

RU

Лингво-семантическое моделирование переводческих трансформаций
в условиях макротипологической асимметрии
русского и арабского языков

Аль Дауд Джаафар, Козеренко Е. Б.

Аннотация. В данной статье рассматривается проблема преодоления макротипологической асимметрии при переводе текстов с русского языка на арабский и с арабского на русский. Статья посвящена комплексному исследованию механизмов межязыковой передачи смыслов, при которой прямой перенос (транскодирование) оказывается невозможным из-за структурной интерференции. Особое внимание уделено деконструкции сложных глагольных комплексов, метафорических выражений и культурно-маркированной лексики в условиях отсутствия прямых словарных эквивалентов. Цель исследования заключается в разработке комплексной формальной модели лингво-семантических и синтаксических трансформаций для русско-арабской языковой пары на основе грамматик зависимостей с опорой на эмпирическую базу параллельных текстовых корпусов. Научная новизна работы определяется тем, что впервые механизмы преодоления лексико-синтаксической лакуарности между типологически дистантными языками (такие как деметафоризация, межчастеречная транспозиция (перекатегоризация) и синтаксическая реконфигурация) представлены в виде исчисляемых синтаксических деревьев на базе параллельного макротекста. Впервые обоснован прикладной потенциал русско-арабских параллельных массивов с применением принципа корпусной триангуляции для выявления переводческих закономерностей. В результате были выявлены и систематизированы ключевые стратегии адаптации текста (номинальные, семантические и экспликативные). Полученные результаты показали, что достижение прагматической эквивалентности алгоритмически невозможно без структурного расширения или редукции исходного графа, а использование корпусов страхует переводчика от субъективных девиаций и феномена «переводизма».

EN

Linguistic-semantic modeling of translation transformations
in the context of macro-typological asymmetry
of the Russian and Arabic languages

Al Dawood Jaafar, E. B. Kozerenko

Abstract. This paper is focused on the problem of overcoming macro-typological asymmetry in translating texts from Russian into Arabic and from Arabic into Russian. The paper is dedicated to a comprehensive study of the mechanisms of cross-linguistic meaning transfer in cases where direct transfer (transcoding) proves impossible due to structural interference. Particular attention is paid to the deconstruction of complex verbal phrases, metaphorical expressions, and culture-specific vocabulary in the absence of direct dictionary equivalents. The aim of the research is to develop a comprehensive formal model of linguistic-semantic and syntactic transformations for the Russian-Arabic language pair based on dependency grammars, utilizing the empirical data of parallel text corpora. The scientific novelty of the study lies in the fact that, for the first time, the mechanisms for overcoming lexical-syntactic lacunarity between typologically distant languages (such as demetaphorisation, cross-part-of-speech transposition (recategorization), and syntactic reconfiguration) are presented in the form of computable syntactic dependency trees based on a parallel macrotext. Furthermore, the applied potential of Russian-Arabic parallel arrays is substantiated for the first time by applying the principle of corpus triangulation to identify translation patterns. As a result, the key text adaptation strategies (nominal, semantic, and explicative) were identified and systematized. The results obtained showed that achieving pragmatic equivalence is algorithmically impossible without the structural expansion or reduction of the source graph, and that the use of corpora safeguards the translator from subjective deviations and the phenomenon of "translationese".

Introduction

The linguistic phenomenon of translation and the automatic processing of typologically distant languages necessitate the search for qualitatively new formal grammatical and corpus-based solutions. At the current stage of computational linguistics development, traditional methods of direct transcoding prove insufficient for the adequate transmission of semantics when comparing languages of the inflective type characterized by different mechanisms of inflection realization (the predominance of extroflexion in Russian and introflexion in Arabic).

The relevance of this research is driven by the acute need to develop transductive translation models capable of overcoming deep structural interference between the Russian and Arabic languages through the mathematical parameterization of syntactic shifts. The relevance of the study is further determined by the demand within modern translation studies to transition from conjectural classifications of transformations to a rigorous descriptive statistical analysis based on objective data from parallel corpora. The identification of latent patterns in cross-linguistic transfer meets the needs of computational linguists and developers of machine learning systems.

To achieve this goal, the following objectives were set:

1. to analyze the methodology of working with parallel corpora and text alignment algorithms under conditions of macro-typological asymmetry;
2. to identify and describe the mechanisms for overcoming lexical-syntactic lacunarity in translation from Russian into Arabic and vice versa;
3. to verify the identified transformations using extracted corpus concordances;
4. to perform graph-based modeling of syntactic shifts using formal dependency grammar.

To accomplish the objectives defined in this work, the following research methods were employed: comprehensive and purposive sampling was applied for the extraction of empirical material from multilingual databases; the comparative-contrastive (typological) method was used for the explication of macro-typological asymmetries in Russian and Arabic syntax; the corpus-based method, in synergy with the concept of data triangulation, served as the primary tool for verifying contextual equivalents; the method of formal (graph) modeling was utilized to construct computable dependency trees.

The research materials comprise reference books, dictionaries, and parallel text arrays extracted from the multilingual open-source database *Tatoeba*, which ensures a high degree of pragmatic naturalness in translations.

The theoretical framework of the study is grounded in foundational works in the fields of translation theory, functional linguistics, and computational linguistics. The research relies on the postulates of the transductive "Meaning ↔ Text" paradigm by I. A. Mel'čuk (2012, 2021) and Yu. D. Apresyan (1995), the concept of formal syntax modeling by V. S. Khrakovsky (2024), as well as the theories of equivalence and translation transformations by M. Baker (2018) and E. Alymova (2022). The study of the mechanisms of semantic shifts is based on the research of A. M. Mutlaq (2024) in the field of pragmatics and metonymy. The evolution of natural language processing algorithms and network technologies is comprehended through the prism of works by M. B. Rarenko (2021) and L. N. Beliaeva, et al. (2019). Finally, an important aspect was the determination of principles for working with parallel corpora and text alignment, which became possible thanks to the contributions of S. Sharoff, et al. (2023), H. Kemppanen (2025), and M. Saad, et al. (2025).

The practical significance of this study lies in the fact that the proposed methodology for corpus analysis and the graph models of syntactic shifts can be utilized by computational linguists to optimize Machine Translation (MT) systems and algorithms for automatic Arabic text processing. The obtained data can be applied in higher education institutions specializing in the humanities, specifically in specialized courses on translation studies and contrastive syntax, as well as in the educational and methodological training of Arabic translators based on the concept of Data-Driven Learning.

Discussion and results

Macro-typological asymmetry as a factor of structural interference

The analysis of the architectural vectors of bilingual databases reveals that, under conditions of Russian-Arabic macro-typological asymmetry, the algorithm of cross-linguistic transfer lacks the property of reversibility. Moreover, when working with corpus data, a researcher inevitably encounters the phenomenon of "translationese" – a statistically significant structural divergence between a translated text and a text originally produced by a native speaker. As S. Sharoff notes, the emergence of such a deviation is determined by the very nature of cross-linguistic transfer, in particular by the tendency towards explicitation, which constitutes "the need to provide more information in the translated text for what remains implicit in the source text" (Sharoff et al., 2023, p. 130).

Substantiating the morphological causes of this asymmetry, M. Saad and co-authors emphasize that Arabic is a "highly inflected language" in which "certain Arabic terms can be agglutinated" (Saad et al., 2025, p. 7). Due to such syncretism, a single Arabic word form can act as the functional-semantic equivalent of a multi-component Russian phrase. Obviously, the processing of such complex morphosyntactic complexes requires abstracting from surface syntax and establishing correspondences at the level of deep semantic roles using formal dependency grammars.

In justifying the mathematical apparatus of our research, we rely on the concept of I. A. Mel'čuk (2021, p. 17), who categorically states that "there is no other, non-dependency, syntax. The syntactic structure of a sentence cannot

be described in any other way – for instance, in terms of constituents, or phrases". Thus, the use of constituency grammars (which fragment text into linear blocks) obscures the true hierarchy of the language, whereas dependency grammar allows the translation system to construct a universal semantic-syntactic graph, free from the nation-specific topline envelope. Consequently, "the internal semantic status of a sentence is formed by elementary meanings" (agent, patient), acting as an invariant during translation, which determines the application of formal dependency grammars (Khrakovsky, 2024, p. 17).

During the corpus-driven modeling, it was established by us that the architectonics of translation strategies in the Russian-Arabic pair is objectively classified into the macro-categories of nominal, semantic, and explicative transformations.

Nominal transformations: pragmato-deictic transformation and preservation of the original

Taking into account the pragmatic nature of deixis, it should be stated that the introduction of deictic elements in the Arabic-Russian language pair frequently acts as an instrument of syntactic compression. Within the terminological apparatus of modern translation studies, this mechanism, realized within the nominal macro-category, is most reasonably qualified as a pragmato-deictic transformation. The strategy involves the replacement of a fully meaningful lexeme of the source text with a deictic pointer. Verification of this strategy on the material of the Tatoeba corpus reveals that such an approach allows avoiding structural overloading. For example, the Arabic utterance «هَذَا هُوَ السَّبَبُ فِي أَنِّي أَكْرَهُهُ» (Lit.: this is the reason why I hate him) is translated into Russian as «**Вот** за что я его ненавижу» (That's what I hate him for). In this case, the fully meaningful Arabic noun «السَّبَبُ» (reason) is replaced by the Russian deictic particle «Вот». Graph modeling shows that in the Russian language, this particle occupies the position of a specifier, determining the referential link with the entire subsequent proposition, which allows modeling a concise construction while preserving the pragmatic invariant.

Source text (Arabic):

هَذَا هُوَ السَّبَبُ فِي أَنِّي أَكْرَهُهُ.

Transcription: (hādhā huwa al-sabab fī annanī akrahuhu) (Tatoeba: ID 1924054).

Target text (Russian):

Вот за что я его ненавижу (Tatoeba: ID 1924044).

To visualize the syntactic shift caused by the introduction of the deictic marker, let us turn to graph modeling (Figure 1 – Arabic original, Figure 2 – Russian translation).

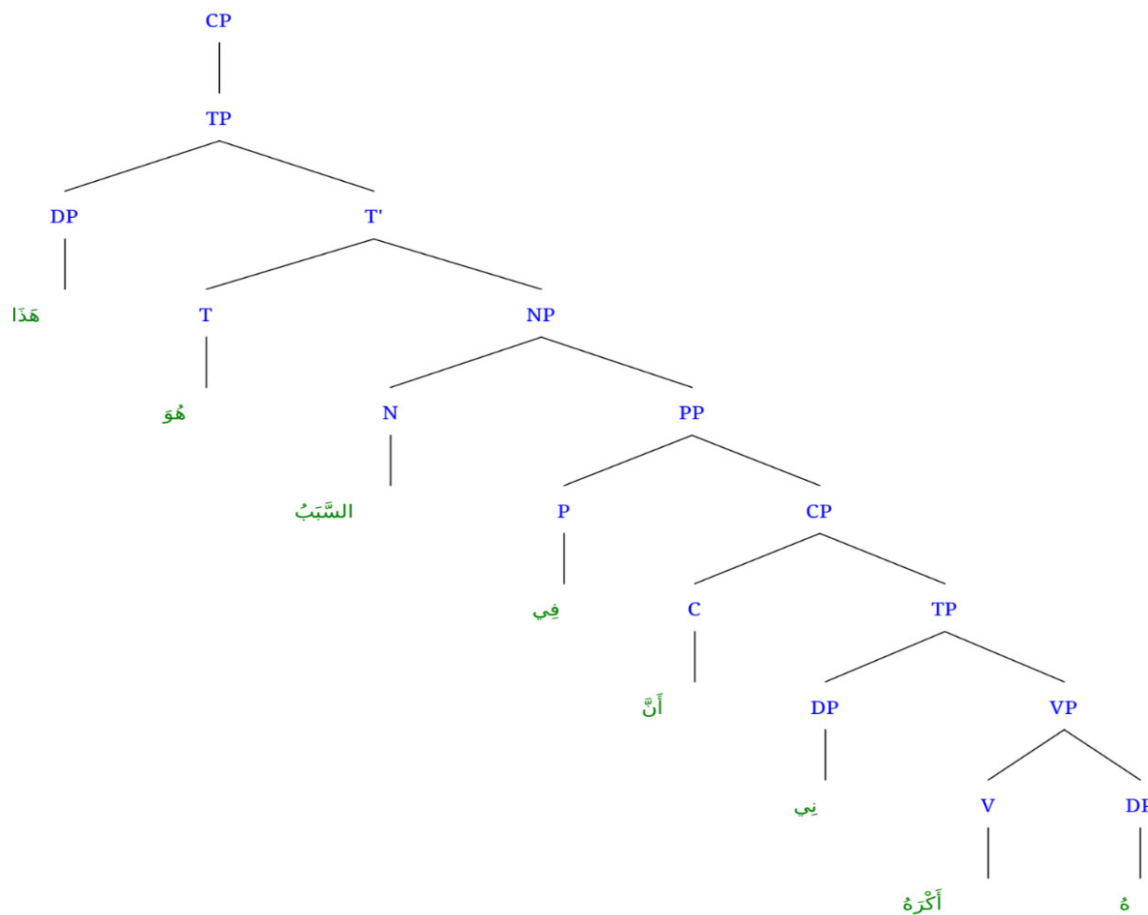


Figure 1. Arabic original

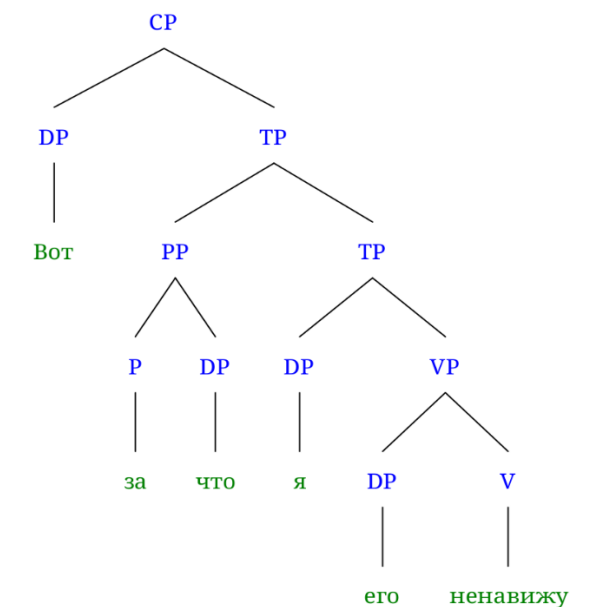


Figure 2. Russian translation

In turn, the strategy of preserving the source text element (untranslated transfer or transliteration) is conceptualized as the direct transfer of a lexeme, ensuring the preservation of ethnocultural or terminological identity. As the analysis demonstrates, this technique dominates in the transfer of highly specialized, culturally marked vocabulary and religious-historical terminology. Thus, in the example «من أي سورة هذه الآية؟» (From which surah is this ayah?), the Arabic lexemes «سورة» (chapter of the Quran) and «آية» (verse of the Quran) are transferred into the target text in the form of transliteration, without descriptive expansions and semantic adaptation. Such a decision eliminates the risk of losing culturally significant information, avoiding semantic divergence.

Source text (Arabic):

من أي سورة هذه الآية؟

Transcription: (min 'ayyi sūratin hādhihi l-'āyatu?) (Tatoeba: ID 4284119).

Target text (Russian):

Из какой **суры** этот **аят**? (Tatoeba: ID 4284120).

Semantic transformations: demetaphorisation and synonymic shift

The achievement of pragmatic adequacy logically requires deviation from the literal meaning. Determined by profound divergences in the ways of conceptualizing reality, semantic transformations algorithmically compensate for macro-typological distance. Operating in opposition to the strategy of metaphorisation is the mechanism of demetaphorisation, in which a figurative unit of the source text is substituted by a purely descriptive equivalent.

As the analysis of corpus data proves, vivid Arabic images are frequently reduced to usual statements in the Russian language. For instance, the spatial-game metaphor in the Arabic sentence «الكرة الآن في ملعبك» (Lit.: The ball is now in your stadium/field!) undergoes partial demetaphorisation in the Russian translation: «Мяч сейчас на твоей стороне!» (The ball is now on your side!). The volumetric image of a physical structure (stadium) is eliminated, giving way to the abstract geometric noun "side". The removal or reduction of the trope decreases the aesthetic multidimensionality of the text; however, in strictly informative texts, the transmission of pure denotative information prevails over the preservation of stylistic form.

Source text (Arabic):

الكرة الآن في ملعبك!

Transcription: (al-kuratu l-'āna fī mal'abika!).

Literal translation: The ball is now **in your stadium / playing field!** (Tatoeba: ID 1602999).

Target text (Russian):

Мяч сейчас **на твоей стороне!** (Tatoeba: ID 4615038).

Of particular research interest is synonymic translation. As noted by I. B. Rubert and co-authors, when searching for an adequate equivalent, "it is not always sufficient to transform the surface structures of the source language into the structures of the target language, but it is necessary to conduct a deep analysis of the linguistic correspondences <...>, that is, to determine the content core" (Rubert et al., 2021, p. 172). The analysis of the translation of verbs of verbalization reveals the replacement of the Arabic verb «أعلن» (to announce, to declare) in the syntagma «أعلن حبه» (He announced his love) with the Russian quasi-synonym «признаться» (to confess) («Он признался в любви» – He confessed his love). The use of a direct equivalent in the Russian language would be characterized by excessive

pathos and would violate natural lexical collocability, whereas synonymic substitution based on a common content core demonstrates the filigree use of the resources of the recipient language.

Source text (Arabic):

أَعْلَنَ حُبَّهُ.

Transcription: ('a'ḥana ḥubbahu).

Literal translation: *He announced/declared his love* (Tatoeba: ID 371449).

Target text (Russian):

Он признался в любви (Tatoeba: ID 5491690).

Explicative transformations: cross-part-of-speech transposition (recategorization) and omission

The architectonics of translation strategies includes a category of complex explicative transformations that determine the fundamental morphosyntactic restructuring of the source text. The theoretical rationale for the necessity of such shifts stems directly from macro-typological differences in the mechanisms of coining new vocabulary. Investigating this phenomenon, F. B. Karim emphasizes that, in contrast to Russian suffixal word formation, "in the Arabic model of word production, internal inflection operates" (Karim, 2024, p. 43-44). Due to such an asymmetry between consonantal roots and affixal systems, the direct grammatical transfer of concepts often proves impossible, which forces the translator to resort to cross-part-of-speech transposition (recategorization).

Empirical verification confirms the high frequency of cross-part-of-speech transposition. The most representative illustration is the transposition according to the "Noun → Adverb" model. In the Arabic sentence «لَقَدْ تَسَلَّقْتُمُ الشَّجَرَةَ بِسُرْعَةٍ», the adverbial modifier of manner is expressed by the prepositional-nominal group «بِسُرْعَةٍ» (with speed). In the Russian target text («Вы быстро забрались на дерево» – You climbed the tree quickly), this group is transformed into the concise adverb «быстро» (quickly). The construction of dependency trees captures the elimination of the prepositional and nominal nodes in favor of a single adverbial node, which confirms the thesis regarding the asymmetry of derivational models: Arabic syntax requires an analytical adverbial phrase, whereas the Russian language realizes this semantics via suffixation.

Source text (Arabic):

لَقَدْ تَسَلَّقْتُمُ الشَّجَرَةَ بِسُرْعَةٍ.

Transcription: (laqad tasallaqtumu š-šāgarata bi-sur'atin).

Literal translation: *You climbed the tree with speed / with swiftness* (Tatoeba: ID 7353459).

Target text (Russian):

Вы **быстро** забрались на дерево (Tatoeba: ID 4924346).

To formalize this syntactic shift, let us turn to graph modeling (Figure 3 – Arabic original, Figure 4 – Russian translation).

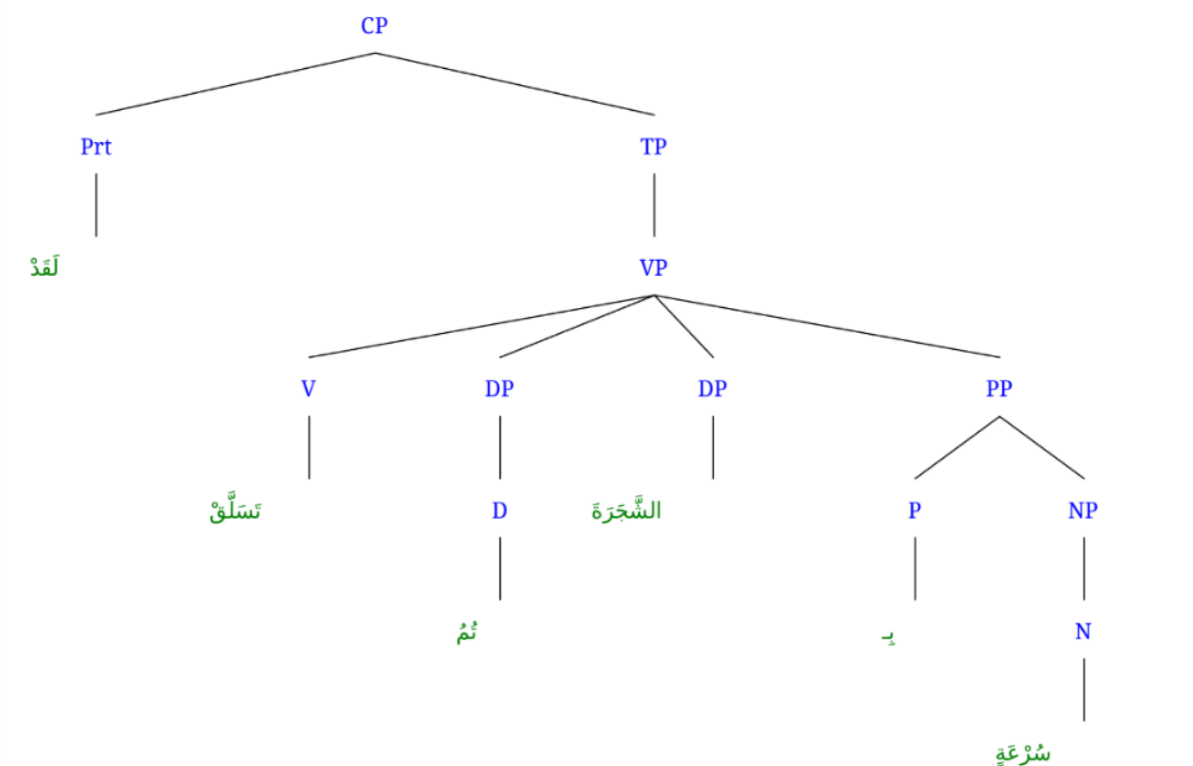


Figure 3. Syntactic model of the Arabic sentence

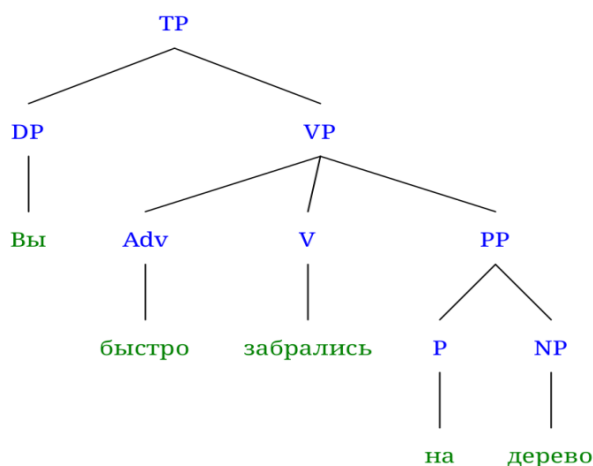


Figure 4. Syntactic model of the Russian sentence

No less significant is the recategorization of substantivized forms. The Arabic construction «الْعَاقِلُ مَنْ اتَّعَطَّ بِغَيْرِهِ» (the understanding one), is translated into Russian using a substantivized adjective: «Умный тот, кто учится у других» (The smart one is he who learns from others). The formal model demonstrates that, to achieve pragmatic adequacy, the translator resorts to cross-part-of-speech transposition (in this example, replacing the morphological category of the participle with the category of the adjective), which is dictated by the usual preferences of the recipient language.

Source text (Arabic):

الْعَاقِلُ مَنْ اتَّعَطَّ بِغَيْرِهِ.

Transcription: (al-‘āqilu mani tta‘aẓa bi-ḡayrihi).

Literal translation: The *understanding* one / The *comprehending* one is he who learns from others (Tatoeba: ID 12157819).

Target text (Russian):

Умный тот, кто учится у других (Tatoeba: ID 12157824).

The visualization of this categorical conversion is presented in Figures 5 and 6.

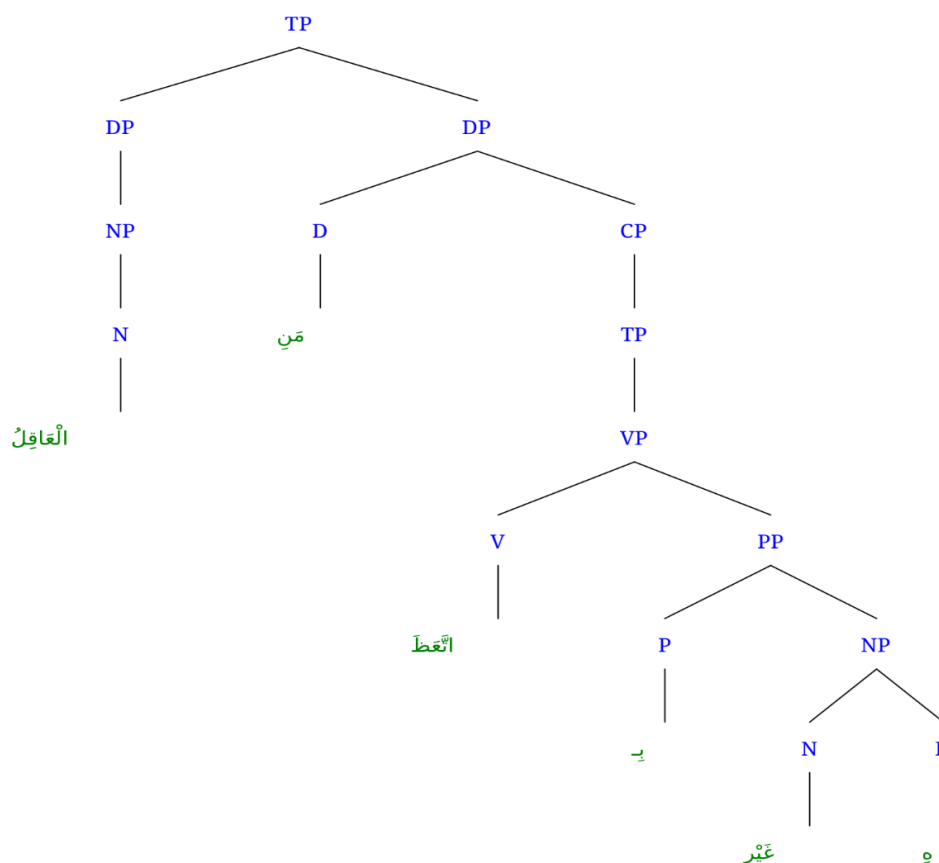


Figure 5. Syntactic model of the Arabic sentence

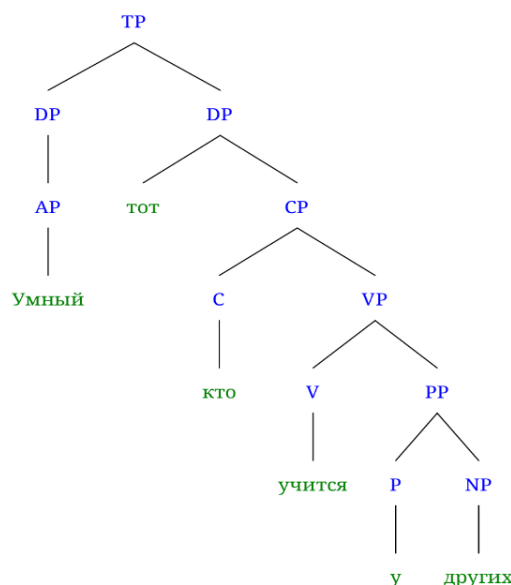


Figure 6. Syntactic model of the Russian sentence

Special attention should be paid to strategies of omission. As shown by the analysis of the ultra-compact Arabic verbal form «أَنْزِلْهُمْهَا» (containing an interrogative particle, a subject prefix, a verbal stem, and a cascade of two pronominal objects), when translated into Russian («Неужели нам следует заставить вас?» – Should we really force you?), the position of the second incorporated object is completely eliminated. On the dependency graph, this transformation is marked by an empty category node (Gap $\emptyset$). Such a rupture in the chain of predicate-argument relations leads to the emergence of a referential gap. It seems appropriate to note that the cited example from the Tatoeba corpus raises legitimate questions regarding the completeness of the translation correspondence. Obviously, the considered solution reflects the subjective specificity of a particular corpus example (an individual translation deviation recorded on the Tatoeba platform) rather than a normative translation strategy. Nevertheless, the analysis of such controversial variants clearly confirms the thesis that the strategy of omission requires utmost caution and is permissible only provided there is full compensation of meaning by the macro-context.

Source text (Arabic):

أَنْزِلْهُمْهَا؟

Transcription: ('a-nu-lzimu-kumu-hā?) (Tatoeba: ID 4087145).

Target text (Russian):

Неужели нам следует заставить вас? (Tatoeba: ID 5212445).

For a visual explication of the structural gap, let us turn to graph modeling (Figure 7 – morphosyntactic decomposition of the Arabic form, Figure 8 – model of the Russian analytical correspondence indicating the eliminated object).

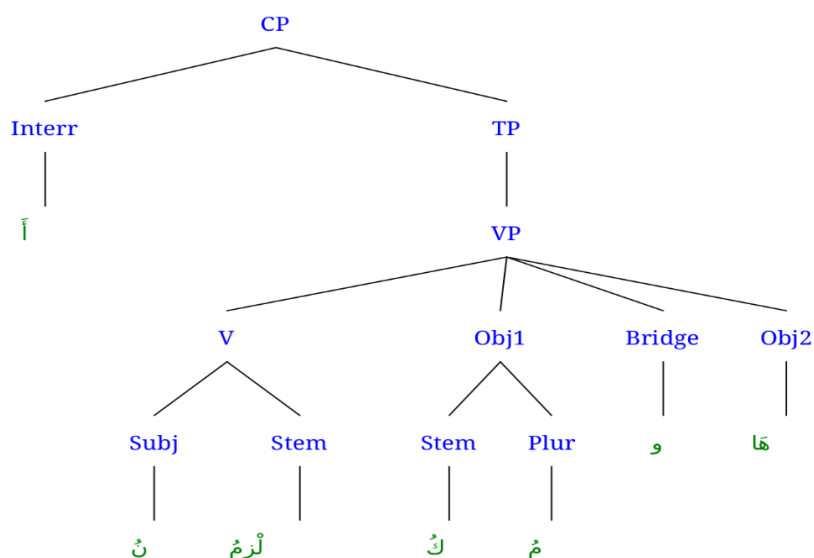


Figure 7. Arabic original

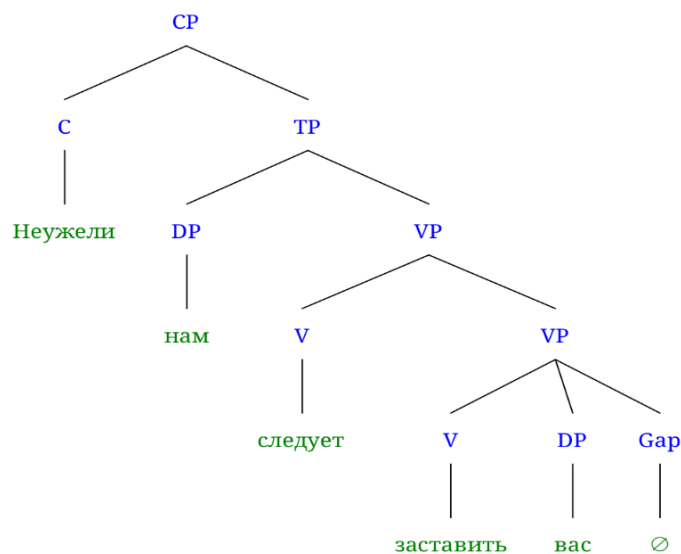


Figure 8. Russian translation

Summarizing the results of the conducted analysis, it should be stated that overcoming macro-typological interference in the Russian-Arabic language pair is of a systemic nature. As graph models and corpus concordances demonstrate, nominal, semantic, and explicative transformations rarely function in isolation. Most often, they act as a complex compensatory mechanism that allows the translator to adapt the text to the grammatical and usual norms of the recipient linguaculture. It is obvious that the rejection of linear transcoding in favor of deep syntactic and semantic reconfiguration (up to the deliberate omission of incorporated elements) becomes a necessary condition for preserving the pragmatic invariant of the original. The use of formal dependency grammars in synergy with corpus data provides objective verification of these processes, minimizing the influence of the subjective factor when selecting translation solutions.

Conclusion

Summing up the linguistic-semantic analysis and the formal modeling of translation transformations, the following conclusions can be formulated:

Firstly, while analyzing the methodology of working with parallel corpora, it was established that algorithmic text alignment in the Russian-Arabic language pair requires abstracting from surface linear syntax. The application of hybrid approaches and corpus triangulation allows overcoming macro-typological asymmetry, ensuring the objective extraction of data regardless of the quantitative mismatch of sentence boundaries.

Secondly, the mechanisms for overcoming lexical-syntactic lacunarity were identified. It has been proven that achieving pragmatic equivalence between the Arabic and Russian languages under conditions of macro-typological asymmetry of their inflective mechanisms (introflexion and extroflexion) is commonly impossible without the systemic application of compensatory mechanisms. The identified transformations are objectively stratified into three macro-categories: nominal (calquing, transliteration, pragmato-deictic transformations), semantic (generalization, specification, metaphorisation), and explicative (cross-part-of-speech transposition, descriptive translation).

Thirdly, the identified transformations were successfully verified using extracted corpus concordances. Corpus statistics confirm that the choice of a specific strategy (for example, replacing a neutral lexeme with a figurative metaphor or using converse antonymic translation) is not a random deviation, but typically acts as a normative mechanism for adapting the text to the usual expectations of the target audience, safeguarding the translator from the phenomenon of "translationese."

Fourthly, graph modeling of syntactic shifts based on formal dependency grammars was carried out. The construction of dependency trees mathematically proved that the translation of Arabic *masdars*, participles, and ultra-compact verbal forms requires cross-part-of-speech transposition (recategorization), the elimination of incorporated objects (omission), or syntactic expansion. The translator algorithmically dismantles the source actant structure and reconstructs it anew, submitting to the valency constraints of the recipient language.

Prospects for further research are seen in expanding the empirical base by involving specialized Learner Translator Corpora. A promising direction is the use of bilingual topic modeling methods and contextual vector representations for the algorithmic identification of semantic prosody, as well as the integration of the obtained graph models into the architecture of modern neural machine translation systems to improve the quality of Arabic text processing.

Материалы исследования | Research materials

1. Tatoeba: multilingual open database of sentences and translations. <https://tatoeba.org/en/>

Источники | References

1. Апресян Ю. Д. Избранные труды. М.: Школа «Языки русской культуры», 1995. Т. II. Интегральное описание языка и системная лексикография.
2. Беляева Л. Н., Камшилова О. Н., Пиотровская К. Р. Сетевые лингвистические технологии: коллективная монография. СПб: Изд-во РГПУ им. А. И. Герцена, 2019.
3. Карим Ф. Б. Суффиксальный способ образования новой лексики в русском и арабском языках и его влияние на достижение переводческой адекватности // Научный форум: Филология, искусствоведение и культурология. 2024. № 7 (83).
4. Мельчук И. А. Язык: от смысла к тексту. М.: Языки славянской культуры, 2012.
5. Мутлак А. М. Ироническая метонимия в русском и арабском языках // Journal of the College of Languages. 2024. № 49. <https://doi.org/10.36586/jcl.2.2024.0.49.0145>
6. Паренко М. Б. Машинный перевод: от перевода «по правилам» к нейронному переводу (обзор). // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 6. Языкознание: реферативный журнал. 2021. № 3. <https://doi.org/10.31249/ling/2021.03.05>
7. Руберт И. Б., Киселева С. В., Миронова М. Ю. Перевод многозначных и синонимичных терминов английского языка: возможности когнитивного подхода // Вестник Волгоградского государственного университета. Серия 2. Языкознание. 2021. Т. 20. № 4. <https://doi.org/10.15688/jvolsu2.2021.4.14>.
8. Храковский В. С. Очерки по общему и арабскому синтаксису / отв. ред. Г. В. Церетели. Изд. 2-е, стереотип. М.: ЛЕНАНД, 2024.
9. Alymova E. Cross-cultural lacunarity and translation techniques: a corpus-based study of English, Russian and Spanish: PhD thesis. Glasgow: University of Glasgow, 2022. –
10. Baker M. In Other Words: A Coursebook on Translation. 3rd ed. London; New York: Routledge, 2018.
11. Kemppanen H. Empirical and Quantitative Approaches to Corpus Translation Studies // The Routledge Handbook of Corpus Translation Studies / ed. by D. Li, J. Corbett. London; New York: Routledge, 2025.
12. Mel'čuk I. A. Ten Studies in Dependency Syntax. Berlin; Boston: Walter De Gruyter Mouton GmbH, 2021.
13. Saad M., Langlois D., Smaïli K. Building and Aligning Comparable Corpora. // arXiv.org. 2025. 4 Aug. <https://doi.org/10.48550/arXiv.2508.02555>
14. Sharoff S., Rapp R., Zweigenbaum P. Building and Using Comparable Corpora for Multilingual Natural Language Processing. Cham: Springer Nature Switzerland, 2023. <https://doi.org/10.1007/978-3-031-31384-4>

Информация об авторах | Author information**RU****Аль Дауд Джаафар¹****Козеренко Елена Борисовна²**, к. филол. н.¹ Российский университет дружбы народов имени Патриса Лумумбы, г. Москва² Российский университет дружбы народов имени Патриса Лумумбы, г. Москва;
Федеральный исследовательский центр «Информатика и управление» РАН, г. Москва**EN****Al Dawood Jaafar¹****Elena Borisovna Kozerenko²**, PhD¹ RUDN University, Moscow² RUDN University, Moscow;

Federal Research Center 'Computer Science and Control' of the RAS, Moscow

¹ al-dawood-j@rudn.ru, ² kozerenko@mail.ru**Информация о статье | About this article**

Дата поступления рукописи (received): 20.06.2026; опубликовано online (published online): 30.07.2026.

Ключевые слова (keywords): переводческие трансформации; параллельные корпуса; Русско-арабская языковая пара; лакуарность; грамматики зависимостей; translation transformations; parallel corpora; Russian-Arabic language pair; lacunarity; dependency grammars.