



DOI: 10.22363/2313-2299-2026-17-2-517-533

EDN: LMPPPM

Research Article / Научная статья

Exploring Cross-Cultural Knowledge Transfer: AI-Integrated Semantic Framework

Maria I. Kiose^{1,2}, Zhenishkul S. Khulkhachieva¹,
Artem V. Barmin¹, Berd Amirkhanov¹

¹Moscow State Linguistic University, Moscow, Russian Federation

²Institute of Linguistics RAS, Moscow, Russian Federation

maria_kiose@mail.ru

Abstract. The study develops a semantic framework with AI-integrated component to identify the knowledge domains transferred from donor into recipient language. Two types of knowledge transfer are considered, direct and indirect, which provide the transfer via loanwords and via their collocates in discourse. The developed framework is used to explore cross-cultural knowledge transfer from Russian as donor language into Kyrgyz as recipient language. To proceed, a clustering Word2vec algorithm is applied to identify the loanword frequent collocates, next, the semantic classes of the collocates of more and less frequent loanwords are determined. The results featuring the semantic classes distribution in the collocates (overall 8,515 in two lists) in web discourse disclose similar and distinct domains of knowledge transfer from Russian into Kyrgyz. The common knowledge domains are mental sphere and knowledge, people, buildings, appliances and routine objects, special event, transport, texts, class names, collective objects, sport, money and finance, measure units, interaction, space and location, substances. Indirect transfer follows two patterns, discrete found in most frequent loanwords which mediate a restricted number of knowledge domains, and continuous found in less frequent loanwords which mediate a larger number of knowledge domains.

Keywords: semantic clustering; semantic tagging; cross-cultural knowledge transfer; loanword; donor language; recipient language

Founding: The research is part of the FSFU-2025-0004 Project “Assessment of cultural integration and disintegration processes in CIS countries: the study of communicative practices” which is being carried out at Moscow State Linguistic University.

Authors’ contribution: The authors contributed equally to this article and preparation of the article text.

Conflict of interests: The authors stated that there were no conflicts of interest.

Article history: Received: 20.01.2026; Accepted: 10.02.2026.

© Kiose M.I., Khulkhachieva Zh.S., Barmin A.V., Amirkhanov B., 2026



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

For citation: Kiose, M.I., Khulkhachieva, Zh.S., Barmin, A.V., & Amirkhanov, B. (2026). Exploring Cross-Cultural Knowledge Transfer: AI-Integrated Semantic Framework. *RUDN Journal of Language Studies, Semiotics and Semantics*, 17(2), 517–533. <https://doi.org/10.22363/2313-2299-2026-17-2-517-533> EDN LMPPPM

Изучение кросс-культурного трансфера знания: семантическое решение с встроенным ИИ-компонентом

М.И. Киосе^{1,2}  , Ж.С. Хулхачиева¹ ,
А.В. Бармин¹ , Б. Амирханов¹

¹Московский государственный лингвистический университет, Москва, Российская Федерация

²Институт языкознания РАН, Москва, Российская Федерация

 maria_kiose@mail.ru

Аннотация. Цель исследования — разработка семантического решения с встроенным ИИ-компонентом, с помощью которого определяется состав сфер знания, подвергающихся межкультурному трансферу из языка-донора в язык-реципиент. Рассматриваются два типа трансфера знания: прямой и непрямой, обеспечивающие передачу знания с помощью самих заимствованных слов и их коллокатов в дискурсе. Выработанное решение позволяет установить состав таких сфер и особенности их распределения в отношении заимствованной лексики из русского языка в киргизский. Задача решается с применением алгоритма Word2vec и последующей семантической разметки коллокатов заимствованных слов с учетом их частотности в языке-реципиенте. Анализ семантических классов двух собранных датасетов контекстов заимствованных слов и их коллокатов (общим числом 8,515 в двух датасетах более и менее частотных заимствований) в новостном дискурсе выявил сходные и отличительные сферы прямого и непрямого трансфера знания из русского языка в киргизский. Основными сферами являются следующие: ментальная сфера и знание, лица, здания и сооружения, инструменты и приспособления, бытовые предметы, мероприятия, транспортные средства, тексты, классы объектов, множества и совокупности объектов, спорт, деньги и финансы, единицы измерения, взаимодействие и взаимоотношение, пространство и место, вещества и материалы. Установлены два паттерна непрямого трансфера знания: дискретный, характерный для наиболее частотных заимствований, проявляющийся в том, что с их помощью происходит трансфер знания только в отношении ограниченного количества сфер знания, и непрерывный, характерный для менее частотных заимствований, проявляющийся в том, что с их помощью обеспечивается трансфер знания в отношении значительно большего количества сфер знания.

Ключевые слова: семантическая кластеризация, семантическая разметка, межкультурный трансфер знания, заимствование, язык-донор; язык-реципиент

Финансирование: Исследование выполняется в Московском государственном лингвистическом университете в рамках Госзадания Минобрнауки России, проект FSFU-2025-0004 «Диагностика процессов культурной интеграции и дезинтеграции в странах СНГ: анализ коммуникативных практик».

Вклад авторов: равнозначен на все этапах исследования и подготовки статьи к публикации.

Конфликт интересов: авторы заявили об отсутствии конфликтов интересов.

История статьи: дата поступления: 20.01.2026; дата приема в печать: 10.02.2026.

Для цитирования: Kiose M.I., Khulkhacheva Zh.S., Barmin A.V., Amirkhanov B. Exploring Cross-Cultural Knowledge Transfer: AI-Integrated Semantic Framework // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2026. Т. 17. № 2. С. 517–533. <https://doi.org/10.22363/2313-2299-2026-17-2-517-533> EDN LMPPPM

Introduction

Cross-cultural knowledge transfer is commonly explored in language via loanwords [1–3]; their rate in recipient language is highly indicative of direct transfer from donor to recipient culture. Meanwhile, in discourse the context in which the loanword frequently appears also becomes charged with the transferred knowledge, since the collocates inherit the semantic structure of loanwords, which can be viewed as a form of adoption (imposition) or indirect knowledge transfer [4; 5]. To explore the knowledge transfer via loanwords in large discourse datasets, complex semantic frameworks are required to identify the loanword collocates and their semantic classes displaying the transferred knowledge domains into the recipient language and culture. This study aims to develop the AI-integrated semantic framework to explore cross-cultural transfer performed directly and indirectly via loanwords, and to further reveal the domains of inherited knowledge from Russian into Kyrgyz, two typologically different although politically and culturally interrelated discourses.

One of the potent AI-based natural processing methods is semantic clustering which presupposes identifying groups of objects displaying similarity to each other and difference to other objects in other groups [6]. Existing clustering frameworks establish the foundation for efficient combining AI-based semantic clustering algorithms and other semantic methods in applied linguistics [7; 8]. This study adopts AI-based semantic clustering within a complex framework to explore cross-cultural direct and indirect transfer. The framework comprises 1) AI-based clustering algorithm of Word2vec revealing word clusters exposed in their collocates and 2) semantic taxonomy analysis to attribute the semantic classes (related to knowledge domains) to the collocates. This complex framework is intended to identify the extent of cross-cultural transfer of two types, direct expressed in the semantic integration of loanwords into a different language and culture, and indirect expressed in the semantic integration of loanword collocates. We hypothesize that higher and lower frequency loanwords function in a singular way in cross-cultural interaction.

The study is structured as follows. Theoretical framework section presents the problem of cross-cultural direct and indirect transfer and the developed methodological decision based on natural processing methods of semantic clustering. Methods and procedure section shows the data and the research steps of the study. Results section presents the research outcomes and their implications for cultural transfer exploration. Final remarks section briefly summarizes the key findings of the study.

Theoretical Framework.

Direct and Indirect Cross-Cultural Knowledge Transfer

Cultural knowledge transfer is addressed as the integration of culturally specific knowledge domains into a donor culture which can be observed in the distribution of single words and expressions [9]. It also appears in the distribution of loanwords and the knowledge domains they refer to, since “culturally specific transfer of domains and conceptual modifications <...> appear in their exporting and importing from one culture into another” [10. P. 5]. It is notable that both the loanwords which designate a new knowledge domain and an already existing one are considered in the studies [2; 3]. Consequently, we expect that identifying the knowledge domains related to these loanwords will allow to reveal the knowledge modifications in recipient culture. Developing these ideas further, we presume that apart from ascertaining the presence of loanwords in a recipient language, their collocates distribution also matters since it evidences of the degree of culture integration attributed to specific knowledge domains [8]. Therefore, two types of knowledge transfer can be observed, direct, via the knowledge domains of loanword, and indirect, via the loanword collocates. In the present study, we diagnose the integration degree and affected knowledge domains of both direct and indirect transfer in two languages, Kyrgyz and Russian which are known to have close cultural contacts.

Single studies address the problems of lexical transfer via loanwords in Russian and Kyrgyz languages [11]; meanwhile these studies do not appeal to the structure of inherited knowledge. They commonly explore the loanword transfer within two frameworks: lexicography aimed to compile the dictionaries of loanwords, and lexical semantics aimed at attributing the loanword meaning. Following these findings, we can explore the direct and indirect knowledge transfer in contemporary discourse adopting the lists of loanwords identified and listed in the compiled etymological dictionaries of the Kyrgyz language¹; meanwhile, these dictionaries being compiled at the end of the XXth century (there are no up-to-date etymological dictionaries of Kyrgyz) do not reveal recent inputs into Kyrgyz, which is definitely a research limitation.

Semantic Framework with AI-Integrated Clustering Decision

To explore cross-cultural transfer in large datasets, we have to develop the semantic framework capable of processing the loanwords in contexts. This problem is currently addressed via natural processing methods which have now become integrated into applied linguistic studies, with semantic clustering being one of the most potent in grouping words. Semantic clustering performs three major tasks: a) underlining the pattern and insight into the data, b) identifying the degree of

¹ Yudakhin, K.K. (1985). *Kyrgyz-Russian Dictionary*. Frunze: Glavnaya redaktsiya Kirgyzskoy Sovetskoy Entsiklopedii. (In Russ.); Karasaev, Kh.K. (1986). *The Dictionary of Loanwords: 5100 words*. Frunze: Kyrgyz Sovet. (In Kyrgyz).

similarity between data points, c) data organization and summarization through cluster prototypes [6]. Multiple studies present the evolution and trends in data clustering; to proceed, they focus on algorithms for grouping data sets used in statistics, computer science and machine learning (review see [12]). Importantly, clustering decisions have become complex or hybrid methods integrating both AI-based techniques and other clustering methods (e.g. automatic K-means clustering and semantic clustering) [13]. Semantic clustering is at present a natural language processing method employed in investigating semantic relationship through the shared meanings the words maintain in context [14; 15]; it allows to cluster words through word embeddings (Word2vec) or through transformer-based models (BERT) by calculating their similarity in the vector space.

In applied linguistics, semantic clustering has been extensively used in recent decades for different purposes. In Divjak [16], a clustering decision for lexical groups was developed. It was later extended to cluster the languages via the use of perception semantics in verbs, where D. Divjak applied random forest model to identify the predictors of each perception type in the samples [17]. It is notable that a linguistic foundation for clustering methods in lexical groups lie in prototypicality effects explored vastly since the pioneering works of Rosch [18] and Geeraerts [19] who claimed that the centers of semantic categories could be revealed in psychological experiments. These ideas were taken further due to developing the corpus linguistic methods dealing with big data [20]. At present, LLM (large language models) operate performing natural language processing including text understanding and generation (for review see [21]), for instance to perform text clustering based on deep semantic understanding [6; 22]. Semantic clustering as another of natural language processing methods is now extensively used in education and psychology studies [7; 23]. Developing its applied potential in cross-cultural studies, we address the view promoted by Kuhn et al [14] who regard lexical semantic clusters as groups of source artifacts that use similar vocabulary; therefore, these are the common knowledge domains which underlie cluster formation.

Meanwhile, identifying the clusters of loanwords and their collocates in exploring indirect transfer is not enough to reveal these knowledge domains. Additionally, semantic taxonomy has to be implemented to annotate the collocates and to further determine the semantic classes related to these knowledge domains. To explore the semantic classes of both loanwords and their target (frequently used) collocates, their overall taxonomy is required. Several major approaches have been developed to proceed, which are lexical-semantic analysis, frame analysis, image-schema, qualia analysis, which display the taxonomic classes of words within different types of conceptual domains. In [5] a complex semantic framework with AI-integrated clustering decision was already tested with the second component being qualia-analysis. Since in this study we intend to identify the knowledge domains inherited from or mediated by the donor culture, we appeal to lexical-semantic taxonomy proposed in the National Corpus of the Russian Language

(NCRL)² and identify the distribution of these taxonomic classes attributed to the loanwords in two datasets, with higher and lower corpus frequency, and to their target collocates. Following the postulates of cultural transfer studies [10], we presume that revealing the knowledge distribution in related or distant cultures and languages allows to identify the cultural transfer routes and the affected knowledge domains.

Methods and Procedure

Following the decisions elaborated above, we propose a two-component semantic framework to explore indirect cross-cultural transfer via loanwords. To obtain a list of collocates with their numerical scores, we use an AI-based algorithm of semantic clustering (Word2vec). To further identify the knowledge domains mediated by indirect transfer, we apply the semantic taxonomy protocol devised in NCRL, to annotate the semantic classes of collocates. Since we explore both direct and indirect transfer, we apply semantic taxonomy to annotate both loanwords in two lists, more and less frequent, as part of direct transfer analysis, and to annotate loanword collocates in two lists attributed to more and less frequent loanwords. We presume that higher and lower frequency loanwords promote knowledge transfer via different domains of direct and indirect transfer thus functioning in a singular way in cross-cultural interaction.

The research procedure involves the following steps. At step 1, we compile two lists of loanwords from Russian into Kyrgyz following their corpus frequency. At Step 2, we compile the datasets with their corpus contexts. At Step 3, via semantic clustering procedure the target collocates with their numerical scores are identified. At Step 4, we determine the semantic taxonomic distribution of both the domains of inherited in loanwords and modulated knowledge in collocates, to further contrast the data and verify the hypothesis claiming the existing differences in their distribution in two lists. Consequently, the outcomes of the study include i) two lists of loanwords from Russian into Kyrgyz distinct in their corpus frequency, ii) two datasets of corpus contexts of these loanwords, iii) two lists of target collocates of these loanwords with their numerical scores obtained via AI semantic clustering, iiii) the determined differences in the taxonomic distribution of both the domains of inherited and mediated knowledge transferred from Russian into Kyrgyz.

The research data comprise two macro blocks: a) a compiled list of loanwords borrowed from Russian into Kyrgyz which fall into two lists different in their web corpus frequency, b) two compiled datasets of their contexts elicited from the web corpus. To prepare a list of loanwords from Russian into Kyrgyz, two existing dictionaries of the Kyrgyz language were addressed: 1) “The Dictionary of

² National Corpus of the Russian Language. <https://ruscorpora.ru/page/instruction-semantic> (accessed: 10.01.2026). (In Russ.).

Loanwords” compiled by Kh. Karasaev³ which contains 5.100 loanwords; 2) “The Dictionary of the Kyrgyz Language” compiled by K. Yudakhin⁴ which contains 100.000 words. To obtain the datasets of loanword collocates, we addressed Kyrgyz Web text corpus (Kyrgyzstan)⁵ which is part of the open-access Leipzig Corpora, and which comprises 3.036.362 sentences and 36.581.130 tokens. A programmed decision was further elaborated to form the context dataset, since the corpus does not have a lemmatizer⁶. Next, AI-based clustering decision was applied to cluster the loanwords via their collocates. Word2vec word-embedding algorithm was used to vectorize the datasets data, which allowed to determine the closest neighbors of each loanword in each of its word forms in each of the datasets. We further manually performed semantic annotation of target collocates in two lists of target words. To proceed, we applied the taxonomy developed for NCRL. In accordance with it, single semantic classes were attributed to autosemantic parts of speech. While this taxonomy is easier attributed to names, its use to semanticize attributes, adverbs and verbs is less regulated since their meaning is highly contextualized. For instance, in NCRL the semantics of *актуальный* (topical) is restricted only to one class ‘relative’, which will not allow to identify its semantic attribution. In our case, when dealing with target collocates, the procedure of context analysis could not be accomplished. For this reason, only names were classified in this study.

Results.

Direct Knowledge Transfer: Knowledge Domains in Loanwords

All the words in Karasaev⁷ and in Yudakhin⁸ marked as having Russian as immediate etymon were selected, their number being equal to 1.204. Next, their corpus frequency in the base form was identified, which allowed to segment the data into three lists, highly frequent loanwords (64, corpus frequency in the base form ranging from 800 to 13,500), less frequent (74, corpus frequency in the base form ranging from 300 to 799) and infrequent (corpus frequency in the base form not exceeding 299). The loanwords from Lists 1 and 2 were further annotated using the semantic taxonomy presented in NCRL.

In Figure 1 we present the semantic taxonomic distribution of the loanwords in two lists.

³ Karasaev, Kh.K. (1986). *The Dictionary of Loanwords*: 5100 words. Frunze: Kyrgyz Sovet. (In Kyrgyz).

⁴ Yudakhin, K.K. (1985). *Kyrgyz-Russian Dictionary*. Frunze: Glavnaya redaktsiya Kirgizskoy Sovetskoy Entsiklopedii. (In Russ.).

⁵ Kyrgyz Web Text Corpus. Leipzig Corpora URL: <https://wortschatz.uni-leipzig.de/en> (accessed: 10.01.2026).

⁶ The authors are grateful to Daniil Sinkevich for his help with developing the software used to compile the context dataset.

⁷ Karasaev, Kh.K. (1986). *The Dictionary of Loanwords*: 5100 words. Frunze: Kyrgyz Sovet. (In Kyrgyz).

⁸ Yudakhin, K.K. (1985). *Kyrgyz-Russian Dictionary*. Frunze: Glavnaya redaktsiya Kirgizskoy Sovetskoy Entsiklopedii. (In Russ.).

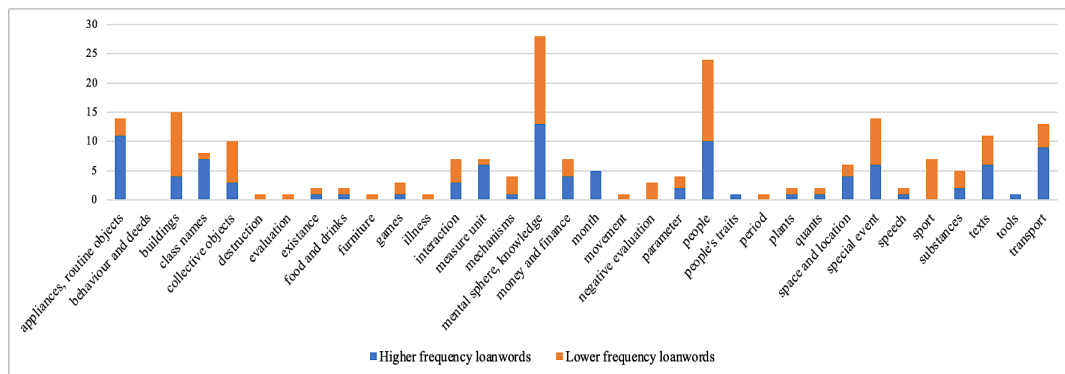


Figure 1. Semantic taxonomic distribution of the loanwords, abs.

Source: compiled by Maria I. Kiose, Zhenishkul S. Khulkhachieva, Artem V. Barmin, Berd Amirkhanov.

As seen, the most frequent semantic classes of loanwords are 1) mental sphere, knowledge, 2) people, 3) buildings, 4) appliances, routine objects, 5) special event, 6) transport. “Mental sphere, knowledge” is expressed in *план* (plan), *физика* (physics), *экономика* (economics); in multiple cases it is represented integrally with the class “people”, e.g. in *профессор* (professor), *парламент* (parliament), *партия* (party), where the latter two also manifest the class “collective objects”. The semantic class “buildings” is found in *клуб* (club), *аэропорт* (airport), *база* (base). The class “appliances, routine objects” appears in *карта* (card, map), *мандат* (ticket), *товар* (goods), *техника* (tools, technics), where the latter two additionally manifest the class “class names”. The loanwords *олимпиада* (Olympiad), *семинар* (seminar), *экзамен* (exam) serve to represent the semantic class “special event”, while the sphere “transport” is found in *поезд* (train), *велосипед* (bike), *такси* (taxi).

The major differences in the distribution of these semantic classes within the two lists of loanwords lie in the classes “transport”, “appliances, routine objects”, “month”, “class names”, “measure unit”, which prevail in higher frequency loanwords list; while the classes “buildings”, “collective objects”, “sport” display prevalence in the list of lower frequency words. Importantly, as the results show, lower frequency words display the single cases of use attributed to multiple semantic classes related to states, communication, evaluation. It is noticeable that the loanwords related to the class “food” are less numerous and less frequent, which shows the prevalence in Kyrgyz culture to its own culinary traditions. These results may serve to identify the semantic classes concomitant with the knowledge domains majorly affected by direct knowledge transfer within the Kyrgyz culture, and the ones less affected, still presumably due to their less significant role in culture.

Indirect Knowledge Transfer: Knowledge Domains

To compile the datasets of contexts with more and less frequent loanwords, we used the Kyrgyz Web text corpus (Kyrgyzstan)). Meanwhile, since this corpus is not equipped with a lemmatizer, the lists of word forms were compiled; they

contained 825 word forms for 64 most frequent loanwords, and 878 word forms for 74 less frequent ones. The two compiled datasets of contexts with more and less frequent loanwords comprise 348.084 contexts (Dataset 1) and 145.039 contexts (Dataset 2).

At this step, to perform the AI-based semantic clustering of loanwords and their neighbors (collocates), we applied a word-embedding algorithm using *word2vec* R-package v.0.4.0. in RStudio v.2024.12.1+563 with R v.4.3.2. The following parameters of the model were set: type of word2vec algorithm (type) = skip-gram; dimensions of word vectors (dim) = 300; number of times a word should occur to be considered as part of the training vocabulary (min_count) = 1. The model was trained on web samples using the “predict” function in the *word2vec* R-package to determine the five nearest neighbors for each of the loanword forms based on their vector similarity in the 300-dimensional space. We also collected the numerical scores of vector similarity between each of the loanwords and their five nearest neighbors. To visualize the results, we uploaded the three-column matrix to Gephi v.0.10.1. — an open-source software for network visualization and analysis and then used the Fruchterman-Reinhold layout algorithm to better optimize the resulting network.

Figure 2 shows a fragment of semantic net with the word cluster produced by the loanword *максу* (taxi) displaying the semantic class “transport” prevailing within the loanwords.

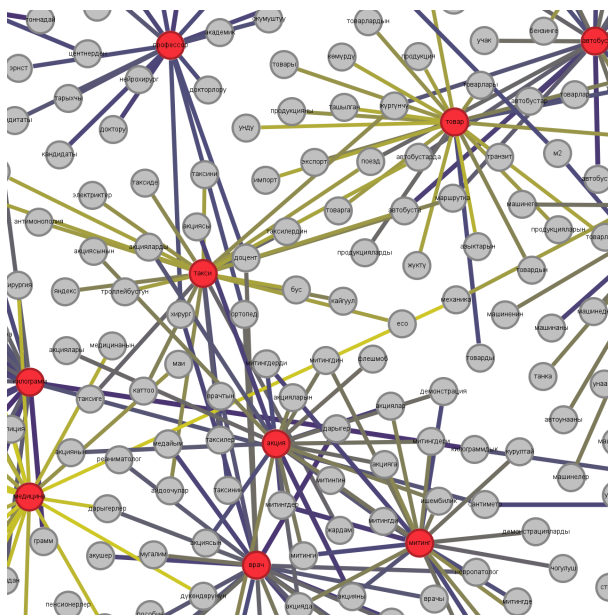


Figure 2. Fragment of Semantic Net

Source: compiled by Maria I. Kiose, Zhenishkul S. Khulkhachieva, Artem V. Barmin, Berd Amirkhanov.

As seen, the loanword *максу* (taxi) forms semantic nets with the loanwords related to other knowledge domains via semantic classes, here *транспорт*

(professor) and *врач* (doctor) related to mental sphere and people, *акция* (action, share) related to special event, money and finance, *товар* (goods) referring to appliances, routine objects, collected objects, *килограмм* (kilo) attributed to measure unit, etc., which overall gives rise to multiple semantic connections expressed in collocates.

Two lists of target collocates for more and less frequent loanwords (Dataset 1 and Dataset 2) were further obtained, with the first containing 4,125 words and the second containing 4.390 words. Next, we proceeded to semantic classification of target collocates of these loanwords in both lists. All the target collocates in two lists were attributed to one or more of these taxonomy classes following their basic dictionary meaning. To attribute the semantic class to a lexeme, we applied a two-stage procedure. First, we addressed the basic dictionary meaning / meanings of a word, which allowed to identify a single semantic class or a group of semantic classes included in the taxonomy. Next, we consulted the semantic classes attributed to a lexeme in NCRL. At this step, we specified or added the semantic classes to the words. Using a one-step procedure (addressing only NCRL) appeared less reliable. For instance, *опера* (opera) is semanticized as 1) music and drama genre, 2) staged performance; therefore, the semantic classes of mental sphere, special event and space and location are attributed to it. It is notable that in NCRL *опера* (opera) has the following semantic attribution: semantics: a piece of music, object name; additional semantics: notional name; where the class “a piece of music” although attributed is not included in the semantic taxonomy, and other two classes are utterly common and do not display semantic taxonomic distribution. Therefore, although in major cases we consulted the semantic taxonomy of single words presented in NCRL, we mostly addressed the meanings listed in the dictionary to attribute the semantic classes. Below we present the results displaying the distribution of the semantic classes corresponding to affected knowledge domains.

List one with the collocates of most frequent loanwords includes 4.126 target collocates; with 3.525 nouns and nominals which were further semanticized using NCRL semantic protocol. To annotate the semantic classes in these collocates, the same procedure applied to annotate the loanwords was used; 5.392 tags were performed. Figure 3 shows the distribution of semantic classes tagged in the collocates.

The results show that the collocates semantic classes can be further grouped considering their frequency in the compiled collocates list. The collocates of higher frequency (with values exceeding one third of the highest frequency value) manifest the classes “mental sphere, knowledge”, “people”, “appliances, routine objects”, “texts”. The collocates of lower frequency (with values exceeding one tenth of the highest frequency value) relate to the classes “special event”, “collective objects”, “transport”, “money and finance”, “class names”, “measure unit”, “buildings”, “month”, “parameter”, “space and location”, “substances”, “interaction”, “mechanisms”. The least frequent collocates relate to a wide range of semantic

classes attributed both to tangible and intangible values. Additionally, several occasional classes were identified but due to their infrequent use were not involved in the diagram.

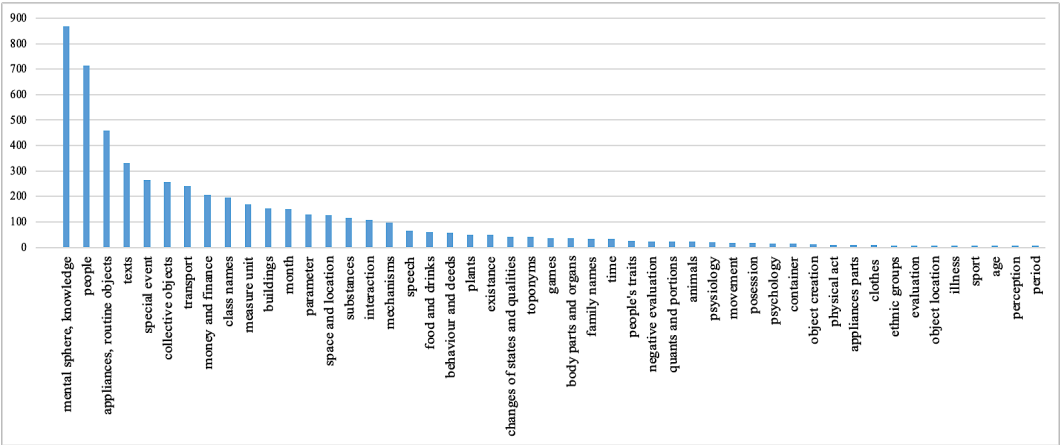


Figure 3. Semantic Taxonomy of the List one Loanword Collocates, abs.
Source: compiled by Maria I. Kiose, Zhenishkul S. Khulkhachieva, Artem V. Barmin, Berd Amirkhanov.

Next, we contrasted these classes distribution with the ones tagged in the loanwords of higher frequency (Figure 4), in both cases relative values were used.

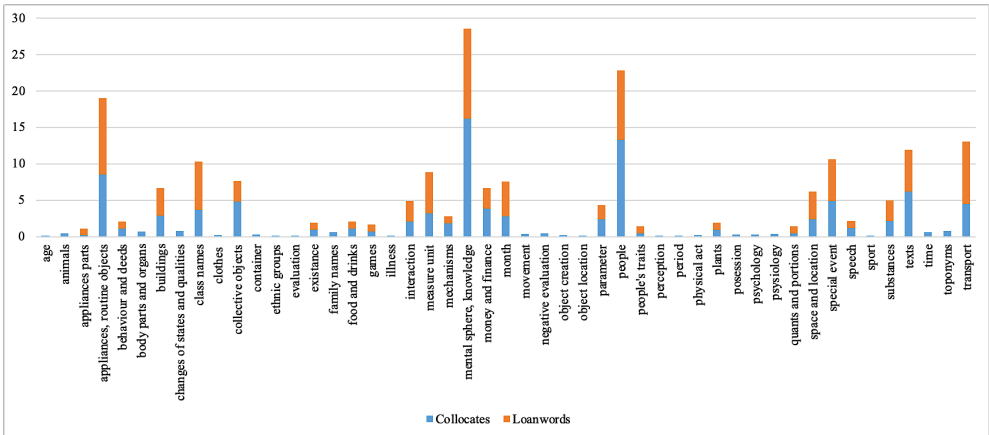


Figure 4. Contrastive Semantic Taxonomic Distribution of the Loanwords and Collocates, rel.
Source: compiled by Maria I. Kiose, Zhenishkul S. Khulkhachieva, Artem V. Barmin, Berd Amirkhanov.

With the loanword semantic taxonomy manifesting direct transfer and the collocates semantic taxonomy manifesting indirect transfer, the data help contrast the direct and indirect knowledge transfer affected by highly frequent loanwords. The results show that the difference in the distribution of higher frequency classes lies in “class names”, “measure unit”, “month”, “space and location”, “transport” which prevail in direct transfer, and “collective objects”, “money and finance”,

“texts” which prevail in indirect transfer; meanwhile, multiple χ^2 tests did not show any significant differences in their distribution.

Next, we addressed the distribution of semantic classes within the collocates of less frequent loanwords. List two with the collocates of less frequent loanwords includes 4.390 target collocates; with 3.067 nouns and nominals which were further semanticized. It is notable that List two with collocates involves a larger number (compared to List one) of names (of companies, brands) which were not semanticized. To annotate the semantic classes in these collocates, the same procedure applied to annotate the loanwords was used; 3.874 tags were found. Figure 5 shows the distribution of semantic classes tagged in the collocates.

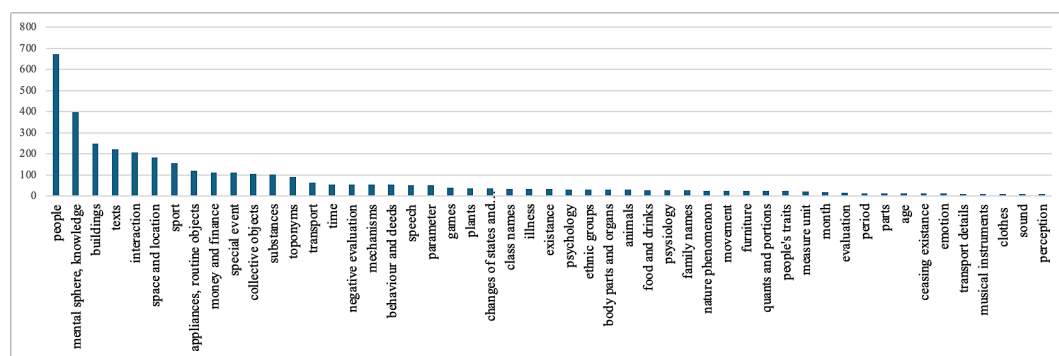


Figure 5. Semantic Taxonomy of the List Two Loanword Collocates, abs.

Source: compiled by Maria I. Kiose, Zhenishkul S. Khulkhachieva, Artem V. Barmin, Berd Amirkhanov.

The results show that the collocates of higher frequency (with values exceeding one third of the highest frequency value) manifest the classes “people”, “mental sphere, knowledge”, “buildings”. The collocates of lower frequency (with values exceeding one tenth of the highest frequency value) relate to the classes “texts”, “interaction”, “space and location”, “sport”, “special event”, “collective objects”, “money and finance”, “substances”, “appliances, routine objects”, “toponyms”. Infrequent semantic classes are numerous; meanwhile, we observe the balanced used of tangible and intangible objects.

Importantly, we observe a slightly different picture if contrasted with the distribution of List one loanword collocates. First, we notice a decrease in the frequency of two highly manifested classes, which are “mental sphere, knowledge” and “people”, which cannot be linked to the decrease in the presence of these spheres in the List two loanwords since they are equally frequently present in both lists. Presumably, as the results show, while most frequent loanwords display a directed effect, less frequent loanwords encompass a wider range of semantic classes. These differences show two distinct patterns of indirect transfer — discrete with single affected knowledge domains found with frequent loanwords, and continuous with multiple affected knowledge domains found with less frequent loanwords. It is noticeable although that Mann-Whitney U test did not show significant differences in the data distribution in two collocate lists ($U=1743$,

$p=0.765$), which proves the prevalence of the same semantic classes and infrequent presence of similar ones.

Next, we contrasted these classes distribution with the ones tagged in the loanwords of lower frequency (Figure 6), in both cases relative values were used.

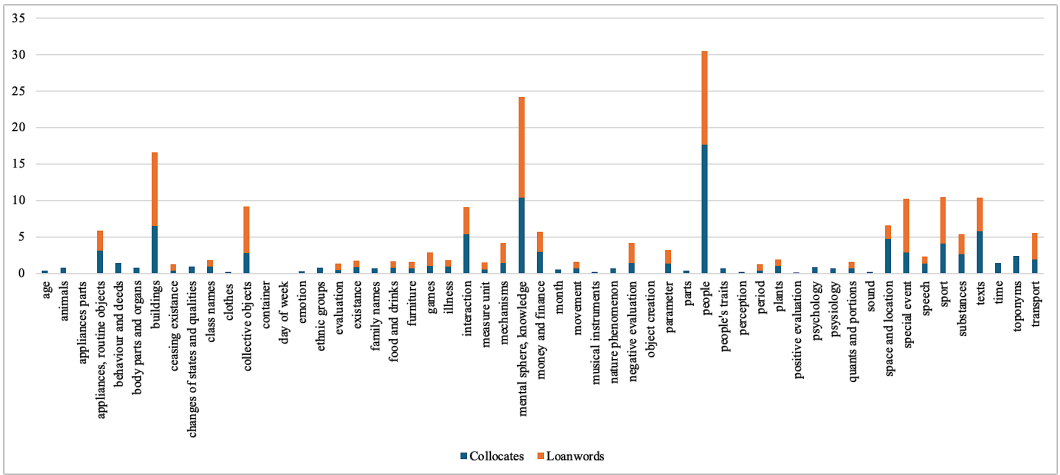


Figure 6. Contrastive Semantic Taxonomic Distribution of the Loanwords and Collocates, rel.
Source: compiled by Maria I. Kiose, Zhenishkul S. Khulkhachieva, Artem V. Barmin, Berd Amirkhanov.

The results allow to identify the differences in direct and indirect transfer via lower frequency loanwords. Higher differences are observed in the semantic classes “special event”, “collective objects”, “sport”, “transport” which prevail in direct transfer, and the classes “interaction”, “space and location”, which prevail in indirect transfer. In line with the results obtained with higher frequency loanwords, multiple χ^2 tests did not show any significant differences in their distribution due to a large number of classes.

Meanwhile, the contrastive analysis helped formulate several regularities in the direct and indirect knowledge transfer. First, there are two semantic classes which are largely affected by both direct and indirect knowledge transfer in higher and lower frequency loanwords, which are “mental sphere, knowledge” and “people”; other less affected classes are “appliances, routine objects” and “buildings”. Second, most affected semantic classes display regular correspondences in the effects of direct and indirect transfer, which means that the frequency rate of these semantic classes is similar in direct and indirect transfer. Third, both object and abstract names (representing tangible and intangible values) serve to participate in direct and indirect knowledge transfer; additionally, no regulations were found in their distribution mediated by higher and lower frequency loanwords.

The study helped rank the semantic classes according to their regularity in affecting and being affected by knowledge transfer. Two semantic scales displayed below manifest the direct and indirect knowledge transfer from Russian into Kyrgyz

language and culture featuring the frequency effects displayed in modern web discourse.

Direct knowledge transfer domains (most frequent semantic classes, cases): Mental sphere, knowledge (28) > People (24) > Buildings (15) > Appliances, routine objects (14), Special event (14) > Transport (13) > Texts (11) > Collective objects (10) > Class names (8) > Sport (7), Money and finance (7), Measure unit (7), Interaction (7) > Space and location (6) > Substances (5), Month (5).

Indirect knowledge transfer domains (most frequent semantic classes, cases): People (1387) > Mental sphere, knowledge (1263) > Appliances, routine objects (576) > Texts (551) > Special event (374) > Collective objects (361) > Money and finance (317), Interaction (315) > Space and location (307), Transport (302) > Buildings (296) > Class names (230) > Substances (216) > Measure unit (190) > Parameter (180) > Month (169) > Sport (161) > Mechanisms (151) > Toponyms (132).

Overall, the obtained results do not fully support the research hypothesis claiming that higher and lower frequency loanwords promote knowledge transfer via different domains of direct and indirect transfer thus functioning in a singular way in cross-cultural interaction. The distribution of semantic classes corresponding to knowledge domains clearly showed that while there is a distinction in several knowledge domains prevailing within the collocates of either more or less frequent loanwords (e.g., the redistribution in the classes “people” and “mental sphere, knowledge”), the key knowledge domains remain stable. This means that these knowledge domains are most susceptible to being mediated by knowledge transfer, presumably due to their high activity in language and culture. The results also validate the significance of the notion of knowledge transfer proposed and developed in [9; 10] since it proved to serve a valid methodology to explore the distribution of knowledge domains within a mediated culture. Additionally, they show that the lexical decision of loanwords proposed in [1–3] can serve to explore not only direct, but also indirect knowledge transfer.

Most importantly, the AI-integrated framework was found valid to explore indirect knowledge transfer, which paves the way for new applications of semantic clustering. Therefore, besides the applications shown in [7; 14; 23], cross-cultural integration can be efficiently determined with its help. Additionally, the second component of the developed semantic framework, which is semantic taxonomy application proposed in NCRL, was found equally effective in attributing semantic classes to collocates and identifying the distribution of affected knowledge domains. Meanwhile, the research limitation is definitely the necessity to use the etymological dictionaries^{9,10} [which do not enlist the loanwords of the XXIst century.

⁹ Karasaev, Kh.K. (1986). *The Dictionary of Loanwords*: 5100 words. Frunze: Kyrgyz Sovet. (In Kyrgyz).

¹⁰ Yudakhin, K.K. (1985). *Kyrgyz-Russian Dictionary*. Frunze: Glavnaya redaktsiya Kirgyzskoy Sovetskoy Entsiklopedii. (In Russ.).

Final Remarks

The study proposed a two-component semantic framework to explore cross-cultural knowledge transfer mediated by more and less frequent loanwords. The framework comprising AI-based semantic clustering algorithm (Word2vec) and semantic taxonomy analysis allowed to determine the knowledge domains of direct and indirect knowledge transfer. Two scales of most frequent knowledge domains susceptible to knowledge transfer were deduced, with the knowledge domains corresponding to semantic classes “people” and “mental sphere, knowledge” being most frequent. Additionally, the study determined several regularities in direct and indirect knowledge transfer. Importantly, indirect transfer follows two patterns, discrete found in most frequent loanwords which mediate a restricted number of knowledge domains, and continuous found in less frequent loanwords which mediate a larger number of knowledge domains. Overall, the results contribute to developing the procedure of AI-based methods integration into semantic frameworks, with semantic clustering being efficient in exploring cross-cultural knowledge transfer. The research perspectives are seen both in verifying the AI-based semantic framework in cultural studies, and in exploring cross-cultural transfer in other languages and cultures.

References / Список литературы

1. Myers-Scotton, C. (2002). *Contact linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.
2. Myers-Scotton, C. (2006). *Multiple Voices: An Introduction to Bilingualism*. Malden, MA: Blackwell.
3. Haspelmath, M. (2009). Lexical Borrowings: Concepts and Issues. In: M. Haspelmath, U. Tadmor (eds.) *Loanwords in the World's Languages* (pp. 35–54). Berlin: De Gruyter.
4. Thomason, S.G. (2001). *Language Contact*. Washington, D.C.: Georgetown University Press.
5. Kiose, M.I., Khulkhachieva, Zh.S., Izyumskaya-Kapitonova, V.V., & Barmin, A.V. (2025). Lexical-Semantic Clustering in the Diagnostics of Culture Integration in Discourse. *Research Result. Theoretical and Applied Linguistics*, 11(4), 63–84. <https://doi.org/10.18413/2313-8912-2025-11-4-0-4> EDN: UBIXHD
6. Singh, J., & Singh, D. (2024). A Comprehensive Review of Clustering Techniques in Artificial Intelligence for Knowledge Discovery. *Advanced Engineering Informatics*, 62, 102799. <https://doi.org/10.1016/j.aei.2024.102799> EDN: QKLLAC
7. Pérez-Serrano, M., Noguerol-López, M., & Duñabeitia, J. A. (2022). Effects of Semantic Clustering and Repetition on Incidental Vocabulary Learning. *Frontiers in Psychology*, 13, 997951.
8. Zhu, Y., Yang, L., Xu, K., Zhang, W., Song, Z., Wang, J., & Yu, P.S. (2025). LLM-MemCluster: Empowering Large Language Models with Dynamic Memory for Text Clustering. *Computation and Language*. <https://arxiv.org/abs/2511.15424> (accessed: 15.01.2026).
9. Iriskhanova, O.K., & Kiose, M.I. (2016). Technologies of Interdisciplinary Terms Transfer into Language Studies. In: V.V. Feschenko (Ed.) *Linguistics and Semiotics of Cultural Transfers: Methods, Principles, Technologies* (pp. 151–180). Moscow: Cultural revolution. (In Russ.). EDN: YQAGOL

- Ириханова О.К., Киосе М.И. Технологии трансфера междисциплинарных терминов в лингвистику // Лингвистика и семиотика культурных трансферов: методы, принципы, технологии / отв. ред. В.В. Фещенко. М. : Культурная революция, 2016. С. 151–180.
10. Feschenko, V.V., & Bochaver, S.Yu. (2016). Theory of Cultural Transfers: from Translation Studies — through Cultural Studies — to the Theory of Language. In: V. V. Feschenko (Ed.) *Linguistics and Semiotics of Cultural Transfers: Methods, Principles, Technologies* (pp. 5–35). Moscow: Cultural revolution. (In Russ.). EDN: YFKVZR
Фещенко В.В., Боцавер С.Ю. Теория культурных трансферов: от переводоведения — через cultural studies — к теоретической лингвистике // Лингвистика и семиотика культурных трансферов: методы, принципы, технологии / отв. ред. В. В. Фещенко. М. : Культурная революция, 2016. С. 5±–35.
 11. Kambaralieva, U.D., & Sternin, I.A. (2021). *Russian and Kyrgyz Communicative Behaviour*. Voronezh: Ritm. (In Russ.). EDN: JNYREE
Камбаралиева У.Д., Стернин И.А. Русское и киргизское коммуникативное поведение. Воронеж : Ритм, 2021.
 12. José-García, A., & Gómez-Flores, W. (2016). Automatic Clustering Using Nature-Inspired Metaheuristics: a Survey. *Applied Soft Computing*, 41, 192–213.
 13. Nguyen, C.D., & Cios, K.J. (2008). GAKREM: A Novel hybrid Clustering Algorithm. *Information Sciences*, 178 (22), 4205–4227.
 14. Kuhn, A., Ducasse, S., & Gırba, T. (2007). Semantic Clustering: Identifying Topics in Source Code. *Information and Software Technology*, 49(3), 230–243.
 15. Witschard, D., Jusufi, I., Martins, R.M., Kucher, K., & Kerren, A. (2022). Interactive Optimization of Embedding-Based Text Similarity Calculations. *Information Visualization*, 21(4), 335–353. <https://doi.org/10.1177/14738716221114372> EDN: JNAUAD
 16. Divjak, D. (2010). *Structuring the Lexicon: A Clustered Model for Near-Synonymity*. Berlin: De Gruyter.
 17. Divjak, D. (2015). Exploring the Grammar of Perception. A Case Study Using Data from Russian. *Sensory Perceptions in Language and Cognition. Special issue of functions of Language*, 22(1), 44–68. <https://doi.org/10.1075/fol.22.1.03div> EDN: URKXYP
 18. Rosch, E. (1977). Human Categorization. In: *Studies in Cross-Cultural Psychology*, N. Warren (Ed.) (Vol. 1, pp. 1–49). New York: Academic Press.
 19. Geeraerts, D. (1988). Where Does Prototypicality Come from? In: *Topics in Cognitive Linguistics*, B. Rudzka-Ostyn (Ed.) (pp. 207–229). Amsterdam, Philadelphia: John Benjamins.
 20. Soloviev, V.D. (2015). Possible Mechanisms of Change in the Cognitive Structure of Semantic Rows. In: *Language and thought. Modern cognitive linguistics* (pp. 478–487). Moscow: Languages of Slavic Culture.
 21. Kolmogorova, A.V., & Margolina, A.V. (2024). Written vs Generated Text: “Naturalness” as a Textual and Psycholinguistic Category. *Research Result. Theoretical and Applied Linguistics*, 10(2), 71–99. <https://doi.org/10.18413/2313-8912-2024-10-2-0-4> EDN: CCESHE
 22. Hadifar, A., Sterckx, L., Demeester, T., & Develder, C. (2019). A Self-Training Approach for Short Text Clustering. In: *Proceedings of the 4th Workshop on Representational Learning for NLP (RepL4NLP-2019)* (pp. 194–199). Italy: Florence.
 23. Troyer, A.K., Moscovitch, M., & Winocur, G. (1997). Clustering and Switching as Two Components of Verbal Fluency: Evidence from Younger and Older Healthy Adults. *Neuropsychology*, 11(1), 138–146.

Information about the authors:

Maria I. Kiose, Dr. Sc. (Philology) (Advanced Doctorate), Associate Professor, Chief Researcher of the Centre for Socio-Cognitive Discourse Studies, Moscow State Linguistic University (38 build 1 Ostozhenka str., Moscow, Russian Federation, 119034); Leading Researcher; Laboratory for Multichannel Communication, Institute of Linguistics, RAS

(1 B. Kislovsky In., Moscow, Russian Federation, 125009); *Research interests*: cognitive semantics, indirect naming, multimodal linguistics; *e-mail*: maria_kiose@mail.ru
ORCID: 0000-0001-7215-0604. Scopus AuthorID: 56642747500; ResearcherID: AAB-7989-2019, SPIN-code: 4419-0090, AuthorID: 341964.

Zhenishkul S. Khulkhachieva, PhD in Philology, Associate Professor, Head of Department of Languages and Cultures of CIS and FSU countries, Head of Ch. Aitmatov Centre for Kyrgyz language and culture, Moscow State Linguistic University (38 build 1 Ostozhenka str., Moscow, Russian Federation, 119034); *Research interests*: linguodidactics, teaching methodology, cognitive linguistics; *e-mail*: j.khulkhachieva@linguanet.ru
ORCID: 0009-0000-3506-2007. SPIN-code: 7897-5998, AuthorID: 1014042.

Artem V. Barmin, Junior Researcher of the Laboratory for Cognitive Studies of Communication of Moscow State Linguistic University (38 build 1 Ostozhenka str., Moscow, Russian Federation, 119034); *Research interests*: semantic memory, network science; *e-mail*: art.barmin1@gmail.com
ORCID: 0000-0001-9658-6621. Scopus Author ID: 58113747500; ResearcherID: AGL-9794-2022; SPIN-code: 4233-4593, AuthorID: 1146203.

Al-Berd I. Amirkhanov is a Research Assistant at the Center for Sociocognitive Research of Discourse, Moscow State Linguistic University (38 build 1 Ostozhenka str., Moscow, Russian Federation, 119034); *Research interests*: political economy, regional development, cross-cultural knowledge transfer; *e-mail*: berd582@mail.ru

Сведения об авторах:

Киосе Мария Ивановна, доктор филологических наук, доцент, главный научный сотрудник Центра социокогнитивных исследований дискурса, Московский государственный лингвистический университет (119034, Российская Федерация, г. Москва, ул. Остоженка, д. 38 с. 1); ведущий научный сотрудник Лаборатории мультимедийной коммуникации, Институт языкознания РАН (125009, Российская Федерация, г. Москва, Б. Кисловский переулок, д. 1); *сфера научных интересов*: когнитивная семантика, не прямое наименование, мультимодальная лингвистика; *e-mail*: maria_kiose@mail.ru
ORCID: 0000-0001-7215-0604. Scopus Author ID: 56642747500; ResearcherID: AAB-7989-2019; SPIN-код: 4419-0090, AuthorID: 341964.

Женишкуль Саматовна Хулхачиева, кандидат филологических наук, доцент, заведующий кафедрой языков и культур стран СНГ и ближнего зарубежья ИМОиСПН, директор центра киргизского языка и культуры им. Ч. Айтматова, Московский государственный лингвистический университет (119034, Российская Федерация, г. Москва, ул. Остоженка, д. 38 с. 1); *сфера научных интересов*: лингводидактика, методика преподавания, когнитивная лингвистика; *e-mail*: j.khulkhachieva@linguanet.ru
ORCID: 0009-0000-3506-2007. SPIN-код: 7897-5998, AuthorID: 1014042.

Бармин Артем Вячеславович, младший научный сотрудник Лаборатории когнитивных исследований основ коммуникации, Московский государственный лингвистический университет (119034, Российская Федерация, г. Москва, ул. Остоженка, д. 38 с. 1); *сфера научных интересов*: семантическая память, наука о сетях; *e-mail*: art.barmin1@gmail.com
ORCID: 0000-0001-9658-6621. Scopus Author ID: 58113747500; ResearcherID: AGL-9794-2022; SPIN-код: 4233-4593, AuthorID: 1146203.

Амирханов Аль-Берд Исламович, лаборант-исследователь Центра социокогнитивных исследований дискурса, Московский государственный лингвистический университет (119034, Российская Федерация, г. Москва, ул. Остоженка, д. 38 с. 1); *сфера научных интересов*: политическая экономия, региональное развитие, межкультурный трансфер знаний; *e-mail*: berd582@mail.ru