



ӘУЕ КӨЛІГІ ЖӘНЕ ТЕХНОЛОГИЯЛАР
ВОЗДУШНЫЙ ТРАНСПОРТ И ТЕХНОЛОГИИ
AIR TRANSPORT AND TECHNOLOGY

UDC 81'33:656.7

IRSTI 73.37.01; 16.21.33

https://doi.org/10.53364/24138614_2026_42_3_4

A. Koshekov^{1*}, Y. Kuanov¹

¹Civil Aviation Academy, Almaty, Kazakhstan

*E-mail: a.k.koshekov@gmail.com

DIGITAL TRANSFORMATION OF PROFESSIONAL AVIATION
MULTILINGUAL COMMUNICATIONS: A CORPUS-BASED STUDY

Abstract. Digital transformation is influencing linguistic communication within the aviation industry; however, linguistic data remains limited. This study examines these changes based on 120 publicly available trilingual texts related to the aviation sector, comparing two time periods: 2018–2019 and 2024–2025. Preprocessing revealed cross-linguistic similarities. Text processing was carried out using a rule-based, Unicode-aware pipeline. Lexical richness was assessed using the MATTR-50 metric, while sentence length and digital terminology were analyzed via five conceptual categories and document-level variance analysis. Independent sample assessment involved Mann-Whitney U-tests with the Holm correction applied across the entire text. Variance analysis of terminology was conducted using Fisher's tests. The results indicate that sentence length decreased in both Russian and English—though the changes remained statistically significant after the Holm correction—whereas no changes were observed in Kazakh. While digital terminology was present in the earlier period, modern terms place greater emphasis on digital concepts such as AI, modeling, and learning.

Keywords: digital transformation, aviation discourse, multilingual corpus, digital aviation terminology, professional standardization.

Introduction.

The digital transformation of the aviation industry represents a modern approach to the advancement of scientific research. Currently, the implementation of digital systems involves the use of AI, modeling, and decision-support tools across virtually all critical aspects of aviation [1]. These changes are transforming not only the technological processes themselves but also the specialized discourse through the description, regulation, training, and coordination of professional activities. This study examines these linguistic shifts as a consequence of the industry's intensive digitalization.

The aviation industry is one of the most suitable fields for such research, given that professional communication within it is standardized and multilingual. The international aviation environment encompasses shared concepts, globally used abbreviations, and regulated terminology. English serves as the language of international aviation communication, establishing requirements for safety, precision, brevity, and mutual understanding [2]. The International Civil Aviation Organization (ICAO) mandates a high level of English proficiency for flight crews and air traffic controllers [3]. Corpus-based studies of aviation English reveal specific functional and linguistic characteristics, as well as a procedural function [4].

However, aviation communication is not limited to radio exchanges. Professional communication also encompasses news, safety-related messages, aeronautical documentation,

regulatory documents, and educational matters. Regarding adherence to national standards, most English-language terms are adopted through loanwords, abbreviations, and linguistic adaptation. A study conducted in the Republic of Kazakhstan on the interplay of trilingual terminology revealed that establishing equivalence and standardization across these languages is challenging [5].

Consequently, Kazakhstan offers a convenient multilingual environment for professional aviation communication. Kazakh and Russian serve as the languages of national professional communication, while English is used for professional training, radio communication, and global terminology. National regulations permit trilingual radio communication. This linguistic configuration creates an environment in which the same process can be represented using different linguistic structures [6].

Most previous studies have focused primarily on the use of aviation English, radiotelephony phraseology, proficiency assessment, training strategies, and terminology. For instance, Bieswanger (2016) identifies specialized aviation registers in his work [7], while Ishihara and Prado (2021) highlight English as a language critical to safety [8]. Subsequently, the research focus shifted toward the study of specialized vocabulary, the development of specific practical tasks, and the analysis of multilingual document structures. Drayton and Coxhead (2023) demonstrated the importance of aviation vocabulary in enhancing safety systems through language training methodologies [9]. Pacheco et al. (2023) showed the feasibility of teaching aviation English using authentic corpora and outlined the limitations involved in compiling such corpora [10]. Carvalho et al. (2023) examined corpus-based methods through the study of multilingual documentation in the context of lexicographical and professional training [11]. Rigaud Peixoto et al. (2023) illustrated how aviation terminology can differ between institutional and specialized literature [12].

A review of the literature relevant to this study reveals a specific gap in knowledge: a lack of corpora that adequately reflect the growing role of digitalization within the multilingual aviation environment. This article addresses that gap, recognizing that digital transformation entails more than merely an increase in the frequency of these terms in texts. The aviation sector was already high-tech prior to the widespread adoption of artificial intelligence; early studies examined automated air traffic control systems, electronic navigation databases, cybersecurity measures, and decision-support systems. Contemporary developments are characterized by the reorientation of these technologies toward the use of artificial intelligence.

Another issue addressed in this article is the use of direct translation—specifically, identical identifiers—on the multilingual websites of official bodies. Analyzing such texts can result in an inflated measure of interlingual similarity. To address this challenge, the study proposes distinguishing between the diachronic analysis of a single language and descriptive methods of cross-linguistic comparison, while also incorporating a check for dependencies between identifiers and events.

The aim of this study is to identify and quantify diachronic changes in trilingual aviation discourse - specifically those linked to the digital transformation of the aviation industry at the institutional level. To achieve this aim, the researchers address the following objectives:

1. Assessing lexical and structural changes characteristic of institutional aviation discourse across two time periods: 2018–2019 and 2024–2025.
2. Assessing changes in the distribution of digital aviation terminology and technical concept groups for each language.
3. Assessing structural and terminological trends within the trilingual corpora.
4. Assessing indicators of the international standardization of aviation terminology within institutional trilingual discourse.

This study combines an analysis of diachronic structural indicators with an assessment of the variance of concept groups, incorporating controls for multiple comparisons and group

variance. To this end, it employs multiple comparison controls and examines dependencies influenced by translation quality and specific events.

Materials and research methods.

The study employs a comparative diachronic corpus analysis method. The research material consists of texts related to professional aviation activities from two time periods: 2018-2019 and 2024-2025. These periods were selected to reflect the onset of the industry's digital transformation and its current stage of development. The texts were compiled into three language datasets: KZ, RU, and EN. Based on this design, six sub-corpora were created: KZ-1, KZ-2, RU-1, RU-2, EN-1, and EN-2. The final collection comprises 120 aviation articles - specifically, 20 articles per sub-corpus. The total size of the corpus is determined as follows:

$$N = L \times P \times C \times n, \quad (1)$$

where L is the number of languages, P represents time intervals, C denotes thematic categories, and n is the number of documents per category.

The target metric is the number of documents rather than the volume of text. This is primarily due to the fact that the institutional texts are limited to corporate news or professional informational updates; consequently, word count is treated merely as an empirical characteristic of the corpus, while statistical analysis relies on metrics calculated for each individual document.

The dataset was sourced from the public archives of RSE "Kazaeronavigatsia." A single information source was deliberately chosen to ensure a unified publication environment, thereby eliminating the influence of varying editorial policies across different institutional bodies. This approach also facilitated the balancing of language samples based on temporal and thematic criteria. Nevertheless, the results obtained cannot be automatically generalized to the entire spectrum of aviation communication in Kazakhstan.

To mitigate the impact of thematic variations, the composition of each sub-corpus was balanced across four thematic categories: air traffic management and navigation; flight safety and regulatory compliance; professional training; and digital aviation technologies. Five documents were selected for each category, resulting in a total of 20 texts per time period and language.

Corpus quality assessment was conducted via a two-stage verification process. In the first stage, parallel records sharing identical identifiers were identified and marked by comparing identifiers across languages. The second stage involved a manual review of events; this helped identify events belonging to the specified time period that nonetheless had unique identifiers for each language. Thus, it can be concluded that identical identifiers do not equate to statistical data independence.

Due to the parallelism of events across language samples, direct statistical tests could not be employed in the final analysis. The unit of inference in the model is the document, and statistical analysis is based on comparing early and late periods for each sub-corpus. Convergences and divergences are assessed descriptively using models of direction, magnitude, and scope.

Articles were stripped of non-textual page elements, and a deterministic rule pipeline was applied to them. A token was defined as a contiguous sequence of letters and digits, preserving hyphens and apostrophes, though individual punctuation marks were removed. Sentence boundaries were determined by terminal punctuation marks. A unified set of rules was applied to all three languages to avoid dependencies on inconsistent language-specific taggers or morphological analyzers. Lexical diversity was assessed using the Moving-Average Type-Token Ratio (MATTR), a metric less sensitive to text length than traditional measures [13]. MATTR calculations were based on case-insensitive word surface forms, using a 50-token window and a step size of one token.

$$TTR_i = \frac{V_i}{w}. \quad (2)$$

The final MATTR value is calculated as follows:

$$MATTR_{50} = \frac{1}{k} \sum_{i=1}^k \frac{V_i}{50}. \quad (3)$$

If a document's length fell below this threshold, the type-to-token ratio calculated for the entire sequence was used for segmentation. This rule applies across the entire range of languages. Average length is calculated as the ratio of the number of tokens to the total number of segmented sentences. Surface word forms are used as lemmas to avoid inconsistencies in lemmatization quality.

Given the morphological structure of the languages, MATTR values were interpreted with caution. For instance, the nature of Kazakh morphology can lead to increased variation in surface forms. Consequently, conclusions are drawn based on diachronic changes within a single language, rather than by comparing MATTR scores across languages.

Statistical metrics were calculated at both the document and sub-corpus levels. For each level, MATTR-50 values and average sentence lengths were compared using the Mann-Whitney test. Effect size was determined via rank-biserial correlation, noting the observed direction of change across the two time periods. Result robustness was verified using the Holm correction. To ensure model transparency, p-values adjusted by the Holm correction were used.

An analysis of terminological variance compared the proportions of documents present in the early and late periods for each language set using Fisher's tests. In the analysis of concept families, no p-values exceeding 0.05 were observed; consequently, the findings are interpreted as descriptive shifts in terminology. Direct Kruskal-Wallis tests were excluded from the final analysis because preliminary results indicated partial cross-linguistic dependency.

Results and their discussions

The experiment is built around a data set including 120 professional aviation texts taken from the official resource of the RSE “Kazaeronavigatsiya”. The six sub-corpora are balanced and have the same number of documents, as well as thematic composition. Preliminary analysis showed partial cross-language parallelism. As a result, the conclusions are limited to the framework of intralingual diachronic comparison.

As a result of the tokenization, the data set is presented in Table 1. It also presents the main structural indicators: MATTR-50 and the results of the Mann-Whitney test (without corrections) for each of the subcorpora. After introducing the Holm correction into the model, the parameters did not retain statistical significance. As a result, the data obtained indicate a change in lexical diversity within the subcorpora. The values of the MATTR indicator do not reflect the large lexical diversity of the subcorpora due to different morphological structures.

Table 1 - Structural characteristics of subcorpora

Language	Period	Tokens	MATTR-50	Mean sentence length
English	2018-2019	4 266	0.740	29.91
English	2024-2025	5 009	0.758	24.90
Russian	2018-2019	3 301	0.886	27.19
Russian	2024-2025	4 821	0.896	19.93
Kazakh	2018-2019	1 715	0.903	18.03
Kazakh	2024-2025	2 578	0.915	16.90

Figure 1 illustrates the shift in the directions of diachronic changes.

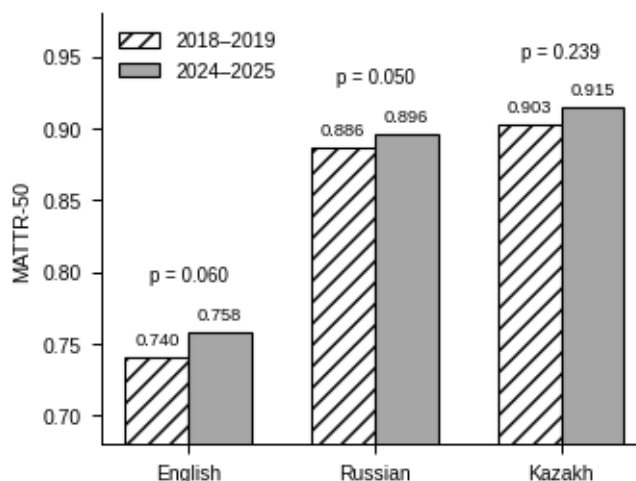


Figure 1 – Diachronic variation in MATTR-50 for multilanguage aviation discourse

MATTR-50 shows an increase across all language corpora. However, it is worth noting that the magnitude of change and the statistical significance vary for each of them. For the Russian language, the p-value is quite close to the standard threshold, whereas for the other languages, the results remained statistically insignificant. Corpus-based studies demonstrate that English possesses a specialized register characterized by specific lexical, functional, and discourse features. This study extends that approach by moving beyond the established research questions surrounding radiocommunication, thereby allowing the structural dynamics of institutional texts to be measured as well.

The most notable change concerns sentence length. For instance, in English, the figure stood at 29.91 words in 2018–2019 but dropped to 24.9 words in the second period. Russian showed a more substantial shift, with the average length falling from 27.19 to 19.93 words, while Kazakh exhibited only minor changes, moving from 18.03 to 16.9 words. Table 2 presents these changes, incorporating additional statistical analysis.

Table 2 – Structural characteristics of subcorpora

Language	Indicator	U-mean	Raw p	Holm p	r_{rb}	Interpretation
English	MATTR-50	130.0	0.060	0.212	0.350	Increase, not robust
English	Mean sentence length	275.5	0.042	0.212	0.378	Decrease, nominal only
Russian	MATTR-50	127.0	0.050	0.212	0.365	Increase, not robust
Russian	Mean sentence length	314.0	0.002	0.013	0.570	Decrease, robust
Kazakh	MATTR-50	156.0	0.239	0.479	0.220	Increase, not significant
Kazakh	Mean sentence length	182.5	0.646	0.646	0.088	Decrease, not significant

Applying the Holm correction affects the scope of the conclusions drawn. The effect of sentence length remains significant for Russian but not for English. The MATTR result for Russian is nominally close to 0.05. Thus, the statistical results are primarily linked to changes in sentence

length in Russian; a similar trend with nominal statistical significance is observed for English, whereas the indicator remains stable for Kazakh.

This shift in sentence structure aligns with the current trend toward brevity in communication; however, the diachronic nature of the study does not account for technological causal factors. Based on the document analysis, it can be concluded that aviation discourse prioritizes precision, clarity, and conciseness. Consequently, the observed linguistic reductions can be interpreted as language changes driven by the extensive adoption of digital technologies. Diachronic changes in sentence length for each corpus are shown in Figure 2.

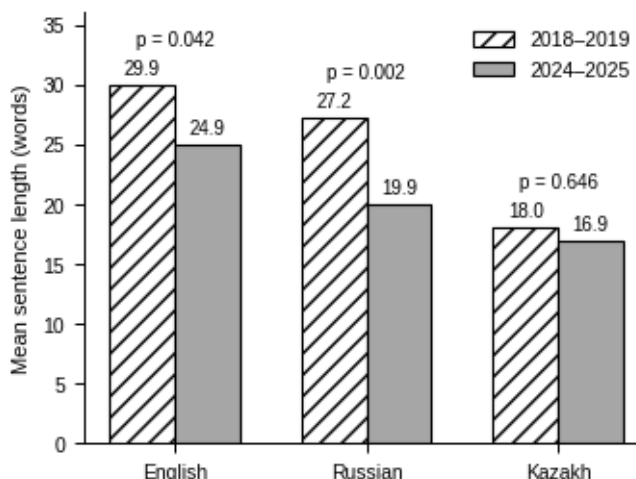


Figure 2 – Mean sentence length for multilanguage aviation discourse

The authors do not perform a final comparative analysis, as partial parallelism was identified during the preprocessing stage. In the descriptive experiment, the sub-corpora retain their initial metrics and exhibit distinct diachronic trajectories. Russian demonstrates a pronounced and consistent trend toward compression; English shows a less distinct direction of change; and Kazakh maintains a stable profile. These differences justify an interpretation based on "selective standardization."

The initial hypothesis regarding the impact of digital transformation on lexical frequency was not confirmed by the experiment. Such terminology was already in widespread use as early as 2018–2019. Aviation English discourse had previously addressed terms such as "automated control systems" and "software modernization." Consequently, digital transformation is understood not as a shift between discourses, but rather as a redistribution and diversification of technologies.

To determine specific characteristics, the input distribution was divided into five conceptual groups. Figure 3 illustrates the distribution of these groups across the sub-corpora, accounting for language-specific features and timeframes.

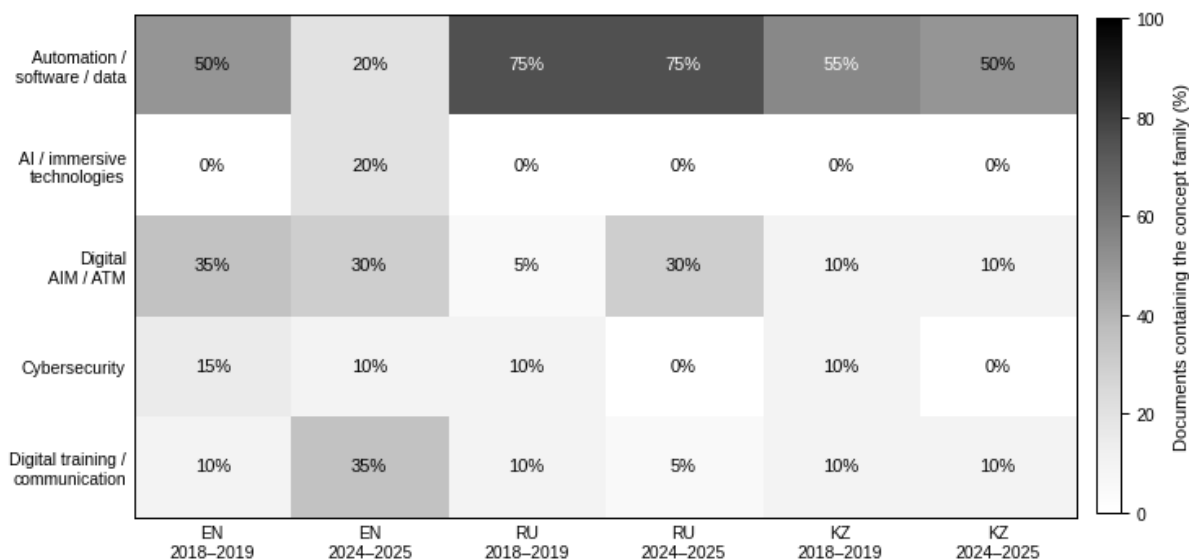


Figure 3 – Document coverage of aviation digital concepts

This visualization demonstrates the early integration of digital concepts into the professional aviation context - for instance, through the application of automation, software, and digital air navigation concepts. Changes in the sub-corpora are presented in Table 3.

Table 3 – Document coverage of aviation concepts

Language	Concept family	2018-19, %	2024-25, %	Change, pp
English	Automation/software/data	50	20	-30
English	AI/immersive technologies	0	20	+20
English	Digital AIM/ATM	35	30	-5
English	Digital training/communication	10	35	+25
Russian	Automation/software/data	75	75	0
Russian	AI/immersive technologies	0	0	0
Russian	Digital AIM/ATM	5	30	+25
Russian	Digital training/communication	10	5	-5
Kazakh	Automation/software/data	55	50	-5
Kazakh	AI/immersive technologies	0	0	0
Kazakh	Digital AIM/ATM	10	10	0
Kazakh	Digital training/communication	10	10	0

The changes shown in Table 3 do not reach the level of statistical significance according to Fisher's test. Consequently, such terminological shifts should be viewed as descriptive variations in usage rather than as changes in specific details.

One such example is the following shift in terminology: in 2024, articles discussing the acquisition of a new simulator use the term "pseudo-pilot" instead of "artificial intelligence." Materials regarding participation in the Airspace World 2024 conference highlight the transformation of the aviation industry through the application of AI and the Skyline-X platform. Professional training utilizing AI technologies incorporates 3D modeling and next-generation simulator systems.

An analysis of the English-language corpus suggests a shift from "digital infrastructure" to "AI." However, the analysis using Fisher's test yielded statistically insignificant results, indicating

a descriptive trend rather than a confirmed shift. This pattern aligns well with recent corpus-based linguistic studies, which demonstrate that technical vocabulary and linguistic data play a pivotal role in training and testing within the aviation sector [14].

An analysis of the Russian language corpus reveals a different picture. Documents concerning automation, software, and data processing account for 75% of the total volume across both time periods, while the scope of digital technology application has expanded from 5% to 30%. Fisher's p-value was 0.091; although this figure did not reach the generally accepted threshold of significance, it points to important changes that warrant verification using a larger language corpus.

The developed model demonstrates the increasing integration of digital technologies into air navigation processes. Such integration is viewed as a form of auxiliary information system. A prime example is Performance-Based Navigation (PBN), which plays a crucial role in the digital transformation of air navigation in Kazakhstan; this involves the use of satellite navigation and the modernization of airspace infrastructure. Thus, the transformation is characterized by a shift from mere digital infrastructure to aviation operations that integrate digital technologies.

The analysis conducted in this study highlights parallel processes in the standardization of international terminology. Various terms and acronyms—such as ICAO, ATM, ATC, and PBN—are increasingly being used in professional discourse within national languages. Research carried out in Kazakhstan demonstrates a correlation between international and national standards and the translated professional terminology [15].

The results obtained in this study do not demonstrate uniformity across linguistic constructs. Although the dataset exhibits signs of cross-linguistic dependency, no statistical analysis of the differences between linguistic distributions was performed. Instead, the study compares the directions and magnitudes of changes identified through diachronic analysis.

The MATTR metric shows an increase across all sub-corpora. This observation aligns with previous studies of aviation communication—though those studies took context into account [16]. The findings also resonate with research on multilingual corpora, where a shared technical domain fosters terminological standardization while respecting language-specific characteristics [11, 12]. The heterogeneous nature of these changes is illustrated in Figure 4.

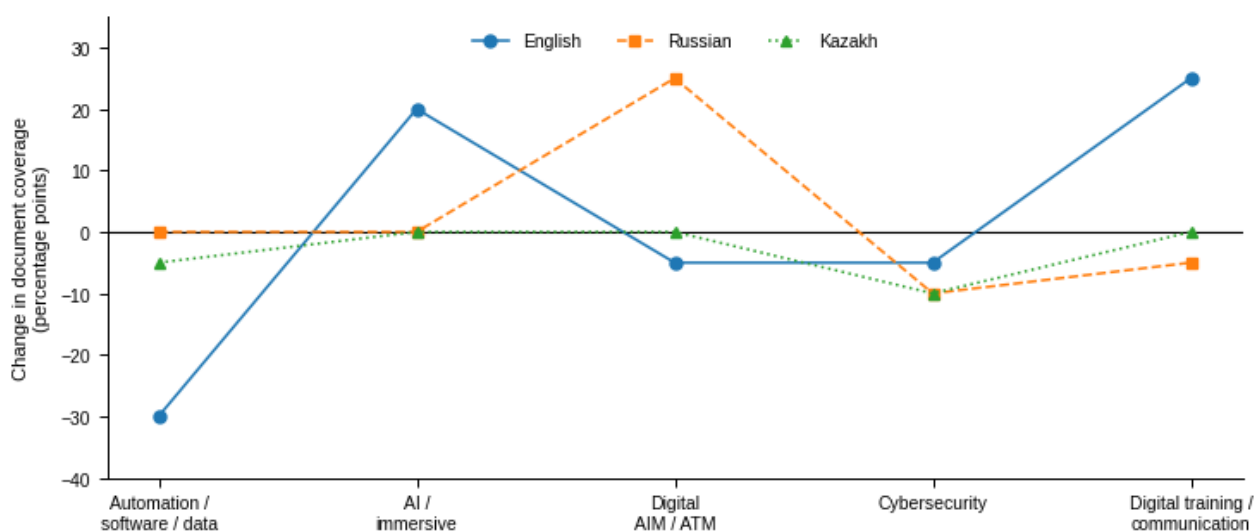


Figure 4 – Diachronic changes in documents

English shows the most pronounced shift toward AI technologies compared to traditional concepts. In Russian, the scope of topics related to digital ATM systems has expanded, whereas Kazakh demonstrates stability across all concept groups. This asymmetry serves as further evidence against the hypothesis of multilingual diversity.

In conclusion, selective convergence—viewed from a descriptive perspective—reflects a

combination of evolving trends and professional context. However, it is worth noting that the languages examined follow similar trajectories. Based on these results and experience, one can conclude that while international standards establish common benchmarks, their linguistic implementation remains language-specific.

Conclusion

The study examined diachronic changes in institutional literature in three languages actively used in Kazakhstan. The reviewed literature is divided into two time periods 2018-19 and 2024-25 for comparative analysis. The findings show how digital transformation is selectively influencing changes in aviation discourse. The most significant result was a reduction in the length of sentences in Russian. The English language has the same tendencies, and the Kazakh language remains stable in this regard. Lexical diversity improves description in all three languages, but the overall language pattern remains statistically significant.

Analysis of language terminology showed that digital transformation is considered as a redistribution of concepts, and not as digital vocabulary. In the texts of the first time period, terms associated mainly with automation are used, and modern discourse is associated with the integration of artificial intelligence terminology. General international terminology has standardization of concepts taking into account the specifics of linguistic discourse.

The proposed study allows us to expand corpus research in the aviation industry. They often involve learning radio communication and aviation English. This work also shows the importance of separating direct translations from the main comparison of the corpus in an interlingual environment. In the future, work will be carried out to expand the data array. There will also be a detailed study of the borrowing process, phraseology and human-machine interaction.

Funding

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23490578 “Development of linguistic support and mobile application for a Kazakh-English-Russian thesaurus-type dictionary of modern aviation terms regulated by ICAO, EASA, IATA standards”).

References

1. Abu Talib, M., Nasir, Q., Dakalbab, F., & Saud, H. (2025). Future Aviation Jobs: The Role of Technology in Shaping Skills and Competencies. *Journal of Open Innovation: Technology, Market, and Complexity*, 11(2), 100517. DOI: 10.1016/j.joitmc.2025.100517.
2. Kim, H. (2026). Aviation English as a Global Lingua Franca. DOI: 10.1017/9781009660785
3. ICAO. Language proficiency. <https://www.icao.int/operational-safety/language-proficiency>.
4. Zhao, W. (2023). A corpus-based study on aviation English from the perspective of systemic functional linguistics. *Discourse & Communication*, 17(5), 630–661. DOI: 10.1177/17504813231171644.
5. Zharkynbekova, S., Galiyeva, B., Ayupova, G., Bekova, M., & Shakhputova, Z. (2026). Language as a Value Category in Kazakhstan's Multilingualism: A Sociolinguistic Analysis of Ethnolinguistic Identity and Language Attitudes Among Different Population Groups. *International Journal of Bilingualism*. DOI: 10.1177/13670069261462817.
6. Karimova, A., Zhumataeva, E., Baigozhina, Z., & Akizhanova, D. (2026). Foreign Language Learning Environment and Communicative Competence Development in Kazakhstan. *Education Sciences*, 16(2), 298. DOI: 10.3390/educsci16020298.
7. Bieswanger, M. (2026). World Englishes, Applied Linguistics, and Air Traffic Control Communication. *World Englishes*. DOI: 10.1111/weng.70016.

8. Ishihara, N., & Prado, M. (2021). The Negotiation of Meaning in Aviation English as a Lingua Franca: A Corpus-Informed Discursive Approach. *The Modern Language Journal*, 105(3), 639–654. DOI: 10.1111/modl.12718.
9. Drayton, J. F., & Coxhead, A. (2022). Plain Language or Anything But? *Journal of Aviation/Aerospace Education & Research*. DOI: 10.15394/jaaer.2022.1908.
10. Pacheco, A., & Tales Figueiredo Silva. (2024). The Challenges of Pilot Language Training for Effective Aeronautical Communications in Multicultural Contexts. *Journal of Aviation/Aerospace Education and Research/The Journal of Aviation/Aerospace Education & Research*, 33(2). DOI: 10.58940/2329-258x.2011.
11. Carvalho, M., Terenzi, D., Tosqui-Lucks, P., & Zhou, Z. (2023). LANGUAGE OF AIRCRAFT DOCUMENTATION: USING A MULTILINGUAL CORPUS IN COLLABORATIVE WORK IN LEXICOGRAPHY AND PEDAGOGY. *Journal of Teaching English for Specific and Academic Purposes/The Journal of Teaching English for Specific and Academic Purposes*, 041–041. DOI: 10.22190/jtesap230103005c.
12. Rigaud Peixoto, R. A. J. (2023). The interface between specialized translation and institutional translation: a selection of candidate terms validated by Aeronautical Meteorology corpora. *Applied Corpus Linguistics*, 100051. DOI: 10.1016/j.acorp.2023.100051.
13. Allkivi, K. (2026). Towards interpretable models for language proficiency assessment: Predicting the CEFR level of Estonian learner texts. *Journal of Responsible Technology*, 25, 100162. DOI: 10.1016/j.jrt.2026.100162.
14. Muawanah, U., Marini, A., & Sarifah, I. (2024). The interconnection between digital literacy, artificial intelligence, and the use of E-learning applications in enhancing the sustainability of Regional Languages: Evidence from Indonesia. *Social Sciences & Humanities Open*, 10, 101169. DOI: 10.1016/j.ssaho.2024.101169.
15. Anayatova, R., Tulekova, G., Koshekov, A., Koshekov, K., & Levchenko, N. (2024). Professional and Communicative Competences in Using Linguistic Aspects of the State Language in the Sphere of Civil Aviation of Kazakhstan. *Psycholinguistics*, 36(2), 63–89. DOI: 10.31470/2309-1797-2024-36-2-63-89.
16. Yalvaç, G. (2026). A bibliometric analysis of airline service quality. *Social Sciences & Humanities Open*, 13, 102417. DOI: 10.1016/j.ssaho.2025.102417.

КӘСІБИ АВИАЦИЯНЫҢ КӨП ТІЛДІ БАЙЛАНЫСТАРЫНЫҢ ЦИФРЛЫҚ ТРАНСФОРМАЦИЯСЫ: КОРПУС НЕГІЗІНДЕГІ ЗЕРТТЕУ

***Аңдатпа.** Цифрлық трансформация авиация саласындағы тілдік коммуникацияға әсер етуде; дегенмен, тілдік деректер шектеулі болып қала береді. Бұл зерттеу авиация саласына қатысты 120 жалпыға қолжетімді үш тілді мәтінге сүйене отырып, екі уақыт кезеңін салыстыра отырып, осы өзгерістерді зерттейді: 2018–2019 және 2024–2025. Алдын ала өңдеу тілдерінің ұқсастықтарды анықтады. Мәтінді өңдеу ережесі негізделген, Юникодты ескеретін құбырды пайдалану арқылы жүзеге асырылды. Лексикалық байлық MATTR-50 метрикасы арқылы бағаланды, ал сөйлем ұзындығы мен цифрлық терминология бес тұжырымдамалық санат және құжат деңгейіндегі дисперсиялық талдау арқылы талданды. Тәуелсіз үлгіні бағалауға Холм түзетуі бүкіл мәтінге қолданылған Манн-Уитни U-тесттері кірді. Терминологияның дисперсиялық талдауы Фишер тесттерін қолдану арқылы жүргізілді. Нәтижелер сөйлем ұзындығының орыс және ағылшын тілдерінде де азайғанын көрсетеді - дегенмен өзгерістер Холм түзетуінен кейін статистикалық тұрғыдан маңызды болып қалды - ал қазақ тілінде ешқандай өзгеріс байқалмады. Сандық терминология ертерек кезеңде болғанымен, қазіргі терминдер жасанды интеллект, модельдеу және оқыту сияқты сандық ұғымдарға көбірек мән береді.*

***Түйін сөздер:** сандық трансформация, авиациялық дискурс, көптілді корпус, сандық авиациялық терминология, кәсіби стандарттау.*

ЦИФРОВАЯ ТРАНСФОРМАЦИЯ МНОГОЯЗЫЧНОЙ КОММУНИКАЦИИ В ПРОФЕССИОНАЛЬНОЙ АВИАЦИИ: КОРПУСНОЕ ИССЛЕДОВАНИЕ

Аннотация. Цифровая трансформация оказывает влияние на языковую коммуникацию в авиационной отрасли; однако лингвистические данные остаются ограниченными. В данном исследовании изучаются эти изменения на основе 120 общедоступных трехязычных текстов, относящихся к авиационному сектору, с сравнением двух временных периодов: 2018–2019 и 2024–2025 гг. Предварительная обработка выявила межъязыковые сходства. Обработка текста проводилась с использованием основанного на правилах конвейера, учитывающего Unicode. Лексическое богатство оценивалось с помощью метрики MATTR-50, а длина предложений и цифровая терминология анализировались с помощью пяти концептуальных категорий и дисперсионного анализа на уровне документа. Независимая оценка выборки включала U-тесты Манна-Уитни с поправкой Холма, применяемой ко всему тексту. Дисперсионный анализ терминологии проводился с использованием тестов Фишера. Результаты показывают, что длина предложений уменьшилась как в русском, так и в английском языках — хотя изменения оставались статистически значимыми после поправки Холма, — тогда как в казахском языке изменений не наблюдалось. В то время как цифровая терминология присутствовала и в более ранний период, современные термины уделяют больше внимания цифровым понятиям, таким как ИИ, моделирование и обучение.

Ключевые слова: цифровая трансформация, авиационный дискурс, многоязычный корпус, цифровая авиационная терминология, профессиональная стандартизация.

Information about the authors

Koshekov Abay	PhD, Associate Professor, Department of Aviation Engineering and Technologies, Civil Aviation Academy, Almaty, Kazakhstan, E-mail: a.k.koshekov@gmail.com
Kuanov Yerkanat	Master, Supervisor, Center for Scientific Achievements, Civil Aviation Academy, Almaty, Kazakhstan, E-mail: y.kuanov@agakaz.kz

Авторлар туралы мәлімет

Кошеков Абай Кайратович	PhD, «Азаматтық авиация академиясы» АҚ, «Авиациялық техника және технологиялар» кафедрасының ассоц. профессоры, Алматы, Қазақстан, E-mail: a.k.koshekov@gmail.com
Куанов Ерқанат Ерболұлы	Магистр, «Ғылыми жетістіктер» орталығының басшысы, Алматы, Қазақстан, E-mail: y.kuanov@agakaz.kz

Сведения об авторах

Кошеков Абай Кайратович	PhD, ассоц.профессор кафдеры «Авиационная техника и технологии» АО «Академия гражданской авиации» Алматы, Казахстан, E-mail: a.k.koshekov@gmail.com
Куанов Ерқанат Ерболұлы	Магистр, Руководитель «ЦНД», E-mail: y.kuanov@agakaz.kz