**Multiple Prefixation in Uzbek Scientific Terminology:**

**A Lexicological Analysis**

**Abstract**

This study examines multiple prefixation in Uzbek—cases in which two or more left-edge combining forms precede a single base within scientific and technical vocabulary. Although prefixation is historically marginal in Uzbek, borrowed combining forms of Persian–Tajik and European origin have become increasingly integrated and productive in modern terminology. We develop a replicable segmentation protocol grounded in three diagnostics: (i) productivity across types, (ii) recoverability of the base as an independent lexeme/root, and (iii) stable semantic scope contributed by each left-edge element. Applying the protocol to a curated inventory (N = 69 multi-prefix lexemes), we find 64 double-stacks, 3 triple-stacks, and 2 quadruple-stacks. Recurrent stacking preferences include *radio+tele* (the most frequent pair), and right-headed derivation regularly selects category-forming suffixes after the stack; the most common bases are *graf* (33 items) and *metr* (8 items). Borderline items with opaque or non-recoverable bases are treated conservatively as unanalyzed complexes. We distinguish borrowed combining forms from native prefixes and formalize a practical template [Prefix A] + [Prefix B] + BASE + (Suffix). The study closes a descriptive gap in Uzbek morphology and provides pattern summaries (Tables 1–2) and a normalized, segmented inventory (Appendix A) to support lexicography, corpus annotation, computational processing, and teaching in terminology-rich domains.

**Keywords:** multiple prefixation; combining forms; Uzbek morphology; morphemic segmentation; lexicography; corpus linguistics.

**Introduction**

The linguistic evolution of Uzbekistan reflects the country’s complex cultural and historical dynamics. As a multilingual nation with a rich heritage, Uzbekistan has implemented active language planning and policy measures to promote both national identity and linguistic diversity. One of the most prominent features of these efforts is the advancement of English as a foreign language. Recognizing English as a global medium for communication, commerce, and technological advancement, the Uzbek government has launched several initiatives to improve English language proficiency, especially in educational and professional contexts. While Uzbek is the sole official state language, Russian continues to serve as a lingua franca among ethnic minorities and is widely spoken as a second native language. Since gaining independence, Uzbekistan has experienced a significant increase in the influence of English across various societal domains (Rahmanova and Ekşi 2023).

This growing linguistic diversity has contributed to the emergence of unique lexical formations in the Uzbek language. Among these is the relatively rare but linguistically significant phenomenon of multiple prefixation—the attachment of two or more prefixes of foreign origin to a single root. This structure is atypical within Turkic languages, which are predominantly suffixing and agglutinative by nature. While traditional grammar references acknowledge the existence of individual prefixes in Uzbek, limited scholarly attention has been given to cases where multiple prefixes co-occur in a single lexeme.

Examples such as astrogeofizika, radioelektrograf, and biogidroakustika illustrate how layered prefixes have become embedded in the Uzbek lexicon, especially within scientific and technical terminology. In these complex formations, each prefix contributes a specific semantic or grammatical function to the word as a whole. However, the systematic analysis of such structures remains underdeveloped in Uzbek morphological research.

This study seeks to address this gap by analyzing multi-prefix lexical items drawn from authoritative Uzbek explanatory, orthographic, and morpheme dictionaries. Using descriptive, classificatory, comparative, and componential methods, the research aims to provide a typologically informed framework for the segmentation and interpretation of these complex forms. In doing so, the study contributes to the broader understanding of morphological adaptation and lexical innovation in under-resourced and agglutinative languages like Uzbek.

To guide this investigation, the following objectives have been formulated:

This study aims to:

* Systematically identify and classify Uzbek words containing multiple prefixes, particularly those of foreign origin;
* Analyze the structural and semantic roles of each prefix within a single lexical unit;
* Examine the productivity and morphotactic behavior of prefix stacking in Uzbek;
* Bridge a critical gap in Uzbek morphological studies by offering a typologically informed framework for multi-prefix analysis.

Building on these objectives, the research is guided by the following questions:

1. What types of prefixes are most commonly stacked within a single lexeme in Uzbek, and what are their linguistic origins?
2. How does the presence of multiple prefixes affect the morphological segmentation and semantic interpretation of Uzbek words?
3. What morphological patterns and typological constraints govern prefix stacking in an agglutinative language like Uzbek?

**Literature Review**

In recent years, students have increasingly turned to artificial intelligence (AI) to support their academic tasks, including language learning. While AI has proven effective in many digitized languages (Mengliev and Gulyamova 2022; Abdurasulovich and Rajabovich 2019) — facilitating tasks such as word structure analysis, interlingual translation (Mirdjonovna et al. 2024; Shakirjanovna et al. 2023; Adalı, Rajabovich and Mirdjonovna 2022), and the resolution of grammatical, syntactic, and lexical issues — its application in under-resourced languages, such as Uzbek, remains limited. One unresolved challenge involves the segmentation of words into their structural components. In particular, the occurrence and function of multiple prefixes within a single lexical root remain insufficiently explored.

Although G‘ulomov’s Uzbek Language Grammar discusses prefixation, it does not address the phenomenon of multiple prefix usage (O‘zbekiston SSR Fanlar Akademiyasi 1975). However, this gap was partially addressed two years later in the Morpheme Dictionary of the Uzbek Language, which acknowledged that prefixes can be stacked, as illustrated in the term gidro/geo/log/iya (G‘ulomov et al. 1977). These early references suggest that layered prefixation has been present in Uzbek linguistic discourse since at least 1977. Nonetheless, systematic morphological analysis of such formations is still underdeveloped.

Corpus-based lexical studies have the potential to support the expansion of linguistic resources in underrepresented languages (Abdurasulovich and Rajabovich 2019), especially given that “the field of corpus linguistics is one of the most crucial topics in linguistics and related disciplines” (Mengliye, Hamroyeva and Abdullayeva 2023). In response, this study draws upon a wide body of international research on linguistic corpora and morphological analysis (Hwang 2024), as well as foundational work in lexicography (Lew 2024), prefixation (Schledewitz 2023), and morphemics (Behzadnia et al. 2023).

Recent developments in morphological analysis for low-resource and agglutinative languages have reignited scholarly interest in the complexities of word formation. Although prefixation is typically associated with Indo-European languages, cross-linguistic research increasingly highlights the growing productivity of borrowed prefixes in typologically unrelated languages, a trend driven by globalization and the internationalization of scientific terminology (Schledewitz 2023; Haspelmath 2020). Within this context, the phenomenon of multiple prefixation — where two or more foreign-origin prefixes attach to a single root — has emerged as an important area of investigation, particularly in hybrid lexical structures.

Several studies stress the importance of understanding morpheme sequencing and compatibility within borrowed compound formations. For example, Behzadnia et al. (Behzadnia et al. 2023) demonstrate how morphemic awareness enhances the acquisition of neologisms, particularly in scientific domains. In a related study, Divate (Divate 2024) proposes a hybrid morphological model capable of capturing compound morpheme interactions in low-resource languages — a framework that resonates with the analytical needs of Uzbek linguistic research.

Moreover, advances in computational tools have facilitated the morphological parsing of complex word forms. NLP applications such as UDPipe and morphological richness analyzers, as described by Straka & Straková (Straka and Straková 2017) and Hwang (Hwang 2024), illustrate how automated systems can perform effective morphological decomposition, even for opaque lexemes. However, such technologies are either unavailable or insufficiently adapted for Uzbek, necessitating manual, linguistically informed segmentation methods, as adopted in this study.

Linguistic typology research further underscores the uniqueness of prefixal morphotactics across languages (Haspelmath 2020; Bickel and Nichols 2013), warning against overgeneralized models. Uzbek, which does not employ native prefixation but increasingly adopts foreign-prefix layers, presents a compelling case for expanding existing typological frameworks.

Finally, the expanding role of AI in morphological segmentation and alignment (Adalı, Rajabovich and Mirdjonovna 2022) suggests that morphologically rich languages like Uzbek could eventually benefit from AI-driven analysis — provided that adequate annotated datasets exist. This reinforces the importance of foundational studies like the present one, which aim to contribute both empirical data and theoretical insight to support future computational linguistics tools.

**Methodology**

This study adopts a mixed-methods design that integrates descriptive linguistic analysis, corpus validation, and morphemic segmentation informed by contemporary morphological theory. The core objective is to examine the structural properties and frequency of Uzbek lexemes that incorporate multiple combining forms[[1]](#footnote-1)—primarily of foreign origin—within a single word form. Such formations are typologically rare in Turkic languages, including Uzbek, where prefixation is historically non-native but has gained visibility through lexical borrowings, especially in scientific and technical fields.

The lexical data were systematically collected from multiple authoritative Uzbek sources, including explanatory, orthographic, and morphemic dictionaries such as *O‘zbek tili grammatikasi* (1975), *O‘zbek tili morfem lug‘ati* (1977), and Mengliyev’s *So‘z yasalishi o‘quv lug‘ati* (2007). To ensure the inclusion of contemporary and domain-specific vocabulary, additional terms were extracted from specialized linguistic corpora, and a manually curated dataset was developed to reflect real-world usage in scientific and technological discourse. Only those lexemes that unambiguously exhibited two or more prefixal morphemes were included in the final sample.

Each selected word underwent componential morphemic analysis to determine its internal structure. The analysis involved the segmentation of each lexeme into prefixal morphemes (e.g., *astro-, geo-, bio-*), root or base morphemes (e.g., fizik, graf, metr), and suffixal morphemes (e.g., *-ist, -iya, -ik*). The segmentation criteria were grounded in cross-linguistic principles of morpheme identification, giving priority to morphological elements that hold clear lexical, grammatical, or derivational value (Haspelmath 2020). Special attention was given to borrowed prefixes that have undergone structural and functional adaptation in Uzbek, thus marking a shift from phonological integration to productive morphological assimilation.

The prefixal elements were further analyzed typologically, focusing on their etymological origin, level of productivity, and functional roles within the compounds. Prefixes were grouped into categories such as lexicalized borrowings (e.g., *avto-*, *elektro-*), scientific and technical affixes used predominantly in academic discourse (e.g., *termo-*, *foto-*), and hybrid constructs formed by multi-source elements (e.g., *radio/tele/graf/ist*). This classification enabled a deeper investigation into both morphotactic compatibility and semantic layering.

To validate the empirical robustness of the analysis, all segmented forms were cross-referenced with digitized Uzbek language corpora, including national corpus prototypes, academic textual databases, and technical documentation repositories. Variables such as word frequency, contextual distribution, and morphological variation were documented to evaluate the productive capacity and structural behavior of multi-prefix constructions.

The theoretical anchoring of the study draws upon the item-and-arrangement model (Bloomfield 1933; Hockett 1954), Distributed Morphology (Halle & Marantz 1993), and principles of typological morphology (Spencer & Zwicky 2001). These frameworks provided the tools to distinguish morphotactic sequences from surface-level phonological strings, particularly in cases of complex borrowing and affix stacking. Additionally, where feasible, AI-assisted morphological parsers adapted from tools such as UDPipe and spaCy were deployed to verify segmentation accuracy. While these tools are not yet fully optimized for the Uzbek language, custom rule-based modules enabled partial adaptation and provided an auxiliary verification layer.

This study employed a qualitative linguistic analysis combined with computational validation to investigate the presence and function of multiple prefixes in Uzbek compound lexemes, particularly those of foreign origin. The selection of lexical items was conducted through purposive sampling from existing Uzbek explanatory and orthographic dictionaries, with an emphasis on words containing more than one prefix element (e.g., "*astrogeofizika*," "*gidrogeologiya*").

To ensure a structured morphological breakdown, we manually segmented each lexeme into its constituent morphemes, identifying roots, derivational prefixes, and inflectional elements based on existing morphological theories and references in Uzbek grammar sources (O‘zbekiston SSR Fanlar Akademiyasi 1975). Particular attention was given to the identification of prefix-like components that, although not native to Turkic languages, have been borrowed primarily from Indo-European languages due to scientific and technical expansion.

Where feasible, we implemented AI-supported morphological parsers such as UDPipe and spaCy, which were partially adapted through rule-based modules to suit the structure of the Uzbek language. These tools provided secondary verification layers for the manually segmented lexemes, although their optimization for Uzbek remains limited (Mengliyev et al. 2021). The rule-based approach proved especially useful in identifying consistent patterns of segmentation, particularly in complex formations that involve two or more prefixes.

This was especially relevant considering that prefixation, in general, is not typical for Uzbek or other Turkic languages. A key challenge in this research was the presence of multiple prefixes before a single root in scientific and technical terms. For instance, in the word "*astrogeofizika*," the component “*astro-*” also appears in other derivatives such as “*astrografik*” and “*astrofizik*,” suggesting its productivity and prefix-like behavior in modern Uzbek. The lexeme “*fizik*” (physics), on the other hand, exists independently in Uzbek dictionaries, underscoring that “*astro-*” operates as a modifier attached prior to a root. This analysis aligns with historical accounts that many of the prefixes currently found in Uzbek have been borrowed from Persian-Tajik and other Indo-European languages (O‘zbekiston SSR Fanlar Akademiyasi 1975).

The formation of words and their internal structure was treated as a foundational step in conducting morphological analysis. Recognizing that language is built on specific character sequences governed by syntagmatic rules, we acknowledged the need to distinguish true linguistic units from coincidental character chains. Lexemes — as the minimal meaningful units — were analyzed in light of their symbolic combinations, which obey both linguistic construction and external symbolic rules (Mengliyev et al. 2021).

In conclusion, this methodological framework allowed for the development of a database of Uzbek compound words containing multiple prefixes, with both linguistic and computationally verified segmentation, thereby laying groundwork for future applications in AI-driven corpus linguistics.

**Results and Discussion**

The analysis shows that multiple prefixation—the attachment of two or more foreign-origin elements to a single Uzbek base—occurs systematically in scientific and technical vocabulary. Although Turkic languages are predominantly suffixing and lack native prefixation, Uzbek has incorporated a set of left-edge formatives (e.g., *aero-, geo-, foto-, radio-, tele-, baro-, termo-, kardio-, xromo-, paleo-*) through borrowing and domain expansion. Descriptive traditions differ: for instance, *O‘zbek tili grammatikasi* treats *aeroport* as a borrowing and refrains from internal division (O‘zbekiston SSR Fanlar Akademiyasi 1975), whereas Berdialiyev and subsequent pedagogical sources recognize recurrent left-edge elements as derivational (foreign) prefixes when they are productive and semantically stable (Berdialiyev and Ermatov 2022). In this study, such elements are treated as prefixal morphemes when they (i) recur across a family of lexemes, (ii) attach to an attested or recoverable base, and (iii) contribute a consistent semantic scope.

Segmentation decisions were guided by three diagnostics: productivity, base independence, and evidence from right-edge derivation. Productivity is established by the recurring presence of *aero-, foto-, geo-, radio-, tele-* and others across numerous items. Base independence is verified in Uzbek lexicography for stems such as *fizik* ‘physics’, *graf* ‘writing/drawing’, *metr* ‘measure’, *fon* ‘sound’, *kimyo/ximiy(a)* ‘chemistry’, *klub* ‘club’. Right-edge morphology—-*ist, -iya, -ik, -a, -lar*—reliably signals part-of-speech and supports boundary placement. Accordingly, items such as *aerofotogeodezist → aero/foto/geodezist, aerogelioterapiya → aero/gelio/terapiya,* and *aerogeofizika → aero/geo/fizik/a* are segmented in line with Uzbek practice, which treats frequently recurring initial elements as derivational prefixes while recognizing the right-hand segment as the base (O‘zbekiston SSR Fanlar Akademiyasi 1975). Here *gelio-* conveys a solar (‘sun’) meaning consistent with borrowed scientific usage. Where available, Mengliyev’s word-formation guidance provides independent confirmation: *ampermetr → amper/metr* motivates *termometr → termo/metr*, and by analogy *aerotermometr → aero/termo/metr* (Mengliyev 2007). More generally, measurement bases (e.g., *metr*) and writing/recording bases (e.g., *graf*) show strong compatibility with stacked scientific prefixes, often followed by category-forming suffixes (e.g., *-ik, -iya*).A qualitative scan of the inventory (Appendix A) reveals several robust pattern families. One productive schema is a double scientific prefix layered over a measurement/writing base with or without a suffix: *aero/geo/fizik/a; baro/termo/graf; baro/mano/metr; giper/geo/metr/ik; izo/geo/termik; xromo/foto/graf/iya; paleo/geo/graf/ik ~ /iya; zoo/geo/graf/ik ~ /iya*. As treated in the Uzbek sources, *mano-* is analyzed in items like *manometr, sfigmomanometr, mikromanometr*, and contributes a measurement-related sense in combinations such as *baro/mano/metr*.

A second cluster comprises radio/tele series with stable bases and derivations: *radio/tele/fon; radio/tele/graf/ist ~ /iya; radio/tele/metr/ik ~ /iya; radio/tele/mexanik/a; radio/aero/navigatsiya; radio/dez/informatsiya; radio/gidro/meteorolog/ik; radio/avto/graf/ik ~ /iya; radio/foto/graf/iya*.

A third set shows vehicle/automation plus scientific stacks: *avto/gemo/terapiya; avto/lito/graf/iya; avto/moto/klub; avto/moto/lotereya; avto/moto/maktab; avto/poli/ploidiya.* Finally, *hydro/geo* combinations are common: *gidro/aero/ion/izatsiya; gidro/aero/mexanik/a; gidro/geo/kimyo ~ /ximiya; gidro/izo/gips/lar; gidro/per/oksid*—with *per-* justified by limited but consistent recurrence (*gidroperoksid, perxlorat*) and a transparent base (o*ksid*) (see Table 1).

**Table 1. Recurrent stacks of left-edge combining forms in Uzbek: [Prefix 1] + [Prefix 2] + BASE + (Suffix), segmented examples and typical domains**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Prefix 1 | Prefix 2 | Base | Suffix | Example (segmented) | Notes / Typical domain |
| *aero-* | *geo-* | *fizik* | *-a* | *aero/geo/fizik/a* | air + earth + physics → noun |
| *aero-* | *termo-* | *metr* | — | *aero/termo/metr* | air + heat + measurement device |
| *baro-* | *termo-* | *graf* | — | *baro/termo/graf* | pressure + heat + recording |
| *baro-* | *mano-* | *metr* | — | *baro/mano/metr* | pressure + “mano-” gauge (device) |
| *giper-* | *geo-* | *metr* | *-ik* | *giper/geo/metr/ik* | excess + earth + measure → adjective |
| *izo-* | *geo-* | *termik* | — | *izo/geo/termik* | equal/similar + earth + heat (adj.) |
| *xromo-* | *foto-* | *graf* | *-iya* | *xromo/foto/graf/iya* | color + light + writing → noun |
| *paleo-* | *geo-* | *graf* | *-ik* | *paleo/geo/graf/ik* | ancient + earth + writing → adjective |
| *zoo-* | *geo-* | *graf* | *-iya* | *zoo/geo/graf/iya* | animal + earth + writing → noun |
| *radio-* | *tele-* | *fon* | — | *radio/tele/fon* | radio + distance + sound |
| *radio-* | *tele-* | *graf* | *-ist* | *radio/tele/graf/ist* | radio + distance + writing → agent |
| *radio-* | *tele-* | *metr* | *-ik* | *radio/tele/metr/ik* | radio + distance + measure → adjective |
| *bio-* | *gidro-* | *akustik* | *-a* | *bio/gidro/akustik/a* | life + water + sound → noun |
| *avto-* | *gemo-* | *terapiya* | — | *avto/gemo/terapiya* | self/auto + blood + therapy |
| *fito-* | *geo-* | *graf* | *-iya* | *fito/geo/graf/iya* | plant + earth + writing → noun |
| *etno-* | *geo-* | *graf* | *-ik* | *etno/geo/graf/ik* | ethnicity + earth + writing → adjective |

While these formations are right-headed, ordering at the left edge is not random. A frequent tendency is domain/semantic classifier → measurement/recording-friendly domain → base, yielding sequences like *aero-/geo-/fizik, xromo-/foto-/graf, giper-/geo-/metr,* where the first prefix sets a domain, the second narrows functional scope, and the base provides the lexical core. This pattern accords with typological accounts that caution against projecting a single universal template for prefix stacking, since languages display idiosyncratic ordering and compatibility constraints (Schledewitz 2023; Haspelmath 2020). In Uzbek, compatibility is modulated by the availability of an attested base and by suffixal requirements: items lacking a recoverable base (e.g., *geodez*) are conservatively analyzed with the larger right-edge unit (e.g., *geodezist*) as the base, whereas items with clear bases admit segmentation.

Borderline and non-segmentable cases are treated conservatively. In particular, although *-olog* is recognizable, we do not segment *meteorolog*; accordingly, *radiogidrometeorologik* is analyzed as *radio/gidro/meteorolog/ik* rather than segmenting *-olog* as an independent morpheme. We apply two filters: (i) no base, no cut (e.g., *geodez* is not independently attested; division stops at *geodezist*), and (ii) lexical opacity blocks segmentation even where a string such as getero- is productive elsewhere (e.g., *geterodinlamoq*). These decisions reflect standard lexicographic practice and Uzbek word-division rules (O‘zbekiston SSR Fanlar Akademiyasi 1975). Conversely, elements such as *brone-* (‘armored’)—though not always separately glossed like *avto*-—show patterned occurrence and compositional meaning (e.g., *brone/avto/mobil*), supporting their treatment as prefixal in compounds.

Right-edge suffixes behave regularly and strengthen segmentation decisions. *-ist* forms agentive/professional nouns *(…/graf/ist*), *-iya* nominalizes complex bases (…/*graf/iya; …/metriya*), *-ik* derives adjectives *(…/metr/ik; …/graf/ik*), *-a* nominalizes, and -lar marks plurality. In getero/poli/birik/ma/lar, for example, the sequence *-ik → -ma → -lar* transparently yields a plural nominal from a verbal stem, demonstrating that stacked prefixes integrate smoothly with Uzbek derivational and inflectional morphology (see Table 2).

**Table 2. Right-edge suffixes following stacked combining forms in Uzbek: functions, typical POS outcomes, and segmented examples**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Suffix | Function | Typical POS outcome | Example (segmented) | Notes |
| *-ist* | Agent / professional derivation | Noun (agent) | *radio/tele/graf/ist* | Productive in borrowed scientific titles |
| *-iya* | Nominalization of complex bases | Noun | *xromo/foto/graf/iya*; *radio/tele/metr/iya* | Common in technical/scientific terms |
| *-ik* | Adjectival derivation | Adjective | *giper/geo/metr/ik*; *astro/foto/graf/ik* | Often marks technical adjectives |
| *-a* | Nominalizer / category marker | Noun | *aero/geo/fizik/a*; *radio/tele/mexanik/a* | Signals nounhood of complex base |
| *-lar* | Plural marker | Noun (plural) | *getero/poli/birik/ma/lar* | Attaches after derivation/nominalization |
| *-uv* | Deverbal noun (result/process) | Noun (verbal noun) | *radio/tele/bosh/qar/uv* | Parallel to *-ish*; register-dependent choice |
| *-ish* | Verbal noun (action/process) | Noun (verbal noun) | *radio/tele/bosh/qar/ish* | Highly productive Uzbek masdar |
| *-ma* | Result/state nominalizer from verb | Noun (result) | *getero/poli/bir/ik/ma/…* | Often combines with *-lar* for plurality |
| *-viy* | Adjectival derivation from noun | Adjective | *bio/geo/kimyo/viy* | Common in scientific adjectives (e.g., *kimyoviy*) |

Returning to the research questions, we first address **RQ1**. The distribution shows that the most frequent left-edge elements are scientific/technical borrowings—*aero-, geo-, foto-, radio-, tele-, baro-, termo-, kardio-, xromo-, paleo-—*alongside domain labels such as *izo-, giper-,* and *brone-*. As Table 1 and Appendix A illustrate, these prefixes recur productively with well-attested bases (*metr, graf, fizik, kimyo/ximiy(a), fon, klub*), forming stable series like *radio/tele/fon; radio/tele/graf/ist ~ /iya; aero/geo/fizik/a; xromo/foto/graf/iya.* Etymologically, they are primarily Greek/Latin (often via Russian scientific registers), with some items traceable to Persian–Tajik or Slavic sources; their recurrence and semantic stability support treating them as derivational prefixes in Uzbek when attached to an attested base (O‘zbekiston SSR Fanlar Akademiyasi 1975).

Building on this distributional picture, **RQ2** concerns segmentation and meaning. The data support a layered analysis in which stacked prefixes contribute additive domain semantics at the left edge, a recoverable base anchors the lexical core, and right-edge suffixes determine the grammatical category. Thus, *aero- ‘air’ + geo- ‘earth’ + fizik* yields *aero/geo/fizik/a* (noun), *radio- + tele- + graf + -ist* yields *radio/tele/graf/ist* (agent noun), and *xromo- + foto- + graf + -iya* yields *xromo/foto/graf/iya* (noun). The three diagnostics already applied—prefix productivity, base independence in Uzbek lexicography, and derivational evidence from suffixes—guide boundary placement (see Table 2). Where a supposed base is not independently attested (e.g., *geodez*), segmentation stops at the larger unit (*geodezist*); where the interior is opaque (e.g., *geterodinlamoq*), further cutting is avoided. Conversely, even if an element can occur independently in some contexts (e.g., *avto-*), its patterned use at the left edge of compounds supports a prefixal analysis in multi-part words (O‘zbekiston SSR Fanlar Akademiyasi 1975).

Concerning **RQ3**, the observed stacks conform to a consistent, right-headed template—**[Prefix A] + [Prefix B] + BASE + (Suffix)**—with ordering preferences that mirror semantic scope: a broader domain prefix typically precedes a more specific technical prefix before combining with the base (e.g., *aero-/geo-/fizik; xromo-/foto-/graf; giper-/geo-/metr*). These tendencies interact with Uzbek morphotactics via practical constraints: no base, no cut; opaque segments are not split; and the result must remain compatible with Uzbek suffixal morphology *(-ist, -iya, -ik, -a, -lar*). In effect, Uzbek maintains its agglutinative, right-headed character while integrating borrowed left-edge morphemes that behave derivationally once assimilated—an outcome consistent with typological expectations and prior descriptions (O‘zbekiston SSR Fanlar Akademiyasi 1975).

Taken together, the patterns support a right-headed template**: [Prefix A] + [Prefix B] + BASE + (Derivational/Category Suffix)**. Here, Prefix A/B contribute domain semantics (e.g., *aero- ‘air’, geo-* ‘earth’, *foto-* ‘light’, *radio-* ‘radio’, *tele-* ‘distance/television’, *baro-* ‘pressure’, *termo-* ‘heat’, *kardio-* ‘heart’), the BASE is attested *(graf, metr, fizik, kimyo/ximiy(a), fon, klub, akustik, meteorolog*), and the suffix assigns grammatical category. This aligns with general morphological theory on derivation vs. compounding and with typological caution against overgeneralizing prefixal patterns across languages (Schledewitz 2023; Haspelmath 2020). For Uzbek—an agglutinative language with borrowed prefixal material—the results suggest a hybrid morphotactics: native suffixal architecture plus imported left-edge morphemes that behave derivationally once integrated.

The findings also indicate a route to harmonize dictionary practice: recurring left-edge elements with demonstrable productivity and stable semantics should be marked as prefixal in entries and segmented when attached to attested bases, while opaque formations remain whole. This would improve headword consistency, cross-references, and user guidance in explanatory, orthographic, and morpheme dictionaries (Lew 2024). The curated inventory in Appendix A provides a normalized starting set of segmentations to support such standardization.

From a computational perspective, recognizing prefix stacks is essential for accurate tokenization, lemmatization, and morphological tagging. Existing pipelines (e.g., UDPipe) and morphological richness analyzers can be adapted to respect these morpheme boundaries when language-specific rules and lexicons are added (Hwang 2024; Straka and Straková 2017). The present inventory and diagnostics can seed rule-based modules and training data for Uzbek-specific parsers, improving robustness in technical registers. Prior Uzbek computational efforts (e.g., analyzers and aligners) provide a base for integration (Mirdjonovna et al. 2024; Shakirjanovna et al. 2023; Mengliyev et al. 2021; Adalı, Rajabovich and Mirdjonovna 2022), while insights from morphemic knowledge (Behzadnia et al. 2023) and hybrid models for low-resource languages (Divate 2024) indicate practical design choices for segmentation workflows.

Finally, because morphemic awareness supports the learning of complex terminology—especially in scientific discourse—systematic segmentation of multi-prefix lexemes can assist terminology instruction, dictionary use, and reading strategies in higher education (Behzadnia et al. 2023). Clear, consistent segmentation also aids translator training and corpus annotation, with downstream benefits for domain-specific L2 instruction. In sum, Uzbek exhibits a consistent right-headed, multi-prefix template in technical vocabulary. When prefixal elements recur productively and attach to attested bases, they behave as derivational morphemes in Uzbek, even though prefixation is non-native typologically. The results reconcile divergent descriptive practices (O‘zbekiston SSR Fanlar Akademiyasi 1975), align with word-formation guidance (Mengliyev 2007), and accord with string-to-morpheme constraints in written language (Mengliyev et al. 2021). The diagnostics and the Appendix A inventory provide a practical foundation for lexicographic normalization and computational processing of Uzbek multi-prefix constructions.

**Conclusion**

This study provides a clear, unified account of multiple prefixation in Uzbek—cases where two or more foreign-origin elements attach to a single base. Drawing on authoritative dictionaries and corpus checks, and using transparent diagnostics (prefix productivity, base independence, and right-edge derivation), we showed that recurring left-edge elements (e.g., *aero-, geo-, foto-, radio-, tele-, baro-, termo-, kardio-, xromo-, paleo-*) function as derivational prefixes when they combine with attested bases (e.g., *graf, metr, fizik, kimyo/ximiy(a), fon, klub*) and regular category-forming suffixes (e.g., *-ist, -iya, -ik, -a, -lar*). The results converge on a consistent right-headed template—[**Prefix A] + [Prefix B] + BASE + (Suffix)**—with stable ordering tendencies that mirror semantic scope. In short, Uzbek retains its agglutinative character while productively integrating borrowed left-edge morphology.

Beyond establishing the phenomenon, the paper delivers practical outputs for research and tooling. The pattern summary (Table 1) and suffix functions (Table 2), together with the normalized inventory (Appendix A), provide ready-to-use resources for lexicographic standardization and for initializing Uzbek-specific NLP components (tokenization, lemmatization, morphological parsing). The analysis also reconciles divergent descriptive practices by specifying when segmentation is warranted (recurrent, semantically stable prefixes + attested base) and when it should be avoided (no independent base; internal opacity).

Taken together, these contributions supply a replicable method and a reusable dataset that support future work in Uzbek morphology, lexicography, and language technology. They also offer immediate pedagogical value: consistent segmentation of complex, multi-prefix terms can facilitate terminology instruction, translator training, and corpus annotation across scientific and technical domains.

**Limitations**

This study is subject to several constraints.

* The dataset is dominated by dictionary-attested and technical/scientific items; colloquial usage and diverse genres are underrepresented.
* Limited corpus coverage means that frequency and productivity estimates for specific prefix stacks are provisional.
* Because Uzbek morphological tooling is still developing, many boundaries required manual expert judgment, limiting scalability and full automation.
* The study does not include native-speaker or psycholinguistic evidence to test acceptability and interpretation of stacked forms.
* We did not implement or report inter-annotator agreement under a formal annotation protocol, so reproducibility can be improved.

**Recommendations**

To build on this research, the following recommendations are proposed:

* Future studies should broaden the corpus beyond dictionary-attested technical items and should include brief native-speaker surveys or tasks to confirm segmentation choices and semantic interpretations across registers.
* A public guideline for morpheme boundaries—aligned with this study’s diagnostics (prefix productivity, base independence, right-edge derivation)—should be published, together with a small, double-annotated benchmark reporting inter-annotator agreement.
* The normalized inventory (Appendix A) and patterns (Tables 1–2) should be integrated into a morphological analyzer that explicitly models stacked prefixes, with evaluation reported at morpheme-boundary precision/recall and targeted error analysis.
* Explanatory, orthographic, and morpheme dictionaries should consider marking recurrent left-edge elements as prefixal where warranted and adopt consistent segmentation conventions in line with the [Prefix A] + [Prefix B] + BASE + (Suffix) template.

**Contribution to the Field**

This work (i) documents an under-described phenomenon in Turkic morphology; (ii) clarifies when stacked left-edge elements act as derivational prefixes; (iii) articulates a consistent right-headed template with workable constraints; and (iv) delivers reusable resources—Tables 1–2 and Appendix A—that enable replicable analysis, lexicographic standardization, and tooling for Uzbek.

**Data Availability Statement**

All data supporting the findings—including the segmented inventory (Appendix A, Table A1) and pattern summaries —are contained in the article and its appendix. Additional materials are available from the corresponding author upon reasonable request.

**Conflict of Interest**

The authors declare no commercial or financial relationships that could be construed as a potential conflict of interest.

**Funding**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Acknowledgments**

We thank colleagues for comments on Uzbek morpheme segmentation and acknowledge the use of publicly available explanatory, orthographic, and morpheme dictionaries that informed the dataset. Any remaining errors are our own.

**Ethics Statement**

This study did not involve human participants, personal data, or interventions requiring ethics approval.

**Supplementary Material**

Appendix A (Table A1) provides the complete inventory of multi-prefix lexemes with recommended segmentation. If the journal requires separate files, upload Appendix A as “Supplementary Material.”

# References

Abdurasulovich, K. R., & Rajabovich, M. B. (2019). The role of the parallel corpus in linguistics, the importance and the possibilities of interpretation. International Journal of Engineering and Advanced Technology, 8(5s3), 388–391. https://doi.org/10.35940/ijeat.e1083.0785s319

Adalı, E., Rajabovich, M. B., & Mirdjonovna, K. S. (2022). Comparison of Uzbek–Turkish inflection suffixes. 2021 6th International Conference on Computer Science and Engineering (UBMK). https://doi.org/10.1109/ubmk55850.2022.9919603

Behzadnia, A., Ziegler, J. C., Colenbrander, D., Bürki, A., & Beyersmann, E. (2023). The role of morphemic knowledge during novel word learning. Quarterly Journal of Experimental Psychology, 77(8), 1620–1634. https://doi.org/10.1177/17470218231216369

Berdialiyev, A., & Ermatov, I. (2022). Hozirgi o‘zbek adabiy tili. Tashkent: Tamaddun.

Bickel, B., & Nichols, J. (2013). Inflectional synthesis of the verb. In M. S. Dryer & M. Haspelmath (Eds.), The World Atlas of Language Structures Online. Leipzig: Max Planck Institute.

Bloomfield, L. (1933). Language. New York: Henry Holt and Company.

Divate, M. S. (2024). Hybrid Morph-Analysis model for Marathi. In Lecture Notes in Networks and Systems (pp. 323–335). https://doi.org/10.1007/978-3-031-64779-6\_31

G‘ulomov, A., va boshq. (1977). O‘zbek tili morfem lug‘ati. Toshkent: O‘qituvchi.

Halle, M., & Marantz, A. (1993). Distributed Morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), The View from Building 20 (pp. 111–176). Cambridge, MA: MIT Press.

Haspelmath, M. (2020). Understanding morphology (2nd ed.). London: Routledge. https://doi.org/10.4324/9780429294186

Hockett, C. F. (1954). Two models of grammatical description. Word, 10(2–3), 210–234. https://doi.org/10.1080/00437956.1954.11659524

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. Retrieved from https://spacy.io/

Hwang, H. (2024). Development of morphological diversity in second-language Korean: An NLP analysis using the Korean Morphological Richness Analyzer 1.0. System, 121, 103260. https://doi.org/10.1016/j.system.2024.103260

Lew, R. (2024). Dictionaries and lexicography in the AI era. Humanities and Social Sciences Communications, 11(1). https://doi.org/10.1057/s41599-024-02889-7

Mengliyev, B. (2007). O‘zbek tilining so‘z yasalishi o‘quv lug‘ati. Toshkent.

Mengliyev, B., Shahabitdinova, S., Khamroeva, S., Gulyamova, S., & Botirova, A. (2021). The morphological analysis and synthesis of word forms in the linguistic analyzer. Journal of Language and Linguistic Studies, 17(1), 558–564. https://doi.org/10.52462/jlls.37

Mengliev, B., & Gulyamova, S. (2022). Linguistic modeling of polysemous words for the semantic analyzer of the Uzbek language. 2021 6th International Conference on Computer Science and Engineering (UBMK), 1–5. https://doi.org/10.1109/ubmk55850.2022.9919553

Mengliyev, B. R., Hamroyeva, Sh., & Abdullayeva, O. (2023). Scopus-based bibliometric analysis on corpus linguistics for the period of 2017–2021. E3S Web of Conferences, 413, 03008. https://doi.org/10.1051/e3sconf/202341303008

Mirdjonovna, K. S., Mengliyev, B., Kholova, M., & Gulyamova, S. (2024). The 80/20 principle in Morphemics–Morphology in the educational corpus of the Uzbek language. 2021 6th International Conference on Computer Science and Engineering (UBMK), 1020–1023. https://doi.org/10.1109/ubmk63289.2024.10773585

O‘zbekiston SSR Fanlar Akademiyasi. (1975). O‘zbek tili grammatikasi. Toshkent: Fan.

Rahmanova, G., & Ekşi, G. Y. (2023). English-medium instruction in higher education in Uzbekistan: Views on effectiveness, career prospects and challenges. World Journal of English Language, 13(5), 458. https://doi.org/10.5430/wjel.v13n5p458

Schledewitz, D. (2023). Verbal prefix do- in Russian and Ukrainian. Poljarnyj Vestnik, 26, 24–34. https://doi.org/10.7557/6.7189

Spencer, A., & Zwicky, A. M. (Eds.). (2001). The Handbook of Morphology. Oxford: Blackwell. https://doi.org/10.1002/9781405166348

Straka, M., & Straková, J. (2017). Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In Proceedings of the CoNLL 2017 Shared Task (pp. 88–99). Association for Computational Linguistics. https://doi.org/10.18653/v1/K17-3009

**Appendix A — Table A1. Inventory of Uzbek multi-prefix lexemes**

**(recommended segmentation)**

|  |  |
| --- | --- |
| **Lexeme** | **Recommended segmentation** |
| aerofotogeodezist | *aero/foto/geodezist* |
| aerogelioterapiya | *aero/gelio/terapiya* |
| aerogeofizika | *aero/geo/fizik/a* |
| aerotermometr | *aero/termo/metr* |
| agrozootexnika | *agro/zoo/texnik/a* |
| astrofotometrik | *astro/foto/metr/ik* |
| astrofotometriya | *astro/foto/metr/iya* |
| avtogemoterapiya | *avto/gemo/terapiya* |
| avtolitografiya | *avto/lito/graf/iya* |
| avtomotoklub | *avto/moto/klub* |
| avtomotolotereya | *avto/moto/lotereya* |
| avtomotomaktab | *avto/moto/maktab* |
| avtopoliploidiya | *avto/poli/ploidiya* |
| baromanometr | *baro/mano/metr* |
| barotermograf | *baro/termo/graf* |
| barotermometr | *baro/termo/metr* |
| biogeokimyoviy | *bio/geo/kimyo/viy* |
| biogidroakustika | *bio/gidro/akustik/a* |
| broneavtomobil | *brone/avto/mobil* |
| fitogeografiya | *fito/geo/graf/iya* |
| geteropolibirikmalar | *getero/poli/birik/ma/lar* |
| gidroaeroionizatsiya | *gidro/aero/ion/izatsiya* |
| gidroaeromexanika | *gidro/aero/mexanik/a* |
| gidrogeokimyo | *gidro/geo/kimyo* |
| gidrogeoximiya | *gidro/geo/ximiya* |
| gidroizogipslar | *gidro/izo/gips/lar* |
| gidroperoksid | *gidro/per/oksid* |
| gipergeometrik | *giper/geo/metr/ik* |
| izoantigenlar | *izo/anti/gen/lar* |
| izogeotermik | *izo/geo/termik* |
| radioaeronavigatsiya | *radio/aero/navigatsiya* |
| radiodezinformatsiya | *radio/dez/informatsiya* |
| radiogidrometeorologik | *radio/gidro/meteorolog/ik* |
| radioteleboshqarish | *radio/tele/bosh/qar/ish* |
| radioteleboshqaruv | *radio/tele/bosh/qar/uv* |
| radiotelefon | *radio/tele/fon* |
| radiotelegrafist | *radio/tele/graf/ist* |
| radiotelegrafiya | *radio/tele/graf/iya* |
| radiotelemetrik | *radio/tele/metr/ik* |
| radiotelemetriya | *radio/tele/metr/iya* |
| radiotelemexanika | *radio/tele/mexanik/a* |
| astrofotografik | *astro/foto/graf/ik* |
| astrofotografiya | *astro/foto/graf/iya* |
| astrogeografiya | *astro/geo/graf/iya* |
| antropogeografiya | *antropo/geo/graf/iya* |
| agrogeografiya | *agro/geo/graf/iya* |
| biobibliografik | *bio/biblio/graf/ik* |
| biobibliografiya | *bio/biblio/graf/iya* |
| biogeografiya | *bio/geo/graf/iya* |
| etnogeografik | *etno/geo/graf/ik* |
| etnogeografiya | *etno/geo/graf/iya* |
| fototelegrafiya | *foto/tele/graf/iya* |
| makrofotografiya | *makro/foto/graf/iya* |
| makroradiografiya | *makro/radio/graf/iya* |
| mikrobarograf | *mikro/baro/graf* |
| paleogeografik | *paleo/geo/graf/ik* |
| paleogeografiya | *paleo/geo/graf/iya* |
| radioavtografik | *radio/avto/graf/ik* |
| radioavtografiya | *radio/avto/graf/iya* |
| radiofotografiya | *radio/foto/graf/iya* |
| sotsiogeografiya | *sotsio/geo/graf/iya* |
| stereofotografiya | *stereo/foto/graf/iya* |
| stereokartografiya | *stereo/karto/graf/iya* |
| teleelektrokardiograf | *tele/elektro/kardio/graf* |
| vibrokardiografiya | *vibro/kardio/graf/iya* |
| xromofotografiya | *xromo/foto/graf/iya* |
| xromolitografiya | *xromo/lito/graf/iya* |
| zoogeografik | *zoo/geo/graf/ik* |
| zoogeografiya | *zoo/geo/graf/iya* |

1. In this paper, we use the term **combining forms** for borrowed left-edge confixoids (e.g., *aero-, geo-, radio-*). For brevity, we occasionally refer to them as “prefixes,” but analytically they are treated as combining forms distinct from native prefixes. [↑](#footnote-ref-1)