IEEE
TÜRKİYE SECTION

UBMK'25

Bildiriler Kitabı
Proceedings

Editör Eşref ADALI

10. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı

10th International Conference on Computer Science and Engineering

17-18-19 Eylül (September) 2025 İstanbul - Türkiye

IEEE TÜRKİYE SECTION

UBMK'25

Bildiriler Kitabı
Proceedings

Editor Eşref ADALI

10. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı

10th International Conference on Computer Science and Engineering

17-18-19 Eylül (September) 2025 İstanbul - Türkiye

Media type	Part Number	ISBN	Online ISSN
XPLORE COMPLIANT	CFP25L97-ART	979-8-3315-9975-1	2521-1641
CD-ROM	CFP25L97-CDR	979-8-3315-9974-4	

10. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı (UBMK'2025)

10th International Conference on Computer Science and Engineering

17-18-19 Eylül 2025 -İstanbul-Türkiye
17-18-19 September 2025 - İstanbul-Türkiye

IEEE Catalog Number : CFP25L97-ART

ISBN : 978-8-3315-9975-1

Additional copies mey be ordered from:
Curran Associates, Inc
57 Morehouse Lane Red Hook, NY 12571 USA
Phone: (845) 758 0400
Fax: (845) 758 2633
E-mail: curran@proceeding.com

UBMK’2025’ye Hoşgeldiniz
Welcome to UBMK’2025

Sevgili Katılımcılar:

UBMK uluslararası nitelikli konferans serisi, 1990 yılından beri düzenli olarak yapılmakta olan Bilgisayar Mühendisliği Bölüm Başkanları toplantılarında alınan bir kararla on yıl önce başlamıştır. Konferansın 10.su IEEE-UBMK-2025 bu yıl 17-18-19 Eylül, 2025 günlerinde İstanbul Teknik Üniversitesinin ev sahipliğinde düzenlemiştir.

IEEE-UBMK-2025 konferansına bu yıl Almanya, Amerika Birleşik Devletleri, Azerbaycan, Fransa, Irak, İngiltere, İsveç, İtalya, Kanada, Kazakistan, Kırım, Kırgızistan, Rusya, Özbekistan, Tataristan, Taylant, Ürdün ve Türkiye’den 610 dolayında bildiri gönderilmiş ve bu bildiriler Türk ve yabancı 250 hakem tarafından değerlendirilmiştir.

Her bildiri en az iki hakem tarafından incelenmiş ve uzlaşma olmadığı durumlarda üçüncü bir hakemin değerlendirmesine başvurulmuştur. Bildiri başına düşen ortalama hakemlik 2,3 olmuştur. Bu değerlendirmelerin sonunda 327 bildirinin sözlü olarak sunulması uygun bulunmuştur. Kabul edilen ve sunulan bildiriler içerik ve kalite ölçünlerini sağlaması durumunda IEEE Xplore’da yayımlanacaktır.

Konferans çalışmalarında, Bilgisayar Mühendisliği Bölüm Başkanları Danışma Kurulu olarak görev almışlardır. Bildirilerin değerlendirilmesi Bilim Kurulu üyeleri tarafından yapılmıştır. Konferansın düzenlenmesi ise Yürütme Kurulunun önerileri doğrultusunda, Düzenleme Kurulu tarafından yapılmıştır.

Son olarak, konferansın başarılı bir şekilde yürütülmesi için tüm olanaklarını sunan İstanbul Teknik Üniversitesi Rektörü Sayın Prof. Dr. Hasan Mandal’a teşekkür ediyoruz. Ayrıca Düzenleme Kuruluna, bildirileri titizlikle değerlendiren Bilim Kurulu Üyelerine ve değerli araştırmalarının sonuçlarını bilişim camiası ile paylaşan bildiri sahiplerine teşekkürlerimizi iletiriz.

Prof. Dr. Eşref ADALI
UBMK-2025 Konferans Başkanı ve Bildiri Kitabı Editörü

Dear Participants:

The UBMK international conference series started nine years ago with a decision taken at the Computer Engineering Department Heads (BMBB) meetings, which have been held regularly since 1990. The 10th edition of the conference, UBMK'25, was held this year on October 17-18-19, 2025, hosted by İstanbul Technichal University.

This year, approximately 610 papers were submitted to the IEEE-UBMK-2025 conference from Germany, the United States, Azerbaijan, France, Iraq, the United Kingdom, Sweden, Italy, Canada, Kazakhstan, Crimea, Kyrgyzstan, Russia, Uzbekistan, Tatarstan, Thailand, Jordan, and Turkey, and these papers were evaluated by 250 Turkish and foreign referees.

Each paper was evaluated at least by two referees, and in cases where there was no consensus, a third referee was consulted. At the end of these evaluations, 327 papers were accepted for oral presentation. Accepted and presented papers will be submitted for inclusion into IEEE Xplore subject to meeting IEEE Xplore's scope and quality requirements.

During the conference, Heads of Information Engineering Departments took part in the Advisory Board. The evaluation of the papers was made by the members of the Scientific Committee. The conference was organized by the Organizing Committee in line with the recommendations of the Executive Committee.

Finally, we would like to thank İstanbul Technical University Rector Prof. Dr. Hasan Mandal for his continued support for the success of the conference. In addition, we would like to thank the Organizing Committee, the Scientific Committee Members who carefully evaluated the papers, and the owners of the papers who shared the results of their valuable research with the informatics community.

Prof. Dr. Esref ADALI
UBMK'25 Conference Chair and Proceedings Editor

Düzenleyenler / Organizer

Destekleyenler / Sponsors

Automatic Identification of Phrase Structures in Uzbek Texts

Elov Botir Boltayevich
Dept. of Computational Linguistics and Digital Technologies
Tashkent State University of Uzbek Language and Literature named Alisher Navo'i
Tashkent, Uzbekistan
elov@navoiy-uni.uz

Abdullayeva Oqila Xolmo'minovna
Dept. of Computational Linguistics and Digital Technologies
Tashkent State University of Uzbek Language and Literature named Alisher Navo'i
Tashkent, Uzbekistan
abdullayevaoqila@gmail.com

Abdullayeva Nazokat Isayevna
Dept. of Computer Systems The Samarkand branch of Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
Samarkand, Uzbekistan
nazokat_abdullayeva_78@mail.ru

Saodat Israilova
Dept. of Uzbek Linguistics
National University of Uzbekistan named after Mirzo Ulugbek
Tashkent, Uzbekistan
israilovas404@gmail.com

Nizomova Fotima Bo'rixo'ja qizi
Dept. of Computational Linguistics and Digital Technologies
Tashkent State University of Uzbek Language and Literature named Alisher Navo'i
Tashkent, Uzbekistan
nizomovafotima@navoiy-uni.uz

***Abstract*—In recent years, various software systems have been developed in the Uzbek language to perform text processing tasks aimed at enhancing the effectiveness of machine translation, information retrieval, syntactic parsers, language learning tools, and lexicographic systems. In our study, we developed a program for the automatic identification of phrase structures, which represents one of the essential initial steps for building a syntactic parser for the Uzbek language. The availability of such a tool is crucial not only for accurate meaning extraction and syntactic analysis but also for solving alignment problems in translation systems. We investigated the syntactic templates of phrase structures in Uzbek and the patterns of dependency and head-modifier combinations. Based on this, we designed a rule-based system grounded in linguistic models and theoretical frameworks. The program was implemented using Python libraries and natural language processing (NLP) algorithms adapted for the Uzbek language. Uzbek phrase structures include governable, agglutinative, and concordant types. The system evaluates the effectiveness of identifying all three types of phrase structures in sentences, and the average accuracy of the program's performance was measured. A corpus of 100,000 Uzbek sentences was collected and analyzed. A rule-based approach was used to develop the rule set for phrase structure identification.**

Keywords—Phrase structure, syntactic pattern, tagging, modeling, linguistic model, POS tagging, morphological analysis.

I. INTRODUCTION

Modeling each grammatical rule and structure enables more efficient language learning and utilization in computer systems. Therefore, the tagging and modeling of phrase structures has become a relevant topic in the field of computational linguistics, contributing significantly to the development and advancement of natural language processing (NLP) applications [1]. In many cases, the smallest syntactic unit that carries semantic meaning in a text is not an individual word, but rather a phrase. Accurately identifying the boundaries of phrase structures plays a crucial role in the creation of modern language corpora, especially in the detection of terminological phrases, free word combinations, and idiomatic expressions. Automating this process requires reliance on the grammatical dependency rules inherent in the language, which guide the identification of syntactic relationships within text data.

In the Uzbek language, phrase structures are classified based on both the grammatical nature and the syntactic structure of their components. When classified according to grammatical properties, the categorization relies on the part of speech of the head (governing) word and the syntactic function of the dependent component. Phrase structures occur in both fixed (idiomatic) and free forms; free phrases, in turn, are formed through either coordinate or subordinate relations [2]. Structurally, Uzbek phrase constructions are largely similar to those found in other languages, including English. Both typically consist of a head word and a dependent (auxiliary) word [3]. In both languages, the phrase is classified according to the part of speech of the head word. With regard to structure, both languages distinguish between simple and complex phrases. However, unlike Uzbek, compound phrases are also recognized in English. English includes several distinct types of phrases, such as noun phrases, verb phrases, adjective phrases, adverbial phrases, prepositional phrases, verbal phrases, and absolute phrases, as well as expressions like phrasal verbs, collocations, and idioms [4]. In contrast, Uzbek phrase types differ somewhat, typically including noun phrases, verb phrases, adjective phrases, adverbial phrases, and modal-word phrases [5].

Methods

The realization of phrase structures within linguistic syntactic patterns is governed by significant syntactic regularities. Linguistic syntactic patterns (LSPs) represent a cognitive repertoire — a form of mental linguistic wealth — that speakers subconsciously rely on. The number of such patterns is highly limited. For example, the literature suggests that the number of fundamental sentence construction templates does not exceed 10 to 15. Research further indicates that there may be approximately 40 identifiable phrase structure templates. It is through the combinatorial application of these 50–60 LSPs that speakers are able to generate millions of concrete phrase structures and sentences in actual speech.

Linguistic syntactic patterns (LSPs) are further divided into patterns that generate nominative and communicative speech units. Phrase structures function as nominative syntactic units and differ in their linguistic (systemic) and speech (actualized) realizations. The realization of LSPs in phrase structures follows the language–speech continuum, which reflects a transition from general structures to specific, concrete utterances through intermediate representations.

For example:

[W → W]

[W + $_{\text{morphological marker}}$ → W + $_{\text{morphological marker}}$]

[$W_{\text{genitive case}}$ → $W_{\text{possessive suffix}}$]

[$\text{Noun}_{\text{genitive case}}$ → $\text{Noun}_{\text{possessive suffix}}$]

[Common noun $_{\text{in genitive case}}$ → Common noun $_{\text{with possessive suffix}}$]

[Common noun $^{\text{as dependent}}$ + $_{\text{genitive}}$ → Common noun $^{\text{as head}}$ + $_{\text{possessive}}$]

In the phrase Madinaning yozuvi (Madina's writing), the general LSP is represented by the structure [W → W], whereas the more specific LSP is [Common noun as dependent + genitive case → Common noun as head + possessive suffix].

O. Abdullayeva and S. Khudayarova have expressed the view that, in Uzbek linguistics, phrase structures are considered complex units based on their structural formation, the grammatical relations they establish, and the fact that in some cases they are equivalent to a single word, while in others they functionally correspond to a full sentence [6]. They note that it is difficult to assign a single fixed template to such units.

J. Ibragimov notes that in modeling phrase structures within Uzbek texts, he relies on generally recognized linguistic labels, and provides the following illustrative examples. In the sentence "Ustodning muborak ko‘zoynaklarini taqib, xalqimning o‘tmishiga qaradim" ("Wearing the honorable teacher's glasses, I looked into my people's past"), he analyzes the phrase "o‘tmishiga qaradim" as follows: o‘tmishiga qaradim – N + SFN + V + SFV, where: N = noun (OT), SFN = syntactic forms of the noun, V = verb, SFV = syntactic forms of the verb. A submodel of the typical structure "o‘tmishiga qaradim" can be presented as:

N + SG + PS_aff + Acc_CS_aff + V + TS_aff + ShS_aff, where: N = noun, SG = singular number, PS_aff = possessive suffix, Acc_CS_aff = directional case suffix (jo‘nalish kelishigi), V = verb, TS_aff = tense suffix (from "Tense"), ShS_aff = person-number suffix (representing both person and number agreement); taqib qaramoq – V + V + SFV, where: V = verb, SFV = syntactic forms of the verb. xalqimning o‘tmishiga qaradim – Nd + SFN + Nk + SFN + V + SFV, where the structure reflects nested dependencies among possessive and directional case forms; ustodning ko‘zoynaklari – N + SFN + N + SFN, where each N stands for a noun and each SFN represents its syntactic form; muborak ko‘zoynaklari – ADJ + N + PS_aff, where: ADJ = adjective (sifat), N = noun, PS_aff = possessive suffix [7].

In addition, S. Nazarova, in her research, developed a model for noun+noun pattern-based phrase constructions in Uzbek. The patterns are illustrated as follows:

1) [Noun $_{\text{in genitive case}}$ → Noun $_{\text{with possessive suffix}}$ = $^{\text{as dependent}}$ noun + $^{\text{as head}}$ noun]
Example: *kitobning varag‘i* (the page of the book)

2) [Noun $_{\text{in genitive case}}$ → Adjective $_{\text{with possessive suffix}}$ = $^{\text{as dependent}}$ noun + $^{\text{as head}}$ adjective]
Example: *daraxtning mo‘rti* (the brittle part of the tree)

3) [Adjective $_{\text{in genitive case}}$ → Noun $_{\text{with possessive suffix}}$ = $^{\text{as dependent}}$ adjective + $^{\text{as head}}$ noun]
Example: *gulning/qizilining hidi* (the smell of the flower / the red one), and ets. [8].

II. Results And Discussion

An analysis of linguistic-syntactic pattern (LSP) combinations in phrase structures revealed that there are a total of 452 distinct LSPs in the Uzbek language. This analysis was carried out based on materials from the Uzbek language corpus (https://uzschoolcorpara.uz/). When examining the LSP statistics of phrase structures in POS-tagged sentences within the corpus, it was found that 36 distinct part-of-speech (POS) category combinations are involved in the formation of phrase structure templates. The dataset of phrase structures was compiled using POS-tagged simple and compound sentences from the Uzbek language corpus.

In the Uzbek language, the possible linguistic-syntactic patterns (LSPs) of prepositional phrase (boshqaruvli so‘z birikmasi), non-prepositional phrase (bitishuvli so‘z birikmasi), and possessive case phrase (moslashuvli so‘z birikmasi) structures were identified. To enable the annotation of these LSPs within the corpus, we developed a dedicated tagging system. A total of 428 LSPs were identified for prepositional phrase structures. Each pattern was analyzed in terms of how part-of-speech (POS) tags are combined in the output of the Uzbek morphological analyzer. For example, the LSPs of the phrases guldan nozik (more delicate than the flower) and onamga sovg‘a (a gift for my mother) correspond to the structures Ndan+JJ and Nga+N, respectively. When processed by the Uzbek morphological analyzer, these appear as N+JJ and N+N POS tag sequences. To label such prepositional phrase structures, we proposed the tag SBbosh. A search conducted in the Uzbek language corpus using this pattern retrieved a dataset of 1,252 phrase structure instances corresponding to governable combinations (Table I).

TABLE I. Linguistic-Syntactic Pattern and Model of the Prepositional Phrase Structure

T/r	Linguistic-Syntactic Pattern of a Phrase Structure	POS tag in the morphological analyzer	Corpus-derived linguistic dataset	General tagging of a prepositional phrase structure
1.	$N_{k.sh.}$+VB	N+VB	75	SBbosh
2.	$P_{k.sh}$+VB	P+VB	200	SBbosh
3.	$P_{k.sh}$+JJ	P+JJ	200	SBbosh
4.	$JJ_{k.sh}$+VB	JJ+VB	154	SBbosh
5.	$N_{k.sh}$+JJ	N+JJ	29	SBbosh

Total:	428 ta	12	1252	

An investigation of possible variants of possessive case phrase structures revealed a total of 14 distinct linguistic-syntactic patterns (LSPs), based on analysis and examples. For these structures, the use of the tag SBmos was proposed. Analyses were conducted on the Uzbek language corpus, and morphological POS tagging was identified as a key factor in detecting such phrase structures. For instance, in concordant phrases such as Aqllining fikri (the opinion of the intelligent one) and Kelajakning ishonchi (Future's trust), the morpheme-level morphological analyses yielded the patterns JJning + Ni and Nning + Ni, respectively. A total of 960 instances of this phrase type were collected as a dataset from the Uzbek language corpus (Table II).

TABLE II. LINGUISTIC-SYNTACTIC PATTERN AND MODEL OF THE POSSESSIVE CASE PHRASE STRUCTURE

T/r	Linguistic-Syntactic Pattern of a Phrase Structure	POS tag in the morphological analyzer	Corpus-derived linguistic dataset	General tagging of a possessive case phrase structure
1.	N*q.k*+N*e.sh*	N+N	100	SBmos
2.	N*q.k* +P*e.sh*	N+P	50	SBmos
3.	$P_{q.k}$ +$UH_{e.sh}$	P+UH	7	SBmos
4.	$N_{q.k}$ +$UH_{e.sh}$	N+UH	11	SBmos
5.	$P_{q.k}$ +$N_{e.sh}$	P+N	241	SBmos
Total:	14 ta	14 ta	960 ta	

An analysis of the linguistic-syntactic patterns (LSPs) of non-prepositional phrase structures revealed several problematic aspects. Qualitative non-prepositional phrase does not exist in the Uzbek language. In phrase structures formed via agglutination, the head element cannot be an adjective. When the components of a phrase structure appear in such a form, the grammatical relation no longer expresses a conceptual unit (i.e., a phrase), but rather constitutes a complete proposition. For example, constructions like bino katta (the building is big) or ko'ylak chiroyli (the dress is beautiful) are not phrase structures, but instead are classified as simple sentences of the type S + P. In ko'ylak chiroyli, ko'ylak functions as the subject, and chiroyli functions as the nominal predicate. Similarly, bino katta follows the S + NP syntactic template. After analyzing all possible variants of non-prepositional phrases in speech, it was determined that 10 LSPs exist for this type of structure. Sentences in the corpus were POS-tagged using a morphological analyzer, and based on the morphological analysis, syntactic patterns for phrase structure identification were developed. For instance, in the non-prepositional phrases o'zi bilmoq (to know oneself) and oltin soat (gold watch), the corresponding POS tag combinations identified in the corpus were P + VB and N + N, respectively. For such constructions, the syntactic label SBbitishuv was proposed. A total of 1,885 instances of non-prepositional phrase structures were collected from the Uzbek language corpus (Table III).

TABLE III. LINGUISTIC-SYNTACTIC PATTERN AND MODEL OF THE NON-PREPOSITIONAL PHRASE STRUCTURE

T/r	Linguistic-Syntactic Pattern of a Phrase Structure	POS tag in the morphological analyzer	Corpus-derived linguistic dataset	General tagging of a non-prepositional phrase structure
1.	JJ+N	JJ+N	502	SBbitishuv
2.	NUM+N	NUM+N	523	SBbitishuv
3.	P+N	P+N	128	SBbitishuv
4.	JJ+VB	JJ+VB	26	SBbitishuv
5.	VB+N	VB+N	212	SBbitishuv
Total:	10	10	1885	

In this study, a total of 452 linguistic-syntactic patterns of phrase structures in the Uzbek language were identified. Using the bigram and trigram functions of the Uzbek morphological analyzer corpus, 36 types of phrase combinations were tested in the software environment. To empirically demonstrate the existence of these combinations in Uzbek, 4,028 phrase structures were extracted.

Automatic Identification of Phrase Structures in Uzbek Texts

The task of identifying phrase structures is considered relevant not only within the domain of traditional syntax, but also in modern computational linguistics and natural language processing (NLP). Indeed, the analysis of complex syntactic structures, as well as their automatic extraction and classification, plays a critical role in the effective functioning of numerous applied systems — such as machine translation, text indexing, semantic search, and linguistic analysis platforms. In the Uzbek language, phrase structures are primarily formed on the basis of three types of linguistic relations: Prepositional phrase (boshqaruvli so'z birikmasi), non-prepositional phrase (bitishuvli so'z birikmasi), and possessive case phrase (moslashuvli so'z birikmasi). Therefore, the automatic phrase detection system developed in this study was designed based on these theoretical foundations.

The program employs a rule-based approach for identifying phrase structures. For each type of phrase, a specific syntactic template was developed. These models take into account both the morphological and syntactic dependencies between words. As a result, in order to detect prepositional phrase, non-prepositional phrase, and possessive case phrase structures, it was essential to develop rule systems grounded in theoretical linguistics that reflect the morphosyntactic features of each type. The software was implemented by designing a set of structured patterns and rules as outlined below.

Linguistic-Syntactic Pattern and Rule of prepositional phrase Structure

1. Example pattern: N + [–ni / –ga / –da / –dan / postpositions] + VB. Rule: If a noun (N) in the text is followed by a specific case suffix (such as –ni, –ga, –da, –dan) or a postposition (e.g., bilan, uchun, haqida, bilan birga) and is then followed by a verb (VB), this construction is identified as a prepositional phrase structure.
2. Example pattern: P + [–ni / –ga / –da / –dan / postpositions] + VB. Rule: If a pronoun (P) occurs with a specific case suffix (–ni, –ga, –da, –dan) or a postposition, and is then followed by a verb (VB), it forms a prepositional phrase structure.
3. Example pattern: P + [–ni / –ga / –da / –dan / postpositions] + JJ. Rule: If a pronoun (P) is followed by a case suffix or a postposition (e.g., bilan, uchun, haqida) and then an adjective (JJ), the resulting

construction is considered a prepositional phrase structure.

4. Example pattern: P + [–ni / –ga / –da / –dan / postpositions] + MD. Rule: If a pronoun (P) is accompanied by a case suffix or a postposition (e.g., bilan, uchun, haqida), and then followed by a modal word (MD, such as kerak, lozim, mumkin, shart), it is classified as a prepositional phrase structure.

Linguistic-Syntactic Patterns and Rules of possessive case Phrase Structures

1. Example pattern: N + [–ning] + N + [possessive affix: –m, –im, –ng, –ing, –si, –miz, –imiz, –ngiz, –ingiz, –lari]. Rule: If a noun (N) appears in the genitive case (-ning), and is followed by another noun bearing a possessive suffix, the resulting construction is considered a possessive case phrase structure.
2. Example pattern: N + [–ning] + P + [possessive affix]. Rule: If a noun (N) in the genitive case (-ning) is followed by a pronoun (P) that bears a possessive suffix, the phrase is identified as a possessive case phrase structure.
3. Example pattern: N + [–ning] + Num + [possessive affix]. Rule: If the first word is a noun (N) in the genitive case (-ning), and is followed by a numeral (Num) that carries a possessive suffix, the construction is considered a possessive case phrase.
4. Example pattern: N + [–ning] + VB (verbal noun) + [possessive affix]. Rule: If the first word is a noun in the genitive case (-ning), and is followed by a nominalized verb (VB) that carries a possessive suffix, the resulting structure is classified as possessive case phrase.

Linguistic-Syntactic Patterns and Rules of non prepositional Phrase Structures

1. Example pattern: JJ + N. Rule: If an adjective (JJ) is immediately followed by a noun (N) in the text, with no intervening postpositions or conjunctions, the resulting sequence is identified as an non-prepositional phrase structure.
2. Example Pattern: Num + N. Rule: If a numeral (Num) is directly followed by a noun (N) without any other elements (such as postpositions or conjunctions) between them, they form an non-prepositional phrase structure.
3. Example pattern: P + N. Rule: If a pronoun (P) directly precedes a noun (N) with no intervening elements, and no postpositions or connectives are present, this construction is classified as an non-prepositional phrase structure.
4. Example pattern: N + N. Rule: If two nouns (N + N) appear consecutively, with no other words (e.g., postpositions or conjunctions) between them, and they are tightly semantically bound to form a new meaning, they constitute an non-prepositional phrase structure.

Linguistic Aspects Necessary for the Effective Operation of an Automatic Phrase Identification System

To ensure the effective performance of an automatic phrase identification system, the following linguistic aspects must be taken into account:

Part-of-Speech Tagging: Determining the grammatical category of each word—such as noun, adjective, verb, numeral, pronoun, etc.—plays a crucial role in identifying the type of phrase.

Morphological Analysis: Identifying the grammatical forms within a word (e.g., case endings, possessive, and plural suffixes) is essential for recognizing syntactic relationships. Special attention must be given to affixes such as –ning, –ni, –ga, –dan, –i, –si, –imiz, –lari, as they are critical in detecting and classifying phrase structures.

Immediate Word Proximity: Especially in non-prepositional phrases, it is important that no other words intervene between two elements. For example, katta qora mushuk (a big black cat) may be considered a phrase composed of three tightly linked words.

Pattern Matching: Each pair or triplet of words is compared against pre-defined linguistic templates. If the conditions of a given pattern are met, the corresponding phrase type is identified.

This process can be represented through the following procedural flowchart (Fig. 1.):

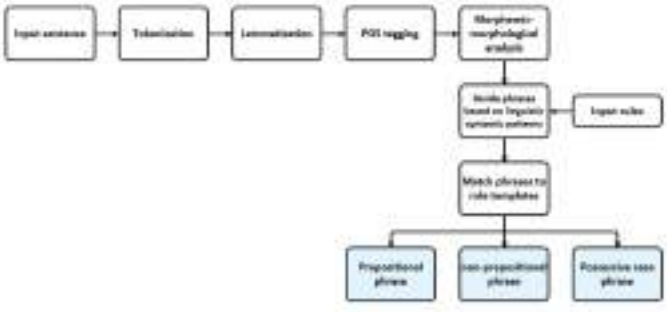

Fig. 1. Overall Block Diagram of the Program Workflow

The system executes the following primary processes (Table IV):

TABLE IV. MAIN FUNCTIONS OF THE SOFTWARE MODULE

T/r	Function	Imported Library
1.	Tokenization	Nltk, re
2.	POS tagging	Nltk.pos_tag, spacy
3.	Lemmatization	Spacy, stanza, pymorphy2
4.	Morphological Analysis	Stanza, pymorhy2
5.	Matching linguistic-syntactic structures	Python's conditional statements (if, elif, else) and regular expressions (re)
6.	Classification of word combinations (phrases)	Categorization of phrase types (Prepositional phrase, non-prepositional phrase, possessive case phrase) according to their matched linguistic-syntactic templates
7.	Generating Output	Displayed via the terminal and graphical user interface

Program Architecture

The program is written in a modular structure, with each component performing a specific task. The core module

analyzes the input text and identifies phrase structures. Subsequently, it matches them to the appropriate pattern-based models.

User Interface of the Uzbek Phrase Identification System

The interface of the system for automatic phrase identification in Uzbek text allows users to input sentences, process them through the analysis pipeline, and view the detected phrase structures based on linguistic-syntactic templates (Fig. 2 and Fig. 3.).

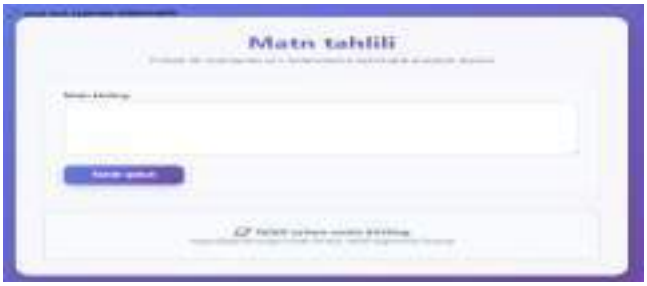

Fig. 2. User Interface of the Automatic Phrase Identification System for Uzbek Texts

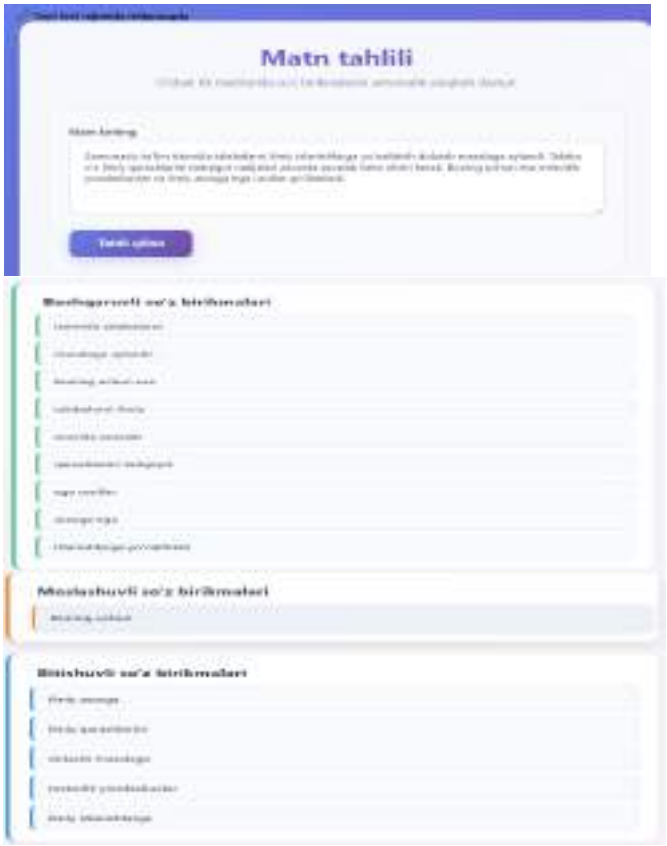

Fig. 3. Analysis and Results

III. Study Of The Problem

As part of our research, we tested 100 sentences for each type of phrase structure in the developed program for automatic phrase identification in Uzbek texts. The accuracy and precision during the execution of the system showed high accuracy for Prepositional phrase (boshqaruvli) phrase structures, while the possessive case phrase (moslashuvli) phrase structures was significantly lower. For non-prepositional phrase (bitishuvli) phrase structures, the accuracy exceeded 50%. Below, the process of testing the examples and obtaining the results is analyzed.

Accuracy of Identifying Prepositional Phrase Structures

- Total number of sentences: 100
- Number of actual prepositional phrase: 100
- Number of correctly identified Prepositional phrase by the system: 76

Based on this, the accuracy is calculated proportionally.

Formula:

$$Accuracy\ (Precision\ or\ Effectiveness) = \left(\frac{Correctly\ Detected\ Instances}{Total\ Number\ of\ Real\ Phrase\ Instances}\right) \times 100\%$$

$$Accuracy\ (Precision\ or\ Effectiveness) = \left(\frac{76}{100}\right) \times 100\% = 76\%$$

Incorrectly Identified and Unidentified Instances:

100%−76%=24%

The system correctly identified 76 out of 100 Prepositional phrase structures, resulting in an accuracy rate of 76%. This is considered a satisfactory outcome for a rule-based approach. However, to improve the system's performance in subsequent stages, it is necessary to analyze the remaining 24% of incorrect or missed cases.

1) Accuracy of Identifying possessive case phrase Structures

- Total number of sentences: 100
- Actual number of possessive case phrase structures: 126
- Correctly identified by the system: 24
-

Formula:

$$Accuracy\ (Precision\ or\ Effectiveness) = \left(\frac{Correctly\ Detected\ Instances}{Total\ Number\ of\ Real\ Phrase\ Instances}\right) \times 100\%$$

$$Accuracy\ (Precision\ or\ Effectiveness) = \left(\frac{24}{126}\right) \times 100\% \approx 19.05\%$$

Incorrectly Identified and Unidentified Instances:

100%−19.05%=80.95%

The system correctly identified only 24 possessive case phrases, resulting in an accuracy rate of 19%. This relatively low accuracy suggests that the current model for identifying possessive case phrase structures requires further refinement and improvement.

2) Accuracy of Identifying non-prepositional phrase Structures

- Total number of sentences: 100
- Actual number of non-prepositional phrase structures: 94
- Correctly identified by the system: 53

Formula:

$$Accuracy\ (Precision\ or\ Effectiveness) = \left(\frac{Correctly\ Detected\ Instances}{Total\ Number\ of\ Real\ Phrase\ Instances}\right) \times 100\%$$

$$Accuracy\ (Precision\ or\ Effectiveness) = \left(\frac{53}{94}\right) \times 100\% \approx 56.38\%$$

Incorrectly Identified and Unidentified Instances: 100%−56.38%=43.62%

The system correctly identified 56.38% of the non-prepositional phrase structures. This result indicates a moderate level of performance, suggesting that the current model is functional and useful; however, the presence of approximately 44% unrecognized cases highlights the need for further analysis and refinement of rule-based components.

To improve the system's precision, it is necessary to develop additional rules capable of capturing currently unrecognized constructions and to consider the integration of statistical and machine learning-based models. This would enhance both the accuracy and processing speed of the phrase identification system.

3) Average Efficiency of the System by Phrase Type

- Prepositional phrase: 76%
- possessive case phrase: 19.05%
- non-prepositional phrase: 56.38%

Formula:

$$SAverage\ Accuracy = \left(\frac{S1 + S2 + S3}{n}\right)$$

$$Average\ Accuracy = \left(\frac{76 + 19.5 + 56.38}{3} = \frac{151.43}{3} \approx 50.48\%\right)$$

Analysis of the Overall Efficiency of the System

Based on the results above, the system demonstrated varying levels of effectiveness in identifying Prepositional phrase, possessive case phrase, and non-prepositional phrases in Uzbek language texts. The accuracy for detecting Prepositional phrases was relatively high at 76%, while for possessive case phrases, it was the lowest at 19.05%. In the case of non-prepositional phrases, the system achieved a moderate accuracy of 56.38%.

By calculating the arithmetic mean of these individual results, the overall efficiency of the system was determined as follows:

$$Average\ Accuracy = \left(\frac{76 + 19.5 + 56.38}{3} = \frac{151.43}{3} \approx 50.48\%\right)$$

This indicates that the system operates with an overall accuracy of 50.48%, meaning that it produces a correct result in approximately one out of every two instances. This level of performance may be considered acceptable for an initial prototype stage, but highlights the need for further improvement — particularly in enhancing the accuracy of identifying possessive case phrases, expanding the rule base, and integrating machine learning components.

Moreover, the analysis results provide insights into which models and approaches yield higher precision in the development of a syntactic parser. The obtained average score can serve as a benchmark for evaluating the effectiveness of hybrid approaches (rule-based + machine learning) in future implementations.

IV. Conclusion

A specialized software and linguistic infrastructure has been developed for the identification of phrase structures in the Uzbek language. The proposed system is based on models formulated in accordance with the syntactic and morphological rules of Uzbek. Using a set of rule-based models, the program identifies grammatical dependencies between words and classifies them into the corresponding phrase types — Prepositional, non-prepositional, or possessive case constructions. The study revealed that the automatic detection of phrase structures using a rule-based approach faces challenges such as complex syntactic relations and lexical ambiguity (homonymy). This research represents an initial step toward the development of a full syntactic parser. In subsequent phases, the system is expected to be improved by integrating rule-based methods with machine learning techniques, thereby enhancing its performance and linguistic accuracy.

References

[1] O. Abdullayeva and F. Nizomova, "O‘zbek tilida moslashuv yo‘li bilan birikkan so‘z birikmalarini modellashtirish masalasi," O‘zbekiston: til va madaniyat. Kompyuter lingvistikasi, "O‘zbekiston: til va madaniyat" akademik jurnali, elektron nashr, Toshkent: ToshDO‘TAU, 2025, p. 34

[2] R. Sayfullayeva, B. Mengliyev et al., Hozirgi o‘zbek adabiy tili, Toshkent, 2009, pp. 305–324.

[3] G. Booij, Word-formation in construction grammar. Construction Morphology, Oxford University Press, 2010, p. 10.

[4] R. Huddleston and G. K. Pullum, The Cambridge Grammar of the English Language, Cambridge University Press, 2002, ch. 5–8.

[5] S. M. Khamroeva, O. X. Abdullayeva, N. S. Matyakubova, A. A. Rakhmanova, and D. K. Khudoyqulova, "Alignment of Phrases in Uzbek-English Languages," 2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE), Novosibirsk, Russia, 2024, pp. 1830–1833. doi: 10.1109/PIERE62470.2024.10805057.

[6] O. Abdullayeva and S. Xudayarova, "O‘zbek tilshunosligida so‘z birikmasiga ta'rif, tavsif va tasnif masalasi," International scientific–theoretical conference on the topic: Problems of Research and Education of the Uzbek Language, myscience.uz, 2023, pp. 84–97.

[7] J. Ibragimov, "O‘zbek tili matnlarida so‘zlarning bevosita (kontakt) va bilvosita (distant) birikuvchanligini dastlabki dasturiy modellashtirish," in Zamonaviy dunyoda innovatsion tadqiqotlar: Nazariya va amaliyot, masofaviy onlayn konferensiya materiallari, 2023. https://doi.org/10.5281/zenodo.7515431.

[8] S. Nazarova, Birikmalarda so‘zlarning erkin bog‘lanish omillari: filol. fan. nomz. diss. avtoreferat, Toshkent, 1997, p. 26.

[9] I. A. Azimova, K. M. Mavlonova, O. I. Saidaxmedova, R. K. Atamuratov, and O. Abdullayeva, "The order of morpheme acquisition in Uzbek language (examples of Chinese students who learning Uzbek as a second language)," E3S Web of Conferences, vol. 413, p. 03006, 2023. doi: 10.1051/e3sconf/202341303006.

[10] A. R. Kadirjanovich, A. N. Isayevna, P. G. Gulomjonovna, and P. S. Botir qizi, "Improving the Methodology of Teaching Specialized Subjects in The Preparation of Future Computer Engineering on The Basis of Innovative Technologies," 2023 8th International Conference on Computer Science and Engineering (UBMK), Burdur, Turkey, 2023, pp. 125–130. doi: 10.1109/UBMK59864.2023.10286726.