

THE INFLUENCE OF PREFIXATION ERRORS IN TEXTS ON SEMANTIC CHANGES

Narzullayeva Muborak Sherzod qizi

Independent Researcher, Gulistan State University, Uzbekistan

E-mail: narzullayevamuboraksherzodovna@gmail.com

Аннотация

The article analyses prefixation errors in Uzbek texts and the semantic changes they produce. Prefixation in Uzbek is confined to a small inventory of elements of Persian, Arabic and international origin, and the meaning of a prefixed word is normally compositional, which makes every deviation in the choice or the segmentation of a prefix a direct source of semantic distortion. Six types of error are distinguished: the choice of a wrong prefix from a synonymous set; pleonastic marking, in which a prefix and the suffix -siz express the same negation twice; the substitution of a prefixal formation for a suffixal one where the two are not interchangeable; false segmentation of stems that only appear to be prefixed; orthographic errors at the prefix boundary; and mechanical calquing of foreign prefixes in translation. The consequences are described at four levels: denotative, modal and evaluative, scopal, and terminological. Legal and official texts are examined as the register in which the terminological consequence is most costly, with the pair *noqonuniy* and *g'ayriqonuniy* as the central example. A separate section treats the effect of false segmentation on automatic morphological analysis, where stripping *be-* from *bemor* destroys the stem and every operation built on it. The article concludes with an editorial procedure for prevention.

Keywords: Prefixation errors, semantic change, Uzbek language, error typology, false segmentation, pleonasm, legal terminology, translation, automatic morphological analysis, editing.

Introduction

The meaning of a prefixed word in Uzbek is in most cases compositional. The prefix contributes a constant semantic operation and the stem contributes the content on which that operation is performed, so that *beg'am* is derived from *g'am*

for sorrow by the operation of privation and noaniq is derived from aniq for precise by the operation of negation [1]. Compositionality has a consequence that is central to the present study. Where meaning is computed from parts, an error in a part is not a stylistic blemish but a change of meaning. A misplaced prefix does not make a text clumsy; it makes the text say something else.

Error analysis has established that errors are systematic and that their systematicity is the reason they can be described and prevented [2]. Prefixation errors in Uzbek are systematic in exactly this sense. They arise from a small number of structural conditions: the inventory of prefixes contains near-synonyms whose distribution is conventional rather than logical; the language possesses native suffixes that express the same meanings and therefore invite substitution; and a substantial group of loanwords begins with a syllable identical to a prefix without containing one. Each of these conditions produces its own class of error and its own class of semantic consequence.

The aim of the article is to establish a typology of prefixation errors in Uzbek texts and to determine the semantic changes each type produces. The tasks are to describe the structural conditions that generate the errors; to classify the errors by their formal mechanism; to classify the resulting semantic changes by level; to analyse the consequences in the register where they are most costly, that is, in legal and official texts; to describe the effect of one error type on automatic language processing; and to formulate a procedure of prevention. The material consists of prefixed formations of the modern literary language, of variant usages observed in the press and in official documents, and of the segmentation problems posed by the loanword stratum.

2. Structural conditions that generate the errors

The first condition is the presence of near-synonymous prefixes. Negation and privation in Uzbek are expressed by be-, no- and g'ayri-, and in the international layer by anti-, de- and dis-. These elements are not distributed by a semantic rule that a speaker can apply, but by lexical convention established for each stem. There is no general principle from which it follows that qonuniy for lawful combines with no- and with g'ayri- while huda combines only with be-. The speaker must know the item, and where knowledge fails, substitution follows [3]. The second condition is the competition with the native suffixes -siz, -li, -dor and -dosh, described in the literature on Uzbek word formation as a source of doublets

such as besuv and suvsiz or hamkasb and kasbdosh. Because the members of many pairs are close in meaning, the speaker assumes free interchangeability. The assumption is correct for part of the pairs and false for the rest, and the error appears at exactly the points where the two members have diverged.

The third condition is the presence of stems that begin with a prefix-like syllable without containing a prefix. Uzbek has a considerable number of such items: bemor for an ill person, nozik for delicate, notiq for orator, nomzod for candidate, nonushta for breakfast, antiqa for antique, sertifikat for certificate, bezak for ornament and beshik for cradle [4]. Any operation that segments by surface form, whether performed by a language learner, an editor or a program, is exposed to false analysis here.

The fourth condition is the unsettled orthography of the international layer. The elements anti-, post-, mikro-, veb- and kiber- entered recently, and normative decisions on solid, hyphenated and separate spelling have not covered every case, so variation is present in published texts. The fifth condition is translation. English, Russian and Uzbek do not map their negative and privative affixes onto one another one to one, and mechanical correspondence produces formations that are well formed but wrong.

3. A typology of prefixation errors

Type one is the choice of a wrong prefix from a synonymous set. The speaker selects an element with the correct general meaning but the wrong lexical distribution. Formations of this type are transparent and are usually understood, which is precisely why they pass unnoticed into edited text. The substitution of no- for be- in a stem that conventionally takes be-, or of anti- for the established Uzbek privative in a term that already exists, belongs here.

Type two is pleonastic marking, in which the same negation is expressed twice. The prefix be- or no- is combined with the privative suffix -siz on one stem, so that the word contains two exponents of one operation. Logically such a form should yield a double negation and therefore an affirmative reading, but in practice it is read as a single reinforced negation, which is the reason the error is stable and reproduces itself. A related subtype combines a prefixal negation with the analytical negation emas, as when a writer negates noaniq once more instead of using the positive aniq.

Type three is substitution between the prefixal and the suffixal member of a doublet where the two have diverged. This is the type with the highest cost in specialist registers, because the divergence in such pairs is normally terminological. The pair *noqonuniy* and *g'ayriqonuniy* is examined separately in section five.

Type four is false segmentation. A stem that contains no prefix is analysed as if it did, and the analysis is then acted upon: a supposed positive counterpart is constructed, a supposed stem is used in isolation, or a supposed antonym is coined. From *bemor* one may construct a non-existent stem; from *notiq* one may construct a supposed positive; from *antiqa* one may derive a supposed non-prefixed form. The error is rare in the speech of native speakers with a wide vocabulary and frequent in the writing of learners, in hasty journalism, and in automatic processing [5].

Type five is orthographic error at the prefix boundary: separate writing of a solid formation, omission or insertion of a hyphen, and inconsistent treatment of the same element within one text. The formal consequence is that the reader must decide whether a syntactic group or a single word is intended, and this decision is not always recoverable from the context.

Type six is calquing in translation. A foreign prefix is rendered by a fixed Uzbek equivalent regardless of the item. English *non-*, *un-*, *in-*, *dis-* and *anti-* and Russian *ne-*, *bez-*, *anti-* and *protivo-* are mapped onto *no-* or *be-* mechanically, and the result is a well-formed Uzbek word that either does not exist in usage or exists with a different meaning.

4. Levels of semantic change produced by the errors

The first level is denotative. The error changes the class of objects to which the word refers. This is the direct effect of type three and of part of type one, and it is the level at which the damage is easiest to demonstrate, because two different established words with two different established referents are involved.

The second level is modal and evaluative. Uzbek prefixation includes an evaluative class, *bad-* and *xush-*, and a privative class that is logically neutral. The replacement of an evaluative prefix by a neutral one removes the speaker's assessment from the text, and the replacement in the opposite direction inserts an assessment that the author did not make. A description that used *xush-* and is edited into a privative formation loses its positive evaluation without any visible

change in the propositional content. This level is characteristic of publicistic and literary registers, where the evaluative charge is often the point of the sentence [6].

The third level is scopal. Prefixal negation applies to the property expressed by the stem, whereas the analytical construction with *emas* negates the predication, and the two differ in what falls inside the negation. In a sentence with a complex predicate the difference is detectable: the prefixal form ascribes a negative property to the subject, the analytical form denies that a property holds. The pleonastic forms of type two are a scopal problem as well, since a form containing two exponents of negation is formally ambiguous between a reinforced negative and a cancelled one, and only convention resolves it.

The fourth level is terminological. Here the change is not in the reference of a word in a sentence but in the identity of a term in a system. A term is defined by its place in a set of related terms, and the substitution of one prefix for another moves the item to a different place or creates an item with no place at all. The consequence is that the text ceases to be usable for the purpose for which terms exist, which is the unambiguous identification of a notion across documents.

The four levels are not mutually exclusive, and a single error frequently produces effects at more than one of them. A substitution inside a doublet in a statutory text changes the denotation of a word in the sentence and at the same time changes the identity of a term in the system, and the second effect outlasts the first, because the sentence is read once while the term is used in every document that refers to the instrument. The ordering of the levels adopted here is therefore an ordering by durability of the consequence rather than by the severity of the individual case: a denotative change is confined to the text in which it occurs, whereas a terminological change propagates through everything built on that text.

5. The terminological consequence in legal and official texts

The clearest case in Uzbek is the pair *noqonuniy* and *g'ayriqonuniy*. Both are formed from *qonuniy* for lawful, both are rendered into English as unlawful or illegal, and both are current in the language of legislation and of official documents. They are nevertheless not interchangeable in every context, because each has become fixed in particular collocations and in particular normative formulations. The prefix *no-* is of Persian origin and belongs to the older and more integrated layer; *g'ayri-* is of Arabic origin and belongs to the bookish

register, and in official usage it appears in formulations of a more abstract and classificatory character.

The practical consequence is that a substitution inside this pair inside a legal text is not a stylistic edit. Where a normative act uses one of the two forms in a definition, that form becomes the designation of the defined notion, and every subsequent occurrence must repeat it. An editor who introduces the other form creates the appearance of two notions where the drafter intended one, or conversely merges two notions that the drafter distinguished. The reader of the text, who is required to apply it, has no means of establishing which of the two situations obtains, because both forms are well formed and both are attested in the same register.

The same mechanism operates in other official pairs. The negation of qonuniy is not the only case in which Uzbek offers a prefixal and an analytical option: qonunga xilof for contrary to the law is a third designation of the same general area, and it is a syntactic group rather than a word. The three means are distributed across normative texts by convention, and the convention is what the drafter must follow. The general rule that follows for editorial practice is simple to state: in a normative text the prefixed term is not a word to be improved but a designation to be reproduced [7].

The consequences of type two errors in this register are of a different order. A pleonastic double negation in a statutory provision produces a formally ambiguous sentence, and in a document that will be applied by officials and courts, formal ambiguity is a defect that can be raised in argument. The consequences of type five, the orthographic errors, are less severe but not negligible, since inconsistent spelling of a term within one document weakens the identity of the term and complicates its retrieval in electronic legal databases.

A separate observation concerns the interaction of prefixation errors with abbreviation, which is frequent in official Uzbek. Where a designation containing a prefixed term is abbreviated, the prefix is normally represented by its initial letter, and the resulting abbreviation is opaque with respect to the prefix that produced it. If the underlying term is later written with a different prefix, the abbreviation remains unchanged and the connection between the two is lost for the reader. The abbreviation then survives as the only stable designation, while the full form varies from document to document, which is the reverse of the relation that terminological practice requires.

6. Errors in translation into and out of Uzbek

Translation is the largest single source of prefixation errors in contemporary Uzbek texts, because the volume of translated material in administration, technology and journalism is large and because the affixal systems of the languages involved do not correspond. English distinguishes non-, un-, in- and dis-, and the choice among them is lexically determined for each stem. Russian distinguishes ne-, bez-, anti- and protivno-. Uzbek offers be-, no-, g'ayri-, the suffix -siz, and the analytical construction with emas. A translator working under time pressure adopts a default correspondence, and the default then applies to items for which it is wrong [8].

Three patterns are typical. The first is over-generalisation of no- as the equivalent of English non- and un-. The pattern gives correct results in a large number of cases, which is why it is adopted, and incorrect results wherever the Uzbek item conventionally takes be- or a suffixal formation. The second is the transfer of the international prefix instead of the existing Uzbek term, so that a text uses a formation with anti- where Uzbek already has an established designation with a native or a Persian element. The result is not incomprehensible, but it introduces a second designation for a notion that already had one, which is a terminological loss.

The third pattern is the failure to render the evaluative component. English and Russian frequently express evaluation lexically, whereas Uzbek can express it affixally with bad- and xush-, and a translation that ignores this possibility produces a neutral text where the original was evaluative. The reverse also occurs, and it is more damaging: an Uzbek evaluative prefix rendered by a neutral foreign equivalent deletes an assessment that the source text made. Translation quality assessment normally treats such deletions as shifts of a kind that must be recorded, and the affixal nature of the carrier makes them easy to overlook [9].

The register in which type six errors accumulate fastest is the language of the press and of institutional communication, because both work with translated source material under time pressure and both produce texts that are then quoted. A prefixed formation that appears in a widely reproduced news item acquires currency independently of whether it was well chosen, and once it is current it begins to compete with the established designation. The mechanism is ordinary lexical diffusion, but its starting point in these cases is an error, and the result is a doublet whose members differ in nothing except their origin. Terminological

work therefore has to intervene early, at the point of translation, since intervention after diffusion requires the removal of a form that speakers already use.

7. The effect of false segmentation on automatic processing

Type four errors have a consequence outside human editing that deserves separate treatment. Automatic morphological analysis of Uzbek is normally organised as stem identification plus affix stripping, since the agglutinative structure of the language makes this approach efficient for suffixes. If the analyser also strips prefixes by surface form, every item of the false analysability group is corrupted. Stripping *be-* from *bemor* leaves a sequence that is not a stem of Uzbek. The analyser then either fails or, worse, assigns the item to a wrong lemma, and every downstream operation inherits the error: search retrieves wrong documents, frequency counts are distorted, and machine translation produces an unrelated equivalent [10].

The scale of the problem is determined by the size of the false analysability group, and that group is not small in Uzbek, because the Persian and Arabic strata contributed many words whose initial syllable coincides with a prefix. The solution is not statistical but lexical. The analyser requires an explicit exception list of stems that must never be segmented at the left edge, and that list can only be compiled by etymological and lexicographic work of the kind described in the first two articles of this series. This is a case in which historical knowledge has a direct engineering application.

A second processing consequence follows from type five. Orthographic variation at the prefix boundary multiplies the surface forms of one lexical item, so that a term written solid in one document and hyphenated in another counts as two types rather than one. In corpus work this depresses the frequency of the term and can remove it from a frequency-based term extraction list altogether. Normalisation before analysis is therefore necessary, and the normalisation rules must be built from the same lexical material [11].

8. Prefixation errors in teaching and in learner writing

The error types described above are distributed unevenly across writers. Types one, three and six predominate in the work of professional writers and translators, who know the inventory but not the lexical distribution of its elements. Types two and four predominate in the writing of learners, both of learners of Uzbek and of

Uzbek speakers writing in English or Russian, because these types arise from the over-application of a rule that the learner has correctly identified but incorrectly generalised.

Three contrastive illustrations show what is at stake. The illustrations are constructed for the purpose of exposition and are not corpus data. In the first, a text describes a territory as *besuv* and an editor replaces the word with *suvsiz*. Both forms are well formed and both are translated as waterless, but the prefixal form is normally used of a permanent characteristic of the territory and the suffixal form of a situational absence of water, so that the replacement converts a geographical property into a temporary condition. The direction of the difference is presented here as probable, since it rests on collocational observation rather than on a systematic register study.

In the second illustration, a description uses *badbo'y* for foul-smelling and a translator renders it with a neutral expression for having a smell. The propositional content survives and the evaluation does not. The reader of the target text learns that the object has a smell but not that the author condemned it, and no element of the target text signals the loss. In the third illustration, a writer replaces *noaniq* with the analytical *aniq emas* inside a complex predicate. The prefixal form ascribes the property of imprecision to the subject; the analytical form denies that the property of precision holds. In a simple sentence the two are equivalent for practical purposes, and in a sentence with modal or quantified elements they are not, because the negation takes a different scope.

For teaching, three conclusions follow. First, the productive prefixes must be presented together with their lexical distribution and not as free operations, because the learner who is told that *no-* negates will produce type one errors at every point where usage requires *be-*. Second, the exception list of falsely analysable stems must be taught explicitly and early, since *bemor*, *notiq*, *nozik*, *nomzod*, *nonushta* and *antiqa* are frequent words that a learner will encounter long before acquiring the etymological knowledge needed to segment them correctly. Third, the doublets with *-siz*, *-li*, *-dor* and *-dosh* should be presented in pairs with their register values, since the assumption of free interchangeability is the source of type three errors and that assumption is created by the way the two patterns are usually taught in isolation from each other [7].

A fourth conclusion for teaching concerns the order of presentation. Because the prefixal formations of Uzbek belong to the loanword stratum and cluster in

written registers, they are usually introduced late in a course, after the productive native suffixes are established. The falsely analysable items, however, are frequent words of the everyday vocabulary and are introduced early. The learner therefore meets *bemor*, *nozik* and *nonushta* long before meeting the prefix that their initial syllables resemble, and when the prefix is finally introduced it is applied retroactively to material already in the learner's memory. Reversing this order is not possible, since the frequency of the words is what determines their place in a course. The alternative is an explicit warning at the point where the prefix is introduced, listing the frequent items that must not be segmented.

9. Prevention: an editorial procedure

The typology yields a procedure with five checks, ordered by the ease with which each can be performed. The first check is lexical verification of every prefixed item against a dictionary. This eliminates types one and four at once, because both consist in the use of a form that either does not exist or exists with a different meaning, and both are detectable without reference to context.

The second check is the search for double exponence. Every item containing *be-* or *no-* is examined for the suffix *-siz* and for the analytical *emas* in the same predicate. This eliminates type two mechanically and can be automated with a simple pattern.

The third check is terminological consistency. In legal, technical and official texts every prefixed term is compared with its form in the source instrument, and any variation is treated as an error to be corrected rather than as a stylistic choice to be preserved. This addresses the most costly consequences, described in section five.

The fourth check is the review of translated passages against the source, with attention to the three patterns described in section six, and in particular to the evaluative prefixes, whose loss is invisible in the target text alone. The fifth check is orthographic normalisation of the international elements throughout the document, so that one element has one spelling in one text. The order of the checks matters for efficiency: the first two are formal and fast, the last three require judgement and access to external material.

The procedure presupposes a resource that Uzbek does not yet possess in a usable form, namely a list of prefixed formations of the literary language with their stems, their productivity status and their register values, together with the

exception list of falsely analysable items. The individual components of such a resource exist: explanatory dictionaries record the formations, etymological dictionaries record the origins, and terminological standards record part of the official usage. What is missing is their integration into a single machine-readable list. Until it exists, the first and the third checks of the procedure depend on the personal knowledge of the editor, which is why prefixation errors of types one and three are the ones that most often survive editing in practice.

The procedure is deliberately formulated for the editing stage rather than for the drafting stage. Prefix choice at the moment of writing is governed by the writer's lexical knowledge, and that knowledge cannot be supplied by a rule, since the distribution of *be-*, *no-* and *g'ayri-* across stems is conventional. What can be supplied is a systematic check applied to the finished text, and the five checks above are ordered so that the two that require no external knowledge come first. In institutional practice this ordering matters, because the first two checks can be delegated or automated while the remaining three require a qualified editor with access to the source instruments.

10. Conclusion

Prefixation errors in Uzbek are not peripheral errors of form. Because the meaning of a prefixed word is computed from its parts, an error in the prefix changes what the text states. The six types established here differ in their mechanism but converge in this effect.

The four levels of semantic change identified in the study are ordered by the register in which each is most damaging. Denotative change affects all texts. Modal and evaluative change affects publicistic and literary texts, where the affixal evaluation carried by *bad-* and *xush-* is often the point of the utterance. Scopal change affects any text with complex predication and is aggravated by the pleonastic forms of type two, which are formally ambiguous between reinforced and cancelled negation. Terminological change affects legal, official and technical texts, and it is the most costly, because it damages the identity of a term across a system of documents rather than the meaning of one sentence.

Two findings have applications beyond editing. The first is that the pair *noqonuniy* and *g'ayriqonuniy*, and pairs of the same structure, must be treated in normative texts as fixed designations and not as stylistic variants, since a substitution inside such a pair creates or destroys a distinction that the drafter

made. The second is that the group of stems with false analysability, which includes bemor, notiq, nozik, nomzod, nonushta, antiqa and sertifikat, is the point at which prefixation errors pass from human texts into automatic processing, where a single wrong segmentation propagates through search, corpus statistics and machine translation. Both findings point to the same requirement: a codified list of Uzbek prefixed formations and of their exceptions, compiled on an etymological basis and made available in a form that editors and language technology can use.

References

1. Haspelmath M., Sims A. D. Understanding Morphology. 2nd ed. London: Hodder Education, 2010. 366 p.
2. James C. Errors in Language Learning and Use: Exploring Error Analysis. London: Longman, 1998. 304 p.
3. Hojiyev A. O'zbek tilida so'z yasalishi. Toshkent: O'qituvchi, 1989.
4. Rahmatullayev Sh. O'zbek tilining etimologik lug'ati. Toshkent: Universitet, 2000.
5. Booij G. The Grammar of Words: An Introduction to Linguistic Morphology. 3rd ed. Oxford: Oxford University Press, 2012. 353 p.
6. Plag I. Word-Formation in English. Cambridge: Cambridge University Press, 2003. 240 p.
7. Bodrogligeti A. J. E. A Grammar of Modern Literary Uzbek. Munich: LINCOM Europa, 2003. 480 p.
8. Newmark P. A Textbook of Translation. New York: Prentice Hall, 1988. 292 p.
9. Baker M. In Other Words: A Coursebook on Translation. 3rd ed. London: Routledge, 2018. 390 p.
10. Jurafsky D., Martin J. H. Speech and Language Processing. 3rd ed. Stanford, 2023.
11. Manning C. D., Schutze H. Foundations of Statistical Natural Language Processing. Cambridge (Mass.): MIT Press, 1999. 680 p.