A.B. Ormanova[1]* , M.L. Anafinova[1] ,
S.D. Atanasova-Divitakova [2]

[1] L. N. Gumilyov Eurasian National University, Kazakhstan, Nur-Sultan
[2] St Cyril and St Methodius University of Veliko Turnovo, Bulgaria, Veliko Turnovo
*e-mail: ormanova_ab_1@enu.kz

CORPUS BASED RESEARCH IN TERMINOLOGY

Language exists and develops due to intralinguistic and extralinguistic factors. In connection with the global situation, most world dictionaries have a common feature – the emergence of new terms characterizing the spread of the pandemic, which has had a significant impact on the development of the human language. Since 2020, the term "lockdown", previously used only in prison terminology, has begun to emerge in common vocabulary. The term is gaining popularity and is often applied by media. However, the semantics of this term is not fully revealed. In this regard, our research interest lies in the study of "lockdown" from the point of view of linguistic science in general and the achievements of corpus linguistics in particular. For the analysis, we reveal its origin and formation, present the definitions of "lockdown", proposed by world dictionaries, and the search results in the corpora of the Kazakh, Russian and English languages. Additionally, we carry out a corpus analysis of the search term in the research corpus created with the #LancsBox program. Thus, we receive up-to-date data on the use of the term "lockdown" in the current Kazakh linguistic environment. The results cover concordance lines, occurrence, frequency lists, linguistic markings, and the relevant context.

Key words: corpus tool, terminology, corpora, term, lockdown.

А.Б. Орманова[1]*, М.Л. Анафинова[1], С.Д. Атанасова-Дивитакова[2]

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Қазақстан, Нұр-Сұлтан қ.
[2] Әулие Кирилл және Мефодий Велико-Тырново университеті, Болгария, Велико-Тырново қ.
*e-mail: ormanova_ab_1@enu.kz

Терминологиядағы корпустық зерттеулер

Тіл түрлі тілішілік және экстралингвистикалық факторлардың әсерінен бар және дамиды. Жаһандық жағдайға байланысты әлемдік сөздіктердің көпшілігінің ортақ ерекшелігі бар – пандемияның таралуын сипаттайтын жаңа терминдердің пайда болуы, олар адам тілінің дамуына елеулі әсер етті. 2020 жылы бұрын тек түрме терминологиясында қолданылатын "локдаун" термині жалпы сөздік қорда қолданыла бастады. Бұл термин тез танымал бола бастады және оны саясаткерлер мен БАҚ жиі қолданады. Алайда бұл терминнің семантикасы толық ашылмаған. Осыған байланысты, біздің зерттеу қызығушылығымыз "локдаун" терминін жалпы лингвистикалық ғылым тұрғысынан және жеке корпус лингвистикасының жетістіктері тұрғысынан анықтауымыз керек. Талдау үшін біз "локдаун" терминнің шығу тегі мен қалыптасу процесін ашамыз, әлемдік сөздіктер ұсынған «құлыптау» анықтамаларын және қазақ, орыс және ағылшын тілдерінің корпустарында іздеу нәтижелерін ұсынамыз. Сонымен қатар, біз #LancsBox компьютерлік бағдарламасының көмегімен құрылған зерттеу корпусында іздеу терминіне корпус талдауын жүргіздік. Осылайша, қазіргі қазақ тілдік ортасында "лакдаун" терминінің қолданылуы туралы өзекті деректерді алдық. Нәтижелер үйлесімділік сызықтарын, пайда болуын, жиіліктер тізімін, лингвистикалық белгілерді, контексті қамтиды.

Түйін сөздер: корпус құралы, терминология, корпус, термин, локдаун.

А.Б. Орманова[1]*, М.Л. Анафинова[1], С.Д. Атанасова-Дивитакова[2]

[1]Евразийский национальный университет им. Л. Н. Гумилева, Казахстан, г. Нур-Султан
[2]Великотырновский университет Святых Кирилла и Мефодия, Болгария, г. Велико-Тырново
*e-mail: ormanova_ab_1@enu.kz

Корпусные исследования в терминологии

Язык существует и развивается под воздействием разнообразных внутриязыковых и экстралингвистических факторов. В связи с глобальной ситуацией для большинства мировых словарей характерна общая особенность – появление новых терминов, характеризующих

распространение пандемии, что оказало значительное влияние на развитие человеческого языка. В 2020 году термин «локдаун», который ранее был использован только в тюремной терминологии, стал применяться и в общеупотребительной лексике. Термин быстро приобретает популярность и часто используется в СМИ. Однако семантика данного термина не раскрыта в полной мере. В связи с этим наш исследовательский интерес заключается в изучении термина «локдаун» с точки зрения лингвистической науки в целом и достижении корпусной лингвистики, в частности. В процессе анализа мы раскрываем происхождение и процесс образования термина, представляем определения термина, предложенные мировыми словарями, и результаты поиска в корпусах казахского, русского и английского языков. Дополнительно мы провели корпусный анализ искомого термина в созданном нами исследовательском корпусе с помощью программы #LancsBox. Таким образом, мы получили актуальные данные об использовании термина «локдаун» в нынешней казахстанской лингвистической среде. Результаты охватывают линии конкордации, встречаемость, списки частот, лингвистические маркировки, контекст.

Ключевые слова: инструмент корпуса, терминология, корпус, термин, локдаун.

Introduction

A language exists and develops forced by diverse intra and extra linguistic factors. Due to the present global situation, all spheres of human's life are affected, and language is not an exception. Quite a few of the words on the world's dictionaries shortlists have one thing in common: the pandemic. Something that has changed everyone's lives so profoundly, leaving no country or continent untouched. The pandemic has got a significant impact on the human language, enlarging it with a new lexicon. We can currently observe the dynamic nature of Terminology, when terminologization and de-terminologization of the general language occur. Terms initially being a part of a specific domain due to extra-linguistic factors may shift to common language vocabulary. Meanwhile, words may narrow the meaning by acquiring terminological value.

Contemporary science proposes the following definitions of Terminology and a term. Terminology is a set of terms associated with a professional field of activity (area of knowledge, technology, management, and culture) related to each other at the conceptual, lexical-semantic, derivational and grammatical levels (Zhuravleva, 1995: 4). In comparison, a term is meant as a word or a subordinate phrase having special meaning, expressing and forming a professional concept and used in learning and mastering scientific and vocational objects and the relations between them (Utemgaliyeva, Bektemirova, Rivers, 2019: 190).

Following McEnery (2020), linguistic research grants from using corpora firmly. Firstly, large amounts of data tell the research about tendencies, frequencies, and average or typical in real-life language use. Secondly, corpora also reveal rare or exceptional cases that we would not get from looking at single texts or introspection. Moreover, human researchers may make mistakes and are slow compared to computers that are much quicker and more accurate.

So, corpus linguistics instruments are beneficial in research in Terminology as the corpora show relevant use of the terms, settle how they suit and reflect the purpose of communication. Our study focuses on the term "lockdown", which has gained much popularity recently. What does "lockdown" mean? The term's occurrence in the world's corpora, its definition and origin, and the morphological structure stand for our paper's research issues.

Materials and methods

Terminology today has successfully adopted an approach to collecting lexical data based on corpora. According to Sager (1990:58), "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 analyse Terminology in its communicative, i.e. linguistic context, leads to some new theoretical assumptions and also to new methods of compilation and representation".

Traditionally, linguists have used the term "corpus" to designate a body of naturally-occurring (authentic) language data, which can be used as a basis for linguistic research (Garside, Leech, McEnery, 2013:11).

Following McEnery (2020), linguistic research grants from using corpora strongly. Firstly, large amounts of data tell the research about tendencies, frequencies, and what is normal or typical in real-life language use. Secondly, corpora also reveal rare or exceptional cases that we would not get from looking at single texts or introspection. Moreover, human researchers may make mistakes and are slow

compared to computers that are much quicker and more accurate.

Corpora can be general or specific, synchronic or diachronic (Brezina, 2018). A general corpus is a sample of language taken from a vast population. In the case of a general corpus, the population consists of all of the languages people produce during a particular time. It reflects the language as such, while a specific one focuses on a particular genre, author or area of language use. A synchronic corpus samples language around one point in time, whereas a diachronic corpus includes language across different periods.

Our study applied seven world corpora for exploring the term "lockdown", as shown in Table 1. When we searched a term of interest, we got a wealth of information on it, including frequency information, definition, translation, link to audio, images, and videos, synonyms, related word forms, topics, collocates, clusters, and concordance / KWIC lines.

Table 1 – Overview of corpora used

#	Corpus	Size	Representativeness
1	iWeb: The intelligent Web-based Corpus	14 billion	22 million web pages
2	Corpus of Contemporary American English	1 billion	Written and spoken Formal and informal
3	Coronavirus Corpus	725 million	online newspapers and magazines
4	Russian National Corpus	600 million	Written and spoken Formal and informal
5	British National Corpus	100 million	Written and spoken Formal and informal
6	Almaty Corpus of Kazakh language	40 million	Written and spoken Formal and informal
7	National Corpus of Kazakh Language	30 million	Written and spoken Formal and informal

1. "Lockdown" in iWeb

The iWeb corpus contains fourteen billion words in twenty-two million web pages. It is related to other English corpora that English-Corpora.org created to offer unparalleled insight into variation in English. The corpus generated a list of all instances of "lockdown" in the form of a concordance. There are 17,186 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 used (Figure 1).

2. "Lockdown" in the Corpus of Contemporary American English (COCA)

COCA is the only large, genre-balanced corpus of American English. It is probably the most widely-used corpus of English, and it is related to many other corpora of English. The corpus contains more than one billion words of text (25+ million words each year 1990-2019) from eight genres: spoken, fiction, popular magazines, newspapers, academic texts, and (with the update in March 2020): TV and Movies subtitles, blogs, and other web pages. COCA presented 1,445 frequencies of "lockdown". According to statistics, the highest frequency is in TV and media texts. The corpus provides data on collocates, concordance lines, context, clusters and gives an overview of the terms of interest (Figure 2).

3. "Lockdown" in the Coronavirus Corpus

The 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. It is first released in May 2020; currently, it is about 727 million words in size, and it continues to enlarge. The corpus provided a term occurrence and a range of various phrases with "lockdown". There are 198,308 frequencies, with 583 unique. The most frequent occurrence goes to "lockdown" in both singular and plural forms, with 180,120 and 15,659 respectively, while the least one is "lockdown based" with only four frequencies. Each line in the Frequency List supplies the concordance lines. Clicking on any concordance line directs to the relevant expanded context (Figure 3).

Figure 1 – The term "lockdown" searched in the iWeb corpus
https://www.english-corpora.org/iweb/

Figure 2 – The term "lockdown" searched in the Corpus of Contemporary American English.
https://www.english-corpora.org/coca/

Figure 3 – The term "lockdown" searched in the Coronavirus Corpus
https://www.english-corpora.org/corona/

4. "Lockdown" in the Russian National Corpus

The Russian National Corpus is a corpus of the modern Russian language with more than 600 million words. The corpus includes original prose texts representing the Russian literary language and translated works, poetic texts, and texts representing non-literary forms of the modern Russian language: colloquial (recordings of oral speech, public and non-public), dialect. The Russian National Corpus did not provide any appropriate data on the searched term.

5. "Lockdown" in the British National Corpus (BNC)

Created by Oxford University Press in the 1980s – early 1990s, BNC contains 100 million words of texts from various genres (e.g. spoken, fiction, magazines, newspapers, and academic). It is related to many other corpora of English, which offer unparalleled insight into variation in English. The corpus showed two occurrences of "lockdown" in the texts not earlier than 1994.

6. "Lockdown" in the National Corpus of the Kazakh language and the Almaty Corpus of Kazakh language

The volume of the text base of the National Corpus of the Kazakh language is 30 million words. The collected texts were obtained from five Kazakh language styles: art style, scientific style, journalistic style, business style, conversational style. The corpus owns data on concordance, lemmatization, and linguistic markings such as morphological, word-formative, lexical, phonetic, morpho-semantic.

The Almaty Corpus of Kazakh language is one of the versions of the National Corpus of the Kazakh language as a reference system on the basis of exten-

sive fund of the marked texts of literary Kazakh. Indeed, the corpus is constantly being supplemented, updated both quantitatively and qualitatively. Currently, the size of the corpus is more than 40 million word tokens. The corpus texts were marked through the automatic morphological analyzer; 86 % of word forms of the corpus were parsed.

Nonetheless, neither the National Corpus nor the Almaty Corpus provides any relevant data on "lockdown" in the Kazakh language.

Besides, not limiting with the corpora and dictionaries, we have to mention Terminkom.kz, a unified terminological electronic database of terms in science, education, technology and economics, and political and social life in Kazakhstan. The site thrives for 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 the field of Kazakh Terminology as well. Regarding our study, the terminological database provides neither the definition nor an occurrence of the term "lockdown" in the Kazakh language.

Thus, having searched for the term of interest, we conclude that the term "lockdown" is recorded in English corpora. When it comes to Kazakh, we find no data.

Literature review

Language is a reflection of the world around us, and the global pandemic has dominated 2020. "Lockdown" is not a word of the year to celebrate,

but it sums up the year for most of the world. With its heavy, clunking syllables and heavier associations, it is the condition we have most worried about in 2020 (Shariatmadari, 2020).

World dictionaries aim to define the word that best characterizes the past year. According to Collins Dictionary (https://collinsdictionary.com/), "lockdown" is a word of the year, and it determined people's lives across the world in 2020. In a report "The year of Lockdown", the term is defined as "the imposition of stringent restrictions on travel, social interaction, and access to public spaces", and its usage has boomed over the last year. The 4.5bn-word Collins Corpus, which contains written material from websites, books and newspapers, and spoken material from radio, television and conversations, registered a 6,000 % increase in its usage. In 2019, there were 4,000 recorded instances of "lockdown", while in 2020, the number soared to more than a quarter of a million. Collins language content consultant Helen Newstead says: "We have chosen "lockdown" as our word of the year because it encapsulates the shared experience of billions of people who have had to restrict their daily lives in order to contain the virus. Lockdown has affected the way we work, study, shop, and socialize.

According to the Oxford English Dictionary (OED) (https://www.oed.com/), 2020 is not a year that could neatly be accommodated in one single "word of the year", so in "Words of an Unprecedented Year" report OED proposed more expansively on the phenomenal breadth of language change and development over the year.

In Kazakhstan, the language dictionary sources (lugat.kz, emle.kz, termincom.kz, sozdikqor.kz, sozdik.kz, tilalemi.kz) do not provide statistical data on the issue. The term "lockdown" appeared on the news platforms Tengrinews.kz, in the report "*Head of the Ministry of Health Alexey Tsoi presented several scenarios for the introduction of lockdown in Kazakhstan*" on July 1. A.Tsoi used an unusual concept instead of the common term "*quarantine*". Since then, the term "lockdown" has been broadly introduced in Kazakhstan's information space.

To start our study, we focus on the ways how the world dictionaries define "lockdown".

1. A lockdown is a requirement for people to stay where they are, usually due to specific risks to themselves or others if they can move freely. The term "stay-at-home" or "shelter-in-place" is often used for lockdowns that affect an area, rather than specific locations (https://en.wikipedia.org/wiki/Lockdown);

2. A) A lockdown is a security measure in which those inside a building or area are required to remain confined in it for a time;

B) A lockdown is the imposition of stringent restrictions on travel, social interaction, and access to public spaces (https://www.collinsdictionary.com/dictionary/english/lockdown);

3. A) A lockdown is the confinement of prisoners to their cells for all or most of the day as a temporary security measure

B) A lockdown is an emergency measure or condition in which people are temporarily prevented from entering or leaving a restricted area or building (such as a school) during a threat of danger;

C) A lockdown is a temporary condition imposed by governmental authorities (as during the outbreak of an epidemic disease) in which people are required to stay in their homes and refrain from or limit activities outside the home involving public contact (such as dining out or attending large gatherings) (https://www.merriam-webster.com/lockdown)

4. A) A lockdown is an occasion or time when prisoners are locked in their cells;

B) A lockdown is an occasion or time when access to a place is restricted because of some danger;

C) A lockdown is a time when large numbers of people are ordered to stay at home either most or all of the time. https://www.macmillandictionary.com/dictionary/british/lockdown

5. A) A lockdown is an emergency situation in which people are not allowed to freely enter, leave, or move around in a building or area because of danger;

B) A lockdown is a period of time in which people are not allowed to leave their homes or travel freely, because of a dangerous disease. https://dictionary.cambridge.org/dictionary/english/lockdown?q=lockdown

As for the Kazakh dictionaries, there has no relevant definition provided so far. On the platform massaget.kz there are two articles concerning the topic of the term "lockdown" in the information space use. The article dated July 7 states the new term "lockdown" appeared first in the speech of the head of the Ministry of Health of Kazakhstan and since then has been in everyday use (Slyambekov, 2020). The article dated November 11 informed shared the information about Word of the Year 2020 by Collins Dictionary (Usenova, 2020). No Kazakhstani linguists have taken that issue into the research yet.

Our research follows the Cambridge Dictionary definition that calls "lockdown" an emergency when people cannot freely enter, leave, or move around in a building or area because of danger.

Speaking about the historical background, "lockdown" was originally a piece of prison vocabulary. It described the confinement of prisoners to their cells for all or most of the day as a temporary security measure. According to Merriam-Webster, the first known use of "lockdown" in print was in 1973.

Terms generally may be formed as simple words/complex phrases with specific morpho-syntactic and morpho-semantic features that may be unknown in general language, but that can be specific to certain domains. Term formation depends on the functional role of designations in domain communication. The interaction between domain languages and general language is very dynamic. There is a constant flux of lexical material in both directions that define the processes such as:

- *de-terminologization* is when specialized terms are incorporated into general language as widely known words;

- *terminologization* is when common words become (a part of) terms (Guidelines for Terminology policies, 2005).

In most cases, the words/terms do not lose their original meaning but instead acquire a dual meaning, thus becoming polysemantic.

Considering the term "lockdown" originating from prison vocabulary, it currently appears as a common language. Thus, de-terminologization stands for forming the term "lockdown", as shown in Figure 4.
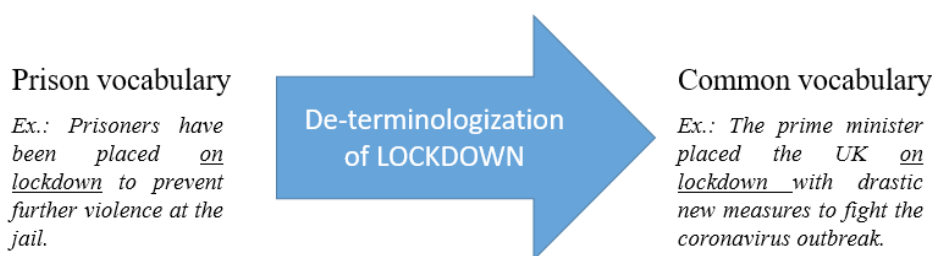
Prison vocabulary

Ex.: Prisoners have been placed <u>on lockdown</u> to prevent further violence at the jail.

De-terminologization of LOCKDOWN

Common vocabulary

Ex.: The prime minister placed the UK <u>on lockdown</u> with drastic new measures to fight the coronavirus outbreak.

Figure 4 – Term formation of "lockdown"

The most productive methods of creating words in term formation are semantic, syntactic and morphological (Kasatkina, 2020). To define a structural type of a word and accomplish a morphological analysis, we implied the analyses of constituent structure proposed by L. Bloomfield (Langendoen, 1998). In English, "lockdown" is formed as a compound noun consisting of the lexical morpheme "lock", which means *to shut, to block, to fix*, and the grammatical morpheme "down" to strengthen the lexical meaning.

Since English compounds have the unity of morphological and syntactical functioning, they are used in a sentence as one part of it, and only one component may change grammatically. According to the way, components are joined together, "lockdown" stands for morphological, where a linking element join components. The example below can prove it.

Example: A third of the world's population now under *lockdown*. (bbc.co.uk)

According to the meaning of the whole compound, "lockdown" goes to non-idiomatic, because its components are not different in their meaning from corresponding free phrases. According to the relations between the components, "lockdown" is characterized by subordinate relations. In "lockdown", the first component is the semantic and structural center, while the second is subordinate. By order of the components, there is a direct order within the components in "lockdown". Thus, the term "lockdown" is an instance of compounding in English.

Considering the Kazakh language, the term "lockdown" has emerged through transliteration from English. It has a form "локдаун" in Cyrillic and "lokdaýn" in Latin letters. In the texts of the information space, the term occurs in Cyrillic letters mostly, rarely in English, as shown in the sentence below:

Example: Что такое «*локдаун*» и чего ждать от него казахстанцам? (What is «*lockdown*" and what to expect from it for Kazakhstanis?)

In Kazakh, "lockdown" is translated as "isolation"; but the term "lockdown" is thoroughly used in Kazakh speech, becoming a representative of borrowings.

Results and discussion

Since corpora of Kazakh do not offer information on the term of interest – "lockdown", we managed to build the research corpus (hereafter RC) of nearly 300 thousand words with the help of a corpus tool #LancsBox 5.1.2 (http://corpora.lancs. ac.uk/lancsbox/docs/pdf/LancsBox_5.1_manual. pdf). Following V. Brezina, one of the developers, this software 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, Weill-Tessier, McEnery, 2020).

To process the RC, we loaded 251 publications from the online news platform Tengri News published in 2020. Our research used wildcard searches with a special character – * to find all variants of "lockdown". The program adds the following annotation: 251 files, 292881 tokens, 51500 types, 51706 lemmas.

All analyses were produced using the functional tools: KWIC, GraphColl, Whelk, Words, Ngrams, Text. In our research, we were doing two searches as the term appeared both in Latin and Cyrillic letters; we employed asterisk with it (*) – "lockdown*" and "локдаун*" in the texts of our created corpus. As two variants are eligible, we present results individually.

1. KWIC

The KWIC tool generates a list of all instances of a search term in a corpus in a concordance form. It is applied for finding the frequency of a word/ phrase in a corpus or different word classes such as nouns, verbs, adjectives; finding complex linguistic structures such as the passives, split infinitives using 'smart searches'. Finally, the primary function is sorting, filtering and randomizing concordance lines in a corpus.

We apply the frequency list and examine the concordance lines. The search term "lockdown" occurs six times with the relative frequency of 0.20 per 10k words in 1 out of 251 texts. As for "локдаун", it occurs 25 times with the relative frequency of 0.85 per 10k words in 7 out of 251 texts. Overall, "lockdown" emerges 31 times with the relative frequency of 1.05 per 10k words in 7 out of 251 texts.

2. GraphColl

The GraphColl tool identifies collocations with an item of interest and displays them in a table and collocation graph or network. It is applied for finding and visualizing the collocates and colligations of a word or phrase, identifying shared collocates and summarizing discourse in terms of its 'aboutness'.

The tool allows to identify and visualize the collocations. In Figure 9, one collocate (preposition in/в) of "lockdown" in the right position, and three collocates of "локдаун" (preposition in/в, conjunctions and/и, that/что) are displayed. There is one shared collocate (в). There are three dimensions presented in a graph: the strength of collocation (6.0) as indicated by the particularly association measure, the frequency of collocation (6) and the position (right). In Figure 5, the closer the collocate was to the node (that is "lockdown"), the more strongly they were associated. The collocate position "in" around the node "lockdown" was the actual position in which the collocate appeared in the text right of the node. The following context can indicate it.

Ex.: "...Алексей Цой представил несколько сценариев введения локдауна в Казахстане. Он использовал…"
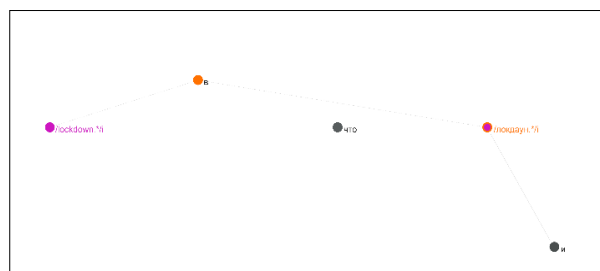

Figure 5 – Collocation network: lockdown and локдаун in the research corpus

3. Whelk

The Whelk tool provides information about how the search term is distributed across corpus files. It is used to find absolute and relative frequencies of the search term in corpus files and then filter the results according to different criteria; sort files according to the search term's absolute and relative frequencies.

Using this tool, we view the distribution of "lockdown"/"локдаун» inside the corpus and individual text files in the corpus. There are two panels; the top one displays the concordance lines, including our search terms, while the bottom panel showed the distribution among the individual files. The concordance lines on the top displayed the information on the occurrence of the search terms, in the bottom, the tool provided the information on the number of tokens in each file and the frequency. "Lockdown" showed frequency 6, relative frequency per 10k that was 80.75371 in the short article of July 1, where there are only 743 tokens. The article's title is "What is "lockdown" and what to expect from it for Kazakhstanis?"

As for "локдаун", its occurrence is 25 (0.85 per 10k) in 7 texts. The highest frequency is observed in the texts (743 tokens) on July 1 with the frequency of 18 (relative frequency is 242.26111 per 10k). The frequency 2 related to the article "Coronoid and quarantine card. What new words have appeared due to COVID-19?" (683 tokens) on July 7. The lowest frequency 1 could be found in fives articles in September, October, November that may mean the decline in that term's usage.

4. Words

The Words tool allows in-depth analysis of frequencies. It is applicable for computing frequency and dispersion measures for the item of interest and visualising frequency and dispersion in corpora. The tool compares corpora using the keyword technique and visualises them.

The Words tool demonstrated the most frequent type in the corpus that is the preposition "в" (in) whose frequency is 9680 and dispersion is 0.0380298. As for "локдаун", its frequency is 13 and dispersion is 14.551373, while the indicator of the frequency of "lockdown" is lower that is 6, and 15.811388 for dispersion in the corpus.

5. Ngrams

The Ngrams tool allows in-depth analysis of frequencies of ngram types, lemmas and POS categories, as well as comparison of corpora using the key ngram technique. It is used to compute and visualise frequency and dispersion measures for ngram types, lemmas and POS tags; compare corpora using the key ngram technique. We do not apply that tool for our research as we find it unreasonable to compare the two corpora in different languages with a significant discrepancy in size.

6. Text

The Text tool enables an in-depth insight into the context in which a word or phrase is used. It is used for viewing a text or a search term in full context, previewing a corpus as a run-on text, and checking different levels of annotation of a text/corpus.

We compiled 251 texts loaded in the #LancsBox that allows previewing all the corpus files ready together to be processed for our research. By navigating, we find all cases of occurrence "lockdown"/"локдаун". The tool helps define the exact articles where our terms of interest appear with the date and context (Figure 6). The search term was mentioned especially in July 2020, when the prerequisites for the local lockdown in the country were discussed.

Our research interest is to investigate currently much used "lockdown" from linguistic and corpus linguistics' perspectives. To analyse this term, we reveal its definitions by the world's dictionaries, its origin, term formation process, search for this term in the seven corpora of three languages – Kazakh, Russian and English. Thus, the English corpora provided frequencies, concordance lines, collocations, context. In contrast, nor Kazakh neither Russian corpora recorded the usage of "lockdown". Not finding any data on this term in Kazakhstani linguistic sources, we were forced to build the research corpus with the #LancsBox program developed by Lancaster University. The corpora tool supplied with the data on frequency, concordance lines, dispersion, collocates, and the context in Kazakhstan's current linguistic environment. Overall, the research corpus carried 251 online articles. In the table below, we want to present our findings – the characteristics of both "lockdown" and "локдаун".

Figure 6 – The "Text" mode possessing "локдаун" in the research corpus

Table 2 – Research corpus based characteristics of "lockdown" and "локдаун"

A tool	Characteristics	
	«lockdown»	«локдаун»
KWIC	occurs 6 times in 1 out of 251 texts; relative frequency 0.20 per 10k words.	occurs 25 times in 7 out of 251 texts; relative frequency 0.85 per 10k words.
GraphCol	1 collocate of lockdown (в)	3 collocates of локдаун (в, что, и)
Whelk	frequency 6; relative frequency 80.75371 per 10k; in the text (743 tokens) on July, 1;	occurrence 25 relative frequency 0.85 per 10k; in the text (743 tokens) on July, 1; the highest frequency (18) in the article (743 tokens) on July, 1, (relative frequency 242.26111 per 10k); the lowest frequency (1) in five texts (19.09.2020, 10.10.2020, 15.10.2020, 5.11.2020, 22.11.2020)
Words	frequency 6; dispersion 15.811388.	frequency 13; dispersion 14.551373.
Ngrams	not applicable	not applicable
Text	in the text 1.07.2020	in the texts: 1.07.2020, 7.07.2020, 13.09.2020, 9.10.2020, 12.10.2020, 30.10.2020, 3.11.2020.

Thus, Table 2 shows that the terms "lockdown" and "локдаун" appeared in use starting from July onwards. Comparing the occurrence in the online articles, "локдаун" in Cyrillic had four times higher rates than "lockdown". The frequency of "локдаун" was twice higher than the same of "lockdown". The distribution of texts with «локдаун» was prevailing and included not only publishing in July but other months of 2020. All these characteristics prove that «локдаун» has become a general language, it has not lost its original meaning, but instead acquired a dual meaning, thus becoming polysemantic.

Conclusion

The pandemic has led to the emergence of new words and phrases in languages around the globe. Some terms may change their primary meaning, others are already officially recorded in dictionaries, while some still exist as jargons and neologisms. Terminology has successfully applied the new approach to reveal and analyze terms with the help of corpus linguistics instruments.

Our research focused on the term "lockdown", which has gained much popularity recently. Its meaning, occurrence in the world's corpora, definitions in English, Russian and Kazakh, origin and morphological structure were in focus in the study. New transformations of the definition of this term based on multilingual corpora were taken into con-

sideration. "Lockdown" was originally a piece of prison vocabulary, but due to the terminologization, currently, the meaning of the term shifted irrevocably: in most people's minds, it is now a public health measure. Following Merriam-Webster dictionary, we propose the following definition of lockdown as "a temporary condition imposed by governmental authorities (as during the outbreak of an epidemic disease) in which people are required to stay in their homes and refrain from or limit activities outside the home involving public contact". Also, we apply seven language corpora for exploring the term "lockdown" in Kazakh, English and Russian with presenting concordance lines, occurrence, frequency lists, linguistic markings, and the relevant context. Since Almaty Corpus of Kazakh language and National Corpus of Kazakh Language do not offer sufficient information on usage of "lockdown", the research corpus was built with the help of a corpus tool #LancsBox 5.1.2 program. The program adds the following annotation: 251 files, 292881 tokens, 51500 types, 51706 lemmas. The publications were downloaded from the online news platform TengriNews published in 2020. All analyses were produced using the functional tools: KWIC, GraphColl, Whelk, Words, Ngrams, Text. More specifically the term emerges 31 times with the relative frequency of 1.05 per 10k words. Our research has shown that the term "lockdown" is in wide use in online resources and has become a part of our daily lives.

References

BBC [Electronic resource] : Accessed: https://www.bbc.com/ – 26.11.2020

Brezina V. (2018) Statistika v korpusnoj lingvistike. Prakticheskoe rukovodstvo [Statistics in Corpus Linguistics: A practical guide]. *Cambridge University Press*, 15p.

Brezina V., Weill-Tessier P., McEnery T. (2020). #LancsBox 5.x [software]. Available at: http://corpora.lancs.ac.uk/lancsbox

Cambridge dictionary [Electronic resource]: Accessed: https://dictionary.cambridge.org/dictionary/english/lockdown?q=lockdown/ – 17.12.2020

Collins dictionary [Electronic resource] : Accessed: https://collinsdictionary.com/ – 2.12.2020

Corpus of Contemporary American English [Electronic resource] : Accessed: https://www.english-corpora.org/coca/ – 26.11.2020

Dictionary by Merriam-Webster [Electronic resource] : Accessed: https://www.merriam-webster.com/lockdown/ – 17.12.2020

Garside R., Leech G., McEnery T. (2013) Annotaciya korpusa [Corpus Annotation]. *Routlegde* 250p.

iWeb: The 14 Billion Word Web Corpus [Electronic resource] : Accessed: https://www.english-corpora.org/iweb/ – 26.11.2020

Kasatkina M.S. (2020) Metody` terminoobrazovaniya v sovremennom anglijskom delovom diskurse [Methods of term formation in modern English business discourse] [Electronic resource] : Accessed: https://cyberleninka.ru/article/n/methods-of-term-formation-in-modern-english-business-discourse -20.01.2021

Langendoen D. (1998) Glava "Blumfild" V knige: E`nciklopediya Kognitivnoj lingvistiki Massachusetskogo texnologicheskogo instituta [Chapter "Bloomfield" In book: The MIT Encyclopedia of Cognitive Linguistics]. *UniPress*, p.150-183

Macmillan dictionary [Electronic resource] : Accessed:https://www.macmillandictionary.com/dictionary/british/lockdown/ – 17.12.2020

McEney T. (2020) Korpusnaya lingvistika: Metod, Analiz, Interpretaciya. [Corpus Linguistics: Method, Analysis, Interpretation]. *Lancaster University* https://www.futurelearn.com/courses/corpus-linguistics/8/steps/868172 [accessed: 10.11.2020]

Oxford English Dictionary [Electronic resource] : Accessed: https://www.oed.com/ -12.12.2020

Rukovodyashhie principy` terminologicheskoj politiki (2005) Formulirovanie i realizaciya terminologicheskoj politiki v yazy`kovy`x soobshhestvax. Podgotovleno Infoterm YuNESKO [Guidelines for terminology policies. Formulating and implementing terminology policy in language communities. Prepared by Infoterm UNESCO [Electronic resource] : Accessed: https://unesdoc.unesco.org/ark:/48223/pf0000140765 – 12.01.2021

Sager J.C. (1990) Prakticheskij kurs po obrabotke terminologii [Practical Course in Terminology Processing] / *J.C. Sager. – Amsterdam: John Benjamins Publishing*, 264 p.

Shariatmadari D. (2020) God izolyacii [The year of lockdown] [Electronic resource] : Accessed: https://blog.collinsdictionary.com/language-lovers/the-year-of-lockdown/ -15.01.2021

SozdikQor [Electronic resource] : Accessed: https://sozdikqor.kz/ – 20.12.2020

Sozdik.kz [Electronic resource] : Accessed: https://sozdik.kz/ – 20.12.2020

Termincom [Electronic resource] : Accessed: https://termincom.kz/ – 20.12.2020

The Coronavirus Corpus [Electronic resource] : Accessed: https://www.english-corpora.org/corona/ – 26.11.2020

Tilalemi [Electronic resource] : Accessed: https://tilalemi.kz/ – 20.12.2020

Utemgaliyeva N.A., Bektemirova S.B., Rivers W.P. (2019) Akademicheskie usloviya Bolonskogo processa. [Academic terms of the Bologna process]. *Eurasian Journal of Philology: Science and Education.* №2 (174). 2019

Wikipedia [Electronic resource] : Accessed: https://en.wikipedia.org/wiki/Lockdown/ – 17.12.2020

#LancsBox 5.1 manual [Electronic resource] : Accessed: http://corpora.lancs.ac.uk/lancsbox/docs/pdf/LancsBox_5.1_manual.pdf/ – 1.11.2020

Алматы қазақ тілі корпусы (АҚТК) [Electronic resource] : Accessed: http://web-corpora.net/KazakhCorpus/search/?interface_language=kz/ – 26.11.2020

Журавлева Т.А. (1998) Особенности терминологической номинации. -Донецк: АООТ Торговый дом "Донбасс".

Национальный корпус казахского языка http://qazcorpus.kz/indexru/ – 26.11.2020

Национальный корпус русского языка [Electronic resource] : Accessed: https://ruscorpora.ru/new/search-murco/ – 26.11.2020

Слямбеков Қ. (2020) Локдаун деген не? [Electronic resource] : Accessed: https://massaget.kz/mangilik_el/janalyiktar/62594/ – 10.01.2021

Усенова Н. (2020) Ағылшын сөздігі жыл сөзін анықтады [Electronic resource] : Accessed: https://massaget.kz/layfstayl/bilim/gumanitarly-ylymdar/63720/ – 12.01.2021

References

Almaty Corpus of Kazakh language (АҚТК) [Electronic resource] : Accessed: http://web-corpora.net/KazakhCorpus/search/?interface_language=kz/ – 26.11.2020

BBC [Electronic resource] : Accessed: https://www.bbc.com/ – 26.11.2020

Brezina V. (2018) Statistics in Corpus Linguistics: A practical guide. Cambridge University Press, p15

Brezina V., Weill-Tessier P., McEnery T. (2020). #LancsBox 5.x [software]. Available at: http://corpora.lancs.ac.uk/lancsbox

Cambridge dictionary [Electronic resource]: Accessed: https://dictionary.cambridge.org/dictionary/english/lockdown?q=lockdown/ – 17.12.2020

Collins dictionary [Electronic resource] : Accessed: https://collinsdictionary.com/ – 2.12.2020

Corpus of Contemporary American English [Electronic resource] : Accessed: https://www.english-corpora.org/coca/ – 26.11.2020

Dictionary by Merriam-Webster [Electronic resource] : Accessed: https://www.merriam-webster.com/lockdown/ – 17.12.2020

Garside R., Leech G., McEnery T. (2013) Corpus Annotation. Routlegde Guidelines for terminology policies (2005) Formulating and implementing terminology policy in language communities. Prepared by Infoterm UNESCO [Electronic resource] : Accessed: https://unesdoc.unesco.org/ark:/48223/pf0000140765 – 12.01.2021

iWeb: The 14 Billion Word Web Corpus [Electronic resource] : Accessed: https://www.english-corpora.org/iweb/ – 26.11.2020

Kasatkina M.S. (2020) Methods of term formation in modern English business discourse https://cyberleninka.ru/article/n/methods-of-term-formation-in-modern-english-business-discourse Langendoen D. 1998 Chapter "Bloomfield" In book: The MIT Encyclopedia of Cognitive Macmillan dictionary [Electronic resource] : Accessed:https://www.macmillandictionary.com/dictionary/british/lockdown/ – 17.12.2020

McEney T. 2020 Corpus Linguistics: Method, Analysis, Interpretation. Lancaster University [Electronic resource] : Accessed: https://www.futurelearn.com/courses/corpus-linguistics/8/steps/868172 – 10.11.2020

National Corpus of the Kazakh Language http://qazcorpus.kz/indexru/ – 26.11.2020

Oxford English Dictionary [Electronic resource] : Accessed: https://www.oed.com/ -12.12.2020

Russian National Corpus [Electronic resource] : Accessed: https://ruscorpora.ru/new/search-murco/ – 26.11.2020

Sager J.C. Practical Course in Terminology Processing / J.C. Sager. – Amsterdam: John Benjamins Publishing, (1990) – 264 p.

Shariatmadari D. (2020) The year of lockdown [Electronic resource] : Accessed: https://blog.collinsdictionary.com/language-lovers/the-year-of-lockdown/ 15.01.2021

Slyambekov K. (2020) Lockdown degen ne? [What is lockdown] [Electronic resource] : Accessed: https://massaget.kz/mangilik_el/janalyiktar/62594/ – 10.01.2021 (in Kazakh)

SozdikQor [Electronic resource] : Accessed: https://sozdikqor.kz/ – 20.12.2020

Sozdik.kz [Electronic resource] : Accessed: https://sozdik.kz/ – 20.12.2020

Termincom [Electronic resource] : Accessed: https://termincom.kz/ – 20.12.2020

The Coronavirus Corpus [Electronic resource] : Accessed: https://www.english-corpora.org/corona/ – 26.11.2020Tilalemi [Electronic resource] : Accessed: https://tilalemi.kz/ – 20.12.2020

Usenova N. (2020) Ağylşyn sözdigı jyl sözın anyqtady [The English Dictionary has defined the word of the year] [Electronic resource] : Accessed: https://massaget.kz/layfstayl/bilim/gumanitarly-ylymdar/63720/ – 12.01.2021 (in Kazakh)

Utemgaliyeva N.A., Bektemirova S.B., Rivers W.P. (2019) Academic terms of the Bologna process. *Eurasian Journal of Philology: Science and Education.* №2 (174). 2019

Zhuravleva T.A. (1995). Osobennosti terminologicheskoj nominacii. [Peculiarities of terminological nomination] Monography. Moskva, pp. 3-4 (In Russian)

Wikipedia [Electronic resource] : Accessed: https://en.wikipedia.org/wiki/Lockdown/ – 17.12.2020

#LancsBox 5.1 manual [Electronic resource] : Accessed: http://corpora.lancs.ac.uk/lancsbox/docs/pdf/LancsBox_5.1_manual.pdf/ – 1.11.2020