Corpus Study of Newspaper Abbreviations

Puškarić, Nikolina

Sveučilište u Zadru

Odjel za anglistiku
Sveučilišni prijediplomski studij Anglistike

Nikolina Puškarić

Corpus Study of Newspaper Abbreviations

Završni rad

Zadar, 2024.

Sveučilište u Zadru

Odjel za anglistiku
Sveučilišni prijediplomski studij Anglistike

Corpus Study of Journal Abbreviation

Završni rad

Student/ica:
Nikolina Puškarić

Mentor/ica:
doc. dr. sc. Frane Malenica

Zadar, 2024.

Izjava o akademskoj čestitosti

Ja, **Nikolina Puškarić,** ovime izjavljujem da je moj **završni** rad pod naslovom **Corpus Study of Newspaper Abbreviation** rezultat mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mojega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 16. rujan 2024.

Table of Contents

1. Introduction

Abbreviation is a common form of word formation based on reducing pre-existing words to produce new ones, usually less formal versions of the original words used. Abbreviations are highly productive word forms, with new abbreviations relating to different topics appearing each year. There are three main types of abbreviations: clippings, contractions, and initialisms. Initialisms and their use in the newspaper domain will be the main focus of this paper.

This paper is outlined as follows: In Section 2, a few definitions of abbreviations will be presented along with their general taxonomy and disagreement between various linguists.
Its subsection 2.1. will deal with the historical context of abbreviations in general, mentioning the first time they appeared in different languages, talking about the first dictionaries of abbreviations, and emphasizing the problems concerning these dictionaries. Not only that, but the change in their purpose throughout history will be discussed. The main focus of subsection 2.2. will be special types of abbreviations called initialisms and their classification into acronyms and alphabetisms, while in subsection 2.3. the differences and similarities between them will be presented. The subsection 2.4. will deal with special types of acronyms called backronyms. Their definitions and examples will be provided.

Several previous works will be presented in Section 3. Some of these papers analysed abbreviations in text messages and compared them either between two languages (English and German) or two sexes (males and females). Others dealt with abbreviations in the documents related to the European Commission's mobility or even abbreviations in relation to journalism and social media.

The main topic of this paper is a corpus study of newspaper abbreviations, particularly initialisms, which will be dealt with in Section 4. I am interested to see the frequency of their usage across 2 newspaper types (broadsheets and tabloids) and 3 decades (the 1990s, 2000s, and 2010s). It will be done on the sample of 19 abbreviations manually filtered from the Sketch Engine's corpus English Broadsheet Newspapers 1993-2021 (SiBol). The research sample and methodology will be presented in subsections 4.1. and 4.2. These sample abbreviations will be analysed in Section 4.3. of this paper. After researching the frequency for each one of the 19 abbreviations across 2 newspaper types and 3 decades, in Section 4.4., the normalized frequency of all of the sample abbreviations will be presented and compared.

2. Abbreviation

Abbreviation is a popular form of word formation that involves reducing already existing words to create new ones that are typically more informal versions of the original words, as in (1) and (2). According to the Encyclopaedia Britannica (1876, p. 26), abbreviation is "a letter or group of letters, taken from a word or words, and employed to represent them for the sake of brevity". According to the Oxford English Dictionary, "abbreviation is a shortened form of a spoken word or written symbol; a part of a word or symbol standing for the whole".

(1) ETA = 'Estimated Time of Arrival'

(2) TBC = 'To Be Confirmed'

Three highly productive ways in which abbreviation is involved in English word formation can be distinguished: initialisms, blends, and clippings. The focus of this study will be on initialisms. However, there is a problem when it comes to the systematic study of abbreviation because there is a lack of consistency and even overlapping within a general taxonomy; that is the reason why many linguists still disagree on it (Cannon, 1989, p. 106). According to Plag (2018, p. 125), abbreviations are in reality similar to blends "because both blends[1] and abbreviations are amalgamations of parts of different words".

An example of a blend would be (3). They are also similar to truncation[2] since they involve the loss of material and not the addition of it, like, for example, affixation. An example for truncation would be (4). On the other hand, abbreviation is different from both blending and truncation in terms of prosody. When it comes to abbreviation, orthography is the most important, not the prosodic categories as in the other two types of word formation.

(3) brunch = 'breakfast + lunch'

(4) carb from 'carbohydrate'

When it comes to terminology, Cannon (1989, p. 106) emphasizes the confusion; he states that abbreviation often implies any type of shortening, including contractions. It is used inconsistently for both the item and the process of word-formation that produces the item. He proposed replacing the term abbreviation with the common term shortening "as the name of the division[3] that produces blends, acronyms, abbreviations, and other reduced items" (1989, p. 107).

[1] The act of making something shorter or quicker, especially by removing the end of it
[2] A word formed by combining two other words
[3] A taxonomic term meaning 'a grouping of similar word-formation categories

2.1. Historical context of abbreviation

The history of abbreviations goes way back with abbreviations constantly changing its forms creating confusion in its general taxonomy. “Abbreviations are not only of linguistic value but also of historical as they may contain archaic elements, being formed with ancient models of contraction. Abbreviations are not artificially created linguistic units for language economy but they eliminate the contradiction between the modern consciousness and limited lexical resources of any language” (Zerkina et al. 2015, p. 138).

For the first time, abbreviations occurred in Sumerian (Cannon, 1989, p. 99). There are a few examples of abbreviations in Hebrew as in (5) and (6).

(5) INRI = ‘Jesus Nazarenus Rex Judaeorum’

(6) SPQR = ‘Senatus populusque Romanius’

The first version of dictionaries of initialisms appeared in the fifteenth century. *Modus Legendi Abbreviations* was probably the first one as well as one of the most famous dictionaries of initialisms. It was published in 1475. It consisted of examples that were pronounced as the whole words instead of as reduced forms, which can be seen in examples (7) and (8) (1989, p. 100). Other examples of dictionaries of initialisms would be *Lexicon Diplomaticum* by Johann Ludolph Walter (1745), *Manuel Tironien* by Aime Ambroise Joseph Feutry (1775), *Dictionnaire desabréviations*by Alphonse Antoine Louis Chassant (1846), and Courtenay’s and Macgregor's dictionaries from 1885.

(7) nobc = ‘nobiscum’

(8) gle = ‘generale’

The problem was that all of these dictionaries consisted of only a few real abbreviations; most of the examples were clippings, as abovementioned (7) and (8) and also (9).

(9) Aar = ‘Aaron’

Abbreviations gradually started to serve particular audiences, as demonstrated by Martin's Latin and French abbreviations employed in English historical manuscripts (1892) and Cordingley's mercantile collection (1902). (Cannon, 1989, p. 100)

In 1867, De La Rue collected a short list for the *Annual Report* of the Smithsonian Institution; this list included American abbreviations as in (10), (11), (12), and (13). It was a typical American tendency to simplify complicated terms to the simplest acronyms.

(10) P.D.Q. = ‘pretty damn quick’

(11) N.G. = ‘National Guard’

(12) C.O.D. = ‘cash/collect on delivery’

(13) O.K. = ‘oll korrect’ (*all correct)

During World War One, the list of abbreviations was created more methodically. In 1918, military abbreviations were classified. An example of an abbreviation invented during World War One would be (14). However, it was not World War One but World War Two that marked the beginning of the actual increase of intentionally produced initialisms. Many new abbreviations appeared in the United States Army Technical Manuals; some of them would be (15) and (16) as the longest ones in their corpus.

(14) A.B. & Q. = ‘appropriation for barracks and quarters’

(15) F.O. = ‘forward observer’[4]

(16) USAMSMADHS = ‘United States Army Medical Service Meat and Dairy Hygiene School’

Throughout history, initialisms have also served a different purpose than their original medieval necessity for economy and efficiency. Several early linguists believed that initialisms were an essential, de-stigmatized component of technical vocabulary.

The growing popularity of initialisms in the 20th century can be seen through the growing quantity and size of dictionaries, some of which have been profitable enough to be expanded into later editions, as well as the growing number of specialized dictionaries. There has been an increase in the number of general foreign initialisms dictionaries in the second half of the 20th century, as in French, Russian, German, and Portuguese.

Cannon (1989, p. 105) concludes by saying that “in just 75 years, the number of initialisms in numerous languages worldwide has exploded from tens of thousands to perhaps 800,000 recorded in dictionaries, besides those that are still unrecorded”.

2.2. Initialisms

Initialisms are alphabet-based abbreviations, as shown in (17), (18), and (19). The most common way of forming abbreviations is by taking the initial letters of the sequences of a couple of words to make up a new word, as in (20) and (21).

(17) ATM = ‘Automated (or Automatic) Teller Machine’

(18) NMHU = ‘Not Much Here, You?’

[4] A soldier who directs artillery fire

(19) BRB = ‘Be Right Back’

(20) POTUS = ‘President of the United States’

(21) AWOL = ‘Absent Without Official Leave’

Even though initialisms are operations that can be used to produce new words (just like morphological operations), Haspelmath (2002, p. 25) claims that “they do not fall under morphology because the resulting new words do not show systematic meaning-sound resemblances of the sort that speakers would recognize”. He continues by giving an example as shown in (22): “If we know that CD is an abbreviation of compact disc, we do not know it as a result of unconscious language acquisition, but because we were explicitly told so” (2002, p. 25).

(22) CD = ‘Compact Disc’

Booij (2005, p. 21) expands on what Haspelmath concluded by saying that “if you do not know a certain acronym, there is no way to find out about its meaning on the basis of your knowledge of the language”. All of this means that if something is a process of word creation, it does not necessarily mean that it is also a process of word formation. In other words, word creation differs from word formation in that the new word is recoverable from that of its constituents, and usually, it is an intentional form of language use (Booij, 2005, p. 25). Once again, he is stating that “the lack of transparency of these words serves to create in-crowd groups who understand these shortened words, and so they have an important sociolinguistic value” (2005, p. 26).

In addition to that, Plag (2018, p. 127) also claims that using an acronym could communicate a social significance. He points out that “within certain groups of speakers, the use of an abbreviation can be taken as a marker of social identity: speaker and listeners, but not outsiders, know what the speaker is talking about”. Moreover, when talking about all of these types of initialisms, or alphabetical combinations, Gregg (1987, p. 122) says that “ABCs primarily belong to the peripheral vocabulary of the language. Not usually found in formal writing, they may drop out of the language when their temporary usefulness is finished”. As shown in (23), this is the abbreviation that is used within a TikTok community whose meaning outsiders would not understand.

(23) NPC = ‘non-playable character’

Furthermore, there are many difficulties when it comes to ABCs since their formation is unpredictable and different from the usual operations of lexical production. Laurie Bauer (1983, p. 237) presents these difficulties by saying the following: “The lack of predictability

in acronyms stems from at least two sources. Firstly, the phrase from which the acronym is taken is treated with a certain amount of freedom to permit the acronym to arise". For example, in (24) only the first part of a compound adjective (all-purpose) provides a letter for the acronym, while in (25) both parts of Anglo-Saxon provide a letter for the acronym (Gregg, 1987, p.22).

(24) BASIC = 'Beginner's All-purpose Symbolic Instruction Code'

(25) A.S.C. = 'Anglo-Saxon Chronicle'

On the other hand, in (26) the particle 'as' provides the A in the acronym, but in (27) the particle 'of' is not permitted to provide a letter (otherwise the acronym would be FOIST, which is presumably far less effective as the name of a trade union). It seems that the interests of the acronym are the deciding factor in what the "initial letters" of the phrase will be taken to include" (1987, p. 122).

(26) GRAS = 'Generally Recognized As Safe'

(27) FIST (Federation of Inter-State Truckers),

Initialisms can be divided into two categories: acronyms and alphabetisms, the difference between those two will be explained in the next subsection.

2.3. Acronyms and alphabetisms – differences and similarities

When talking about these two types of initialisms, Gregg (1987, p. 121) says that "the cover term to stand for either of these categories will be the compound alphabetical combination". This term "alphabetical combination" (or its abbreviated form 'ABC') was coined by him in this exact study.

The main difference between acronyms and alphabetisms (ABCs) is in the way they are pronounced. Acronyms are read orthoepically, which means that they are read as one word, as in (28) (29) and (30).

(28) AIDS = 'Acquired Immune Deficiency Syndrome'

(29) NATO = 'North Atlantic Treaty Organization'

(30) UNICEF = 'United Nations International Children's Emergency Fund'

On the other hand, alphabetisms are read, as the name says, alphabetically, which means that each letter in the string is pronounced as shown in (31), (32), and (33).

(31) FBI = 'Federal Bureau of Investigation'

(32) CIA = 'Central Intelligence Agency'

(33) USA= ‘United States of America’

The letters in alphabetisms can represent whole words, as in (34) and (35). However, they can also represent constituents in a compound or just parts of a word. Examples of that would be (36) and (37).

(34) VIP = ‘Very Important Person’

(35) OED = ‘Oxford English Dictionary’

(36) ID = ‘identification’

(37) TV = ‘television’

According to López Rúa (2006, p. 677), “in contrast with the regular behaviour of alphabetisms, the orthoepic requirements of acronyms allow some freedom in the selection of initials, thus adding to the complexity of the category”, as shown in the example (38).

(38) DISPLAY = ‘Digital Service Planning Analysis’

On the contrary, she goes on by claiming that “central but not prototypical alphabetisms comprise items displaying irregularities in the choice of initials, forms whose source is a word instead of a phrase”, as demonstrated in (39) (2006, p. 677).

(39) ECG = ‘electrocardiogram’

When it comes to peripheral alphabetisms, they consist of extraneous or rare items like numerals, symbols, complete word forms, or letters representing phases, types, or series, as in (40) (López Rúa, 2006. p. 677).

(40) UVA = ‘Ultraviolet A’

According to Silvestre and Villalva (2015, p. 115), “there is a great disagreement over what the terms initialism, alphabetism, and acronym represent and what they should be used”. They base their classification of abbreviations on López Rúa’s (2006). According to her, there are two types of abbreviations: simple and complex. Simple ones include only proper abbreviations (those that appear in the written form only), while complex ones include blends, clippings, and initialisms, which are then divided into acronyms and alphabetisms. All types of complex abbreviations can appear not only in the written but also in the spoken form. However, Silvestre and Villalva (2015, p. 115) address the problem of taxonomy, especially when it comes to the terms for initialisms, acronyms, and alphabetisms, claiming that they “are often used interchangeably or wrongly dubbed as abbreviations or shortenings (e.g., in Plag 2003; Jackson and Zé Amvela 2005)”.

Furthermore, when talking about acronyms, Carstairs-McCarthy (2002, p. 65) mentions that “the use of capital letters in the spelling of some of these words reflects the fact that

speakers are aware of their acronym status. It does not follow that any string of capital letters represents an acronym".

However, some words that were originally created as acronyms are no longer written with capital letters; these words, for most of the speakers, are not related to the words they were abbreviations of in the first place, as in (41) (Plag, 2018, p. 125).

(41) radar = 'Radio Detection and Ranging'

According to Plag (2018, p. 126), "acronyms, being pronounced like regular words, must conform to the phonological patterns of English, which can create problems in applying regular reading rules if the reading out would result in illegal phonological words". He gave an example of an alphabetism 'BBC' (42). He continues by saying that it is not likely that BBC will ever become an acronym "because [bbk] or [bbs] feature an illegal word-internal combination of sounds in English" (2018, p. 127).

(42) BBC = 'British Broadcasting Corporation'

Nevertheless, English speakers sometimes make these 'illegal word-internal combinations' pronounceable; in other words, they create new acronyms. It is really popular when it comes to linguistic conferences and their names. An example of that can be seen in (43). (Mattiello, 2013, p. 89)

(43) NWAVE [ɛnweiv] = 'New Ways of Analyzing Variation in English'

New acronyms are produced freely, especially by administrators and scientists, and also for the names of different organizations. Usually, the syllable that is left over gives us enough detail to recognize the word it is an abbreviation of, however, this is not always the case, as seen in the example of United Airlines' low-cost carrier as shown in (44). This carrier got the name as the abbreviated form of the word 'United' but using only the last three letters, which is not the ordinary way of shortening words (Delahunty and Garvey, 2004, p. 137). They mention that "advertisers make prolific use of acronyms and often try to make them pronounceable as ordinary words".

(44) Ted = 'Uni**ted** Airlines'

The next significant factor contributing to acronyms' lack of predictability is that not all abbreviations that could be acronyms are treated as such, and there does not appear to be any particular explanation for why certain ABCs should be ignored. As has already been discussed in this paper, BBC cannot be pronounced as a word since it is different from the rules that define the structure of English words in phonological terms. However, according to

Gregg (1987, p. 122), “the same is not true of examples (45) and (46). Usage alone would seem to make the difference, and it is not clear what factors influence the variant usages”.

(45) GOM = ‘Grand Old Man’

(46) OD = ‘Over-Dose’

He also gave an example of American journalist William Safire's story from 1986. It is a story about the name of the committee, as can be seen in (47). According to Gregg (1987, p. 123), “the predictable alphabetism "CRP" was formed (Barnhart, 1980, p. 125), but the unexpected acronym "CREEP" was also formed using the vowel "ee" of "reelect." It will turn out, however, that if Safire's group had had a chance to read this paper, they still might have avoided that embarrassing acronym, since the use of the vowel following an initial consonant is a precedented method of creating an acronym”.

(47) CREEP = ‘Committee to Reelect the President’

He goes on by giving yet another example. He indicates that "a better example of the unpredictability of the formation of an acronym for a rock formation found on the moon from the phrase, "Potassium, RareEarth Elements, Phosporus," as shown in (48), it begins with a 'K' because that is the chemical symbol for potassium” (1987, p. 123).

(48) KREEP = ‘Potassium, Rare Earth Elements, Phosporus’

Nevertheless, there are a series of interrelated general tendencies in word formation as well as specific definable tendencies in the formation of ABCs that will make predictions about the probable forms of alphabetisms and acronyms. To conclude, despite all of this, several closely linked general word formation tendencies as well as specific measurable tendencies associated with the creation of alphabetical compounds can be used to predict potential forms of acronyms and alphabetisms.

2.4. Backronyms

Backronyms are special types of acronyms. Those are words that already have a meaning but are reinterpreted as backronyms as shown in (49), (50), (51), (52), and (53).

(49) USA PATRIOT ACT = ‘Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act’

(50) MADD = ‘Mothers Against Drunk Driving’

(51) SAD = ‘Seasonal Affective Disorder’

(52) SPECTRE = ‘Special Executive for Counter-intelligence, Terrorism, Revenge and Extortion’

(53) COBRA = ‘Cabinet Office Briefing Room A’

Backronyms fall into two categories: those that are created with a predetermined set of words in mind, selecting the words to fit the acronym, and those that falsely claim to explain the origins of a word by using its first few letters. Possibly the most well-known example of a broad backronym of this kind would be (54). When traveling on a P&O ship from England to India during the Raj, posh passengers would, if it was possible, reserve a cabin on the ship's port side for the outbound trip and its starboard side for the return trip. They were so protected from the sun's heat each time. Therefore, "Port Out Starboard Home" was thought to be the meaning behind the abbreviation POSH, which the cabins supposedly stood for. The song "Posh" by Lionel Jeffries, which was featured in the 1968 film Chitty Chitty Bang Bang, may have contributed to the misconception's popularity. Even though there are a few theories, the true origin is unknown (Schleifer, 2004, p. 6).

(54) POSH = ‘Port Out, Starboard Home’

Another interesting backronym would be (55). The bacronym "START" as a whole is intriguing because it was created not solely as a term to refer to a possible disarmament treaty between the US and the USSR but also possibly to suggest that the US side intended to make a new, serious effort in disarmament negotiations with the former Soviet Union at a time when many questioned the US government's determination in seriously pursuing disarmament.

(55) START = ‘Strategic Arms Reduction Talk’

Along the way, the 'START' program replaced what had previously been the 'SALT' (56) disarmament initiative, which had not been successful. These findings demonstrate that political discourse participants believe it is crucial to name a phenomenon in a specific way to win a political debate. The primary idea of this kind of approach, according to Plag (2018, p. 127), “is that the name used for a given phenomenon will influence the language user's concept of and attitude toward that phenomenon”.

(56) SALT = ‘Strategic Arms Limitation Talks’

The examples of "START" and "SALT" further make it hard to determine whether abbreviations are new lexemes or allomorphs of an existing lexeme. Allomorphs are different realizations or new surface forms of the same lexeme. In other words, if certain bacronyms share the same meaning as the original words, it makes it more suited to analyse them as

allomorphs, compared to analysing them as different lexemes (Mattiello, 2013. p. 91). However, he goes on by saying that in the cases of "START" and "SALT," we could point out that the abbreviation is not entirely related to the base word's meaning because the abbreviation has a different connotation, which results in the creation of a new lexeme (2013, p. 91).

3. Corpus-based studies of abbreviations

In this section, I will analyse some of the previously conducted research on frequency of abbreviations using corpus. Bieswanger (2007) did research on abbreviations in text messages and compared them between English and German. He used a contrastive method, using two corpora of text messages, one in English and one in German. He wanted to know whether the average number of abbreviations in text messages differs within these two languages and if there were any preferences for specific types of abbreviations in English and German. These messages were mostly written by students or by people who just graduated from university. These students were both male and female. The German corpus consisted of 1500 text messages sent in 2001 and was subdivided by the age and gender of the senders. However, Bieswanger narrowed down the number of messages to ensure compatibility, and he used only 387 of them, sent by both males and females from ages 17–30. This way, English and German corpora would have similar senders. He covered six categories of shortenings: initialisms, clippings, contractions, letter/number-homophones, phonetic spellings, and word-value characters. The results of his research showed that shortenings are much more frequent in English than in German. English senders use almost six times more shortenings than German senders since the tokens showed that there are 5.57 shortenings per English message, as opposed to 0.86 in German messages, which proved his first hypothesis. He found that there are certain preferences for certain kinds of abbreviations. German senders use more initialisms; the use of clippings is similar between both corpora, while phonetic spellings and contractions are much more common in English than in German. The most frequent category of abbreviations in English is letter-/number-homophones (1.64 per average). This group of shortening is not present in the German corpus at all, but that may be because German as a language does not have the same potential for that group. The final number of shortenings in German text messages is extremely low, while in English text messages it is fairly high.

Similar research was conducted by Lyddy et al. (2013). They did their research on spelling in texts written by students (139), of both sexes, and all of them had English as their first language, in which each participant gave them 10 messages from the previous week. This method created a collection of 936 text messages, or 13 391 words (tokens), and 676 non-word units (such as emojis, symbols, or multiple punctuation marks). The categories of non-standard spelling were initialisms, contractions, emoticons, clippings, accent stylization,

misspellings, letter/number homophones, missed punctuation and missed capitalization, onomatopoeic expressions, nonconventional phonetic spellings, semantically unrecoverable words, and, of course, shortenings. The results showed that there were 3296 instances of non-standard spelling, which is 25% of the total word number. Missed capitalization was the most frequent group (22%), followed by accent stylizations (19%), and a large number of spellings in these text messages included phonetic abbreviations.

Fabijanić and Malenica (2021) conducted corpus-based research on abbreviations in the documents related to the European Commission's mobility. They provided a typological overview of abbreviations found in the corpus of Erasmus+ mobility documents. The corpus consisted of 4 million words. After narrowing down false positives and words inappropriate or inadequate for their research, they were left with 122 abbreviations. They looked at the creation of alphabetisms and acronyms. The narrower sense of creation implied "a symmetrical one-to-one relationship between the words in the source phrases and their initial graphemes, which are used in the creation of a particular abbreviation" (Fabijanić, Malenica, 2021, p. 53). The broader sense implied the omission of particular words of a phrase and the use of syllables in the creation of these abbreviations. 74 out of 122 abbreviations (60.66%) belonged to the narrow sense, and 48 (39.34%) belonged to the broader sense. Their hypothesis was that longer abbreviations would be pronounced as acronyms because of a linguistic economy principle. They could not verify all 122 abbreviations, but only 59. Their second hypothesis was whether the omission parts were linked to the length of the abbreviation and their type. The results proved their hypothesis from their previous research that the omission of parts of a source phrase is more notable in the formation of acronyms than in the formation of alphabetisms. In their research, they proved that the majority (60.66%) of abbreviations are created in a way that every word of a source phrase is represented by one grapheme. The rest of the abbreviations deviated from this prototype.

Askman (2020) conducted research on the language journalists use on Twitter. He examined linguistic features of electronic discourse and collected the data from well-known news platforms—CNN, the New York Times, and The Washington Post. The data was collected during the US presidential elections in 2020. They chose one reporter from each media house based on their activity on Twitter (commenting and posting). The key focus of their research was on acronyms, initialisms, contractions, abbreviations, and first-person pronouns. The corpus for the reporter of The Washington Post consisted of 4325 words; the corpus of the CNN reporter consisted of 6520 words; and the corpus of the New York Times

reporter consisted of 2164 words. This sums up to 1399 words from posts, and they added 11 610 words from news articles from these media houses. All of these shortened constructions are considered to be informal but are often used on Twitter because of the limit of characters for each post or comment. Askman found that contractions are the most commonly used type of shortening, making up 3% of the total number of words. The second most used are abbreviations, with around 2% of occurrence. They are followed by initialisms with also 2% occurrence, but a little bit lower than abbreviations. First-person pronouns had a 1.5% frequency rate, while acronyms are in the last place with only about 0.5% of occurrence. The results showed that shortenings are used more in tweets of these journalists, making up 10% of tweets in total. In comparison, shortenings only make up 1.8% of words in real articles/posts of the official sites of these media houses.

4. Corpus Study of Newspaper Abbreviations

My research deals with the frequency of use of abbreviations in the newspaper domain. In Section 2, I mentioned what abbreviations were, the way they were formed, the change in their general taxonomy throughout history, and the disagreement about it. I also mentioned types of abbreviations most significant for this paper called initialisms and their subtypes, acronyms and alphabetisms. Their name derives from the fact that they only consist of initial letters, and that is the reason why I chose them for my research. When it comes to other pieces of research, in Section 3, I mentioned a few previously done studies that are in some way related to mine and can help us better understand it.

This research is based on the following questions:

Research question 1: Has the frequency of use of abbreviations in British broadsheet newspapers increased or decreased throughout the years?

Research question 2: Are abbreviations more commonly used in broadsheet or tabloid newspapers?

In my research, I used the corpus *English Broadsheet Newspapers 1993-2021 (SiBol), which* is an online corpus from the Sketch Engine site that consists of 858, 566, 374 words. Currently, this corpus provides research of 18 different broadsheet and tabloid newspapers throughout 8 different years. Using the above-mentioned corpus, I researched the use of abbreviations, particularly initialisms, in the newspaper domain, their general frequency, and their normalized frequency throughout three different decades (the 1990s, 2000s, and 2010s) and two different newspaper types (broadsheets and tabloids). I used the CQL query: [word="[[:upper:]]{2,10}"] which enabled me to focus only on the words written in capital letters.

The sample abbreviations I manually filtered from this corpus are: *UK* (United Kingdom), *TV* (television), *US* (United States), *BBC* (British Broadcasting Corporation), *NHS* (National Health Service), *EU* (European Union), *ITV* (Independent Television), *PM* (Prime Minister), *SNP* (Special Needs Plan), *MP* (Member of Parliament), *FTSE* (Financial Times Stock Exchange), *BP* (British Petroleum), *UN* (United Nations), *EC* (European Community), *VAT* (Value Added Tax), *IRA* (Individual Retirement Account), *DVD* (digital video disc), *BT* (British Telecom), and *RBS* (Royal Bank of Scotland).

4.1. Methodology and research sample

The first thing I did in Sketch Engine's SiBol corpus was that I selected the option 'Concordance'. It is a feature that provides the user with a list of all instances of the searched word in a corpus. After that, I used CQL, which is a complex query that enabled me to use search criteria I specifically needed for my research. I used the following query: [word="[[:upper:]]{2,10}"] which enabled me to focus only on words written in capital letters, in this case initialisms (acronyms and alphabetisms).

To obtain initialisms from two different newspaper types and three different decades, I used the above-mentioned query five times. Each time, I used a random sample of 10,000 words, which I then had to filter manually to select only initialisms and get rid of the words in the headlines written in capital letters. Most of these false positives were the surnames of the famous footballers or the names of different football clubs from different leagues.

For the first query, I used the whole subcorpus of The Daily Telegraph as the representative sample of all broadsheet newspapers and for the second one, I used The Sun's subcorpus as the representative sample of all tabloid newspapers. The reason why I chose these two exactly is because these were the only two newspapers that provided enough data for each decade.

For the last three queries, I used The Daily Telegraph once again. However, each time I researched subcorpora of different years (1993, 2005, and 2013.) to see if the use of abbreviations in British broadsheet newspapers increased or decreased throughout the years.
I chose 1993 as the representative sample of the 1990s, 2005 as the representative sample of the 2000s, and 2013 as the representative sample of the 2010s since that is the way the *SiBol* corpus is organized.

After manually filtering initialisms and removing false positives, I conducted five different lists of initialisms. For the whole subcorpus of The Sun, I conducted a list of 309 items; for the whole subcorpus of The Daily Telegraph, I conducted a list of 989 items. For the subcorpus of 1993, I conducted a list of 744 items; for 2005, a list of 850 items; and for 2013, a list of 538 items. I then reduced each list to the top 10 most frequently used initialisms of each year and newspaper type. After doing all of this, I downloaded these 5 lists to Excel to check if some of the initialisms were reoccurring in different lists, and it happened to be the case.

After excluding the reoccurring ones, I got the final list of the 19 sample initialisms, which were the main focus of this study. In Excel, I analysed the normalized frequency of each abbreviation from the 5 lists to answer my research questions. This allowed me to compare the frequency in 2 different newspaper types and 3 different decades and to see in which newspaper type and decade they were used most or least frequently if they existed in the first place. A research sample is presented in the tables below. In Table 1, we can see the most frequently used abbreviations in the broadsheet and tabloid subcorpora.

Table 1. Top 10 most frequent abbreviations in the broadsheet and tabloid subcorpora

Tabloid		Broadsheet	
Abbreviation	Freq. per million	Abbreviation	Freq. per million
UK	0.1407	UK	0.43102
TV	0.08323	US	0.39931
US	0.07035	BBC	0.24177
BBC	0.04062	EU	0.1189
NHS	0.03369	NHS	0.1189
EU	0.02873	TV	0.10206
ITV	0.02675	MP	0.06837
PM	0.02477	FTSE	0.05747
SNP	0.0218	BP	0.04558
MP	0.01784	UN	0.04459

In Table 2, we can see the most frequently used abbreviations in the 1990s, 2000s, and 2010s.

Table 2. Top 10 most frequent abbreviations in the broadsheets in the 1990s, 2000s and 2010s

1990s		2000s		2010s	
Abbreviation	Freq. per million	Abbreviation	Freq. per million	Abbreviation	Freq. per million
US	0.26852	UK	0.39634	UK	0.45678
EC	0.2279	US	0.37256	US	0.34977
BBC	0.20015	EU	0.18727	BBC	0.33194
UN	0.19619	BBC	0.15457	EU	0.13178
UK	0.1189	TV	0.11791	TV	0.1189
MP	0.09909	NHS	0.09611	NHS	0.10899
VAT	0.09116	FTSE	0.0862	BT	0.05648
TV	0.08819	MP	0.07729	FTSE	0.05053
IRA	0.06341	DVD	0.06738	ITV	0.04261
ITV	0.05945	UN	0.05945	RBS	0.03666

4.2. Analysis

In this chapter, I will analyse the data collected from the Sketch Engine. The frequency per million (normalized frequency) of the manually filtered abbreviations in the 2 newspaper types and 3 different decades will be analysed. The reason why I chose normalized frequency is because the sizes of the 5 subcorpora are not the same, and normalized frequency allows for comparisons across different corpora. At the end of this chapter, it will be clear which newspaper type and which decade has the highest frequency of abbreviations.

As can be seen in Table 3, I have calculated the normalized frequency of all of the sample abbreviations in two different newspaper types. This enables us to answer the first research question and to see if the abbreviations are more commonly used in the broadsheet or tabloid newspapers.

Table 3. Normalized frequency of the sample abbreviations in 2 different newspaper types

Type of newspaper	Normalized frequency
Broadsheet	0,001901
Tabloid	0,000517

I have also calculated the normalized frequency of the sample abbreviations in three different decades. It provides a better view of their change in frequency throughout the decades and helps us answer the second research question.

Table 4. Normalized frequency of the sample abbreviations in the broadsheet newspapers throughout the decades

Decade	Normalized frequency
1990s	0,001543
2000s	0,001827
2010s	0,001781

As for the first research question, these 19 sample abbreviations are more frequently used in the broadsheet newspapers than in the tabloid ones. As we can see from Table 3, the difference between the frequencies of use of abbreviations between these two newspaper types is huge. In other words, the abbreviations in the broadsheets (0.001901) appear almost 4 times more than in the tabloids (0.000517). Even though tabloid newspapers tend to write more about pop culture whose integral parts are jargon words and abbreviations, it may be so that abbreviations are more present in the broadsheets because they use abbreviations for world political, economic, and financial organizations on a daily basis. For example, the abbreviation EU (European Union) is four times more frequent in broadsheet newspapers than in tabloid ones. On the other hand, even though television is a part of pop culture, it is still

more frequently used in the broadsheets than in the tabloids since it is still the most popular telecommunication medium.

Concerning the second research question, which is if the frequency of use has increased or decreased throughout the decades, it has been proven wrong. According to Table 4, we can see that the frequency of use of abbreviations did change throughout the decades, but the change is not consistent. Abbreviations were a lot less used in the 1990s (0.001543) than in the beginning of the 21st century. However, even though there is a small difference between the 2000s (0.001827) and the 2010s (0.001781), the frequency of use of abbreviations was the highest in the 2000s and not the 2010s.

To sum it all up, the frequency of their use has not increased or decreased throughout the decades as was assumed; it was rather inconsistent. We can definitely say that if we compare the 20th and the 21st centuries, we can see a bigger difference, which shows us that abbreviations became more popular in the last couple of decades. It could also be the fact that new abbreviations appear each year and that some of them did not even exist in the 1990s. For example, the abbreviation DVD (Digital Video Disc) did not appear once in the subcorpus of the 1990s since DVDs came out at the end of the decade. However, this abbreviation was pretty frequent in the subcorpora of the 2000s and 2010s. On the other hand, the abbreviation EC (European Community) was most frequently used in the 1990s since it is a former economic association that was then replaced by the European Union in 1993. This also leads to the conclusion that the use of abbreviations and their frequencies can show us the circumstances of a particular country in a particular period.

In Table 5 below, results for all 19 sample abbreviations are presented.

Table 5. Normalized frequency of the sample abbreviations across 2 newspaper types and 3 decades

Normalized frequency					
Abbreviation	Type of newspaper		Decade		
	Broadsheet	Tabloid	1990s	2000s	2010s
UK	0.43102	0.1407	0.1189	0.39634	0.45678
TV	0.10206	0.08323	0.08819	0.11791	0.1189
US	0.39931	0.07035	0.26852	0.37256	0.34977
BBC	0.24177	0.04062	0.20015	0.15457	0,33194
NHS	0.1189	0.03369	0.05549	0.09611	0.10899
EU	0.1189	0.02873	0.00099	0.18727	0.13178
ITV	0.04162	0,02675	0.05945	0.03567	0.04261
PM	0.01486	0.02477	0.00595	0.00991	0.00396
SNP	0.03072	0.0218	0.00297	0.00099	0.00099
MP	0.06837	0.01784	0.09909	0.07729	0.00892
FTSE	0.05747	0.01189	0.00099	0.0862	0.05053
BP	0.04558	0.00198	0.01288	0.03963	0.01982
UN	0.04459	0.00099	0.19619	0.05945	0.02576
EC	0.03567	0	0.2279	0.00495	0.00198
VAT	0.03765	0.00099	0.09116	0.01784	0.00991
IRA	0.01883	0.00099	0.06341	0.03369	0.00099
DVD	0.01189	0	0	0.06738	0.00793
BT	0.04261	0.00694	0.03666	0.04162	0.05648
RBS	0.0218	0	0	0.0109	0.03666

The most frequently used abbreviation in the broadsheets, tabloids, 2000s, and 2010s is the abbreviation UK (United Kingdom), as can be seen in Table 6, which makes a lot of sense since most of the newspapers from the SiBol corpus are British, as, for example, The Daily Mail and The Daily Telegraph. Only in the 1990s, the abbreviation UK was the fifth when it comes to its frequency of use.

The most frequently used abbreviation in the 1990s was US (United States). US, just like UK, is among the most frequently used ones since some of the newspapers from the

corpus are American, like, for example, the New York Times and Washington Times. When it comes to its frequency across different newspapers and decades, according to Table 6, the abbreviation UK had the highest frequency in the 2010s (0.45678) and the lowest frequency in the 1990s (0.1189), which shows us that its change in frequency is consistent with it consistently decreasing, and it was more frequently used in the broadsheets than in the tabloids.

Another abbreviation that shows up among the most frequent abbreviations in the SiBol corpus is TV (television). Even though tabloid newspapers write more about pop culture and television, according to Sketch Engine, it is more frequently used in broadsheet newspapers (0.10206) than in tabloid ones (0.08323). Its frequency also has increased throughout the decades, with its lowest frequency in the 1990s (0.08819) and its highest frequency in the 2010s (0.1189).

As I already mentioned, the abbreviation US (United States) was the most frequently used abbreviation in the 1990s. It is used a lot more in broadsheets (0.39931) than in tabloids (0.07035), and its change throughout the decades is inconsistent. Even though it was the most frequently used abbreviation in the 1990s (0.26852), in comparison to the 2000s and 2010s, it was used the least, but its frequency was higher in the 2000s (0.37256) than in the 2010s (0.34977).

BBC (British Broadcasting Corporation), since it is the most important broadcaster in the UK, is another frequently used abbreviation in the newspaper domain. Just like in all of the examples above, it is a lot more frequently used in broadsheet newspapers (0.24177) than in tabloid newspapers (0.04062), and its frequency throughout the decades is inconsistent. Its frequency was the highest in the 2010s (0.33194), but it was the lowest in the 2000s (0.15457).

The abbreviation NHS (National Health Service) is more frequently used in broadsheets (0.1189) than in tabloids (0.03369), and its change is consistent. It has increased throughout the years, with its lowest frequency in the 1990s (0.05549) and its highest frequency in the 2010s (0.10899).

EU (European Union), as one of the most significant international organizations, is among the most frequent ones in the corpus. It is a lot more frequently used in broadsheets (0.1189) than in tabloids (0.02873) since there is a lot more talk about politics in the broadsheet newspapers, and its change in the frequency of use is inconsistent with its lowest frequency in the 1990s (0.00099) and its highest frequency in the 2000s (0.18727).

Once again, just like BBC (British Broadcasting Corporation), the abbreviation ITV (Independent Television), which is a famous group of British television companies, is also among the most frequently used abbreviations in the newspaper corpus, with its frequency being higher in broadsheet newspapers (0.04162) than in tabloids (0.02675) and its change in frequency being inconsistent with its highest frequency in the 1990s (0.05945) and its lowest frequency in the 2000s (0.03567).

Even though broadsheet newspapers tend to write more about politics, according to Sketch Engine's SiBol corpus, the abbreviation PM (Prime Minister) is more frequently used in tabloids (0.02477) than in broadsheets (0.01486), and its change in frequency throughout the decades is inconsistent. It was most frequently used in the 2000s (0.00991) and least frequently used in the 2010s (0.00396).

SNP (Special Needs Plan)[5] also has a higher frequency in broadsheets than in tabloids, while its frequency of use in the broadsheets throughout the decades was really low, with its highest normalized frequency in the 1990s (0.00297) and the same normalized frequency in the 2000s and 2010s (0.00099).

The abbreviation MP (Member of Parliament) is more frequently used in the broadsheet subcorpus (0.06837) than in the tabloid one (0.01784), and its change in frequency is consistent. It has decreased in the frequency of use throughout the decades, with its highest frequency in the 1990s (0.09909) and its lowest frequency in the 2010s (0.00892).

FTSE (Financial Times Stock Exchange) is also more frequently used in broadsheet newspapers (0.05747) than in tabloid ones (0.01189) since broadsheets tend to write more about topics concerning the economy, and its change is inconsistent with its pretty low frequency in the 1990s (0.00099), but its highest frequency in the 2000s (0.0862).

BP (British Petroleum), as the most famous British oil and gas company, is among the most frequent ones in the whole corpus. Once again, it is more frequently used in the broadsheet subcorpus (0.04558) than in the tabloid one (0.00198), and its change in frequency is inconsistent with its lowest frequency of use in the 1990s (0.01288) and its highest frequency in the 2000s (0.03963).

The abbreviation UN (United Nations), just like EU, is one of the most famous political international organizations. It has a much higher frequency of use in the broadsheets (0.04459) than in the tabloids (0.00099), and its change is consistent. It has consistently

[5] Special kind of Medicare Advantage plan that has benefits that cover special health care or financial needs

decreased throughout the decades, with its highest frequency in the 1990s (0.19619) and its lowest frequency in the 2010s (0.02576).

One more abbreviation concerning world politics among the most frequent ones in the broadsheet subcorpus would be EC (European Community). It was frequently used in the broadsheet (0.03567) but is non-existent in the tabloid subcorpus.
Its change throughout the decades is consistent; its frequency has decreased consistently, with its highest frequency in the 1990s (0.2279) and its lowest frequency in the 2010s (0.00198).

VAT[6] (Value Added Tax) is a lot more frequently used in the broadsheets (0.03765) than in the tabloids (0.00099), since this is also an economic term, and its frequency of use throughout the decades has decreased consistently, with its highest frequency in the 1990s (0.09116) and its lowest frequency in the 2010s (0.00991).

IRA (Individual Retirement Account), just like other sample abbreviations concerning politics, is a lot more frequently used in the broadsheet subcorpus (0.01883) than in the tabloid one (0.00099). The change in its frequency is consistent; it has consistently decreased throughout the decades, with its highest frequency in the 1990s (0.06341) and its lowest frequency in the 2010s (0.00099).

DVD (Digital Video Disc) is not mentioned at all in the tabloid subcorpus. It was also non-existent in the 1990s, as DVDs first appeared at the end of that decade. However, it was among the top 10 most frequent ones in the 2000s (0.06738) since it was the decade in which they were really popular with other technologies taking over in the 2010s.

The abbreviation BT (British Telecom) is a lot more frequently used in broadsheet newspapers (0.04261) than in tabloid ones (0.00694), and its change in frequency is consistent. The frequency of its use has increased throughout the decades, with its lowest frequency in the 1990s (0.03666) and its highest frequency in the 2010s (0.05648).

The last sample abbreviation I analysed is RBS (Royal Bank of Scotland). It is more frequently used in the broadsheet subcorpus (0.0218) since it is non-existent in the tabloid one with its frequency being 0 and its frequency of use has increased consistently throughout the decades, with its highest frequency in the 2010s (0.03666) and its frequency being 0 in the 1990s.

[6] VAT can be written in two ways, using lower-case or upper-case letters

5. Conclusion

This paper explains what abbreviations are, how they are made, and the problem with their classification. It also refers to different linguists and their inability to agree on what abbreviations are since there is an overlapping between abbreviations, blends and truncation. This section also gives an insight into the historical context of abbreviations, their first usage in Sumerian, and its first dictionaries that included words which were not real abbreviations. Later on, special types of abbreviations called initialisms and their subtypes (alphabetisms and acronyms) are more thoroughly explained since initialisms (alphabet-based abbreviations) were the main focus of my research. In section 3, some previous corpus-based research I referred to and used as a background for mine were mentioned. All of them researched abbreviations in different domains from mine. After analysing those pieces of research, I conducted my own. I researched the frequency of use of abbreviations in the newspaper corpus across 2 different newspaper types (broadsheets and tabloids) and 3 decades (1990s, 2000s, and 2010s). When presenting the results, I firstly focused on each of the abbreviations across two different categories (newspaper type and decade). In the end, I presented normalized frequencies for all of the sample abbreviations across the same 2 categories to answer my last two research questions. I concluded that the frequency of use of abbreviations is a lot higher in the broadsheet newspapers than in the tabloid ones. As for the last research question, the results showed that the change in the frequency of their use throughout the decades is inconsistent with the highest frequency in the 2000s and the lowest frequency in the 1990s. However, there was not enough data for each decade and newspaper type and the sample was too small. With more data from different years and newspaper types, the results would be probably be different.

6. References

Aronoff, M. (1983). *A Decade of Morphology and Word Formation.* Annual Review of Anthropology, Vol. 12 (1983), pp. 355-375. https://doi.org/10.1146/annurev.an.12.100183.002035

Askman, D. (2020). *Analysing Journalists Written English on Twitter: A comparative study of language used in news coverage on Twitter and conventional news sites.* Department of English. Bachelor Degree Project English. Stockholm University

Bauer, L. (2021). *An Introduction to English Lexicology.* Edinburgh University Press.

Bieswanger, M. (2007). *2 abbrevi8 or not 2 abbrevi8: A contrastive analysis of different shortening strategies in English and German text messages.* In Floyd, Simeon et al. (eds.) Texas Linguistic Forum, Volume 50.

Booij, G. (2005) *The Grammar of Word. Oxford Textbooks in Linguistics.*Oxford Linguistics.

Cannon, G. (1989). *Abbreviations and Acronyms in English Word-formation.* Duke University Press.

Carstairs-McCarthy, A. (2002). *An Introduction to English Morphologx: Words and Their Structure.* Edinburgh University Press.

Fabijanić, I., & Malenica, F. (2021). *Corpus-based typology analysis of abbreviations in the European Commission's Mobility related documents.* Bulletin of the Transilvania University of Braşov Series IV: Philology and Cultural Studies, Vol. 14(63) No. 2 https://doi.org/10.31926/but.pcs.2021.63.14.2.4

Glossary of Abbreviations and Acronyms. (2015). International Atomic Energy Agency.

Gregg, A. L. (1987). *Tendencies in the Formation of Alphabetical Combinations: Acronyms and Alphabetisms.* Wichita State University.

Haspelmath, M. (2002). *Understanding Morphology (Understanding Language Series).*Oxford University Press.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language.* Cambridge University Press.

Jackson, H., & Amvela, E.Z. (2000). *Words, Meaning and Vocabulary: An Introduction to Modern English Lexicology.* A & C Black.

López Rúa, P. (2004). *Acronyms & Co.: A typology of typologies.* Universitat de les Illes Balears. Dept. Of Spanish, Modern and Latin Philology. Estudios Ingleses de la Universidad Complutense., vol.12, 109-129

López Rúa, P. (2006). *Nonmorphological Word Formation.* Elsevier Ltd.

Lyddy, F., & Farina, F., & Hanney, J., & Farrell, L., & O'Neill, N. K. (2014). *An Analysis of Language in University Students' Text Messages*. Journal of Computer-Mediated Communication, Volume 19, Issue 3, Pages 546-561.

Mattiello, E. (2013). *Extra-grammatical Morphology in English: Abbreviations, Blends, Reduplicatives, and Related Phenomena.* Walter de Gruyter GmbH.

Morton, R.W., & Abramson, M. M., & Collins, T. D., & Puzey., K. C. (2008). *United States Government Printing Office Style Manual.* Claitor's Pub Division.

Parkhurst, F. A. (1917). *Symbols.* John Wiley & sons, inc.

Plag, I. (2018). *Word-Formation in English (2nd edition). Cambridge Textbooks in Linguistics.* Cambridge University Press.

Quirk, R., & Greenbaum, S., & Leech, G., & Svartvik, J. (1972). *A Grammar of Contemporary English.* Longman Group Ltd.

Quirk, R., & Greenbaum, S., & Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language.* Longman Group Limited.

Schleifer, R. (2004). *Alchemical Calques or the Transmutation of Language.* Verbatim Quarterly Magazine.

Silvestre, J. P., & Villalva, A. (2015). *Planning non-existent dictionaries. Dictionarística Portuguesa IV.* Centro de Linguística de Universidade de Lisboa. Universidade de Aveiro.

Spencer, A. (1991). *Morphological Theory: An Introduction to Word Structure in Generative Garmmar.* Blackwell Publishers Ltd.

Yule, G. (2010). *The Study of Language (4th edition).* Cambridge University Press.

Zerkina, N., & Kostina, N., & Pitina, S. A. (2015). *Abbreviation semantics.* GlobELT: An International Conference on Teaching and Learning English as and Additional Language. Procedia – Social and Behavioral Sciences. Elsevier.

Zwicky, A. M., & Pullum, G. K. (1987). *Plain Morphology and Expressive Morphology.*Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society, pp. 330-340.

Corpus Study of Newspaper Abbreviations

Abstract

This paper provides corpus-based research on English abbreviations in the newspaper domain done on a sample of 19 abbreviations whose total frequencies across two different newspaper types (broadsheets and tabloids) and three different decades (the 1990s, 2000s, and 2010s) were compared. The corpus used for this research is an online corpus from the Sketch Engine site called *English Broadsheet Newspapers 1993-2021 (SiBol).* The results show that abbreviations are more frequently used in the broadsheet newspapers than in the tabloid ones and that their frequency of use was the highest in the 2000s.

Key words

abbreviations, newspapers, corpus study, broadsheets, tabloid newspapers

Korpusno istraživanje novinskih kratica

Sažetak

Ovaj rad pruža korpusno istraživanje o kraticama u engleskom jeziku unutar novinske domene provedenom na uzorku od 19 kratica čije su normalizirane frekvencije uspoređene između dviju vrsta novina (novina velikog formata i tabloidnih novina) te između tri različita desetljeća (1990-e, 2000-e i 2010-e). Korpus korišten za ovo istraživanje je korpus s internetske stranice Sketch Engine naziva *English Broadsheet Newspapers 1993-2021 (SiBol)*. Rezultati pokazuju da su kratice učestalije korištene u novinama velikog formata nego u tabloidnim novinama te da je njihova frekvencija bila najviša u 2000-ima.

Ključne riječi

kratice, novinarstvo, korpusno istraživanje, novine velikog formata, tabloidne novine