ISSN 1799-2591 (print); ISSN 2053-0692 (online)
Theory and Practice in Language Studies, Vol. 16, No. 8, pp. 2670-2680, August 2026
DOI: https://doi.org/10.17507/tpls.1608.13

Linguistic Foundations for Automatic Simplification of Kazakh Texts

Manshuk Mambetova
Department of Turkology and Language Theory, Al-Farabi Kazakh National University, Almaty, Kazakhstan

Akmaral Karymkhan*
Department of Turkology and Language Theory, Al-Farabi Kazakh National University, Almaty, Kazakhstan

Balnur Nurlangazykyzy
Department of Turkology and Language Theory, Al-Farabi Kazakh National University, Almaty, Kazakhstan

***Abstract*—Universal and multilingual large language models cannot correctly simplify Kazakh texts in accordance with internal linguistic patterns. Using LLM without a linguistically sound simplification model leads to uncontrolled generation and increases the risk of lexical and grammatical errors. Therefore, there is a need to create formalized linguistic patterns for Kazakh text simplification, which will facilitate the creation of automatic simplification systems within LLM. This can be achieved through the analysis of manually adapted materials. A corpus of 20 Kazakh texts was annotated by the authors for the analysis, each presented in three versions such as the original (O) and two successive adaptation levels (A1 and A2). At the A1 stage, syntactic transformations predominate, while A2 demonstrates lexical stabilization. The text size is comparable at all levels, which ensures control and reliability of the analysis. The process is formalized through a two-level tagging system: lexical (LS) and syntactic simplification (SS) that records each operation. Frequency analysis revealed a limited core of dominant techniques, supplemented by peripheral layers, forming structured polyoperational architecture. Transitions from A1 to A2 demonstrate a process shift toward lexical changes while maintaining discursive coherence. The results emphasize the importance of linguistic expertise, operation typologies, and manual annotation for the interpretability and quality of systems. Hybrid approaches with humans in the loop improve the accuracy and reliability of ATS.**

***Index Terms*—text simplification, lexical simplification, syntactic simplification, ATS**

I. Introduction

Automatic text simplification (ATS) is one of the tasks of NLP which is aimed at creating simplified (simple) version of an original (complex) text through linguistic (mainly on lexical and syntactic level) modifications by reducing its complexity without changing its core ideas and meaning (Shardlow, 2014; Alva-Manchego et al., 2020; Al-Thanyyan & Azmi, 2021). The primary goal is to improve the accessibility of complex texts to a wider readership (Devaraj et al., 2021).

Simplified texts are especially in demand in teaching foreign languages. The effectiveness of their use in FLT is supported by several studies (Crossley et al., 2012; Coxhead & Boutorwick, 2018). The main obstacle of using authentic texts in teaching reading is their high textual complexity – lexical, syntactic and discursive. Besides this, in international practice, simplified text is successfully used not only for educational purposes (Crossley et al., 2012), but also for adaption of reading materials for children (De Belder & Moens, 2010; Kajiwara et al., 2013), processing scientific and specialized texts (Ermakova et al., 2022), adaptation of hard-to-read materials for people with various cognitive impairments (Carroll et al., 1998; Rello et al., 2013; Kover et al., 2014; Evans et al., 2014; Chen et al., 2017), and for people with a low level of functional literacy in their native language (Watanabe et al., 2009).

This is evidenced by the development of websites and parallel corpora focused on the tasks of ATS in different languages (*ASSET, Newsela English* for English; *FIRST, Simplext, Newsela Spanish* for Spanish; *SIMPITIKI, PaCCSS-IT* for Italian; *CLEAR, WikiLargeFR* for French; *EasyJapanese, Jados* for Japanese; *PorSimples* for Brazilian Portuguese, *Simple German, TextComplexityDE* for German; *The Corpus of Basque Simplified Texts (CBST)* for Basque; *DSim* for Danish, *SimplifyUREval* for Urdu, *RuWikiLarge, RuSimpleSentEval, RuAdapt* for Russian; *SloTS* for Slovene; *SAMER, Saaq al-Bambuu* for Arabic (Ryan et al., 2023). However, this process is still being done manually by language experts, teachers and linguists in low-resource languages, such as Kazakh. This is due to the limited number of language models, annotated parallel corpora and linguistic systems developed considering typological specifics of the Kazakh language. Even a specialist in the field of language teaching has a very low speed of text simplification. This entails large time costs and does not apply to a large number of texts available on the network. Therefore, automation of this process is a task that requires quick solutions.

* Corresponding Author. Email: k.akmaral2309@gmail.com

The Kazakh language belongs to the Turkic language family and is characterized by an agglutinative type of system. In the absence of formalized algorithms for ATS for the Kazakh language, it is impossible to assume that universal and multilingual large language models will be able to correctly simplify Kazakh texts in accordance with internal linguistic patterns without distorting the content. Using LLM without a linguistically sound simplification model leads to uncontrolled generation and increases the risk of lexical and grammatical errors.

The experience of high-resource languages shows that the emergence of neural network and LLM-oriented systems was preceded by extensive linguistic work: parallel corpora of original and simplified texts were manually created, simplification rules were formalized, and readability levels were developed. Based on these resources, rule-based automatic simplification systems were formed, followed by neural network models.

Thus, for the Kazakh language, the primary task is to create linguistically sound corpora and simplification models, without which the implementation of LLM-oriented ATS systems cannot be effective and reliable.

This topic is now dynamically developing in Kazakh linguistics. In general, the advantage of simplified texts has been pedagogically proven in numerous research papers, which indicates the need to form a system of adapted texts in teaching the Kazakh language (Mambetova et al., 2024). Research on Kazakh parallel corpus for machine translation "KazParC" (Yeshpanov et al., 2024) and generation models for paraphrasing (Kassenkhan et al., 2024) show that transformational models for the modern Kazakh language are able to effectively generate and paraphrase only under the condition of special linguistic and technical adaptation. The first major project that proved the technical possibility of ATS of the Kazakh language within NLP is the KazSim (Toleu et al., 2025). The authors of the project created the first parallel corpus of 8,709 pairs of simple and complex sentences using heuristic selection and LLM augmentation. This demonstrated the possibility of automating Kazakh-language materials. However, such projects are largely data-driven and lack explicit language simplification model, characterized by a model-adapted system which still shows that ATS requires linguistic selection and justification. To do this, the text for the Kazakh language requires the formalization of lexical, morphological and syntactic formations necessary for simplification, and a description of the level of complexity. Therefore, this paper aims to systematize the main general linguistic mechanisms of transformation for ATS by analyzing manually annotated corpora.

II. Literature Review

From a linguistic perspective, text simplification should be understood as a complex, hierarchically organized process affecting several interconnected levels of linguistic structure, including lexicon, syntax, discourse, and pragmatics. At the vocabulary level, simplifying transformations involve replacing low-frequency or conceptually loaded units with more common equivalents, using hypernymic generalizations, or introducing clarifying constructions, which are particularly significant in texts with pronounced terminological density (Koptient et al., 2019).

Syntactic simplification is implemented through a number of operations, among which the key ones include the transformation of passive structures into active ones, segmentation of complex sentences, reduction of nested subordination constructions, and reorganization of word order to increase the transparency of syntactic relations (Brunato et al., 2022).

At the discourse level, coreference resolution mechanisms, adaptation of the discourse marker system, and ensuring the continuity of information flow both within and between paragraphs play a central role, directly impacting the coherence of a simplified text (Devaraj et al., 2021). The pragmatic dimension of simplification, in turn, requires consideration of genre specifics and the characteristics of the addressee: for example, texts aimed at children allow and even require proximization strategies, including direct addressing of the reader, whereas in encyclopedic genres, such techniques are considered inappropriate (Bott & Saggion, 2014).

These linguistic principles underlie modern typologies of simplification operations, which continue to be used as a theoretical and methodological basis for manual data annotation and evaluation of the effectiveness of ATS systems.

The development of ATS has gone through several stages: from rule-based systems to statistical models, and then to neural and LLM architectures. Early approaches relied on linguistic typologies and detailed language analysis, focusing primarily on syntactic transformations such as the removal of subordinate constructions and secondary components (Dras, 1999), but their application was limited.

With the transition to statistical and neural methods, the role of explicit linguistic interpretation has significantly weakened. Modern neural models (seq2seq, T5, BART) are trained end-to-end on parallel corpora, without directly modeling the simplification operations (Martin et al., 2020). Even in controlled CEFR-labeled scenarios, simplification often relies on superficial parameters such as sentence length and lexical frequency, without taking into account the deep structure of the text (Flores et al., 2023; Maddela & Alva-Manchego, 2025).

LLMs, including ChatGPT, are capable of simplification without specialized training (Kumar et al., 2020), but they are susceptible to systemic errors such as confusing simplification with summarization, excessive abbreviations, repetitions, and the addition of non-existent information (Vásquez-Rodríguez et al., 2023). These problems arise from the lack of explicit linguistic control, particularly at the discourse and pragmatic levels.

Fully ATS systems still face a number of persistent challenges. First and foremost, these are semantic distortions, for instance, even modern models do not always preserve the original meaning, especially when working with texts containing complex semantic relationships (Agrawal & Carpuat, 2024). An additional threat is "hallucinations" which is

the addition of information not present in the original, making such systems unsafe for use in sensitive fields, including medicine and law (Devaraj et al., 2021).

Another significant drawback is their ignorance of the discourse structure of the text. Focusing on individual sentences without considering the global context leads to reduced coherence, repetitions, and the loss of logical transitions (Sun et al., 2021; Vásquez-Rodríguez et al., 2023).

Finally, automatic systems are poorly tailored to the needs of specific audiences. Readability metrics rely on superficial indicators and poorly reflect actual text comprehension, while different reader groups require different simplification strategies, limiting the practical applicability of universal solutions (Tanprasert & Kauchak, 2021).

Linguists have developed typologies of simplification operations used for manual annotation and analysis. These approaches contrast with line editing, where text is processed as a sequence of tokens with formal operations of addition, deletion, and retention (Alva-Manchego et al., 2017; Cardon & Grabar, 2020). In contrast, linguistic typologies describe simplification as the transformation of structurally and functionally significant units, including operations such as specification and voice change (Bott & Saggion, 2014; Grabar & Saggion, 2022).

Such typologies ensure interpretable and reproducible annotation, clear instructions improve the consistency of annotators and allow simplifications to be explained from a linguistic perspective. Linguistic expertise is also important for quality control of corpora, as manual annotation takes into account grammatical correctness and semantic coherence, reducing noise in the data.

An important contribution of linguists is the development of evaluation protocols focused on error analysis rather than superficial metrics. Unlike BLEU, SARI, and BERTScore, such approaches identify specific violations such as incorrect deletions, additions, paraphrases, and ungrammatical statements as well as semantic losses, which are detected using text comprehension tests (Yamaguchi et al., 2023).

Linguistic expertise is also necessary for the multilingual adaptation of simplification systems. Studies show that simplification metrics and strategies vary significantly across languages, and many operations are language- and genre-specific, making direct transfer of English-language models ineffective (Gonzalez-Dios et al., 2018; Koptient et al., 2019; Scholz & Wenzel, 2025). Therefore, linguists adapt typologies of operations and annotation schemes to specific languages.

Finally, hybrid approaches combining automated methods with human intervention are considered promising. Incremental simplification and training on small, linguistically annotated datasets improve the manageability, coherence, and quality of results, while reducing semantic and discourse biases (Grabar & Saggion, 2022; Kumar et al., 2020).

Modern research shows that hybrid architectures that combine automatic methods with linguistically informed human intervention offer the greatest potential for ATS. For example, the progressive document simplification (ProgDS) approach replicates expert strategies and processes text step by step, at the discourse, thematic, and lexical levels, increasing the manageability of transformations and preserving coherence (Kumar et al., 2020). Training on small datasets annotated by linguists significantly improves the understanding of LLMs and the quality of simplification. Human intervention also enables iterative refinement of results and the reduction of semantic and discourse biases (Grabar & Saggion, 2022).

However, fundamental challenges remain in the field of automatic simplification, directly related to the insufficient role of linguistic expertise. First, document-level simplification quality assessment remains underdeveloped. Existing metrics, including D-SARI, are sensitive to text length and poorly capture semantics and coherence, while coherence models do not always distinguish between medium and low levels of quality (Lai & Tetreault, 2018; Vásquez-Rodríguez et al., 2023). This highlights the need for linguistically motivated metrics that take into account both local and global text structure.

Second, research remains predominantly English-centric and rarely considers cultural and linguistic differences in the perception of simplicity. Overcoming these differences requires an analysis of language and culture specific simplification strategies. Third, current systems offer little support for personalization and poor adaptation to the user's language proficiency and cognitive characteristics. Finally, automatic linguistic annotation remains unstable and requires subsequent review, necessitating linguistically informed post-editing procedures (Yamaguchi et al., 2023).

Automated linguistic annotation remains a challenging task: LLMs are unstable when tagging simplification operations and require subsequent validation (Yamaguchi et al., 2023). This necessitates linguistically sound post-editing and quality control procedures to create reliable and scalable corpora.

To sum up, this review demonstrates that linguistic expertise is not an intermediate step, but rather the foundation of automated text simplification. Linguists are developing a theoretical understanding of simplification as a multi-level process, developing operation types, assessment protocols, and ensuring cross-linguistic adaptation. The limitations of fully automatic approaches such as meaning distortion, loss of coherence, and weak audience orientation are directly related to the lack of explicit linguistic control.

Research on hybrid systems confirms that the integration of linguistic instructions, small-data learning, and iterative feedback improves the quality and reliability of results. Further development requires recognition that textual simplicity is a cognitively and culturally conditioned category and cannot be reduced to formal metrics.

III. Research Materials and Corpus Construction

The practical part of the study is based on a specially developed corpus focused on studying the mechanisms of manual text simplification. 20 texts, each presented in three versions (original, A1, A2) from collections No. 1 and No. 2 of "Adapted texts in Kazakh" (2024) compiled by Mambetova M. and Kuzembekova Zh. were selected as the corpus material. The source texts are characterized by a high degree of lexical and syntactic complexity and generally correspond to advanced levels of language proficiency, approximately to C1–C2.

The corpus is characterized by genre diversity and includes works of fiction, educational materials, and news texts. This combination allows us to consider the influence of genre specificity on the selection and implementation of simplification strategies. The adapted reflect a gradual reduction in complexity. As a result, each trio of texts forms a sequential chain of controlled simplification, creating conditions for analyzing changes on both a qualitative and quantitative level.

Table 1
Manually Annotated Corpus Characteristics

Parameter	O (original text)	A1	A2
Number of texts	20	20	20
Number of total sentences	572	518	591
Average number of sentences per text	27	26	30
Number of total tokens	7516	4 138	5 900
Average number of tokens per text	376	207	295

The total size of the corpus is several thousand words and is distributed unevenly between the versions (see Table 1). The original texts retain the largest volume, while the simplified versions demonstrate a gradual reduction due to lexical and structural compression. This organization of the material allows us to trace the dynamics of the transformation of text parameters at different stages of adaptation. The manual preparation of the simplified versions guarantees their grammatical correctness and pedagogical relevance, ensuring the reliability of the corpus as an empirical basis for subsequent testing and evaluation of ATS methods.

IV. Results and Discussions

A. Linguistic Model of Simplification: Alignment and Operation Types

Types of structural alignment in Kazakh simplified texts

It was mentioned above that ATS is detected based on corpus analysis. The central stage of such an analysis is alignment, since without establishing semantic correspondences between the original and simplified versions, it is impossible to formalize exactly *what is being changed* and *how.* Alignment is used to correlate text segments and then annotate basic simplification operations.

In several English-language studies based on the Newsela corpus, sentence alignment procedures were introduced, on the basis of which the main types of correspondences were described (1→1, 1→N, N→1, etc.) (Xu et al., 2015; Xu et al., 2016). Štajner and Saggion (2018) proposed a typology of simplification operations and implemented alignment as the basis for building a simplification model. In the context of the presence of developed parallel corpora (complex – simplified versions), alignment in English practice acts primarily as a technical stage and is used for automatic training of simplification models on ready-made data.

This study uses the opposite – linguistically motivated paradigm. Since there are currently no ready-made parallel corpora for the Kazakh language, simplified texts are not extracted from existing resources but are formed manually. In other words, a structural simplification model was built through manual markup, after which a system of alignment and linguistic simplification mechanisms were formalized.

To implement this procedure, all 20 texts, both original and simplified versions, were segmented into separate sentences and numbered (O1, O2, O3..., A1_1, A1_2..., A2_1, A2_2...). Further, these sentences were combined into semantic pairs representing carriers of the same semantic unit. The analysis of the changes occurring in such pairs allowed us to identify typical sequences of operations characteristic of the simplification of the Kazakh language.

The original and simplified versions were considered as one semantic pair, subject to the following criteria:

1. The key is the coincidence of the predicate, type and participants of the situation in the correlated variants.

2. Structural – the presence of lexical and/or syntactic changes in the simplified version (in the absence of structural changes, the transformation is interpreted as a rewrite, not as a simplification);

3. Partial overlap – even with structural changes, removal of individual components, or addition of explanations, at least one semantic core must be preserved.

A total of 572 sentences were disturbed in the source text, in A1 – 518, in A2 – 591, from which 480 semantic pairs were formed. Correlation of the number of sentences of the original and simplified texts included in one semantic pair allowed us to systematize the types of alignment for the Kazakh language (see Table 2).

As shown in Table 2, a total of six types of alignment were identified:

1. 1→1 – deriving one sentence in a simplified version from one sentence in the original text;
2. 1→N – deriving two or more sentences in a simplified version from one sentence of the original text;
3. N→1 – deriving one sentence in a simplified version from two or more sentences of the original text;
4. N→N – deriving two or more sentences in a simplified version from two or more sentences of the original text;
5. 1→0 – deletion of one sentence of the original text in a simplified version;
6. N→0 — deletion of two or more sentences of the original text in a simplified version.

The frequency of these types for the two levels is expressed as a percentage relative to the total number of pairs. These data are used to identify dominant and auxiliary strategies in the Kazakh language simplification model.

TABLE 2
STRUCTURAL ALIGNMENT TYPES

No	Alignment type	O → A1 (%)	O → A2 (%)
1	1→1	47.9%	62.7%
2	1→N	15%	11.4%
3	N→1	2.5%	3.1%
4	N→N	8.1%	9.5%
5	1→0	23.3%	12%
6	N→0	2%	0%

In both simplified variants, the predominant form of alignment is the 1→1 structure (A1 – 47.9%, A2 – 62.7%), which indicates a tendency to maintain the "one sentence = one sentence" correspondence during the simplification process. The higher frequency of this type in version A2 indicates a greater degree of preservation of the original sentences at this level compared to A1. This difference is considered as one of the fundamental indicators of the gradation of simplification levels in the Kazakh language.

The second typical feature is that type 1→0 (deletion) occurs at the A1 level almost twice as often as at the A2 level (A1 – 23.3%, A2 – 12%), which indicates a more intensive reduction of content at the initial simplification level. Thus, A1-level texts are formed primarily by actively reducing content, rather than by reconstructing structures.

An additional indicator for the A1 level is the higher frequency of type 1→N (division), reflecting the tendency to split complex syntactic constructions into several simple sentences. Types N→1 and N→0 are characterized by low frequency and, although they are not dominant, perform an auxiliary function in the simplification structure.

The alignment of changes according to the world practice standard allowed us to more clearly define the linguistic mechanisms and their frequency for the Kazakh language, this can also serve as a basis for automation.

B. Linguistic Simplification Mechanisms

While the types of alignment in the simplification process make it possible to determine what happens to sentences, the next step is to describe and explain exactly how simplification is implemented within these structures. In this regard, further analysis is aimed at identifying internal language operations. In this context, the operation is considered as a minimal linguistic action in the simplification process. Such operations are implemented at two levels – lexical and syntactic, which function in interrelation.

The simplification system for levels A1 and A2 is formalized based on quantitative frequency indicators of operations performed at these levels. The combination of frequency indicators of these operations forms a linguistic model that can be used in the training of automatic simplification systems.

In the framework of this study, 2 stable frequency zones are identified: 1) frequent; 2) rare and unstable operations. Accordingly, from this we deduce the model into two main zones, which can be interpreted as core and peripheral. This gradation forms the basis for adapting the simplification system while maintaining the semantic integrity of the text.

For a reader with an insufficient level of language proficiency when reading the source text, unfamiliar or rare words are the main obstacle to understand the text. In addition, variable word meanings, contextual dependence, and idiomatic expressions also make it difficult to interpret the text. Lexical changes are aimed at reducing cognitive and semantic load and are achieved through reworking the vocabulary. Along with lexical complexity, sentence structure (subordinate complex sentences, complex grammatical constructions) also requires simplification. These are syntactic operations.

In our simplification process, we grouped 7 lexical and 9 syntactic transformation operations (see Table 3). Data annotation was carried out manually by linguists, each operation was assigned an internal tag (L1, L2 ...). The following analysis focuses on their distribution and functional role within the model.

TABLE 3
FREQUENCY OF LINGUISTIC OPERATIONS

Operation type	Operation tag	Meaning	Tag frequency in A1 simplification	Tag frequency in A2 simplification	O → A1 (%)	O → A2 (%)
Lexical	L1	Replacement of an abstract word with concrete	7	2	1.4%	0.4%
	L2	Replacement of word with its widely known synonym	103	143	21.4%	29.7%
	L3	Complex word deletion	220	107	45.8%	22.2%
	L4	Explanatory word addition	4	63	0.8%	13.1%
	L5	Removal of specialized terminology	9	10	1.8%	2%
	L6	Reference clarification	3	2	0.6%	0.4%
	L7	Adding linking (cohesive) elements	17	33	3.5%	6.8%
Syntactic	S1	Sentence splitting	107	88	22.9%	18.3%
	S2	Sentence merging	12	8	2.5%	1.6%
	S3	Word reordering	1	2	0.2%	0.4%
	S4	Subordinate structures simplification	1	2	0.2%	0.4%
	S5	General syntax simplification without an explicit SPLIT / MERGE operation.	188	231	39.1%	48.1%
	S6	Proposition deletion	121	51	25.2%	10.6%
	S7	Tense simplification	3	4	0.6%	0.6%
	S8	Replacement of derived forms with non-derived forms	1	1	0.2%	0.2%
	S9	Complex case form simplification	3	2	0.6%	0.4%

At the lexical level, the most frequent changes for A1 level are the deletion of a compound word (45.8%) and the replacement of a word with a synonym (21.4%). For level A2, these types of lexical simplification also remain the leading ones, however, in the reverse hierarchy. When comparing the levels, the deletion of a complex word at the A1 level is carried out almost twice as often as at the A2 level (45.8% → 22.2%). This indicates that at the A2 level most of the complex words are preserved and replaced by alternative units, whereas at the A1 level there is strongly pronounced lexical reduction. This is also evidenced by the higher relative frequency of synonymous substitution operations at the A2 level (29.7%).

Another significant inter-level difference is frequent use of the lexical addition with an explanatory function at the A2 level (13.1%), whereas at the A1 level this operation is rarely used (0.8%). As an auxiliary lexical operation, the addition of connecting (cohesive) elements (3.5% and 6.8%, respectively) is highlighted, which play an important role in ensuring the coherence of the text (for example, the addition of words such as *therefore, that's why* (in Kazakh), etc.).

At the syntactic level, the most frequent operation for both levels is the general simplification of the syntactic structure of a sentence (A1 – 39.1%, A2 – 48.1%), which includes changes that are not limited solely to the mechanisms of division and unification. The next most frequent operation is the splitting of a complex sentence into several simple ones (A1 – 22.9%, A2 – 18.3%).

The A1 level is also characterized by a high frequency of proposition deletion (25.2%), whereas at the A2 level this indicator is significantly lower (10.6%). This reflects the tendency to delete fragments containing secondary information, because they do not contribute to the core propositional content of the text. The remaining types of syntactic operations belong to the peripheral layer of the model and perform an auxiliary function in the process of manual simplification.

The conclusion following from the data obtained is that for the Kazakh language, lexical reduction and syntactic reduction at the A1 level form a single reductive mechanism, and at the A2 level, lexical preservation and syntactic restructuring form a conservation-reorganization model.

C. Empirical Implementation of the Simplification Model

1 Lexical transformations

The most common type of lexical simplification in the transition from the original to level A1 was the deletion of complex words (L4), accounting for 45.8% of all lexical changes. This operation is aimed at reducing semantic complexity for beginning language learners.

For example, in the educational text "The Solar System" the original sentence O8 was simplified to level A1 by removing complex words such as *«едәуір, эклиптика, жазықтығына, көлбеу»* (Eng.: "significantly", "ecliptic", "plane" and "inclination"). These words were deemed too difficult for beginner learners to understand (see Table 4).

TABLE 4
COMPLEX WORD DELETION (L4)

Pair ID	Sentence ID	Text name	Sentence
08	Original 8 = A1 (8) + A1 (9) = A2 (9)	The Solar System (kaz.: Күн жүйесі)	O8 Барлық кіші планеталар да үлкен планеталар қозғалған бағытта Күнді айнала қозғалады, бірақ олардың орбиталары едәуір созылыңқы және эклиптика жазықтығына көлбеу орналасады. *Eng: All the small planets also orbit the Sun in the same direction as the larger planets, but their orbits are significantly elongated and located obliquely to the ecliptic plane*
			A2 (9) Барлық кіші планеталар Күнді айнала қозғалады, бірақ олардың орбиталары созылыңқы және эклиптика бетіне қиғашынан орналасады. *Eng: All the smaller planets move around the Sun, but their orbits are elongated and located obliquely to the surface of the ecliptic.*
			A1 (8) Барлық кіші планеталар Күнді айнала қозғалады. A1 (9) Олардың орбиталары біраз созылыңқы. *Eng: All the smaller planets revolve around the Sun. Their orbits are slightly elongated.*
429	Original 456 = A1 (407) = A2 (455)	Boy Chasing a Bird (kaz.: Құс қуған бала)	0456 Құнанына сүйеніп тұрған Сәтжанның аяғына тұмсығын сүйей Бөрібасар да ентігін баса алмай келіп шөкесінен жатты. *Eng: Boribasar, unable to catch his breath, came up to Satzhan's feet, leaning against his legs.*
			A2 (455) Құнанына сүйеніп тұрған Сәтжанның аяғына тұмсығын сүйей Бөрібасар да тізерлеп жатты. *Eng: Boribasar came to Satzhan's feet, leaning against his legs.*
			A1 (407) Сәтжанның қасына Бөрібасар тізерлеп жатты. *Eng: Boribasar was kneeling next to Satzhan.*

Similarly, in the fictional text “Boy Chasing a Bird”, the phrases «*сүйеніп тұрған, тұмсығын сүйей, ентігін баса алмай, шөкесінен»* (Eng.: “leaning against the beak” and “breathing heavily”, “kneeling down”) in sentence O456 were removed during simplification, allowing the original 16-word sentence to be reduced to 5-words while preserving the main plot action (see Table 4).

These examples clearly demonstrate how removing complex words effectively reduces the lexical complexity of a text without significantly losing information.

Replacing words with more common synonyms (L2) was the most common lexical technique used to simplify texts to level A2, accounting for 29.7% of all lexical changes. This method aims to adapt the text to the lexical competence of readers at this level, ensuring the use of words and expressions familiar and common in everyday speech.

For example, in the text “The Boy Chasing a Bird” the word *«елең алу»* (to startle) in sentence O415 was replaced with the more familiar *«есту»* (to hear), as the former is rare in everyday speech and may be difficult for beginning readers to understand. Similarly, the expression «*аяқ бауын тартқылады*» (Eng.: “inclined to fly”) was transformed into *«ұшқысы келді»* (Eng.: “wanted to fly”) as the original wording is typical of literary texts and does not meet the standards of level A1. Such substitutions improve the text's accessibility, simplify comprehension, and simultaneously preserve the key semantic content (see Table 5).

TABLE 5
REPLACEMENT OF WORD WITH ITS WIDELY KNOWN SYNONYM (L2)

Pair ID	Sentence ID	Text name	Sentence
397	Original 415 = A1 (368) = A2 (415)	Boy Chasing a Bird (kaz.: Құс қуған бала)	O415 Қаз дыбысынан елең алғандай ителгісі талпынып, тұғырынан жерге қонып, аяқ бауын тартқылады. *Eng: As if because of the sound of a goose, the falcon stood up from the pedestal to the ground and pulled on his shoelaces.*
			A2 (415) Қаз дыбысынан елең еткен ителгісі талпынып, аяқ бауын тартқылады. *Eng: The sound of goosebumps made his falcon twitch and pull on her shoelaces.*
			A1 (368) Сәтжанның ителгісі қаздың дауысын естігеннен кейін тезірек ұшқысы келді. *Eng: Satzhan's falcon wanted to fly faster after hearing the goose's voice.*
267	Original 274 = A1 (264) = A2 (246)	Totan Batyr (kaz.: Тотан батыр)	O274 Күресте, ат бәйгесінде, алтын қабақ атуда, жаяу жарыста озып шықсын. *Eng: May he succeed in wrestling, horse racing, shooting gold, and hiking.*
			A2 (246) Күресте, ат жарыста, садақ атуда, жаяу жарыста озып шықсын. *Eng: Let him excel in wrestling, horse racing, archery, and foot racing.*
			A1 (264) Күресте, ат жарыста, садақ атуда, жаяу жарыста мені жеңіп шықсын. *Eng: May he beat me in wrestling, horse racing, archery, and hiking.*

In a number of cases, synonym substitution included replacing obsolete or rarely used words with modern, commonly used equivalents, making the text more accessible to modern readers and corresponding to an A2 level of linguistic proficiency. For example, in the text "Totan Batyr", the old expression «*алтын қабақ»* (Eng.: archaic version of "archery"), which is virtually unused in modern language, was replaced with the modern version «*садақ ату»* ("archery").

A less frequently used text simplification technique for levels A1 and A2 is reference clarification (L6) for rarely encountered or historical words (A1 – 0.4%; A2 – 0.6%). For example, in the work "I was ashamed" by the prominent Kazakh writer Baubek Bulkyshev, there are words that require clarification for adequate comprehension by students. For example, the original sentence

«Мен де ыстық күнде күпі, басыма бөрік киіп, белімді буынып алып жүруге әбден үйреніп алдым»

(Eng: I, too, got used to the fact that on a hot day I wear a hat, a beret on my head and carry it around my waist)

has been simplified with the addition of explanations for complex lexical items indicated using numerical references:

«Мен де ыстық күнде күпі[1], басыма бөрік[2] киіп жүруге үйреніп алдым» (*1-Түйенің немесе қойдың жүнінен жасалған қалың ұзын сырт киім. 2-Қалың бас киім*).

(*Eng: I also got used to wearing a hat[1] on a hot day and a beret[2] on my head" (1 – thick long outerwear made of camel or sheep wool. 2 – thick headdress).*

Reference clarification was also used to preserve cultural authenticity of the material. For instance, during the adaptation of the fairy tale "Totan Batyr" special attention was paid to preserving culturally specific terms related to Kazakh tradition. While simplifying the sentence

Жолына ақсарбас[1] шалып аттандырып, бірнеше күншілік жерден жолдастары кері қайтады.

(Eng: Before setting off, the "aksarbas" sacrificial ceremony was performed, and the friends who conducted it returned a few days later)

the word *«ақсарбас»* was not simplified or replaced with a more common expression; instead, it remained in the text and was explained in footnotes as *1- Береке, бақыт үшін берілетін садақа, құрбандық (мал сойып, ас беру)* (Eng.: Sacrificing animals for better times and providing foods for others, charity) to clarify its meaning and cultural context. This strategy ensures an understanding of cultural realities while simultaneously preserving the educational value of the text for language learners, demonstrating how adaptation can combine simplification with cultural awareness.

2 Syntactic transformations

Syntactic transformations affected the organization of the utterance and were aimed at reducing formal complexity. The most common operations were the splitting complex syntactic constructions into simpler ones, as well as the merging of sentences when this ensured a more linear and predictable structural unfolding. Additionally, word order reorganizations were observed with a focus on the canonical model, the reduction of subordinate constructions, and a general simplification of syntactic structure without explicit splitting or merging. In some cases, sentences not essential to the main content were excluded, verb tenses were simplified, derived morphological forms were replaced with non-derivative ones, and complex case realizations were reduced.

The most common syntactic adaptation strategy involves simplifying syntactic complexity without explicit sentence splitting or merging (S5). This type of transformation accounts for 39.1% of all syntactic changes at level A1 and increases to 48.1% at level A2. This approach involves simplifying the internal structure of a sentence through more straightforward syntactic design, eliminating overloaded constructions, and reducing the utterance to a more transparent form, which improves the readability of the text while maintaining its semantic integrity.

Splitting complex sentences into several simpler statements (S1) is a key syntactic simplification technique when adapting texts for levels A1 and A2, accounting for 22.9% of changes at level A1 and 18.3% at level A2. For instance, in the tale "Totan batyr", a 10-word sentence is not technically very long. However, the specific syntactic organization and the way the components of the utterance are connected significantly complicate its comprehension on first reading. Therefore, during the adaptation process, a decision was made to split this construction into two independent sentences. This transformation increased the syntactic transparency of the text, facilitated decoding of semantic connections, and brought the utterance structure in line with the cognitive abilities of readers at level A2 (see Table 6).

TABLE 6
SENTENCE SPLITTING (S1)

Original	A2
O236 Бірақ сонша алыс жолға ешбір адам жолдас болып баруға жарамайды. (Eng.: But no one is suitable to accompany him on such a long journey.)	A2 (206) Күнікейге баратын жол өте алыс болыпты. (Eng.: The road to Kunikei was very long.) A2 (207) Сондықтан Тотан өзімен бірге баратын адамды таба алмады. (Eng.: Therefore, Totan could not find anyone to go with him.)

Along with sentence division, the reverse strategy of sentence merging (S2) is also used in the syntactic adaptation process. However, according to the data obtained, this technique is used extremely rarely, accounting for 2.5% of changes at level A1 and 1.6% at level A2. The limited prevalence of this operation is explained by the specific nature of simplification at the initial levels, where priority is given to reducing syntactic load and linearity of structure rather than increasing its complexity through declarative enlargement.

TABLE 7
SENTENCE MERGING (S2)

Original	A1
O86 Бір тұсында олар тым ұсақ: 70 сантиметрдің о жақ, бұ жағында ғана. O87 Ал енді бірінде өте ірі – мүйізінің өзі 4 метрден асады. *Eng: In one place, they are quite small — only about 70 centimeters. And in the other, they are very large: the horns alone exceed 4 meters.*	A1 (88) Олар кейбір суреттерде өте кіші, ал басқаларында өте үлкен. *Eng: In some images they are very small, while in others they are very large.*

As an illustration, consider the example from the text "Petroglyphs: A World Heritage Entrusted to Our People" where two original sentences were combined into one. The merging is not mechanical, syntactic simplification is accompanied by lexical simplification, resulting in the combined construction being a simple sentence with a transparent structure. This approach demonstrates that merging is only possible with parallel simplification of vocabulary and syntax, which maintains the text's accessibility for language learners at levels A1 and A2 (see Table 7).

V. CONCLUSION

Aligning a text with a target proficiency level requires substantial processing, known as simplification, which prior to the introduction of LLMs was carried out mainly manually by linguists and language specialists.

The article defines the main simplification mechanisms detected at A1 and A2 levels for Kazakh. This was determined by solving tasks consisting of several stages: 1) the theoretical framework of the ATS process was defined, the principles of text simplification were identified based on the study of researchers and various corpora. This is important because the Kazakh language has not been studied enough for the ATS, namely, there are no purely linguistic simplification mechanisms; 2) 20 texts in Kazakh were manually adapted to two simple levels and divided into sentences, which were then combined into semantic pairs depending on their semantic correspondence (original → A1 → A2); using these semantic pairs, the types of alignment structures and the frequency of linguistic operations were extracted, which are used to systematize and formalize rules and model for both levels. The analysis revealed that manual simplification implemented through the coordinated application of lexical and syntactic strategies. These strategies aim to reduce linguistic and cognitive complexity while preserving the core meaning and communicative function of the source texts. Sequential adaptation from the original to A1 and A2 ensures a controlled reduction in complexity and allows for the identification of stable patterns of linguistic transformation. This data can then be integrated into automatic simplification formalizations. This research paper can become a source on the Kazakh language, capable of filling theoretical and methodological gaps in NLP, LLM and ATS issues. Given that Kazakh is still considered a low-resource language, our study formalizes these operations as a multilevel linguistic simplification model.

REFERENCES

[1] Agrawal, S., & Carpuat, M. (2024). *Do Text Simplification Systems Preserve Meaning? A Framework for Human Evaluation Using Reading Comprehension. Transactions of the Association for Computational Linguistics*, *12*, 432–448. https://doi.org/10.1162/tacl_a_00653

[2] Al-Thanyyan, S. S., & Azmi, A. M. (2021). Automated text simplification: A survey. *ACM Computing Surveys (CSUR*), *54*(2), 1–36. https://doi.org/10.1145/3442695

[3] Alva-Manchego, F., Bingel, J., Paetzold, G. H., Scarton, C., & Specia, L. (2017). Learning how to simplify from explicit labeling of complex–simplified text pairs. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing* (Vol. 1, pp. 295–305). Asian Federation of Natural Language Processing. Retrieved January 16, 2026, from https://aclanthology.org/I17-1030/

[4] Alva-Manchego, F., Scarton, C., & Specia, L. (2020). Data-driven sentence simplification: Survey and benchmark. *Computational Linguistics*, *46*(1), 135–187. https://doi.org/10.1162/coli_a_00370

[5] Bott, S., & Saggion, H. (2014). Text Simplification Resources for Spanish. *Language Resources and Evaluation*, *48*(1), 93–120. https://doi.org/10.1007/s10579-014-9265-4

[6] Brunato, D., Dell'Orletta, F., & Venturi, G. (2022). Linguistically-Based Comparison of Different Approaches to Building Corpora for Text Simplification: A Case Study on Italian. *Frontiers in Psychology*, *13*, 1–19. https://doi.org/10.3389/fpsyg.2022.707630

[7] Cardon, R., & Grabar, N. (2020). French biomedical text simplification: When small and precise helps. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 710–716). International Committee on Computational Linguistics. https://doi.org/10.18653/v1/2020.coling-main.62

[8] Carroll, J., Minnen, G., Canning, Y., Devlin, S., & Tait, J. (1998). Practical simplification of English newspaper text to assist aphasic readers. In *Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology* (pp. 7–10). Association for the Advancement of Artificial Intelligence. Retrieved January 16, 2026, from https://www.researchgate.net/publication/2740075_Practical_Simplification_of_English_Newspaper_Text_to_Assist_Aphasic_Readers

[9] Chen, P., Rochford, J., Kennedy, D. N., Djamasbi, S., Fay, P., & Scott, W. (2017). Automatic text simplification for people with intellectual disabilities. *Artificial Intelligence Science and Technology*, 725–731. https://doi.org/10.1142/9789813206823_0091

[10] Crossley, S. A., Allen, D., & McNamara, D. (2012). Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research*, *16*(1), 89–108. https://doi.org/10.1177/1362168811423456

[11] Coxhead, A., & Boutorwick, T. J. (2018). Longitudinal vocabulary development in an EMI international school context: Learners and texts in EAL, maths, and science. *TESOL Quarterly*, *52*(4), 1000–1023. https://doi.org/10.1002/tesq.450
[12] De Belder, J., & Moens, M. (2010). Text simplification for children. In *Proceedings of the SIGIR Workshop on Accessible Search Systems* (pp. 19–26). Retrieved January 16, 2026, from https://www.researchgate.net/publication/228736550_Text_simplification_for_children
[13] Devaraj, A., Marshall, I., Wallace, B., & Li, J. (2021). Paragraph-Level Simplification of Medical Texts. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4972–4984). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.395
[14] Dras, M. (1999). *Tree adjoining grammar and the reluctant paraphrasing of text* [Doctoral thesis, Sydney: Macquarie University].
[15] Ermakova, L., SanJuan, E., Kamps, J., Huet, S., Ovchinnikova, I., Nurbakova, D., Ara´ujo, S., Hannachi, R., Mathurin, E., Bellot, P. (2022). Overview of the CLEF 2022 SimpleText Lab: Automatic simplification of scientific texts. *Lecture Notes in Computer Science*, *13390*, 470–494. https://doi.org/10.1007/978-3-031-13643-6_28
[16] Evans, R., Orasan, C., & Dornescu, I. (2014). An evaluation of syntactic simplification rules for people with autism. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations* (pp. 131–140). Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-1215
[17] Flores, L., Huang, H., Shi, K., Chheang, S., & Cohan, A. (2023). Medical text simplification: optimizing for readability with unlikelihood training and reranked beam search decoding. *Computer science*, *1*. https://doi.org/10.48550/arXiv.2310.11191
[18] Gonzalez-Dios, I., Aranzabe, M. J., & Diaz de Ilarraza, A. (2018). The corpus of Basque simplified texts (CBST). *Language Resources and Evaluation*, *52*(1), 217–247. https://doi.org/10.1007/s10579-017-9407-6
[19] Grabar, N., & Saggion, H. (2022). Evaluation of Automatic Text Simplification: Where Are We Now, Where Should We Go From Here. *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles*, *1*, 453–463. Retrieved January 16, 2026, from https://aclanthology.org/2022.jeptalnrecital-taln.47/
[20] Kajiwara, T., Matsumoto, H., & Yamamoto, K. (2013). Selecting proper lexical paraphrase for children. In *Proceedings of the 25th Conference on Computational Linguistics and Speech Processing (ROCLING 2013)* (pp. 59–73). ACLCLP. Retrieved January 16, 2026, from https://aclanthology.org/O13-1007/
[21] Kassenkhan, A. M., Mukazhanov, N. K., Nuralykyzy, S., & Kalpeyeva, Z. B. (2024). Text generation models for paraphrase on Kazakh language. *Vestnik KazUTB*, *1*(22). https://doi.org/10.58805/kazutb.v.1.22-249
[22] Kover, S. T., Haebig, E., Oakes, A., McDuffie, A., Hagerman, R. J., & Abbeduto, L. (2014). Sentence comprehension in boys with autism spectrum disorder. *American journal of speech-language pathology*, *23*(3), 385–394. https://doi.org/10.1044/2014_AJSLP-13-0073
[23] Koptient, A., Cardon, R., & Grabar, N. (2019). Simplification-induced transformations: Typology and some characteristics. In *Proceedings of the 18th BioNLP Workshop and Shared Task* (pp. 309–318). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-5033
[24] Kumar, D., Mou, L., Golab, Ł., & Vechtomova, O. (2020). Iterative edit-based unsupervised sentence simplification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 7918–7928). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.707
[25] Lai, A., & Tetreault, J. (2018). Discourse Coherence in the Wild: A Dataset, Evaluation and Methods. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue* (pp. 214–223). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-5023
[26] Martin, L., Fan, A., de la Clergerie, É., Bordes, A., & Sagot, B. (2020). MUSS: Multilingual unsupervised sentence simplification by mining paraphrases. *Computation and Language.* https://doi.org/10.48550/arXiv.2005.00352
[27] Maddela, M., & Alva-Manchego, F. (2025). Adapting Sentence-Level Automatic Metrics for Document-Level Simplification. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1, pp. 6444–6459). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.naacl-long.327
[28] Mambetova, M., Karymkhan, A., & Abdikadyrova, A. (2024). Adapting kazakh texts for language learners: application of adaptation techniques. *Eurasian Journal of Philology. Science and Education*, *195*(3), 58–67. https://doi.org/10.26577/EJPh.2024.v195.i3.ph06
[29] Rello, L., Bayarri, C., Górriz, A., Baeza-Yates, R., Gupta, S., Kanvinde, G., Saggion, H., Bott, S., Carlini, R., & Topac, V. (2013). DysWebxia 2.0! More accessible text for people with dyslexia. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility* (pp. 1–2). Association for Computing Machinery. https://doi.org/10.1145/2461121.2461150
[30] Ryan, M. J., Naous, T., & Xu, W. (2023). Revisiting non-English text simplification: A unified multilingual benchmark. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 4898–4927). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.269
[31] Scholz, K., & Wenzel, M. (2025). Evaluating readability metrics for German medical text simplification. In *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 6049–6062). Association for Computational Linguistics. Retrieved January 16, 2026, from https://aclanthology.org/2025.coling-main.405/
[32] Shardlow, M. (2014). A Survey of Automated Text Simplification. *International Journal of Advanced Computer Science and Applications*, *4*, 58–70. https://doi.org/10.14569/SpecialIssue.2014.040109
[33] Sun, R., Jin, H., & Wan, X. (2021). Document-level text simplification: Dataset, criteria and baseline. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 7997–8013). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.630
[34] Štajner, S., & Saggion, H. (2018). Data-Driven Text Simplification. In *Proceedings of the 27th International Conference on Computational Linguistics: Tutorial Abstracts* (pp. 19–23). Association for Computational Linguistics. Retrieved January 16, 2026, from https://aclanthology.org/C18-3005/
[35] Toleu, A., Tolegen, G., & Ualiyeva, I. (2025). Fine-tuning large language models for Kazakh text simplification. *Applied Sciences*, *15*(15), 8344. https://doi.org/10.3390/app15158344

[36] Tanprasert, T., & Kauchak, D. (2021). Flesch-Kincaid is not a text simplification evaluation metric. In *Proceedings of the First Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)* (pp. 1–14). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.gem-1.1

[37] Vásquez-Rodríguez, L., Shardlow, M., Przybyła, P., & Ananiadou, S. (2023). document-level text simplification with coherence evaluation. In *Proceedings of the Second Workshop on Text Simplification, Accessibility and Readability* (pp. 85–101). Retrieved January 16, 2026, from https://aclanthology.org/2023.tsar-1.9/

[38] Watanabe, W. M., Junior, A. C., Uzêda, V. R., Fortes, R. P. de M., Pardo, T. A. S., & Aluísio, S. M. (2009). Facilita: Reading assistance for low-literacy readers. In *Proceedings of the 27th ACM International Conference on Design of Communication* (pp. 29–36). Association for Computing Machinery. https://doi.org/10.1145/1621995.1622002

[39] Xu, W., Callison-Burch, C., & Napoles, C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, *3*, 283–297. https://doi.org/10.1162/tacl_a_00139

[40] Xu, W., Napoles, C., Pavlick, E., Chen, Q., & Callison-Burch, C. (2016). Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, *4*, 401–415. https://doi.org/10.1162/tacl_a_00107

[41] Yamaguchi, D., Miyata, R., Shimada, S., & Sato, S. (2023). Gauging the gap between human and machine text simplification through analytical evaluation of simplification strategies and errors. In *Findings of the Association for Computational Linguistics: EACL 2023* (pp. 359–375). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-eacl.27

[42] Yeshpanov, R., Polonskaya, A., & Varol, H. A. (2024). KazParC: Kazakh parallel corpus for machine translation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (pp. 9633–9644). ELRA and ICCL. https://doi.org/10.48550/arXiv.2403.19399

Manshuk M. Mambetova — Candidate of Philological Sciences, Associate Professor at the Department of Turkology and Language theory, faculty of Philology, Al-Farabi Kazakh National University. Her research interests include text adaptation, teaching Kazakh as a foreign language, lexicology, and language theory.

Akmaral A. Karymkhan* — PhD student of the educational program "8D02303 — Linguistics" at the Department of Turkology and Language theory, faculty of Philology, Al-Farabi Kazakh National University. Her doctoral research focuses on developing a linguistic model for automatic text simplification.

Balnur Nurlangazykyzy — Lecturer at the Department of Turkology and Language theory, faculty of Philology, Al-Farabi Kazakh National University. Her research interests include text adaptation and foreign language teaching (FLT, ESL).