INTEGRATION OF NEUROLINGUISTIC AND COMPUTATIONAL LINGUISTIC APPROACHES IN NATURAL LANGUAGE PROCESSING FOR UZBEK

Dushamova Nodirabegim

– Independent Researcher

Abstract. This article examines the integration of neurolinguistic and computational linguistic approaches in natural language processing for Uzbek. Human speech perception, lexical comprehension, contextual analysis, and the identification of semantic relationships are comparatively analyzed in relation to modern natural language processing technologies. The agglutinative characteristics of Uzbek, its morphological complexity, polysemy, and context-dependent changes in meaning are considered important factors in the development of artificial intelligence models. The article also discusses the theoretical and practical possibilities of combining neurolinguistic knowledge with computational linguistic methods.

Keywords: neurolinguistics, computational linguistics, natural language processing, Uzbek language, speech comprehension, semantic analysis, artificial intelligence, context, morphology.

The rapid development of digital technologies and artificial intelligence has expanded the possibilities of automatically processing human language, understanding textual meaning, and communicating effectively through natural language. Natural Language Processing (NLP) is currently one of the principal technological fields underpinning machine translation, voice assistants, automatic text analysis, question-answering systems, and dialogue models. Language comprehension involves more than recognizing words or applying grammatical rules. When understanding speech, humans integrate phonological, morphological, syntactic, and semantic information. Furthermore, the conversational situation, previous linguistic experience, general knowledge, and communicative purpose influence the interpretation of meaning. While neurolinguistics investigates the aspects of

these processes associated with brain activity, computational linguistics develops methods for formally modeling linguistic units and analyzing them automatically.

The integration of these two fields is particularly important for natural language processing in Uzbek. Uzbek is an agglutinative language that expresses grammatical meanings through suffixes and is characterized by diverse word forms, relatively flexible word order, homonymy, and context-dependent semantic relationships. Consequently, directly transferring methods developed for resource-rich languages such as English to Uzbek does not always produce satisfactory results.

Neurolinguistics is an interdisciplinary field that studies the processes associated with language perception, comprehension, and production in the human brain. Its research encompasses speech-sound recognition, lexical activation, grammatical structure processing, meaning construction, and discourse comprehension. In understanding speech, humans coordinate information from different linguistic levels. For example, in the sentence *"Talaba kitobni o'qituvchiga berdi"* ("The student gave the book to the teacher"), the grammatical forms of the sentence constituents help determine who performed the action, what object was transferred, and to whom the action was directed. In this process, the accusative case marker in *"kitobni"* ("the book") and the dative case marker in *"o'qituvchiga"* ("to the teacher") serve as important grammatical indicators.

Neurolinguistic research examines language processing not as a single, fixed sequence of stages but as a system of interacting processes. Information about word meaning, syntactic structure, and context may influence one another in parallel under certain circumstances. In his work *Neurolinguistics*, Giosuè Baggio analyzes major theoretical approaches to language acquisition, speech processing, and brain activity associated with language. These approaches provide a theoretical basis for comparing human language comprehension with the operating principles of artificial intelligence models. However, the human brain and artificial neural networks are not identical systems. Although certain similarities exist in their computational principles, the biological, social,

and experience-based dimensions of human language comprehension cannot be equated entirely with the statistical outputs of artificial intelligence.

Computational linguistics develops methods for analyzing, modeling, and automatically processing linguistic units using computers. Its principal tasks include tokenization, morphological analysis, part-of-speech tagging, syntactic parsing, named entity recognition, semantic analysis, and the identification of semantic relationships between texts.

Morphological analysis plays an important role in computational processing of Uzbek. For example, the word *"kitoblarimizdan"* ("from our books") can be segmented into *"kitob + lar + imiz + dan."* Here, the stem, plurality, possession, and ablative case are expressed within a single word form. If a model can identify these components, its ability to interpret the grammatical function of the word and its relationships within a sentence can improve.

The UzbekPOS dataset introduced by Sharipov, Kuriyozov, and Vičič is an important practical resource for part-of-speech tagging in Uzbek. It contains annotated sentences drawn from texts across different domains and can be used to develop morphological analysis, syntactic parsing, and natural language processing models.

Research on named entity recognition in Uzbek has also employed specially annotated corpora and neural networks. In particular, datasets developed to identify people, organizations, and other named entities in Uzbek texts expand the possibilities of automatic linguistic analysis. Nevertheless, identifying a word's part of speech or detecting a named entity is not equivalent to fully understanding a text's meaning. Semantic comprehension also requires consideration of relationships between words, the communicative purpose of a sentence, preceding context, and implicit meanings.

The integration of neurolinguistic and computational linguistic approaches involves applying knowledge about human language comprehension to the design and evaluation of natural language processing models. This integration does not aim to reproduce

biological processes in their entirety; rather, it seeks to incorporate cognitive and linguistic factors that are important for understanding language.

First, context-based analysis is essential. People often determine the meaning of a word according to its linguistic environment. For example, in the sentences *"Uning boshi og'ridi"* ("His/her head hurt") and *"Tog'ning boshi qor bilan qoplangan"* ("The summit of the mountain is covered with snow"), the word *"bosh"* has different meanings. Contextual language models attempt to distinguish these meanings by analyzing surrounding linguistic units and the overall meaning of the sentence.

Second, incorporating morphological knowledge into models is particularly important for Uzbek. Analyzing grammatical meanings expressed through suffixes helps identify syntactic relationships between words. This can be achieved through morphological tagging, lemmatization, grammatical feature encoding, and the application of linguistic rules.

Third, integrating semantic and pragmatic information is necessary. For example, the sentence *"Derazani ochib yuboring"* ("Please open the window") is grammatically expressed as an imperative, but in actual communication it may function as a polite request. To interpret this correctly, a model must consider not only the lexical meanings of words but also the communicative situation.

Fourth, human-in-the-loop evaluation is an important methodological tool. Model outputs should be evaluated not only through automatic metrics but also by native speakers and linguists. This approach helps identify responses that are grammatically correct but semantically inaccurate or incorrectly interpreted in context.

The integration of neurolinguistic knowledge and computational linguistic methods can help address several practical challenges in Uzbek. The first area is automatic speech recognition and speech-to-text conversion. In these systems, it is important to consider pronunciation variants, stress, pauses, speech rate, and contextual information together.

Particular attention must be paid to reductions in spontaneous speech, dialectal expressions, and loanwords.

To conduct research in this area, theoretical analysis, corpus analysis, comparative analysis, and experimental evaluation can be employed. In the initial stage, a collection of semantically complex Uzbek sentences can be compiled. This dataset may include polysemous words, grammatical homonymy, figurative meanings, irony, speech acts involving requests, and context-dependent expressions.

In the next stage, these examples can be enriched with morphological and syntactic annotations. Subsequently, the performance of three types of models can be compared: a model based on simple word sequences, a contextual language model, and a model enhanced with morphological features. Evaluation can employ such criteria as word-sense disambiguation accuracy, contextual interpretation accuracy, identification of grammatical relationships, and agreement with human evaluators.

For a more in-depth application of the neurolinguistic approach, researchers may investigate human language comprehension behavior through experiments involving sentence-reading times or semantic choices. If methods for measuring brain activity, including electroencephalography (EEG) or functional magnetic resonance imaging (fMRI), are employed, they must be conducted in accordance with appropriate ethical requirements and scientific methodologies.

An important limitation is that a direct and complete correspondence between human brain activity and the internal computational processes of artificial intelligence models cannot be assumed. Comparing these processes requires clearly defined research questions, controlled experiments, and cautious interpretation.

Key future priorities for Uzbek include expanding spoken-language corpora, creating semantically annotated datasets, improving morphological analysis tools, accounting for dialectal and stylistic variations, and evaluating models in real communicative situations.

The integration of neurolinguistic and computational linguistic approaches in natural language processing for Uzbek enables the interconnected study of human language comprehension and the operating principles of artificial intelligence models. Neurolinguistics explains the cognitive and biological foundations of speech comprehension, while computational linguistics helps apply this knowledge to automatic analysis and modeling tasks.

The agglutinative structure of Uzbek, its morphological richness, polysemy, and context-dependent semantic characteristics require specialized linguistic resources and adapted computational methods. Therefore, enhancing contextual models with morphological and syntactic knowledge, evaluating their outputs with the participation of native speakers, and developing corpora that cover diverse communicative situations are among the most pressing priorities.

In conclusion, collaboration between these two fields provides a promising scientific foundation for improving the quality of automatic speech and language processing in Uzbek. However, the success of such systems depends not only on model architecture but also on the quality of linguistic data, the evaluation methodology, and the accurate consideration of the complexity of human language comprehension.

References

1. Hojiyev, A. *Tilshunoslik terminlarining izohli lug‘ati* [Explanatory Dictionary of Linguistic Terms]. Tashkent: O‘zbekiston Milliy Ensiklopediyasi, 2002.
2. Nurmonov, A. *O‘zbek tilshunosligi tarixi* [History of Uzbek Linguistics]. Tashkent: O‘zbekiston, 2002.
3. Mengliyev, B. R. *Hozirgi o‘zbek adabiy tili* [Contemporary Standard Uzbek]. Tashkent. Publication details not specified in the original reference.
4. Ne’matov, H., and R. Rasulov. *O‘zbek tili sistem leksikologiyasi asoslari* [Fundamentals of Systemic Lexicology of the Uzbek Language]. Tashkent: O‘qituvchi, 1995.

5. Mahmudov, N. *Til va nutq* [Language and Speech]. Tashkent. Publication details not specified in the original reference.