Proceedings of the 6th International Conference on Image Processing and Capsule Networks (ICIPCN-2026)
DVD Part Number: CFP26VY4-DVD; ISBN: 979-8-3315-9980-5

Analysis of Models for Building Dependency Parsing in Agglutinative Languages

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.oqila@navoiy-uni.uz

Norkuziev Doniyorbek Bakhtiyorovich
Dept.of Psychology
Uzbekistan Oriental University
Tashkent, Uzbekistan
norqoziyevdoniyorbek@gmail.com

Bayjonov Furqat Baxramovich
Dept.of Continuous Education Pedagogy
Oriental University
Tashkent, Uzbekistan
furqat100190@gmail.com

O‘tkirova Fotima Baxtiyorovna
Master’s Student, Dept.of Computational Linguistics and Digital Technologies
Tashkent State University of Uzbek Language and Literature named Alisher Navo’i
Tashkent, Uzbekistan
otkirovafotima456@gmail.com

***Abstract*—In the domain of computational linguistics, syntactic analysis—that is, the comprehension of the interrelationship between words-plays a pivotal role in natural language processing (NLP). Dependency parsing, a fundamental component of computational linguistics, offers a methodical approach to syntactic analysis. On a global scale, one of the most advanced and vital branches of NLP is dependency parsing, which focuses on detecting dependency relations between words. This technique is of particular significance in applications such as machine translation and sentiment analysis. The present article discusses the construction of dependency parsing models for agglutinative languages, their application to Uzbek sentence structures, the differences from other language types, and an analysis of foundational data resources and models for the Uzbek language. In this study, a BERT-based dependency parsing model is employed to analyze Uzbek sentence structures. The article discusses the construction of dependency parsing models for agglutinative languages and their application to the Uzbek language, highlighting differences from other language types. Furthermore, it provides an analysis of foundational linguistic resources and pretrained language models relevant to Uzbek. The paper delineates the distinguishing characteristics of grammatical dependencies in Uzbek sentences by leveraging contextual embeddings generated by the BERT model and juxtaposes these findings with traditional syntactic parsing approaches. Finally, the study evaluates the advantages and future potential of BERT-based dependency parsing for low-resource and agglutinative languages such as Uzbek.**

Keywords—Dependency parsing, tagging, dependency, syntactic analysis, model, and agglutinative language.

I. INTRODUCTION

The initial prototypes of language technologies in global linguistics began to emerge in the mid-20th century. One such method is the automatic syntactic analysis of written texts, known as parsing. Parsing constitutes a significant part of research in natural language processing (NLP). Automatic syntactic analysis functions as the preliminary processing stage for language applications that work with rich linguistic data, such as programs requiring semantic representations. Consequently, numerous language technologies are contingent upon advancements in syntactic analysis. At present, the majority of research conducted in the Uzbek language focuses on the restoration of natural language, the creation of automatic language systems, and the development of various grammatical, syntactic, and morphological models.

There are advanced approaches and algorithms for implementing dependency parsing. The advent of artificial intelligence and deep learning technologies has led to significant advancements in dependency parsing, resulting in enhanced accuracy and efficiency. Transformer architectures (e.g., BERT and GPT) have demonstrated particular efficacy in comprehending the syntactic and semantic structures of language.

In addition to these advancements, a variety of tools and software parsers are available for performing dependency parsing in global linguistics. For instance, SpaCy is recognized for its speed, accuracy, user-friendly API, and support for multiple languages. The dependency parsing module of the Stanford Natural Language Processing (NLP) suite employs advanced statistical models and academic approaches and is capable of analyzing languages such as Spanish, English, and French. Stanza is another open-source tool that utilizes models like BERT and others to perform dependency parsing through an intuitive interface [1].

II. RELATED WORKS

The study of dependency parsing holds both practical and theoretical significance in global linguistics. In the late 20th and early 21st centuries, numerous theoretical works on this subject were conducted. During this period, grammar systems evolved based primarily on formalized, rule-based structures.

In the 1950s, Lucien Tesnière introduced the concept of *dependency grammar*, proposing that the syntactic structure of a language could be expressed through systems of relationships. His work “Structure des relations de

dépendance dans le langage" (1959) laid the foundation for modern dependency models in linguistics [2].

Buchholz and Marsi unified 13 treebanks representing different languages into a single multilingual dependency format [3]. Ballesteros and Nivre extended the CoNLL-X dataset by integrating the Penn Treebank using an algorithm that transformed it into Stanford-style annotations [4].

Ruket Cakici worked on enhancing the effectiveness of the existing Sabancı Treebank and addressed several issues within the system. His study provides insight into analyzing Turkish dependency models using the METU-Sabancı Treebank, and it compares direct and indirect approaches. The author favors using Collins' parsing principles in dependency parsing due to their generative and maximum-coverage characteristics [5].

Foth and Menzel's combination of data-driven and rule-based components increased dependency parsing accuracy in German syntax by 92% [6]. Olteanu and Moldovan approached the analysis of prepositional phrases through isolation-based parsing, achieving an accuracy rate of 92.85% [7].

Turney and Pantel introduced the *distributional hypothesis*, which addresses certain characteristics of natural language [8]. Michael Collins, a prominent researcher in machine learning and NLP, played a vital role in advancing statistical approaches to dependency parsing and contributed to the development of several well-known parsers [9].

Markus Dreyer collaborated with J. Nivre on major projects such as MaltParser and Stanford Parser, which have addressed key challenges in linguistic research [10].
In Uzbekistan, several scholars have also worked on developing dependency parsing methods and linguistic templates. O. Abdullayeva and S. Khudayarova studied syntactic templates of Uzbek word combinations and their modeling [11]. O.J. Khidirov analyzed algorithms and systems for syntactic tagging of word combinations in Uzbek corpora [12]. S. Nazarova examined the syntactic structure of nominal phrases in Uzbek [13]. Sh.G. Kahramonovna studied how polyfunctional words form head-dependent relations with other words in the text, as a linguistic factor for semantic analysis tools [14].

An important practical contribution of this research is that it represents one of the first attempts to train and evaluate a dependency parsing model specifically for the Uzbek language using a linguistically grounded and empirically validated framework. Previous studies have primarily focused on constructing limited linguistic resources or exploratory annotations, without establishing a robust training pipeline or providing systematic performance evaluation. As a result, the proposed approach fills a critical gap between linguistic resource creation and model-based syntactic analysis for Uzbek.

III. MAIN BODY

To date, various models have been developed to implement dependency parsing. These models may be adapted for either a single language or multiple languages. While multilingual models can improve results in some cases, they may reduce accuracy when retrained across various languages. The more languages involved, the more complex the challenge—especially when the languages differ significantly in linguistic features.

For this reason, recent research has increasingly focused on analyzing models designed for agglutinative languages and selecting the most suitable ones for Uzbek. Globally, a wide range of models are available for identifying dependencies within sentences. Below, we present a classification and analysis of such models.

Recent research has significantly improved the quality and impact of academic and practical work by incorporating these models. However, challenges remain, such as accurately identifying word meanings and accounting for complementary components in dependency parsing. Agglutinative languages, with their complex grammatical structures, unique word-affix relationships, morphological richness, and multiple literal or figurative meanings, present specific challenges for dependency parsing and NLP research in general. In several agglutinative languages—including Finnish, Estonian, Hungarian, Indonesian, Japanese, Kazakh, Korean, Turkish, and Uyghur—various models have been tested for dependency parsing. These languages serve as benchmarks in comparative studies, which we will now analyze in more detail.

In the study conducted by Muchahit Oltintash and Cuneyd Tantug', practical measures were undertaken to enhance dependency parsing performance for nine agglutinative languages [15]. To improve dependency detection, they proposed two key optimization factors.

The first involves the use of subword units based on word representations. Subwords are considered an effective solution in neural machine translation (NMT) for addressing the open vocabulary problem [16]. These subunits also help mitigate challenges related to low-resource languages and out-of-vocabulary words in machine learning. Subwords are language-independent and serve as universal tokens within neural networks.

The second factor introduces the use of a sentence-level semantic token, in addition to regular token features. This allows sentences with the same meaning but different word order to be parsed using consistent dependency structures. The study evaluates the effect of these two factors on dependency parsing accuracy using nine widely spoken agglutinative languages: Estonian, Finnish, Hungarian, Indonesian, Japanese, Kazakh, Korean, Turkish, and Uyghur.

The researchers applied an improved version of the biaffine model originally developed by Dozat and Manning [17]. This model was the first successful attempt to integrate deep analysis into the system. It assigns probabilities at each step and uses specific tags and scores to determine the parsing actions for each word. Subsequent enhancements to the model were proposed by Weiss et al. (2015) [18] and Andor et al. (2016) [19], who incorporated neural network-based conditional random fields. Dyer et al. (2015) proposed an alternative architecture using LSTM networks to model the stack and buffer, achieving state-of-the-art results by enabling phrase-level parsing.

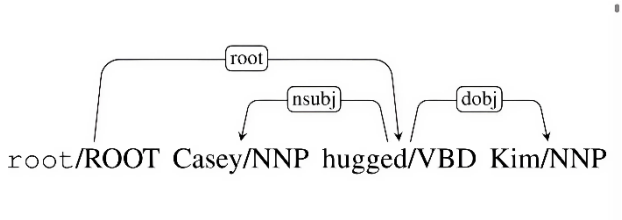

Fig. 1. Example of Deep Biaffine parsing [20].

Figure 1 above shows the analysis result in the Deep Biaffine model. The improved Deep biaffine model was enhanced with encoders and classifiers based on LSTM (Long Short-Term Memory) networks. Uniform hyperparameters were applied to all agglutinative languages in the study. However, since a tagging system has not yet been developed for the Uyghur language, no results were produced for it. Pretrained word embeddings were applied in the following order: ELECTRA, BERT, ELMo, and Word2Vec. That is, the model first attempted to use the ELECTRA language model (LM); if no ELECTRA model was available for the given language, it switched to BERT. If neither ELECTRA nor BERT were available, it used ELMo; if ELMo was also unavailable, it defaulted to Word2Vec embeddings.

For Estonian, Finnish, Hungarian, Indonesian, and Japanese, the BERT model produced successful results. In the case of Kazakh, none of the first three models yielded acceptable results, and the model was ultimately trained using Word2Vec embeddings. For Korean and Turkish, ELECTRA achieved high accuracy from the initial training phase. Table 1 shows alternative language models for 9 agglutinative languages.

TABLE I. MODELS TRAINED FOR 9 AGGLUTINATIVE LANGUAGES.

№	TIL	MODEL
1	Estoniya	BERT
2	Finlyandiya	BERT
3	Vengriya	BERT
4	Indoneziya	BERT
5	Yaponiya	BERT
6	Qozoq	Word2vec
7	Koreya	ELEKTRA
8	Turk	ELEKTRA
9	Uyg'ur	ELMO

For Finnish, multilingual BERT models were initially used. However, large-scale training on unannotated corpora caused a drop in performance. Later, an advanced BERT-based model known as FinBERT was developed, which became the most effective model for POS tagging, named entity recognition, and dependency parsing in Finnish [21]. FinBERT was evaluated on three Finnish corpora—Turku Dependency Treebank, FinnTreeBank, and Parallel UD—and achieved accuracy rates exceeding 90% on all datasets. For instance, it scored 91.93% on Turku Dependency Treebank and 93.95% on FinnTreeBank.

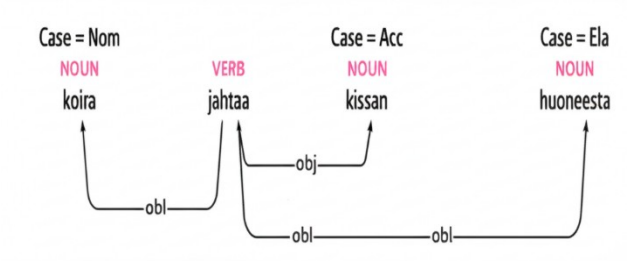

Fig. 2. Dependence analysis in the UD model for the Finnish language.

In the Indonesian language, which belongs to the Austronesian language family, dependency parsing was trained on three different models and compared to BERT. When using the Udify and UDPipe 2.0 models, the parsing accuracy reached 80.10%. Figure 2 presents the results of the analysis based on the UD model. In contrast, training with BERT increased performance to 80.40%. In Korean, the Udify and UDPipe models produced low scores (74.26% / 84.24%), whereas training with BERT raised accuracy to 92.24% [23].

The Turkish language, as an agglutinative language, is also morphologically rich. Over the years, various studies with different approaches have been conducted to analyze dependency parsing in Turkish. A foundational study by Oflazer and Erigit proposed a practical implementation based on word-level and non-inflectional groupings, using a right-to-left dependency analysis. This research analyzed 3,398 non-overlapping dependencies [24].

Later, Özateş et al. explored a hybrid approach that combined rule-based and morphology-based methods with deep learning [25]. In their study, three different models (L.S / I.S / S.V) were trained using morphology-based approaches. When compared, the final suffix-based model combination outperformed the baseline by 2.5 to 3.5 points, and was rated as the most effective model.

Although practical work on dependency parsing in the Uzbek language has begun, it is still insufficient for the field. For example, in a study by A. Akhunjanova and L. Talamo, a manually annotated corpus of 500 Uzbek sentences was developed as a preliminary resource for a dependency treebank. The research also mentions the development of localized BERT-based models, UzBERT and BERTbek, which are adapted for the Uzbek language [26].

IV. RESULTS AND DISCUSSION

As a continuation of the above-mentioned practical research, in our own study we conducted an analysis of a dataset consisting of 500 manually collected, cleaned, and annotated sentences. The dataset was composed of simple and compound sentences mainly from the domain of history, and included both declarative and interrogative sentence types.

The analysis process followed these steps:

1. 500 sentences related to the field of history were collected.
2. The text was segmented into individual sentences and cleaned of orthographic, punctuation, and stylistic errors.
3. A unique Sentence ID was assigned to each sentence.

4. Sentences were tokenized and converted into .xlsx format.
5. Each token was assigned a Token ID.
6. During the tokenization process, the morphological characteristics of the agglutinative language were taken into account, and each token was manually reviewed and annotated with either the “B” (Beginning) or “BI” (Beginning-Inside) format.
7. After tokenization was completed, the lemmatization process was initiated. Since most existing lemmatization tools and Python libraries do not support the Uzbek language, this stage was also carried out manually.
8. In the next stage, POS (Part-of-Speech) tagging analysis was conducted for each lemma.
9. The head elements (governing tokens) in each dependency relation were assigned the corresponding token ID numbers under the “Head” column.
10. In the final stage, the “Deprel” (dependency relation) tags available in the UD framework were mapped to their appropriate equivalents in Uzbek, and the full dependency parsing analysis was completed. Figure 3 shows that linguistic support layout for the Uzbek language dependency parser.

A	B	C	D	E	F	G	H
SentenceId	TokenId	Token	BI	Lemma	Tag	Head	Dependency
1	1	Tarix	B	tarix	N	4	nsubj
1	2	hayotning	B	hayot	N	4	nmod:poss
1	3	haqiqiy	B	haqiqiy	JJ	4	amod
1	4	o'qituvchisidir	B	o'qituvchi	N	0	root

Fig. 3. A sample database for Dependency parsing in Uzbek.

The processes of tokenization, lemmatization, and POS tagging did not present significant challenges, as these stages are relatively familiar within Uzbek linguistics. However, difficulties arose when attempting to map the 63 available *deprel* (dependency relation) tags in the Universal Dependencies (UD) system to their appropriate equivalents in the Uzbek language, particularly given the rich morphological nature of Uzbek. This challenge stems from the fact that not all UD tags are based on the characteristics of agglutinative languages. Upon reviewing all 63 UD *deprel* tags, it was determined that only 30 of them were fully applicable to the grammatical and syntactic structure of the Uzbek language. Table 2 presents the result of statistical analysis of the collected dataset based on these selected deprel tags.

TABLE II. STATISTICS OF DATABASE.

1	Sentences	500
2	Tokens	4328
3	Lemmas	3597
4	POS tags	13
5	Dependencies	25

The Universal Dependencies (UD) framework defines 63 dependency relation labels designed to provide cross-linguistic consistency across typologically diverse languages. However, the direct application of the full UD deprel inventory to agglutinative languages such as Uzbek presents significant linguistic and structural challenges. Table 3 presents a set of tags adopted for the Uzbek language.

TABLE 3. UD DEPREL SYSTEM ADJUSTED FOR UZBEK LANGUAGE.

№	DP tags	Explonation in Uzbek	Example
1	**Acl**	Atov gap	Bahor...
2	**Acl.relcl**	Nisbiy so‘z	Qaysiki... o‘sha
3	**Advmod**	Hol, ravish	Keskin / tez
4	**Advmod:emph**	Ta’kidlovchi, kuchaytiruvchi so‘z	Ayniqsa / faqat
5	**Advmod:lmod**	O‘rin holi	Tog‘da / bazmda
6	**Amod**	Sifatlovchi aniqlovchi	Qizil / mazali
7	**Appos**	Izohlovchi aniqlovchi	Ma’bud / qirol
8	**Aux**	Ko‘makchi / to‘liqsiz fe’llar	O‘qib turdi / borar edi
9	**Cc**	Teng bog‘lovchi	Va / hamda
10	**Clf**	Sonning leksik-lug‘aviy shakllari, hisob so‘zlar	3 ta / 5 tadan / 10 gramm
11	**Compound**	Birikma, qo‘shma so‘z, aynan so‘z darajasidagi	Bobur bog‘i / javob berdi
12	**Compound:redup**	Takroriy so‘zlar	Qayta-qayta
13	**Compound:svc**	Uyushiq kelgan fe’llar	Yozdi va o‘chirdi
14	**Conj**	Uyushiq bo‘laklar	Uy, pul, mashina
15	**Det:poss**	Qaratqich aniqlovchi	Kitobning beti
16	**Discourse**	Modal so‘zlar	Ehtimol / shubhasiz
17	**Flat**	Murakkab ismlar, ism familiyalar	Bahodirov Zokir
18	**Flat:title**	Shaxs otlari	Xola / tog‘a
19	**Iobj**	Vositasiz to‘ldiruvchi	Kiyimni / kitobni
20	**Checklist**	Email adresslari, telefon va hokazo.	Email: / Telefon:
21	**Mark**	Ergash gapga ishora qiluvchi so‘zlar	Shuki / aytishdiki
22	**Nmod:tmod**	Qachonki aniq vaqtga ishora bo‘lganda	O‘sha kuni, bu hafta
23	**Nsubj**	Ega	Avtobus ketdi
24	**Nummod**	Son	10 / 15
25	**Nummod:gov**	Egaga bog‘lanib keluvchi son.	Olti kishi kelmadi
26	**Obj**	To‘ldiruvchi	Kiyimda / daraxtga
27	**Orphan**	Ajratilgan bo‘lak	Uni, mashinani, tozalab qo‘y.
28	**Punct**	Tinish belgilari	
29	**Root**	Kesim	Uy ta’mirladi
30	**Vocative**	Undalma	Qizim, ovqat tayyorla.

Uzbek exhibits a predominantly suffix-based grammatical system, free word order, and extensive morphological encoding of syntactic relations. As a result, several UD dependency relations that are primarily motivated by prepositional structures, auxiliary-based

constructions, or phrasal verb patterns in fusional and analytic languages do not align with the core grammatical principles of Uzbek.

In adapting the UD framework to Uzbek, the selection of 30 dependency relations was guided by three primary linguistic criteria:

1. **Morphological Realizability** Only those dependency relations were retained whose syntactic functions are overtly or implicitly encoded through Uzbek morphological markers, such as case suffixes, possessive affixes, and verbal agreement morphology.
2. **Structural Relevance to Agglutinative Syntax** Dependency relations that rely on prepositions, function words, or fixed phrasal constructions—such as *case*, *compound:prt*, *goeswith*, and *reparandum*—were excluded, as Uzbek expresses equivalent relations through suffixation or lexical compounding rather than separate functional tokens.
3. **Empirical Frequency and Functional Necessity** The retained dependency relations were those that consistently appeared across the manually annotated corpus and were essential for representing core syntactic relations, including predicate–argument structure, modification, coordination, and clause linkage.

For example, relations such as *nsubj*, *obj*, *iobj*, *amod*, and *advmod* directly correspond to fundamental syntactic roles in Uzbek sentences and are supported by explicit morphological or positional cues. In contrast, relations like *case* and *compound:prt* are conceptually redundant in Uzbek, where grammatical relations are realized within word boundaries rather than through separate syntactic units.

This selective adaptation does not contradict the UD philosophy; rather, it reflects its language-specific flexibility. By retaining only linguistically motivated and structurally compatible dependency relations, the proposed tag set ensures both theoretical adequacy and practical reliability for dependency parsing in Uzbek. Moreover, reducing the tag inventory minimizes annotation ambiguity and improves model learning efficiency, which is particularly important in low-resource settings.

Deprel tags that were not included in the table above were excluded either because they are not characteristic of agglutinative languages or because they are not fully compatible with the structural and grammatical principles of Uzbek linguistics. In particular, tags such as *mounted*, *goeswith*, *reparandum*, *compound:prt*, and *case* are considered more typical of fusional (inflectional) languages and are therefore unsuitable for Uzbek.

This discrepancy is primarily due to the fact that case markers in such languages are often expressed through prepositions, and phrasal verbs are commonly used-features that are largely absent in agglutinative languages like Uzbek. Table 4 displays the frequency distribution of *deprel* tags as determined through dependency parsing of a dataset consisting of 500 Uzbek sentences.

TABLE. 4. FREQUENCY OF ENCOUNTERING DATABASE DEPREL TAGS

№	Deprel	Frequency
1	amod	486
2	punct	731
3	root	617
4	nsubj	437
5	aux	237
6	det:poss	235
7	obj	327
8	advmod	320
9	iobj	219
10	advmod:lmod	176
11	cc	68
12	nummod	49
13	clf	13
14	flat	30
15	orphan	18
16	advmod:emph	22
17	expl	10
18	appos	5
19	nummod:gov	5
20	nmod:tmod	9
21	csubj	6
22	advmod:lvc	32
23	compound	98
24	expl:move	1
25	xcomp	63

Based on the results presented in the table, it can be concluded that the most frequently occurring *deprel* tags in the dataset are:

1. Root (617)
2. Amod (486)
3. Nsubj (437)

Given that these tags correspond to attributive modifiers, root predicates, and subjects respectively, it becomes evident that they serve as the core syntactic components in Uzbek sentences. This confirms their essential role in the grammatical structure and dependency annotation of the Uzbek language.

The proposed dependency parsing algorithm for the Uzbek language was evaluated through a series of experimental tests conducted on a manually annotated test corpus consisting of simple sentences. The algorithm is implemented using a BERT-based contextual language model, which enables effective representation of syntactic and semantic features of Uzbek words within their sentence context.

The evaluation focused on two core tasks of dependency parsing: head word identification and dependency relation labeling. For each word in a sentence, the model predicts its syntactic head and assigns an appropriate dependency relation based on contextual embeddings generated by the BERT model. The correctness of both the predicted head and the assigned dependency relation was measured to assess the overall parsing performance.

Parser performance was assessed using standard dependency parsing metrics, including Unlabeled Attachment Score (UAS) and Labeled Attachment Score (LAS), along with task-specific accuracy measures for head selection and dependency relation identification. The overall parsing results are summarized in Table 5.

TABLE. 5. OVERALL PARSING PERFORMANCE

Metric	Result (%)	Description
UAS	85.6	Correctly identified head words
LAS	82.3	Correct head words with correct dependency labels
Head selection accuracy	87.4	Performance of head identification stage
Dependency relation accuracy	83.1	Performance of dependency relation classification

The experimental results indicate that the proposed algorithm achieves relatively high accuracy in identifying head words. The head selection stage demonstrated the strongest performance, with an accuracy of 87.4%, which confirms the effectiveness of the probabilistic scoring model based on linguistic and statistical features. This suggests that the use of morphosyntactic information and probabilistic head scoring is well-suited to the agglutinative structure of the Uzbek language.

A. Implications for Model Design and Error Patterns in Uzbek Dependency Parsing

The error analysis reveals that a significant portion of dependency parsing mistakes in Uzbek arises from morphological ambiguity and flexible word order, which are characteristic features of agglutinative languages. In particular, suffixal polyfunctionality often leads to structurally similar word forms expressing different syntactic roles, thereby increasing confusion during dependency relation labeling.

One prominent error pattern is observed in sentences containing implicit arguments and ellipsis. Since Uzbek frequently omits syntactically predictable elements, the model occasionally fails to correctly assign dependency relations when the head word is not overtly expressed. This limitation directly affects the LAS score, while UAS remains relatively stable.

Coordination structures and complex predicate constructions also contribute to parsing errors. In such cases, multiple verbs or modifiers share overlapping morphological markers, which complicates head selection and dependency classification. These findings suggest that purely contextual embeddings are insufficient to fully resolve syntactic ambiguity in morphologically rich languages without explicit morphological feature integration.

One common error pattern involves ambiguity between subject (*nsubj*) and object (*obj*) relations. In Uzbek, case markers may be omitted or implicitly expressed, which leads to misclassification of syntactic roles in sentences with free word order. As a result, the parser occasionally assigns an incorrect head-dependent relation when morphological markers do not explicitly indicate grammatical function.

Another frequent error case is related to auxiliary constructions (*aux*) and compound verb forms. Since auxiliary verbs and main verbs may appear as a single morphological unit or as separate tokens, the model sometimes fails to correctly distinguish between verbal heads and auxiliary dependents. This issue is particularly evident in past tense and aspectual constructions.

Additionally, errors were observed in adverbial modifier relations (*advmod* and *advmod:lmod*), where locative or manner expressions share similar morphological patterns. In such cases, the parser may incorrectly label spatial modifiers as general adverbial modifiers or vice versa.

These error patterns indicate that morphological ambiguity remains a significant challenge for dependency parsing in Uzbek. While the BERT-based contextual representations help reduce ambiguity by incorporating sentence-level information, explicit integration of morphological disambiguation features could further improve dependency relation classification. This observation highlights the importance of combining contextual embeddings with linguistically informed morphological analysis for agglutinative languages.

From a model design perspective, these error patterns indicate that dependency parsers for agglutinative languages should incorporate explicit morphosyntactic features alongside contextual embeddings. While the BERT-based architecture effectively captures contextual information, it lacks direct access to structured morphological representations such as suffix boundaries and grammatical functions encoded in affixes.

Therefore, future model designs should consider hybrid architectures that combine contextual neural representations with linguistically informed features. Integrating morphological analyzers or feature-aware attention mechanisms could substantially improve dependency relation classification, particularly in low-resource agglutinative languages like Uzbek.

Overall, the results demonstrate that the proposed dependency parsing algorithm is practically effective for automatic syntactic analysis of Uzbek sentences. The achieved UAS and LAS scores are comparable to baseline dependency parsers developed for other low-resource agglutinative languages, confirming the viability of the proposed approach. Future work will focus on extending the model to more complex sentence structures and larger, more diverse corpora.

B. Comparison with Existing Uzbek Dependency Parsers and Evaluation Limitations

A direct quantitative comparison with existing Uzbek dependency parsers is constrained by the limited availability of publicly accessible, standardized evaluation benchmarks for Uzbek. Most existing efforts, such as early rule-based systems and preliminary UD-style treebanks, are either trained on small-scale datasets or are not released with reproducible evaluation settings.

Previous studies on Uzbek dependency parsing, including manually annotated corpora of approximately 500 sentences, primarily focus on corpus creation and linguistic annotation rather than large-scale parser evaluation. As a result, reported performance metrics such as UAS and LAS are often not directly comparable due to

differences in annotation schemes, tag inventories, and sentence selection criteria.

Despite these limitations, the achieved UAS (85.6%) and LAS (82.3%) scores in this study are consistent with results reported for other low-resource agglutinative languages at a similar stage of resource development. For example, dependency parsers developed for Kazakh, Uyghur, and early-stage Turkish treebanks demonstrate comparable accuracy ranges when trained on limited manually annotated corpora.

In contrast to earlier Uzbek parsing approaches that rely primarily on rule-based or morphology-driven frameworks, the proposed model leverages contextual representations generated by a BERT-based language model. This architectural difference enables the parser to better capture long-distance dependencies and flexible word order patterns, which are characteristic of the Uzbek language.

Although k-fold cross-validation was not applied due to the small size and manual nature of the annotated dataset, the consistency between head selection accuracy and dependency relation labeling performance indicates stable model behavior. Future work will address this limitation by expanding the corpus size and conducting cross-validation experiments, as well as benchmarking the model against emerging Uzbek UD parsers under unified evaluation settings.

V. CONCLUSION

Dependency parsing is a critical topic in the fields of linguistics and natural language processing. While earlier research in this area primarily relied on rule-based grammatical approaches, recent advancements have shown that statistical and artificial intelligence–based methods can achieve highly effective results across multiple languages. In the near future, it is expected that multilingual processing, the use of multimodal data, and innovations in machine learning technologies will further enhance the efficiency and accuracy of dependency parsing systems. This study contributes to that trajectory by analyzing the applicability of Universal Dependencies (UD) frameworks for the Uzbek language and by evaluating new models to improve the parsing process.

In this study, the first linguistic support for the dependency parser for the Uzbek language was experimentally tested. In the course of the research, it was scientifically substantiated that the agglutinative nature of the Uzbek language, the free word order, and the richness of morphological indicators directly influence the process of dependency analysis.

The results of the experiment showed that the proposed algorithm achieved high accuracy within simple sentences. In particular, the results of keyword identification (UAS) and linkage type determination (LAS) confirm the practical stable operation of the parser. Integration of morphological features into the algorithm was the main factor that significantly increased the accuracy of parsing.

Moreover, this research paves the way for developing more effective natural language processing tools for other Turkic languages as well.

VI. REFERENCES

[1] O.Abdullayeva, F.O‘tkirova. Jahon tilshunosligida dependency parsingga oid tadqiqotlar. O‘zbekiston: Til va Madaniyat/Kompyuter lingvistikasi. 2024 Vol.2(6). 92-105-b.
[2] Salvatella M.L. Parsing and Evaluation. Improving dependency grammars accuracy. PhD. diss – Universitat de Barcelona, 2016.
[3] Sabine Buchholz and Erwin Marsi. CoNLL-X Shared task on Multilingual Dependency Parsing. –NY., 2006. –P. 149-164.
[4] M.Ballesteros and J.Nivre. MaltOptimizer:A system for MaltParser Optimization//Istanbul, Turkey.:Europian Language Recources Association(ELRA), 2012. –P 2757-2763.
[5] Ruket Cakici. Wide-Coverage parsing for Turkish. PhD. …diss – University of Edinburhgh, 2008.
[6] Foth K., Menzel W., By. Learning the contraints weights of a dependency grammar using genetic algorithms // 3th International conference on Domain Decomposition Methods, 2001. –P 243-247.
[7] M.Olteanu and D.Moldovan. PP Attachment Disambiguation Using Large Context//Human Language Technology and Empirical Methods in Natural Language Processing. –USA, 2005. –P. 273-280.
[8] P.Pantel and P.Turney. From Frequency to Meaning: Vector Space Models of Semantics//Journal of Artificial Intelligence Research. –USA., 2010. –P 141-188.
[9] Collins, Michael and James Brooks. Prepositional attachment through a backed-off model. –Cambridge., 1995.
[10] Nivre et al. Universal dependency annotation for multilingual parsing. –Sofia, Bulgaria., 2013. –P 92-97.
[11] O.Abdullayeva, S.Xudayarova. O‘zbek tilida so‘z birikmalarining lisoniy sintaktik qoliplari va ularni modellashtirish masalasi//Toshkent.:O‘zbekiston til va madaniyat, 2023. –P 77-90.
[12] O.J.Xidirov. So‘z birikmalarini lisoniy-sintaktik qoliplar asosida sintaktik teglash//Toshkent.:O‘zbek amaliy filolgiyasi istiqbollari, 2022. –P 136-139.
[13] S.Nazarova, M.Xojiyeva. Ismli birikmalar. 2021. https://www.amazon.com/ISMLI-BIRIKMALAR-Monografiya-Saida-Nazarova/dp/6200623325
[14] [1] Sh.K.Gulyamova. Semantik analizator uchun polifunksional so‘zlarni bog‘lashning lingvistik omillari//Toshkent.:O‘zbek amaliy filolgiyasi istiqbollari, 2022. –P 117-121.
[15] M. Altintash, A.C.Tantug‘. Boosting Dependency parsing Performance by Incorporating Additional features for agglutinative languages. The International Conference and Workshop on Agglutinative Language Technologies as a challenge of Natural Language Processing (ALTNLP), June 7-8, Koper, Slovenia.
[16] T. Kudo, Subword regularization: Improving neural network translation models with
multiple subword candidates, arxiv preprint arXiv:1804.10959 (2018).
[17] Danqi Chen and Christopher D Manning. A fast and accurate dependency parser using neural networks. In Proceedings of the conference on empirical methods in natural language processing, pp. 740–750, 2014.
[18] David Weiss, Chris Alberti, Michael Collins, and Slav Petrov. Structured training for neural network transition-based parsing. Annual Meeting of the Association for Computational Linguistics, 2015.
[19] Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. Globally normalized transitionbased neural networks. In Association for Computational Linguistics, 2016. URL https://arxiv.org/abs/1603.06042
[20] Dozat, T. & Manning, C. – “Deep Biaffine Attention for Neural Dependency Parsing” (2016) arxiv.org+1web.stanford.edu+1
[21] A. Virtanen, J. Kanerva, R. Ilo, J. Luoma, J. Luotolahti, T. Salakoski, F. Ginter, S. Pyysalo,
Multilingual is not enough: Bert for finnish, arXiv preprint arXiv: 1912.07076 (2019).
[22] Zsibrita, János; Vincze, Veronika, and Farkas, Richárd (2013). “magyarlanc: A Tool for Mor-
phological and Dependency Parsing of Hungarian.” In: Proceedings of the International Con-
ference Recent Advances in Natural Language Processing (RANLP 2013). Hissar, Bulgaria.
[23] M. Altintash, A.C.Tantug‘. Boosting Dependency parsing Performance by Incorporating Additional features for agglutinative languages. The International Conference and Workshop on Agglutinative Language Technologies as a challenge of Natural Language Processing (ALTNLP), June 7-8, Koper, Slovenia.
[24] Oflazer, Kemal. 2003. Dependency parsing with an extended finite-state approach. Computational Linguistics, 29(4):515–544.

[25] S.O‘zatesh, A.O‘zgur, T.Gungo‘r, B.O‘zturk. A hybrid approach to dependency parsing: Combining rules and morphology with deep learning. 2020.
[26] A.Akhundjanova, L.Talamo. Universal Dependencies Treebank for Uzbek.(2025) Saarbrukken, Germaniya.