботирова. [Морфологический анализ и синтез словоформ в лингвистическом анализаторе.](#)
12. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б. 22. – 105 b 20.
13. Хамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. Фил.фан.док.(DCs) дис.аftoref. – Тошкент, 2021. – 65 б.
14. Хамроева, Ш., & Гулямова, Ш. Полифункционал сўзлар ва уларнинг семантик анализатори хусусида. In «Узбекские национальные образовательные здания теоретическое и практическое создание вопросы" Международная научно-практическая конференция 2022, June. (Vol. 1, No. 1).
15. Ҳожиев А. Ўзбек тилида кўмакчи феъллар. Монография. – Тошкент: Фан, 1966. – 223 б. – 3 б.