



LUNDS
UNIVERSITET

The Use of Americanisms in British English, with
a Specific Focus on *Gotten*: A Corpus Study

Jennifer Melckersson
ENGK01
Autumn 2017
Supervisor: Eva Klingvall

Abstract

This essay aims to, with the support of an extensive corpus study, investigate the use of *gotten* in British English. *Gotten* is an alternative past participle of *get* which occurs mainly in American English, but is British in origin. The main goals of this essay are to investigate if and how this word is used in current British English, as well as how its use may have developed since the late 20th century, with a secondary goal of evaluating the usefulness of corpus studies in the context of recent regional language change. The claim that *gotten* is not used in British English is investigated through comparing the instances of *gotten* in the 1994 edition of the British National Corpus to instances in the more recent BNC2014. The *gotten* tokens are also sorted into categories for the purpose of determining how the word is used. As the BNC2014 purely consists of spoken language, tokens in a sample of the written Corpus of Global Web-based English are also categorized.

Results of the study indicate that *gotten* has seen an increase in usage in British English over the last few decades, though it is still rare and nonstandard. If the samples used in this study were to be considered indicative of English as a whole, it could also be concluded that *gotten* is being used in a more diverse manner in terms of meaning than previously, both in British and American English. We have also been able to see in this study that corpora which claim to be representative of a specific country's dialect can face significant challenges in making sure that their data stems from the desired dialect.

Table of Contents

1. Introduction	1
2. Background	2
2.1 The history of <i>gotten</i>	2
2.2 Different uses of <i>gotten</i>	3
2.3 Previous research on the influence of American English on British English.....	6
3. Methods	8
4. Materials	9
4.1 BNC1994 & BNC2014.....	10
4.2 COCA.....	10
4.3 GloWbE.....	10
5. Results & Discussion	11
5.1 Individual challenges & interpretation of the results.....	14
5.1.1 GloWbE.....	14
5.1.2 BNC1994.....	15
5.1.3 BNC2014.....	16
5.2 Further categorization & implications.....	16
5.2.1 Idioms.....	16
5.2.2 Acquisition & possession.....	17
5.2.3 Other categories.....	18
5.3 Processes of language change.....	19
6. Conclusions	21
References	23

1. Introduction

In recent decades, there have been significant amounts of research into differences between American English (AmE) and British English (BrE). In addition to this there have also been several studies on how and to what extent AmE may be influencing BrE (Gonçalves, Loureiro-Porto, Ramasco & Sánchez, 2017). Are these two varieties of English becoming more or less similar? Is one changing due to influence from the other? The general consensus seems to be that AmE has considerable influence over BrE, being as Leech, Hundt, Mair and Smith describe, “one of the major moving forces on BrE” (2009, p. 253). Additionally, a corpus investigation conducted by Gonçalves et al. concludes that the “influence [of AmE] is felt even within the UK borders” (2017, p. 1). In other words, AmE exerts a certain amount of influence over BrE, but the extent of this has not yet been exhaustively researched. For example, the field of lexico-grammar, wherein one studies “the grammar of individual words”, is one area where there is room for more research into differences between AmE and BrE (Tottie, 2009, p. 342). This is what this study focuses on.

Gotten, one of the two past participles of the verb *get*, is often considered one of the most distinguished differences between AmE and BrE (Fowler's Concise Dictionary, 2016). It occurs on several lists of so-called “Americanisms”, words or phrases stemming from the United States that have come to be used in BrE (“Americanisms: 50 of your most noted examples”, 2011). However, *gotten* as a past participle of *get* did not originate in the United States, it originated in Britain (Oxford English Dictionary, 2016). At some point, it fell out of use there, while it continued to be used in the United States. The aim of this study is to examine, with support from corpus data, whether *gotten* has returned to use in British English, in what manner it is used, and what the future might hold for its usage. A secondary goal is to show advantages as well as limitations of corpus studies when investigating recent changes in regional varieties of the English language.

I begin by outlining the origin of *gotten* and its development over the centuries, followed by relaying the different meanings it can have in language use, and the different categories of meaning I have chosen to make use of. I then review previous studies and literature dealing with corpus linguistics and language change in connection with the influence of AmE on BrE. The results of this study have been that *gotten* is in fact used in certain areas of BrE, and that there is a significant upward trend in frequency of usage. However, the results have also shown that when investigating language change through corpus study, it is still important to

take other factors into consideration, such as the history of the word(s) being examined, and the accuracy of the corpora employed. This study also sheds light on some of the inherent limitations of corpora based on separating varieties of English from one another, as well as corpora that derive their data solely from the internet.

2. Background

2.1 The history of *gotten*

As mentioned, *gotten* originates from BrE. The Oxford English Dictionary (2016) dates it back to 1382 (OED), which means that it predates the now customary past participle of *get* in BrE, *got*. As described in this essay, there is significant academic consensus that *gotten* is now not used in standard BrE. The 2004 edition of Fowler's *Modern English Usage* even states that "[n]othing points more clearly to the North Americanness of a person than the ability to use the past participle forms *got* and *gotten* in a natural manner" (Fowler & Burchfield, 2004, p. 338). One might question, however, if *gotten* would be used more readily in any other dialects of BrE. There is little mention of this in the grammars, usage guides, or dictionaries consulted for this essay, though it does appear in passing in Fowler's *Dictionary* (2016). The entry for *gotten* there says that "[a]lthough it was once in regular use, *gotten* had been extinct in standard language in **southern Britain**" [emphasis mine] (2016). This raises the question: what has been different in, presumably, northern BrE? As mentioned, however, this is a perspective sorely lacking from other sources about the subject, and perhaps something for a future study to examine. As it stands, I would judge this question to be difficult to answer with the use of corpus investigations as employed in this essay, especially considering the regional separation of data that would be required.

As for the arguable extinction of *gotten* in BrE, it is difficult to pinpoint exactly when this lexical item disappeared. Two successful grammars of the time can perhaps provide a general time frame. For example, Robert Lowth (1762) (as cited by the British Library) prescribes *gotten* as the correct past participle of *get*. In contrast, Lindley Murray's grammar from 1795 includes *gotten* in a list of "obsolete" verb conjugations (p. 117). Whether or not this conjugation was entirely obsolete by the end of the 18th century, it seems reasonable on the grounds of these well-regarded grammars to state that it experienced a significant decline during the 18th century. A differing account from Marckwardt puts the end of sanctioned use of *gotten*

in BrE somewhat earlier, already in the “middle or late part of the seventeenth century” (Bernstein, 1965, p. 206).

Meanwhile in the United States, the use of *gotten* seems to have been a point of contention among linguists and grammarians. In Richard Grant White's words,

There is among some persons not uneducated or without intelligence a doubt about the past participle of *got--gotten*, which produces a disinclination to its use. I am asked, for instance, whether *gotten*, like *proven*, belongs to the list of ‘words that are not words.’ (1870, p. 118)

To this he says: “Certainly not” (1870, p. 118). In what can in some respects be called a contradictory account made just one year earlier, Richard Meade Bache (1869) states that “Gotten is in English still, but it is nearly obsolete” (p. 134). Despite calling it “nearly obsolete”, he goes on to concede that “some speakers and writers have an unaccountable partiality for it” (p. 134). Thus, what these 19th century usage guides tell us is that at this point in time *gotten* was used, though not without controversy or perhaps uncertainty, at least among those who crafted these mostly prescriptivist guides. As we get into the 20th century, that picture begins to change. Drawing on work by Marckwardt, Theodore Bernstein simply says that while BrE has discontinued its use of *gotten*, “Americans have continued to use it up to the present” (1965, p. 206).

Thus, even though both varieties of English had experienced similar trends with regards to *gotten* in late modern English, American English speakers found themselves in the 20th century using both *got* and *gotten* as past participles of *get*, while British English speakers used only *got* with an overwhelming majority. But why did *gotten* not die out in AmE as well? There are of course several reasons for this, but one potential explanation is that *gotten* has a distinct utility, which brings us to the different meanings it has the potential to convey.

2.2 Different uses of *gotten*

In order to investigate how *gotten* is used in BrE, we must be able to systematically look at its different utilities. As mentioned, in AmE there are sometimes different meanings associated with *got* and *gotten*. *Got* therefore remains in AmE despite the use of the other participle. The original meaning of *gotten* as it has been used in AmE is something like *obtain*, *acquire*, or

receive (Trudgill and Hannah, 2013, p. 61). Originally, *gotten* was used to convey this meaning, while *got* was used in all other contexts. Eventually, however, this distinction became blurry, and *gotten* began to be used in other contexts as well. Biber et al. (2002) put it like this:

In American English there is a meaning difference between *has/have got* and *has/have gotten*: *has/have got* usually refers to current possession, while *has/have gotten* means that something has been acquired or that a change of state has occurred (p. 160).

In this paragraph, three distinct ways of using *got* or *gotten* are described. From Biber et al. I have chosen to call them **possession**, **acquisition**, and **change of state**. But this is not the complete picture, as other ways of using *get* are listed earlier in their work. In fact, after reviewing this and other literature, I arrived at seven categories of usage for *get* which are useful for this study. What this means is that an occurrence of *gotten* is analyzed in a different way depending on the meaning it takes on. This in order to answer one of the main questions of this study: if *gotten* is used in BrE, how is it used? It is likely that *gotten* is only used in certain contexts and not in others. That is, after all, how it works in AmE. Using these distinct categories in the analysis portion of this study will serve as an indication of how similarly *gotten* is used in BrE to AmE, or, indeed, how differently.

As we have seen, there are several different ways of categorizing the uses of *gotten*. One could also differentiate between main verb usage and auxiliary usage, an aspect afforded much significance by Marckwardt and Walcott (1938, p. 87), or make use of concepts such as semantic categories, as – to name but one example – Biber et al. do (2002). According to Biber et al. (2002), *get* has a "wide range of meanings and grammatical patterns" (p. 111). To simplify, they state that *get* can be an activity verb, a causative verb, an occurrence verb, or a mental verb (pp. 111-112). I considered making use of this system, and took inspiration from it, but all of these categories are not of use in this study. Furthermore, they do not include all uses of *get*, and Biber et al. have to state in addition to these categories the auxiliary use of *get* (the so-called *get-passive*), the semi-modal use (as in *have got to*), the simple possession use, and the idiomatic use. Biber et al.'s categories are also too wide, encompassing broad categories of meaning instead of the narrower ones required for this particular investigation. Thus, these categories can provide perspective, but cannot be the main or only categories used for this study.

Instead it is Algeo's (2006) distinction between dynamic and static senses that serves as the starting point for the categories used in this study. In grammar, a verb describing something that changes or moves is dynamic, whereas a verb describing something that is constant, or an unchanging state, is static. Algeo (2006) mainly emphasizes the static nature of current possession, resulting in the use of *got*, and the dynamic nature of senses like **acquisition**, **change of state** and what he calls the sense in which one is "permitted" to do something, resulting in the use of *gotten* (pp. 14-15). To illustrate this permission sense, he uses the following example and definition: "*I've gotten to go* = I have become able to go" (p. 14). Additionally he also mentions various senses which are mainly distinguished through things like what the thematic role of the subject or object is, which are not necessarily relevant here. In order to encompass all meanings of what Algeo calls *permit*, I have chosen to call this sense **causation**. This because "verbs of facilitation or causation" is a semantic domain that according to Biber et al., encompasses a wider range of meaning than just permission, such as "*allow, cause, enable, force, help, let, require, and permit*" (1999, p. 363). Thus, having combined both Biber et al. and Algeo's categorization, we are left with the dynamic senses **acquisition**, **change of state**, and **causation**, and the static sense of **possession**. I argue that at least three further static categories are needed to fully describe all uses of *get*, *got*, or *gotten*. One of them is the semi-modal use of *got* as mentioned earlier, for example *I have got to*, which I have for simplicity's sake chosen to call **obligation**. Another is the **idiomatic** use, which can serve many different functions depending on the fixed expression in which the word appears. For ease of reference, the seven categories are listed below:

1. Acquisition
2. Change of state
3. Causation
4. Understanding
5. Possession
6. Obligation
7. Idiomatic

In this division, the static senses of **understanding**, **possession** and **obligation** are the ones which are used with *got* as their sole past participle in AmE. With regards to **possession** and **obligation**, this can be explained by the fact that here is where there can be significant difference in meaning between *got* and *gotten*. As discussed earlier, where *got* means

possession, *gotten* would mean **acquisition**. It can be argued that the same is largely true for **obligation** versus a subset of the **causation** category – that of permission or facilitation. For example, ‘I have got to go to London’ is an expression of obligation, but ‘I have gotten to go to London’ expresses some kind of permission or having been given a desirable opportunity. For this reason this particular subset of the category can never be interchangeable with **obligation**. When this occurs, this can be used as a tool to help categorize the tokens in my data for this essay.

To the distinction between *got* and *gotten* and its different meanings, Theodore Bernstein (1965) remarks that “[i]n its use of *gotten* in addition to *got* American English may well have the advantage” (p. 206). As Bernstein says, to have this distinction available within the language can be beneficial, because it can make it easier to discern different meanings that may not be immediately available through context alone. This is one possible reason why *gotten* had its resurgence in the United States. But is such a resurgence possible in the United Kingdom? And if so, why? Before discussing the methods through which this study is conducted, we will take a look at the mechanics of language change, and, more specifically, what research can tell us about language change brought about by influence from one variant of English to another.

2.3 Previous research on the influence of American English on British English

Although very little research has been conducted on a potential revival of *gotten* in BrE, there is plenty of evidence in the form of corpus-based research that AmE is exerting more and more influence not only over BrE, but over the English language as a whole (Gonçalves et al., 2017, p. 8). In order to execute their investigation, Gonçalves et al. used Twitter and Google Books as corpora. From Twitter, they derived their most current data and their global data, creating a corpus of the English language as used by people all over the world. Google Books provided them with the diachronic aspect of their data, in the form of English in books of both British and American origin (2017, p. 3). With the help of this data, Gonçalves et al. (2017) were able to conclude that “the past two centuries have clearly resulted in a clear shift in vocabulary and spelling conventions from British to American” (p. 8). This applies to English in a global sense rather than to BrE specifically, but through their historical data, Gonçalves et al. were also able to find that this trend is somewhat mirrored in BrE itself. It is described in the following way:

Interestingly, while the ratio timelines within the United Kingdom had been towards becoming ever more British, we find **a significant change of trend**

in the last 20 years of our dataset, corresponding to the period after the fall of the Berlin wall and the end of the Cold War that left America as the world's only superpower. [emphasis mine] (Gonçalves et al., 2017, p. 8)

Gonçalves et al. are not the only ones to see the advantage afforded to modern linguistics by modern corpora. In her research on the differences between AmE and BrE, Tottie (2009) argues that “thanks to the availability of bigger, better and more accessible computerized corpora, we are likely to find out much more about American-British differences” (p. 341). It follows, then, that the same corpora could be used to test whether something that is prevalent in AmE might be increasing in popularity in BrE. Golmann (2008) takes a list of words traditionally seen as “Americanisms” and uses corpora to do exactly this. The list of words comprising Golmann’s (2008) material for this study was taken from a study by Modiano (p. 13). Golmann takes issue with Modiano’s selection of lexical items, stating for example that several of the items in Modiano’s study are of Old English origin, not American (2008, p. 21). This conflict becomes an important part of Golmann’s research, and his results show that “only slightly less than half (seven out of sixteen) of the lexical items offered by Modiano as ‘Americanisms’, can actually be verified as such by their distribution in the corpus” (2008, p. 27). This leads him to conclude that corpora can be useful in investigations such as these, but that the “historical dimension of language contact” must also be taken into consideration when one is interpreting the results of a corpus investigation.

Another potentially problematic aspect to be aware of is that correlation is not necessarily causation. There are several different factors at play which affect ongoing language change in BrE. However, the influence of AmE may be one of these factors. In order to decide how to interpret the results of this study, we must therefore look at general processes of language change. Leech et al. (2009) call these processes “patterns of change” (p. 252), different patterns that can be observed when looking at diachronic corpora. They list the following six patterns:

- a) **Regionally specific change**, “where one variety changes in a certain respect but the other does not” (Leech et al., 2009, p. 252).
- b) **Convergent change**, where the two language varieties grow more similar than before.
- c) **Divergent change**, where the two language varieties grow more different than before.
- d) **Parallel change**, “where the two varieties change in the same direction and to a similar extent” (Leech et al., 2009, p. 252).

- e) **Different rates of change**, where the two varieties change in the same direction but at different rates.
- f) **Follow-my-leader**, “where one variety, moving in the same direction as the other, takes the lead, which the other follows” (Leech et al., 2009, p. 252).

They bring up one of these patterns as particularly relevant to the possibility of Americanization: **convergent change** (2009, pp. 253-254). Additionally, these patterns of change often overlap, and any one particular grammatical phenomenon can be said to show tendencies of several of these patterns (Leech et al., 2009, p. 253). Another pattern often observed in conjunction with **convergent change** is the so-called **follow-my-leader** pattern, “where [for example] AmE is ahead or changing more quickly than BrE” (2009, p. 254). In relatively small projects it can be difficult to test for these patterns, especially if one means to include all six. However, they will be considered in section 5.

3. Methods

The use of corpus investigations in research is a dynamic method. Corpus investigations have been used in the past with great results, and are especially employed when the object of study is a single word or phrase, such as in this study. As mentioned in the background, the effectiveness of using corpus studies to determine influence of one variety of a language on another has also been investigated. Tottie (2009) and Golmann (2008) both note its usefulness, while Golmann emphasizes that it is important to be aware of the limitations of this method.

The main material for this study consists of five different datasets from four corpora. The claim that *gotten* is not used in BrE is investigated through comparing the instances of this word in the original British National Corpus (hereafter BNC1994), which consists of written and spoken material from the second half of the 20th century, to instances in the more recent BNC2014, which is spoken only (Love, Dembry, Hardie, Brezina & McEnery, 2017). The spoken section of the BNC1994 is of roughly equal size to that of the BNC2014, that is to say, they both consist of around 110 million words. The equivalence in size makes it so that comparison of frequency between the two is appropriate.

Beyond the realm of frequency, the categories described in the background section are used to distinguish between traditional uses of *gotten*, in the **acquisition** sense, and broadened uses of

gotten, in any other sense. In other words, uses which fall into any of the **non-acquisition** categories are seen as evidence of the broadened meaning of *gotten*, as described by Trudgill and Hannah (2013). This serves as an indicator of how *gotten* is used in BrE. However, we also want to put this in the context of Leech et al.'s (2009) processes of language change, and in order to do that, how *gotten* is used in BrE today needs to be compared with how *gotten* is used in AmE. For that reason, use of the Corpus of Contemporary American English (COCA) is employed. I use samples of the COCA dataset to compare the ratio of **acquisition** uses to other uses of *gotten* in AmE in 1994 versus 2014. Being able to compare the ratio of **acquisition** to other uses in BrE and AmE will enable us to determine which of Leech et al.'s (2009) patterns of language change are applicable to the use of *gotten* in BrE. This is because the theories employed by Leech et al. (2009) require knowledge of language change in both languages to be able to describe how and why the language is changing.

The final aspect of the analysis itself is the additional dataset of *gotten* instances in the British section of the Corpus of Global Web-based English (GloWbE). A sample of this is employed to further investigate the manner in which *gotten* is used in BrE, and quite crucially, to provide a written sample of this usage to add to and contrast with the purely spoken form of the BNC2014. A more detailed description of the datasets and corpora follow in the material sections below.

But first, a note on the process of categorization. Categorizing the tokens is a manual process, which is one reason why the samples chosen are small in size. In order to determine which token belongs to which category, I rely on both the immediate context and, if necessary, the extended context provided by the corpus metadata. I am also able to use a few tests to support my categorization. As mentioned in section 2.2, where **acquisition** is concerned, *gotten* often takes on a different meaning than *got*. In these cases, I can substitute *gotten* for *got*, and if the meaning changes, I know I am dealing with an **acquisition** token. This of course raises the issue of distinction between static possession and dynamic acquisition, and whether or not BrE makes the same kind of distinction that AmE does. See section 5.2.2 wherein I argue that it does. Using the same principle, I can also test if a token is **causation** or **obligation**, though this has a more limited use in this particular study.

4. Materials

4.1 BNC1994 & BNC2014

The BNC1994 was created in the 90's, and consists of texts and transcripts created or published in the "later part of the 20th century" ("What is the BNC?", 2009). Although still useful in terms of contemporary BrE, it is likely that a certain amount of change has occurred within BrE since it was compiled. This is one reason why work has been underway to create a corresponding corpus of BrE with data stemming from this century. The spoken section of this corpus, the spoken BNC2014, was made available to the public in September of 2017, and consists of data collected between 2012-2016.

For this study, the first 100 instances of *gotten* in the BNC1994 and BNC2014 are examined. As there are less than twenty occurrences of *gotten* in the spoken section of the BNC1994, the spoken and written sections combined are used to provide this point of contrast, so that comparison between the two in terms of content will be possible.

4.2 COCA

COCA (Davies, 2008-) is a corpus of AmE that one might describe best as well-balanced. On the website itself it is also called "probably the most widely-used corpus of English" (COCA). It consists of over 530 million words, more or less evenly distributed between five main sections: spoken, fiction, magazines, newspapers, and academic texts. The data was also collected evenly from every year from 1990 up to and including 2015, making it possible to look at results from a single year, and contrast them with those of another year, which is what I have done.

For this study, the first 100 instances of *gotten* in the 1994 section of the corpus and the 2014 section of the corpus are examined. 1994 was chosen because it is the year of completion for the original BNC, and because of being an even twenty years before 2014. 2014 was chosen because it is also the median year of the BNC2014. Since the British and American data points come from similar times, we can compare the way both languages change, as that change relates to the research question.

4.3 GloWbE

While the other corpora employed in this study are fairly mainstream, the GloWbE (Davies, 2013) is slightly different. The entirety of the GloWbE data is derived from the internet, and the corpus consists of 1.9 billion words. This data was sourced from twenty different countries.

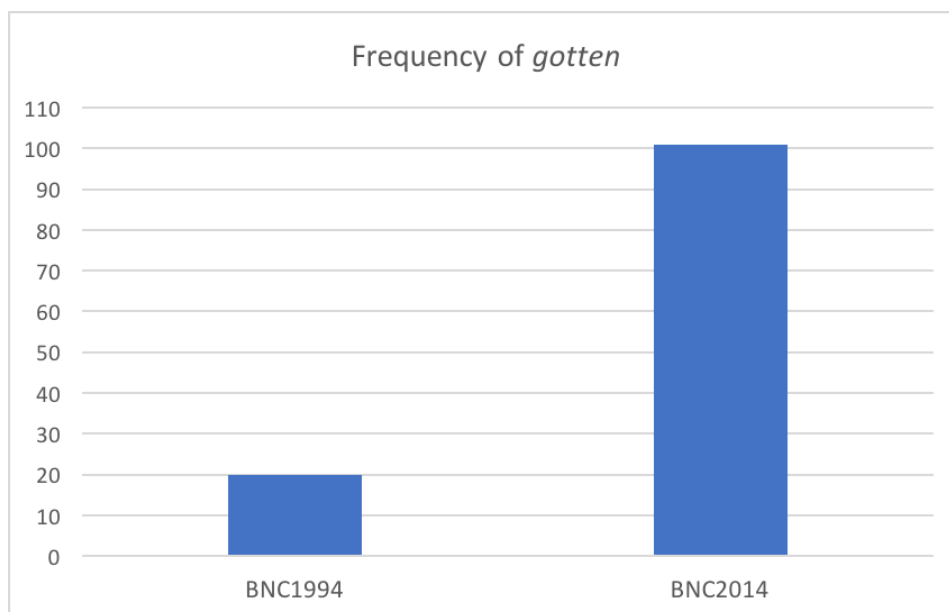
As the website states, this provides a “unique” corpus that “allows you to carry out comparisons between different varieties of English” (GloWbE). The data in the GloWbE is not divided into sections in the same way as, for example, in COCA. Instead it is either divided by country or into the categories **general** and **blog**.

There is also the issue of ascertaining the origin country of a website, a problem which they address on the website of the corpus (GloWbE). However, an issue that they do not address is that of quoting within a text. I found this to be a substantial problem when looking for results of the key word that this study is based on. Although, this is mainly a problem within newspaper or magazine articles, where, for example, an American person may be quoted in an article written by a Brit. By only looking at texts in the **blog** category, there is a way to mostly circumvent this issue. This is what I have done.

To sum, a sample of 100 tokens, the same sample size as all the other datasets, was derived from the intersection of **blog** and **Great Britain** in the GloWbE. As with the other samples, the first 100 tokens were selected without prejudice. It is of course necessary to keep in mind when looking at these results that the nature of the internet is such that people from all backgrounds are interacting on all kinds of websites. However, as I will argue, there is merit in this particular method as well. Both advantages and disadvantages of this kind of corpus will be discussed in this essay in the context of corpus linguistics, and in the context of the results of this particular study.

5. Results & Discussion

Looking at the frequency of *gotten* in British corpora or British sections of corpora, the claim that *gotten* is not used in BrE at all is plainly false, at least if claimed without any kind of caveat. In the spoken section of the BNC1994, which is equal in size to the spoken BNC2014, there are 20 occurrences of *gotten*. In the purely spoken BNC2014, there are 101 occurrences of *gotten*. Both of these numbers may be small in terms of raw frequency (20 out of 12 million words versus 101 out of 11 million), but this nonetheless shows an 80% increase in use of the word, as displayed in the diagram below.



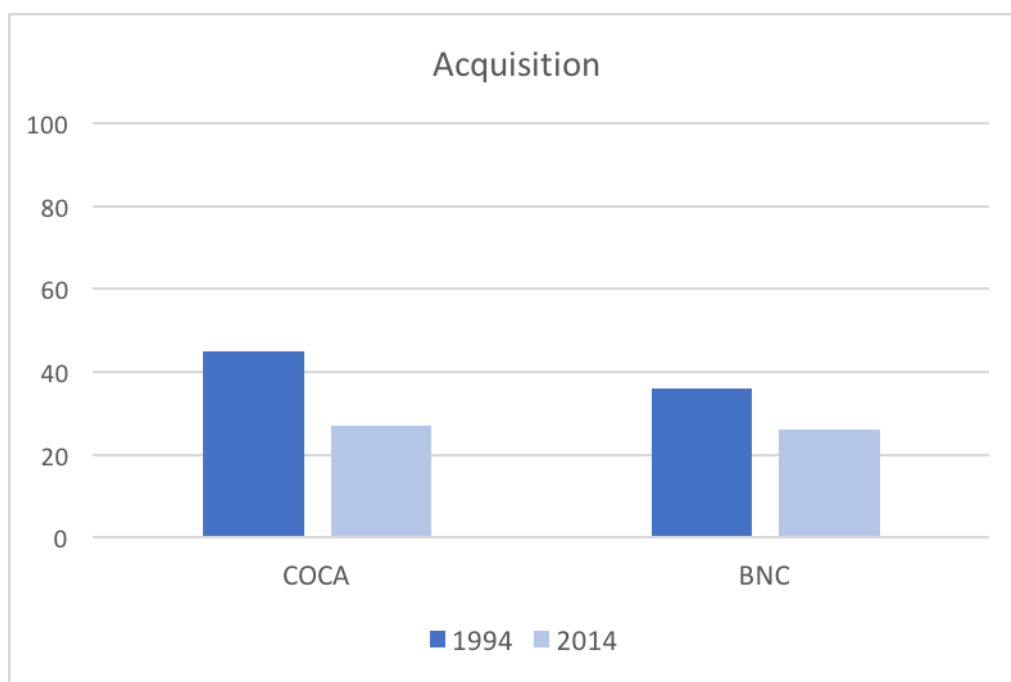
a) This diagram shows the 20 occurrences of *gotten* in the spoken section of the BNC1994 and the 101 occurrences in the BNC2014.

While there is no direct diachronic dimension of the GloWbE corpus, and with its close to 2 billion words it is substantially larger than the others, it does contain 5576 *gotten* tokens in its **Great Britain** section. This data and the data of the spoken BNC2014 combined tells us that *gotten*, at least to a certain extent, is used in both spoken and written BrE. It is important to note, however, that the GloWbE consists of forum entries and blog posts, and so can be said to represent an informal register of the language. Where more formal written BrE is concerned, *gotten* may be even less common than in informal written BrE. As noted before, for example, the GloWbE corpus at large also includes a great amount of newspaper articles, which are often of a more formal nature. This more formal section of the GloWbE contains what we might call ‘false positive’ instances of *gotten*, as British newspaper articles have been noted to quote speakers of AmE (see section 5.1 for a more thorough discussion of this).

Now that we have been able to establish that *gotten* can be found in the language of British English speakers, we can turn to the question of how it is used. An initial overview of the results indicates that the idea of a broadening usage of *gotten*, as posited by Trudgill and Hannah (2013), is supported by the American data. This is as we would expect. We can observe this by looking at the number of *gotten* tokens which fall into the **acquisition** category compared with those that do not. Specifically, 45 out of 100 *gotten* tokens in COCA1994 were used in the **acquisition** sense, whereas only 27 out of 100 in COCA2014 were used in the **acquisition** sense. In the data, the number of **acquisition** tokens have clearly decreased in 2014 compared

to 1994.

The British data is more difficult to disambiguate considering the dubious nature of the origin or reason behind some of the uses of *gotten* in, specifically, the BNC1994. It is impossible to find 100 tokens in this set not containing any verifiable false positives, such as examples where *gotten* is used as part of a quote of an American speaker or as part of a quote of an older variety of English. Nevertheless, were we to include these tokens, we would see that 32 out of 100 are **acquisition**. However, because of the false positives the results are presented after their exclusion, making the numbers 23 out of 64. In the diagram below these numbers are represented by percentages, so as to emulate parity with the other figures. The diagram clearly shows that **acquisition** use has decreased in both the COCA dataset and the BNC dataset, thereby implying a simultaneous increased usage of **non-acquisition** senses.



b) This diagram illustrates the decrease in use of the **acquisition** sense, going from 45% to 27% in the COCA samples and from 36% to 26% in the BNC.

To further illuminate this aspect of the analysis, we might look at the results from the GloWbE. In the GloWbE, 25 tokens out of the 100 token sample can be sorted into the **acquisition** category. This can be seen as an indicator that around a fourth of all *gotten* tokens in the GloWbE **Great Britain** and **blog** section fit into this category. This is fairly consistent with the ratio of **acquisition** to **non-acquisition** in the spoken BNC2014. We can thus establish that these results indicate that the **acquisition** sense makes up around 25% of all uses of *gotten*

in BrE. As the frequency in COCA2014 is 27 out of 100, it would seem that AmE usage is similarly distributed. Before discussing the way this usage of *gotten* has changed and how it might be used in the future, we take a look at the challenges presented by the individual corpora.

5.1 Individual challenges and interpretation of results

The different corpora used for this essay all presented their own challenges and tokens whose categorization was ambiguous or merits further explanation. Awareness of these challenges is required in order to accurately interpret the results of this corpus investigation, which is why these challenges are discussed in the following sections, along with an elaboration of the results themselves.

5.1.1 GloWbE

As mentioned, the GloWbE data was chosen largely due to a need to provide a written perspective on BrE as it pertains to the research question. More specifically, the intersection between **Great Britain** and **blog** was chosen to as far as possible eliminate occasions where *gotten* was used as part of a quote of, for example, an American speaker, such as in (1). The GloWbE gives the opportunity to visit the website where the data is pulled from, and it is in doing so one can discern that the following is a quote from an American novelist published in an online article of the British newspaper The Guardian.

(1) She has published two brilliant novels since she's been with me and neither has **gotten** anywhere near the recognition they so richly deserve. (GloWbE)

Furthermore, a noticeable and recurring problem with the GloWbE data stemming from newspaper articles is that in the data collected by the corpus, so-called 'comment sections' are also included. As we know, any person from mostly any country can read an online article published in another country and write a comment to appear below. This is the case in, for example, (2), where it is apparent from the text that the writer of the comment is of American origin.

(2) How many lives would have been saved if our addicts (I am writing as an American here) could have **gotten** treatment and counseling for their problems instead of getting labeled as a criminal and ostracized as a criminal? (GloWbE)

The decision to only include data from the **blog** section was made in an effort to circumvent these issues, and I believe from a cursory look at the websites responsible for the sample data that this has been largely successful.

5.1.2 BNC1994

There are few occurrences of *gotten* in the BNC1994 that I can confidently say were uttered or written by a speaker of BrE without them having either quoted a speaker of a different variant of English or deliberately tried to emulate a different variant of English. The latter occurred mainly in fiction novels that take place in America or feature an American main character, as in (3). There were also several examples of an older variety of English being quoted, such as in (4), where through the extended context provided by the corpus we can infer that the author is quoting 16th century philosopher Thomas Hobbes. Awareness of where the language in a corpus comes from is an aspect advised by both Golmann and Leech et al., and as we can see here it is crucial in interpreting the results of this study, both in terms of the data itself and in terms of the processes of language change which we will look at in section 5.4.

(3) because I want to tell you about another new idea I've **gotten**. Now I'm not a writer. (H0M, BNC1994)

(4) 'because they are given us immediately by nature, and not **gotten** by ratiocination, they are not philosophy.' Though true science (ABM, BNC1994)

Of the first 100 *gotten* tokens (out of 110), 64 tokens remain after filtering out American or pre 20th century ones. However, these 64 also include ambiguous examples which may well be American in nature as well, only it is more difficult to prove this or find it out through the BNC metadata or other resources available. The fact that so many of the texts, segments and quotes I happened to look at in this corpus were of American origin or explicitly made to reflect AmE could be slightly troubling in regards to this type of corpus. This reflects two distinct flaws in this corpus, the first being that speakers of AmE are quoted in it, and the second being British speakers emulating AmE. I suspect many corpora suffer from this exact issue. It is difficult if not impossible to find out how extensive this problem is, and whether or not it – in any significant way – affects the merit of this corpus or others like it, but it does show that it is not only recent-day corpora such as the GloWbE which suffer from being unable to separate variants of English from one another.

5.1.3 BNC2014

In the BNC2014 dataset, there were also a few tokens which were difficult to disambiguate. (5), for example, can be interpreted to be within the scope of the **acquisition** sense, but the speaker seems to correct themselves and ends up saying *got* instead. As the intended meaning itself is clear, this token has been included in the **acquisition** category, though it does deserve further thought. The speaker in question uses *gotten* in a way that is completely grammatical in AmE, but corrects themselves. It is possible that for this British English speaker, *gotten* is a semantically transparent alternative of *got*, but considered ungrammatical or incorrect.

(5) I haven't **gotten** I haven't **got** the excess (S346, BNC2014)

(6) oh there's one of these you definitely would have **gotten** (S6BS, BNC2014)

Entries like (6) can seem ambiguous at first glance, needing greater context than the transcription itself provides to be able to say with absolute certainty how *gotten* is being utilized. Context is often more than just what is said. It can also be what speakers are doing, where they are, their body language, etc. Thankfully the BNC2014 includes some such information in its metadata, and by looking at that we can see that this exchange occurs in the context of a game of Trivial Pursuit, a trivia game of sorts. Interestingly, this scenario occurred several times in the BNC2014 data, and in each case this information is used to conclude, with reasonable certainty, that these entries do fit into the **acquisition** category.

Other tokens which were by virtue of ambiguity excluded from a certain category or other are for example ones where neither the extended context, the corpus metadata, nor the sentence itself can provide a clear interpretation. This includes for example (7), where the person making the transcriptions wrote in *UNCLEARWORD* directly following *gotten*, presumably because it was impossible to discern what was being said. This also makes it impossible to categorize this token and others like it.

(7) most people who were there like like seven days who've **gotten** --UNCLEARWORD (S7SU, BNC2014)

5.2 Further categorization & implications

5.2.1 Idioms

As mentioned by Biber et al., *get* and its inflections occur with fixed and formulaic expressions. They mention expressions such as *get away with*, and *get rid of* (2002, p. 112). Thus, a note on these expressions is warranted. *Gotten away with* appears in my data ten times; see (8) for an example from the GloWbE dataset. From these small datasets – or samples as the GloWbE and COCA datasets are – there does not seem to be a clear trend in frequency of this expression separated by language variety. *Gotten away with* occurs four times in both the COCA2014 dataset and the GloWbE dataset, and one or zero times in the other datasets. *Gotten rid [off]* occurs five times in the BNC2014, 3 times in the BNC1994, one time in COCA1994 and zero times in the remaining two datasets. It is worth keeping in mind that any discrepancy here could simply be due to the small size of the samples. Phrasal verbs such as *get used to* and *get on with* can also be included in this category, both of which occur with *gotten* as a past participle in several of the datasets; see for example (9) and (10) from the BNC2014.

(8) That says clearly these guys have **gotten** away with it. That ought to change. That means legislation. (GloWbE)

(9) er yeah I can take because I've **gotten** used to the normal honey (S5XD, BNC2014)

(10) it makes me feel really good because I've always **gotten** on really well with black -- ANONnameF (SKPP, BNC2014)

5.2.2 Acquisition & possession

When it comes to other categories mentioned in this essay, there are some claims made by for example Algeo (2006) that I would like to further examine with the help of these categories. Algeo writes that *gotten* can be used in BrE, though rarely and – he argues – differently to how it is used in AmE. For example, he mentions just as Fowler's dictionary (2016) does, that BrE use of *gotten* can be dialectal, and he also says that *gotten* is “occasionally used interchangeably with got”, as in “Haven't you gotten your key? = ‘Don't you have your key?’” (p. 14). This implies that BrE does not make the distinction between static possession and dynamic acquisition such as we have established that AmE does. This is not directly supported by my data. In my British data, examples such as (11) and (12) can be found, which clearly make the distinction between having acquired something and simply possessing it. This, and other examples not listed, indicates that this distinction is present in BrE as well as AmE.

(11) I love how many ales I've **gotten** this year (S23A, BNC2014)

(12) If you hadn't interviewed, and **gotten** the job as a result of your stunning looks, then the

other person (GloWbE)

However, while the distinction between static and dynamic is an important aspect of examining how *gotten* is used in BrE, other categories such as the **change of state** category are also very common in my data, and deserve further exploration.

5.2.3 Other categories

It is clear from my data that the four main categories of usage of *gotten* are **acquisition**, **change of state**, **idiomatic** and **causation**. As the **idiomatic** category can overlap with other categories, it must be judged on its own, as it is in section 5.2.1, where idiomatic expressions involving *get* and its inflections are discussed. I will not be able to present exact numbers for **change of state** and **causation** tokens, as some tokens can be interpreted to belong to both of these categories. General trends of frequency and rough numbers are presented instead.

Causation seems fairly rare in relation to **acquisition** and **change of state**, but appears in certain formulations like (13) and (14). While **acquisition** seems to have been the most common category in the earlier datasets, in both the BNC2014, GloWbE, and COCA2014 datasets, **change of state** is more common. For example, in the GloWbE, **change of state** occurs over 50 times whereas **acquisition** only occurs 25 times. The BNC2014 and COCA2014 look very similar, both with around 60 **change of state** tokens and 26 **acquisition** ones. A frequently occurring and easily recognizable form of **change of state** is of the *gotten better* or *gotten worse* kind, such as (15) and (16). Another type of **change of state** token is of the (17) variety, though there are more types in addition to these. Considering the broad nature of this category, it is not surprising that it applies to so many of the tokens.

(13) someone else's life as well because he's never **gotten** to see long term consequences or had to make and educated important decision. (GloWbE)

(14) it's been awesome. Very blessed. Like you said, I've **gotten** to play with so many great artists, and become friends with so many great-- (COCA2014)

(15) to work, but technology has changed and cameras have **gotten** smaller, better and more consistent. (GloWbE)

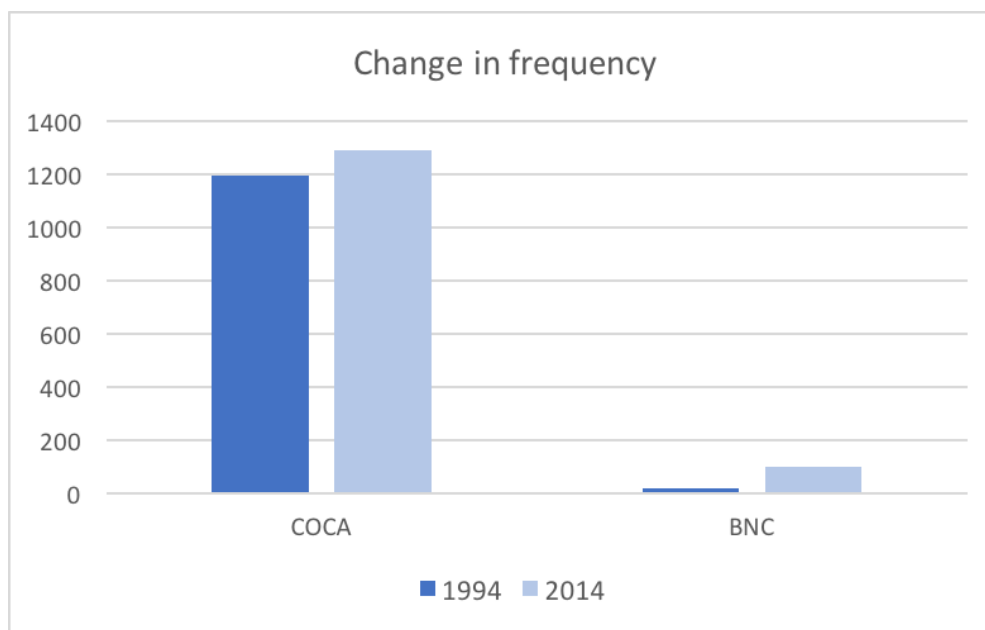
(16) it 's really (.) **gotten** darker over the years which is sad (STMM, BNC2014)

(17) I think that after you've **gotten** some of your editing done you could be ready to submit. (GloWbE)

Some of these tokens could have been exchanged for *got* tokens without any change in meaning. Even so, *gotten* is used. It is possible that this occurs less in BrE than in AmE, and that *gotten* is used more readily in BrE in cases where it achieves a meaning distinct from *got*. Interestingly, this could be considered directly in contention with Algeo's (2006) statement discussed in the previous section. However, this is speculation based on trends in the data which are difficult to quantify.

5.3 Processes of language change

There are six patterns of language change discussed in Leech et al. (2009). As Leech et al. state, several patterns can apply to one phenomenon, or apply only to a certain degree as opposed to fully. When it comes to *gotten*, it is possible to go back to the point of divergence between AmE and BrE to get an idea of the kind of change that has taken place since then. As we established, it was not as simple as *gotten* disappearing from BrE while becoming more prevalent in AmE. It took some time for *gotten* to lose its presence in BrE, and at the same time it was, at least by grammarians, frowned upon in AmE. Only after a certain amount of time did *gotten* become accepted in AmE. In other words, this can be considered an example of **divergent change**, where variants of the language eventually changed in opposite directions. But, in the past few years, *gotten* has seen notable increase in usage in BrE (80%, according to the spoken BNC2014), while it has increased only a small amount in AmE (12%, according to COCA). Important to note, however, is that these percentages indicate very different figures – as depicted in the diagram below.



c) This diagram illustrates the difference in frequency of *gotten* in these corpora. It shows *gotten* appearing 1193 times in COCA1994 and 1290 times in COCA2014, as well as 20 times in the BNC1994 and 101 times in the BNC2014. It is intended to illustrate that while the numbers themselves in the BNC corpora are small, the increase is large.

In the terminology of Leech et al., this can be interpreted as **convergent change**, where “BrE has changed in the direction of AmE, whereas AmE has changed only trivially” (2009, p. 253). Were *gotten* to continue to increase in frequency in BrE, it could also be seen as the pattern termed **different rates of change**, where the change in AmE slows down and allows BrE to “catch up” (Leech et al., 2009, p. 253). For BrE to truly catch up with AmE in this regard in the future, a paradigm shift in BrE usage would be required. If this were to happen, it would also be an example of the **follow-my-leader** pattern, where one variety of the language “takes the lead” (p. 253). To a certain extent, however, **follow-my-leader** already applies to what we can observe in this study, in that AmE and BrE are moving in the same direction, and one is further along in this development than the other.

The kind of change is also fairly clear in terms of the manner in which *gotten* is used. In my American data, usage of *gotten* in the **acquisition** sense has decreased, where usage in other senses has increased. In our BrE data, the ratio of **acquisition** to **non-acquisition** has also decreased, but slightly less. This would lead us to call this an example of **different rates of change**. Considering the difficulty in interpreting the BNC1994 data, however, this may have to be taken with a grain of salt. What we can say with relative certainty is that these results show that the current ratio of **acquisition** to **non-acquisition** in BrE – at least as indicated by

the data – is entirely reflective of the same ratio in AmE. This means that BrE usage of *gotten* mirrors the results of the broadening of usage that AmE has gone through. It is thus safe to say that, from a cursory inspection, *gotten* works in largely the same way in BrE, when it is used there, as in AmE.

As for how this may or may not change in the future, there is a potential clue in the corpora used in this study. Considering the diverse nature of these corpora, it can be important to note that while they are comparable in terms of size and year of origin, they consist of different levels of formal language. Additionally, both the BNC1994 and the GloWbE datasets contain written language whereas the rest of the datasets are only spoken. What does this mean for the results of this study? Leech et al. state that “many grammatical innovations start out in spoken English and then gradually spread into writing” (2009, p. 12). Spoken English and informal English also seem to go hand in hand, as in – for example – the spoken BNC2014, which consists of transcripts of conversations mainly in informal situations. However, the **blog** section of the GloWbE consists of informal written language, not spoken. These corpora both include use of *gotten*, an inflection that could be considered an overt Americanism (Fowler & Burchfield, 2004, p. 338). The fact that this word can be found in these BrE corpora is perhaps an indication that *gotten* could increase in frequency and continue to gain traction in spoken and informal language, and as Leech et al. say, this is the kind of thing that could eventually lead to the adoption of a lexical item in more formal, written settings as well.

6. Conclusions

In conclusion, *gotten*, an inflection considered emblematic of AmE, has seen an increase in usage in BrE over the last few decades. While it is still rare and considered nonstandard, the 80% increase observed in the BNC2014 indicates that *gotten* is being used in BrE today to an extent that it was not in 1994. As language change often starts in spoken English, such as the BNC2014, and moves into written English, it is possible that this increase could, in the future, lead to an increased usage of *gotten* in written BrE as well.

If the samples used in this study could be considered indicative of English as a whole, we could also conclude that *gotten* is being used slightly differently today compared to how it was used more than 20 years ago, both in BrE and in AmE. It more often takes on a meaning other than

the original **acquisition** meaning, such as the common and fairly broad **change of state** meaning. Perhaps there will be even further broadening of the term in the future.

The secondary goal of this study has to do with evaluating the advantages and disadvantages of this type of investigation in the context of regional language change. When it comes to this aspect, we have been able to see in this study that corpora which claim to be representative of a specific country's dialect can face significant challenges in making sure that their data stems from the desired dialect. English in particular, as a global language, seems to face this issue. My original thought was that this was a fairly new problem, brought about by the global nature of the internet, and corpora like GloWbE, which mainly make use of the internet. Looking at the BNC1994 has shown that this is not the case, as the older BNC1994 contains many instances of AmE – though perhaps for different reasons than a corpus like the GloWbE does. This, to me, shows that this is a challenge which perhaps faces all of these corpora. In corpus investigations where the frequency of the object of study is high, this may not be a particularly significant issue. In studies such as this, however, where this frequency is low, I believe it is essential to be aware of this and to investigate how it will affect the results.

As for the reason behind these results, it is difficult to measure why any particular Americanism makes its way into BrE. Possibly it is due to the increasing influence of AmE on English in the global scene. With this in mind, I believe there is room for a more substantial and broader study of Americanisms in BrE to investigate trends, possible generalizations, and how the prevalence of American culture can influence BrE. For such an investigation making use of corpus methods, it is my hope that this study can provide some useful background.

References

Primary sources

British National Corpus (XML edition). Retrieved from

<https://cqpweb.lancs.ac.uk/bncxmlweb/>

Davies, Mark. (2008-). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Retrieved from <https://corpus.byu.edu/coca/>.

Davies, Mark. (2013) *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE)*. Retrieved from <https://corpus.byu.edu/glowbe/>.

Love, R., Dembry, C., Hardie, A., Brezina, V. and McEnery, T. (2017). The Spoken BNC2014: designing and building a spoken corpus of everyday conversations. In *International Journal of Corpus Linguistics*, 22(3), pp. 319-344.

Spoken BNC2014. Retrieved from <https://cqpweb.lancs.ac.uk/bnc2014spoken/>.

Secondary sources

Algeo, J. (2006). *British or American English?: A Handbook of word and grammar patterns*.

“Americanisms: 50 of your most noted examples”. (2011, July 20). *BBC News*. Retrieved from <http://www.bbc.com/news/magazine-14201796>

Bernstein, T. M. (1965). *The Careful Writer: A Modern Guide to English Usage*. New York: The Free Press.

Biber, D., Conrad, S., Leech, G. (2002). *Longman Student Grammar of Spoken and Written English*. Harlow: Longman.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finnegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Fowler, H. W., Burchfield, R. W. (2004). *Fowler's modern English usage*. Oxford: Oxford University Press.

Golmann, M. (2008). *Investigating British and American English: Dictionary research and corpus investigation* (Thesis).

Gonçalves, B., Loureiro-Porto, L., Ramasco, J. J., Sánchez, D. (2017). *The Fall of the Empire: The Americanization of English*. Retrieved 25 Jan. 2018, from <http://arxiv.org/abs/1707.00781>

Gotten. (2016). In Butterfield, J. (Ed.). *Fowler's Concise Dictionary of Modern English Usage*. :Oxford University Press. Retrieved 22 Oct. 2017, from

<http://www.oxfordreference.com.ludwig.lub.lu.se/view/10.1093/acref/9780199666317.01.0001/acref-9780199666317-e-1666>.

- Gotten. (2016). In *Oxford English Dictionary*. Retrieved 22 Oct. 2017, from <http://www.oed.com.ludwig.lub.lu.se/view/Entry/80242?redirectedFrom=gotten&>
- Leech, G., Hundt, M., Mair, C., Smith, N. (2009). *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.
- Lowth, R. (1762). *A Short Introduction to English Grammar, with critical notes*. Cambridge: Folsom, Wells, and Thurston. Retrieved 20 Jan. 2018, from <http://www.bl.uk/learning/timeline/item126709.html>
- Marckwardt, A. H. (1958). *American English*. New York: Oxford University Press.
- Marckwardt, A. H., Walcott F. (1938). *Facts about current English usage*. New York: Appleton-Century-Crofts.
- Meade Bache, R. (1869). *Vulgarisms and Other Errors of Speech*. Philadelphia: Claxton, Remsen, and Haffelfinger.
- Murray, L. (1795). *English Grammar*. New York: Collins and Perkins.
- Tottie, G. (2009). How different are American and British English grammar? And how are they different?. In Rohdenburg, G., Schlüter, J. (Eds.), *One Language, Two Grammars? Differences between British and American English* (pp. 341-363). Cambridge: Cambridge University Press.
- Trudgill, P., Hannah, J. (2013). *International English: A guide to the varieties of Standard English*. Abingdon: Routledge.
- “What is the BNC?”. (2009). British National Corpus. Retrieved 1 Dec. 2017, from <http://www.natcorp.ox.ac.uk/corpus/>
- White, R. G. (1870). *Words and Their Uses, Past and Present. A Study of the English Language*. New York: Sheldon and Company.