Journal of Language Pedagogy and Innovative Applied Linguistics
December 2025, Volume 3, No. 2, pp: 78-80
ISSN: 2995-6854

The Importance of Linguistic-Syntactic Pattern Annotation and Corpus Creation in the Uzbek Language

Pokiza Nurmamatova *
Samarkand State Institute of Foreign Languages

Abstract
This article highlights the system of existing linguistic-syntactic patterns (LSP) in the Uzbek language, their role in sentence construction, and the linguistic and technological significance of the markup (annotation) process. If a sentence, as the main unit of communication in language, is a syntactic structure completed in terms of content, purpose, and intonation, then LSQ is an abstract pattern of phrases that form a sentence. The article shows ways to deeply study the morphological, syntactic, and semantic features of LSQs by dividing them into components, marking them with tags, and creating an individual corpus.
Key Words: *linguistic-syntactic pattern (LSQ), marking, corpus, tagging, morphological analysis, semantic proportionality, valency, metaphor, parser, thesaurus, automatic analysis, Uzbek language.*

Paper/Article Info

Reference to this paper should be made as follows:

Nurmamatova, P. (2025). The Importance of Linguistic-Syntactic Pattern Annotation and Corpus Creation in the Uzbek Language. Journal of Language Pedagogy and Innovative Applied Linguistics, 3(2), 78-80.
https://doi.org/10.1997/j53n3565

** Corresponding Author*
DOI: https://doi.org/10.1997/j53n3565

Arizona, USA

In Uzbek syntax, word combinations function as the fundamental structural units from which sentences are constructed. These combinations serve as the "building materials" of sentence formation and represent stable linguistic–syntactic patterns entrenched in the speaker's cognition. Although LSQs constitute a fixed system, they provide the basis for producing an unlimited number of speech realizations. Understanding the internal structure of LSQs and annotating them within linguistic corpora is crucial for the systematic study of Uzbek syntax.

LSQs operate as abstract syntactic molds comparable to construction blocks. Research in Uzbek linguistics confirms that speakers rely on cognitive templates for converting lexemes into meaningful combinations and sentences [1, p. 75]. Despite their structural stability, LSQs generate limitless linguistic outputs, enabling flexible yet rule-governed speech production.

Annotating LSQs allows researchers to decompose word combinations into their constituent parts, identify their syntactic functions, and determine the relationships between roots and affixes. The Uzbek language contains 18 invariant LSQs, which considerably simplify the annotation process and make it scalable for corpus linguistics. Tagging therefore offers a standardized mechanism for the automatic analysis of LSQs, facilitating both linguistic description and computational applications.

Corpus-based annotation provides insights beyond mere frequency counts. When parts of speech are assigned to each lexical item, researchers can determine usage patterns across discourse types. Furthermore, linguistic tagging assigns a unique code to each word form [2 p.28], enabling the systematic study of morphological and syntactic behavior. Morphological tagging is particularly important in identifying grammatical features such as case, possession, and derivation, which in turn supports the automatic analysis of LSQs, e.g., do'stimning uyi ('my friend's house'). Semantic compatibility plays a central role in the evaluation of LSQs. Lexical valency requires that a governing lexeme select semantically appropriate dependents. Although many LSQs may be formally generated—such as structures involving a noun in the accusative case plus a verb—only a subset represents meaningful combinations (e.g., qo'ylarni ekdi 'planted the sheep' is formally well-formed but semantically illogical). Thus, LSQ analysis must incorporate both formal and semantic criteria.

Meaning extension, particularly metaphorical shift, must also be considered. For instance, in oltin kuz ('golden autumn'), the lexeme oltin undergoes metaphorization and functions as an attributive modifier rather than a noun. Such cases require precise tagging to reflect both syntactic and semantic transformations.

Syntactic parsers and thesaurus-based lexical resources play an essential role in LSQ annotation. Thesauri capture the full range of lexical behavior in a language and support the identification of lexical valency patterns. Their integration into annotation enables the extraction of LSQs from large text collections, decomposition into structural components, and detailed analysis of their syntactic and semantic properties.

The identification and annotation of Uzbekistan's 18 fixed LSQs create opportunities to:

- systematically analyze morphological, syntactic, and semantic patterns;
- uncover the relationship between word combinations and sentence constituents;
- construct foundational datasets for machine translation and natural language processing;
- develop more accurate language models grounded in linguistic structure.

These outcomes highlight the importance of LSQ-based annotation for both theoretical linguistics and applied computational research.

LSQs represent abstract syntactic patterns that guide speech production and shape the structure of Uzbek word combinations. Their systematic identification and annotation using corpus-based methods allow for deeper insights into the morphological, syntactic, and semantic properties of the language. The integration of parsers and thesauri further enhances the precision of LSQ analysis. These developments contribute significantly to linguistic theory and provide essential resources for the advancement of Uzbek language technologies

References

[1]. Sayfullayeva R., Mengliyev B., Boqiyeva G., Qurbonova M., Yunusova Z., Abduzalova M. Modern Uzbek Literary Language. Tashkent: Fan va texnologiya, 2009.- p.286

[2]. Zakharov V., Mengliyev B., Khamroyeva Sh. Corpus Linguistics: Building and Using Corpora. Tashkent, 2021.
[3]. Hozirgi_ozbek_adabiy_tili //https://namdu.uz/media/Books/pdf/2024/06/20/NamDU-ARM-6806- /Дата обращения 09.12.2025г.
[4]. Страничка Википедия// https://ru.wikipedia.org/wiki /Дата обращения 09.12.2025г.