

MAPPING THE UZBEK GRAMMATICAL TRADITION ONTO UNIVERSAL DEPENDENCIES: CONVERGENCES AND MISMATCHES

Khamdamova Dilnoza

Karshi State University, doctoral student

DOI: <https://doi.org/10.5281/zenodo.21247993>

Abstract: The paper examines how the categories of the Uzbek national grammatical tradition the system of principal and secondary sentence members correspond to the universal syntactic relations of the Universal Dependencies (UD) framework. It is shown that the mapping is systematically asymmetric: some traditional categories converge directly with UD relations, others split into several relations, and certain UD notions have no counterpart in the national tradition. The findings provide a linguistic foundation for the syntactic annotation of Uzbek simple sentences.

Keywords: Uzbek, Universal Dependencies, syntactic annotation, sentence members, grammatical tradition, dependency relations, simple sentence.

Introduction

The development of dependency treebanks for low-resource languages inevitably raises the question of how universal annotation schemes relate to national grammatical traditions. For Uzbek, this question is particularly acute: the language possesses a well-established syntactic tradition describing the simple sentence through the system of sentence members (*gap bo'laklari*), while emerging resources such as UD_Uzbek-UT [1] and UzTreebank [7] adopt the cross-linguistic relation inventory of Universal Dependencies [3, 5]. The present paper systematically compares the two category systems and identifies three types of correspondence: direct convergence, category splitting, and categorical absence. Following W. Croft, cross-linguistic annotation must rest on functional equivalence rather than on the transfer of language-specific categories [2]; the mapping proposed here is intended as such an equivalence-based bridge.

The Two Category Systems

The Uzbek tradition, developed in the works of A. G'ulomov, G'. Abdurahmonov, and their successors [4, 6], analyzes the simple sentence through two principal members (*bosh bo'laklar*) the subject (*ega*) and the predicate (*kesim*) and three secondary members (*ikkinchi darajali bo'laklar*): the object (*to'ldiruvchi*), the attribute (*aniqlovchi*), and the adverbial modifier (*hol*). The system is semantic-functional in orientation: members are identified by the questions they answer and the functions they perform.

Universal Dependencies, by contrast, describes the sentence as a set of binary head-dependent relations drawn from a universal inventory (*nsubj*, *obj*, *obl*, *amod*, *advmod*, etc.), governed by lexicalist principles: content words serve as heads, while function words attach to them via dedicated relations (*case*, *cop*, *aux*) [3, 5].

Convergences, Splits, and Absences

The correspondence between the two systems is summarized in Table 1.

Table 1. Correspondence between Uzbek sentence members and UD relations

Uzbek tradition	UD relation(s)	Type of correspondence
Ega	nsubj	direct convergence
Kesim	root (head), not a relation	categorical shift

Vositasiz to’ldiruvchi	obj	direct convergence
Vositali to’ldiruvchi	iobj / obl	split
Aniqlovchi	amod / nmod / det / nummod	split
Hol	advmod / obl / advcl	split
— (bog’lama)	cop	absence reinterpretation
—	case (postpositions)	absence

Three configurations deserve comment.

First, *direct convergence*: *ega* maps onto *nsubj*, and the direct object marked by *-ni* onto *obj*, e.g. *Talaba maqolani yozdi*, where *talaba* → *nsubj*, *maqolani* → *obj*. Here annotation may rely on the national tradition without modification.

Second, *category splitting*: the traditional *aniqlovchi* covers adjectival modification (*qiziq kitob* → *amod*), possessive constructions (*talabaning kitobi* → *nmod*), demonstratives (*bu kitob* → *det*), and numerals (*ikki kitob* → *nummod*). One traditional category thus corresponds to at least four UD relations, and the annotator must draw distinctions the tradition does not require. The same holds for *vositali to’ldiruvchi*, which UD divides between the core argument relation *iobj* and the oblique relation *obl*, a distinction based on argumenthood that the national tradition, treating all case-marked and postpositional complements uniformly, does not encode.

Third, *categorical absence and shift*: the most fundamental mismatch concerns *kesim*. In the national tradition the predicate is a sentence member coordinate with the subject; in UD the predicate is not a labeled relation but the structural head (root) on which all other members depend. Conversely, UD relations such as *cop* and *case* have no traditional counterpart: the copular element treated in the tradition as part of a composite nominal predicate (*bog’lama*) receives a separate *cop* relation in UD, and postpositions traditionally auxiliary words within the member they accompany are annotated as case dependents.

Implications for Annotation

The asymmetry of the mapping has direct methodological consequences. Where convergence is direct, annotation guidelines for Uzbek can be formulated in terms familiar to nationally trained linguists. Where categories split, guidelines must supply explicit operational criteria (morphological form, argumenthood tests) absent from the tradition. Where categories are absent or shifted, annotation requires conscious reconceptualization most notably, abandoning the coordinate subject–predicate model in favor of head-dependency. Ignoring these asymmetries produces inconsistent annotation; making them explicit, as attempted here, turns the national tradition from an obstacle into a resource for treebank development.

Conclusion

The comparison demonstrates that the Uzbek grammatical tradition and Universal Dependencies are neither isomorphic nor incompatible: their categories stand in a systematic but asymmetric correspondence comprising direct convergences (*ega*–*nsubj*, *vositasiz to’ldiruvchi*–*obj*), one-to-many splits (*aniqlovchi*, *hol*, *vositali to’ldiruvchi*), and categorical shifts (*kesim*–*root*, *bog’lama*–*cop*). An explicit mapping of this kind constitutes a necessary linguistic foundation for consistent syntactic annotation of the Uzbek simple sentence and, more broadly, a model for reconciling national grammatical traditions with universal annotation frameworks.

References



1. Akhundjanova A., Talamo L. Universal Dependencies Treebank for Uzbek // Proceedings of the Second Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL 2025). – Tallinn, 2025.
2. Croft W. Typology and Universals. 2nd ed. – Cambridge: Cambridge University Press, 2003. – 341 p.
3. de Marneffe M.-C., Manning C. D., Nivre J., Zeman D. Universal Dependencies // Computational Linguistics. – 2021. – Vol. 47, No. 2. – P. 255–308.
4. G‘ulomov A., Asqarova M. Hozirgi o‘zbek adabiy tili. Sintaksis. – Toshkent: O‘qituvchi, 1987. – 256 b.
5. Nivre J., de Marneffe M.-C., Ginter F., Hajič J., Manning C. D., Pyysalo S., Schuster S., Tyers F., Zeman D. Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection // Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020). – Marseille: ELRA, 2020. – P. 4034–4043.
6. Sayfullayeva R. R., Mengliyev B. R., Boqiyeva G. H. va boshq. Hozirgi o‘zbek adabiy tili. – Toshkent: Fan va texnologiya, 2009. – 416 b.
7. Sharipov M. et al. UzTreebank: Dependency Treebank for the Uzbek Language // IEEE Proceedings. – 2025.