E-ISSN: 2997-9439

American Journal of Education and Evaluation Studies

https://semantjournals.org/index.php/ AJEES

| Research Article

Check for updates

Alignment of Word Combinations in Uzbek and English

Matyakubova Noila Shakirjanovna
Doctoral student of TSUULL
matyakubovanoila@navoiy-uni.uz

Abstract: In various fields of natural language processing (NLP), especially in machine translation, achieving accurate alignment of parallel texts requires a deep understanding of the expressions, grammatical structures, and semantics in both the source and target languages. Although Uzbek and English share a structural similarity in their use of head and subordinate words within phrases, English uniquely features compound word combinations as well. In Uzbek, word combinations are classified based on the grammatical properties and structures of their components. Specifically, the classification relies on the part of speech of the dominant element and the syntactic role of the subordinate element. During alignment, differences arise due to the grammatical characteristics of the words making up the phrases in each language and their syntactic positions within sentences.

Keywords: alignment, word combination, tokenization, source language (SL), target language (TL), grammatical structure, semantics.

Introduction.

Word combinations are groups of words that convey specific meanings within a sentence and are distinguished from single words by their distinct grammatical structure. In many areas of natural language processing (NLP), particularly in machine translation, accurately aligning parallel texts requires a thorough understanding of the expressions, grammatical structures, and semantics in both the source and target languages [1]. When the grammatical frameworks of two languages differ significantly, substantial structural and semantic differences can arise. English and Uzbek are examples of such languages, and these differences can create challenges in achieving proper alignment [2]. Compiling a comprehensive set of Uzbek-English word combinations and utilizing machine learning techniques can enhance the performance of alignment tools, machine translation systems, and various educational resources that work with parallel corpora.

Structure of word combinations.

In Uzbek, word combinations are categorized based on the grammatical properties and structure of their components. Specifically, classification by grammatical nature considers the part of speech of the dominant element and the syntactic role of the subordinate element [3]. Uzbek word combinations are divided into stable and free forms, with free combinations further classified

based on whether they are formed through coordination or subordination. Structurally, they closely resemble English word combinations, consisting of a head word and a dependent word [4]. In both languages, the overall phrase is typically named after the part of speech of the head word. Structurally, word combinations are referred to as simple or complex in both Uzbek and English; however, English additionally features compound combinations, which Uzbek lacks. English phrases include various types such as noun phrases, verb phrases, adjective phrases, adverbial phrases, prepositional phrases, verbal phrases, absolute phrases, phrasal verbs, collocations, and idioms. In contrast, Uzbek phrases are generally categorized into noun, verb, adjective, adverbial, and modal combinations. When translating parallel texts, word combinations in the source language (SL) may appear either as phrases or as separated elements in the target language (TL) [5]. This phenomenon often introduces complexities and challenges in the alignment process.

The process of aligning English and Uzbek word combinations presents a complex set of challenges that stem from deep grammatical, semantic, and structural differences between the two languages (See: Fig. 1). One of the primary difficulties lies in the way phrases are formed and function within sentences. English relies heavily on prepositions, strict word order, and compound structures to build meaning, while Uzbek depends largely on case endings, suffixes, and postpositions, allowing for more flexible sentence construction [6]. As a result, a direct one-to-one mapping between English and Uzbek word combinations is often impossible.

Table 1. Alignment of English-Uzbek Word Combinations

Type	English word combination	Uzbek equivalent	Notes on alignment
Noun phrase (NP)	a black cat	qora mushuk	Adjective-noun order is reversed; structure is similar.
Verb phrase (VP)	is running	yugurmoqda	Uzbek uses a single verb form with suffixes instead of auxiliary + verb.
Adjective phrase	full of hope	umidga to‘la	Semantic equivalence, but structural shift due to postposition.
Adverbial phrase	very quickly	juda tez	Almost direct alignment; both use intensifier + adverb.
Prepositional phrase	on the table	stol ustida	English uses preposition; Uzbek uses postposition via suffix.
Compound noun	toothbrush	tish cho‘tka(si)	Compound noun in English becomes a noun + noun phrase in Uzbek.
Phrasal verb	give up	voz kechmoq	No verb-preposition equivalent; translation requires a completely different verb.
Idiomatic expression	break the ice	Munozarani yengillashtir-moq	Literal translation fails; requires idiomatic or descriptive equivalent.
Modal combination	must go	borishi kerak	Uzbek uses verb + modal suffix, not auxiliary verb.
Stable collocation	make a decision	qaror qabul qilmoq	Verb-noun collocation differs lexically and grammatically.
Free word combination	interesting book	qiziqarli kitob	Direct alignment; same adjective-noun structure.
Absolute phrase	weather permitting	ob-havo ruxsat etsa	Requires subordinate clause in Uzbek, no exact syntactic equivalent.

In English, many phrases are tightly fixed, such as phrasal verbs (look after, give up) and idiomatic expressions (break the ice, spill the beans), where the meaning cannot be derived from the individual words. Translating or aligning these to Uzbek often requires expanding them into longer, more descriptive constructions, potentially disrupting the compactness of the original English phrase. Moreover, compound nouns in English, which combine multiple concepts into a single lexical unit (e.g., snowstorm, sunlight), have no exact syntactic equivalent in Uzbek and typically need to be paraphrased, sometimes leading to loss of nuance.

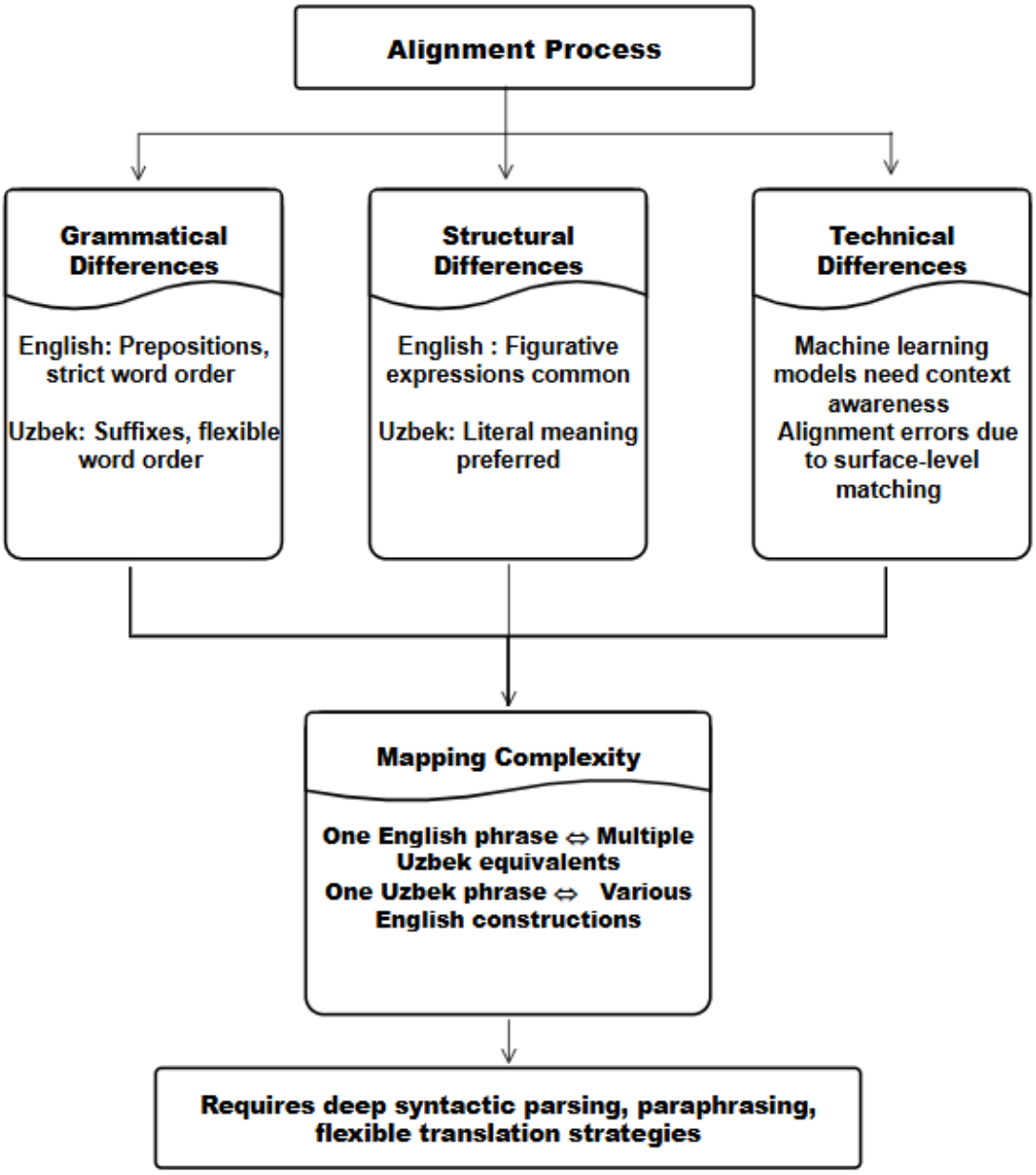

Fig. 1. Challenges in English-Uzbek Word Combination Alignment

One of the critical challenges is that in Uzbek, the syntactic relationship within word combinations is often indicated by morphological markers rather than fixed word order, whereas in English, positional relationships are key. This difference means that during alignment, it is not enough to match words based on lexical meaning; the grammatical functions and dependencies must also be analyzed and reconstructed accurately in the target language. Furthermore, in the case of stable collocations and idioms, English often favors metaphorical or figurative phrasing [7], while Uzbek may use more literal expressions, making semantic preservation particularly

difficult. Additionally, the nature of free and stable combinations differs between the two languages: while English makes frequent use of relatively rigid, memorized phrases, Uzbek tends to allow freer construction based on syntactic rules, adding another layer of complexity to alignment.

While aligning parallel texts, an English word combination may correspond not to a single Uzbek phrase, but to multiple possible renderings depending on the context, intended nuance, and grammatical role within the sentence. Conversely, Uzbek free word combinations may map onto different types of English phrases depending on stylistic choices or syntactic needs. Another complication arises when Uzbek modal word combinations, which express necessity, possibility, or obligation, are aligned with English modal verb constructions, requiring not just lexical but functional correspondence. These structural, grammatical, and semantic disparities often necessitate sophisticated alignment strategies that go beyond surface-level matching and involve deep syntactic parsing, context-sensitive translation, and sometimes even rephrasing or restructuring to preserve meaning.

Machine translation systems and alignment tools, when dealing with English-Uzbek parallel corpora, must therefore be trained not only on direct translation pairs but also on flexible patterns of paraphrase, syntactic transformation, and semantic equivalence, which can vary widely even within a single document [8,9]. In many cases, errors in alignment occur precisely because automatic systems fail to account for these deeper grammatical and functional distinctions, especially when encountering collocations, idiomatic expressions, and compound constructions unique to English. Consequently, building a robust alignment model between English and Uzbek requires an extensive database of phrase-level correspondences, context-aware machine learning techniques, and nuanced linguistic rules that bridge the structural and conceptual gaps between the two languages.

Conclusion.

The alignment of English and Uzbek word combinations is a linguistically intricate task that goes far beyond simple word-to-word translation. The grammatical, structural, and semantic disparities between the two languages introduce significant complexity, particularly when dealing with idiomatic expressions, phrasal verbs, compound structures, and modal constructions. These challenges highlight the need for sophisticated, context-sensitive alignment strategies that integrate morphological analysis, syntactic parsing, and semantic modeling. Relying solely on surface-level or lexical matching is insufficient for achieving accurate and meaningful alignment. Therefore, to build effective English-Uzbek alignment systems, whether for machine translation, bilingual corpus development, or NLP tools which is crucial to incorporate rich linguistic resources, deep learning techniques, and a nuanced understanding of both languages' phraseological systems. Only through such comprehensive approaches can alignment systems preserve both the form and function of word combinations across linguistic boundaries.

References

1. MacCartney B., Galley M., Manning Ch. "A Phrase-Based Alignment Model for Natural Language Inference". Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing. 2008.
2. Matyakubova N., Xamroyeva Sh., Dauletov A., Mengliyev B., Adali E. "Algorithm of Creating The "Uzbek-English Aligner" Program". 6.IEEE - UBMK-2023 VIII. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı - XII. Pp 114-118. 2023.
3. Abdullayev F.A. "Grammar of the Uzbek language". Vol. 2, Syntax. pp18-39. 1976.
4. Brinton L. J. "The Structure of Modern English" John Benjamins Publishing Company Amsterdam /Philadelphia.2000.

5. Arase Y., Tsujii J., “Compositional Phrase Alignment and Beyond”. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, November 16–20, 2020.
6. Abdullayeva O., Xudayarova S. “The problem of definition, description and classification of the phrase in Uzbek linguistics”. International scientific-theoretical conference on the topic: «Problems of research and education of the Uzbek language».
7. Aarts B., Haegeman L., “English Word Classes and Phrases”, Bas Aarts and Liliane Haegeman. January 2008.
8. Nwet K. Th. “Developing Word to Phrase Alignment for Myanmar-English Machine Translation”, 13th International Conference on Computer Applications 2015.
9. Sennrich R., Martin V. “Iterative, MT-based sentence alignment of parallel texts.” In: NODALIDA 2011, Nordic Conference of Computational Linguistics, Riga, 11 May 2011.