

A Linguistic Interference in Information Space Terms: A Corpus-Based Study in Kazakh

Assel B. Ormanova*

Department of General Education Disciplines, Astana IT University, Astana, Kazakhstan

Madina L. Anafinova

Foreign philology, L. N. Gumilyov Eurasian National University, Astana, Kazakhstan

Abstract—Recently the interaction of Kazakh and English has become a highly debatable issue in Kazakhstan linguistics. Language contacts tend to result in possible errors both in communication and perception of the linguistic worldview. The paper deals with the analysis of linguistic interference and the way it is represented in Kazakh terminology. Language corpus outcomes stand for the research instrument. #LancsBox 5.1.2 program builds and processes the corpus that involves 1,238 texts from the five Kazakhstani online newspaper platforms' websites. The research provides the statistics on the number of texts, concordance lines, frequency, collocations, and analysis of the Kazakh terms that feature interference such as *galamtor*, *indet*, *onlain*, *oflain*. There are characteristics of the most frequently used terms and their overview in the major corpora of Kazakh and English. The study analyzes influence at phonetic, morphological, syntactical, lexical, and semantic levels in information space terms. Thus, our research presents a novel framework to study linguistic interference through contrastive analysis based on corpus processing outcomes at different language levels in a multilingual environment in Kazakhstan.

Index Terms—interference, term, corpus tool, corpora, Kazakh

I. INTRODUCTION

Today, due to the policy of trilingualism in Kazakhstan, where the Kazakh, Russian, and English languages interact and influence each other, such a language phenomenon as linguistic interference has become a topic of debate and research interest in the country. Such a trinity of languages is essential for social harmony in multinational Kazakhstan (Zhikeeva, 2017, p.37).

Does linguistic interference feature Kazakh terminology? If so, to what extent? How can the language corpora reveal it? Therefore, linguistic interference and its differentiation types in the terminology of global information space based on the corpora outcomes stand for the research interest in this paper. We analyse the terms featuring linguistic interference in Kazakh due to the influence of English. Our focus lies on information space terms that are extracted with a corpus instrument.

There are two research questions in the present study as follows:

1. How can the language corpora reveal linguistic interference in Terminology?
2. What are the differential features of linguistic interference in information space terms in the multilingual environment?

Following the research questions, we put two objectives:

1. To build the corpus and process it through the corpus tool #LancsBox developed at the Centre for Corpus Approaches to Social Science (CASS) at Lancaster University;
2. To define, analyse, and describe linguistic interference and its differential features at the language levels comparing Kazakh and English.

II. LITERATURE REVIEW

A. Linguistic Interference

Linguistic interference was first proposed by Weinreich (1953) and studied in the frames of language-contact studies (Rosenzweig, 1972; Odlin, 1989; Heine & Kuteva, 2005; Karlinskiy, 2011). According to the behavioristic approach, in learning L2, speakers transfer the 'habit' of the L1 into L2. Additionally, affecting second/foreign language learning by the learner's native language is called interference of mother tongue (Thyab, 2016). The term, interference, is defined as 'the automatic transfer, due to habit of the surface structure of the first language onto the surface of the target language' (Dulay & Burt, 1976, p.71). Thus, linguistic interference refers to the errors committed by speakers in the bilingual environment.

* Corresponding Author

Relying on the definitions above, interference is regarded as the transfer of the norm of the first language upon the target language norm, where grammar, lexis, phonology, sentence structure, and meaning are influenced. If the target language is different from L1, this transfer results in interference or negative transfer; on the contrary, if L1 and L2 are similar, a positive transfer occurs (Goswami, 2020, p.146). Language interference is a deviation from the literary norms in a bilingual speech under the influence of another language within the language contact (Anafinova & Ormanova, 2020). According to the Linguistic encyclopedia (LES, 1990), interference is the interaction of language systems in a bilingual environment, emerging either through language contact or when an individual learns a foreign language; expressed in deviations from the norms and system of a second language being influenced by a native one". Specifically, we consider interference as the interaction of language systems in a bilingual environment, expressed in deviations from the norms and system of L2 being influenced by L1.

Research studies concerning linguistic interference are as follows.

Pozdnyakova (2012) conducted research related to teaching bilingual students the Russian language. She proposed that the main method of overcoming interference in the speech of bilingual students was and remains the analysis of the results of their speech activity, or, more simply, the analysis of errors in the speech of students at different levels - lexical, phonological, morphological, syntactic. This analysis should be carried out in several stages. At the same time, it must be taken into account that, when mastering a language (both native and foreign), mistakes in speech are made by everyone - both monolinguals and bilinguals, so not every mistake in speech can be considered to have arisen under the influence of a foreign language system. In the study there is a detailed description of typical mistakes caused by linguistic interference and an analysis of teaching methods to eliminate linguistic interference in a speech of bilingual students.

Kaweera (2013) reviewed the theoretical concept of interlingual interference of the mother tongue, Thai to the target language, English and intralingual interference found in EFL student writing in the Thai context with the attempt to define the existence of errors according to their sources. She exemplifies some frequent errors normally found in Thai student writing based on three perspectives of interlingual interference (lexical, syntactic and discourse interference) and seven aspects of intralingual interference (false analogy, misanalysis, incomplete rule application, exploiting redundancy, overlooking cooccurrence restrictions, hypercorrection and overgeneralization). In the research she proved that errors found in Thai student writing are mainly influenced by both interlingual and intralingual causes.

Another significant study of linguistic interference is Goswami (2020). This work examines the errors committed by Sylheti speakers during their speaking in English that can be attributed to the interference of the rule of the mother tongue in learning the second language. This interference can be of phonological, morphological and sentence structure. The domain of the work is exclusive to phonological interference of L1 on L2 through the contrastive analysis of the phonological aspects of Sylheti Bangla and English. The author intends to improve the status of the English teaching-learning process compared to contemporary practices.

Though a substantive number of researches have been done on linguistic interference, not a single work is available in the literature on the interference in the terminology of Kazakh caused by English. Most research deals with second language acquisition, but not the language structure in particular.

In Kazakhstan, the issue finds its reflection in the School of Language Contacts and its founder – a linguist Avram E. Karlinskiy whose works have contributed to Kazakhstani contactology. He is the developer of the dialing analysis method aimed to model the speech of a bilingual by predicting and experimentally checking the speech interference (Karlinskiy, 2011, p. 55). The followers N. S. Pak, D. D. Shaibakava, Z. Zh. Aukhadieva, A. I. Rabinovich, L. N. Kovylna, V. T. Kirshner, S. Gazieva have investigated the processes of speech interference, integration, and the convergent development of languages (Narozhnaya, 2018, p.108). Interference reveals in cases of deviation from the norm in a bilingual speech of L2 under the influence of L1 on the phonological, grammatical, and lexical levels of the language due to language contact (Abazova, 2019, p.127).

Taking into consideration the differentiation of linguistic interference, there are several views. Weinreich (1979, p.22) defines three types of interference as phonetic, grammatical, and lexical, while Karlinsky (2011, pp. 219, 238, 251) proposes grammatical, lexical, and syntagmatic. In addition, semantic, stylistic, intralingual, and interlingual interference (Kaweera, 2013, p.9) are distinguished. In our study, we rely on the classification proposed by Alimov (2015, p.31), who differentiates this linguistic phenomenon into five levels: phonetic, morphological, syntactical, lexical, and semantic. Thus, considering his views, we analyze five linguistic interference types and sustain them with terms taken from Kazakhstani information space sources.

B. Information Space Sources

In its turn, information space is viewed as a specific environment in the globalized information society that can provide a great range of data sources. It represents a collection of resources and infrastructures that make up the state and inter-state computer networks, telecommunications systems, public networks, and other cross-border communication channels (Dobrovolskaya, 2014, p.141). While analyzing it as a space of social interaction, it is possible to allocate the following fundamental factors:

- an audience of a resource which possesses valuable general orientations;
- the social importance of information for the emergence and development of interpersonal interaction;
- the technical environment as the appropriate means supporting the process of interaction (Karaseva, 2016, p.111).

Information sources may be observations, people's speeches, documents, pictures, etc. They can be divided into two broad categories: documentary (primary, secondary, tertiary) and non-documentary (formal, informal) sources (Varshney, 2011).

Kazakhstani information sources include official and non-official Internet web platforms, presented at local, state, and national levels. Their number is increasing due to the digitalization of society and the great opportunities they provide. From a linguistic perspective, the primary role of information sources is to represent a current language situation in society and show relevant linguistic means used in real communication.

C. Corpus Approach to Terminology

Since information space is always in progress, it stimulates language development and enlarges it with a new lexicon. Terminology, as the study of a specific vocabulary, has successfully adopted an approach to collecting lexical data based on corpora. According to Sager (1990, p.58), "by being studied in the context of communicative situations, terms are no longer seen as separate items in dictionaries or part of a semi-artificial language deliberately devoid of any of the functions of other functional items. The increasing tendency to analyze terminology in its communicative, i.e., linguistic context, leads to some new theoretical assumptions and also to new methods of compilation and representation".

Research grants much through computational technologies as the benefits of using linguistic corpora are in the high data representativeness in the real-life context and the possibility of repeated use of a once-created corpus. Current research in the area of computational terminology is mainly aimed at structuring the output of term extractors to access further levels of knowledge (Drouin, p.2004).

Traditionally, linguists have used the term "corpus" to designate a body of naturally occurring (authentic) language data that can be used as a basis for linguistic research (Garside et al., 2013, p.1). It is made up of written texts and spoken discourse. A corpus represents a particular language, and the computer processes it. With increasing computers' power and capability, corpora have improved significantly in size, variety, and ease of access.

III. METHODOLOGY

A. Data Collection

We present cross-linguistic and one-year longitudinal data from a study of linguistic interference in information space terms through a methodology of contrastive analysis in Kazakh and English, where the corpus is applied as an instrument of linguistic research.

When it comes to data collection, we focus on online newspaper platforms highly rated among the public in Kazakhstan (see Table 1). For building the corpus, we compile the online publications from five platforms. The selection of the online newspaper platforms is due to the ratings on <https://aqparat.info/feed>.

TABLE 1
OVERVIEW OF KAZAKHSTANI ONLINE NEWSPAPER PLATFORMS USED IN THE CORPUS

Online newspaper platform	Access	Publications	Tokens	Types	Lemmas
Business – Mir Kazakhstan	http://businessmir.kz/	291	159,253	30,895	33,581
Tengri News	https://tengrinews.kz/	278	326,659	55,059	63,112
Egemen Qazaqstan	https://egemen.kz/	273	73,935	43,705	46,760
Forbes Kazakhstan	https://forbes.kz/	245	263,921	64,890	76,796
Kazahstanskaya Pravda	https://kazpravda.kz/	151	57,682	15,374	17,051
Total: 5		1,238	881,450	122,513	143,780

According to the table above, the corpus comprises 1,238 articles taken from Business – Mir Kazakhstan, Tengri News, Egemen Qazaqstan, Forbes Kazakhstan, Kazahstanskaya Pravda. These articles total up to 881,450 tokens which correspond to roughly 122,513-word forms, and 143,780 lemmas. The size of the corpus although still modest can guarantee that the articles discuss a wide range of subjects and that their content is heterogeneous. The selected articles were published over twelve months in 2020. We consider the corpus to reflect the non-technical usage of the terms, though its specialized field is journalistic.

B. Data Analysis

To analyse the corpus, we use #LancsBox 5.1.2 – a program developed by the Centre for Corpus Approaches to Social Science (CASS) at Lancaster University. It is a new generation corpus tool that allows easy analysis and visualization of corpus data. Throughout the tool, #LancsBox offers powerful searches at different levels of corpus annotation using i) simple searches, ii) wildcard searches, iii) smart searches, and iv) regex searches (Brezina et al., 2020).

To mention, whereas the tool's supported languages are Chinese, English, French, and Russian, we run it in Russian and managed to get sufficient data for further analyses. The interface of #LancsBox with the Annotation of 1,238 files, 881,450 tokens, 122,513 types, and 143,780 lemmas is presented below (see Figure 1).

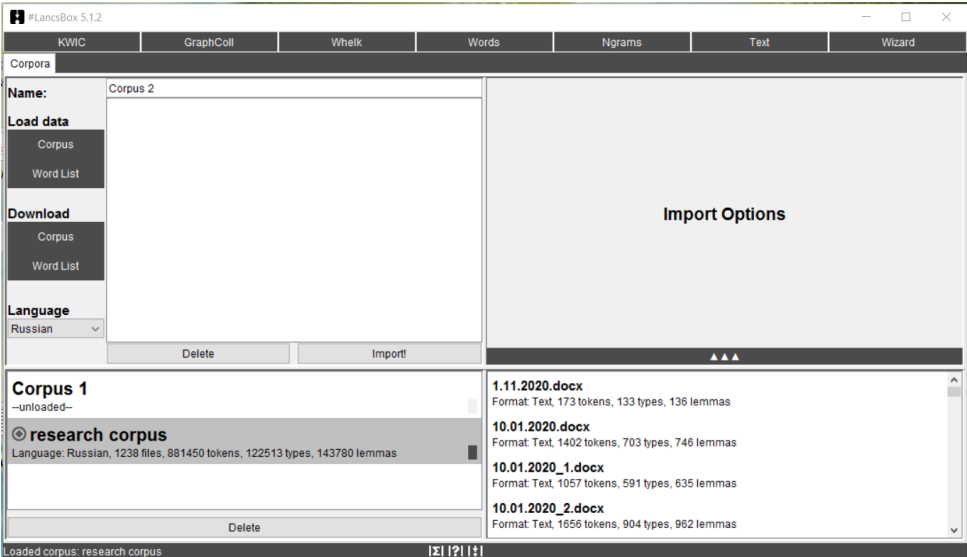


Figure 1 The Interface of #LancsBox Program.

After finding out the most frequently used entries in the corpus, we focus on terms that feature linguistic interference at different language levels. Following Alimov’s differentiation of linguistic interference, we prove the emergence of interference in the information space terms on five levels: phonetic, morphological, syntactical, lexical, and semantic.

IV. RESULTS

A. The Built Corpus Outcomes

While processing the software, we use wildcard searches with a particular character - * (asterisk) to find all variants of searched terms. In this prospect, we build a list of the most frequently used information space terms with the relevant wordforms (see Figure 2).

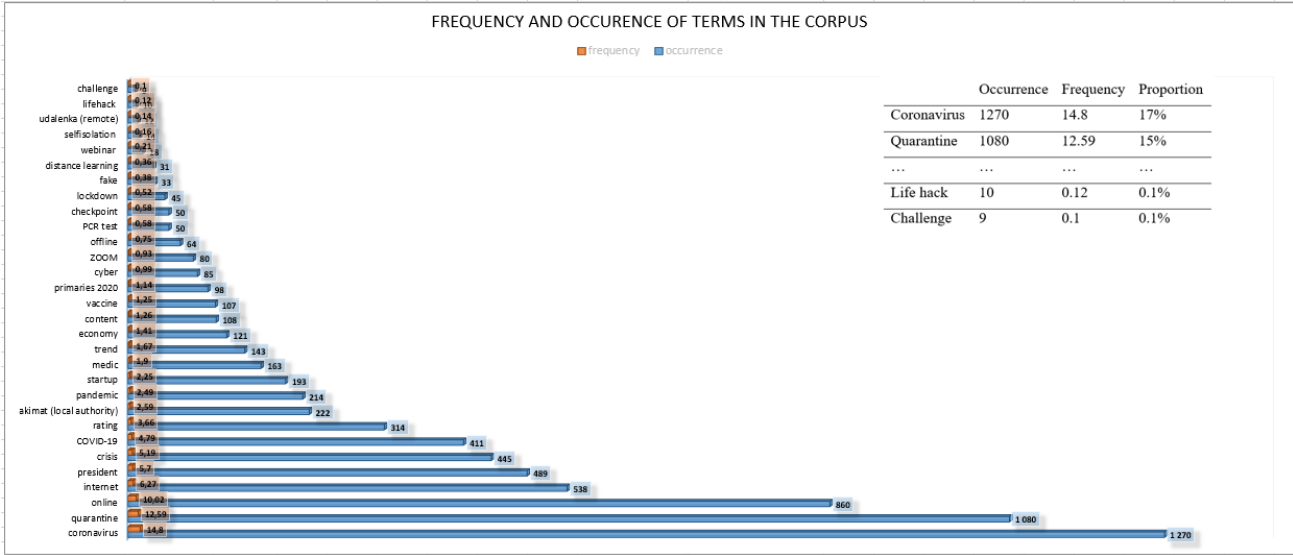


Figure 2 The Number of Occurrences of Information Space Terms in the Corpus.

The corpus analysis program generates a list of all words with the number of occurrences and frequency. We extract thirty terms out of the whole list as the most frequently used. Figure 2 indicates that *coronavirus* and *quarantine* account for the most frequently used entries in the corpus, with more than one thousand occurrences each in the texts. At the same time, *life hack* and *challenge* show minor numbers, 10 and 9 cases, respectively. The frequency of *coronavirus* and *quarantine* is 14,8 and 12,59, while for *life hack* and *challenge* is 0,12 and 0,1, respectively. By proportion, *coronavirus* composes 17% and *quarantine* with 15%, while *life hack* and *challenge* make 0.1% among thirty highly used terms of the corpus. Overall, occurrences, frequencies, and proportions of all thirty terms in English and Kazakh are presented below (see Table 2).

TABLE 2
THE MOST FREQUENTLY USED TERMS IN THE CORPUS

In English	In Kazakh	Occurrence	Frequency	Proportion
coronavirus	koronavirus	1,235	14,01	17
quarantine	karantin	1,080	12,59	15
online	onlain	860	10,02	12
internet	internet, ğalamtor	538	6,27	7
president	prezident	489	5,7	7
crisis	daǵdarys	445	5,19	6
COVID-19	COVID-19	411	4,79	6
rating	reitiń	314	3,66	4
akimat (local authority)	ákimdik	222	2,59	3
pandemic	pandemia, indet	214	2,49	3
startup	startap	193	2,25	3
medic	medik	163	1,9	2
trend	trend	143	1,67	2
economics	ekonomika	121	1,41	2
content	kontent	108	1,26	1
vaccine	vaksina	107	1,25	1
primaries 2020	bastauys 2020	98	1,14	1
cyber	kiber	85	0,99	1
ZOOM	ZOOM	80	0,93	1
offline	oflain	64	0,75	1
PCR test	PTR-test	50	0,58	1
checkpoint	blokpost	50	0,58	1
lockdown	lokdaun	45	0,52	1
fake	feik	33	0,38	0
distance learning	qasyqtyqtan oqytu	31	0,36	0
webinar	vebinar	18	0,21	0
selfisolation	samoizol ásia	14	0,16	0
udalenka (remote)	udalenka	12	0,14	0
lifehack	laifhak	10	0,12	0
challenge	chellenj	9	0,1	0

In addition, based on our findings, we may propose Word of the Year 2020 in Kazakhstan which stands for *coronavirus*. It appears 1235 times with a relative frequency of 14.01 in 353 out of 1,238 publications on Kazakh online news platforms during the twelve months of the year. The term shares the following wordforms: *corona*, *coronavirus* with the morphological suffixes, “koronavirýsnaia infektsiia” (coronavirus infection) (see Figure 3).

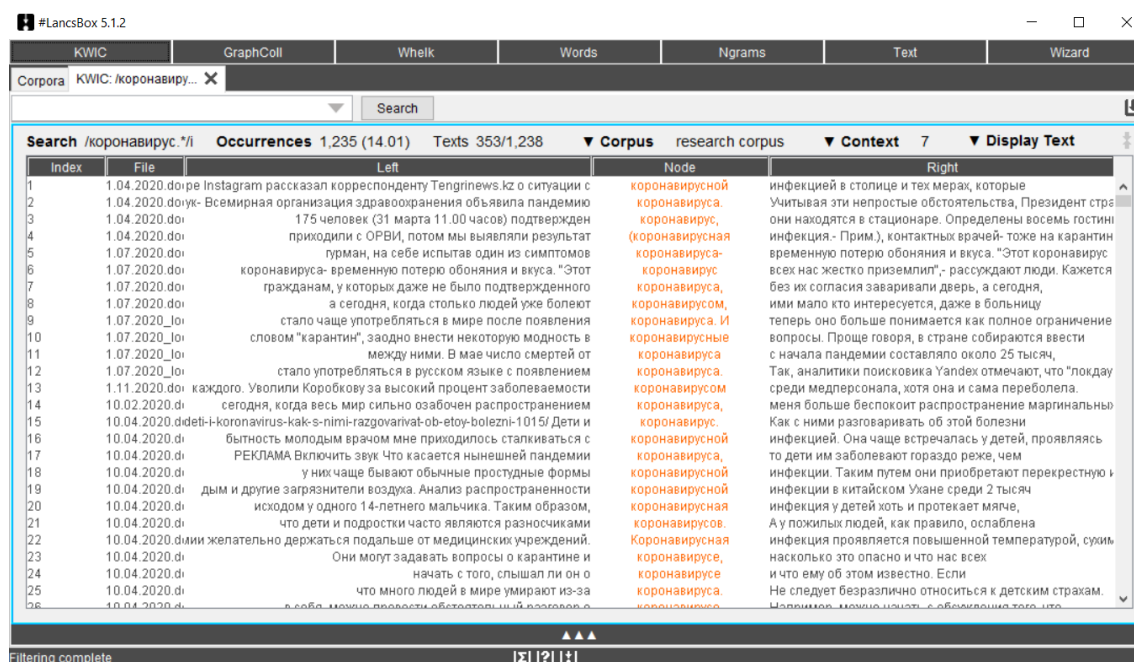


Figure 3 #LancsBox Processing the Term “coronavirus”

Thus, we detect the entries with the highest frequency among the Kazakhstani information space terms in 2020. The prevailing majority of them is the transliterated borrowings from English. The Kazakh language adopts them from English through Russian which plays as a mediator in the linguistic process. Some are pure of the Kazakh language

origin, like *akimat* which stands for a 'local authority'. Another term *udalenka* is authentic for Russian and hardly can be understood by others as it is a reduced form of *udalennyi* that means remote.

B. Searching the Major Corpora of Kazakh and English

We observe the linguistic situation of the term coronavirus and its reflection in the major corpora of Kazakh and English. They provide much information, including frequency, definition, translation, links to audio, images, videos, synonyms, related word forms, topics, collocates, clusters, and concordance / KWIC lines (see Table 3). Further below, there is an overview of each corpus from the perspective of searching coronavirus.

TABLE 3
OVERVIEW OF CORPORA USED

Corpus	Access	Size	Representativeness	Data on "coronavirus"
iWeb: The intelligent Web-based Corpus	www.english-corpora.org/iweb/ 15 January 2021	14 billion	22 million web pages	1547 frequencies, concordance lines, web pages, context
Coronavirus Corpus	www.english-corpora.org/corona/ 15 January 2021	725 million	online newspapers and magazines	1,712,023 frequencies, collocations, context, can be downloaded
Almaty Corpus of Kazakh language	www.web-corpora.net/KazakhCorpus/ 15 January 2021	40 million	Written and spoken Formal and informal	mentioned once in 2013
National Corpus of Kazakh Language	www.qazcorpus.kz 15 January 2021	30 million	Written and spoken Formal and informal	linguistic markings, article author's name, its title, source, details, features

(a). "Coronavirus" in National Corpus of Kazakh Language

The volume of the text base of National Corpus of Kazakh Language is 30 million words. The collected texts were obtained from five Kazakh language styles (art style, scientific style, journalistic style, business style, and conversational style). When searching for a word, a list of texts marked with meta-markings is displayed on the screen. The corpus owns data on concordance, lemmatisation, and linguistic markings (morphological, word-formative, lexical, phonetic, morpho-semantic). The site is available to everyone.

The corpus provides seventy-four documents with *coronavirus* and information about linguistic markings, the name of the author of the article, its title, source, details, and features (see Figure 4).

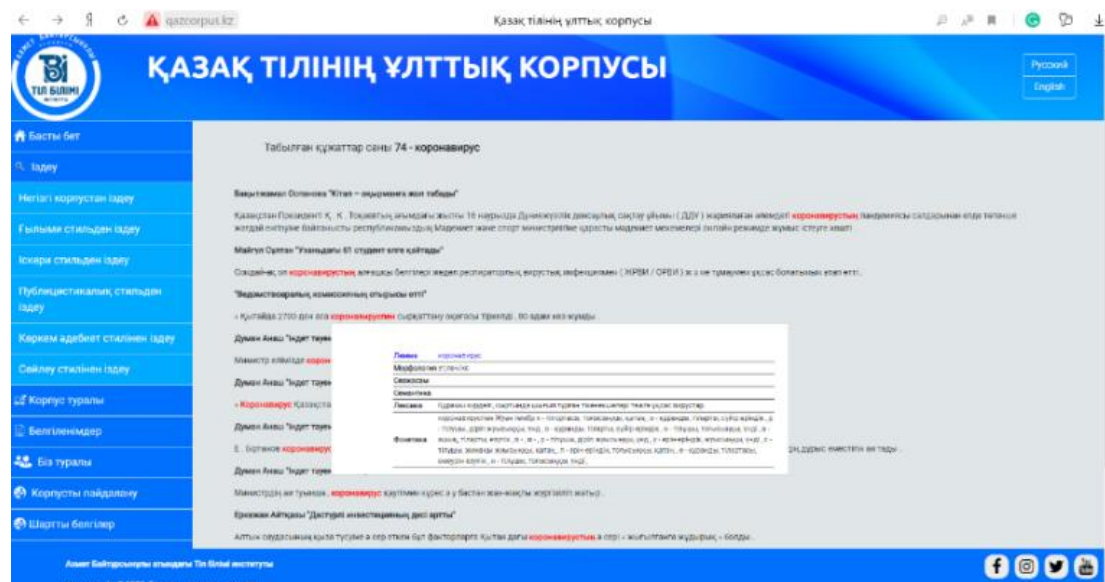


Figure 4 The Interface of National Corpus of Kazakh Language With the Search of *coronavirus* (accessed 15 January 2021)

(b). "Coronavirus" in Almaty Corpus of Kazakh Language

Almaty Corpus of Kazakh language is one of the versions of National Corpus of Kazakh language as a reference system based on the extensive fund of the marked texts of literary Kazakh. The corpus is continuously supplemented and updated quantitatively and qualitatively. Besides, the corpus's search functionality is significantly improved. Currently, the size of the corpus is more than 40-million-word tokens. The corpus texts are marked through the automatic morphological analyzer, 86% of the corpus' word forms are parsed. The homonymy in the corpus is not removed, i.e. all possible analysis options without a context are assigned to each word form. It is a convenient tool for

scientific research, developing manuals and workbooks of the Kazakh language, independent studying of Kazakh, providing the majority of word forms with lexical-morphological analyses and the Russian/English translation equivalents. The annotated corpus is supplied with grammatical and bibliographic marking. The corpus includes literary, scientific, publicist texts and more than 100 classical Kazakh literature works.

Almaty Corpus of Kazakh language shows the only occurrence of *coronavirus* in the publication dated no later than 2013, which means there is no relevance to the global situation due to the pandemic.

(c). “*Coronavirus*” in *iWeb*

iWeb corpus contains fourteen billion words in twenty-two million web pages. It is related to other corpora of English that English-Corpora.org created to offer unparalleled insight into variation in English. There are three main ways to search the corpus. First, a researcher can browse a frequency list, searches by word form, part of speech, ranges, and pronunciation. Second, a researcher can search by an individual word and see collocates, topics, clusters, websites, concordance lines, and related words for each of these words. Third, a researcher can search for phrases and strings (*ism, un*able, verb-ed, buy * adj noun, and others). The corpus is optimized for speed, so searches are quick.

It generates a list of all instances of *coronavirus* in the form of a concordance. Overall, it shows 1547 frequencies and concordance lines, the web pages, and the context. Clicking on the web page sends a researcher to the whole article where the term is employed.

(d). “*Coronavirus*” in the *Coronavirus Corpus*

Coronavirus Corpus is a definitive record of the social, cultural, and economic impact of COVID-19 in 2020 and beyond. The corpus shows what people are saying in online newspapers and magazines in twenty different English-speaking countries. The corpus was first released in May 2020, and currently, it is about 727 million words in size, and it continues to enlarge. The Coronavirus Corpus allows a researcher to see the frequency of words and phrases in 10-day increments since January 2020, such as social distancing, flatten the curve, Zoom, pandemic, and others. It enables finding out the collocates (nearby words) and the patterns in which a word occurs. A researcher can also compare different time periods and compare the 20 countries in the corpus. Clicking on the links in the search form to the left for context-sensitive help can provide the range of queries that the corpus offers.

Coronavirus Corpus provides 1,712,023 frequencies. The term occurs more than one million times in the corpus. Collocations can be seen in the *Keyword in Context* display. For the search <*coronavirus**> there are collocations as *coronavirus-related* (14,802), *coronaviruses* (10,387), *coronavirus-induced* (2,096), *coronavirus-hit* (1,389), *coronavirus-infected* (650), and others. While, for the search <**coronavirus*> shows the following collocations as *#coronavirus* (1,803), *post-coronavirus* (1,793), *pre-coronavirus* (1,607), *anti-coronavirus* (1,165), *non-coronavirus* (901). The frequencies of the collocations are given in brackets. Moreover, the corpus provides the number of occurrences under every twelve months in 2020, and it can be downloaded for offline use.

(e). “*Coronavirus*” in *Other Kazakhstani Information Space Sources*

To add, we expand our search not limiting to the corpora, but considering other information space sources. So, we would like to mention the internet platform Termincom.kz for dealing with terms only. It is a unified terminological electronic database of terms in science, education, technology and economics, political and social life in Kazakhstan. The website thrives on implementing new requirements for the systematization of Kazakh terminology and the unification of terminological Kazakh vocabulary. Users can find approved and existing terms and use them concerning their activities; they can get acquainted with scientific and reference literature published in Kazakh terminology. The terminological database provides the occurrence of *coronavirus* in the categories: medicine, ecology, economy, education, national policy, business and law regulations, food industry, and consumer services. There is no definition of *coronavirus* in Kazakh, and its spelling is like *коронавирус* in Russian and two variants in Kazakh: *коронавирус* and *coronavirus*. The primary Kazakh language dictionary sources (lugat.kz, emle.kz, sozdikqor.kz, sozdik.kz, tilalemi.kz) propose the translation for coronavirus as *коронавирус* for a noun, *коронавирустық* for an adjective.

Hence, we manage to get relevant data on the search term through searching corpora. The Kazakh language is presented by National Corpus of Kazakh Language and Almaty Corpus of Kazakh language, English is by iWeb: The intelligent Web-based Corpus and Coronavirus Corpus. Compared to Kazakh, English databases provide more entries to be taken into research.

V. DISCUSSION

This section outlines the information space terms that feature phonetic, orthographic, grammatical, lexical, and semantic interference according to Alimov’s framework.

A. *Phonetic Interference*

Phonetic interference is characterized by the presence or absence of differences in the primary language, different pronunciations of equivalent phonemes; or different phoneme systems (Abazova, 2019, p.147). Looking at the corpus entries, the pronunciation of *coronavirus* in Kazakh and English demonstrates interference at the phonetic level.

(1) *Koronavirus /korona'virus/: Taraz «sary aimaqqa» endi.* [transl. 'Coronavirus: Taraz entered the "yellow zone"']. [<https://lat.egemen.kz/> (in Kazakh, accessed 20 November 2020)]

(2) *Precautions were taken to try to limit the spread of coronavirus /kə'rəʊnəvaɪrəs/* [<https://www.oxfordlearnersdictionaries.com/> (in English, accessed 29 January 2021)]

The Kazakh term *koronavirus* is the English replication. It is pronounced the same way as in English, except for the third syllable – vowel sound /i/ instead of /ai/ due to the Kazakh language's phonetic rules. Other terms show the same interaction between the two languages: the English *cyber* /'sərbə(r)/ changes into Kazakh /'kiber/, the English *crisis* /'kraɪsɪs/ is pronounced as /krizis/ in Kazakh. It means that the recipient language adopts the term under its phonological system principles that go to possible phonological errors by the Kazakh speakers.

B. Grammatical Interference

Grammatical interference implies grammatical norms of L1 to be used by analogy in the recipient language. It may become a source of errors in punctuation, morphological and syntactic structures in the language. For example, in Kazakh, there are *pandemia* as a noun and its derivative *pandemialyq* as an adjective for only one form *pandemic* for both speech categories in English. So, in Kazakh, it may occur when *pandemic* is wrongly interpreted either a noun or an adjective. Moreover, this term is translated as *indet* into Kazakh, but it is of less use in communication due to the influence of English. The following examples can demonstrate the English term *pandemic* (5, 6) and its Kazakh invariants *pandemia* (n) (4, 5) and *pandemialyq* (adj) (3) in information space sources.

(3) *Pandemialyq tūmauğa qarsy otandyq vaksina daiyndau tehnologialaryn ...* [transl. Qarjylandyrudy qajet etpeidi] [https://adilet.zan.kz/kaz/docs/R090000120_ (in Kazakh, accessed 29 January 2021)]

(4) *Vaksinalau pandemiany (noun) aýzdyqtai ala ma?* [transl. 'Can vaccination stop the pandemic?'] [<https://lat.egemen.kz/> (in Kazakh, accessed 25 January 2021)]

(5) *Experts warned of the impending threat of a global pandemic (noun) if the virus was not contained.* [www.collinsdictionary.com (in English, accessed 29 January 2021)]

(6) *There's no denying the pandemic (adjective) changed style for good.* [www.merriam-webster.com (in English, accessed 29 January 2021)]

C. Orthographic Interference

Orthographic interference occurs when there is a transfer of the rules for writing in L1 into L2. It leads to spelling errors and visual inconsistencies. From this perspective, due to the influence of one language's spelling norms, words are written in another language. Thus, to confirm this statement, we illustrate the term *offline* which is spelt differently but pronounced identically.

(7) *Oflain (ózindik, bólektengen) rejimi.* [transl. 'Offline (own, dedicated) mode'] [<https://kk.wikipedia.org/> (in Kazakh, accessed 31 January 2021)]

(8) *He suggested that we should take the discussion offline.* [<https://dictionary.cambridge.org/> (in English, Accessed 31 January 2021)]

The provided above show the variations in spelling. To Kazakh the term *oflain* is borrowed from the English *offline*, where it is written with a doubled consonant. Eventually, it loses the doubled consonant *ff*, though the pronunciation in both languages is similar.

Other examples of spelling interference can be witnessed in the following terms. *Startup* in English and *startap* in Kazakh, where Kazakh spelling norms affect the English borrowing in substituting the vowel /u/ with /a/ in the second syllable. Double consonant *cc* and silent /e/ in English *vaccine* change into *ks* and *a* in Kazakh *vaksina* respectively. English *medic* is spelt like *medik*, English *content* becomes *kontent* in Kazakh. The English term *lockdown* is borrowed as *lokdaun*, *fake* becomes *feik* in Kazakh, English *webinar* turns into *vebinar*, challenge as *chellenj*, and *life hack* is borrowed as *laifhak* as one term. Every pair of terms is similar in pronunciation in the given languages but they bear differences in spelling due to the impact of Kazakh orthographic norms.

D. Lexical Interference

It typically means the direct transfer of the vocabulary of one language system to another. Lexical interference occurs with the use of foreign-language words and phrases in the speech of L2. Karlinskiy (2011, p. 180) admits the occurrence of this type mostly, saying "this pattern is primarily since the vocabulary and phraseology of any language are an open system, less strictly organized in comparison with the units and rules of the phonetic and grammatical levels".

Table 4 shows the information space terms featuring lexical interference as they are borrowed into Kazakh by the relevant word forms with minor changes at the phonetic and morphological levels. There are twenty-two items that constitute two-thirds of all entries.

TABLE 4
INFORMATION SPACE TERMS SHOWING LEXICAL INFLUENCE

English term	=	Kazakh term
coronavirus	=	koronavirus
quarantine	=	karantin
online	=	onlain
internet	=	internet, ǵalamtor
president	=	prezident
COVID-19	=	COVID-19
rating	=	reiti ñ
pandemic	=	pandemia, indet
startup	=	startap
medic	=	medik
trend	=	trend
economics	=	ekonomika
content	=	kontent
vaccine	=	vaksina
cyber	=	kiber
ZOOM	=	ZOOM
offline	=	oflain
lockdown	=	lokdaun
fake	=	feik
webinar	=	vebinar
lifehack	=	laifhak
challenge	=	chellenj

Let us focus on two terms in Kazakh that suffer from linguistic interference. First, *ǵalamtor* (*internet* in English) shows the influence of the English language greatly regardless of possessing the Kazakh equivalent, native speakers still use English borrowing. Secondly, *indet* (*pandemic* in English) is mostly substituted by its Russian equivalent *pandemia*. Online newspaper platforms apply *internet* and *pandemia* in most cases.

(9) *Qazaqstan Respublikasynda tūmau pandemiasyna den qoiudyñ ūltyq josparyn bekıtu turaly*. [transl. On approval of the national flu pandemic response plan in the Republic of Kazakhstan https://adilet.zan.kz/kaz/docs/R090000120_ (in Kazakh, accessed 2 February 2021)]

(10) *Pandemia – qauip ārı mūmkındık*. [transl. Pandemic-threat and opportunity. <https://egemen.kz/article/254006-pandemiya-%E2%80%93-qauip-ari-mumkindik> (in Kazakh, accessed 2 February 2021)]

(11) *30 krupneiřih torgovyh internet-plořadok – 2020* [transl. 30 largest online trading platforms - 2020 https://forbes.kz/leader/30_krupneyshih_torgoviyh_internet-ploščadok_-_2020_1604068634/ (in Russian, accessed 2 February 2021)]

Analysing our findings, we strongly agree with Karlinsky's idea that a language's vocabulary suffers another language's influence in the language contact process to the greatest extent. Besides, we admit that together with the dominant lexical interference, the search terms experience orthographic interference resulting in possible spelling errors.

E. Semantic Interference

This type of language interaction appears at the level of meaning under the impact of the source language. The polysemy, homonymy, and synonymy of grammatical forms are common for it, standing for the reasons for its occurrence. The main semes' inadequate identification is most often noted when using prefixes, suffixes, prepositions, conjunctions, various parts of speech, forms of tense, and mood (Nikiforova & Gredina, 2014, p.635). The specific situation of communication determines the meaning of the term. To correctly use the terms, it is necessary to determine the semantic expediency of their use according to the scope of the concept's semantic structure (Anafinova, 2012, p.71).

To illustrate, we propose examples of semantic interference in *corona* that has come to possess additional meaning due to the pandemic. Initially, the direct meaning refers to the king's or queen's headwear. *Corona* bears a pandemic-related meaning related to coronavirus (12), while the same term conveys the concept of a head-covering (13, 14, 15).

(12) *The word "Corona" has acquired a new meaning within the new realities. "He has a corona," it is said of someone who contacted the coronavirus*. [<https://tengrinews.kz/> accessed 5 February 2021]

(13) *She is not on a diet, does not wear high-heeled shoes, does not worry in front of the cameras, and does not take offence at malicious comments on social networks. This is how Regina Vandyshva lives, who received two titles last year - "Miss Almaty" and "Miss Kazakhstan". However, what about the stories that beauties get their crowns*

(corona in the original text) thanks to money or patrons? "This is a stereotype. It is destroyed, " - confidently said Regina Vandysheva in an interview with Forbes Woman [https://forbes.kz/ accessed 5 February 2021]

(14) The Akimat of the East Kazakhstan region has received the international award "Golden Crown (corona in the original text) of Quality" of the company Business Initiative Directions, Kazinform reports regarding the press service of the East Kazakhstan region. [https://tengrinews.kz/ accessed 5 February 2021]

(15) The crown (corona in the original text), decorated with diamonds and other precious stones, was exhibited in a particular case. [https://dictionary.cambridge.org/dictionary/ accessed 5 February 2021]

The given examples correspond to the dual meaning of *corona* proving semantic interference. People do not imply its direct meaning as a crown but understand it as pandemic disease. We may assume that this transferred meaning is relevant due to the COVID-19 spread. As soon as the global and local situations improve, the reduced form *corona* with the meaning of *coronavirus* may not be used so far.

Overall, having analyzed the information space terms, we reveal linguistic interference features at different language levels. The given examples convey phonetic, orthographic, grammatical, lexical, and semantic interference according to Alimov's differentiation of linguistic interference. It results from the language contact where English serves as a source language, whereas Kazakh is a recipient one.

VI. CONCLUSION

In contemporary Kazakh linguistics, the issue of interference between Kazakh and English has become a highly discussed issue in the scholarly community. We propose the study on linguistic interference focusing on its differential features in information space terms. Language corpus outcomes stand for research instrument, while Alimov's typology is served as a framework.

The corpus built and processed with #LancsBox program, extracts the list of the most frequently used terms from major online newspaper platforms' websites in Kazakhstan. Most of the terms are borrowed from English, while there are those of Kazakh and Russian origin. Amongst all information space terms, the term *coronavirus* possesses the highest numbers in occurrence and frequency that may define it as the Word of the Year 2020 in Kazakhstan. In addition, searching the global corpora, we present an overview of the term *coronavirus*. We admit that two domestic corpora are relatively small and need to be enlarged to provide sufficient data for research. Following Alimov's framework, we analyse the terms that feature interference at the phonetic, orthographic, grammatical, lexical, and semantic levels.

We consider that the findings of the study are important in several respects. At first, they provide another dimension to studies on the linguistic interference in Kazakh from a typological perspective. Secondly, further studying can benefit from using the corpus outcomes as they give the linguistic situation in information space sources in 2020 in Kazakhstan.

Overall, the paper has aimed to analyse linguistic interference in information space terms through corpus tools comparing Kazakh and English. More specifically, we presented: 1. Kazakhstani information sources; 2. the corpus built and processed with #LancsBox program; 3. the search of the term in the foreign and domestic corpora; 4. linguistic interference and its differential features in the information space terms in a multilingual environment.

ACKNOWLEDGEMENTS

The authors wish to thank the developers and mentors of the course "Corpus Linguistics: Method, Analysis, Interpretation" at Lancaster University for sharing the theoretical and practical material, a corpus tool program #LancsBox 5.1.2. (McEnery 2020)

REFERENCES

- [1] Abazova, M. M. (2019). Language interference (on the example of the speech of the Kabardino-Circassian Diaspora abroad and the literary Kabardino-Circassian language). *Modern studies of social problems*, 4, 123-136.
- [2] Alimov, V. V. (2015). *Theory of translation: A manual for linguists-translators*. Moscow: Lenand.
- [3] Anafinova, M. L. (2012). *International discourse and specific terminology*. Kostanay: Kostanay Printing House.
- [4] Anafinova, M. L., & Ormanova, A. B. (2020). Language interference: overviewing the fundamentals. *Bulletin of Kokshetau state university*, 3, 260-266.
- [5] Brezina, V., Weill-Tessier, P., & McEnery, T. (2020). *#LancsBox 5.x [software]*. Retrieved November 12, 2020, from <http://corpora.lancs.ac.uk/lancsbox/>.
- [6] Dobrovolskaya, I. A. (2014). The concept of "information space": various approaches to its study and features, *Bulletin of Peoples' Friendship University of Russia. Series: Literary Studies. Journalism*, 4, 140-147.
- [7] Drouin, P. (2004). *Detection of domain specific terminology using corpora comparison*. Retrieved January 12, 2020, from https://www.researchgate.net/publication/228736875_Detection_of_domain_specific_terminology_using_corpora_comparison/.
- [8] Dulay, H., & Burt, M. (1976). Creative construction in second language learning and teaching. *Language Learning*, 4, 65-79.
- [9] Goswami, A. (2020). Changing Contours: The Interference of the Mother Tongue on English Speaking Sylheti Bengali. *Journal of English as an International Language*, 2, 145-177
- [10] Garside R., Leech G., & McEnery T. (2013). *Corpus Annotation*. Taylor & Francis: Routledge, 2013, 1-4. ISBN 13:978-0-583-29837-8
- [11] Heine B., & Kuteva T. (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.

- [12] Karaseva, N. V. (2016). Information space as a factor of youth socialization, *Bulletin of Karaganda state university*, 84(4), 111-116.
- [13] Karlinskiy, A. E. (2011). *Interaction of languages: bilingualism and language Contacts*, Almaty: Abylai Khan Kazakh University of International Relations and World Languages.
- [14] Kaweera, C. (2013). Writing error: a review of interlingual and intralingual interference in EFL context, *English language teaching*, 7, 9-18.
- [15] Linguistic Encyclopedia (LES). (1990). Editor-in-Chief V. N. Yartseva. Moscow "Soviet Encyclopedia" 1990. Retrieved February 15, 2021, from <http://tapemark.narod.ru/les/>.
- [16] McEnery, T. (2020). *The course 'Corpus Linguistics: Method, Analysis, Interpretation'*. Retrieved November 12, 2020, from <https://www.futurelearn.com/courses/corpus-linguistics/8/steps/868172>.
- [17] MOOC "Corpus Linguistics: Method, Analysis, Interpretation". Retrieved November 12, 2020, from <http://corpora.lancs.ac.uk/lancsbox/help.php#tabs1/>.
- [18] Narozhnaya, V. D. (2018). Development of Kazakhstani linguistic science at the present stage, *Bulletin of Kokshetau State University*, 1, 107-116.
- [19] Nikiforova, K. A., & Gredina, I. V. (2014). Interference in the terminology of the oil and gas industry and thermodynamics, *Young scientist*, 7 (66), 635-636.
- [20] Odlin, T. (1989). Earlier thinking on transfer. In *Language Transfer: Cross-Linguistic Influence in Language Learning (Cambridge Applied Linguistics)*, pp. 6-24. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524537.004
- [21] Pozdnyakova, A.A. (2012). Linguodidactic methods of overcoming interference in the speech of bilingual students. *Teacher of the XXI century*, 1(3), pp.161-165.
- [22] Rosenzweig, V. Yu. (1972). *Language contacts*. Leningrad: Nauka.
- [23] Sager, J. C. (1990). *A practical course in terminology processing*. Amsterdam: John Benjamins.
- [24] Thyab, R. A. (2016). Mother tongue interference in the acquisition of English articles by L1 Arabic students. *Journal of Education and Practice*, 7(3), 1-4
- [25] Varshney, A. (2011). *Overlapping in secondary sources of Information in Social Science 1995-2000: An Evaluative study*. Retrieved February 14, 2011, from <http://shodhganga.inflibnet.ac.in:8080/jspui/handle/10603/40587>.
- [26] Weinreich, U. (1979). *Language contacts*. Translated from English and commented by Yu. A. Zheluktenko; Intro. article by V. N. Yartseva. Kiev: Vyshcha shkola. Publishing house at Kiev. Uni.
- [27] Zhikeyeva, A. R. (2017). Features of the formation and development of trilingualism in The Republic of Kazakhstan on the example of Kostanay region. *Pedagogical Education in Russia*, 11, 37-40.



Assel Bakhytovna Ormanova graduated from Kostanay State University, Kazakhstan, in 2008, as Bachelor of Foreign Languages. Obtained MD at Omsk State University, Russia, in 2013. Currently a PhD candidate.

She is a Lecturer at Astana IT University, at the Department of General Educational Disciplines, Astana, Kazakhstan. The author of a textbook, ten articles. The current research interests include terminology, bilingualism, corpus linguistics, LSPs, computer sciences, etc. ORCID ID 0000-0003-3265-6111



Madina Latypovna Anafinova graduated from North Kazakhstan University of Petropavl, Kazakhstan, in 1979. Obtained CandSc Degree in General Linguistics in Abylai Khan Kazakh University of International Relations and World Languages of Almaty, Kazakhstan, in 2010. Passed a scientific internship under the Bolashak program in University of Cambridge, the UK, in 2013.

She is an associate Professor at the Foreign Philology Department, Faculty of Philology, L. N. Gumilyov Eurasian National University, Kazakhstan. The author of two monographs, two dictionaries, four textbooks, more than fifty articles. The current research interests include: lexicology, problems of terminology development, ESP, etc.

Ms. Anafinova supervises research of master and PhD seeking students. ORCID ID 0000-0002-8523-0010