

MEASURING SUBSTANDARD LEXIS SHIFTS IN UZBEK RUSSIAN TRANSLATION USING ALIGNED CORPORA

Makhmudova Yulduz Askarovna

Lecturer, Department of Russian Language and Literature

Bukhara State University

ORCID 0009-0002-4369-2433

Abstract. The paper measures how substandard lexis (argot, jargon, colloquial and vulgar items) changes when Tahir Malik's novel *Shaytanat* moves from Uzbek into Russian. A sentence-aligned parallel corpus of the novel and its 2016 Russian edition supplies the data. Every substandard token receives a shift label, after which the distribution across five categories, namely neutralisation, register shift, compensation, explicitation and omission, is counted. Neutralisation prevails, which supports Toury's law of growing standardisation for this low-resource pair. Criminal argot resists neutralisation more than colloquial items, because Russian offers ready equivalents inherited from a shared Soviet-era substandard stock.

Keywords: corpus-based translation studies, substandard lexis, translation shifts, neutralisation, standardisation, parallel corpus, Uzbek Russian translation, criminal argot, Tahir Malik, *Shaytanat*, low-resource language pair.

Introduction. Criminal fiction lives by the speech of its outlaws. When Tahir Malik built *Shaytanat*, a long novel about the drug trade, contract killing and organised crime in Uzbekistan during the final Soviet years and the first years of independence, he filled it with the coded talk of the underworld [12]. Asadbek, its mafia chief, gathers around him a vocabulary that no dictionary of literary Uzbek would sanction. Veiled words for murder, prison terms, coarse street nicknames, borrowings from Russian thieves' cant, all of it crowds the pages. Bukhara philologists who assembled an author corpus of the book counted these units and sorted them under dialect words, foreign and vulgar words and criminal argot [11; 14]. In 2016 a Tashkent house issued the Russian version, in the translation of T. Narmirzayev and M. Alimov [13]. Every marked word in the source then met one outcome or another in the target text, whether its coarseness survived, softened toward neutral speech, moved to another point in the sentence or vanished.

This study measures those outcomes. Within a sentence-aligned parallel corpus of the Uzbek original and the Russian translation, every substandard token receives a label according to what happened to its markedness. Five categories hold the data, namely neutralisation, register shift, compensation, explicitation and omission. Counting the labels yields a distribution, and the distribution answers a concrete question, whether the translators pulled

the coarse speech of the book toward standard Russian or kept it rough. Prior corpus work on translated language predicts the first answer. Whether a low-resource pair with a shared history of criminal slang behaves the same way is the open question, since Russian and Uzbek thieves' cant grew together under one state for seventy years, and that common stock may hand the translator ready equivalents a distant pair would lack.

Methods and Literature Review. The material was prepared in several passes. The Uzbek source and the Russian target were digitised from the printed editions and cleaned of layout noise, after which the two texts were aligned at sentence level with automatic tools of the kind used for Uzbek parallel data by N.Z. Abdurakhmonova [16] and in the Uzbek Kazakh set of B. Allaberdiev and colleagues [19], and then corrected by hand, a correction that proved unavoidable because Uzbek word order (subject object verb) and Russian word order (subject verb object) produce frequent one to many sentence mappings in literary prose. Substandard tokens were identified in the source through the classes of V.V. Khimik, who separates argot, jargon and urban vernacular, within the social differentiation of Russian described by L.P. Krysin [7; 15], and the Russian equivalents in the target were checked against the dictionary of V.M. Mokienko and T.G. Nikitina [10]. Each aligned pair was then classified by the kind of shift its markedness underwent, following the shift typology of J.C. Catford and the comparative model of K.M. van Leuven-Zwart [4; 5]. Frequencies were computed for every category, and the shares were compared across the groups of substandard vocabulary to see which groups resisted neutralisation and which yielded to it. Manual coding was performed twice on a control sample to check agreement, and the strategy labels for the target solutions were drawn from the catalogue of L. Berezowski [9]. The treatment stayed descriptive, reporting counts and proportions rather than inferential tests, because the aim was to describe one translator pair's practice on one book, not to generalise across a population of translations.

Corpus method entered translation study through M. Baker, who saw in machine-readable collections a way to watch translation behave. In her programmatic paper she held that translated texts deserve study in their own right, writing that «the starting point of this paper is that translated texts record genuine communicative events and as such are neither inferior nor superior to other communicative events in any language. They are however different, and the nature of this difference needs to be explored and recorded» [1, p. 234]. The claim opened a search for recurrent features of translated language, among them the levelling of marked forms. G. Toury gave that search its descriptive frame and its most cited generalisation, the law of growing standardisation, stated as the observation that «in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by a target repertoire» [2, p. 268]. Substandard speech, being the least habitual material a translator meets, sits where this law bites hardest. S. Laviosa turned the idea into countable measures and reported that translated language shows lower lexical variety and a narrower vocabulary than original writing, an outcome she filed under simplification and convergence [3]. Where the counting starts, though, is the single shift, and here J.C. Catford supplied the founding definition, holding that «by shift

of level we mean that a SL item at one linguistic level has a TL translation equivalent at a different level» and adding the split into level shifts and category shifts that later analysts refined [4, p. 73]. K.M. van Leuven-Zwart built the finest instrument for prose, a comparative model in which «both micro- and macrostructural shifts in translation can furnish indications of the translational norms adopted by the translator, his interpretation of the original text and the strategy applied during the process of translation» [5, p. 151]. A. Chesterman later sorted the recurrent features into two families, calling those found by comparing source with translation S universals and those found by comparing translation with native writing T universals, which places the levelling of coarse speech among the S universals [6]. On the target side the vocabulary of the coarse itself has been mapped in Russian, where V.V. Khimik set argot against jargon as two bodies of unequal reach [7], and where V.D. Devkin grounded the study of lowered vocabulary in the theory of colloquial speech, describing it as «непринуждённая, спонтанная, несколько сниженная, устная, преимущественно ситуативно обусловленная, чаще диалогическая речь» [8, p. 11]. The reference work for the target was the large dictionary of Russian jargon by V.M. Mokienko and T.G. Nikitina, which records twenty five thousand words and seven thousand set expressions and names translators among its intended users [10]. For strategies of rendering non-standard varieties the catalogue of L. Berezowski holds the ground, its neutralisation heading naming the move by which a non-standard feature returns as standard, a move that «includes the rendering of swearwords as neutral words» [9, p. 89]. Nearest to the present material stands the Bukhara work on an author corpus of Shaytanat, which showed what such a collection makes visible in the vocabulary of one novelist [11].

Results. Coding the aligned pairs brought out one pattern above the others. Marked words in the source tend to reach the reader of the Russian text with their coarseness reduced or gone, which is the levelling that descriptive translation study long predicted. As M. Baker set out the promise of watching this happen at scale, she held that «large corpora will provide theorists of translation with a unique opportunity to observe the object of their study and to explore what it is that makes it different from other objects of study, such as language in general or indeed any other kind of cultural interaction» [1, p. 235]. The book gives that observation plenty to work on. Its underworld does not talk in the words of a school primer, and the translators had to decide, again and again, how far the primer would win. Most often it won.

Killing is never named plainly by Malik's criminals. They send a man off, they make him vanish, they dispatch him to asfalasofin, the lowest pit of the other world. Bukhara researchers who traced this vocabulary read the whole coarse layer of the book against the categories of lexicology, observing that «mukammal badiiy matnni lisoniy tahlil qilish jarayonida uning tilidagi leksik birliklarni leksikologiya qonun qoidalariga monand ravishda ma'nodosh so'zlar, shakldosh so'zlar, zid ma'noli so'zlar, ko'p ma'noli so'zlar, tarixiy va arxaik so'zlar, yangi yasalmalar, shevaga oid so'zlar, chet va vulgar so'zlar ajratib olinadi va asarda qanday maqsadga yo'naltirilganligi izohlanadi» [14]. In Russian the veiled charge often thins. The set phrase that hides a murder behind the idea of travel keeps its surface in a

rendering such as *отправить*, yet *asfalasofin*, a word soaked in the geography of religious punishment, has no ready Russian twin and tends to arrive as a plain word for the beyond or to fall away. Where the source only hints, the target frequently says the thing outright.

Not every marked word loses its edge. The items that Uzbek thieves had themselves taken from Russian cross back with ease, because the target already owns them. *Gastrol*, the criminal trip to work another town, is Russian *гастроль* used in exactly the thieves' sense, so the translator sets it down unchanged and the coarse charge stays whole. The unchanged *гастроль* is the case J.C. Catford put outside the class of shifts, formal correspondence holding across the two languages, whereas a level shift, in his definition, means that «a SL item at one linguistic level has a TL translation equivalent at a different level» [4, p. 73], which is what happens when a single coarse Uzbek word is opened out into a neutral Russian phrase with an explaining clause. The same ease covers *zemlyagim*, an address form built on Russian *земляк*, which returns to *земляк* without loss. Here the shared past of the two cants does real work. Table 1 sets out the five outcomes with an illustration of each drawn from the book.

Table 1. Five outcomes of substandard items in the Russian text of *Shaytanat*

Shift category	Effect on markedness	Illustration from the novel (Uzbek into Russian)	Shift category	Effect on markedness
Neutralisation	The marked item returns as a neutral standard word and the coarse charge is lost	o'qilon, prison slang for a long term, rendered by a neutral Russian word for a sentence	Neutralisation	The marked item returns as a neutral standard word and the coarse charge is lost
Register shift	Coarseness is kept but recalibrated, a vulgarism softened to mild colloquial speech or raised	a blunt insult toned down to a milder Russian colloquial word	Register shift	Coarseness is kept but recalibrated, a vulgarism softened to mild colloquial speech or raised
Compensation	A charge lost at one word is restored elsewhere, often by an added epithet or a footnote	the insult inside G'ilay (squint) or Kesakpolvon (clod-wrestler) carried by an added phrase or a gloss	Compensation	A charge lost at one word is restored elsewhere, often by an added epithet or a footnote

Explicitation	A veiled or coded item is made overt in the target	asfalasofinga jo‘natish, to send to the lowest pit, spelled out as plain killing	Explicitation	A veiled or coded item is made overt in the target
Omission	The marked item is deleted and no target counterpart appears	a coarse tag dropped where the sentence carries the sense without it	Omission	The marked item is deleted and no target counterpart appears

Nicknames carry a rough charge that travels badly. Malik tags minor criminals by their defects and their trades, so a reader of the Uzbek meets Chittako, Kesakpolvon, G‘ilay and Chuvrindi, each name a small insult worn in place of a surname. The squint inside G‘ilay, the clod-wrestler inside Kesakpolvon, these sit on the surface of the Uzbek word and fade the moment the name is transliterated into Cyrillic without a gloss. A translator who keeps the sound loses the sting. L. Berezowski named the reverse move plainly, defining neutralisation as the return of a non-standard feature as standard, a move that «includes the rendering of swearwords as neutral words» [9, p. 89], and a footnote or an added epithet that restores the sting works against exactly that pull, buying back meaning at the price of brevity, which is compensation of one kind. Coarse common words behave otherwise again, since a blunt term for a fool or a drunkard in the source usually meets a blunt Russian counterpart, and the coarseness survives because the target vernacular holds a rich supply of such words.

The behaviour of the argot proper repays a closer look. V.V. Khimik drew the line between the two main bodies of coarse vocabulary, holding that «арго – это закрытая лексическая подсистема специальных номинаций, обслуживающих узкие социально-групповые интересы, чаще всего профессиональные. Арготизмы – рациональные номинации-терминоиды, используемые в практических интересах профессии, ремесла, дела. Жаргон – значительно более широкое понятие, полуоткрытая лексико-фразеологическая подсистема» [7]. The prison term o‘qilon, a coiled name for a long sentence behind bars, belongs to the closed body Khimik describes, and closed terms bound to one country's penal world rarely own a foreign match. Such items neutralise most readily. The half-open jargon of insult and appetite, shared across the border, survives better.

Table 2. Groups of substandard vocabulary and their predominant outcome

Group	Examples in the source	Predominant outcome in Russian	Reason
Uzbek-specific criminal argot	o‘qilon and other closed penal terms	neutralisation	closed to one penal world, no ready Russian match

Euphemisms for killing	jo‘natib yuborish, gum qilish, asfalasofinga jo‘natish	neutralisation or explication	culture-bound veiling with no Russian twin
Coarse nicknames	Chittako, Kesakpolvon, G‘ilay, Chuvrindi	transliteration with compensation	the sound is kept and the insult restored by a gloss
Russian-borrowed slang	gastrol (гастроль), zemlyagim (from земляк)	direct transfer with no shift	the target already owns the item
Urban vernacular and vulgarisms	blunt words for a fool, a drunkard, a coward	survival, coarseness kept	the target vernacular is rich in equivalents

Taken across the groups, the direction holds steady. The closed argot of the Uzbek penal world and the culture-bound euphemisms move toward neutral Russian most often, the shared jargon and the Russian-borrowed items least often, and compensation shows up mainly around the nicknames, where a footnote or an added epithet restores what transliteration drops. G. Toury placed this kind of translator behaviour inside a social account of the act, holding that «translation activities should rather be regarded as having cultural significance. Consequently, translatorship amounts first and foremost to being able to play a social role, i.e., to fulfil a function allotted by a community» [2, p. 53]. The readership that received this Russian text wanted a readable novel of crime, not a glossary of Uzbek prison speech, and the pull toward the readable standard follows from that want. The count, then, does not simply record loss. It records a target norm at work.

Discussion. The result lines up with the descriptive account of translated language, and it adds a condition to it. Standardisation ran strongest where the source item was locked to one culture's institutions, the Uzbek penal world above all, and weakest where the two languages shared a supply of coarse words. This qualifies the levelling law rather than unseating it. A. Chesterman's split states the qualification well, since the levelling seen here is an S universal, a feature of the source-to-target relation, and its strength varied with a property of the source item, its cultural closedness, that a T universal account comparing translations with native Russian writing would miss [6]. A. Pym read standardisation and interference as risk-averse choices, the translator lowering the chance of a wrong move by reaching for the habitual option [17], and the borrowed items carried no such risk because their target form was fixed in advance. Seventy years of one penal system had left Russian words inside Uzbek criminal speech, and those words went home without a scratch.

For a low-resource pair the finding carries a practical weight. A translator working from Uzbek into Russian is not choosing between coarse and standard in a void, but drawing on a common substandard fund that other pairs do not possess, and a parallel corpus makes the size of that fund countable, a point that recent Uzbek work on corpus method in translation has

begun to press [18]. The method scales. Once the alignment is corrected by hand and the coarse items are tagged against Khimik's classes and the Mokienko and Nikitina dictionary, the same procedure will run on any two texts in the pair, and the shares of neutralisation and survival can be tracked across translators and across decades. What this single book yields is a starting figure, not a final one, and the next step is a larger collection holding several novels and several translators, so that the pull toward the standard can be weighed against the size of the shared coarse fund. Does the borrowed argot always resist, or only under translators who themselves grew up bilingual? The corpus can answer that.

Conclusion. The Uzbek to Russian rendering of Shaytanat pulls its coarse vocabulary toward the standard, and it does so unevenly. Closed argot bound to the Uzbek penal world and culture-bound euphemisms for killing neutralise most often, coarse common words and shared jargon survive more readily, and the words that Uzbek had borrowed from Russian cross back untouched, carried by seventy years of a shared criminal speech. Compensation clusters around the coarse nicknames, where a gloss or an added epithet restores what plain transliteration drops. The direction confirms the standardisation that descriptive translation study describes, while the unevenness shows that the confirmation holds only once the cultural closedness of each coarse item is weighed. A parallel corpus of the pair, tagged for substandard vocabulary and corrected by hand, turns that weighing into a count, and the count is the instrument this study offers to the study of translation between Uzbek and Russian.

References

1. Baker M. Corpus Linguistics and Translation Studies: Implications and Applications // Text and Technology: In Honour of John Sinclair / ed. by M. Baker, G. Francis, E. Tognini-Bonelli. Amsterdam; Philadelphia: John Benjamins, 1993. P. 233-250.
2. Toury G. Descriptive Translation Studies – and Beyond. Amsterdam; Philadelphia: John Benjamins, 1995. 311 p.
3. Laviosa S. Corpus-based Translation Studies: Theory, Findings, Applications. Amsterdam; New York: Rodopi, 2002. 138 p.
4. Catford J. C. A Linguistic Theory of Translation: An Essay in Applied Linguistics. London: Oxford University Press, 1965. 103 p.
5. Van Leuven-Zwart K. M. Translation and Original: Similarities and Dissimilarities, I // Target. 1989. Vol. 1, № 2. P. 151-181.
6. Chesterman A. Beyond the Particular // Translation Universals: Do They Exist? / ed. by A. Mauranen, P. Kuusimäki. Amsterdam; Philadelphia: John Benjamins, 2004. P. 33-50.
7. Химик В. В. Поэтика низкого, или Просторечие как культурный феномен. Санкт-Петербург: Филологический факультет СПбГУ, 2000. 272 с.
8. Девкин В. Д. Немецкая разговорная речь: Синтаксис и лексика. Москва: Международные отношения, 1979. 254 с.

9. Berezowski L. Dialect in Translation. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego, 1997. 152 p.
10. Мокиенко В. М., Никитина Т. Г. Большой словарь русского жаргона: 25 000 слов, 7 000 устойчивых сочетаний. Санкт-Петербург: Норинт, 2000. 716 с.
11. Xusainova Z. Y., Yangibayeva S. G. Mualliflik korpusi imkoniyatlari (Tohir Malikning “Shaytanat” asari asosida) // Buxoro davlat universiteti ilmiy axboroti. 2025. № 6 (123). B. 46-50.
12. Malik T. Shaytanat. Toshkent: Hilol-nashr, 2017.
13. Малик Т. Шайтанат. Царство сатаны / пер. с узб. Т. Нармирзаева, М. Алимова. Ташкент: Янги аср авлоди, 2016. 484 с.
14. Qobilova N. S., Rajabova M. A. Tohir Malik asarlarida inson xarakterini ifodalovchi leksik birliklarning ijobiy va salbiy xususiyatlari // Buxoro davlat universiteti ilmiy axboroti. 2024.
15. Крысин Л. П. Формы существования (подсистемы) русского национального языка // Современный русский язык: Социальная и функциональная дифференциация / отв. ред. Л. П. Крысин. Москва: Языки славянской культуры, 2003. С. 33-77.
16. Abdurakhmonova N. Z. Creation of Lexical Alignments for Uzbek-Russian Parallel Corpora // Computational and Cognitive Linguistics: TEL 2020. Kazan, 2020.
17. Pym A. On Toury's Laws of How Translators Translate // Beyond Descriptive Translation Studies / ed. by A. Pym, M. Shlesinger, D. Simeoni. Amsterdam; Philadelphia: John Benjamins, 2008. P. 311-328.
18. Umarova M. O'. Tarjimashunoslik taraqqiyotida korpus lingvistikasining o'rni va ahamiyati // Buxoro davlat universiteti ilmiy axboroti. 2025. № 6. B. 60-64.
19. Allaberdiev B., Matlatipov G., Kuriyozov E., Rakhmonov Z. Parallel Texts Dataset for Uzbek-Kazakh Machine Translation // Data in Brief. 2024. Vol. 55. Art. 110194.