

The Classification of Lexical Units Formed on the Basis of Whole–Part Relations (on the Example of Uzbek)

Safarova Mahfuza Rajabovna

Independent researcher of Bukhara State University

Senior English teacher of school No. 17, Bukhara district,

Bukhara region

Email: safarovam@gmail.com

Abstract:

The lexical units in Uzbek that are driven by whole-part relations (partonymic units) are described and classified in this paper. Following an overview of the theoretical underpinnings of mereological relations and lexical semantics, it is suggested categorizing Uzbek lexical items that encode whole-part relations into four main lexical-semantic groups: metaphorical/derivative partonymy, member-collective relations, physical-part meronyms, and substance-whole relations. In addition to discussing diagnostic criteria for membership and providing brief implications for lexicography and lexical database development, the paper uses Uzbek examples to illustrate each group. A brief list of pertinent works of literature is given.

Keywords: partonymy, meronymy, holonymy, lexical semantics, Uzbek, UzWordNet, mereology

Introduction

Lexical semantics places a strong emphasis on studies of lexical relations. Word sense representation in dictionaries and lexical databases is influenced by relationships like synonymy, hyponymy/hypernymy, and meronymy/holonymy. Semantic ties where one lexical unit indicates a constituent (part) and another indicates the whole are specifically described by meronymy (part-of) [1]. These relationships, which cut across semantic domains (body parts, artifacts, collectives, and substances), are productive in many languages, including Uzbek. Although theoretical explanations and computational tools for studying Uzbek lexical relations have advanced, linguists, lexicographers, and NLP practitioners can still benefit from a brief classification account that focuses on whole-part motivated lexical units [2-3].

Theoretical Background

The analysis is informed by two theoretical strands that overlap. First, lexical semantics

provides tools for describing the relationships between lexical items (e.g., polysemy, collocational profiles, meronymy/holonymy, and synonymy/hypernymy) and identifying diagnostic tests for semantics (e.g., distributional behavior, semantic entailment, and substitutability in contexts). Second, linguistic categories of meronymy (part-meronym, member-meronym, substance-meronym) correspond to conceptual distinctions (part vs. member vs. substance) that are provided by mereology, the philosophical/logical study of part-whole relations. Three types of meronymy are frequently distinguished in linguistic descriptions: substance meronymy (flour–bread), member meronymy (sheep–flock), and component/part meronymy (wheel–car). The taxonomy suggested below is guided by these distinctions [4].

Methodology

The classification relies on (a) semantic diagnostics (can X exist without Y? is X a necessary structural constituent of Y?), (b) morphosyntactic behavior (case marking, possessive constructions in Uzbek), and (c) distributional evidence (co-occurrence patterns in corpora and lexical network resources such as UzWordNet). For each candidate pair, I apply the following tests:

1. **Part-identity test:** whether the putative meronym denotes a tangible, bounded constituent of the holonym (e.g., *bosh* ‘head’ : *tana* ‘body’).
2. **Membership vs. part test:** whether the relation expresses a member-in-collection (e.g., *qo* ‘y’ ‘sheep’ : *qo* ‘m’ ‘flock’) rather than physical parthood.
3. **Substance test:** whether the relation is substance → whole (material used to make whole, e.g., *temir* ‘iron’ : *karobka* ‘(metal) box’ when interpreted as material).
4. **Figurative/derivative test:** whether the link is metaphorical or derivational (e.g., part of a compound or a grammaticalized expression).

Uzbek examples are drawn from attested usage and descriptive grammars; where applicable, cross-checks are made against published Uzbek lexical studies and UzWordNet entries.

Proposed classification

I propose four primary lexical-semantic groups for Uzbek partonymic units:

Physical-component meronyms (Component/Part)

This group includes lexical units that denote physical components or structural parts of tangible wholes. Typical domains: human/animal anatomy, artifacts, plants.

Examples (Uzbek):

1. *oyoq* ‘foot’ — *tana* ‘body’ (body part).
2. *daraxt* ‘tree’ — *barg* ‘leaf’, *shox* ‘branch’.
3. *kompyuter* ‘computer’ — *ekran* ‘screen’, *klaviatura* ‘keyboard’.

Diagnostics: These meronyms typically participate in possessive constructions (*tananing boshi* ‘the body’s head’) and allow part-placement paraphrases (*X Yning bir qismi xisoblanadi* — ‘X is counted as a part of Y’). They show strong ontological dependence: in many contexts the meronym cannot meaningfully exist as that kind of entity without being a part of a holonym (e.g., *oyoq* outside any animate whole would be an anomalous referent).

Member–collective relations (Member meronymy)

Here lexical items denote members of an aggregate or collection.

Examples:

1. *qo* ‘y’ ‘sheep’ — *qo* ‘m/oxon’ ‘flock’ (regional terms vary).
2. *talaba* ‘student’ — *kurs* ‘class/cohort’ (institutional collectives).
3. *bin* ‘building’ — *mahalla* ‘neighborhood’ in some relational uses.

Diagnostics: Since members are autonomous people who can live independently of the collective, member meronyms are different from component meronyms in that membership is not ontologically constitutive in the same way that a sheep can exist without its flock. Uzbek has particular collective formations and derivations that represent the diagnostic interactions between plural marking and collective nouns in morphosyntactic constructions.

Substance–whole relations (Substance meronymy)

Lexical units that name materials or substances from which wholes are made fall here.

Examples:

1. *temir* ‘iron’ — *temir eshik* ‘iron door’ (material → artifact).
2. *bug‘doy* ‘wheat’ — *non* ‘bread’ (material/ingredient → product) — note cultural and lexicalization patterns affecting whether this is treated as meronymy or causal relation.

Diagnostics: Substitute tests (Does *temir* entail *temir eshik*?) often fail; the relation is not strictly part-of but material-of. In many lexical databases this relation is tagged separately from component meronymy. Context and collocational frequency are crucial to disambiguate.

Metaphorical, grammaticalized and derivative partonymy

Certain lexical items encode partonymic relations by derivation, metaphor, or idiom. These can include synecdochic uses (part standing for whole), idioms, and denominal derivations that incorporate a part term.

Examples:

1. Synecdoche: *qo‘l* ‘hand’ used metonymically to refer to *help* or *labor* in idiomatic expressions.
2. Derivations: compounds where a part term contributes a specialized meaning (e.g., *bosh-qosh* constructions with extended semantics).

Diagnostics: Semantic shift, non-literal interpretation, and morphosyntactic evidence of derivation (affixation, compounding) mark this group. These items are especially important for lexicographers because their partonymic origin may be obscured by semantic drift.

Results and Discussion

The four groups are not mutually exclusive in real language data. For example, *barg* (leaf) is a component meronym of *daraxt* (tree), but in poetic contexts it may be used metaphorically as a member of a collection (many leaves = foliage), or as part of an idiom. Likewise, substance relations often co-occur with component relations in artifact descriptions (a *taxta* ‘board’ is both a part of a table and a material). Clear annotation guidelines are therefore needed for lexicon projects [5].

In the analysis of part-whole (meronymic) relationships within lexical semantics, it becomes apparent that the classical four-way categorization—component, member, substance, and derivative—is not always cleanly separable in real-world language data [6]. While these distinctions are theoretically motivated and useful for computational modeling and lexicography, natural language often exhibits overlap, ambiguity, and context-dependent variation that challenge strict categorization.

For instance, consider the Uzbek noun *barg* (‘leaf’). At a basic, compositional level, it functions as a component meronym of *daraxt* (‘tree’), in that a leaf is physically part of the tree structure. However, in poetic, metaphorical, or idiomatic usage, *barg* can also operate in a member sense—one among many similar entities (e.g., “leaves” forming “foliage”). Moreover, it may appear in metaphorical expressions where its literal referent is not relevant, thus introducing figurative meronymy [7]. Such polyfunctionality is not exceptional, but rather commonplace, especially in morphologically rich and metaphorically vibrant languages like Uzbek.

Similarly, cases of substance–component overlap are frequent in everyday and technical discourse. Take *taxta* (‘board’) as an example. In one context, *taxta* may denote a wooden plank used in construction—a component of a table or floor. In another, it refers to wood as a material substance. This dual role requires lexicons and semantic databases to account for both interpretations simultaneously, possibly linking both senses to a shared semantic root but differentiating their usage conditions via tagged relations or context-sensitive rules [8].

These examples illustrate that meronymy in natural language is not a set of mutually exclusive categories, but rather a semantic continuum, influenced by grammatical constructions, cultural conventions, and discourse contexts. Therefore, the development of language technologies—

particularly in less-resourced languages like Uzbek—necessitates clear, granular annotation frameworks that capture such subtleties.

Implications for Lexicography and Uzwordnet

Given the overlapping and context-sensitive nature of meronymic relations in Uzbek, the design of digital lexical resources such as UzWordNet must adopt a more nuanced approach to semantic classification. Traditional lexicographic representations that lump all part-whole relations into a single “meronym” category fail to capture the important distinctions necessary for Natural Language Processing (NLP) tasks, including syntactic parsing, word sense disambiguation, information retrieval, and semantic similarity measurements [9].

A robust solution would involve tagging meronymic linkages with specific subtypes—namely, component, member, substance, and figurative. This layered approach improves both human interpretability and machine-readability, allowing semantic search engines and AI systems to better understand the nature of the relationships between lexical entries [10].

For instance, if *barg* is annotated as a component of *daraxt*, a member of *yaproqlar* (leaves/foilage), and a figurative element in an idiom like *bargdek titrash* (‘to tremble like a leaf’), then semantic parsers and translation models can more accurately generate, translate, or analyze texts based on the relevant usage context. Similarly, *taxta* can be tagged both as a substance (wood) and as a component (board), with contextual metadata guiding disambiguation.

Such fine-grained tagging also supports lexical inference and semantic expansion in WordNet-based applications.[11] For example, a semantic search for “parts of a table” could automatically retrieve *taxta*, *oyoq* (‘leg’), and *yuza* (‘surface’) if those component tags are present. Likewise, metaphorical uses of parts—such as *qalbning bir bo‘lagi* (‘a piece of the heart’)—could be treated under a distinct figurative meronymy label, enhancing both search precision and metaphor analysis [12].

UzWordNet’s architecture must also accommodate cross-linked senses. A single lexical unit may appear in multiple meronymy subtypes, depending on its sense. This calls for a multi-sense tagging strategy and sense-specific relation mapping, which, while resource-intensive, can dramatically increase the utility of the database for both human users and AI systems.

Although previous foundational work on Uzbek WordNets and related lexical databases has provided basic part-whole mappings, most existing annotations are shallow and lack the semantic granularity necessary for advanced applications. Manual validation and tagging of partonymic subtypes—based on real corpus data and linguistic intuition—remain a crucial next step. Computational methods such as automatic pattern recognition, distributional semantics, and machine learning classifiers can assist this process but will need to be guided by a linguistically grounded annotation schema [13].

Implications for Lexicography and Uzwordnet

Lexical entries and semantic networks can be better organized by using a classification that separates component/member/substance/derivative relations. Meronymic linkages should be tagged with subtypes (component/member/substance/figurative) by UzWordNet and related resources to enhance retrieval and disambiguation in NLP tasks (parsing, question answering, semantic similarity) [14]. Although previous work on Uzbek lexical networks provides a foundation for this kind of fine-grained tagging, partonymic subtypes need to be manually validated.

Cross-Linguistic Considerations

Part-whole differences are encoded differently in many languages (case marking, compounding, morphological derivation) [15]. Partonymic semantics interact with certain morphosyntactic mechanisms found in Uzbek, such as possessive suffixes, relational nouns, and collective markers. Comparative research that contrasts Uzbek patterns with typologically distinct languages would demonstrate language-specific trends and test the four-way division’s universality.

Conclusion

Physical-component meronyms, member-collective relations, substance-whole links, and metaphorical/derivative partonymy are the four divisions into which this article suggested a useful classification of Uzbek lexical words based on whole-part relations. The taxonomy is supported by corpus and lexical database evidence and is based on semantic diagnostics. Adding this classification to lexicographic resources (like UzWordNet) would enhance Uzbek semantic search, disambiguation, and natural language processing applications. Comparative typological studies and a systematic corpus annotation initiative should be part of future research.

References

- [1] Agostini et al., “UzWordNet: A Lexical–Semantic Database for the Uzbek Language,” in ACL Workshop, 2021.
- [2] D. Jurafsky and J. H. Martin, Lectures on Lexical Relations and WordNet, [Lecture Notes].
- [3] S. U. Mustafayeva, “Specific Features of Meronymy in English and Uzbek,” *Biruni Journal*, 2023.
- [4] “Correlation of Lexical-Semantic Relations with Grammatical Features in Uzbek,” *Scientists.uz*, [Online].
- [5] J. S. O’ktamovna, “Theoretical Perspectives on the Partonymic (Holo-Meronymic) Phenomenon,” 2025.
- [6] G. A. Miller, *WordNet: An Electronic Lexical Database*, Cambridge, MA: MIT Press, 1995.
- [7] “General Resources on Lexical Semantics and Meronymy (e.g., Brandeis handouts, Fiveable),” [Online].
- [8] A. Eshmuminov, “Linguistic Maintenance of the Database of Meronyms of the National Corpus of the Uzbek Language,” *American Journal of Language, Literacy and Learning in STEM Education*, vol. 2, no. 10, pp. 144–148, 2024.
- [9] S. Mustafaeva, “The Samenesses and Peculiarities of Meronymy in Uzbek and English Languages,” *American Journal of Language, Literacy and Learning in STEM Education*, vol. 2, no. 3, pp. 605–607, 2024.
- [10] A. Haydarov and M. Barnoeva, “Partonymy in the Uzbek Language: Lexico-Semantic Analysis,” *American Journal of Language, Literacy and Learning in STEM Education*, vol. 2, no. 11, pp. 253–258, 2024.
- [11] U. Salaev, “UzMorphAnalyser: A Morphological Analysis Model for the Uzbek Language Using Inflectional Endings,” *arXiv preprint*, 2024.
- [12] M. Sharipov, E. Kuriyozov, O. Yuldashev, and O. Sobirov, “UzbekTagger: The rule-based POS tagger for Uzbek language,” *arXiv preprint*, 2023.
- [13] M. Sharipov and O. Yuldashov, “UzbekStemmer: Development of a Rule-Based Stemming Algorithm for Uzbek Language,” *arXiv preprint*, 2022.
- [14] B. R. Tajiboyev, “Linguocultural Features of Lexical Units in Uzbek Language,” *Modern Science and Research*, 2024.
- [15] M. Proietti and A. Lenci, “The quasi-semantic competence of LLMs: a case study on the part-whole relation,” *arXiv preprint*, 2025.