

**UNIVERSITY OF TARTU  
DEPARTMENT OF ENGLISH STUDIES**

**BECOMING A NOUN: THE CASE OF CYBER**  
BA Thesis

ILMAR ÜLE  
SUPERVISOR: Assoc. Prof. JANE KLAVAN

University of Tartu  
2022

## ABSTRACT

With the vast development of the cyber realm, the word *cyber* is acquiring new uses as well. This BA thesis investigates the use of *cyber* – standardly an adjective – as a noun. The aim is to establish the frequency of using *cyber* as a noun, document the hypothesised trend towards nominalisation of the word *cyber*, and determine the types of texts driving the habitual use of *cyber* as a noun. This is important as the generation of imprecise umbrella terms (such as *cyber*, used as a noun, *crypto* in the sense of *cryptocurrency*, etc.) contributes to increasing vagueness of such terminology. The temporal scope of the analysis is twenty years (that is, 2000-2020) since the usage of the word *cyber* has markedly grown during that time period.

The thesis is divided into two substantive chapters, followed by a conclusion. The theoretical background chapter (Chapter 2) gives a concise overview of the word formation theories in English and contextualises the current study against the backdrop of the previous pertinent research. The empirical chapter (Chapter 3) introduces the research design and methodology of the study, along with the presentation of the main analytical results of the research. The empirical corpus for this research draws on Nexis Uni database, an extensive catalogue of print and online journals, transcripts of television and radio broadcasts, and other materials. The research procedure entailed obtaining a general count of the string *cyber* in each of the studied years and analysing one thousand documents per each year between 2000 and 2020. A simple Python script was written to use Natural Language Processing and add part-of-speech tags to the corpus, labelling words according to their function in the phrase and machine-processing the context that the search string was found in. The results were then cleaned up using regular expressions and text processing utilities, such as `grep`, `sed` and `tr`.

## Table of Contents

<b>ABSTRACT</b> .....	<b>2</b>
<b>1. INTRODUCTION</b> .....	<b>4</b>
<b>2. THEORETICAL BACKGROUND</b> .....	<b>7</b>
2.1. NOMINALISATION .....	7
2.2. RECOGNISING UNKNOWN NOUNS AUTOMATICALLY .....	10
<b>3. ANALYSIS</b> .....	<b>12</b>
3.1. COMPILING THE CORPUS .....	13
3.2. FINDING THE USES OF NOMINALISED <i>CYBER</i> .....	14
3.3. THE RESULTS .....	20
3.4. INTERPRETING THE RESULTS .....	26
<b>4. CONCLUSION</b> .....	<b>28</b>
<b>LIST OF REFERENCES</b> .....	<b>30</b>
<b>APPENDIX</b> .....	<b>33</b>

*Cyber is a fine platform for language inventiveness to leap from.*

Frank Devine, The Australian

## 1. Introduction

Writing in 2018, the long-time NATO spokesperson Jamie Shea maintained that “the more cyber becomes an inevitable and more decisive element of conflict, the more NATO Allies have an interest in preventing cyberspace from becoming a[.]...hostile domain” (Shea 2018: 40). Referring to *cyber*<sup>1</sup> as a domain and a capability, Shea was using the word as a standalone noun. Normally, we would expect the word to be an adjective, modifying nouns or verbs. Even the Oxford English Dictionary (OED) defines *cyber* as an adjective:

*cyber, adjective.*

Of, relating to, or involving (the culture of) computers, virtual reality, or the internet; futuristic.

Mr Shea is not alone in his use of *cyber*. There are many other examples where *cyber* is used as a noun. Perhaps most eye-catchingly, there was a much-maligned advertisement campaign conducted by the British government which encouraged ballet dancers and other artists to retrain in an undefined discipline called *cyber* (“Dying Swan” 2020; see Figure 8 in Appendix). The shift from adjective to noun is visible also in other languages, for example in French (see Figure 9 in Appendix, “Le cyber au Coeur de la PFUE”).

---

<sup>1</sup> From here onward, when referring to ‘cyber’ as my object of study, the word will be italicised.

The word *cyber* is a short form of *cybernetics*, itself a loan from Ancient Greek. In Ancient Greek *κυβερνήτης* meant “steersman”, derived from the verb *κυβερνάω*, “to steer”. The verb shares the root with *to govern*, which has been borrowed to English via Latin, Old French, and Anglo-Norman. It is also likely that the Ancient Greek word has been borrowed from another language that the seafaring Greeks came into contact with. This is supported by Hjalmar Frisk (1970: 38), who says in his *Griechisches Etymologisches Wörterbuch* that the word is probably of foreign origin.

In its modern English sense, the word was first used by the US mathematician and artificial intelligence pioneer Norbert Wiener, whose book *Cybernetics: Or Control and Communication in the Animal and the Machine* (1948) uses the term to refer to self-regulating mechanisms. This gave rise to various compound words with *cyber* as an adjectival modifier. *Cyberspace* is a term used for the virtual electronic world by the writer William Gibson, particularly in his *Neuromancer* (1984). Those familiar with the British science fiction series *Doctor Who* will know the menacing *cybermen*, first appearing in 1966.<sup>2</sup> From thereon, *cyber* is often used as an adjective modifying the following noun or verb, as in *cyber crime* or *cyber attack*. Alternative compound forms, such as *cyberattack* or *cyber-attack* are common.

To empirically substantiate the layperson observation about the habitual use of *cyber* as a noun, this BA thesis sets out to address the following research questions:

- 1) Apart from anecdotal evidence, how often is *cyber* used as a noun?
- 2) Is there a documentable trend towards nominalisation of *cyber*?
- 3) If there is such a trend, what types of texts are driving the change?

---

<sup>2</sup> The episode is partially viewable at <https://www.bbc.co.uk/doctorwho/classic/episodeguide/tenthplanet/> URL retrieved on April 16, 2022.

To answer these questions, the study proceeds in two steps. The ensuing section outlines the theoretical background of the research, providing a brief insight into word formation in English and proposing a way to recognise nouns with regular expressions, relying on their context. The empirical analysis follows in the second section. I provide the overview of the dataset and methodology, describe the procedures used to process and filter the data, and offer an interpretation of the results. In the conclusion, I reflect on some of the limitations of the study and offer ideas for further research.

This thesis shows that while *cyber* indeed occurs as a noun, such occurrence is nonetheless not very frequent. The trend of using *cyber* as a noun stems from the policy and military spheres and is not a feature in specialist technical writings.

## 2. Theoretical Background

### 2.1. Nominalisation

The English language has numerous ways of generating new words or re-assigning new functions to already existing ones (Marchand 1969; Clark and Clark 1979). The process of turning other types of words into nouns is known as nominalisation (Lees 1968). Nominalisations are “complex nouns that are derived from verbs, adjectives, and other nouns” (Lieber 2016: 3). Nominalisation can happen in multiple ways: it might, for example, entail transcategorisation of structures from verb to noun class, occurring either at the level of the word (e.g., *Sam’s washing of the windows*), pertaining to a structure that lies in between a verb and a full clause (*Sam’s washing the windows*) and consisting of full clauses (e.g., *that Sam washed the windows*) (Heyvaert 2003: 49). Nominalisation may be characterised by its morphological form or type (affixation, conversion) or semantics (Lieber 2016: 15-17).

English nominalisations have been broadly studied by both morphologists (Lieber and Baayen 1999; Melloni 2011; Fradin 2011) and syntacticians (Lees 1969; Chomsky 1970; Aronoff 1976; Borer 2013). Much of the literature on nominalisation accordingly focuses on the reclassification of verbs into nouns. Liesbet Heyvaert (2003), for example, zooms in on deverbal nominalisations, that is on nominal constructions which incorporate a reclassified verbal component. Drawing on Langacker’s seminal work (1991: 22) which emphasises the theoretical significance of nominalisations, she maintains that nominalisations are attention-worthy as constructions in their own right: “they are

reclassifications of non-nominal into *nominal* units or units which have adopted nominal functions both in their external and in their internal functioning” (Heyvaert 2003: 58).

This study shares such ethos. But as the current research’s theoretical interest lies in an adjective becoming a noun, we need to understand what specific mechanisms are at work in the English language to make this kind of conversion possible. Proceeding from the assumption that language is at once functional and cognitive,<sup>3</sup> this study shares the belief that nominalisations should not be regarded as mere “impoverished” clauses, but rather as constructions in their own right that encode ideational, interpersonal and textual meaning (Heyvaert 2003: 41-52). While departing from earlier emphasis on deverbal nominalisations, this thesis takes on board the preference for “a usage-based approach...which views nominalizations as ‘systematically motivated’ constructions which are structured with reference to the conventional units of language, and linked to them in terms of schematization and extension” (Heyvaert 2003: 42).

Zooming in on the conversion of adjectives to a nouns is part of a broader effort in linguistics known as studying derivation by conversion. This process is standardly placed at the crossroads of morphology, syntax, and lexical semantics (Valera 2015). In conversion research the notion “zero-derivation” designates two quite distinct approaches to conversion. For one approach, conversion is realised by adding a zero morpheme, that is a phonologically null suffix to the base. The gist of this line of thinking is captured in the synonymous terms “covert affixation”, “derivation by a zero-morpheme” and “zero-affixation” (Martsa 2013: 11). “Zero-derivation” or “zero-affixation” accordingly describes conversion on analogy with affixation (cf. Marchand 1969; Kastovsky 1980: 213-17, 230,

---

<sup>3</sup> In the sense that language is shaped by the functions it serves, with respect to the specific needs of the language user (Halliday 1970: 142), and it is also an integral part of human cognition (Langacker 1987).

1992: 291, 300; Adams 1973; Lipka 1990; Payne 1997). The other approach regards the process of conversion not as adding a zero-affix to the converting base, but as a lack of affixation on that base instead (Martsa 2013: 11). Quirk et al. (1985: 1558) treat conversion as a derivational process “whereby an item is adapted or converted to a new word class without the addition of a suffix”. For them, “zero derivation” reflects the notion of a “zero” suffix as analogous to actual suffixes. Yet another interpretation places conversion outside derivation, presenting it as “coinage” of a new word, or as lexical creation (Valera 2015; cf. Lieber 1992, 2004, 2005). The idea is that existing words can simply be entered under a new category – just as new words can be introduced to the language. Zero-derivational analyses are challenged by studies which Lieber (2005: 419) has described as “rebracketing or category-change with no concomitant affixation”. Accordingly, the noun *hammer*<sup>NOUN</sup> would change to the verb *hammer*<sup>VERB</sup> with no formal change involved. Arguably, simple analysis would be better placed compared to zero derivational analysis here since in the latter case it is impossible to determine whether the zero affix is a prefix or a suffix (Martsa 2013:33).

In the case of *cyber*, nominalisation does not change the form of the word. The adjective *cyber* looks and sounds exactly the same as its nominalised derivation. No affixes, that is additional morphemes, are added to the word stem to obtain the new meaning. Therefore, this is a zero-affixation process, also known as zero-derivation (Aronoff 1980). Zero-derivation leaves the internal processual status of the word intact but enables it to work as a nominal in larger configurations (Heyvaert 2003: 227). A good example of a noun to adjective zero-derivation would be the word *bridge*. It can denote both the physical object or the act of linking something together. The adjective *military* has gone through a

similar derivation process and can also be found as a group noun (with a *the*), or a regular noun, denoting a member of *the military* (*Oxford English Dictionary Online*).

Usually, nominalisation of adjective takes place when an adjective is used to denote a group, such as *the poor*, *the English*, etc. In this case, the noun is left out (ellipsis) (Günther 2018) and group adjective is acting as the head of a noun phrase (Quirk et al 1985: 421). According to the *Comprehensive Grammar of the English Language* by Quirk et al (1985), ellipsis is a significant mechanism in adjective to noun conversion. It can be explained in terms of a well-established adjective + noun phrase from which the noun has been omitted, or elided. Examples of that would be *daily (newspaper)*, *roast (beef)*, *regular (customer)* (Quirk et al 1985: 1561). However, we cannot always presume that adjective to noun zero-affixation conversion happens through the ellipsis of a well-established noun. In the case of *cyber*, it is not altogether clear what has been elided. It could have been *cyber warfare*, *cyber espionage*, *cyber crime*, etc. Ellipsis presumes that the removed component is somehow recoverable, that people know what goes unspoken. Assuming that nominalised *cyber* carries the same meaning each time it is used, it can be stated that we are not dealing with a classic case of ellipsis but something rather more complex, when multiple different *cyber*<sup>ADJECTIVE</sup> NOUN phrases produce a uniform *cyber*<sup>NOUN</sup> result and the elided word cannot always be guessed. A close and contextual reading would need to accompany a quantitative corpus analysis to grasp the patterns of using the adjective of *cyber* as a noun with more granularity.

## 2.2. Recognising Unknown Nouns Automatically

In order to find out whether *cyber* is being used as a noun or an adjective, we should be able to recognise nouns. As I am going to use a text corpus and pattern matching for

this task, I should be able to describe a noun in regular expressions so that it can be used in automated text processing. As it is not realistic to expect a computer to actually understand a word, I need to be able to describe it according to its function and place in a sentence.

In English grammar, a **head** is the key word that determines the nature of a phrase. This contrasts to modifiers and determiners. For example, in a **noun phrase**, the head is a noun or a pronoun (*cyber warfare*). In an **adjective phrase**, the head is an adjective (remarkably *inadequate*). For a human reader, this is easy to spot. However, to analyse a large number of texts, we need to automate the process. In order to recognise whether *cyber* is a noun or an adjective, we should be able to determine its context and thus deduce the word's function in a phrase.

### 3. Analysis

In computational linguistics an  $n$ -gram is a contiguous sequence of  $n$  items in a given sample of text or speech. The items can be phonemes, syllables, letters, words or word pairs – any string or combination of symbols that yields itself to documentation and analysis.  $N$ -grams are useful for natural language processing, text mining, language modelling, etc. Since we only have one search term – *cyber* – we can call it a *unigram*.

To get a broader overview of the representation of *cyber* in texts, the Google Books Ngram Viewer will be used.<sup>4</sup> This is a free service that searches the full text of books and magazines that Google has scanned, converted to text using optical character recognition (OCR), and stored in its digital database. Books are provided either by publishers and authors through the Google Books Partner Program, or by Google's library partners. A search term only appears in the chart if there are at least 40 hits in the corpus at a certain point of time. Our search term is the character string *cyber*, the time span is 1948 (the publication of Norbert Wiener's book) to 2019 (the most recent year available).

In Figure 1, we can see the  $n$ -gram that Google generated. It tells us that *cyber* was used very little until mid-1990s. Mid-nineties, the increasing ubiquity of computers, and the advent of the online culture is the time when the use of *cyber* is starting to go through a steady growth. The  $n$ -gram shows a slight flattening of the growth curve as we get into 2019. Unfortunately, the data does not cover the last 2,5 years and we cannot see yet if we are past peak-*cyber* or if the word's usage frequency will grow still higher.

---

<sup>4</sup> <https://books.google.com/ngrams>

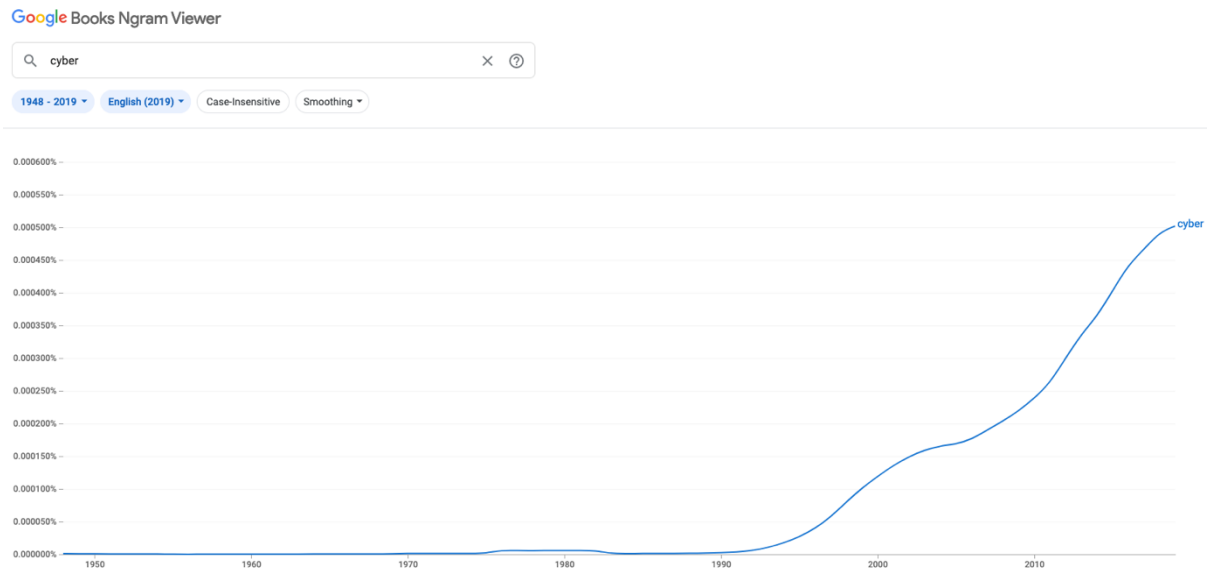


Figure 1. Google Ngram chart for the string "cyber", years 1948-2019

The numbers on the vertical axis of the graphic show us how big a percentage the unigram *cyber* forms of all unigrams in the Google Books dataset. This is not of special interest for this study, but predictably, the percentage is very small.

### 3.1. Compiling the Corpus

In order to conduct an analysis on if and how often *cyber* is used as a standalone noun, I set out to analyse a large number of occurrences of the word throughout a timeframe of twenty years (2000 – 2020) and attempt to find out if there were any signs of it being used as a noun. I chose this timeframe because we can see from Figure 1 that during that time period the word *cyber* went through a marked growth in usage. Accordingly, the research assumption is that if there is an ongoing nominalisation process, it can be best documented by observing the two first decades of the current century.

For the research, I needed a corpus spanning that exact timeframe. Looking at what was available, I decided against corpora like Contemporary Corpus of American English

(COCA) or News on the Web (NOW) simply because their standard student license only allows to conduct a limited number of searches. In order to get a meaningful overview of a change in the language, I needed a big set of data, fully searchable and manageable with automatic text analysis tools, capable of using regular expressions.

I decided to use Nexis Uni,<sup>5</sup> simply because it was available to me. This is a commercially-run database, containing an extensive catalogue of print and online journals, transcripts of television and radio broadcasts, syndicated news, and blogs. All put together, it covers more than 17,000 sources of information. It also contains the US Supreme Court decisions since 1790. It must be borne in mind that Nexis Uni is not a corpus but rather a full document search engine with its own database. As such, it can be used to build a suitable corpus for this research.

### 3.2. Finding the Uses of Nominalised *Cyber*

The research procedure was the following:

1. Use Nexis Uni to obtain a general count of the string *cyber* in each of the years between 2000 and 2020.
2. Download an equal and large number of documents all containing *cyber* from each studied year. I settled on one thousand documents per year, for the period of 2000–2020. This formed the corpus of texts used in this study.
3. Use regular expressions and text editing utilities to capture only the lines containing the actual search string.

---

<sup>5</sup> <https://www.lexisnexis.com/en-us/professional/academic/nexis-uni.page>

4. Use Natural Language Processing to add part-of-speech (POS) tags to the words in the corpus.
5. Weed out irrelevant POS combinations and known phrases where *cyber* acts as an adjective.
6. Go through the results manually, applying additional clean-up rules with regular expressions, refining the results further.
7. Analyse the results.

I searched Nexis Uni for the string *cyber* and downloaded the first 1,000 matching documents from each year, starting from 2000 and ending with 2020. I left all search settings to default, except for language, which I set to English.

The screenshot shows the Nexis Uni search interface. At the top, the search bar contains the term 'cyber'. The left sidebar displays a 'Snapshot' of search results across various categories, with 'News' having the highest count at 10,000+. Below this, there are filters for language (English) and date range (Jan 01, 2020 to Dec 31, 2020). A 'Search within results' section is also visible. The main content area shows a list of search results, with the first three results highlighted. Each result includes a title, a snippet of text, and a 'Preview' button.

Figure 2. Nexis Uni sample search, covering cyber in 2020

Predictably, I ended up with 20,000 documents containing the string *cyber*, grouped by the year. In total, there are 5,793,419 words in the collected dataset. This collection of documents, obtained from Nexis Uni search results, forms the corpus of this study. It must also be said that that Nexis Uni sorts the search result by their relevance, which is highly likely calculated by the frequency with which the search string appears in a text. This is different from a traditional, balanced corpus structure. It also means that the documents are not of equal length, because it is not always the longest documents that contain the highest count of *cyber*. It must also be said that Nexis Uni is a privately managed and commercially available database, which means that the search algorithms are not fully

transparent to the user and may plausibly contain additional logic for improving customer experience.

I then saved the results in plain text files, one file per year, and then added line numbers to retain the original context of the later search results. For refining the results further, I used three simple but powerful text processing utilities:

1. Grep. This is a command-line utility for searching plain-text datasets for lines that match a regular expression. It can be used for inverted searches, finding lines that do NOT contain the search term. This is useful for eliminating unwanted data.
2. Sed, or "stream editor" is a utility that parses and transforms text. It is a line-oriented text processing tool. It takes its input from a data stream or file, applies user-determined transformation rules, and gives back a data stream as an output. This stream can then be directed to another tool for further processing, printed on screen, or into a file.
3. Tr, "translate" or "transliterate" is a tool for replacing a character, a group of characters, or a class of characters with other symbols. It can also be used for character deletion, collapsing whitespaces, etc.

All these tools are native to Unix-like operating systems. Among computer programs, they are remarkably old, with sed tracing its ancestry back to 1965. They are also very well documented, maintained, and free to use.

Using grep, I removed all lines not containing the string *cyber*. I then used grep, sed, and tr to remove compound words containing *cyber*, numerals, and change all text to lowercase. The reason for removing all compound words containing the search string was

that the existence of cyber-compound nouns is evident since at least 1966 and they thus fall out of scope of this research.

Next, I wrote a simple Python script that looped through all words in the text and POS-tagged them. I used the spaCy Python package to achieve this. spaCy<sup>6</sup> is a free and open-source library for advanced Natural Language Processing (NLP) in Python. There are other comparable toolkits available, such as the Natural Language Toolkit (NLTK)<sup>7</sup> package for Python. However, spaCy required fewer dependencies and was thus easier to install and less complex to start working with.

Although spaCy is a remarkably powerful tool, my use of it was very basic. My script only added part-of-speech tags to the corpus, labelling words according to their function in the language and allowing me to continue machine-processing the context that the search string was found in.

```
1 import spacy
2 import sys
3 #Loading English pipeline. Includes tokeniser, tagger, parser, sender, ner, attribute_ruler, lemmatiser.
4 nlp = spacy.load("en_core_web_sm")
5 #changing output to file instead of screen
6 original_stdout = sys.stdout
7 with open("input.txt", "r") as a_file:
8     with open("output.txt", 'a') as f:
9         sys.stdout = f
10 #Looping through words
11     for line in a_file:
12         stripped_line = line.strip()
13         doc = nlp(stripped_line)
14 # printing the original word + the POS tag
15         for token in doc:
16             print(token, token.pos_)
17
18
```

Figure 3. Simple POS tagger with Python and spaCy

Once the corpus had been POS-tagged, I noticed that the word *cyber* was unfamiliar to spaCy and was often tagged incorrectly. This is not unusual for untrained tagging engines and can be improved with some work.

---

<sup>6</sup> <https://spacy.io>

<sup>7</sup> <https://www.nltk.org/>

Here is an example of a correctly tagged instance of *cyber*:

she PRON said VERB people NOUN needed VERB to PART realise VERB that SCONJ  
cyber ADJ bullying NOUN could AUX have VERB unforeseen ADJ consequences  
NOUN

And here is an example of *cyber* getting miscategorised:

cyber NOUN conference NOUN allows VERB participants NOUN to PART join VERB  
the DET meeting NOUN at ADP their PRON convenience NOUN

From these two examples we can see that it is relatively safe to ignore how *cyber* is tagged, as long as the words providing it with a context are tagged correctly. Encouragingly, the words surrounding *cyber* were almost always familiar to the tagger and usually received correct tags. I refined the results by removing obviously irrelevant word pairs where *cyber* cannot be the head of a noun phrase, such as *cyber*+NOUN (e.g., *cyber security*) and *cyber*+VERB (e.g., *cyber bullying*, *cyber shopping*). I also extended the pattern to include longer phrases such as *cyber and electronic warfare* (*cyber+and/or+ANY\_WORD+NOUN/VERB*). I verified all matched patterns before removing them, avoiding “blind” data removal. This is easily achievable with grep pattern matching, using combinations of search words and POS tags as patterns.

I then went through the results line-by-line, cleaning up false positives with pattern-matched regular expressions, so that I would have to remove such phrases only once. In all the steps of text processing, I kept the line numbers intact, so that each corpus entry could be traced back to its original context.

### 3.3. The Results

First, I compiled a simple table with the search result. In the first column of Table 1 we can see how many times the search string *cyber* occurred in the Nexis Uni database in a particular year. The results broadly correspond to the Google Ngram (Figure 1), indicating a slight decrease in the frequency *cyber* from 2018 onwards.

The second column shows us how many times *cyber* occurred in our corpus in a particular year. The third column contains the identified instances where *cyber* serves the function of a noun in the corpus, and the fourth column gives the percentage of nominalised *cyber* of all occurrences of the search string in the corpus, in a particular year.

Year	Total <i>cyber</i> in Nexis Uni	Total <i>cyber</i> in corpus	<i>Cyber</i> as noun in corpus	Percentage of nominalised <i>cyber</i> among all occurrences of <i>cyber</i> in the corpus
2000	25477	4388	15	0,34%
2001	18060	5535	26	0,47%
2002	14373	4605	22	0,48%
2003	13916	4725	26	0,55%
2004	14195	5809	26	0,45%
2005	22209	6941	24	0,35%
2006	27523	7785	26	0,33%
2007	36324	8011	23	0,29%
2008	31296	10151	58	0,57%
2009	38393	14901	82	0,55%
2010	55608	16031	123	0,77%
2011	70332	16303	163	1,00%
2012	82809	18159	269	1,48%
2013	127527	17996	449	2,49%
2014	142698	15848	157	0,99%
2015	191214	16417	295	1,80%
2016	220358	17393	180	1,03%
2017	279301	17625	222	1,26%
2018	286081	15995	104	0,65%
2019	255543	14690	122	0,83%
2020	251023	12606	29	0,23%

Table 1. Occurrences of *cyber*

As we can see from Table 1, the use of the word *cyber* in all its forms grew tenfold in the years 2000 – 2020. 2018 was the year of peak-*cyber*, with 286,081 matches in Nexis Uni (Figure 4).

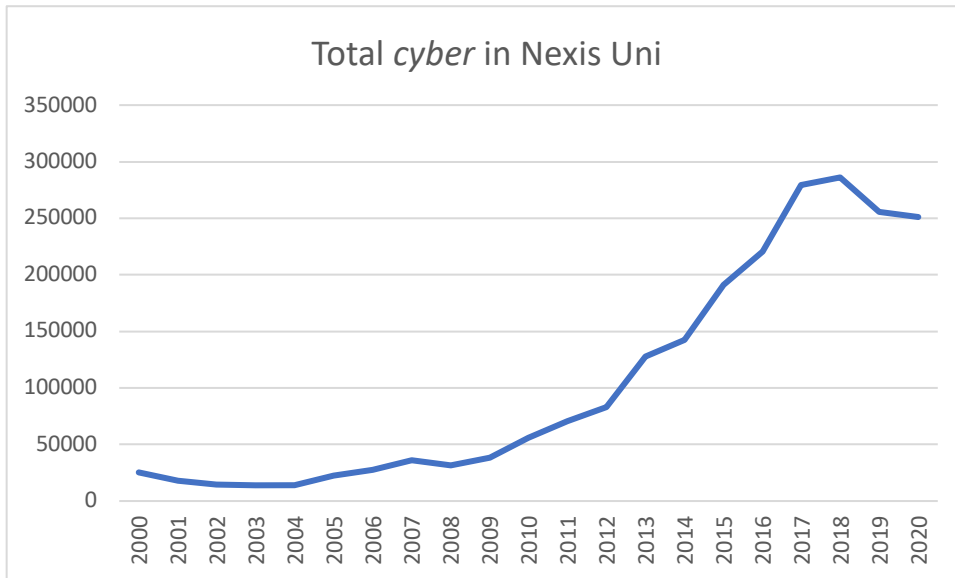


Figure 4. Total cyber in Nexis Uni

As the corpus used in this research is derived from Nexis Uni database through a search engine, the documents that made it into the corpus are the ones that have the most frequent use of the string *cyber*. In Figure 5, we can see that 2012 and 2013 are the years where the most *cyber*-heavy documents occurred, making it into the corpus and hinting at intense discussions on the cyber-topics.

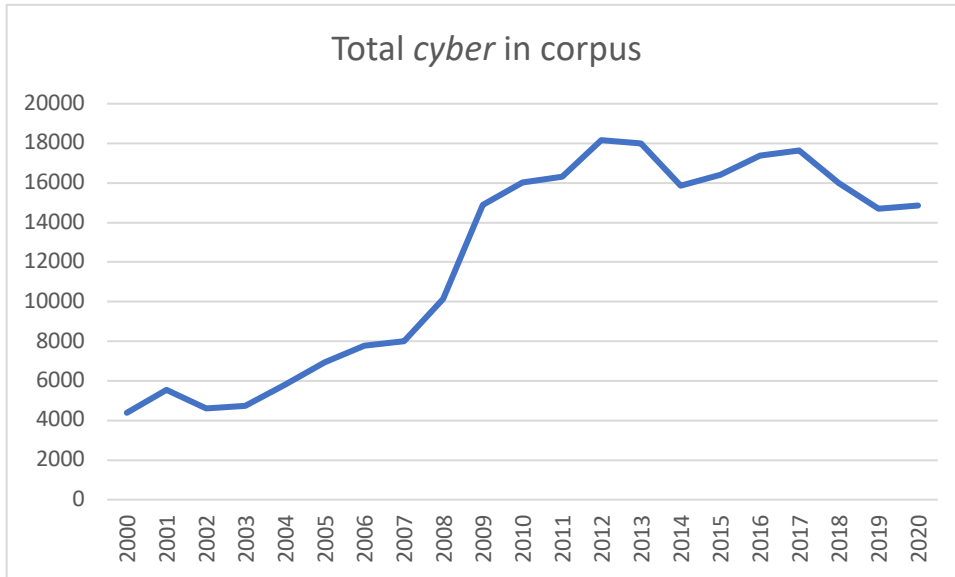


Figure 5. Total cyber in corpus

The word *cyber* was used as a noun in each covered year. The year 2000 yielded only 10 uses of nominalised *cyber* in the corpus. In Figure 6 we can see that from thereon, the use of nominalised *cyber* grows every year, peaking at 449 in 2013 and forming 2.49% of all uses of *cyber* in the corpus. After 2013, the number of occurrences started to decrease, falling to 29 in 2020.

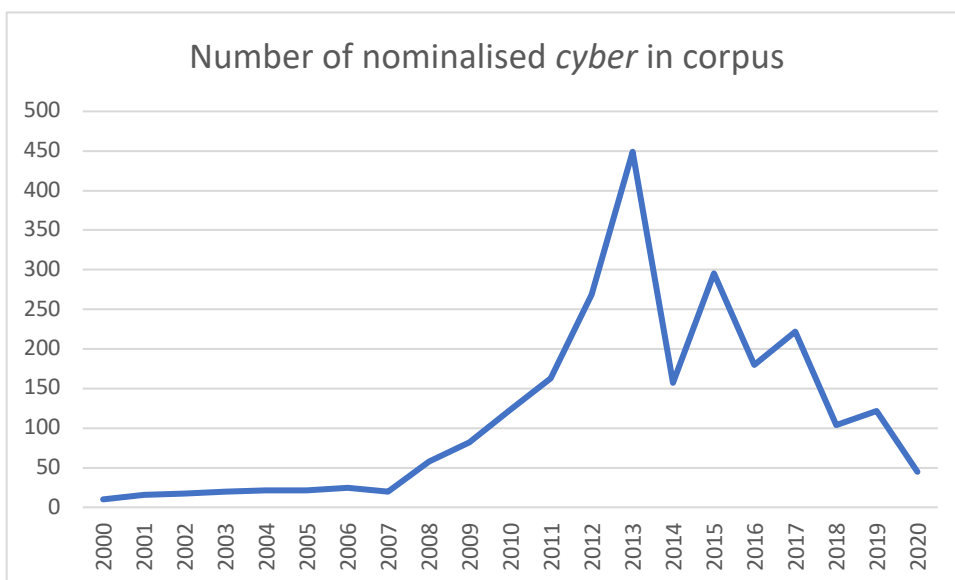


Figure 6. The number of nominalised cyber in the corpus

Finally, I observe what percentage nominalised *cyber* formed of the total number of *cyber* in the corpus. Figure 7 shows us that in 2013, 2.49% of all occurrence of *cyber* in the corpus were nouns. The percentage has generally been declining since then, reaching a low point of 0.3 % in 2020.

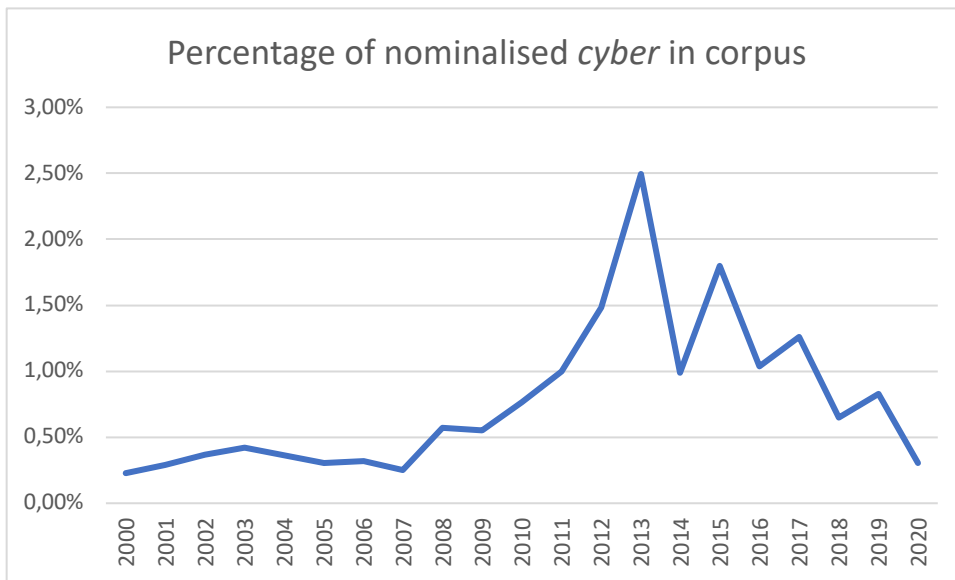


Figure 7. Percentage of nominalised *cyber* in corpus

In figures 5, 6, and 7 we can see a notable increase in the use of *cyber* in 2007, in all its forms. As we remember, 2007 was the year when the first state-on-state cyber attacks took place, targeting Estonia (Ottis 2008), so it is plausible that this gave rise to chatter about *cyber* in general. However, it is equally plausible that the “notch” in the three figures is just representing a naturally increasing use of a term describing a growingly ubiquitous phenomenon.

Next, I set out to get a general overview of which types of texts tend to use *cyber* as a noun. I did that manually, referring back to the original unedited corpus entries. Below, I shall highlight some examples of nominalised *cyber* from years 2000-2020. For easier machine processing and pattern matching, capital letters have been turned into lowercase

and there are extra whitespaces around punctuation characters. Also, *cyber* has been capitalised for easier visual processing.

2000: CYBER is a fine platform for language inventiveness to leap from .

2001: one of the principal problems that we have identified in our report , not just in CYBER but everywhere , is the inability to share information across and up and down lines .

2002: a cartoon spider called CYBER will be helping pupils learn about the national assembly .

2003: a perth academic stalked by CYBER for the past five years expects to be awarded australian record damages this week in an internet defamation case .

2004: [...] to appreciate CYBER as a vector of threat .

2005: " [...] the future is schools that are bricks and CYBER, " he said .

2006: " e - commerce is now the vehicle for delivering a wealth of private sector and government services , " mccarthy said . " but CYBER is now also the vehicle of choice for the bad guys to deliver and organize their services .

2007: " with CYBER , in some cases , we 're going to have to turn things in very short timeframes , often in a matter of days . "

2008: last year , the air force defined CYBER as everything inside the electromagnetic spectrum .

2009: general kehler is accountable to develop a combat capability in the domain of space , and now in the domain of CYBER .

2010: CYBER is also attractive to our adversaries because it is hard to identify the origin of an attack .

2011: we came to realize that nuclear weapons could destroy the world , and CYBER let loose has the same capability to destroy our societies if it is not controlled .

2012: [...] to stop their theft of other kinds of intellectual property through the use of CYBER .

2013: however , CYBER still remains the one area where if there 's a dead body on the ground , there 's no police you call who will run to you and do the forensics and all that .

2014: and i just want to be sure i understand what sort of legal authorization you were talking about when it comes to CYBER because that is in our bailiwick .

2015: [...] iran used CYBER to attack u.s . military networks in december , that would be in the middle of the iranian nuclear negotiations .

2016: we certainly acknowledge that states engage in the use of CYBER as a tool to gain access and knowledge .

2017: mccain : that obviously assumes that the department of homeland security has the capabilities and the authority in order to carry out that requirement , whereas this -- CYBER is warfare . CYBER is warfare . CYBER is an attempt to destroy a democracy . that 's what mr . putin is all about .

2018: to insure CYBER or not to insure CYBER : cautions and caveats for carriers

2019: indeed , krusko added it 's difficult for underwriters to wrap their heads around CYBER when looking at individual risk situations .

2020: " in CYBER ... israel is a global powerhouse in absolute numbers , " said matania .

Taken together, these examples help to qualitatively illustrate a characteristic pattern in the corpus: whereas the early nominalisation of *cyber* is often deliberate by focusing on the word itself, there is a notable uptick associated with the definition of *cyber* as a military domain in the US context from 2008 onwards.

### 3.4. Interpreting the Results

Notably, nominalised *cyber* often occurs in transcripts of spoken text, such as political hearings (US Representative Jim Saxton 2003) or interviews. Nominalised *cyber* is rarely used in periodicals that have gone through an editorial process. The only remarkable cases when this is happening is when the text is discussing the word *cyber* itself, such as the example sentence from the year 2000. Hence, we can draw a conclusion that *cyber* is used as a noun in more spontaneous, verbal discussions and is making its way from there to printed material, such as the examples on Figures 8 and 9 in the Appendix.

A major increase in the use of nominalised *cyber* started in 2008, when the US Air Force introduced the idea of a Cyber Command. This gave rise to a lot of discussion over *cyber* as a military domain, as a new field of military, criminal, and law enforcement activities. Most of these discussions are held by policymakers and military leaders to whom the concept of cyberspace might be a new one and who embrace *cyber* with a high degree of linguistic liberty, deviating from its standard use as an adjective.

The year 2008 also provided us with another explanation of *cyber*, defining it as “everything inside the electromagnetic spectrum” (Lord: Future of Air Force 2008). As a side note, this definition implies that *cyber* now covers some of the aspects of traditional electronic warfare and signals intelligence, making it an umbrella term for radio communications, radar, etc. However, this proposed definition of *cyber* has not taken root in practice.

The US Cyber Command was established in 2010.<sup>8</sup> With that, cyberspace was defined as the fifth military domain, alongside with land, sea, air, space, and information

---

<sup>8</sup> <https://www.cybercom.mil/About/History/>

domains. This can also be seen in the corpus data. Corresponding to that timeframe, there is a definite increase in the number of occurrences of *cyber*, both as an adjective and a noun. Sentences such as “*Although Obama's new review put overall control and coordination of cyber at the White House, it reportedly does not get into the NSA debate*” (Baldor 2009) illustrate the surging use of *cyber* as a noun.

Moving closer to today, we see a slight decrease in the overall frequency of *cyber* in the corpus. Examining the corpus entries from 2019 and 2020, we see that nominalised *cyber* is almost exclusively used in defence, security, or intelligence context. We have phrases such as “*cyber is the ultimate team sport*” (FBI Director Wray 2020) or “*offensive cyber can do a range of things*” (Malnick 2021) or “*we need to firm up our cyber*” (House Oversight and Reform Committee 2021).

The phrase *offensive cyber*, brought forth in the previous chapter is a good illustration of nominalisation of *cyber*. This is a typical noun phrase, where *cyber* serves as the head, taking *offensive* as a modifier.

Being a cyber defence professional, I noticed that my search for nominalised *cyber* did not yield any technical texts. There are no write-ups of actual cyber attacks, no analyses of cyber threat actors, no practical recommendations for beefing up cyber security. These texts are abundant in Nexis Uni and are also represented in our corpus. Yet they yielded no *cyber* as a noun. Observing the search results and taking into account the conspicuous absence of technical texts, we can conclude that the drive towards nominalisation comes from policymakers, treating *cyber* as a field of operation, as a military domain, or indeed, as everything in the electromagnetic spectrum. The texts that describe bits and bytes, cyber attacks, cyber threat actors, and actual attack techniques (e.g., Bitdefender Achieves

Highest Detection Score 2021) rarely use *cyber* as a noun. In fact, we see no such cases in our corpus.

#### 4. Conclusion

This thesis set out to investigate how often *cyber* is used as a noun; whether there is an observable pattern indicating nominalisation of *cyber*, and if indeed, what types of texts are driving such a shift. The analysis conducted for this study allows to conclude that *cyber* does indeed occur as a noun. However, that does not happen very often, with only 2.49% of the total occurrences serving the function of a noun in our corpus in the year of 2013, when the use of nominalised *cyber* peaked.

We can conclude that a drive towards nominalised *cyber* comes from the policy and military fields. Such a use is not detectable in technical writing. The corpus showcases the nominalisation tendency in transcripts of spoken discussion, such as the US congressional hearings, press conferences, and interviews being at the forefront of the observed changes.

We can also see that the use of *cyber* as a noun increased after the creation of the US Cyber Command and the declaration of the fifth operational domain. There is a slight upwards bump after the 2007 cyber attacks against Estonia.

The more recent decrease in the use of nominalised *cyber* can possibly be explained by the fact that discussion about the fifth military domain has died down and the novelty of the realm has worn off. Nominalised *cyber* has found its niche and field of use and is a mainstay in political and military discussions, occasionally making guest appearances in government-sponsored advertisement campaigns or commercial texts.

As any study, this research has also possible limitations. It is conceivable that the current corpus is somewhat influenced by the search algorithms that were used to compile

it. As already mentioned, if we search for the string *cyber* in Nexis Uni to compile our own corpus from the search results, we are served documents that are the best match, that is the ones that have the highest count of the search word. This means that the corpus is more likely to contain certain types of texts, such as transcriptions of conversations and discussions where people discuss a certain topic, repeat the relevant words often and mirror each other's sentences. Accordingly, some potentially interesting use cases of *cyber* perhaps never made it into our corpus because their relatively low count of the matching search term. It would be interesting to repeat the research on a sizeable, balanced and curated corpus of modern English, such as the COCA in its entirety, in order to see if the results would vary.

To facilitate further research on *cyber*, it would also be very useful to train the spaCy Python library with the results of this study. Taking our samples of nominalised *cyber*, tagging them correctly, and feeding them back to spaCy would help the algorithm to automatically recognise nominalised *cyber*, significantly streamlining any further research on the topic.

## List of References

- Adams, Valerie. 1973. *An Introduction to Modern English Word-Formation*. Harlow: Longman.
- Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.
- Aronoff, Mark. 1980. Contextuals. *Language*, 56: 4, 744-758.
- Baldor, Lolita C. 2009. Obama to create cyber czar in awareness effort. *The Associated Press*, May 28.
- Bitdefender Achieves Highest Detection Score in MITRE ATT&CK® Evaluations for Enterprise. 2021. PR Newswire, April 22.
- Borer, Hagit. 2013. *Taking Form*. Oxford: Oxford University Press.
- Chomsky, Noam. 1970. Remarks on nominalization. In R. Jacobs & P. Rosenbaum (eds.) *Readings in English Transformational Grammar*, 184–221. Waltham, MA: Ginn.
- Clark, Eve V. and Herbert H. Clark. 1979. When Nouns Surface as Verbs. *Language*, 55: 4, 767-811.
- Devine, Frank. 2000. Cyber's long journey into currency. *The Australian*, January 15.
- Dying swan or lame duck? Why 'Fatima' the ballerina's next job was tripping up the government. 2020. *Guardian*, October 13. Available at <https://www.theguardian.com/stage/2020/oct/13/dying-swan-or-lame-duck-why-fatima-the-ballerinas-next-job-was-tripping-up-the-government>, accessed April 18, 2022.
- FBI Director Wray Issues Remarks at Boston Conference on Cyber Security. 2020. Targeted News Service, March 6.
- Fradin, Bernard. 2011. Remarks on state denoting nominalizations. *Recherches linguistiques de Vincennes*, 40, 73–99.
- Frisk, Hjalmar. 1970. *Griechisches Etymologisches Wörterbuch*. Vol. II. Heidelberg: Heidelberg University Press.
- Gibson, William. 1984. *Neuromancer*. New York: Ace Books.
- Google Ngram Viewer. Available at <https://books.google.com/ngrams>
- Günther, Christine. 2018. The rich, the poor, the obvious. *Linguistik Aktuell/Linguistics Today*, 246, 77-112.
- Halliday, Michael A.K. 1970. Language structure and language function. In John Lyons (ed.) *New Horizons in Linguistics*, 140/165. Harmondsworth: Penguin Books.
- Heyvaert, Liesbet. 2003. *A Cognitive-Functional Approach to Nominalization in English*. Berlin and New York: Mouton de Gruyter.
- House Oversight and Reform Committee Holds Hearing on Disrupting Hackers and Building Resilience to Cyber Threat. 2021. CQ Transcriptions, November 16.
- Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Kastovsky, Dieter. 1980. Zero in morphology: a means of making up for phonological losses? In Jacek Fisiak (ed.) *Historical Morphology*, 213-250. The Hague: Mouton.
- Kastovsky, Dieter. 1992. The formats change – the problems remain: word-formation theory between 1960 and 1990. In Martin Pütz (ed.) *Thirty Years of Linguistic Evolution: Studies in Honour of René Dirven on the Occasion of his Sixtieth Birthday*, 28-310. Amsterdam: Benjamins.

- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar. Descriptive Application*. Stanford: Stanford University Press.
- Lees, Robert B. 1968. *The grammar of English nominalizations*. Bloomington: Indiana University Press.
- Lieber, Rochelle. 1992. *Deconstructing Morphology*. Chicago: University of Chicago Press.
- Lieber, Rochelle. 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.
- Lieber, Rochelle. 2005. Word formation processes in English. In Pavol Štekauer and Rochelle Lieber (eds.) *Handbook of Word Formation*, 377-422. Dordrecht: Springer.
- Lieber, Rochelle. 2016. *English Nouns: The Ecology of Nominalization*. Cambridge: Cambridge University Press.
- Lieber, Rochelle and Harald Baayen. 1999. Nominalizations in a calculus of lexical semantic representations. In Geert Booij & Jaap van Marle (eds.) *Yearbook of Morphology 1998*, 175–198. Dordrecht: Kluwer Academic Publishers.
- Lipka, L. 1990. *An Outline of English Lexicology*. Tübingen: Niemeyer.
- Lord: Future of Air Force Cyber Command May be Decided within 45 Days. 2008. *Inside the Air Force*, September 12.
- Malnick, Edward. 2021. Britain to carry out offensive cyber attacks; Defence Secretary unveils cyber plan in red wall area. *The Sunday Telegraph*, October 3.
- Marchand, Hans. 1969. *The categories and types of present-day English word-formation: a synchronic-diachronic approach*. München: C.H. Beck.
- Martsa, Sándor. 2013. *Conversion in English: A Cognitive Semantic Approach*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Melloni, Chiara. 2011. *Event and Result Nominals: A Morpho-semantic Approach*. Bern: Peter Lang.
- Nexis Uni. <https://www.lexisnexis.com/en-us/professional/academic/nexis-uni.page>.
- Ottis, Rain. 2008. Analysis of the 2007 Cyber Attacks Against Estonia from the Information Warfare Perspective. Tallinn: Cooperative Cyber Defence Centre of Excellence. Available at [https://ccdcoe.org/uploads/2018/10/Ottis2008\\_AnalysisOf2007FromTheInformationWarfarePerspective.pdf](https://ccdcoe.org/uploads/2018/10/Ottis2008_AnalysisOf2007FromTheInformationWarfarePerspective.pdf), accessed April 18, 2022.
- Oxford English Dictionary Online*.
- Payne, Thomas E. 1997. *Describing Morphosyntax: A guide for Field Linguists*. Cambridge: Cambridge University Press.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. 2<sup>nd</sup> ed. London and New York: Longman.
- Shea, Jamie. 2018. NATO's Role in Baltic Sea Security: Reestablishing Deterrence, Projecting Stability. In Ann-Sofie Dahl (ed.) *Strategic Challenges in the Baltic Sea Region: Russia, Deterrence, and Reassurance*, 33-48. Washington, D.C.: Georgetown University Press.
- U.S. Representative Jim Saxton (R-NJ) Holds Hearing on Cyber-Terror Threats. 2003. *FDCH Political Transcripts*, July 24.
- Valera, Salvador. 2015. Conversion. In Rochelle Lieber and Pavol Štekauer (eds.) *The Oxford Handbook of Derivational Morphology*. Oxford Handbooks Online, DOI: 10.1093/oxfordhb/9780199641642.013.0010

Wiener, Norbert. 1948. *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.

## Appendix

Examples of nominalised *cyber* from outside the corpus.



Figure 8. UK Government campaign for retraining artists



Figure 9. "Le cyber" appears as a masculine noun in a presentation introducing the French presidency of the European Council

## RESÜMEE

TARTU ÜLIKOOL  
ANGLISTIKA OSAKOND

Ilmar Üle  
Becoming a Noun: the Case of Cyber  
Nimisõnaks saamine: *cyber*-i juhtum  
Bakalaureusetöö  
2022  
35 lehekülge

Kübermaailma tormiline areng tähendab, et ka sõna *küber* leiab uusi tähendusi. Käesolev bakalaureusetöö uurib ingliskeelse omadussõna *cyber* kasutust nimisõnana. Töö eesmärk on välja selgitada, kui tihti *cyber* esineb nimisõnana, dokumenteerida selle sõna hüpoteetilist nimisõnastumissuunitlust ning selgitada välja, mis tüüpi tekstid sellist võimalikku muutust juhivad. Selline uurimus on oluline, kuna viimasel ajal on esile kerkinud palju ebatäpseid termineid (*krüpto* krüptovaluuta kohta või *küber* iseseisva nimisõnana), mis kasutavad juba olemasolevaid ja välja kujunenud sõnu uues, segasemas tähenduses. Uurimistöö keskendub aastatele 2000-2020, sest selle aja jooksul on sõna *cyber* kasutus märkimisväärselt kasvanud.

Bakalaureusetöös on kaks sisulist peatükki. Peatükk 2 annab ülevaate teoreetilisest taustast, käsitledes sõnaloometeooriaid inglise keeles ning asetab käesoleva uurimuse varasemate tööde konteksti. Peatükk 3 annab ülevaate meetodikast ning tutvustab uurimuse tulemusi. Uurimistöö kasutab korpuslingvistilisi meetodeid. Töö korpus on koostatud Nexis Uni andmebaasist ning koosneb kahekümnest tuhandest dokumendist, mis sisaldavad otsilõime *cyber*. Korpus märgendati Pythoni programmeerimiskeele spaCy paketi abil ning puhastati regulaaravaldistega.

Uurimus leidis, et *cyber* leiab tõepoolest kasutust ka nimisõnana, kuid mitte väga tihti. Kõige tihedamalt kasutati *cyber*-it nimisõnana aastal 2013. Põhiliselt leiab nimisõnastatud *cyber*-it verbaalsete tekstide üleskirjutustest. Sellised tekstid esindavad põhiliselt poliitika- ja militaarvaldkonda ning nende esinemissagedus on kõige suurem Ameerika Ühendriikide küberväejuhatuse loomise järgsel perioodil. Korpuses leidunud tehnilised tekstid ei kasutanud *cyber*-it nimisõnana, mistõttu võib järeldada, et nimisõnastumistrend tuleb poliitika-, haldus- ja militaarvaldkonnast. On võimalik, et nendes valdkondades on tegemist mingil määral uudissõnaga ning seetõttu käiakse sellega ümber mõnevõrra liberaalsemalt.

### Märksõnad:

Inglise keel, keeleteadus, uudissõnad, korpuslingvistika, arvutilingvistika, nimisõnastumine, nimisõna, omadussõna.

## **Lihtlitsents**

**Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks.**

Mina, Ilmar Üle

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Becoming a Noun: the Case of Cyber,

mille juhendaja on Jane Klavan,

1.1 reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2 üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. Olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Digiallkirjastatud

Ilmar Üle

Grimbergenis, 24.05.2022

## **Autorsuse kinnitus**

Autorsuse kinnitus

Kinnitan, et olen koostanud käesoleva bakalaureusetöö ise ning toonud korrektselt välja teiste autorite panuse. Töö on koostatud lähtudes Tartu Ülikooli maailma keelte ja kultuuride kolledži anglistika osakonna bakalaureusetöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

Digiallkirjastatud

Ilmar Üle

Grimbergenis, 24.05.2022

**Lõputöö on lubatud kaitsmisele.**

Digiallkirjastatud

Jane Klavan

Tartus, 24.05.2022