

CRITERIA FOR DETERMINING TRANSLATION EQUIVALENCE OF STABLE COMBINATIONS IN THE UZBEK–ENGLISH PARALLEL CORPUS

Nurullayeva Oydin Nazirjon qizi

Senior Lecturer at TUAS, PhD

Abstract. The article examines the criteria for determining translation equivalence of stable combinations in the Uzbek–English parallel corpus. The study proceeds from the view that translation difficulties in bilingual corpus construction are not limited to phrasemes alone. They also concern proverbs, sayings, periphrases, collocations, terminological combinations, speech etiquette formulas and discourse patterns. Therefore, the article treats stable combinations as a broader linguistic category that includes semantically, structurally or pragmatically fixed units used in speech as ready-made expressions. The main purpose of the article is to define the principles according to which Uzbek stable combinations may be correlated with their English equivalents in a parallel corpus. The analysis shows that equivalence should not be reduced to lexical coincidence. It must be determined through semantic integrity, functional adequacy, structural correspondence, cultural relevance, stylistic compatibility and contextual validity. The article proposes a classification of equivalence types: full equivalence, partial equivalence, functional equivalence, cultural equivalence, descriptive equivalence, collocational equivalence, terminological equivalence and pragmatic equivalence. Uzbek and English examples are provided for each type. The results may serve as a theoretical basis for developing bilingual lexical resources, translation databases and corpus-based tools for Uzbek–English translation studies.

Keywords: Uzbek language, English language, parallel corpus, stable combination, translation equivalence, phraseme, proverb, saying, periphrasis, collocation, terminological combination, corpus linguistics.

In contemporary linguistics, parallel corpora have become one of the most reliable empirical resources for studying interlingual correspondence. Unlike traditional bilingual dictionaries, a parallel corpus demonstrates not only lexical meaning but also contextual realization, syntactic arrangement, stylistic choice and translation strategy. J. Tiedemann describes OPUS as a large and continually expanding collection of parallel corpora intended for computational linguistics, translation studies and cross-linguistic research [8]. P. Koehn's Europarl corpus also proved that sentence-aligned bilingual and multilingual materials may function as a valuable source for statistical machine translation and contrastive linguistic analysis [3].

For Uzbek linguistics, the development of parallel corpora is especially significant because Uzbek remains underrepresented in many large international corpus infrastructures. Existing studies on Uzbek computational linguistics have already addressed such issues as morphological analysis, lemmatization, automatic translation, educational corpora and phraseological databases. In this regard, O.N. Nurullayeva's dissertation, *Foundations for Creating a Parallel Corpus of Uzbek and English Phrasemes*, is of particular importance. The dissertation develops a theoretical and methodological model of the Uzbek–English phraseme parallel corpus, proposes principles of morphological and semantic description, and forms a database consisting of 420 Uzbek–English phraseme pairs supplied with explanations and contextual examples [4].

Nevertheless, the translation problems that arise in corpus construction are not restricted to phrasemes. A broader layer of language units also requires special attention. These include

proverbs, sayings, periphrases, collocations, terminological combinations, speech etiquette formulas and discourse patterns. For example, the Uzbek expression *og'zi qulog'ida* may correspond to the English grinning from ear to ear; the proverb *Yetti o'lchab, bir kes* may be rendered by Look before you leap; the collocation *qaror qabul qilmoq* corresponds to make a decision; the discourse formula *xulosa qilib aytganda* corresponds to in conclusion. These examples show that different types of stable combinations require different criteria for determining equivalence.

The aim of this article is to define the main criteria for determining translation equivalence of stable combinations in the Uzbek–English parallel corpus. The task is not to describe the technical annotation of such units, but to clarify how their English equivalents should be selected, classified and justified at the linguistic level.

The theoretical study of stable combinations is closely connected with phraseology, lexicology, translation theory and corpus linguistics. A.P. Cowie's edited volume *Phraseology: Theory, Analysis, and Applications* presents phraseology as a field concerned with conventionalized multi-component expressions and their use in lexicography, language learning, stylistics and computational analysis [1]. I.A. Sag and his co-authors emphasize that multiword expressions are a key difficulty for natural language processing because such expressions cannot always be interpreted through the meanings of their individual components [5]. This position is especially relevant for Uzbek–English translation, where literal rendering often leads to semantic distortion.

In international corpus linguistics, several resources have been developed for idiomatic and fixed expressions. MAGPIE is a large English corpus of potentially idiomatic expressions based on automatic pre-extraction and controlled annotation; it contains more than 50,000 instances of idiomatic expressions [2]. EPIE contains 25,206 sentences with lexical instances of 717 idiomatic expressions and includes both idiomatic and literal uses [7]. The PARSEME project focuses on verbal multiword expressions and provides multilingual data for their identification and classification [6]. These studies show that stable multi-component expressions require special treatment in corpus-based language processing.

However, most of these resources are either monolingual or focused mainly on idioms and verbal expressions. The broader class of stable combinations, including proverbs, periphrases, collocations, terminological combinations and discourse formulas, has not been sufficiently studied in the Uzbek–English corpus context. Therefore, the present article extends the scope from phrasemes to stable combinations as a wider linguistic category.

The article applies descriptive, comparative, classificatory and componential methods. The descriptive method is used to define the concept of stable combination and its internal types. The comparative method is applied to Uzbek and English examples in order to reveal the nature of translation correspondence. The classificatory method is used to distinguish the types of equivalence. Componential analysis helps identify semantic elements that are preserved, transformed or lost in translation.

The material consists of Uzbek stable combinations and their possible English equivalents selected from phraseological, paremiological, collocational, terminological and discourse-pragmatic layers. The examples are presented not as a complete corpus database but as model cases for determining equivalence criteria in a future Uzbek–English parallel corpus.

In this article, a stable combination is understood as a multi-component linguistic unit that is reproduced in speech as a ready-made expression and possesses semantic, structural or pragmatic stability. Such a unit may be fully figurative, partially figurative or non-figurative, but it is not entirely free in its lexical composition or communicative function.

Stable combinations include several internal groups:

Type of stable combination	Uzbek example	English equivalent	Dominant feature
Phraseme	ko'ngli tog'dek ko'tarildi	his spirits lifted	figurative meaning
Phraseme	ko'z yumdi	turned a blind eye	idiomatic meaning
Proverb	Yetti o'lchab, bir kes	Look before you leap	didactic generalization
Saying	Oyni etak bilan yopib bo'lmas	The truth will out	figurative folk wisdom
Periphrasis	oq oltin	white gold / cotton	figurative naming
Collocation	qaror qabul qilmoq	make a decision	lexical stability
Collocation	xulosa chiqarmoq	draw a conclusion	restricted lexical choice
Terminological combination	parallel korpus	parallel corpus	terminological stability
Speech etiquette formula	xush kelibsiz	welcome	communicative formula
Discourse pattern	xulosa qilib aytganda	in conclusion	text-organizing function

The table demonstrates that stable combinations differ in their linguistic nature. A phraseme such as *kapalagi uchdi* cannot be translated word for word as *his butterfly flew*; it requires a functional equivalent such as *he was scared stiff*. A collocation such as *e'tibor qaratmoq* is not idiomatic in the strict sense, but its English equivalent is not turn attention; the normative form is pay attention. A discourse formula such as *shuni ta'kidlash joizki* is not figurative, but it has a fixed academic function and corresponds to it should be noted that. Hence, the selection of equivalents in a parallel corpus must be based on the type of stable combination.

Main Criteria for Determining Translation Equivalence

1. Semantic Integrity. The first criterion is semantic integrity. If the meaning of a stable combination is not equal to the sum of its components, the equivalent must be selected according to the whole meaning, not according to individual lexical items. For instance, *og'zi qulog'ida* expresses a person's visible joy. Its natural English equivalent is *grinning from ear to ear*. Both expressions use the image of facial expression and both denote strong happiness. The equivalence here is not word-for-word but image-based and semantic.

Another example is *boshi osmonga yetdi*. A literal translation, *his head reached the sky*, would sound unnatural in English. Depending on the context, the equivalent may be *he was over the moon*, *he was extremely happy*, or *he was filled with pride*. If the context emphasizes joy, *over the moon* is appropriate; if it emphasizes pride or honour, a descriptive equivalent may be preferable.

2. Functional Adequacy. The second criterion is functional adequacy. In many cases, the English equivalent does not reproduce the same image but performs the same communicative function. The Uzbek *kapalagi uchdi* and the English *he was scared stiff* differ in imagery, yet both express sudden fear. Similarly, *yuragi orqasiga tortdi* may be rendered as *his heart sank* when the context refers to disappointment, fear or emotional collapse.

Functional equivalence is especially important for culturally specific units. For example, *oq yo'l* is often used as a blessing or wish before a journey or important undertaking. In English, its equivalent depends on the situation: safe journey, good luck, wishing you success. None of these equivalents reproduces the exact cultural form of the Uzbek expression, but each may fulfil the same communicative function.

3. Structural Correspondence. The third criterion is structural correspondence. Some stable combinations have parallel structural models in Uzbek and English. For example, *qaror qabul qilmoq* corresponds to *make a decision*, *xulosa chiqarmoq* corresponds to *draw a conclusion*, *ta'sir ko'rsatmoq* corresponds to *have an effect*, and *masala ko'tarmoq* corresponds to *raise an issue*. These units are not idioms in the narrow sense, but they are stable collocational combinations. Structural correspondence is particularly useful in building corpus search models. If the Uzbek unit contains a noun plus a light verb, the English equivalent often contains a verb plus an abstract noun. However, the verbs are not freely interchangeable. Uzbek *qaror qabul qilmoq* cannot be rendered as *accept a decision* in ordinary English usage; the correct collocation is *make a decision*. Therefore, collocational stability must be considered a major criterion of equivalence.

4. Cultural Relevance. The fourth criterion is cultural relevance. Some Uzbek stable combinations encode cultural, religious or social meanings that cannot be fully replaced by a direct English expression. For example, *duo olmoq* may be translated as *to receive someone's blessing*. The English expression conveys the main meaning, but it may not fully express the social and spiritual value of receiving a blessing from elders in Uzbek culture.

The expression *non sindirmoq* is another case. Depending on context, it may refer to sharing bread, establishing social closeness or taking part in a traditional act of hospitality. Possible English equivalents include *to share bread*, *to sit down to a meal together*, or *to establish friendly ties*. The equivalent must therefore be selected after analyzing the cultural context. The same applies to *ko'ngil olmoq*. In one context it may mean *to comfort someone*; in another it may mean *to win someone's affection*; in a third context it may mean *to please someone*. Thus, the equivalent cannot be fixed mechanically. It must be chosen according to the semantic and pragmatic situation.

5. Stylistic Compatibility. The fifth criterion is stylistic compatibility. A stable combination may belong to colloquial, literary, official, academic or religious style. Its English equivalent must correspond to the same stylistic level. For example, *og'ziga talqon solmoq* is figurative and colloquial; depending on context it may be rendered as *to keep silent*, *to say nothing*, or *to remain tight-lipped*. The last variant is closer in expressive force.

In academic discourse, formulas such as *shuni ta'kidlash joizki*, *yuqoridagilardan kelib chiqib*, *tadqiqot natijalari shuni ko'rsatadiki* should be rendered by academic English equivalents: it should be noted that, based on the above, the results of the study show that. These are not idioms, but they are stable discourse patterns. Their inclusion in the corpus is justified because they guide the structure of scientific argumentation.

6. Contextual Validity. The sixth criterion is contextual validity. A stable combination may have more than one possible equivalent, and the corpus must reflect this variation. The Uzbek ko'z yumdi may mean physical closing of the eyes in a literal context, but as a stable expression it means ignoring something deliberately. In the second case, the English equivalent is turned a blind eye. Similarly, og'zining suvi keldi may refer to actual appetite or strong desire. In English, his mouth watered works well in food-related contexts, while he longed for it may be more appropriate in non-food contexts.

This criterion is especially important for parallel corpora because the same Uzbek unit may correspond to different English units depending on context. Therefore, equivalence should be recorded not only at the level of dictionary meaning but also at the level of contextual use.

The analysis indicates that stable combinations should not be treated as ordinary word sequences in Uzbek–English corpus construction. Their equivalence depends on several interacting factors: meaning, structure, function, culture, style and context. If the corpus records only surface translation, it cannot provide reliable data for linguistic research or translation practice.

For instance, the pair Yetti o'lchab, bir kes – Look before you leap is not based on lexical similarity. The Uzbek proverb contains the image of measuring and cutting; the English proverb contains the image of looking and jumping. Yet both express the same practical wisdom: one should think carefully before acting. Such cases prove that translation equivalence must be determined at the level of communicative and conceptual meaning.

Similarly, oq oltin may be rendered as white gold in contexts where figurative naming is preserved, but as cotton in contexts where clarity is more important than imagery. Therefore, the parallel corpus should allow for more than one equivalent where the function of the expression changes.

Collocations also require careful treatment. In Uzbek–English translation practice, errors often occur when a collocation is translated literally. The expression kuchli yomg'ir corresponds to heavy rain, not strong rain; e'tibor qaratmoq corresponds to pay attention, not turn attention; imkoniyat yaratmoq corresponds to create an opportunity, not make a possibility. Such units demonstrate that stable combinations are not always figurative, but they are still important for accurate translation.

The same is true of terminological and discourse units. In academic texts, korpus lingvistikasi, parallel korpus, tarjima muqobiligi, semantik yaxlitlik and struktur barqarorlik should have stable English equivalents: corpus linguistics, parallel corpus, translation equivalence, semantic integrity and structural stability. The consistency of these equivalents is essential for scholarly communication.

The study confirms that determining translation equivalence of stable combinations is one of the central theoretical tasks in building an Uzbek–English parallel corpus. Stable combinations include not only phrasemes but also proverbs, sayings, periphrases, collocations, terminological combinations, speech etiquette formulas and discourse patterns. Each group requires its own equivalence criteria.

The article proposes six main criteria for selecting English equivalents: semantic integrity, functional adequacy, structural correspondence, cultural relevance, stylistic compatibility and contextual validity. On this basis, several types of equivalence are distinguished: full figurative equivalence, image-based equivalence, functional equivalence, paremiological equivalence, cultural equivalence, descriptive equivalence, collocational equivalence, terminological equivalence, pragmatic equivalence and discourse equivalence.

The findings show that translation equivalence cannot be reduced to lexical identity. In many cases, the correct equivalent is determined by meaning, function, context and cultural value rather than by formal similarity. Therefore, the Uzbek–English parallel corpus of stable combinations

should be designed as a linguistically informed resource that records not only source and target forms but also the nature of their interlingual correspondence.

The proposed approach may be used in compiling bilingual dictionaries, developing translation memory resources, improving machine translation systems and teaching Uzbek–English translation through corpus-based materials. It may also serve as a theoretical foundation for further research on the classification, contextual use and corpus representation of stable combinations.

REFERENCES

1. Cowie A.P. (ed.). *Phraseology: Theory, Analysis, and Applications*. – Oxford: Oxford University Press, 1998. – 272 p.
2. Haagsma H., Bos J., Nissim M. MAGPIE: A Large Corpus of Potentially Idiomatic Expressions // *Proceedings of the 12th Language Resources and Evaluation Conference*. – Marseille: European Language Resources Association, 2020. – P. 279–287.
3. Koehn P. Europarl: A Parallel Corpus for Statistical Machine Translation // *Proceedings of Machine Translation Summit X*. – Phuket, 2005. – P. 79–86.
4. Nurullayeva O.N. O‘zbek va ingliz tillaridagi frazemalarning parallel korpusini yaratish asoslari. *Filologiya fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi*. – Toshkent, 2026. – 137 b.
5. Sag I.A., Baldwin T., Bond F., Copestake A., Flickinger D. *Multiword Expressions: A Pain in the Neck for NLP // Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science*. – Berlin: Springer, 2002.
6. Savary A., Ramisch C., Cordeiro S., Sangati F., Vincze V., QasemiZadeh B., Candito M., Cap F., Giouli V., Stoyanova I., Doucet A. *The PARSEME Multilingual Corpus of Verbal Multiword Expressions // Multiword Expressions at Length and in Depth*. – Berlin: Language Science Press, 2018. – P. 87–147.
7. Saxena P., Paul S. *EPIE Dataset: A Corpus for Possible Idiomatic Expressions*. – arXiv:2006.09479, 2020.
8. Tiedemann J. *Parallel Data, Tools and Interfaces in OPUS // Proceedings of the Eighth International Conference on Language Resources and Evaluation*. – Istanbul: European Language Resources Association, 2012. – P. 2214–2218.
9. Donaj G., Antloga Š. *ParaDiom: A Parallel Corpus of Idiomatic Texts*. – CLARIN.SI Repository, 2022.
10. *Oxford Dictionary of Idioms*. 3rd ed. – Oxford: Oxford University Press, 2009.
11. *Cambridge Idioms Dictionary*. 2nd ed. – Cambridge: Cambridge University Press, 2006.
12. Rahmatullayev Sh., Mahmudov N., Xolmanova Z., O‘razova I., Rixsiyeva K. *O‘zbek tili frazeologik lug‘ati*. – Toshkent, 2022.