

THE DEGREE OF PRODUCTIVITY OF PREFIXES IN THE MODERN UZBEK LANGUAGE

Narzullayeva Muborak Sherzod qizi

Independent Researcher, Gulistan State University, Uzbekistan

E-mail: narzullayevamuboraksherzodovna@gmail.com

Abstract

The article investigates the productivity of the prefixal formatives of modern Uzbek and proposes a graded, criterion-based scale in place of the undifferentiated inventories found in the descriptive literature. Productivity is understood in the sense established in general morphology as availability for new coinage, a property distinct from the number of types a pattern has already produced. Because a lemmatised and balanced corpus of modern Uzbek is not yet available, a multi-indicator ordinal procedure is adopted, resting on four diagnostics: dictionary type count, formation of hybrids on inherited Turkic bases, attestation in post-2000 terminology, and acceptability of nonce formations. The procedure yields four tiers, ranging from the fully productive *be-*, *no-*, *ser-*, *ham-*, *anti-*, *mikro-*, *eko-* and *super-* to the fossilised *bo-*, *ba-*, *dil-* and *sar-*. It is shown that type frequency alone misclassifies the fossilised group, which confirms the need for a multi-indicator approach. The principal constraint on productivity is identified as rivalry with the inherited suffixes *-siz*, *-li* and *-dosh*, which yield near-synonymous doublets resolved by functional specialisation rather than by elimination; etymological selectivity and stylistic marking act as secondary constraints. The emergent English-mediated formatives *kiber-*, *onlayn-*, *veb-* and *smart-* are identified as the principal growth zone of the system.

Keywords: Uzbek language, prefixation, morphological productivity, affix rivalry, hybrid formation, neologism, terminology, corpus linguistics, hapax legomena, language contact.

Introduction

Descriptions of Uzbek word formation list the prefixal formatives of the language as a single undifferentiated set: *be-*, *no-*, *ser-*, *ham-*, *bad-*, *bo-*, *ba-*, *kam-*, *xush-*, *g'ayri-*, *sar-*, *dil-*, together with the international *anti-*, *super-*, *mikro-*, *post-*, *trans-* and others. Such lists are accurate as inventories but misleading as descriptions of the living language, because the members differ radically in a property that no inventory records: whether they can still be used to make new words.

The difference is easy to demonstrate. The formative *ser-* combines freely with bases of any origin and yields forms that speakers accept without hesitation even when these are not recorded in dictionaries. The formative *bo-* appears in a respectable number of dictionary entries — *bobarakot*, *boadab*, *boaql* — yet produces nothing new, and a coinage of the type



*bosuv is simply rejected. Both are listed as prefixes of Uzbek; only one is a resource of the modern language.

The question of productivity is therefore not a refinement of the inventory but a precondition for its correct interpretation. It is also a practical question. Terminological work in Uzbek — the coining of equivalents in information technology, ecology, law and economics — requires knowledge of which derivational patterns are available, and the recommendation of a pattern that is in fact fossilised produces terms that fail to enter use.

The aim of this article is to establish the relative productivity of Uzbek prefixal formatives, to propose a methodology suited to a language for which large annotated corpora are not yet available, and to identify the factors that constrain productivity in this part of the system.

2. LITERATURE REVIEW

The theoretical framework is supplied by the general literature on morphological productivity. L. Bauer [11] distinguishes the availability of a pattern for new coinage from its profitability, that is, the extent to which it is actually exploited, and stresses that a large existing type inventory is compatible with complete unavailability. R. H. Baayen [12] develops quantitative measures based on the proportion of hapax legomena in a corpus, the underlying reasoning being that a productive pattern generates a long tail of forms occurring only once. I. Plag [13] applies this apparatus to English derivation and demonstrates the divergence between type-based and hapax-based rankings. G. Booij [14] provides the general morphological framework. For Uzbek, the descriptive basis is the systematic account of word formation by A. Hojiyev [1] and his morpheme dictionary [2], together with the morpheme dictionary of A. G'ulomov, A. N. Tixonov and R. Qo'ng'urov [3]; further descriptions are given by Sh. Rahmatullayev [4] and Q. Sapayev [5]. These works record the inventory and note that some formatives are more widely represented than others, but they do not apply an explicit productivity criterion, and consequently do not distinguish a widely attested but fossilised formative from a productive one.

The contact-linguistic literature supplies the diagnostic that is missing. F. Seifart [8] shows that a borrowed affix may be present in a recipient language merely as a component of ready-made complex loanwords, or may have been extracted and become independently productive, and that the two situations are distinguished by the presence or absence of hybrid formations, that is, of derivatives combining the borrowed affix with inherited native bases. His cross-linguistic survey [9] and the volume edited by F. Gardani, P. Arkadiev and N. Amiridze [10] establish the general patterning of such borrowing, and Y. Matras [15] situates it within a theory of contact-induced change. Applied to Uzbek, this diagnostic converts an impressionistic judgement into an observable property.

The gap addressed here is thus twofold: the Uzbek tradition lists formatives without assessing availability, and the productivity literature offers measures that presuppose corpus resources Uzbek does not yet possess. The article attempts to bridge both by constructing an ordinal procedure that is applicable now and replaceable by metric measures later.



3. MATERIALS AND METHODS

The dictionary material was drawn from the five-volume explanatory dictionary of Uzbek [16], and segmentation was verified against the morpheme dictionaries [2; 3]. Post-2000 terminological and neological material was collected separately from specialist glossaries and from contemporary usage in the press and in official documents.

Four indicators were applied to each formative. The first, type count, is the number of distinct lexemes recorded in the dictionary sources; it is retained because it is easily observable, but it is treated as the weakest of the four for the reasons given in Section 4.3. The second, hybrid formation, records whether the formative attaches to inherited Turkic bases; following Seifart [8], this is treated as the strongest single indicator of full integration and hence of availability. The third, terminological attestation, records whether the formative appears in vocabulary coined since approximately 2000. The fourth, nonce acceptability, records whether speakers accept unrecorded formations built on the pattern.

The indicators are combined ordinally rather than metrically. No weights are assigned and no composite score is computed; formatives are grouped into tiers on the basis of their profile across the four indicators, with hybrid formation and nonce acceptability treated as decisive where indicators conflict. The resulting scale is an ordered classification, not a measurement, and is offered as a hypothesis testable against corpus data.

The limits of the procedure should be stated. Judgements of nonce acceptability rest on normative usage and on the absence of rejection in the sources rather than on elicitation from a controlled sample of speakers; a psycholinguistic study would place them on a firmer footing. Dictionary type counts under-represent recent formations, since lexicographic codification lags behind usage, and this bias systematically disadvantages the newest formatives.

4. RESULTS AND DISCUSSION

4.1. The four-tier productivity scale

Application of the four indicators yields the classification presented in Table 1.

Table 1. Productivity tiers of Uzbek prefixal formatives

Tier	Formatives	Hybrids on Turkic bases	Diagnostic behaviour
I. Fully productive	be-, no-, ser-, ham-, anti-, mikro-, eko-, super-	regular, numerous	Freely form nonce words; well attested in post-2000 terminology; accept bases of any origin without etymological restriction.
II. Moderately productive	g'ayri-, kam-, bad-, xush-, post-, trans-, ultra-, makro-, re-, de-	sporadic	Form new words within delimited semantic or terminological domains; nonce formation possible but stylistically marked.
III. Residual / fossilised	bo-, ba-, dil-, sar-, bar-, lo-, bil-	absent	Attested only in a closed set of inherited lexemes (bobarakot, badavlat, dilkash, sarbaland, barvaqt, loqayd, bilvosita); no new formations; nonce coinages rejected.
IV. Emergent	kiber-, onlayn-, veb-, smart-, multi-, mega-, bio-	increasing	Expanding rapidly under English influence; under-represented in dictionaries; orthography not yet stable.

Tier I comprises two groups that are equally productive for different reasons. The Persian-derived *be-*, *no-*, *ser-* and *ham-* are productive because they cover the four most frequently required modificational meanings — privation, negation of a property, abundance and sharedness — and because they have been fully nativised, as their hybrid formations show. The international *anti-*, *mikro-*, *eko-* and *super-* are productive because terminological need in the corresponding domains is high and continuous. Since the two groups are driven by different pressures, their future trajectories are unlikely to coincide: the first group is stable, the second expanding.

Tier II consists of formatives whose availability is real but domain-bounded. *G'ayri-* is productive within legal and scholarly terminology and almost nowhere else; *bad-* is productive in the evaluative and expressive register but not in neutral prose; *post-* and *trans-* are productive in the vocabulary of the social sciences. The bounding is semantic and stylistic rather than structural, and a formative may move between Tier II and Tier I as its domain expands.

Tier IV requires separate treatment and is discussed in Section 4.4.

4.2. Hybrid formation as the decisive diagnostic

The indicator that separates the tiers most sharply is the formation of derivatives on inherited Turkic bases. Table 2 sets out the relevant evidence for the Persian-derived formatives.

Table 2. Hybrid formations on inherited Turkic bases

Formative	Hybrid derivatives	Interpretation
<i>ser-</i>	<i>sersuv</i> ‘well-watered’, <i>serquyosh</i> ‘sunny’, <i>serqirra</i> ‘many-sided’, <i>serqatnov</i> ‘heavily trafficked’, <i>seryog</i> ‘in ‘rainy’	Fully extracted and nativised; combines without etymological restriction.
<i>be-</i>	<i>beto</i> ‘xtov’ ‘incessant’, <i>beo</i> ‘lchov’ ‘immeasurable’, <i>bearmon</i> ‘without regret’	Fully extracted; the privative operator applies to native nominal bases.
<i>no-</i>	<i>noo</i> ‘rin’ ‘inappropriate’, <i>notinch</i> ‘restless, troubled’	Extracted, though the pool of admissible property-denoting native bases is smaller.
<i>ham-</i>	<i>hamyurt</i> ‘compatriot’, <i>hamqishloq</i> ‘fellow villager’, <i>hamso</i> ‘z’ ‘of one word, in agreement’	Fully extracted; competes directly with the inherited suffix <i>-dosh</i> .
<i>bo-</i> , <i>ba-</i> , <i>sar-</i> , <i>dil-</i>	none attested	Not extracted; survive only inside inherited Persian collocations. Their dictionary presence is lexical, not derivational.

The contrast in the last row is the crux of the argument. *Bo-* and *ba-* are recorded in a comparable or greater number of dictionary entries than *ham-*, yet they generate nothing new. A ranking based on type count alone would place them in the middle of the scale; the hybrid criterion places them correctly at the bottom. In Seifart’s terms [8], *ser-*, *be-*, *no-* and *ham-* have completed the passage from indirect borrowing — arrival inside ready-made complex loanwords — to independent productivity in the recipient system, whereas *bo-*, *ba-*, *sar-* and

dil- never made that passage and remain, synchronically, parts of unanalysable lexemes rather than derivational units.

This result has a methodological consequence of general interest. In languages that lack the corpus resources required for hapax-based measures, the presence of hybrid formations offers a qualitative substitute that captures precisely what the hapax measure is designed to capture — the extension of a pattern beyond the set of forms in which it was first encountered.

4.3. Rivalry with inherited suffixes as the principal constraint

The most important factor limiting the productivity of Uzbek prefixation is not structural but competitive. Three inherited and fully productive suffixes occupy semantic territory adjacent to that of the productive prefixes: the caritive -siz, the possessive -li and the sociative -dosh. Every prefixal coinage must therefore justify itself against an available suffixal alternative. Table 3 sets out the resulting doublets.

Table 3. Prefixal–suffixal doublets and the outcome of rivalry

Base	Prefixal form	Suffixal form	Outcome of the rivalry
vafo ‘loyalty’	bevafo	vafosiz	Prefixal form is the default and the base of the abstract noun bevafolik; suffixal form survives as an occasional variant.
g’am ‘sorrow’	beg’am	g’amsiz	Semantic split: beg’am has drifted to ‘carefree, irresponsible’, g’amsiz remains compositional.
hosil ‘yield’	serhosil	hosilli	Terminological split: only serhosil functions as an agronomic term; ser- specifies abundance, -li mere possession.
suv ‘water’	sersuv	suvli	Semantic split: sersuv is gradable and evaluative, suvli neutral and non-gradable.
kasb ‘profession’	hamkasb	kasbdosh	Stylistic split: hamkasb neutral and dominant, kasbdosh elevated and less frequent.
fikr ‘thought’	hamfikr	fikrdosh	Stylistic and semantic split: fikrdosh implies a stronger ideological alliance.
yurt ‘homeland’	hamyurt	yurtdosh	Stylistic split: yurtdosh preferred in publicistic and patriotic registers.
qonuniy ‘lawful’	g’ayriqonuniy	qonunsiz	Terminological split: g’ayriqonuniy is the codified legal term, qonunsiz the general-language equivalent.

Three regularities emerge, and together they explain why a wholly borrowed subsystem has survived centuries of competition with an inherited rival. The prefixal member tends to be the more lexicalised, having drifted further from compositional meaning. It tends to be the more terminologised, being preferred in codified specialist vocabulary. And where the two remain semantically identical, they are differentiated stylistically, the ham- forms being neutral and the -dosh forms elevated.

The rivalry is thus resolved not by the elimination of one member but by functional specialisation. This resolution has a double effect on productivity. On the one hand it guarantees the survival of the prefixal pattern, since each member has acquired a niche. On the other it caps its expansion, since a new prefixal coinage is worth making only where the suffixal alternative would not serve — which is why Uzbek prefixation has remained a modest inventory despite a millennium of contact.

Two secondary constraints operate alongside. Etymological selectivity, the tendency of Persian formatives to prefer Persian and Arabic bases, is weakening, as the hybrid formations of Table 2 show, but is still detectable and still restricts the pool of admissible bases. Stylistic marking restricts the prefixal patterns of the older stratum to evaluative and expressive registers, and correspondingly excludes them from neutral technical prose. The international formatives are exempt from both constraints, which is the structural reason for their faster expansion.

4.4. The emergent stratum

Tier IV represents the principal zone of change in the Uzbek derivational system in the present century. The formatives *kiber-*, *onlayn-*, *veb-*, *smart-* and *multi-* differ from the earlier international layer in their route of transmission: they enter through English directly rather than through Russian, which distinguishes them from the Soviet-era vocabulary and gives them a different phonological shape.

Their status is genuinely unsettled. Orthographic practice fluctuates between solid, hyphenated and separate spellings; the same element may behave as a bound formative in one collocation and as an independent modifier in another; and dictionary codification lags well behind usage, which means that a type count would seriously understate their weight. On the criteria applied here they occupy an intermediate position, but their direction of travel is unambiguous, and the appearance of formations combining them with Uzbek bases indicates that extraction is under way.

A prediction follows that can be tested against future data. If the account given in Section 4.2 is correct, these formatives will be assigned to Tier I as hybrid formations on native bases accumulate, and their orthographic treatment will stabilise on solid spelling at approximately the same time. The correlation between the two developments is not accidental, since both reflect the same underlying change in morphological status.

4.5. Methodological limits and the corpus desideratum

The scale proposed here is ordinal and not metric, and its limitations should not be understated. It ranks formatives relative to one another but does not quantify the distance between tiers, and it cannot detect gradual change in the availability of a single formative over time.

A metric account is possible in principle and would require a lemmatised, balanced and diachronically stratified corpus of modern Uzbek. Given such a corpus, Baayen's productivity index [12] could be computed for each formative as the ratio of hapax legomena containing it to the total token frequency of the pattern, and the ordinal tiers proposed here could be replaced by continuous values. The construction of such a corpus is the most useful single contribution that could be made to the study of Uzbek derivation, and the present scale is offered as a hypothesis to be tested against it rather than as a substitute for it.



Two further refinements would be valuable. Elicitation experiments testing the acceptability of nonce formations with each formative would place the fourth indicator on an empirical footing. And a diachronic study comparing the prefixal inventory of nineteenth-century and twentieth-century Uzbek texts would establish the rate at which Tier III formatives lost availability, which would in turn allow the current trajectory of Tier IV to be projected with greater confidence.

5. CONCLUSION

The productivity of Uzbek prefixal formatives is graded rather than uniform, and falls into four tiers: the fully productive *be-*, *no-*, *ser-*, *ham-*, *anti-*, *mikro-*, *eko-* and *super-*; the domain-bounded *g'ayri-*, *kam-*, *bad-*, *xush-*, *post-*, *trans-* and others; the fossilised *bo-*, *ba-*, *dil-*, *sar-*, *bar-*, *lo-* and *bil-*; and the emergent English-mediated *kiber-*, *onlayn-*, *veb-* and *smart-*.

Type frequency is an unreliable index of productivity for this material, since it misclassifies the fossilised group. The formation of hybrids on inherited Turkic bases is the most reliable single diagnostic available and functions as a qualitative substitute for the hapax-based measures that require corpus resources Uzbek does not yet possess.

The principal constraint on productivity is rivalry with the inherited suffixes *-siz*, *-li* and *-dosh*. Where prefixal and suffixal derivatives coexist on the same base they are resolved by functional specialisation — semantic, terminological or stylistic — rather than by elimination, and this simultaneously guarantees the survival of the prefixal pattern and caps its expansion. Etymological selectivity and stylistic marking act as secondary constraints, and the international formatives, being exempt from both, expand fastest.

The emergent stratum is the zone in which the system is currently changing, and its eventual assignment to the fully productive tier is predicted to coincide with the stabilisation of its orthography, since both reflect the same change in morphological status.

The construction of a lemmatised and balanced corpus of modern Uzbek would allow the ordinal scale proposed here to be replaced by metric productivity measures, and a comparison with Kazakh, Kyrgyz and Uyghur would establish whether the pattern of graded productivity described here is specific to Uzbek or general to the Turkic languages of the region.

REFERENCES

1. Hojiyev, A. (2007). O'zbek tili so'z yasalishi tizimi [The word-formation system of the Uzbek language]. Toshkent: O'qituvchi. 167 p.
2. Hojiyev, A. (1992). O'zbek tilining morfem lug'ati [Morpheme dictionary of the Uzbek language]. Toshkent: O'qituvchi. 225 p.
3. G'ulomov, A., Tixonov, A. N., & Qo'ng'urov, R. (1977). O'zbek tili morfem lug'ati [Morpheme dictionary of Uzbek]. Toshkent: O'qituvchi.
4. Rahmatullayev, Sh. (2006). Hozirgi adabiy o'zbek tili [Modern literary Uzbek]. Toshkent: Universitet.
5. Sapayev, Q. (2009). Hozirgi o'zbek tili: morfemika, so'z yasalishi va morfologiya. Toshkent.
6. Кононов, А. Н. (1960). Грамматика современного узбекского литературного языка. Москва–Ленинград: Издательство АН СССР.
7. Johanson, L., & Csató, É. Á. (Eds.). (2022). The Turkic Languages (2nd ed.). London & New York: Routledge.



8. Seifart, F. (2015). Direct and indirect affix borrowing. *Language*, 91(3), 511–532. <https://doi.org/10.1353/lan.2015.0044>
9. Seifart, F. (2017). Patterns of affix borrowing in a sample of 100 languages. *Journal of Historical Linguistics*, 7(3), 389–431.
10. Gardani, F., Arkadiev, P., & Amiridze, N. (Eds.). (2015). *Borrowed Morphology*. Berlin & Boston: De Gruyter Mouton.
11. Bauer, L. (2001). *Morphological Productivity*. Cambridge: Cambridge University Press.
12. Baayen, R. H. (2009). Corpus linguistics in morphology: Morphological productivity. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook* (Vol. 2, pp. 899–919). Berlin: Mouton de Gruyter.
13. Plag, I. (2018). *Word-Formation in English* (2nd ed.). Cambridge: Cambridge University Press.
14. Booij, G. (2012). *The Grammar of Words: An Introduction to Linguistic Morphology* (3rd ed.). Oxford: Oxford University Press.
15. Matras, Y. (2009). *Language Contact*. Cambridge: Cambridge University Press.
16. O‘zbek tilining izohli lug‘ati [Explanatory dictionary of the Uzbek language] (2006–2008). 5 vols. Toshkent: O‘zbekiston milliy ensiklopediyasi.
17. Haspelmath, M., & Tadmor, U. (Eds.). (2009). *Loanwords in the World’s Languages: A Comparative Handbook*. Berlin & New York: De Gruyter Mouton.
18. Bodrogligeti, A. J. E. (2003). *An Academic Reference Grammar of Modern Literary Uzbek*. Munich: LINCOM Europa.

