

Is e- the new cyber?

A corpus study on fashion cycles in vocabulary

Johan Nylin

Department of English

Bachelor/Magister Degree Project
English Linguistics

Autumn 2013

Supervisor: Frank Bramlett



Stockholms
universitet

Is e- the new cyber?

A corpus study on fashion cycles in vocabulary

Johan Nylin

Abstract

A central area of research in linguistics is the study of changes in vocabulary over time, be it over historical time periods or faster changes within generations. One contributing factor driving such fast changes could be “fashion cycles”, as this is a very general cultural phenomenon. Here, results are reported from a corpus study investigating trends over time in the use of *cyber* as the first part of compound nouns, and of alternatives which carry a similar meaning, such as *e-* as short for *electronic*. It is found that *cyber* was commonly used in the time period 1995-2004. Usage then strongly declined, but there was a new peak in popularity in the last year of available data (2012). Interestingly, *cyber* was initially used in positively charged or neutral contexts (e.g. *cyberspace*), but in recent years mostly in negatively charged words such as *cyberbullying* or *cyber warfare*. The hypothesis that *cyber* has been replaced with *e-* was partially supported (in particular in the case of *e-mail*, but *e-books* is another prominent example of a recent rising trend in vocabulary). However, in most other contexts usage of *e-* actually peaked a few years before the last years of the available corpus data. In general, results were consistent with “fashion cycles” in that the popularity of using *cyber* or *e-*, and in particular of specific words including these compound noun parts, seems to come and go rapidly over time. Interestingly use of *cyber* was seen mostly in negative contexts during later time periods. No such change was apparent in the use of *e-*. An emerging hypothesis partially supported by the data is that words in commercial contexts (e.g. *cyber-business*, *e-business*) rapidly lose their positive charge as they become common and are replaced by other, more novel and more fashionable words. Corpus linguistics is a very powerful tool for investigating such patterns of change in the popularity of words, and the processes behind them.

Keywords

Corpus study, fashion cycles, prosody, compound nouns, fads, word usage, word change.

Contents

1. Introduction.....	1
1.1 Aim.....	1
2. Theoretical background.....	2
3. Materials and methods.....	4
4. Analysis.....	6
4.1 <i>Cyber</i> : general.....	6
4.2 <i>Cyber</i> : source sections.....	6
4.3 <i>Cyber</i> : time sections.....	6
4.4 <i>Cyber</i> : words.....	7
4.5 Prosody of <i>cyber</i>	8
4.6 <i>E-</i> : general.....	9
4.7 <i>E-</i> : source sections.....	9
4.8 <i>E-</i> : time sections.....	10
4.9 <i>E-</i> : words.....	10
4.10 Prosody of <i>e-</i>	11
4.11 Alternatives to <i>cyber</i> and <i>e-</i>	11
5. Summary and conclusions.....	12
References.....	14

1. Introduction

One of the basics of historical linguistics is the study of trends over time in vocabulary. These studies usually focus on long-term changes in word use and slow changes over generations. One possible factor contributing to faster changes in vocabulary might however be “fashion cycles”, short-term changes that manifests as a word gaining temporary popularity and then declining in use again. Fashion cycles, or “fads” is a very widespread cultural phenomenon (Acerbi et al 2012) but there seems to be relatively few studies of short-term changes in vocabulary using this theoretical perspective, even though the last decades has created very favorable conditions for these types of studies. Firstly, advances in corpus linguistics means that there are powerful tools easily available for conducting such studies. Secondly, the increasing use of the internet means that trends can travel quickly over the world and be replaced just as quickly, creating interesting cases to focus such studies on.

In this essay, I study the use of the *cyber* as first part of a compound noun (e.g. *cyberspace*, *cyber-space* or *cyber space*) as well as the similar compound noun part *e-* (as in *e-mail* or *e-book*) to try to determine signs of fashion cycles and to try to draw conclusions regarding societal factors that may influence word use over time.

The particular choice of words to study was inspired by an article on the World Wide Words website by Michael Quinion (1999) where he states that “it seems that e- is the new cyber-” and discusses the (in his opinion) excessive use of *e-* in terms of a fashion that will soon fade. As the blog was written in 1999, the subject provides an opportunity to see if his prediction came through.

1.1 Aim

This essay aims to research change in the use of compounds of the word *cyber* to investigate whether fashion cycle theory is applicable on patterns of change in word use.

Specifically, the aim is to answer the following questions:

- How has the use of the *cyber* changed over time?
- Has *cyber* been replaced? By *e-*? By some other compound noun part?
- Can the changes in word use be understood in terms of fashion cycles?

Towards answering the last question:

- Does popularity of the compound noun parts, and in particular of specific words including them, come and go relatively rapidly?
- Is there any pattern to these cycles that can be linked to non-random variation in the probability of copying word use from others?

2. Theoretical background

Corpus linguistics is a methodology which can be used for descriptive purposes, such as developing a lexical grammar by identifying common patterns in word use (Mahlberg 2006). It can however also be used as a tool for studies from various external theoretical perspectives. An important distinction can thus be made between "corpus-based" and "corpus-driven" research (Mahlberg 2005, cited in Mukherjee 2007: 133). In the former, a corpus is used to explore already existing theories, in the latter it is used to generate hypotheses and theories that can then be tested in other ways. For a corpus-based study such as the one reported here, theory regarding corpus linguistics in itself is not equally relevant as in the case of a corpus-driven study, but it is important to be aware of the basic assumptions that apply when using a corpus to infer a linguistic pattern. These include the general philosophical assumption that lists of words and how they are used in context is a meaningful tool for linguistics, but also the more specific assumption that frequency data of the type that can be gathered from a corpus actually can represent how words are really used (Biber et al 1998; Teubert and Čermáková 2007; Gries 2009). This, in turn, is dependent on various factors spanning from the size of the corpus (and suitability for the question to be answered) to statistical issues such as whether one frequency of word use is really different from another (Mukherjee 2007). In the words of Gries (2009): "As a linguist, you don't just want to talk about frequencies or distributional information, which is why corpus linguists must make a particular fundamental assumption or a conceptual leap, from frequencies to the things linguists are interested in, but frequencies is where it all starts".

The present study is concerned with trends in word use over time, and this subject can serve as an example of how very different theoretical perspectives can be applied to a similar subject in corpus-based linguistics. In historical linguistics, for instance, Petersen et al (2012) used a massive data set from millions of books published over two centuries in seven different languages to show that fewer new words tend to be formed as the size of a corpus expands over time, and the language "cools off". In a study covering more recent changes, Baker (2011) compared trends in word use over four equally-sized corpora of written British English. Studying the years 1931, 1961, 1991 and 2006, he reported on a number of common trends, including a tendency towards less verbose and more informal ways of writing.

One theoretical perspective that does not seem to have received so much attention in corpus-based linguistics is what could be called "fashion theory". In this field attempts are made to explain the "fashion cycles" or "fads" that can be observed in a whole range of cultural phenomena (see Acerbi et al 2012 for some references), ranging from clothing over pop music and popularity of dog breeds to subjects closer to linguistics—such as first names commonly given to infants, or frequency of use of specific keywords in academic vocabulary. The last study, performed by Bentley (2008), comes close to the subject matter of the present study, but was not performed using a corpus but rather a database of academic publications.

The characteristic feature of a “fashion cycle” is that a cultural trait does not become more or less stable once it has been introduced, but instead (often rapidly) increases in popularity, only to be abandoned as quickly. Furthermore, there is often nothing in the intrinsic properties of the trait that can easily explain the fashion cycles; rather explanations must be sought elsewhere in cultural dynamics (Acerbi et al 2012). Thus, a central concept in this field is “copying” of culture from others, and if copying is random or which factors may otherwise influence the rates of copying of a cultural trait. Bentley (2008), for instance, used the ISI Web of Science database to compare use of keywords against a null model of random copying of keywords used earlier by others, and found that there was greater selection of keywords (more deviation from the null model) in the physical sciences than in the social sciences. In this case, this has led to greater stability in keyword use over time in the physical sciences, a kind of balancing selection if an analogy is made with evolutionary theory. In other cases, however, deviations from random copying could be what drives a fashion cycle; such as when a new member of a royal family is named and that first name are for some time preferentially copied by the by parents of new born in the general public. Such examples have led to the theory that copying preferentially takes place from individuals of higher social status, but Acerbi et al (2012) showed that in general such copying does not adequately explain the dynamics of fashion cycles. Instead cycles occur spontaneously if individuals copy not only cultural traits but also the actual preferences for these cultural traits, regardless of the relative status of the individuals involved. In their mathematical models, new cultural phenomena are preferentially copied, but when they become common they are instead less likely than random to be copied.

Returning to corpus linguistics and applying “fashion theory” to temporal trends in the data, an attempt could be made to understand the dynamics in popularity of words by examining how they are used in context and if this could conceivably influence the likelihood of copying. Importantly, a corpus can never provide very much cultural context, and is for this reason of limited use for discourse analysis (Sanderson 2010), but there is often enough context present to at least give some leads. First, popularity in different sections of the corpus and in different source materials can be investigated. Second, collocation with other words can give indications as to how the word is used in sentences and in discourse, providing an idea of how users of the word perceive its meaning (Teubert & Čermáková 2007: 37-50). Furthermore: “Researchers can also compensate for the loss of the original context with their own knowledge of the social, cultural, communicative and pragmatic contexts of the texts they study” (Sanderson 2010: 61). This will be attempted in the present study, in order to hypothesize about processes that can result in non-random copying. As a possible part of such explanations it is also of interest to study collocation with words that can be labelled as “positive” or “negative” (Bednarek 2008), what I will henceforth refer to as “emotional charge”. Collocation studies for such purposes can often be problematic in computer-aided studies on a large dataset, as collocated words far from the target word (e.g. a negation) can change its meaning and charge. Studying the first part of compound nouns, as I have done here, has the advantage of simplifying collocation studies, since the part of

the word following it (or the word immediately to the right when it is used as a standalone word) arguably is by far the most important collocate for the purposes of determining meaning and charge (*cybersex* versus *cyber warfare*, for instance).

Bednarek (2008) provides an interesting discussion of the theoretical issues involved in going from collocation to meaning and emotional charge. She makes the distinction between “semantic preference” and “semantic prosody”, where preference is a wider term (i.e. including prosody) that can be used to indicate any type of frequent collocation with word categories. This could be categories that can be determined more or less objectively, e.g. words related to sports, or to commerce, but also for a tendency to collocate with words that can be seen as being “positive” or “negative”. The term prosody should according to Bednarek be used only for the latter case. She further notes that determining semantic preference outside of prosody is problematic enough, as the meaning of a word can be dependent on context, genre and domain. For prosody the problems are multiplied; the decision to label a word as positive or negative is often difficult to perform objectively, as the same word can be seen as having positive or negative implications in different circles (examples include *hippie* and *liberal*). Examples of this problem will become evident in the Analysis below, but perhaps also that many words have a clear positive or negative charge.

3. Materials and methods

This study used the Corpus of Contemporary American English (henceforth referred to as COCA). COCA was chosen as it is large (450 million words), current (spans 1990-2012), contain texts from a number of genres (spoken, fiction, magazine, newspaper and academic) and is well organized, which might be considered a prerequisite for studying change over time.

Occurrences in the corpus of *cyber* and *e-* as the first parts of compound nouns was the main study object. According to the online Oxford English Dictionary the first documented use of *cyber* was in 1961, and the term was originally used to form “words relating to (the culture of) computers, information technology, and virtual reality, or denoting futuristic concepts”, but later used also to form “terms relating to the Internet”. According to the same source, *e-* was first used as late as in 1988, and it is “prefixed to denote involvement in electronic media and telecommunications (esp. the use of electronic data transfer over the Internet, etc.), usually to distinguish objects or actions from their non-electronic counterparts”.

Cyber is a short form of the adjective *cybernetic*, and *e-* is short for the adjective *electronic*. In English, adjectives + nouns can form single-word compounds through concatenation (e.g. *cyberspace*) or compound nouns in the form of phrases (e.g. *cyber space*), and the compounds are thereafter grammatically treated as nouns (Pakov 2013). In some languages, such as Swedish, concatenation is the rule (*tunnbröd*, *blåvinge*). In

German, both alternatives occur, and the stress on either the adjective or noun as well as inflection of the adjective indicates either a compound or a phrase (Schlücker & Plag 2011). In English, there is no clear distinction between concatenated compounds and phrases, and Pakov (2013) notes that concatenation is not a useful indicator of a “noun compound” (because it is rare in the case of noun + noun constructions), and continues:

Another reason for this criterion not being very reliable for English is variation in spelling: often the same compound can appear orthographically separated, connected with a hyphen, or concatenated, e.g., health care, health-care, healthcare. It would be inconsistent to believe that healthcare and health-care are noun compounds, while health care is not. (Pakov 2013: 295)

The main emphasis that Pakov (2013) puts on compounds is put on those formed by stringing two or more nouns to form “noun compounds”, although the adjective + noun situation is also treated. As the present study is concerned with the latter situation, the term “compound noun” seems more suitable, and by analogy with Pakov (2013) all three alternatives (separated words, hyphenated words, or concatenation) are collectively referred to as variants of the same compound nouns.

A further rationale behind this choice was that an initial search suggested that the three variants are used interchangeably, and that *cyber* as a standalone word is always used to modify the following noun in terms of showing a context (in other words, *cyber* occurs exclusively as a bound form). Moreover, compounds with a space were far less common at 2.05 per million, compared to 12.54 per million for compounds without a space. Compounds using a hyphen were the least common, at a frequency of 1.16 per million (cf. Fig 1).

To find the most common of these variants, the concatenated forms, a search was made using *cyber**, the asterisk being used as a wildcard, and these results were used for the initial coarse survey of the use of *cyber* in different sections of the corpus and over time. The most common lexemes to use *cyber-* were then selected for further study, including all variants of the compound nouns. Brand names and company names were excluded from the study (*Cyber-shot*, *Cyberpower* and *CyberCom*), as proper names were assumed to not follow the same trends as other lexemes. Different derivations were grouped together with their lexical root (e.g. *cyberwar*, *cyberwars* and *cyberwarfare* were all grouped as *cyberwar*). A search was then made for split words beginning with *cyber* (i.e. *cyber* as collocate left 1) and the results were grouped with the most common concatenation form identified earlier (*cyber war* with *cyberwar*).

The search strategy for *e-* was similar, except that I did not expect any instances where *e* was used as a standalone word in a phrase and thus did not search for words collocated to the left of *e*, but only for *e-**. One complication was that it was not really feasible to search for concatenated words beginning with *e*, but without a hyphen, since a search for occurrences of *e** would return every word beginning with the letter *e*. If the tendency for concatenation without a hyphen has changed over time, this could produce artifacts, and I attempted to investigate this possibility by comparing the trends for the particularly common constructs *e-book*, *ebook*, *e-mail* and *email*. As this study was only concerned with instances where *e-* was used as short for *electronic*, there were some instances of irrelevant words that should be excluded from the study: *E-R*, *E-Z*

and *E-STARK*. They were however rare, and were thus for simplicity not excluded from the coarse surveys of usage in different source sections and over time.

Finally, attempts were made to find other possible words that could possibly replace *cyber* as the first part of compound nouns, modifying the following parts of compound nouns to refer to an involvement of electronics, computers or similar. The strategy employed to find such words was to search for the first part (or immediate left-hand collocates) of words that were frequently seen as the second part of words beginning with *cyber-* or *e-* (or as right-hand collocates of the word *cyber*).

The genres and contexts of compounds with *cyber* or *e-* were investigated to a limited degree by studying the distribution over different registers and source texts. Due to the limits of this essay, it was not possible to study each compound individually, but some tentative conclusions could be made.

4. Analysis

4.1 *Cyber*: general

This analysis starts by familiarizing the reader with the source material before presenting conclusions. Firstly, the distribution over different sections and time periods is presented and discussed in short. Secondly, the most common compound nouns that includes *cyber* is presented along with a short discussion. Thirdly, there is a more thorough analysis of the change over time, both in regard to the popularity of *cyber* in general and to the popularity of specific compound nouns that include *cyber*.

4.2 *Cyber*: source sections

Despite what might be assumed, *cyber* and compounds thereof is actually least common in the section Fiction (5.81 occurrences per million words, compared to an average of 12.54 over all registers). Within this section, however, the sub-section SciFi/Fantasy is unsurprisingly the source of most of the tokens (13.83 per million). *Cyber* and compounds thereof is most common in the section Magazines (17.84 per million), where several sub-sections have far more occurrences than the average for all registers. In the sub-section Social/Arts there are 352 tokens, resulting in a frequency per million of 46.01. A great majority of these tokens are from the magazine *Futurist*. The sub-section Science/Technology also feature a great number of occurrences (38.30 per million), as does Financial (29.30) and News/Opinions (19.30). *Cyber* is also common in the section Newspapers (15.56 per million), especially in the sub-section Money, at a frequency of 33.85 per million. To sum up, *cyber* seems to be used mostly in technological and financial contexts, but it is used to some degree in every context.

4.3 *Cyber*: time sections

The distribution over time is, however, less even than the distribution over sections (Fig 1). During the period 1990-1994 compound nouns that include *cyber* are only used with

a frequency of 4.55 per million, but the use increase during the time period to 8.80 per million in 1994. The following period, 1995-1999 (16.66 per million for the period), shows an increase in use until the popularity peak in 1997 (23.03 per million) and then start to decrease. The third period, 2000-2004 (14.20 per million), starts with a second peak in 2000 (21.75 per million) but the following years shows an uneven decrease in popularity that continue into the fourth period, 2005-2009 (9.70 per million). The very last period, 2010-2012 (22.67 per million) displays a new spike in popularity, and the very last investigated year, 2012, has a frequency of 31.59 per million - by far the highest peak.

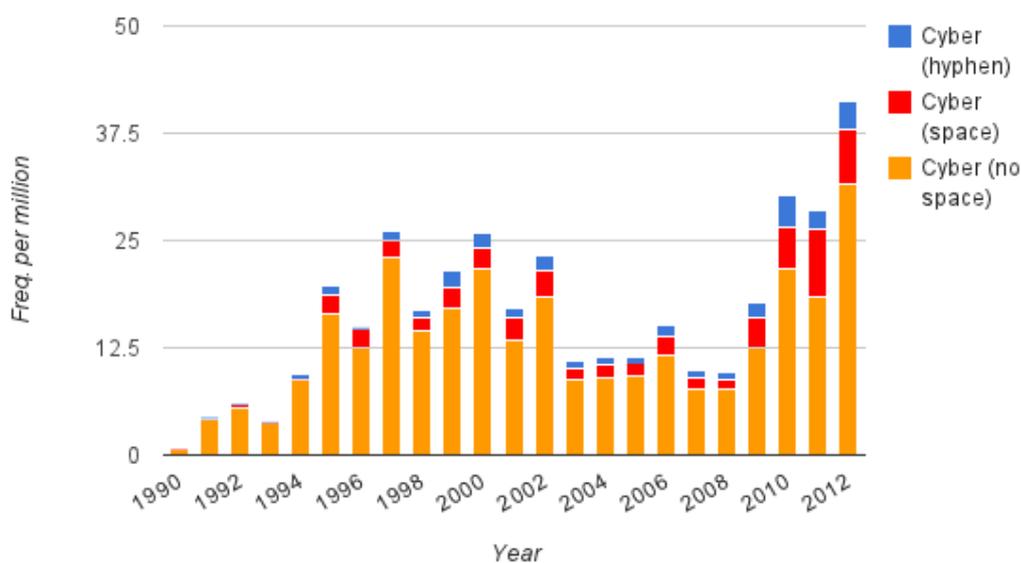


Fig. 1: Use of *cyber* over time

4.4 *Cyber*: words

The eleven most common compounds were (in falling order of occurrences in the most recent time section, 2010-2012) *cyberbully*, *cybersecurity*, *cyberspace*, *cyberattack*, *cybernetic*, *cybercrime*, *cyberwar*, *cyberworld*, *cyberpunk*, *cyberterror* and *cybersex*.

These common compounds were selected for a more comprehensive study, and here all variants of the compound noun were included (e.g. the tokens of *cyberspace*, *cyber-space* and *cyber space* are all added together in the result) and all words with the same word stem are grouped and counted together (*cyberwar*, *cyberwars* and *cyberwarfare* are thus all grouped as *cyberwar*). While there exists more than eleven compounds using *cyber*, these eleven were the only ones common enough for the purpose of studying change over time and distribution over different registers. The twelfth most common compound, *cyberinfrastructure*, only had a total of 22 tokens, all of them in the Academic section and all of them from 2005-2012.

Cyberspace seems to be the most resilient of these compounds regarding use over time. It is the most common of the compounds when studying the average for the entire 22

years covered by COCA, and while its use has declined sharply in the last decade, it remains the third most common compound as of 2012.

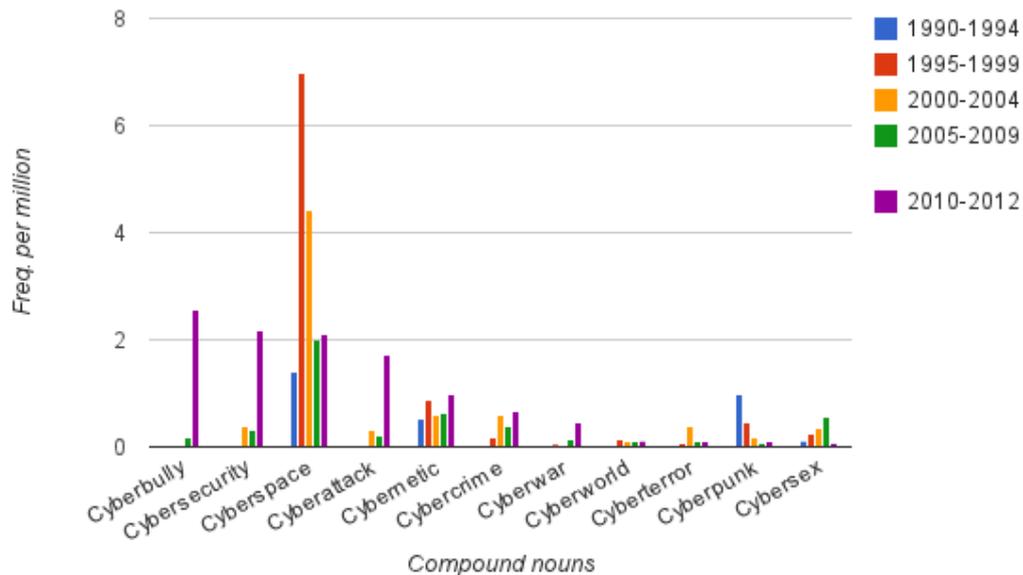


Fig. 2: Change over time in different compound nouns of *cyber*

In figure 2, the 11 most common compound nouns in 2012 are arranged left to right from least to most common. Furthermore, the frequency per million for five time periods are shown. As this figure shows, *cyberspace* had a very sharp spike in use during the period 1995-1999 and remained the most frequent compound of *cyber* in the following periods, before the sharp increase in use of *cyberbully* and *cybersecurity* makes these two compound nouns the most used in the last time period. Curiously, *cybersex* increase in use during the entire researched period before disappearing almost entirely in the last time period. *Cyberterror* is not in frequent use except for the time period 2000-2004. This spike is explained by a very sharp increase after the 9/11 terrorist attacks, after which *cyberterror* becomes one of the most widely used compound nouns of *cyber* for the following years. *Cyberpunk* is one of the most widely used compound in the first time period, second only to *cyberspace*. During the following time periods, the use declines steadily. *Cyberattack*, *cybersecurity*, *cyberbully* and, to a lesser extent, *cyberwar* all follow a similar pattern. After little or no use during the early periods examined, all four compound nouns experience an extreme rise in popularity during the last time period: 2010–2012.

4.5 Prosody of *cyber*

There seem to be a change in the prosody of *cyber* during the examined time period. As figure 2 and table 1 shows, there is a very clear increase in use of *cyberwar*, *cyberattack*

and *cyber bully* and a slight increase in use of *cyber crime* in the last time period examined - all of which are compound nouns that are arguably exclusively used in negative contexts. The earliest time period, however, have several compounds that are used in neutral or occasionally even positive contexts - *cyberspace*, *cybernetic*, *cybersex* and *cyberpunk*. While *cybercrime* is the fifth most common compound in the earliest period, it is used very infrequently compared to the four most common compounds.

Table 1: The 5 most frequent compound nouns of cyber at different time periods

1990-1994		2000-2004		2010-2012	
Cyberspace	1.41	Cyberspace	4.43	Cyberbully	2.56
Cyberpunk	1.00	Cybernetic	0.6	Cybersecurity	2.18
Cybernetic	0.53	Cybercrime	0.59	Cyberspace	2.12
Cybersex	0.12	Cyberterror	0.38	Cyberattack	1.73
Cybercrime	0.02	Cybersecurity	0.37	Cybernetic	1.00

The frequency per million of the five most common compounds of cyber- in three time periods, from all registers in COCA. Note that this includes all variants of compounds: hyphenated, concatenated and phrases.

4.6 E-: general

E-mail is by far the most common compound noun of *e-*, out of a total of 34876 tokens of *e-* in COCA there are 28929 tokens of *e-mail* and forms thereof (*e-mails*, *e-mailed* etc). Thus, any trends in sections for *e-* are mostly shaped by trends regarding the word *e-mail*. While there are some other compounds that starts with *e-*, they are fewer than the compounds of *cyber* and are all over-shadowed by *e-mail*. The second most common compound overall is *e-commerce*, with a total of 762 tokens (1.64 per million) and the third most common is *e-book*, with 459 tokens (0.99 per million).

4.7 E-: source sections

Much like *cyber*, *e-* is least common in the section Fiction, at 17.56 per million compared to the average for all sections: 75.11 per million. As with *cyber*, Science fiction and fantasy are one of the most common sections at 22.90 per million, but unlike *cyber* the section Juvenile has the most frequent use of *e-*, at 23.11 per million. The distribution can also be compared to *cyber* regarding the sections where the word is most common. Similar to *cyber*, *e-* is very common in the section Magazine (95.55 per million), and most common in the sub-section Science/Technology (239.31 per million) followed by Finance (172.54 per million). Unlike *cyber*, however, *e-* does not seem to be very common in the sub-section News/Opinions, only occurring at a frequency of 62.03 per million, below the average for all sections. The section with the greatest frequency of *e-* is the Newspaper-section, especially the sub-section Money (235.58 per million), yet another section where the distribution is comparable to that of *cyber*.

4.8 E-: time sections

There is a steady increase in use of compound nouns using *e-* during most of the years covered by the corpus (Fig. 3). This increase seems to peak in 2000 and then stabilize, which can mostly be explained by an increasing use of the compound *email* (without the hyphen). As figure 4 shows, this change in preference account for much of the decrease in use of *e-mail* and, as stated previously, *e-mail* makes up the great majority of all compounds of *e-*.

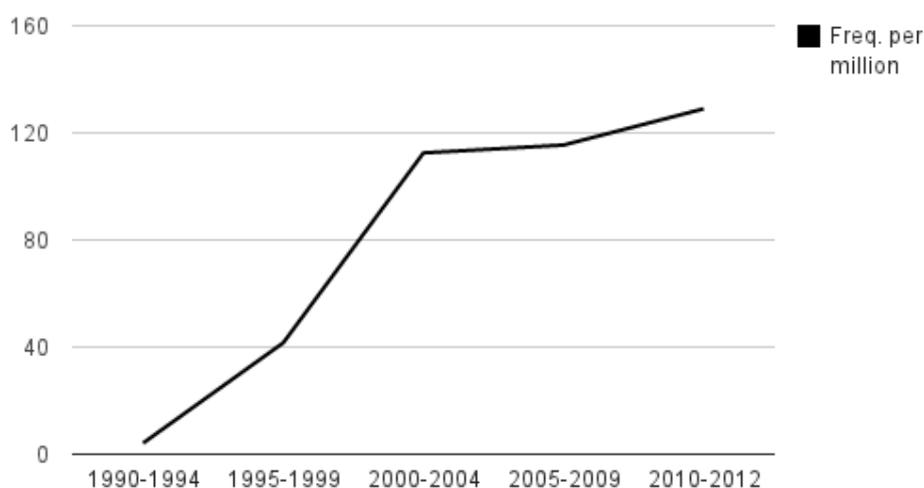


Fig. 3: Frequency of compound nouns beginning with *e-*

4.9 E-: words

The nine most common compounds were (in falling order of occurrences in the most recent section, 2010-2012) *e-mail*, *e-book*, *e-reader*, *e-commerce*, *e-learning*, *e-discovery*, *e-filing*, *e-trade* and *e-business*. As with the compounds of *cyber*, the decision to select nine compounds was based on the fact that these nine compounds were the only ones that occurred frequently enough for this study. There were examples of all nine spelled without a hyphen, usually with a lower case *e* and an upper case in the beginning of the second part of the compound (e.g. *eTrade*), these examples were however very rare and occurred only at a frequency of 0.05 per million, at the most. There are two exceptions to this: *ebook* and *email*. *Ebook* remains the less common form of the compound for the researched period, but *email* increases in use and by 2012 it has actually become exactly as frequently in use as *e-mail*—and if the trend continue it might very well be more common as of 2013 (Fig. 4). One consequence of this change in word form is that the decline in popularity of *e-* in recent years is partly explained by a transition of the most common word of this type to a form without a hyphen, which

cannot be detected without a search specifically for *email*. The fact remains, however, that most constructs with *e-* seem to be disappearing.

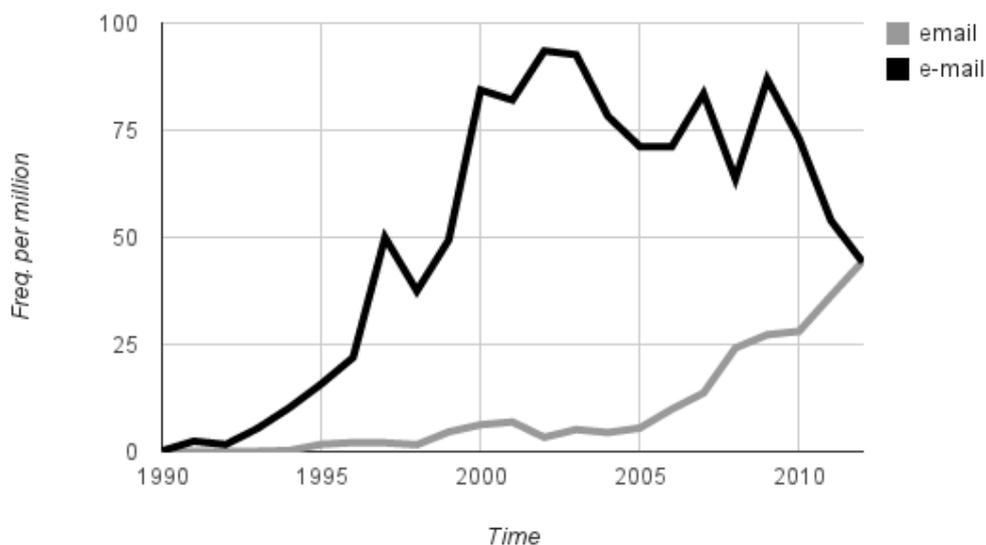


Fig. 4: Use of hyphen in *e-mail* over time

4.10 Prosody of *e-*

Interestingly, there does not seem to be any compound nouns of *e-* with a clear negative prosody. It is possible that negative compounds could show up in a second spike of popularity, as was the case with *cyber*, but there is no indication of such in the corpus.

4.11 Alternatives to *cyber* and *e-*

The search for alternative compound noun parts showing signs of becoming more frequent in later years turned up possibilities such as *digital*, *smart-* (as in *smartphone*, *smartcar*) and *online*, but the results were of very limited interest and relevance, as they were all rare over the years covered by the corpus. The only exception would be the word *smartphone*, which does in fact gain popularity over the last three years. It is possible that this has sparked an increase in use of *smart-* to form other words during 2013, but as COCA does not yet have any data for this year no definite conclusions can be drawn.

5. Summary and conclusions

In this essay I have studied a number of compound nouns that include *cyber* or *e-* as morphemes. Several of these combinations are rare and the numbers can easily be skewed if a single book mentions, for example, *cyberwarfare* a large number of times. Thus, some of the results can be questioned and it is hard to draw definite conclusions. Conclusions will, however, be drawn and might possibly inspire further research on the topic.

There were some clear patterns regarding overall use of *cyber* and *e-* over time. *Cyber* was commonly used in the time period 1995-2004. Usage then strongly declined, whereas constructions with *e-* became more popular. This was noted by contemporary commentators such as Quinion (1999), leading him to ask whether *e-* is the new *cyber*. However, the results of the present study do not wholly support Quinion's prediction in 1999 that *e-* would soon replace *cyber*.

There are indeed some striking examples of constructions with *e-* that has become popular in more recent years, in particular the case of *e-mail* (and most recently also the form *email*). *E-books* is another prominent example of a recent rising trend in vocabulary. However, in most contexts usage of *e-* actually peaked already a few years before the last years of the available corpus data. It could also be argued that the transition of *e-mail* to *email* signals that the word has been assimilated into the language as a common word which no longer evokes the connection with *electronic*; it is simply *email*. Furthermore, regarding *cyber* there was actually a new strong peak in popularity in the last year of available data (2012), even to levels higher than the early peak.

How can such fluctuations in popularity of *cyber*, *e-* and of specific compound nouns including these morphemes be explained? Schlücker & Plag (2011) found that when there is an alternative of using a compound or a phrase (importantly, these are quite distinct grammatical alternatives in German), new concepts tend to be formed as adjective + noun compounds when the adjective is already commonly used in this way in the language, but as a phrase when this is instead the common pattern for that adjective. Their study was an experimental investigation on speakers of German, who were asked to form new words from given adjectives and nouns. If such extrapolation from existing constructs when coining new words is a general phenomenon, as seems likely, it could also drive the formation of new compound nouns using the same already popular adjective. However, this process could explain the rapid rise in popularity of e.g. compounds with *cyber*, as new constructs “feed” on already existing ones, but not the subsequent decline and rise again in popularity that I observed.

I propose that “fashion theory” is an interesting theoretical framework in which to attempt explanations of such fluctuations in popularity of words (and parts of compounds). If such theory applies, the fluctuations would arise from temporal changes in the probability of copying word usage (and word preferences) from other users of the language (Acerbi et al 2012). The similarity to “fashion cycles” in e.g. clothing suggests

a parallel explanation. First, it should however be noted from Bentley's study of the popularity of academic keywords (2008) that "neutral", random, copying can lead to more volatile dynamics than selection of preferred keywords, so something akin to "fashion cycles" can perhaps arise without preferential copying. Thus, the following discussion should only be seen as a very preliminary exploration of the scope for "fashion theory" in linguistics.

A hypothesis in terms of fashion theory can be seen as emerging from my observations, namely that the fluctuations in popularity of *cyber* and *e-* are to a large part driven by the usage of these compound name parts in commercial contexts. Note that both morphemes are common in source sections of the corpus dealing with Finance and Money. It can be speculated that in such commercial contexts word usage is particularly driven by fashion, in that it is crucial to use words that are positively charged. It is part of the "logic of fashion cycles" (Acerbi et al 2012) that cultural phenomena are no longer preferentially copied when they become common, rather the reverse, leading to a rapid decline of the phenomenon. Advertising, copywriting and other commercial contexts, with its constant search for novel ways of describing products to stimulate the interest of buyers and investors, would certainly seem to fit this description.

A comparison between the early and recent peaks in usage of *cyber* leads to the interesting observation that the prosody of *cyber* seems to have changed. *Cyber* was initially used in what is arguably positively charged or neutral contexts (e.g. *cyberspace*), but in recent years it is mostly used in negatively charged words such as *cyber attack*, *cyberbullying* or *cyber warfare*. In line with my "fashion theory" interpretation this could be understood as an early rise and fall in popularity of compound nouns including *cyber* in positive, commercial contexts, with the recent peak being partly unrelated to such contexts and perhaps partly used in commercial contexts when a negative charge is called for as a contrast to the positive solution: the product being sold, e.g. anti-virus programs to counter *cyber attacks*; the consumer needs *cybersecurity*, as *cyber* is considered "bad". It seems likely that the positive charge of a novel buzzword is rapidly depleted to the point where it is no longer preferred in advertising, calling for new constructs, whereas words with a negative charge can in this way linger in the vocabulary even in commercial contexts.

Whether or not these speculations have any truth to them, I hope to have shown that corpus studies are a powerful tool for addressing questions regarding rapid changes in the popularity of words, and perhaps also that "fashion theory" should be considered as a theoretical framework within which to interpret such patterns.

For future research, it might be of interest to study the same compounds in more depth than could be done in this essay, investigating whether they "migrate" to different contexts and genres due to (or thus causing) the change in prosody. Obviously, the generality of the findings should also be examined through further research on word formation and fashion cycles.

References

- Acerbi, A., Ghirlanda, S., Enquist, M. (2012). The logic of fashion cycles. *PLoS One*, 7(3): e32541.
- Baker, P. (2011). Times may change, but we will always have money: Diachronic variation in recent British English. *Journal of English Linguistics*, 39 (1), 65-88.
- Bednarek, M. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory* 4 (2), 119-139.
- Bentley, R.A. (2008). Random drift versus selection in academic vocabulary: an evolutionary analysis of published keywords. *PLoS One*, 3(8): e3057.
- Biber, D., Conrad, S., Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge Univ. Press
- "cyber-, comb. form". OED Online. September 2013. Oxford University Press. <http://www.oed.com.ezp.sub.su.se/view/Entry/250879?rskey=66eiD5&result=2> (accessed October 28, 2013).
- "e-, comb. form". OED Online. September 2013. Oxford University Press. <http://www.oed.com.ezp.sub.su.se/view/Entry/249577?rskey=U18Y1R&result=8> (accessed October 28, 2013).
- Gries, S. (2009). What is Corpus Linguistics?. *Language and Linguistic Compass*, 3. 1-17: 10.1111/j.1749-818x.2009.00149.x
- Mahlberg, M. (2005). *English general nouns. A corpus theoretical approach*. Amsterdam/Philadelphia: John Benjamins.
- Mahlberg, M. (2006). *but it will take time...*points of view on a lexical grammar of English. In: Renouf, A.. & Kehoe, A. *The changing face of corpus linguistics*. Amsterdam: Rodopi. 377-390.
- Mukherjee, J. (2007). Corpus linguistics and linguistic theory: general nouns and general issues. *International Journal of Corpus Linguistics* 12, 131-47.
- Nakov, P. (2013). On the interpretation of noun compounds: syntax, semantics and entailment. *Natural language engineering*, 19: 291-330.

- Petersen, A.M., Tenebaum, J.N., Havlin, S., Stanley, H.E., Perc, M. (2012). Languages cool as they expand: Allometric scaling and the decreasing need for new words. *Scientific Reports*, 2: 943. DOI: 10.1038/srep00943.
- Sanderson, T. (2010). *Corpus, culture, discourse*. Tübingen: Gunt Narr Verlag.
- Schlücker, B., Plag, I. (2011). Compound or phrase? Analogy in naming. *Lingua*, 121: 1539-1551.
- Teubert, W., Čermáková, A. (2007). *Corpus linguistics: A short introduction*. London: Continuum
- Quinion, M. (1999). *The E- prefix*. World Wide Words: Investigating the English language across the world. Page created 16 January 1999, retrieved 25 september 2013. <http://www.worldwidewords.org/topicalwords/tw-eea1.htm>

Stockholms universitet
106 91 Stockholm
Telefon: 08-16 20 00
www.su.se



**Stockholms
universitet**