Topics in Linguistics (2025), 26(1), pp. 68–91
10.17846/topling-2025-0004

COVID-related terms in translated popular scientific articles: A case study on term variation

Wioleta Karwacka*
University of Gdańsk, Poland

Key words

term
term variation
terminology
translation
corpus
popular science

Abstract

This article presents a case study of COVID-related terms in popular science articles. It is a corpus-based analysis of articles published in Scientific American and translated into Polish. Its aim is to observe COVID-related terminology variation in translated popular science texts. The bilingual corpus used in this study comprises 51 pairs of texts published in 2020, 2021, and 2022 in Scientific American and translated into Polish and then published in *Świat Nauki*. The design for this study included corpus creation and compilation, automated term extraction, exploration of the corpus for relevant contexts, equivalent terms and variants (concordance, parallel concordance). The findings confirm variation in medical terminology in popularizations, possibly motivated by language structure and the need for clarity in popular science communication. The study revealed denominative, conceptual and linguistic variants. The most prominent variation concerned the term COVID.

1. Introduction

Terms play a crucial role in specialized communication, and can be considered one of the most prominent features of languages for special purposes. Medical terminology is diverse with its marked presence of variants, including words of Greek and Latin origin, acronyms, eponyms, and complex (multi-word) terms. New terms are coined as medical research advances knowledge of health and diseases. The global COVID pandemic was a period in which the general public could witness the formation of new terms related to the novel disease. This article presents a case study of such terms in popular science articles. It is a corpus-based analysis of articles published in Scientific American and translated into Polish. Its aim is to observe term variants in translated popular science texts.

* Corresponding author

Cite this paper: Karwacka, W. (2025). COVID-related terms in translated popular scientific articles: A case study on term variation. *Topics in Linguistics, 26*(1), 68–91. https://doi.org/10.17846/topling-2025-0004

2. Terms

Terms are central to specialized communication, structuring knowledge and disseminating new findings (Pecman & Kübler 2022). Terms have a linguistic form, a denotative function to mental or real-world objects (Cabré 1999a, 1999b), and – an essential feature – a subject domain (Daille 2017, pp. 11–12; Temmerman and Van Campenhoudt 2011, pp. 23–24). The form and meaning of terms are to an extent determined by the specialized domain and application (Daille 2017; L'Homme 2004). Bowker and Hawkins (2006, p. 83) proposed four principles for terms: monosemy and mononymy, linguistic accuracy, transparency, and conciseness, which is in line with Wüster's traditional General Theory of Terminology. It should be noted, however, that the monosemy and mononymy postulate is criticized as it is in conflict with the transparency and conciseness properties, and also with terminological practice, where homonymy and synonymy are frequently observed (Daille 2017, p. 12). The approaches to terminology which developed later, such as the Communicative Theory of Terminology (Cabré 1999b; Cabré Castellví 1998), emphasize the practical aspects of terminology in use, for instance, communicative needs of term users (Daille 2017, p. 12). What is more, while the traditional approaches, such as the Vienna School of Terminology (and Wüster's General Theory of Terminology) favoured the onomasiological approach to terminology, the more recent socio-cognitive and communicative approaches provide a framework for combining the semasiological and onomasiological perspective, and acknowledging that categories are not always distinctly delineated, and specialized terminologies typically have term synonyms or variants (cf. Temmerman 2000).

Terms are diverse in form: there are simple and complex terms. Simple terms are formed with one stem form, with or without affixes, and complex terms have one or more stem forms and even more stem forms can be added to them (Daille 2017, p. 12). In corpus studies, simple terms i.e. lexical units composed of one graphic unit are called single-word terms (SWT), while complex terms i.e. entities composed of units separated by a space or diacritical sign are called multi-word terms (MWT) (Daille 2017, pp. 12–13; L'Homme 2004; cf. Araúz, Cabezas-García, and Reimerink 2020, p. 2358).

The set of terms applicable to a particular special subject is called *terminology*. The same word may also refer to the principles and conceptual bases for the study of terms and the guidelines of terminographic work (Cabré 1999b).

2.1 Term formation

Forming new terms, known as term formation or neonymy, can proceed along the trajectories of the linguistic processes of derivation, morphological and syntagmatic compounding and syntagmatic formation (Cabré et al., 1998), composition or complex term formation and borrowing (Kageura 2022, p. 462). Those trajectories are observed in various languages, but their extent differs as dominant mechanisms for term formation are language-specific: for

instance, Romance languages show a tendency towards syntagmatic compounding, while German and Japanese lean towards morphological compounding. The French language is an interesting case, where syntagmatic compounding is dominant, but the medical domain shows a tendency towards morphological compounding (Daille 2017, p. 13). Metaphor is also an important phenomenon in term formation (Kageura 2022, p. 462).

Sager (1990; 1997) indicates two types of term formation: primary and secondary term formation. Primary term formation occurs when a name is coined for a new concept without a direct lexical precedent, usually in English, while secondary term formation occurs when a new name is used for a known concept and knowledge is transferred between communities through borrowings and translation, usually from English into other languages (Kageura 2022, p. 462). Standardized criteria for term formation include transparency, consistency, appropriateness, and linguistic economy (Kageura 2022, p. 462).

Seventy to eighty percent of terms in the resources of various domains are complex, e.g. *human translation*, *machine translation*, *statistical machine translation*, and, in fact, complex term formation seems to be the dominant pattern of term formation (Kageura 2022, pp. 462–463). Complex terms are formed based on simple terms so the termhood of complex terms results from the presence of at least one simple term within the complex one (de Santiago González and Grcic Simeunovic 2017, p. 7).

2.2 Term variation

Terminological variation or denominative variation is acknowledged and accepted (Freixa 2006; 2022, pp. 400–401). According to Daille (2017), a term variant is "an utterance which is semantically and conceptually related to an original term". A variant is determined in relation to a listed term in a thesaurus or in a terminological resource, and is semantically and conceptually connected to the original term, which means it can be a synonym of the term (a denominative variant), or reflect a semantic distance from the term (a conceptual variant) (Daille 2017, pp. 29–30).

Variation is not equivalent to synonymy, as synonymy covers the relationship between lexemes while denominative variation covers various types of variation: graphical, orthographical, morphosyntactic, and lexical variation of "lexicalized forms with stability and consensus among the users of units in a specialized domain" (Freixa 2006; 2022, p. 400). Particular variants are chosen in given situations based on the purpose of the text or the context of the communication situation (Bowker 1998; Freixa 2022).

Daille (2017) lists the following properties of variants:

1. a variant always involves at least one term;
2. a variant is obtained by applying at least one linguistic operation which belongs to a mechanism for denominative and conceptual variants;
3. a term can produce several variants;

4. the number of utterances of the term in a text is slightly superior to the number of utterances of the variant. Equality of utterance numbers may be encountered for graphical and denominative variants (Daille 2017).

Term variation is a diverse phenomenon and multiple types of term variants can be observed:

a) orthographic variants not affected by geographical origin and not associated with altering semantics or communicative situations (healthcare, health care);
b) diatopic variants (orthographic variants which do not affect semantics, dialectical variants which may affect semantics, culture-specific variants which affect semantics and the communicative situation, calques, and borrowings);
c) short form variants (abbreviations, acronyms);
d) diaphasic variants (scientific variants, informal variants, domain-specific variants);
e) cognitive variants (dimensional variants, intentional variants);
f) metonymic variants;
g) diachronic variants;
h) non-recommended variants;
i) morphosyntactic variants (Araúz, Cabezas-García, and Reimerink 2020, pp. 2358–2359; Faber and León-Araúz 2016).

A different classification is presented by Beatrice Daille (2017, pp. 37–63):

a) denominative variants (formed through synonymic substitution, simplification, exemplification, and competing patterns);
b) conceptual variants (formed through expansion and anaphorical reduction);
c) linguistic variants (variation mechanisms include graphics and spelling, inflection, derivation, fullback-compounding, modification, coordination, disjunction, and enumeration);
d) variants of register (variation of scientification/popularization, variants of position).

The classifications above offer different perspectives – while the former presents a taxonomy which includes a broad spectrum of formal term variants, the latter proposes a clear categorization encompassing denominative, conceptual, linguistic, and register variants. They can be deemed complementary as they both focus on different mechanisms underlying term variation.

In term variation, varying degrees of equivalence can be observed: total equivalence, conceptual overlapping, inclusion, and, finally, non-equivalence (León-Araúz 2022, p. 482). For instance, the term variants climate change and global warming are not equivalent but are used interchangeably, e.g. *changement climatique* is totally equivalent to *changement du climate*

and partially equivalent to *dérèglement climatique* (León-Araúz 2022, p. 484). What is noteworthy, traditional monoreferential approaches to terminology did not recognize the equivalence issue. It was not until later that equivalence and variation became significant aspects of the study of terminology (León-Araúz 2022, p. 477). While terminographers focus on decontextualized equivalence, much as in bilingual lexicography (context-free equivalence), translators seek context-sensitive equivalence and in that respect terminological equivalence may differ from translation equivalence (León-Araúz 2022, p. 479; Kerremans and Temmerman 2016, p. 59).

2.3 Medical terms

Medical terms, which are central to this study, are present in general and specialized discourse. For instance, English medical terms are found in three categories of medical English: basic English (BE), fundamental medical English (FME), and specialized medical English (SME) (Salager, 1983). Terms can be divided into categories structured along the lay/expert axis. These categories encompass specialized terms and units from the general register utilized in medical communication: dictionary-defined medical terms, co-text-defined medical terms, medical initialisms, medication brand names, and colloquial technical terms (Fage-Butler & Nisbeth Jensen 2016; Montalt et al. 2018). In medical terminology, adherence to monoreferentiality or univocity principles is not always maintained: medical terms incorporate lexical units associated with the general register, doublets, synonyms, and polysemous terms (Mitzkat et al. 2016).

Medical terms show features observed to varying degrees across languages, such as Latin and Greek influences (e.g. *kidneys* vs *renal failure*); affixation (e.g. *cardiac*, *cardiology*, *endocarditis*, *pericarditis*, *cardiopulmonary* etc.); eponymy (*Parkinson's disease*, *Alzheimer's disease*); register-specific variants (*heart attack* vs *myocardial infarction*); and wide-spread use of acronyms (*CBC*, *ESR*, *BP* etc.), which add to overall variation in the domain (Montalt et al., 2018).

3. Popularization articles

Textual terminology, which involves corpora-assisted analysis of terminology, focuses on tracing term usage rather than terminological standards (Condamines and Picton 2022, p. 220), and takes into account the specificity of particular text types or genres in a particular domain (Condamines and Picton 2022, p. 263) (Pecman & Kübler 2022). In this study, we focus on the use of term variants in popular science, also known as popularization of science, pop science, science communication, expository science, public science, and public understanding of science (Manfredi 2019, p. 64; Leane 2016, p. 8). Calsamiglia and van Dijk (2004) understand popularization as a whole range of communicative genres or even events which involve recontextualizing scientific discourse and transforming specialized knowledge into lay terms (cf. Manfredi 2019, p. 65). Perrault (2013) also emphasizes that science

popularization is science-related communication executed through a range of text types addressed at non-specialists. (cf. Manfredi 2019, p. 65). Accordingly, the language of popularization publications is expected to be adapted appropriately to the audience, genre and requirements of the medium.

Traditionally, popularization was viewed as the transfer of knowledge from scientists to laypeople, mediated by experts or journalists (Raffo 2017) and designed to fill a significant gap in non-specialists' knowledge (Myers 2003; Tekgül 2019). This view was reductionist, built on the assumption that scientific language is hermetic and needs to be decoded by a mediator who translates specialized texts into lay language (Manfredi 2019, p. 64). This traditional approach assumed the public's inability to understand and appreciate scientific achievements and took a linear, pedagogical, and paternalistic view of communication (Bucchi & Trench 2012; Raffo 2017). Popularization understood as a one-way transfer of knowledge from the expert to the lay reader ignores the importance of the interaction between all parties involved in popularization (Calsamiglia & Van Dijk, 2004; Tekgül, 2019). Nowadays, popularization is seen as a reformulation of scientific material, which involves interaction and transmission of information, so the communicative dimension is emphasized (Myers 2003; Raffo 2017). Popularization is, in fact, public communication of science, whereby social, political, and cultural contexts are significant in the introduction of new knowledge, as well as the status and importance of scientific facts (Bucchi & Trench 2012; Raffo 2017).

The boundary between specialized and general language is not clear-cut and defining "term" in opposition to "word" is also problematic due to the grey areas and often ambivalent terminological status of the term in question (Raffo 2017). Even though specialized terminology is used in science popularization, replacing terms with colloquial words or paraphrases is said to be one of the features associated with popularization, but the study of this sphere has not yet yielded conclusive results (Raffo 2017). There are available data that are in line with the above thesis, but also indicate the existence of a grey area between specialized terminology and general lexicon (Ciapuscio 2003; Raffo 2017). A partial overlap has also been observed between the vocabulary used in texts for specialists and those intended for laypeople (Raffo 2017), and some analyses even point to a similar level of technicality in popular science texts, textbooks and specialized articles (Bucchi 2008, p. 59; Casadei (1994, cited in Bucchi 2008, p. 59) (Raffo 2017).

4. Aim

The global COVID pandemic brought about new terms for the novel virus and the novel disease. The aim of this corpus study was to observe potential medical term variation, particularly COVID-related terminology variation in original and translated popular science articles, which communicated the new findings to the public.

5. Material and methods

5.1 Material

Quite understandably, the COVID pandemic generated public interest in the novel virus and disease, which was reflected in the number of publications on the subject. One of the ways to avoid diachronic bias is selecting texts from a "discursive moment" (*moment discursif*) as understood by Moirand (2003) – a narrow timeframe when the media discuss a given subject intensely (Moirand, 2003; Raffo, 2017). In this study, the texts were collected from one outlet but only for three years when the pandemic received a lot of media attention and coverage. As a new disease, COVID was not discussed in the media before 2020 and it received a lot of attention in march 2020 until January 2022, and since December 2022 the interest has been minimal. Figure 1 is a Google Trends graph showing the trend for Google web searches for the keyword "covid" from 2019 to 2023.

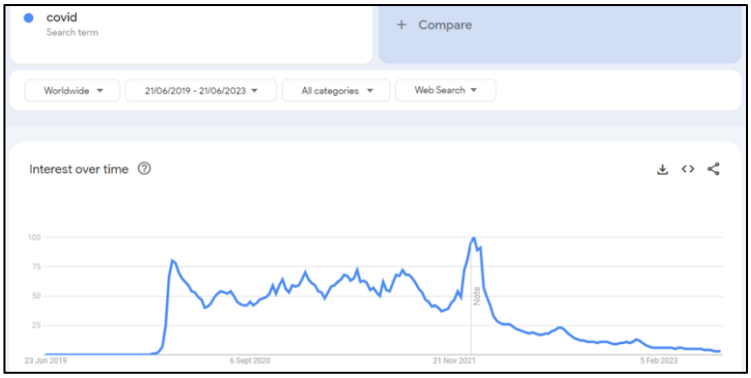

Figure 1. A Google Trends *graph illustrating the increase and decrease in the number of searches for the keyword "covid" from 2019 to 2023. Retrieved on 29 September 2023 from https://trends.google.com/trends/explore?date=2019-06-21%202023-06-21&q=covid*

There is lack of consensus on corpus size (Condamines and Picton 2022, p. 231; Corpas Pastor and Seghiri Domínguez 2010), but the tendency is to compile the largest possible corpora, heterogeneous in genre and domain (Condamines and Picton 2022, p. 231). As a rule of thumb, 20–50 occurrences of each expression ensure sufficient representativeness (Condamines and Picton 2022, p. 231). However, specialized corpora are difficult to collect and compile (Condamines and Picton 2022, p. 232), so the specialized domain is associated with limited corpus sizes, which is a major problem, as is the sparseness of data due to the

limited number of occurrences and shared contexts (Bertels 2022, p. 324). The scarcity of specialized parallel corpora poses yet another significant problem in terminology research (Araúz, Cabezas-García, and Reimerink 2020, p. 2360). Perhaps it is worth noting that smaller corpora can be used for studying text genres in a specialized domain: as Watson (2001) observes, a specific corpus does not have to be very large if it has "a reasonable number of examples of each target word". Therefore, with high frequency words, a corpus of around 50,000 words should produce at least 10 examples of each word. In the case of low-frequency words, the corpus could be expanded to up to 200,000 words. Beyond this point, Watson (2001) postulates, "the law of diminishing returns and the size of the file containing the corpus make further expansion not worth the effort" (Watson, 2001).

The above factors were taken into consideration – the corpus used in this study (bilingual focus corpus) comprises 51 pairs of texts published in 2020, 2021, and 2022 in Scientific American and translated into Polish and then published in *Świat Nauki*. The corpus only contains text bodies, as the Polish texts were re-titled and given new headings by the *Świat Nauki* editors. The translators were not identified by name. The corpus contains all Scientific American articles that were translated into Polish, published between 2020 and 2022, and discussed COVID, SARS-CoV-2, or the pandemic. The corpus was first processed manually by removing titles and headers, and converting files into text-only documents. The texts were then uploaded to Sketch Engine and automatically compiled. The alignment was checked manually, and where texts were not correctly aligned the files were checked, cleaned, and re-aligned. As a result, a bilingual parallel focus corpus was created (focus corpus). The total number of words is 148,353 (172 961 tokens; 76,977 words [88,255 tokens, 51 texts] for the English focus subcorpus, and 71,376 words [84,706 tokens, 51 texts] for the Polish focus subcorpus).

Additionally, to compare preferences for variants in Polish non-translations, a reference domain-specific corpus "Corona" was created in Sketch Engine semi-automatically with its "find texts on the web feature" by selecting popular science texts on the COVID pandemic available online (*COVID* and *pandemic* were used as search words), written in Polish and published in recognized outlets popularizing science (84,755 tokens, 71,029 words), as presented in Table 1.

Large corpora available in Sketch Engine were selected for the calculations of keyness performed by Sketch Engine: English Web 2021 (enTenTen21) (61,585,997,113 tokens; 52,268,286,493 words) and Polish Web 2019 (plTenTen19) (5,216,428,620 tokens; 4,253,636,443 words).

Table 1. The sources of Polish popular science articles in the reference corpus

Source of texts	Frequency (number of texts)	Tokens
pulsmedycyny.pl	12	14563
kopalniawiedzy.pl	9	10079
zdrowie.wprost.pl	9	8856
totylkoteoria.pl	4	6169
forbes.pl	2	6092
scienceinpoland.pap.pl	5	4311
crazynauka.pl	3	4156
naukawpolsce.pl	3	3641
portal.abczdrowie.pl	2	3395
zdrowie.radiozet.pl	5	3297
naukapolska.pap.pl	4	2707
wiadomosci.onet.pl	2	2044
poradnikzdrowie.pl	1	1959
zdrowie.pap.pl	1	1748
tvn24.pl	3	1727
konkret24.tvn24.pl	1	1497
polityka.pl	1	1417
businessinsider.com.pl	1	1074
zdrowie.tvn.pl	1	982
pap.pl	2	966
krokdozdrowia.com	1	855
stanwiedzy.pl	1	669
expressilustrowany.pl	1	565
polsatnews.pl	1	562
o2.pl	2	541
se.pl	1	460
focus.pl	1	423

5.2 Methodology

The methodology design in this study is derived from Pecman and Kübler (2022), Bourigault and Slodzian (1999), and (de Santiago González and Grcic Simeunovic (2017). The stages of a corpus-based and textual approach to terminology with emphasis on text genres by Pecman and Kübler (2022), in turn based on Bourigault and Slodzian (1999), include the creation of a corpus (selection of a genre, corpus arrangement); exploration of corpora for terminological analysis; exploration of corpora to find relevant contexts; context analysis (analysis of contextual environment of terms with respect to genres); and contrastive analysis (analysis of terms and/or their variants with respect to genres) (Pecman and Kübler 2022, p. 282). As far as term extraction methodology is concerned, de Santiago González and Grcic Simeunovic (2017, p. 6) propose the following steps: term extraction, establishing candidate terms; and verification.

The design for this study followed the steps below:

- corpus creation and compilation
- automated term extraction in Sketch Engine (keywords and terms), verification in the *International Classification of Diseases* (ICD) and Coronavirus Glossary of Common Terms and manual corrections
- exploration of the corpus for relevant contexts (concordance)
- exploration of the corpus for equivalent terms and variants (parallel concordance)

In this study terms were extracted automatically with Sketch Engine using its "keywords" and "terms" features, which are defined in the quote below:

> **Keywords**
>
> Keywords are individual words (tokens) which appear more frequently in the focus corpus than in the reference corpus. Any token can qualify for a keyword if it is used more frequently in the focus corpus than in the reference corpus. In reality, the result will include mainly nouns and adjectives because the frequencies of other parts of speech tend to be similar in all texts.
>
> **Terms**
>
> Terms are multi-word expressions which appear more frequently in the focus corpus than in the reference corpus and, additionally, match the typical format of terminology in the language. The format is defined in the term grammar.
>
> The result of term extraction is displayed as lemmas. Gender lemmas are used for languages where the word form of an adjective has to match the gender of the noun.
>
> The simple maths method is used to determine the keyness score of both keywords and terms. It works with normalized (relative, per million) frequencies in the focus and reference corpora. (Kilgarriff et al., 2004, 2014; *Sketch Engine*, n.d.)

The functions allowing for the extraction of keywords and terms are based on the "simple maths" method used in Sketch Engine computations, as explained below:

> **Simple maths** is the keyness score used in Sketch Engine to identify keywords, terms, key n-grams and key word sketch collocations. Simple maths compares the frequencies in the focus corpus with the frequencies in the reference corpus. Alternatively, two subcorpora in the same corpus or in different corpora can be used.
>
> The N value makes the score prefer more frequent or less frequent items.
>
> A higher N value shifts to focus on higher-frequency words (more common words), whereas a lower N value focusses on low-frequency (rarer words). The value should be changed in orders of magnitude, i.e. 0.1, 1, 10, 100, 1000, 10000 etc. Smaller changes rarely produce any noticeable effect.
>
> The statistics is a variation on "word W is so-and-so times more frequent in corpus X than corpus Y". The formula is:
>
> $$\frac{fpm_{focus} + N}{fpm_{ref} + N}$$
>
> where
>
> $\text{fpm}_{\text{focus}}$ is the normalized (per million) frequency of the word in the focus corpus,
>
> fpm_{ref} is the normalized (per million) frequency of the word in the reference corpus,
>
> N is the smoothing parameter ($N = 1$ is the default value). (Kilgarriff et al., 2004, 2014; Sketch Engine, n.d.)

The starting point for the explorations were the single-word and multi-word term lists in the Polish focus subcorpus. Then, parallel concordance was explored to see the terms in the original English focus subcorpus. Next, the English-Polish parallel concordance was explored for term variation in translation. Term variants were also verified in the reference corpus of the Polish non-translations (Corona).

The cut-off point for this analysis is 20 occurrences to ensure sufficient representativeness (Condamines & Picton, 2022). The terms were organized in the descending order of keyness values (Sketch Engine keywords) to capture the keywords specific to the COVID subject domain, i.e. the keywords which are much more frequent in the study corpus than the reference corpus used by Sketch Engine to determine keyness (Polish Web 2019), which at the time of the study did not contain texts published later than in 2019, i.e. did not include texts published at the time or after the pandemic outbreak. Table 2 shows top 15 highest keyness keywords with 20+ occurrences in the study corpus (Polish subcorpus).

Table 2. 15 highest keyness keywords with 20+ occurrences in the Polish focus subcorpus

No	Keyword	Frequency	Keyness
1	covid-19	179	2114,191
2	pandemia	213	1428,76
3	covid	118	1394,054
4	sars-cov-2	92	1087,11
5	fag	48	444,194
6	koronawirusa	30	350,928
7	sars	33	336,8
8	pcr	54	330,051
9	cytokina	50	315,812
10	patogen	87	296,861
11	covid-u	24	284,333
12	koronawirus	20	234,637
13	rbd	22	234,553
14	wirus	430	186,07
15	węchowy	24	183,011

Next, domain-specific terms – the names of the novel disease and virus – COVID / COVID-19, SARS-CoV-2 and *coronavirus* were further explored (sections 6.1–6.3). Additionally, the name of a symptom – anosmia – was analysed, as its frequency in the keyword list in English much exceeded the frequency of the direct equivalent *anosmia* in the Polish list, which predicted some variation (section 6.4). The last step was the analysis of multi-word terms with 20 or more occurrences in the corpus (section 6.5). The categories of variants are based on Daille (2017) and Araúz et al. (2020).

Whenever translation strategies are mentioned, they follow the classification proposed by Chesterman (2016, pp. 91–109):

- syntactic strategies: literal translation, loan, calque, transposition, unit shift, phrase structure change, clause structure change, sentence structure change, cohesion change, level shift, scheme change;
- semantic strategies: synonymy, antonymy, hyponymy, converses, abstraction change, distribution change, emphasis change, paraphrase, trope change and other semantic changes;

– pragmatic strategies: cultural filtering, explicitness change, information change, interpersonal change, illocutionary change, coherence change, partial translation, visibility change, transediting and other pragmatic changes.

The pragmatic strategy of explicitation, which is frequently observed in the corpus and thus referred to in the results in discussion sections, is defined as a change "either towards more explicitness (explicitation) or more implicitness (implicitation)" (Chesterman 2016, p. 105).

6. Results

6.1 COVID

COVID is the name coined by the WHO for the coronavirus disease (WHO n.d.). Covid, Covid-19 and the declension form Covid-u were recognized by the Sketch Engine as separate keywords rather than lemmas of the same word – the result would be 321. The declension forms *COVID-owi* (8 hits), *COVID-em* (12 hits), *COVID-zie* (3 hits) were not recognized as a lemma of "covid" – so the total score for the keyword COVID should in fact be 165, and 344 if we add COVID-19 occurrences.

Both English and Polish show the tendency to use COVID and COVID-19 interchangeably: COVID was translated into COVID or COVID-19 but also – less frequently – coronavirus, pandemic or virus. COVID and COVID-19 are linguistic variants (Daille, 2017) or orthographic variants (Araúz et al., 2020; Faber & León-Araúz, 2016). Coronavirus, virus and pandemic are in fact terms for different concepts (COVID-19 is an abbreviation of the name for a disease – coronavirus disease-19 – caused by a strain of coronavirus, which is a name for one of many virus families) used in translation as near-synonyms, but can hardly be treated as conceptual (Daille, 2017) or metonymic variants (Araúz et al., 2020; Faber & León-Araúz, 2016). We can speculate that replacing concepts (not just terms) is accepted due to the genre (popular science) and therefore may be treated as register-specific or register term variation.

Tables 3, 4 and 5 present the occurrences of variants of 'COVID' and 'COVID-19'.

Table 3. Translation of "COVID" in the corpus

English	Polish	Frequency
COVID	COVID	161
COVID	COVID-19	53
COVID	Ø (implicitation)	9
COVID	wirus	3
COVID	koronawirus	2
COVID	pandemia	1

Table 4. Translation of "COVID-19" in the corpus

English	Polish	Frequency
COVID-19	COVID-19	97
COVID-19	COVID	2
COVID-19	Ø (implicitation)	11
COVID-19	wirus	1
COVID-19	koronawirus	1
COVID-19	pandemia	2
COVID-19	patogen	1

Table 5. Source units which were translated into "COVID-19"

English	Polish	Frequency
COVID-19	COVID-19	97
COVID	COVID-19	52
Ø	explicitation COVID-19	30

In the Polish version, COVID and COVID-19 are variants used as equivalents of COVID. There was a noticeable, but not dominating explicitness change (Chesterman, 2016) towards implicitation to avoid using the abbreviated term COVID or COVID-19 as the context allowed for grasping that implicit information, which is quite typical of popular science as it leads to determinologization.

Examples of implicitation:

(1) a. *COVID nearly shut down the economy*
    b. *niemal całkowitego zatrzymania gospodarki*
    nearly shutting down the economy
(2) a. *COVID deaths*
    b. *zgonów*
    deaths
(3) a. *worsen the symptoms of COVID-19*
    b. *nasilenia choroby*
    exacerbate the disease

There are 115 occurrences of COVID-19 in the English corpus, and 179 in the Polish corpus. There are a few reasons why the number is larger in the Polish corpus: firstly, COVID was translated into COVID-19 52 times (the reverse choice, i.e. translating COVID-19 into COVID, was observed in 2 occurrences), and there were instances of explicitness change towards greater explicitation (Chesterman, 2016), when the term COVID-19 was added, possibly to remove ambiguity.

Examples of explicitation:

(4) a. *the lab-leak theory*
    b. *teza o laboratoryjnej genezie COVID-19*
    the thesis of the laboratory genesis of COVID-19

(5) a. *raging pandemic*
    b. *szalejącej pandemii COVID-19*
    raging pandemic of COVID-19

What is interesting is there was a visible variation concerning inflection. Polish is an inflected language, but acronyms can be inflected or not, depending mainly on phonetic criteria and conventions. As a new word, COVID had not yet been associated with any conventions when the texts were translated, and there are no phonetic contraindications for inflecting, which may explain why both the inflected and non-inflected variants are found in the corpus. There were 23 COVID inflected forms (other than the nominative case) out of 165, so there was a stronger tendency towards non-inflection (for instance, *epidemia COVID, pandemia COVID* rather than inflected *COVID-u*, *długim COVID* rather than *COVID-zie*, *przeciwko COVID* rather than *COVID-owi*).

Table 6. Inflected and non-inflected COVID

	COVID in the Polish focus corpus	**COVID in the Polish focus corpus ("Corona")**
Inflected	23 (14%)	1 (2.5%)
Non-inflected	142 (86%)	35 (97.5%)

COVID-19, in turn, is always non-inflected both in the focus and reference corpus. The Polish reference corpus showed a strong tendency to use COVID-19 over COVID (476 vs. 36 hits).

6.2 SARS-CoV-2 72

SARS-CoV-2 72 occurrences, translated into SARS-CoV-2 (60) and *wirus* SARS-CoV-2 (14 occurrences, examples below). As SARS-CoV-2 stands for severe acute respiratory syndrome coronavirus, and in fact mentions "virus" in its name, adding another "virus" may seem redundant, which means it in fact fosters understanding and makes the text more lay-friendly and helps avoid inflection problems. See for example:

(6) a. *constant exposure to SARS-CoV-2*
    b. *stały kontakt z wirusem SARS-CoV-2*
    constant exposure to virus SARS-CoV-2

(7) a. *infected with SARS-CoV-2*
b. *zakażonych wirusem SARS-CoV-2*
infected with virus SARS-CoV-2

6.3 Coronavirus

The term coronavirus does not seem to have variants and is consistently translated into *koronawirus* with some tendency towards implicitation (explicitness change) and ellipsis without changing the meaning of the passage.

For example:

(8) a. *infected with the novel coronavirus*
b. *przeszło się infekcję*
([one] has endured an infection)
(9) a. *deaths from the novel coronavirus*
b. *liczba zgonów*
number of deaths
(10)a. *coronavirus pandemic*
b. *pandemia*
pandemic

Table 7. Translation of *coronavirus*

English	Polish	Frequency
coronavirus	koronawirus	97
coronavirus	COVID-19	1
coronavirus	virus	1
coronavirus	implicitation	6

To calculate the total frequency of the Polish words *koronawirus* (20 hits) its inflection forms should be added: *koronawirusa* (30 hits) as well as *koronawirusem* (16), *koronawirusowi* (6), *koronawirusy* (17), *koronawirusów* (14), *koronawirusami* (6), k*oronawirusom* (2); the total number of all koronawirus* lemmas in the Polish focus subcorpus is 111. In the English subcorpus, there are 120 occurrences of *coronavirus*. There is a difference between the number of occurrences of *coronavirus* and *koronawirus* because of explicitness changes (for example, *the pathogen – koronawirus*) or translating COVID, COVID-19 into *koronawirus*.

6.4 Anosmia

Anosmia had 22 occurrences in the English subcorpus. It transpired that it was translated into three denominative variants: *anosmia, utrata węchu* (*loss of smell*), *brak węchu* (*lack of smell*) with one sentence that used the term *anosmia* in parenthesis following *utrata węchu*, which is

a descriptive, thus more transparent term. It was consistent with the source text, which also provided an explanation for anosmia. *Utrata węchu* is a standard term, for instance, used in the Polish version of the International Classification of Diseases (ICD). The English version of ICD uses *anosmia*.

Table 8. Translation of *anosmia*

English	Polish	Frequency
anosmia	anosmia	14
anosmia	utrata węchu (anosmia)	1
anosmia	utrata węchu	5
anosmia	brak węchu	2

The search in the reference corpus of Polish non-translations did not produce any results for anosmia in Polish. There were three denominative variants *utrata węchu* (6)(loss of smell), *zaburzenia węchu* (7)(disturbances of smell), *pogorszenie węchu* (1)(worsening of smell).

6.5 Multi-word terms

In the Polish corpus, none of the multi-word terms (MWTs) had more than 20 occurrences, which predicted a considerable deal of variation. In the English corpus, only 7 MWTs had 20 or more hits, as presented in Table 8.

Table 8. Multi-word terms (MWT) with more than 20 occurrences in the English focus subcorpus sorted by keyness score

No	MWT	Frequency	Keyness
1	immune system	64	83,957
2	social distancing	36	67,109
3	immune response	20	65,866
4	t cell	22	61,213
5	infectious disease	25	50,668
6	amino acid	21	37,154
7	public health	35	18,104

Immune system (64 occurrences) is translated into the standard corresponding term *układ odpornościowy* in the vast majority of occurrences (59). The competing form – a calque with identical meaning *system odpornościowy* (a non-recommended variant, c.f. Araúz et al., 2020; Faber & León-Araúz, 2016) is used once and a more general term – *odporność* (immunity) – is used once, too. In 3 occurrences, immune system is a part of another MWT – innate immune system – and is translated into *układ odporności wrodzonej* (system of innate immunity). *Układ odpornościowy* was not automatically recognized as a term in Sketch Engine, but a simple

search produces 73 occurrences of *układ odpornościowy* in all possible inflections – the difference is due to rephrasing, or actually explicitation, as the target *układ odpornościowy* corresponds with source immune activity, response, proteins, damping, immunity.

The translation of *social distancing* (36 occurrences) produces 4 linguistic (Daille, 2017) or morphosyntactic (Araúz et al., 2020; Faber & León-Araúz, 2016) variants: *dystans społeczny* (social distance, 15 occurrences), *dystansowanie społeczne* (social distancing, 14 occurrences), *dystansowanie się społeczne* (social distancing [oneself], 2 occurrences), *społeczne dystansowanie* (social distancing, 1 occurrence).

Immune response (20 occurrences) is translated into two denominative variants – *odpowiedź immunologiczna* (immune response, 13 occurrences) and *reakcja immunologiczna* (immune reaction, 6 occurrences) with one occurrence of *układ odpornościowy* (immune system).

T-cells are always translated into *limfocyty T* (lymphocytes T), without any variation, which is quite surprising as the word *komórka* (cell) is certainly more recognized than the specialized term *limfocyt*. Virtually no variation is observed in translating the term *infectious disease*; it is translated into *choroba zakaźna* (21 occurrences, infectious disease). Of the remaining 4 occurrences: 1 is made implicit as infectious disease is translated into *choroba* (disease, 1 occurrence); borrowing is used in 2 instances when the term infectious diseases is a part of a name of a research centre; and in 1 case a single word *zakaźnik* is used to denote an infectious diseases epidemiologist. No variation is observed in translating the term *amino acid* – it is always translated into a simple term *aminokwas*.

The phrase *public health* recognized as a MWT has 39 occurrences in the corpus and is translated into *zdrowie publiczne* (public health) 22 times; the remaining hits are conceptual variants, which include: *służba zdrowia* (health service, 2 occurrences); *służby medyczne* (medical services, 1 occurrence); *ochrona zdrowia* (health protection, 1 occurrence). The borrowing *public health* is used 9 times in the name of an institution. There are 2 implicitations: *grupy* (groups) and *naukowcy* (researchers), and one apparent mistranslation *rządowe agendy zdrowotne* (government health subsidiaries/agendas).

7. Discussion

This study explored a relatively small specialized corpus for term variation, and although its results cannot be treated as universal linguistic paradigms, we can make certain observations on terminology in translated popularization (c.f. Condamines and Picton 2022, p. 227). The study shows term variation in translated popular science articles discussing the COVID pandemic. The most prominent variation concerned the term COVID and its orthographic and inflected vs non-inflected forms. There are occurrences of denominative variants (e.g. *anosmia, brak węchu*); conceptual variants (e.g. *zdrowie publiczne, ochrona zdrowia*); linguistic variants (COVID, COVID-19); and, to an extent, register specific variation, as the terms *coronavirus*,

pandemic, and *virus* seem to be used interchangeably in popularizations even though they denote different concepts in scientific or medical contexts. Term choices can be motivated conceptually, when a variant indicates aetiology or location, linguistically, when a variant is determined by collocations or socially, i.e. by social conventions and standards (Bowker & Hawkins, 2006).

Term variation is wide-spread, and it also concerns medical texts (Bertels 2022, p. 323). As far as previous studies of translated popularizations are concerned, they show hyponymy, elimination of nominalization, and introducing of nominalization (Manfredi 2019, pp. 185–211). In this study, hyponymy was observed in the relationship between *T-cells* and *limfocyty T*, as *lymphocyte* is a more specific term than *cell*, but it has to be noted that in translation they are equivalent, i.e. they are standard terms used to denote the same concept. Certain differences in term use observed between the subcorpora resulted from translation strategies, especially explicitness change when a term was added to make an implicit concept more explicit or vice versa (e.g. *the pandemic* vs *COVID pandemic*).

The results also confirm that the categories of specialized and lay lexicon are not clear-cut, and that specialized vocabulary is used in texts for lay recipients (c.f. Ciapuscio 2003; Raffo 2017).

The findings in this study are also in line with previous studies confirming that Polish term formation is influenced by English, and used to be influenced by Greek and Latin – there are numerous borrowings from English in Polish terminology, and abbreviations are the largest group of borrowings (Górnicz 2019, p. 274). The term coronavirus is imported to Polish through English from Latin. The abbreviated forms COVID and SARS-CoV-2 are also borrowed from the English terms which were based on Latin.

8. Conclusion

This corpus study focused on observing variation in medical terminology, particularly COVID-related terms, in both original and translated popular science articles. The study revealed denominative, conceptual, and linguistic variants. The most prominent variation concerned the term COVID: COVID and COVID-19 are used interchangeably, with occasional translations to *coronavirus*, *pandemic*, or *virus*. These variations could be attributed to science popularization, which may be related to register-specific variation (specialized vs lay). There was variation in inflection. While Polish is an inflected language, both inflected and non-inflected forms of COVID were found in the corpus. This variation might be due to the absence of established conventions at the time of translation of a relatively new term such as COVID. The term SARS-CoV-2 was consistently translated into SARS-CoV-2 or *wirus* SARS-CoV-2. The term *coronavirus* was consistently translated into *koronawirus* with occasional implicitation and ellipsis. The translation of the term *anosmia* produced considerable variation and so did the translation of multi-word terms. In summary, this corpus study revealed variation in

medical terminology in popularizations, possibly motivated by language structure and the need for clarity in popular science communication.

Limitations

This study was conducted on a relatively small corpus in one language pair, so its conclusions are limited to this pair only, and should be treated as a case study. It does, however, show term variation, which should be further explored on larger corpora in other language pairs, but special attention needs to be paid to avoid diachronic bias in collecting a large number of texts for the corpus.

The large corpus of Polish used in Sketch Engine for the calculation of keyness was pre-pandemic, while the English corpus had already been updated. However, the differences in keyness score were not the focus of this study, so it did not affect the conclusions on the term variants in Polish.

Acknowledgements

This research was funded in whole or in part by the National Science Centre, Poland, grant number 2022/06/X/HS2/00545. For the purpose of Open Access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

References

Araúz, P., Cabezas-García, M., & Reimerink, A. (2020). *Representing Multiword Term Variation in a Terminological Knowledge Base: A Corpus-Based Study. Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 2358–2367.

Bertels, A. (2022). Chapter 14. Terminology and distributional analysis of corpora. In P. Faber & M.-C. L'Homme (Eds.), *Terminology and Lexicography Research and Practice* (Vol. 23, pp. 311–328). John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.23.14ber

Bourigault, D., & Slodzian, M. (1999). Pour une terminologie textuelle. *Terminologies Nouvelles*, *19*, 29–32.

Bowker, L. (1998). Variant terminology: Frivolity or necessity? *8th EURALEXInternational Congress, Liège: University of Liège*, 487–495.

Bowker, L., & Hawkins, S. (2006). Variation in the organization of medical terms: Exploring some motivations for term choice. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, *12*(1), 79–110. https://doi.org/10.1075/term.12.1.05bow

Bucchi, M., & Trench, B. (2012). Science communication research. In *Routledge Handbook of Public Communication of Science and Technology*. Routledge. https://doi.org/10.4324/9780203483794.ch1

Cabré Castellví, M. T. (1998). Do we need an autonomous theory of terms? *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, *5*(1), 4–19. https://doi.org/10.1075/term.5.1.03cab

Cabré, M. T. (1999a). *La terminología: Representación y comunicación elementos para una teoría de base comunicativa y otros artículos*. Instituto universitario de lingüística aplicada.

Cabré, M. T. (1999b). *Terminology: Theory, methods, and applications*. J. Benjamins Pub. Co.

Cabré, M. T., Cormier, M. C., & Humbley, J. (1998). *La terminologie: Théorie, méthode et applications*. les Presses de l'Université d'Ottawa A. Colin.

Calsamiglia, H., & Van Dijk, T. A. (2004). Popularization Discourse and Knowledge about the Genome. *Discourse & Society*, *15*(4), 369–389. https://doi.org/10.1177/0957926504043705

Chesterman, A. (2016). Memes of translation. *Memes of Translation*, 1–237.

Ciapuscio, G. E. (2003). *Textos especializados y terminología* (1. ed). Univ. Pompeu Fabra.

Condamines, A., & Picton, A. (2022). Chapter 10. Textual Terminology: Origins, principles and new challenges. In P. Faber & M.-C. L'Homme (Eds.), *Terminology and Lexicography Research and Practice* (Vol. 23, pp. 219–236). John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.23.10con

Corpas Pastor, G., & Seghiri Domínguez, M. (2010). Size matters: A quantitative approach to corpus representativeness.

Daille, B. (2017). *Term Variation in Specialised Corpora: Characterisation, automatic discovery and applications* (Vol. 19). John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.19

de Santiago González, P., & Grcic Simeunovic, L. (2017). The Polymorphic Behaviour of Adjectives in Terminography. *Meta*, *62*(1), 201–222. https://doi.org/10.7202/1040473ar

Faber, P., & León-Araúz, P. (2016). Specialized Knowledge Representation and the Parameterization of Context. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00196

Fage-Butler, A. M., & Nisbeth Jensen, M. (2016). Medical terminology in online patient–patient communication: Evidence of high health literacy? *Health Expectations*, *19*(3), 643–653. https://doi.org/10.1111/hex.12395

Freixa, J. (2006). Causes of denominative variation in terminology: A typology proposal. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, *12*(1), 51–77. https://doi.org/10.1075/term.12.1.04fre

Freixa, J. (2022). Chapter 18. Causes of terminological variation. In P. Faber & M.-C. L'Homme (Eds.), *Terminology and Lexicography Research and Practice* (Vol. 23, pp. 399–420). John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.23.18fre

Górnicz, M. (2019). *Wewnątrzjęzykowe uwarunkowania zapożyczeń technolektalnych w języku polskim*. Wydawnictwo Naukowe Instytutu Komunikacji Specjalistycznej i Interkulturowej. Uniwersytet Warszawski. http://sn.iksi.uw.edu.pl/wp-content/uploads/sites/306/2019/04/SN-44-Mariusz-G%C3%B3rnicz-Wewnatrzj%C4%99zykowe-uwarunkowania-zapo%C5%BCcze%C5%84.pdf

Kageura, K. (2022). Chapter 21. Terminological growth. In P. Faber & M.-C. L'Homme (Eds.), *Terminology and Lexicography Research and Practice* (Vol. 23, pp. 457–476). John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.23.21kag

Kerremans, K., & Temmerman, R. (2016). Finding the (un) expected?: Quantitative and qualitative comparisons of term variants and their translations in a parallel corpus of EU texts. In *Corpus-based Approaches to Translation and Interpreting: From theory to applications* (pp. 43–63). Peter Lang.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*, 7–36.

Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). The Sketch Engine. *Proceedings of the 11th EURALEX International Congress*, 105–116.

Leane, E. (2016). Reading popular physics: Disciplinary skirmishes and textual strategies. Routledge Taylor et Francis Group.

León-Araúz, P. (2022). Chapter 22. Terminology and equivalence. In P. Faber & M.-C. L'Homme (Eds.), *Terminology and Lexicography Research and Practice* (Vol. 23, pp. 477–502). John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.23.22leo

L'Homme, M.-C. (2004). *La terminologie: Principes et techniques*. Presses de l'Université de Montréal. https://doi.org/10.4000/books.pum.10693

Manfredi, M. (2019). Popular science discourse in translation: Translating 'hard', 'soft', medical sciences and technology for consumer and specialized magazines from English into Italian. Publicacions de la Universitat Jaume I.

Mitzkat, A., Berger, S., Reeves, S., & Mahler, C. (2016). More terminological clarity in the interprofessional field – a call for reflection on the use of terminologies, in both practice

and research, on a national and international level [Text/html]. *GMS Journal for Medical Education; 33(2):Doc36*. https://doi.org/10.3205/ZMA001035

Moirand, S. (2003). Communicative and Cognitive Dimensions of Discourse on Science in the French Mass Media. *Discourse Studies*, *5*(2), 175–206. https://doi.org/10.1177/1461445603005002003

Montalt, V., Zethsen, K. K., & Karwacka, W. (2018). Medical translation in the 21st century—Challenges and trends. *MonTI. Monografías de Traducción e Interpretación*, *0*(10), 27–42.

Myers, G. (2003). Discourse Studies of Scientific Popularization: Questioning the Boundaries. *Discourse Studies*, *5*(2), 265–279. https://doi.org/10.1177/1461445603005002006

Naming the coronavirus disease (COVID-19) and the virus that causes it. (n.d.). World Heath Organization (WHO). Retrieved 27 September 2023, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

Pecman, M., & Kübler, N. (2022). Chapter 12. Text genres and Terminology. In P. Faber & M.-C. L'Homme (Eds.), *Terminology and Lexicography Research and Practice* (Vol. 23, pp. 263–290). John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.23.12pec

Perrault, S. T. (2013). *Communicating Popular Science*. Palgrave Macmillan UK. https://doi.org/10.1057/9781137017581

Raffo, M. (2017). Translation and Popularization: Medical Research in the Communicative Continuum. *Meta*, *61*, 163–175. https://doi.org/10.7202/1038691ar

Sager, J. C. (1990). *A practical course in terminology processing*. J. Benjamins Pub. Co.

Sager, J. C. (1997). 1.2.1 Term Formation. In *Handbook of Terminology Management*. John Benjamins. https://www.jbe-platform.com/content/books/9789027285577-z.htm1.06sag

Salager, F. (1983). The Lexis of Fundamental Medical English: Classificatory Framework and Rhetorical Function (A Statistical Approach). *Reading in a Foreign Language 1:1*, *1*(1), 54–64.

Sketchengine. (n.d.). [Computer software].

Tekgül, D. (2019). Yeni Medyada Çeviri Aracılığıyla Popülerleştirme: Düşünbil Örneği. *Çeviribilim ve Uygulamaları Dergisi*, *25*, 71–86. https://doi.org/10.37599/ceviri.514001

Temmerman, R. (2000). Towards new ways of terminology description: The sociocognitive-approach. Benjamins.

Temmerman, R., & Van Campenhoudt, M. (2011). The dynamics of terms in specialized communication: An interdisciplinary perspective. *Terminology. International Journal of*

Theoretical and Applied Issues in Specialized Communication, *17*(1), 1–8. https://doi.org/10.1075/term.17.1.01tem

Watson, T., R. (2001). Building and using your own corpus and concordance. *ThaiTESOL Bulletin*, *14*(2). https://arts.kmutt.ac.th/call/doc/Building%20and%20Using%20Your%20Own%20Corpus%20and%20Concordance.pdf

World Health Organization (WHO) (1993). *The ICD-10 classification of mental and behavioural disorders*. World Health Organization.

World Health Organization (WHO) (2019/2021). *International Classification of Diseases*, Eleventh Revision (ICD-11), World Health Organization (WHO) 2019/2021 https://icd.who.int/browse11

World Health Organization (WHO). *Naming the coronavirus disease (COVID-19) and the virus that causes it.* (n.d.). Retrieved 27 September 2023, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it