A LEXICOLOGICAL PERSPECTIVE ON THE PANDEMIC: THE CASE STUDY OF ESTONIA

Mai Raet
Tallinn University, EE
mai.raet@tlu.ee

Abstract. The linguistic innovation spurred by the pandemic was remarkable, drawing significant attention in language studies. Although this phenomenon has been studied to some extent in Estonian, the research has primarily been published in Estonian, limiting its accessibility to a broader, non-Estonian-speaking audience. To address this gap, this article identifies the most essential Estonian COVID-19 pandemic-related vocabulary of the public domain 2020–2022. Data was gathered from the Combined Dictionary of Estonian and the Estonian National Corpus 2023, analysis employed mixed methods. A total of 182 lexical items were examined, 91 identified as neologisms, and 22 classified as the 'core lexicon'. 10 semantic categories were established to delineate the main components of Estonian corona discourse. Keyness scores were calculated to monitor changes in the relevance and shifts in thematic focus. This study illustrates how smaller language communities modify their lexicon in response to global crises, aiming to enhance the comprehension of linguistic resilience and language evolution.

Keywords: COVID-19, neologisms, linguistic innovation, dictionaries, corpus linguistics, language change, Estonian language

1. Introduction

On May 5, 2023, three years after COVID-19[1] was declared a global pandemic, the World Health Organization (WHO 2023) announced the end of the global public health emergency. Described as a "major anchor

---

[1] Terminology related to pandemic caused by the SARS-CoV-2 can be confusing (see Raet 2024 for details). In this article, for the sake of brevity, *corona* is used as a catch-all term for the virus, the disease, and the pandemic (e.g., *corona lexicon*). However, this usage is somewhat imprecise when strict terminological distinctions are considered. In general, *COVID-19* refers to the disease, *coronavirus* is used here as a contextual synonym for SARS-CoV-2, the virus causing COVID-19.

point in the twenty-first century" (Frankema & Tworek 2020: 333), the pandemic not only profoundly impacted healthcare and economies but also triggered creativity (Tang, Reiter-Palmon & Ivcevic 2022), including lexical innovation articulated through a surge of neologisms. Thus, from a linguistic perspective, this period provided a unique opportunity to observe the emergence, spread, and assimilation of new lexical items, illustrating the unfolding of linguistic innovation (Trap-Jensen & Lorentzen 2022: 825).

In parallel with the worldwide linguistic adaptations to the pandemic, the Estonian language also diversified its lexis with new terms and expressions. However, there has been a scarcity of scholarly research accessible to international audiences, with only a few studies conducted (e.g., Marling & Käsper 2021; Raet 2023). This study aims to narrow this gap by exploring how the pandemic influenced changes in the Estonian language.

Greater temporal distance enables a clearer understanding of the most notable aspects of past events. This also applies to new lexical items coined to describe various features connected to the pandemic as the acute phase is now over. Therefore, the objective of the article is to identify and analyse the essential vocabulary, including neologisms, used in Estonian public discourse on the pandemic 2020–2022, the primary years of the coronavirus outbreak. Furthermore, to recognise patterns and relationships within the extracted keywords, the study also employs semantic categorisation.

The analysis is based on data from EKI Combined Dictionary of Estonian (CombiDic 2023) and the Estonian National Corpus 2023 (ENC 2023). The latter delves into naturally occurring written texts, offering a view of words as they are used in real-life contexts. In contrast, data from CombiDic 2023 provides insight into institutionalised words, i.e., lexemes incorporated into the CombiDic, reflecting the officially recognised vocabulary. This dual approach results in a more nuanced picture of the linguistic adaptation during the pandemic.

By studying smaller language communities, this research aspires to enhance the understanding of pandemic-related vocabulary as well as linguistic adaptability in general. Additionally, it provides a comparative viewpoint for scholars examining similar phenomena in other languages.

2. COVID-19 pandemic-related lexical innovation: general observations

Threat, urgency, and uncertainty – the three crucial elements of a crisis (Boin, Ekengren, & Rhinard 2010) – create a pressing need for clear communication. This was particularly evident during the recent pandemic, especially at the beginning, as authorities frequently adjusted their daily updates in response to emerging information. This led to numerous new terms, coroneologisms (Roig-Marín 2021), words coined to describe crisis-inflicted realities. As a result, a notable feature of pandemic-driven lexical innovation was the remarkable speed at which novel vocabulary was created and adopted.

Based on the observations of several lexicographers (e.g., Trap-Jensen & Lorentzen 2022; Salazar & Wild 2022; Mihaljević, Hudeček & Lewis 2022), corona lexicon can be broadly divided into three groups: new coinages, specialised terms that gained broader usage, and existing words that acquired new meanings. Firstly, entirely new coinages (neologisms) appeared. Neologisms have been defined in several ways, with different criteria used to identify them in a particular language (Rodríguez Guerra 2016: 529). According to Langemets et al. (2020: 5), neologisms are "words and multiword expressions that have come into use any time during the last two decades, that denote new phenomena in society and that are perceived by users as new". In the pandemic context, a prominent example is the acronym COVID-19, a term described as having "come overwhelmingly to dominate global discourse" (Paton 2020).

The second group of novel vocabulary includes previously existing specialised terms that saw a sharp increase in prominence. These were words often associated with past epidemics, such as 'rate of infection' or 'incubation period', and they became more widely used not only by experts but also by the general public, a phenomenon known as determinologisation. Under normal circumstances, this type of migration of specialised words into common language tends to progress relatively slowly (Meyer & Mackintosh 2000). However, amid the pandemic, this process unfolded exceptionally quickly.

Last but not least, some words acquired additional meanings or nuances. Salazar and Wilde (2022) highlight how the meaning of 'front-line' shifted during the pandemic, becoming more commonly associated

with medical staff and caregivers than previously. Additionally, Collins and Koller (2023: 128) note that 'virus' reverted to its original, biological meaning after being used metaphorically in information technology and content that spreads quickly across social media ('going viral').

Besides the new coinages, the corona lexicon was notable for its global reach. This is unsurprising, as the virus spread worldwide, showcasing the significant influence that viruses can wield when interacting with the dynamics of globalisation (Frankema & Tworek 2020: 333). Media and digital technologies, major forces of globalisation, amplified the use of similar pandemic-related language. The intense focus on the coronavirus and COVID-19 dominated global news narratives, overshadowing other topics (Ng, Chow & Yang 2021). Consequently, the prolonged exposure to the language used in the media shaped public linguistic practices by providing new lexical resources that people creatively incorporated into their communication (Stuart-Smith 2017: 28).

Another notable trait, though not explored in this article, was the extensive use of metaphoric language. Metaphors, as Lakoff and Johnson (1980) posited, play a central role in shaping human cognition, influencing how people perceive and understand the world. During the pandemic, metaphors were extensively used, often framing the virus as an 'invisible enemy' and the global response as a 'war' (cf. Štrkalj Despot & Anić 2021). Even the term 'coronavirus' is based on metaphor (Haddad 2022: 96) as the virus particles appear to resemble a royal crown or solar corona.

Finally, the English language has been central in shaping the global corona vocabulary, including the Estonian vocabulary, as the following investigation will reveal. As mentioned earlier, the pandemic brought specialised medical terms, previously obscure, into common use. In these specialised fields, English is the de facto language (e.g., Gordin 2015), and scientific findings, particularly those concerning SARS-CoV-2 and COVID-19, are primarily published in English, highlighting its significance in creating neologisms (Papp 2022: 149). Consequently, non-English-speaking countries frequently either borrowed directly from English or created calques (literal translations of English terms). However, there have also been instances where minority languages, such as Canadian French, have resisted the dominance of the English by avoiding calques and instead using regional variants (Bowker

2020), i.e., linguistic expressions that are specific to a particular region or dialect of a language.

Although the swift evolution and adoption of new language during the pandemic highlighted resilience, the majority of coroneologisms are context-specific, reflecting the unique challenges of the pandemic. As the situation changes, many of these terms will likely become irrelevant and fade from use (Bueno & Freixa 2022: 81). Additionally, specialized terminology that entered everyday language is expected to return to professional domains (Trap-Jensen & Lorentzen 2022), underscoring the transient nature of crisis-driven language innovations.

Against the backdrop of the abovementioned aspects, this article addresses the following research questions: What keywords, including neologisms, constitute the essential vocabulary of Estonian public discourse on the pandemic, and what are their main lexical characteristics? Additionally, what are the overarching themes that arise from the identified keywords?

In corpus linguistics, a keyword denotes a word that holds statistical significance within a text or a collection of texts. In SketchEngine (Kilgarriff et al. 2014), a tool used later in this study, keywords are individual tokens that occur more frequently in the focus corpus compared to the reference corpus. In more qualitative sense, keywords are the concepts that convey the essence of particular themes, thoughts or discourses (Culpeper & Demmen 2015: 90), hence, reflecting cultural values, creating discursive contexts as well as revealing the scripted lives of people (Levisen & Waters 2017: 5).

3. Material and methods

To explore the posed research questions, the analysis of Estonian corona lexicon is based on two primary sources: EKI Combined Dictionary of Estonian (CombiDic) and the Estonian National Corpus 2023 (ENC 2023). The first one, CombiDic is a dictionary intended for general use, operated via the dictionary writing system Ekilex (Hein et al. 2020). All new lexical items that Estonian lexicographers deem significant enough (cf. Langemets et al. 2020) are introduced in CombiDic for users. As such, these words have an institutionally recognised status. The second source, ENC 2023 is the most extensive

and comprehensive corpus of the Estonian language available, containing around 3.78 billion tokens. It serves as a valuable resource, as in modern linguistics large text corpora are one of the most important sources of material (Muischnek & Lindström 2020: 306), well suited for tracking the emergence and evolution of new lexical items (Collins & Koller 2023: 131), and studying them in naturally occurring contexts (Sinclair 1991: 171).

The sources referred to above, especially ENC 2023, provide the foundation for quantitative analysis. The qualitative part of the analysis focuses on determining whether a word qualifies as a corona-neologism and categorising these words into semantic groups. The classification of coroneologisms can vary depending on the researchers involved. In this article, words sourced from CombiDic are considered neologisms, having been introduced to the general use dictionary between 2020 and 2022. Additionally, special terms that existed previously are also considered new coinages. This aligns with Storjohann and Cimander (2022: 29), who classify terms transitioning from expert domains to general use as neologisms, and Langemets et al. (2020: 5), who describe neologisms as terms users consider new. Similar to deciding whether a word is pandemic-related or not, the process of organising lexical items into categories based on their meanings entails innate subjectivity (cf. Verheyen, Droeshout & Storms 2019), especially in cases where words may straddle multiple categories.

3.1. Extraction principles in CombiDic

The words associated with the COVID-19 pandemic were extracted from Ekilex, the online dictionary system behind the publicly accessible CombiDic on the language portal Sõnaveeb [WordWeb] (Sõnaveeb 2023). To be considered as potential candidates for pandemic-induced neologism, the entries had to a) have a corresponding label assigned to them, showing their registration year (UUS2020, UUS2021, and UUS2022, see Figure 1 for an example), b) be formed using the *koroona* 'corona-' prefix or COVID-19 as a formation component (e.g., *koroonatõend* 'corona certificate', *COVID-19-vaktsiin* 'COVID-19 vaccine'), and/or c) be otherwise connected to the COVID-19 pandemic.

The latter condition posed several challenges, as determining what qualifies as corona-related inherently involves subjectivity. Computer algorithms cannot identify so-called corona words with full autonomy (Trap-Jensen & Lorentzen 2022: 827; Bueno & Freixa 2022) and human judgments also vary, leading to only partial alignment. Hence, to establish more systematic criteria, a word was considered corona-related if it was added to the database during the observed period, indicating its heightened relevance at the time. Additionally, the accompanying example sentence had to reference the pandemic. Example sentences in Sõnaveeb are selected by lexicographers to illustrate entry words in context, carefully chosen to meet established standards for good dictionary examples. Based on the previous authors' works, Kosem et al. (2019) underscore qualities such as authenticity, typicality, informativeness, and intelligibility. Consequently, it is reasonable to assume that the example sentences in Sõnaveeb are coherent and represent a typical context for a given keyword, including its relevance to the pandemic. Therefore, even if a word like *antirekord* 'anti-record' (see Figure 1), existed before the SARS-CoV-2 outbreak, it was included in the list of coroneologisms (cf. Storjohann & Cimander 2022; Langemets et al. 2020). The increased relevance and changing usage of such words during the pandemic, as shown by their inclusion in the CombiDic, highlight shifts in how language described the crisis. Including them provides a more comprehensive understanding of the pandemic's impact on language.
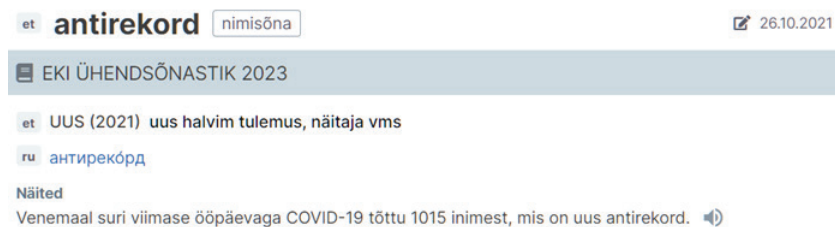

Figure 1. Word entry on Sõnaveeb with the time label UUS (2021) containing reference to the pandemic: "Within the past 24 hours, Russia has reported 1,015 deaths attributed to COVID-19, marking a new anti-record".

3.2. Extraction principles in ENC 2023

To broaden the crisis-related lexicon, the study used ENC 2023 data to identify established lexical items that mark public discourse on the pandemic. Keyword analysis was employed for this purpose because, as Storjohann & Cimander (2022: 25) note, it does more than highlight new vocabulary. It also exposes the discursive focuses, patterns of argumentation, and topicalisations within the various narratives of the discourse.

The identification of typical pandemic-related words was facilitated using the keyword and term extraction function in the corpus query system Sketch Engine (Kilgarriff et al. 2014).[2] This operation requires a reference corpus for comparison to detect elements that are most associated with the focus corpus. Consequently, a sub-corpus of texts from 2019 was compiled and used as the reference to elicit the most prominent keywords from the period 2020–2022, respectively.

One helpful aspect of this function is the keyness score. In Sketch Engine, the keyness score is a statistical measure used to identify prominent words or phrases that are significant to a specific body of text. It is calculated by comparing the frequency of a word or phrase in a focus corpus with its frequency in a reference corpus. A high keyness indicates that a word or phrase appears much more frequently in the focus corpus than would be expected based on its occurrence in the reference corpus. This may indicate a particular topic, theme, or style that is unique or important to the focus corpus.

This study also used keyness score to calculate a metric 'change score' by summing the absolute values of changes in keyness scores. For instance, if the keyness score changed from 2020 to 2022 at a rate of 50-100-80, the change score would be 70. Although this method produced a focused list of prominent terms during the crisis, it required further refinement, adding complexity and subjectivity to identifying corona-specific words. This limitation should be considered when reviewing Supplement B.[3] To assess a word's relevance to the pandemic,

2   For a more detailed description of how the function operates, see the guide on keywords and term extraction at Sketch Engine: https://www.sketchengine.eu/guide/keywords-and-term-extraction/#toggle-id-4-closed.

3   The datasets (Supplements A–E) generated and analyzed during this study are available in Zenodo via DOI: https://doi.org/10.5281/zenodo.11412239

keyword concordance was also examined. The initial keyword extraction produced a dataset requiring further preprocessing to enhance reliability. A notable issue found across all the sub-corpora involved the word-like sequence *koroonaviir*, where the ending 'us' appeared missing, likely to due preprocessing errors. Although *koroonaviir* frequently emerged as an independent keyword, it was excluded from further analysis. This decision was based on how keyness scores are calculated: simply adding the scores of two variants does not accurately reflect their true significance or account for any potential overlap in their usage across different contexts.

It is important to note that the numbers and percentages in the following exposition, concerning both CombiDic and ENC 2023, should not be seen as exact due to potential biases in detecting corona-relatedness, as previously discussed. Nonetheless, although these numbers are not precise, they offer a broad overview of prevailing trends.

4. Results

4.1. Coroneologisms in CombiDic

During 2020–2022, a total of 706 new lexical entries were added to CombiDic. Of these, 91 entries, constituting about 13% of the total, were identified as explicitly pandemic-related. A full list of these words can be found in Supplement A.
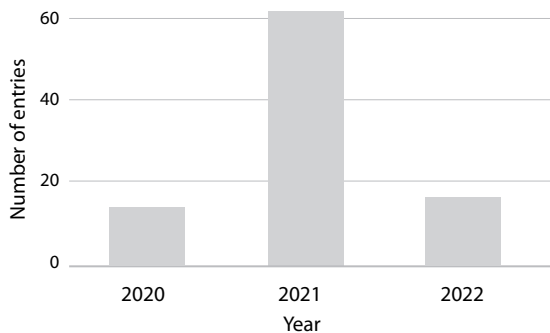

Figure 2. Yearly count of corona-related word entries 2020–2022.

Analysis of corona-related entries in the CombiDic 2020–2022 reveals distinct trends (see Figure 2). In 2020, there were 14 newly introduced entries explicitly linked to the pandemic. The following year, 2021, witnessed a notable surge in coroneologisms, with 61 fresh entries, constituting about 23% of all new terms for that year. However, 2022 experienced a decline, with only 16 new pandemic-specific entries, comprising 5% of the total new terms for the year. While the total number of neologisms in CombiDic increased, the share of pandemic-related entries peaked in 2021 and sharply declined in 2022, indicating reduced significance.

Morpho-lexically, about 80% of the CombiDic list consists of compound words. Among the 91 words, 27 (30%) start with the 'corona-' component.

4.2. Coroneologisms in the ENC 2023

Using the method presented in Section 3.2, the 2019 sub-corpus served as the baseline for extracting data from 2020, 2021, and 2022 individually. To enable comparison with CombiDic, the same number of words were selected, i.e., the first 91 words based on the keyness score (see Supplement B). Table 1 provides the 25 most outstanding words with a keyness score of 17 or higher.

The data from ENC 2023 helps to identify potential candidate words demonstrating lexical innovation, specifically words absent from the 2019 reference corpus. Of the 91 corpus words, 32 had no mentions before 2020 when compared to the 2019 corpus (highlighted in grey in Table 1 and Supplement B). Over half of these 32 words (18) begin with the 'corona-' component.

Morpho-lexically the ENC 2023 list consists of around 65% compounds. Of these 92 words, 18 (20%) start with the 'corona-' component. Thematically, beyond healthcare and science, the ENC 2023 list also includes more general terms. These are words that were used before the pandemic but gained prominence due to it, such as *pandeemia* 'pandemic', *vaktsiin* 'vaccine', and *viirus* 'virus' among others. The list also features names of specific vaccines like AstraZeneca and Pfizer.

Table 1. 25 most prominent corona-related words in ENC 2023 during 2020–2022. Words highlighted in grey had no mentions before 2020. The numbers in the 'Score' column represent the keyness score.

№	Word est	Word eng	Score
1	koroonaviirus	*corona virus*	497.4
2	covid-19	*covid-19*	249.4
3	pandeemia	*pandemic*	106.9
4	koroonakriis	*corona crisis*	83.4
5	koroona	*corona*	78.1
6	koroonapandeemia	*corona pandemic*	50.9
7	nakatunu	*infected*	45.2
8	distantsõpe	*distance learning*	42.9
9	koroonapiirang	*corona restriction*	37.8
10	covid	*covid*	33.1
11	koroonavaktsiin	*corona vaccine*	33.0
12	koroonatest	*corona test*	29.7
13	astrazeneca	*AstraZeneca*	28.6
14	teadusnõukoda	*scientific council*	26.5
15	karantiin	*quarantine*	25.0
16	vaktsiinidoos	*vaccine dose*	24.7
17	tõhustusdoos	*booster dose*	23.4
18	pfizer	*pfizer*	23.3
19	sars-cov-2	*sars-cov-2*	23.2
20	lähikontaktne	*close contact*	21.4
21	koroonaproov	*corona sample*	21.3
22	eneseisolatsioon	*self-isolation*	21.2
23	nakatumine	*infection*	18.5
24	koroonapositiivne	*corona positive*	17.1
25	koroonaaeg	*corona time*	17.0

A quick glance at Table 1 immediately confirms that *koroona-viirus* 'coronavirus' is by far the most prominent lexeme in the aggregated results 2020–2022, with the keyness score around 500. Several synonyms referring to the same concept, such as *pandeemia* 'pandemic' and *kriis* 'crisis' are also present. Additionally, terms like *nakatunu* 'infected', *distantsõpe* 'distance learning' and *koroonapiirang* 'corona restriction' further elaborate on the primary theme, illuminating important details and actions relevant to the crisis in Estonia.

Comparison of CombiDic and ENC 2023 reveals 22 shared terms (see Table 2). This is another possible approach to pinpoint the central elements of the Estonian pandemic vocabulary.

Table 2. Words shared by ENC 2023 and CombiDic, ranked from highest to lowest based on scores from ENC 2023.

№	Word est	Word eng
1	covid-19	*covid-19*
2	koroonakriis	*corona crisis*
3	koroona	*corona*
4	koroonapiirang	*corona restriction*
5	koroonavaktsiin	*corona vaccine*
6	koroonatest	*corona test*
7	tõhustusdoos	*booster dose*
8	koroonaproov	*corona sample*
9	koroonapositiivne	*corona positive*
10	koroonaaeg	*corona time*
11	koroonapass	*corona certificate*
12	koroonapatsient	*corona patient*
13	omikron	*omicron*
14	koroonasurm	*corona death*
15	omikrontüvi	*omicron strain*
16	pcr-test	*pcr-test*
17	koroonajuhtum	*corona case*
18	koroonanakkus	*corona infection*
19	vaktsiinipass	*vaccine certificate*
20	koroonatõend	*corona certificate*
21	koroonapuhang	*corona outbreak*
22	maskikohustus	*mask mandate*

Upon reviewing the list of 22 words, no single topic emerges as dominant, but the presence of synonyms for certain subjects indicates their significance. The following are close or near-synonyms: a) *covid-19* and *koroona* 'corona'; b) *omikron* and *omikrontüvi* 'omicron strain'; c) *koroonatest* 'corona test' and *koroonaproov* 'corona sample'; d) *koroonapass* and *koroonatõend* 'corona certificate' and *vaktsiinipass* 'vaccination certificate'. Each of these topics represents a distinct aspect of responses to the pandemic: a) the virus and the disease at its core causing the pandemic; b) a particular variant of the SARS-CoV-2 virus; c) diagnostics and testing, and d) government vaccination policy documents. Particularly noteworthy is the last topic, including three synonymous terms, thereby signalling heightened prominence.

4.3. Evolution of the pandemic lexicon in ENC 2023: unique words 2020–2022

Table 3 presents 10 most significant corona-related unique words for each year with keyness scores 2.5 or higher (see Supplement D). The majority of unique coronelogisms, about 85%, emerged in 2020. Based on the keyness scores, the most prominent words of 2020 are mainly associated with protective measures, such as *isikukaitsevahend* 'personal protective equipment', *kaitsemask* 'protective mask', and *desinfitseerimisvahend* 'disinfectant'.

Since most of the vocabulary emerged in 2020, it is intriguing to observe what new developments the following two years brought forth. Firstly, there was a drastic decrease in the number of unique words, dropping from 301 in 2020 to only 21 in 2021. The identified words distinctly indicate a transition towards vaccination-related subjects, including unique words such as *vaktsineerimiskeskus* 'vaccination centre', *Janssen* (a vaccine brand), and *kaitsesüst* 'protective shot'. Thus, 2021 truly marked the year of vaccination. For instance, the word 'vaccine' was chosen as Merriam-Webster's Word of the Year 2021, while a more informal term, 'vax', was selected by Oxford Languages for the same honour. According to Merriam-Webster (2024), "few words can express so much about one moment in time", as it was "at the centre of debates about personal choice, political affiliation, professional regulations, school safety, healthcare inequity, and so much more."

By 2022, attention had shifted primarily to the war in Ukraine and the emerging energy crisis, although 24 unique coroneologisms were still identified, albeit with lower keyness scores. The lexical items in 2022 covered various pandemic aspects, with vaccination and its policies still being relevant, e.g., *vaktsineerimiskohustus* 'vaccination obligation', *sundvaktsineerimine* 'mandatory vaccination', *vaktsiiniannus* 'vaccine dose', and *vaktsiinikindlustus* 'vaccine insurance'. All in all, over the three-year period, there is a consistent decrease in the keyness scores, demonstrating a general decline in novelty and reflecting an obvious shift in focus.

Figure 3 presents the 8 keywords that experienced the most drastic changes in keyness scores from 2020 to 2022 in ENC 2023. The indicated 'change score' was calculated by summing the absolute values of the changes in keyness scores for these terms. The trend across all words shows a substantial decrease, suggesting, once more, a reduced prominence in public narratives as the pandemic progressed. The term *koroonaviirus* 'corona virus' underwent the most extreme change, with a score of 939, in contrast to the other terms that declined at a more stable pace.

The highlighted keywords in Figure 3 reveal a disparity in relevance over time, with some gaining high scores in 2020, such as *koroonaviirus* 'corona virus', *covid-19*, *koroonakriis* 'corona crisis', and *nakatunu* 'infected'. These terms represent the initial phase of the pandemic and the immediate response to it. The keywords peaking in 2021 are mainly associated with the vaccine rollout, including specific vaccine brands, and other corresponding terms such as *vaktsiinidoos* 'vaccine dose' and *koroonavaktsiin* 'corona vaccine'.

Table 3. Top 10 most significant unique pandemic-related words 2020–2022 in ENC 2023.

| №	Est	Eng	Score
2020			
1	isikukaitsevahend	*personal protective equipment*	31.7
2	kaitsemask	*protective mask*	27.9
3	desinfitseerimisvahend	*disinfectant*	18.1
4	kriisiaeg	*time of crisis*	17.6
5	maksepuhkus	*payment holiday*	16.6
6	viirusepuhang regionaalosakond	*virus outbreak* *regional department*	15.3
7	nakkusjuhtum	*infection case*	14.8
8	näomask	*face mask*	14.2
9	palgatoetus	*wage subsidy*	13.7
10	koroonanakkus	*corona infection*	12.2
2021			
1	vaktsineerimiskeskus	*vaccination centre*	17.5
2	janssen	*janssen*	15.3
3	kaitsesüst	*vaccine*	14.8
4	nakkusohutus	*infection control*	13.2
5	vaktsineerimisbuss	*vaccination bus*	10.8
6	kaitsepookimine	*vaccination*	10.2
7	digitõend	*digital certificate*	9.7
8	vaktsineerimispass	*vaccination passport*	8.6
9	süst	*injection*	7.4
10	delta	*delta*	5.3
2022			
1	tervisekassa	*health insurance fund*	9.0
2	sümptomaatiline	*symptomatic*	5.9
3	pandeemiaeelne	*pre-pandemic*	5.8
4	covid-tõend	*covid-certificate*	5.8
5	vaktsineerimiskohustus	*vaccination requirement*	5.7
6	koroonapoliitika	*corona policy*	4.8
7	karantiinihotell	*quarantine hotel*	4.32
8	taastekava	*recovery plan*	4.31
9	sundvaktsineerimine	*compulsory vaccination*	4.29
10	vaktsiiniannus	*vaccine dose*	3.82
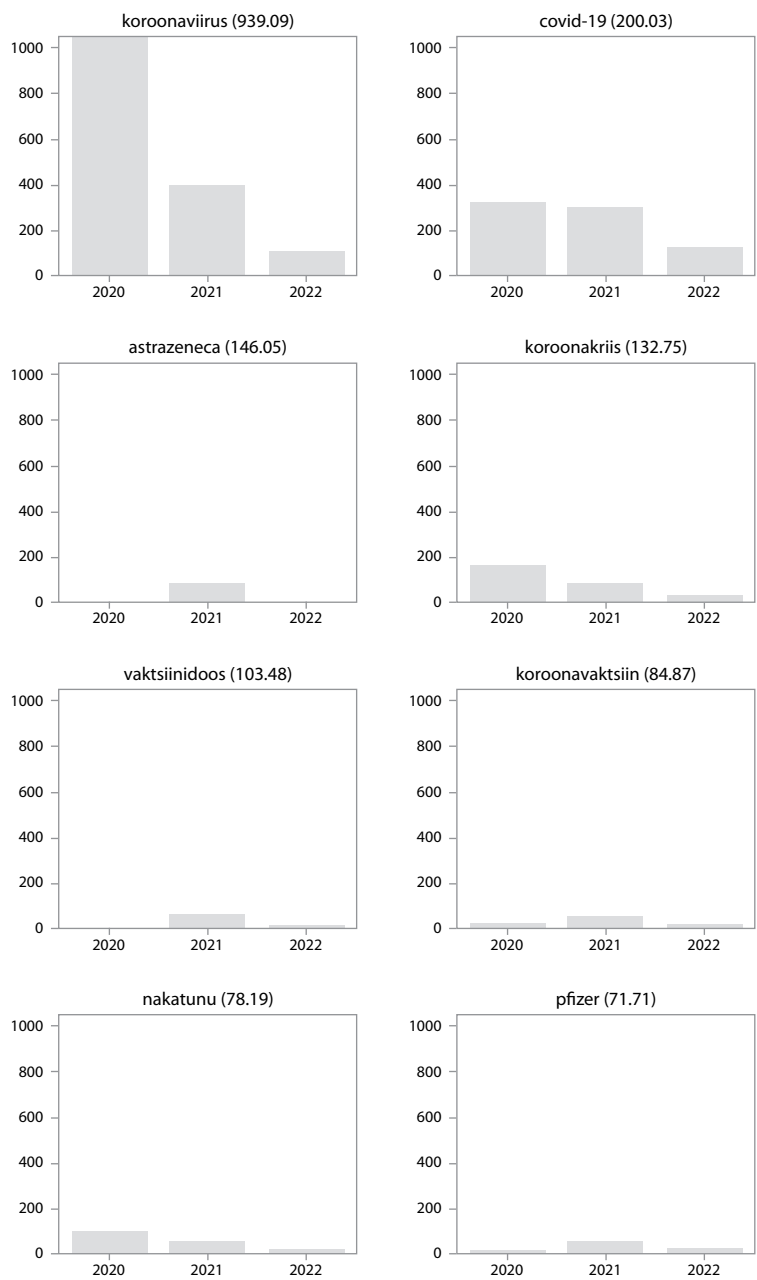
Figure 3. Top 8 keywords with the greatest changes in keyness scores 2020–2022.

5. Lexical characteristics of the Estonian corona-related neologisms

In this article, corona neologisms are represented by the 91 words retrieved from the Ekilex dictionary system (CombiDic). From the lexico-morphological perspective, the list mainly includes monolexemic entries, such as *ogavalk* 'spike protein' and *reisimull* 'travel bubble,' with the notable exception of *pikk COVID* 'long COVID,' a multi-word term. Despite a predominance of monolexemic words, most Estonian corona vocabulary comprises compounds, aligning with the structure of the Estonian language, which consists mostly of compounds and derivatives, accounting for about 90% of the entire lexicon (Kasik 2015: 11), with compounds being more frequent than derivatives. Within these, the prefix *koroona-* 'corona-' appears most frequent occurring in 34 entries and representing about 40% of all pandemic-related entries.

Most newly introduced terms are nouns, totalling 82 words, while a smaller fraction, about 10%, includes five verbs (*laussulgema* 'to lockdown', *sekventsima* and *järjendama* 'to sequence', *kiirtestima* 'to take a rapid test', *sundvaktsineerima* 'to enforce mandatory vaccination', and *järjendama* 'to sequence') and three adjectives (*koroonane* 'having corona', *koroonapositiivne* 'corona positive', and *maskivastane* 'anti-mask'). Although these adjectives can function as nouns in Estonian, they are categorised as adjectives for this study.

Five new keywords have been identified as colloquial in the dictionary: *desoma* 'to disinfect', *antivakser* and *antivaksik* (synonyms for 'anti-vaxxer'), *vaks* 'vax', and *koronts* (slang for corona disease). These terms suggest potentially significant aspects of the pandemic as frequently used complex words often undergo abbreviation (cf. Leech et al. 2009), as seen with *desoma*, a more concise alternative to *desinfitseerima*. Words linked to emotive topics, such as controversial vaccination, also tend to incline towards colloquialisation (Collins & Yao 2018).

When studying Estonian coronaeologisms, the impact of the English language cannot be underestimated, with many, if not most, words directly borrowed from English (e.g., the acronym COVID-19, PCR-test), or translation loans (e.g., *ogavalk* 'spike protein', *enesetestimine* 'self-testing'). As noted in Section 2, the influence of English on corona vocabulary in non-Anglophone countries has been documented in many other languages as well, including, for instance, Bosnian (Delić

& Dedović-Atilla 2022), Arabic (Muassomah 2023), and Japanese (Sarif & Suganda 2020). Despite, or indeed because of the prevalence of English, it can sometimes be challenging to determine whether a specific term originated independently in Estonian or is a loan translation, a notion not exclusive to Estonian (e.g., Mihaljević, Hudeček & Lewis 2022 for Croatian, or Adelstein & de los Ángeles Boschiroli 2022 for Spanish). In addition, it is pertinent to acknowledge Estonia's evolving alignment with the English language. This has been a discernible trend over recent decades (cf. Kruusvall 2015), leading to English becoming the most widely spoken foreign language (Statistics Estonia 2022).

6. Further insights into Estonian corona discourse

Corona pandemic has been showcased as a super discourse (cf. Jakosz & Kałasznik 2022, 2023), meaning it is "composed of various thematically definable strands of discourse" (Jakosz & Kałasznik 2023: 7). One way to identify recurring themes and motifs underlying such a discourse, is by studying the lexicon, i.e., "gain entry to discourse through words" (Née & Veniard 2012). Further, this can be complemented by a corpus-based approach to discourse analysis. Still, one criticism of this approach is that its focus on frequency might merely validate existing knowledge, though, as suggested by Baker (2023: 21), it could also stem from a cognitive bias known as hindsight bias.

A deeper look into Estonian corona discourse can be achieved by analysing the 138 unique terms. Comparing these two lists, ENC 2023 contains slightly more vaccination-related words than the CombiDic list. On the other hand, CombiDic has more words connected to socio-cultural implications such as *koroonapagulane* 'corona refugee', *koroonaeitaja* 'corona denier', *infodeemia* 'infodemic', *ebaravi* 'pseudo-treatment', *reisimull* 'travel bubble'.

Subsequently, the 138 unique words can be grouped into distinct categories. Table 5 highlights this classification, outlining the topics connected to the pandemic in Estonia.

Table 5. Pandemic-related words classified by semantic categories.

Category	ENC 2023	Eki-Lex	Σ	Examples
1. Healthcare interventions and medical care	13	14	27	*desovahend* (sanitizing agent), *koroonaravim* (corona medicine)
2. Vaccination	17	9	26	*vaktsiinitarne* (vaccine supply), *antivaksik* (anti-vaxxer)
3. Transmission	13	2	15	*koroonanakatumine* (corona infection), *-kolle* (corona epicenter)
4. Governmental policies	7	8	15	*laussulgema* (lockdown), *taastekava* (recovery plan)
5. Virus	6	8	14	*pikk-covid* (long covid), *mutantviirus* (mutant virus)
6. Socio-cultural implications	1	9	10	*koroonaeitaja* (corona denier), *infodeemia* (infodemic)
7. Work/school arrangements	4	6	10	*hübriidtöö* (hybrid work), *kodusõpe* (home schooling)
8. Pandemic discourse	6	3	9	*inimkatastroof* (human catastrophe), *viirusekriis* (virus crisis)
9. Testing	2	6	8	*proovivõtupunkt* (sampling point), *kiirtestimine* (rapid testing)
10. Personal states	0	4	4	*terviseärevus* (health anxiety), *tehnostress* (technostress)

The categorisation process is predominantly qualitative, showing variation across different scholarly works in scope and detail of topic delineation. For example, Nam and colleagues (2022) identified 12 distinct semantic categories including religion, food, and clothing, among others, while Storjohann and Cimander (2022) conducted a more detailed examination, elaborating on 18 categories. This article introduces the category "Personal states"; financial measures, such as *palgatoetus* 'wage subsidy', were grouped under the governmental policies. Vaccination is categorised independently, although it could fall under healthcare interventions and medical care. The complete categorisation is available in Supplement E.

From the 10 established categories, the category "Healthcare interventions and medical care" comprises about 20% of the entire dataset,

with the highest combined frequency of 27 mentions across both sources. It is closely followed by "Vaccination", with 26 mentions. Words linked to virus transmission and governmental policies each had a total frequency of 15 mentions, indicating roughly equal importance. For comparison, the categorisation of Korean neologisms (Nam, An & Jung 2022: 49) resulted in the "Politics and Administration" category being the most prominent, accounting for about 33% of all observations.

When examining the ENC 2023 and CombiDic datasets separately, there is a noticeable difference in the number of words in some categories. ENC 2023 places more emphasis on vaccination (17 words) and virus transmission (13 words). In contrast, the CombiDic dataset yields a deeper look into the socio-cultural implications of the pandemic (9 words) and reveals slightly more about testing-related topics (6 words).

In addition to identifying emerging topics from the 138 unique words, they offer broader insights into Estonia's specific circumstances during the pandemic. Words grouped into categories "Socio-cultural implications" and "Personal states" are particularly informative. It becomes evident that the rapid transition to remote work and online learning, prompted by the spread of coronavirus and the resulting isolation, highlighted the concept of *tehnostress* 'technostress' as many individuals, unaccustomed to this digital shift, grappled with the new technology demands. For students, the sudden immersion in virtual environments strained their learning capacity, making it challenging to adapt to digital tools effectively, indicated by the word *õpilünk* 'learning gap'. At the same time, the pandemic created a fertile ground for the spread of misinformation, e.g *infodeemia* 'infodemic' and non-evidence-based treatments (*ebaravi*), contributing to what could only be described as a human disaster (*inimkatastroof*) in terms of public health. Consequently, a significant portion of the population experienced heightened health anxiety (*terviseärevus*).

7. Summary and discussion

Lexicography has consistently shown that substantial social changes drive major linguistic shifts (Paton 2020), a notion that has been especially true during the global coronavirus outbreak. The current study

investigated how these pandemic-related shifts occurred in the Estonian language. Using data from two sources, CombiDic and ENC 2023, the study not only determined the extent of pandemic-driven lexical innovation but also the temporal and thematic trends conveyed by the identified keywords.

In total, 182 corona-related words were detected. Regarding lexical innovation, the 91 words identified in CombiDic were classified as corona-related neologisms. This distinction was based on a key characteristic of neologisms: their newness, as these terms were introduced between 2020 and 2022 (cf. Bueno & Freixa 2022: 74). Additionally, their inclusion in the general-use dictionary signalled formal recognition by language authorities.

From a lexico-morphological standpoint, the neologisms were primarily single-word compound nouns (about 90 %). This is unsurprising, as compounding is a common word-formation strategy in Estonian. In fact, during the pandemic, it was particularly recognised as a productive method of forming neologisms in other languages as well (e.g., Salazar & Wild 2022). The use of *koroona-* 'corona-' as a modifier was especially prevalent. While the term *koroona* existed in the Estonian lexicon before the virus outbreak, primarily referring to the board game corona, words starting with that lexical element can still be considered genuine neologisms, i.e., words that emerged solely because of the pandemic.

The global nature of the pandemic and the dominance of English are evident in the Estonian corona neologisms. Many terms are directly borrowed or adapted through translation loans. On one hand, Estonia has increasingly been becoming English oriented (cf. Statistics Estonia 2022), but this trend was consistently observed in many other non-Anglophone languages. This underscores the dominance of English, particularly in scientific and public health communication.

To pinpoint the central keywords in Estonian corona discourse, data from CombiDic and ENC 2023 were combined and analysed from multiple perspectives. Two datasets shared 22 lexical items in common (listed in Table 2). Despite the lack of a single overarching theme, the presence of synonyms for certain subjects highlights the possibility of their particular importance. These key themes were government vaccination policy documents, virus/disease (especially Omicron variant), and diagnostics/testing. Curiously, despite numerous measures to curb the virus spread (Terviseamet 2023: 69–80), only two specific ones made

it into the top list: *maskikohustus* 'mask mandate' and the synonymous *koroonapass* and *koroonatõend*, both denoting 'COVID certificate'. This may be because these measures were more visible to the public, making them more memorable in discussions. This visibility could be due to their direct effect on daily life, e.g., needing a mask to enter public spaces or a COVID-certificate to access certain services, which in turn led to more media coverage. The second notable difference is that the only virus variant in shared list was Omicron, despite its relatively late appearance in November 2021. The variant's unique characteristics, such as its increased transmissibility, likely heightened its perceived threat. Interestingly, among the 22 shared words there were no explicit references to work/school arrangements, a domain heavily affected by the pandemic (cf. Vitória, Ribeiro & Carvalho 2022). However, the data from ENC 2023 did acknowledge the relevance of the topic, with the *distantsõpe* 'distance learning' ranking eighth based on the calculated keyness score.

The 22 words shared between the general dictionary (CombiDic) and the media-heavy corpus (ENC 2023) suggest these terms transcended specific contexts and were widely adopted across society. This may be due to the shared terms being broad and non-specialsed, making them universally relevant to global pandemic discussions.

To broaden the Estonian corona discourse beyond the 22 central coroneologisms, the 138 unique lexical items (see Supplement C) from both datasets were also analysed. This involved assigning them into ten specific semantic categories (see Supplement E). The inspection revealed a primary focus on healthcare interventions and medical care, including vaccination. In comparison, ENC 2023 placed a greater emphasis on vaccination and virus transmission, while CombiDic provided more insights into the socio-cultural impacts and testing-related topics. The differing emphases between the two sources stem from their distinct natures. CombiDic is designed for general dictionary use, while ENC 2023 mainly consists of media texts, making up about 57% of the corpus. Controversial vaccination policies and anxiety-provoking virus transmission may have received more focus in the media, which adheres to the principles of the "attention economy" (Goldhaber 1997), where news organisations tend to concentrate on sensationalist or negative news (Erikson & Tedin, 2019). This is supported by virologist and head of the COVID-19 research council, Irja Lutsar, who noted in the Health

Board's report (Terviseamet 2023: 11) that in spring 2020, widespread fear among the Estonian public was often amplified by the media, rather than mitigated.

The temporal analysis revealed how the significance and frequency of pandemic-related language evolved over the three-year period. Both datasets captured a sudden increase of corona vocabulary at the beginning of the pandemic. ENC 2023 recorded the most unique corona words in 2020, while CombiDic saw the peak in 2021. This delay is likely due to lexicographers taking more time to select which words to add to the dictionary (cf. Langemets et al. 2020). Both datasets display a sharp decline of new corona entries in 2022. This pattern aligns with the observation that rapid societal changes often spark bursts of neologism creation, which taper off as the novelty fades (Fagan 1987). The decline also reflects a shift in public focus as the pandemic's acute phase receded and other global crises, like the war in Ukraine, gained prominence. The reduction in new terms also indicates a saturation in the pandemic lexicon, with fewer new words needed to describe ongoing events.

The temporal analysis also exposed which keywords underwent the most drastic changes over the three-year period. Based on the changes in keyness scores, three keywords that experienced the most drastic change were *koroonaviirus* 'corona virus', *covid-19* and *AstraZeneca*. Overall, if one Estonian word could be crowned as the standout of the pandemic, it would be *koroonaviirus* 'coronavirus'. Though not a neologism, it was the most salient keyword, undergoing the most significant shifts in importance during the period.

As a limitation, this study focused on essential Estonian corona vocabulary, excluding many unusual but interesting lexical items. Though there are articles on the humorous aspects of Estonian corona language (Hiiemäe et al. 2021; Voolaid 2022), researching rarer, more figurative pandemic-related terms remains a promising area for further study. Additionally, when interpreting results of the current study, the subjectivity of categorisation and the reliance on formal sources like CombiDic and ENC 2023 should be considered. They may overlook informal language use, such as on social media and in everyday conversation, potentially skewing the understanding of real-life language evolution. As noted by Collins and Koller (2023:133), news texts, though increasingly informal (Talbot 2007), are still institutionalised and

standardised, and therefore less likely to showcase the same level of linguistic innovation and non-standard language as other sources.

By contributing to the ongoing dialogue on lexicology and discourse in the context of the coronavirus pandemic, this work aims to enrich the literature on the impact of global events on lesser-known languages. The findings underscore the adaptability of language and the critical role of linguistic innovation in responding to extraordinary real-life challenges. Future research may build upon these insights to further explore the intersection of global communication and local linguistic adaptations in times of crisis.

Acknowledgments

Data availability statement: The datasets (Supplements A–E) generated and analysed in the study are available in Zenodo, accessible via DOI: https://doi.org/10.5281/zenodo.11412239.

References

Adelstein, Andreína & Victoria de los Ángeles Boschiroli. 2022. Spanish neologisms during the COVID-19 pandemic: Changing criteria for their inclusion and representation in dictionaries. In Annette Klosa-Kückelhaus & Ilan Kernerman (eds.), *Lexicography of Coronavirus-related Neologisms*, 93–124. Berlin; Boston: De Gruyter. https://doi.org/10.1515/9783110798081-006.

Baker, Paul. 2023. *Using corpora in discourse analysis*. London: Bloomsbury Academic. https://doi.org/10.5040/9781350083783.

Boin, Arjen, Magnus Ekengren & Mark Rhinard. 2010. The study of crisis management. In Cavelty Myriam Dunn & Victor Mauer (eds.), *The Routledge Handbook of Security Studies*, 452– 462. London: Routledge.

Bowker, Lynne. 2020. French-language COVID-19 terminology: International or localized? In Minako O'Hagan & Julie McDonough Dolmaya (eds.), Special issue of *The Journal of Internationalization and Localization* 7(1–2). 1–27. https://doi.org/10.1075/jial.20014.bow.

Bueno, Pedro J. & Judit Freixa. 2022. Lexicographic detection and representation of Spanish neologisms in the COVID-19 pandemic. In Annette Klosa-Kückelhaus & Ilan Kernerman (eds.), *Lexicography of coronavirus-related neologisms* (Lexicographica: series maior 163), 73–92. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110798081-005.

Collins, Luke C. & Veronika Koller. 2023. *Viral language: Analysing the Covid-19 pandemic in public discourse.* London: Routledge. https://doi.org/10.4324/9781003163459.

Collins, Peter & Xinyue Yao. 2018. Colloquialisation and the evolution of Australian English: a cross-varietal and cross-generic study of Australian, British, and American English from 1931 to 2006. *English World-Wide* 39(3). 253–277. https://doi.org/10.1075/eww.00014.col.

CombiDic = *The EKI Combined Dictionary*. 2023. Langemets, Margit, Indrek Hein, Madis Jürviste, Jelena Kallas, Olga Kiisla, Kristina Koppel, Tiina Leemets, Sirje Mäearu, Tiina Paet, Peeter Päll, Peeter, Maire Raadik, Lydia Risberg, Hanna Tammik, Arvi Tavast, Mai Tiits, Katrin Tsepelina, Maria Tuulik, Tiia Valdre, Ülle Viks, Edgar Sai & Valentina Tubin (eds.). Institute of the Estonian Language. https://doi.org/10.15155/3-00-0000-0000-0000-08C0AL.

Culpeper, Jonathan & Jane Demmen. 2015. Keywords. In Douglas & Randi Reppen (eds.), *The Cambridge Handbook of English Corpus Linguistics*, 90–105. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.006.

Delic, Haris & Elma Dedović-Atilla. 2022. The analysis of the Covid-19 related anglicisms in the Bosnian language – the study of pandemija, infekcija, lokdaun, karantin, klaster, socijalna distanca, and vakcina. *MAP Education and Humanities* 2. 32– 47. https://doi.org/10.53880/2744-2373.2022.2.1.32.

ENC 2023 = *Estonian National Corpus*. 2023. Kristina Koppel, Jelena Kallas, Madis Jürviste & Helen Kaljumäe (eds.). Lexical Computing Ltd. / The Institute of Estonian Language. https://doi.org/10.15155/3-00-0000-0000-0000-08C04M.

Erikson, Robert S. & Kent L. Tedin. 2019. *American Public Opinion: Its Origins, Content, and Impact (10th ed.)*. New York; London: Routledge. https://doi.org/10.4324/9781351034746.

Fagan, David. 1987. On profiles in lexical diffusion. *La Linguistique* 23. 47–69. https://doi.org/10.1016/0024-3841(89)90069-7.

Frankema, Ewout & Heidi Tworek. 2020. Pandemics that changed the world: historical reflections on COVID-19. *Journal of Global History* 15. 333–335. https://doi.org/10.1017/S1740022820000339.

Goldhaber, Michael H. 1997. The attention economy and the net. *(2nd) Draft version of a talk to be presented at the conference on "Economics of Digital Information", Cambridge, MA, Jan. 23–26, 1997.* https://people.well.com/user/mgoldh/AtEcandNet.html (Accessed 2024-05-29).

Gordin, Michael D. 2015. *Scientific babel: How science was done before and after global English*. Chicago: University of Chicago Press.

Haddad, Amal Haddad. 2022. COVID-19 neologisms between metaphor and culture. In Pascal Hohaus (ed.), *Science communication in times of crisis* (Discourse

Approaches to Politics, Society and Culture 96), 91–118. https://doi.org/10.1075/dapsac.96.05had.

Hein, Indrek, Kaur Männiko, Jelena Kallas, Kristina Koppel, Margit Langemets, Tõnis Nurk, Tõnis; Merily Plado, Mari Vaus, Ülle Viks, Arvi Tavast, Martin Laubre,Yogesh Sharma & Hardi Niilo. 2020. Ekilex. *Eesti Keele Instituudi sõnastiku- ja terminibaas*. Eesti Keele Instituut. https://ekilex.ee/.

Hiiemäe, Reet, Mare Kalda, Mare Kõiva & Piret Voolaid. 2021. Koroonakriisi rahvapärased väljendused Eestis: folkloori taaskasutus kui toimetulekuviis. *Keel ja Kirjandus* 12. 1011−1032. https://doi.org/10.54013/kk757a1.

Kosem, Iztok, Kristina Koppel, Tanara Zingano Kuhn, Jan Michelfeit &Carole Tiberius. Identification and automatic extraction of good dictionary examples: the case(s) of GDEX. 2019. *International Journal of Lexicography* 32. 119–137. https://doi.org/10.1093/ijl/ecy014.

Jakosz, Mariusz & Marcelina Kałasznik. 2022. Der Corona-Diskurs und seine Charakteristika. Perspektiven und Forschungsschwerpunkte. In Mariusz Jakosz & Marcelina Kałasznik (eds.), *Corona-Pandemie: Diverse Zugänge zu einem aktuellen Superdiskurs*, 9–22 Göttingen: V&R Unipress. https://doi.org/10.14220/9783737015127.9.

Jakosz, Mariusz & Marcelina Kałasznik. 2023. Corona-Pandemie im Text und Diskurs. In Mariusz Jakosz & Marcelina Kałasznik (eds.), *Fragestellungen, Zugänge und Perspektiven. Corona-Pandemie im Text und Diskurs*, 7–26. Göttingen: V&R Unipress. https://doi.org/10.14220/9783737016285.7.

Kasik, Reet. 2015. *Sõnamoodustus*. Tartu: Tartu Ülikooli Kirjastus.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1. 7–36. https://doi.org/10.1007/s40607-014-0009-9.

Kruusvall, Jüri. 2015. Keelteoskus ja keelte praktiline kasutamine. *Eesti ühiskonna lõimumismonitooring 2015*, 72–86. Tallinn: Ministry of Culture. https://www.praxis.ee/wp-content/uploads/2015/09/6peatykk.pdf (Accessed 2024-05-29)

Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago.

Langemets, Margit, Jelena Kallas, Kaisa Norak & Indrek Hein. 2020. New Estonian words and senses: Detection and description. *Dictionaries: Journal of the Dictionary Society of North America* 41(1). 69–82. https://doi.org/10.1353/dic.2020.0005.

Leech, Geofrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511642210.

Levisen, Carsten & Sophia Waters. 2017. How words do things with people. In Carsten Levisen & Sophia Waters (eds.), *Cultural Keywords in Discourse* (Pragmatics & Beyond New Series 277), 1–23. John Benjamins. https://doi.org/10.1075/pbns.277.01lev.

Marling, Raili & Marge Käsper. 2021. Communicating Covid-19: Framing science and affect in US, French and Estonian traditional media. *ESSACHESS* 14(2). 15–32. https://doi.org/10.21409/z3xp-m289.

Merriam-Webster. 2024. *Merriam-Webster's Word of the Year 2021*. https://www.merriam-webster.com/wordplay/word-of-the-year-2021-vaccine (Accessed 2024-05-29).

Meyer, Ingrid & Kristen Mackintosh. 2000. When terms move into our everyday lives: an overview of de-terminologization. *Terminology* 6(1). 111–138. https://doi.org/10.1075/term.6.1.07mey.

Mihaljević, Milica, Lana Hudeček & Kristian Lewis. 2022. Coronavirus-related neologisms: a challenge for Croatian standardology and lexicography. In Annette Klosa-Kückelhaus & Ilan Kernerman (eds.), *Lexicography of coronavirus-related neologisms* (Lexicographica: series maior 163), 163–190. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110798081-009.

Muassomah, Muassomah. 2023. Language of COVID-19: Language absorption in the pandemic vocabulary from English to Arabic. *International Journal of Arabic-English Studies* 24(1). 173–192. https://doi.org/10.33806/ijaes.v24i1.565.

Muischnek, Kadri & Liina Lindström. 2020. Digitaalsed tekstiandmed ja korpuslingvistika. Metodoloogiline teejuht. In Anu Masso, Katrin Tiidenberg & Andra Siibak (eds.), *Kuidas mõista andmestunud maailma? Metodoloogiline teejuht* (Gigantum Humeris), 306−339. Tallinn: Tallinna Ülikooli Kirjastus.

Nam, Kilim, Jinsan An & Hae-Yun Jung. 2022. The emergence and spread of Korean COVID-19 neologisms in news articles and user comments and their lexicographic description. In Annette Klosa-Kückelhaus & Ilan Kernerman (eds.), *Lexicography of coronavirus-related neologisms* (Lexicographica: series maior 163), 43–72. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110798081-004.

Née, Émilie & Marie Veniard. 2012. Lexical discourse analysis: toward a revival using a semantic approach. *Langage et société* 140 (2). 15–28.

Ng, Reuben, Ting Y. J. Chow & Wenshu Yang. 2021. News media narratives of Covid-19 across 20 countries: early global convergence and later regional divergence. *PLoS One* 16(9). https://doi.org/10.1371/journal.pone.0256358.

Oxford Languages. 2021. *Word of the Year: Vax. A report into the language of vaccines*. Oxford University Press. https://languages.oup.com/wp-content/uploads/oxford-languages-word-of-the-year-2021.pdf (Accessed 2024-05-29).

Papp, Judit. 2022. How the COVID-19 pandemic is changing the Hungarian language: Building a domain-specific Hungarian/Italian/English dictionary of the COVID-19 pandemic. In Annette Klosa-Kückelhaus & Ilan Kernerman (eds.), *Lexicography of coronavirus-related neologisms* (Lexicographica: series maior 163), 147–162. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110798081-008.

Paton, Bernadette. 2020. Social change and linguistic change: the language of Covid-19. *Oxford English Dictionary Blog*. https://www.oed.com/discover/topics-from-the-21st-century (Accessed 2024-05-29).

Raet, Mai. 2023. When negative turns out to be positive: Exploring changes in word associations in the aftermath of the COVID-19 pandemic. *Półrocznik Językoznawczy Tertium* 8(1). 71−98. https://doi.org/10.7592/Tertium.2023.8.1.248.

Raet, Mai. 2024. Terminikasutuse problemaatilisusest COVID-19 näitel [Navigating the challenges of term usage: a case study of COVID-19]. *Emakeele Seltsi aastaraamat* 69. 229−252. http://dx.doi.org/10.3176/esa69.10.

Rodríguez Guerra, Alexandre. 2016. Dictionaries of neologisms: a review and proposals for its improvement. *Open Linguistics* 2(1). 528–556. https://doi.org/10.1515/opli-2016-0028.

Roig-Marín, Amanda. 2021. English-based coroneologisms: a short survey of our Covid-19-related vocabulary. *English Today* 37(4). 193–195. https://doi.org/10.1017/S0266078420000255.

Salazar, Danica & Kate Wild. 2022. The Oxford English Dictionary and the language of Covid-19. In Annette Klosa-Kückelhaus & Ilan Kernerman (eds.), *Lexicography of coronavirus-related neologisms* (Lexicographica: series maior 163), 11–26. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110798081.

Sarif, Irzam & Dadang Suganda. 2020. Interferences of English-Japanese language in the Covid-19 pandemic. *IZUMI: Japanese Language, Literature and Culture Journal* 9(2). 121–127. https://doi.org/10.14710/izumi.9.2.121-127.

Sinclair, John McHardy. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Statistics Estonia. 2022. Population census. 76% of Estonia's population speak a foreign language. https://rahvaloendus.ee/en/news/population-census-76-estonias-population-speak-foreign-language (Accessed 2024-05-29).

Storjohann, Petra & Luisa Cimander. 2022. Annäherung an den Coronadiskurs in der öffentlichen Kommunikation mithilfe von Neologismen. In Mariusz Jakosz & Marcelina Kałasznik (eds.), *Corona-Virus-Pandemie: Diverse Zugänge zu einem aktuellen Superdiskurs*, 25–50. Göttingen: V&R Unipress. https://doi.org/10.14220/9783737015127.25.

Stuart-Smith, Jane. 2017. Sociolinguistic approaches: variationist frameworks. In Colleen Cotter & Daniel Perrin (eds.), *The Routledge handbook of language and media*, 27–43. London: Routledge. https://doi.org/10.4324/9781315673134-4.

Sõnaveeb = Sõnaveeb. 2023. Indrek Hein, Jelena Kallas, Kristina Koppel, Margit Langemets, Kaur Männiko, Tõnis Nurk & Ülle Viks. Developer OÜ TripleDev. Eesti Keele Instituut. Sõnaveeb. http://www.sonaveeb.ee.

Štrkalj Despot, Kristina & Ana Ostroški Anić. 2021. A war on war metaphor: Metaphorical framings in Croatian discourse on Covid-19. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje* 47(1). 173–208. https://doi.org/10.31724/rihjj.47.1.6.

Talbot, Mary. 2007. *Media discourse: representation and interaction*. Edinburgh: Edinburgh University Press.

Tang, Min, Rioni Reiter-Palmon & Zorana Ivcevic. 2022. Creativity and innovation in times of crisis (COVID-19). *Frontiers in Psychology* 13. https://doi.org/10.3389/fpsyg.2022.858907.

Terviseamet. 2023. *COVID-19 pandeemia kirjeldav analüüs ja õppetunnid*. Tallinn: Terviseamet.

Trap-Jensen, Lars & Henrik Lorentzen. 2022. Recent neologisms provoked by COVID-19 in the Danish language and in the Danish dictionary. In Annette Klosa-Kückelhaus, Stefan Engelberg, Christiine Möhrs & Petra Storjohann (eds.), *Dictionar-*

ies and Society. Proceedings of the XX EURALEX International Congress, 12–16 July 2022, Mannheim, Germany, 825–832. Mannheim: IDS-Verlag. https://doi. org/10.14618/ids-pub-11341.

Verheyen, Steven, Elisabeth Droeshout & Gert Storms. 2019. Age-related degree and criteria differences in semantic categorization. *Journal of cognition* 2(1). 1–20. https://doi.org/10.5334/joc.74.

Vitória, Beatriz de Araújo, Maria Teresa Ribeiro & Vânia Sofia Carvalho. 2022. The work-family interface and the COVID-19 pandemic: a systematic review. *Frontiers in Psychology* 13. https://doi.org/10.3389/fpsyg.2022.914474.

Voolaid, Piret. 2022. Kaugõppe kujutamine laustaudimeemides: huumor kui toime-tulekuviis ja enesekaitsestrateegia. *Sirp: Eesti Kultuurileht* 9(3881). 35−37.

WHO 2023 = World Health Organization. 2023. *Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic.* https:// www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coro-navirus-disease-(covid-19)-pandemic (Accessed 2024-05-29).

Kokkuvõte. Mai Raet: Leksikoloogiline vaade pandeemiale: juhtumi-uuring Eesti näitel. Koroonapandeemia tõi kaasa mitmeid uusi sõnu ja väljendeid, mis loodi spetsiaalselt sellega seotud nähtuste nimetamiseks ja kirjeldamiseks. Lisaks sai üldkeeles tavapäraseks kasutada meditsiini ja viroloogia erialakeelde kuuluvaid termineid, mis olid varasemalt laiemale elanikkonnale vähetuntud. Ehkki pandeemiaga seotud sõnavara on maailmas laialdaselt uuritud, puudub eesti keele kohta põhjalikum ülevaade. Seetõttu on siinse töö eesmärk see lünk täita: artiklis identifitseeritakse ja analüüsi-takse Eesti koroonadiskursuse kõige iseloomulikumad keelendid aastatel 2020–2022. Andmete kogumiseks kasutati Eesti keele ühendkorpust 2023 ja EKI ühendsõnastikku 2023. Kokku analüüsiti 182 leksikaalset elementi, neist 22 moodustasid eesti koroonaaja tuumsõnavara, 91 kategoriseeriti koroona-neologismidena. Eesti koroonadiskursuse põhikomponentide tuvastamiseks loodi 10 semantilist kategooriat ning vaatlusaluse perioodi jooksul toimunud temaatiliste nihete kaardistamiseks kasutati Sketch Engine'i esilduvusskoori (keyness score). Siinne artikkel annab ühelt poolt sissevaate Eesti koroona-kogemusse, teisalt illustreerib, kuidas väiksema kõnelejate arvuga keel suudab globaalsetele muutustele adekvaatselt ja kiiresti reageerida.

Märksõnad: COVID-19, neologismid, sõnastikud, korpuslingvistika, keele-muutus, eesti keel