The Americanisation of spoken British English

A corpus-based study of the diachronic and synchronic Americanisation of lexis in spoken British English

Lisa-Marie Morgner

A thesis presented to the Department of Literature, Area Studies and European Languages

UNIVERSITY OF OSLO

June 2021

The Americanisation of spoken British English

A corpus-based study of the diachronic and synchronic Americanisation of lexis in spoken British English

Lisa-Marie Morgner

Department of Literature, Area Studies and European Languages
Faculty of Humanities
University of Oslo

ENG4790 – Master's Thesis in English:
Secondary Teacher Training
30 ECTS Credits
Spring 2021

Supervisor: Kristin Bech

2021

The Americanisation of British English: *A corpus-based study of the diachronic and synchronic Americanisation of lexis in spoken British English*

Lisa-Marie Morgner

http://www.duo.uio.no

Print: Reprosentralen, University of Oslo

Abstract

This thesis presents a twofold investigation of the use of Americanisms in the lexis of spoken British English where both the diachronic and synchronic aspect of Americanisation is examined. Previous research has noted an influence of American English on other varieties of English, but few of them have studied the variety of British English. The aim of this thesis is to find out 1) whether there has been an increase in Americanisms in the lexis of conversational British English in the last twenty years, 2) whether the demographic variables of age, gender and class have an effect on the use of Americanisms, and 3) if there are differences between semantic domains. The study is corpus-based, using the *British National Corpus 1994* (BNC1994) and the *Spoken British National Corpus 2014* (Spoken BNC2014) to investigate the frequency of a selection of American English items. This study demonstrates that British English indeed shows signs of increased Americanisation over the last two decades. The analysis also shows that age, gender and class seem to have an effect on the use of Americanisms. Generally speaking, it is men, speakers aged 0–29 and speakers belonging to classes C2, D and E that reveal the highest frequency of American English items in their spoken language. Furthermore, findings reveal that the semantic categories 'Housing and Household', 'Food and cooking' and 'Transport' exhibit the highest frequencies of Americanisms, both diachronically and synchronically. There are several possible reasons for the Americanisation of spoken British English. In short, it is the continuing influence of the US on culture and media that maintains the influx of American English items into other varieties of English, including British English.

Keywords: Americanisation of English, Americanisms, Spoken British English lexis, Diachronic change, Synchronic variation, Demographic variables, British National Corpus, frequency, Corpus linguistics, Corpus-based comparative approach, Sociolinguistics, Variationist approach

Acknowledgements

First and foremost, I would like to thank my supervisor Kristin Bech, both for introducing me to the joys of English grammar in the first year of my studies, and for her constant guidance, valuable feedback and the amounts of time spent on helping me with this thesis.

I would also like to thank Signe Oksefjell Ebeling for inspiring the research topic of this thesis, and for introducing me to the universe of World Englishes and electronic corpora.

A great thank you to Siham, my best friend, who I met in the first year of my studies and without whom I barely spend a day ever since. Without you, both life and studies would not have been even half as fun as they have been. Thank you for always being there, spending hours and hours of studying together, surviving exams throughout the years, and thank you for hours of laughing and crying. You are awesome!

Thank you, Jon, my partner and best friend, for having infinite amounts of patience and calmness in times when I have not, and for always keeping a cool head. You are my rock!

Thanks to Elke Liebchen, my English teacher in grammar school, who sparked my passion for English and inspired me to become a teacher.

Finally, I would like to thank my family and friends for all their support and encouragement throughout the entirety of my studies.

Thank you.

Oslo, 15 June 2021

Lisa-Marie Morgner

Contents

List of tables

List of figures

List of abbreviation

AmE	–	American English
BrE	–	British English
OED	–	Oxford English Dictionary
RQ	–	Research question

Corpora mentioned

BNC	–	The British National Corpus
BNC2014	–	The British National Corpus 2014
BNC1994DS	–	The demographically sampled spoken component of the British National Corpus 1994
Brown	–	Brown University Standard Corpus of Present-Day American English
COCA	–	Corpus of Contemporary American English
GloWbE	–	Corpus of Global Web-Based English
LOB	–	Lancaster-Oslo-Bergen Corpus

1 Introduction

> The Englishman, whether he knows it or not, is talking and writing more and more American. . . . In a few years it will probably be impossible for an Englishman to speak, or even to write, without using Americanisms, whether consciously or unconsciously (Mencken 1930, as quoted in Algeo 1986, 270).

Even though the British and American peoples share a common language, there are a number of differences between the two varieties. This diversity has its origin in the separation of the two peoples after the arrival of the first settlers on the North American continent and can be traced at all levels of language, especially lexis (Algeo 1986, 271). However, already in 1930 Mencken observes that the two varieties seem to become more alike. He predicts what is nowadays considered to be one of the major ongoing processes of contemporary language change: the Americanisation of English (Gonçalves et al. 2018, 2). Much research focusing on British and American English has been concerned with examining differences between the two varieties. Therefore, this thesis will take another path, namely to study the convergence of the two major streams of English. With the help of large digital collections of naturalistic language data, so-called corpora (more on this in Chapter 4), it has become possible to perform research on the topic of Americanisation on a much larger and more precise scale than before. Previous studies such as Vine (1999), Igboanusi (2003), Hänsel and Deuber (2013), Gonçalves et al. (2018), Korhonen (2020) and Morgner (2020) have all noticed an influence of American English on other varieties of English. Most of them focus on earlier British colonies, but few of them (Gonçalves et al. 2018; Morgner 2020) investigate the earlier colonial power Great Britain itself.

1.1 Aim and scope

The aim of this study is to investigate whether there has been an increase in the use of Americanisms in conversational British English in the last two decades and whether the demographic parameters gender, age and class somehow have an effect on the use of Americanisms. In addition, the influence of American English on different semantic domains is investigated. By adding findings about spoken British English, examining different demographic variables, and examining semantic categories, I hope to add to the discussion of

the field of Americanisation of English. I have chosen to focus on the lexis of conversational British English due to the fact that it is vocabulary that shows most diversity (Algeo 1986, 271) and, at the same time, is most affected by the influence of American English (Korhonen 2020, 198). Also, there is a limited number of studies that carry out research on the lexis of British English by using naturally occurring oral data. To conduct this study, I have chosen to apply a corpus-based comparative approach for the diachronic aspect of this thesis, as well as a variationist approach to study synchronic variation. The data were collected from the demographically sampled spoken component of the *British National Corpus 1994* (BNC1994DS), and the *Spoken British National Corpus 2014* (Spoken BNC2014) (see Chapter 5). The study is mainly a quantitative one: Americanisms will be investigated in terms of their frequency and conclusions will be drawn based on these frequency counts. In addition, comparisons with earlier research will be made and possible reasons for the observed findings will be discussed. A detailed explanation of the method and corpora will be presented in Chapter 4 and 5. This thesis is inspired by Gonçalves et al.'s study (2018), which serves, together with the pre-study by Morgner (2020), as a point of departure. In their studies, they investigate the Americanisation of British English lexis, among other varieties, and find that there indeed is a drift towards the use of American English vocabulary. Therefore, this study seeks to investigate this issue even further, and focuses solely on the lexis of spoken British English.

1.1.1 Research questions

Corresponding to the study's aim, the thesis consists of two parts. Part one is concerned with the diachronic change over the last two decades, whereas part two of the study focuses on synchronic variation. Thereby, the following research questions emerge:

Part one:

> **RQ1**: Has there been an increase in Americanisms in the spoken British English lexicon in the last twenty years? If yes, what causes this increase?
>
> **RQ2:** Are there diachronic differences in frequency between different semantic domains?

Based on previous research (Vine 1999; Igboanusi 2003; Hänsel and Deuber 2013; Gonçalves et al. 2018; Korhonen 2020; Morgner 2020), which found that American English has an influence on other varieties of English, it seems natural to suggest that also British English shows sign of Americanisation over time. My hypothesis is, therefore, that there

indeed has been an increase in Americanisms in the spoken British English lexicon in the last two decades. If the quantitative analysis confirms this hypothesis, possible reasons will be discussed. Moreover, I will analyse whether there are differences in frequency between different semantic domains, and thereby reveal which domains have been influenced the least and the most by American English the last twenty years.

Part two:

> **RQ3:** Who uses Americanisms? Do the demographic variables age, gender, and class have an effect on the use of Americanisms in spoken language?
>
> **RQ4:** Are there synchronic differences in frequency between different semantic domains?

Few studies investigate the use of Americanisms in relation to different demographic variables which makes it difficult to predict the findings for part two. Nonetheless, based on previous research, and considering both the trend of Americanisation (Rayson et al. 1997; Vine 1999; Igboanusi 2003; Korhonen 2020) and other linguistics trends (Rayson et al. 1997; Schmid 2003), I hypothesise that there in general are differences between the groups along all three dimensions. More precisely, my hypothesis is that both younger speakers and speakers belonging to "lower" social grades are most inclined to adopt American English vocabulary. In addition, I analyse how the use of Americanisms in certain semantic domains is distributed between the groups within each variable.

1.2 Thesis outline

This study consists of eight chapters in total. Chapter 1 gives an introduction to the issue of this thesis, presents the aim and scope, in addition to the research questions and hypotheses for this study. In Chapter 2, a variety of previous studies which investigate the Americanisation of English varieties around the world are introduced. Moreover, previous research on the impact of demographic parameters on the use of language is presented. Chapter 3 provides some theoretical background on language change and variation in general, thereafter briefly explains the divergence of British and American English, and finally gives a short introduction to the contemporary linguistic trend of Americanisation. Chapter 4 takes a closer look at corpus linguistics as a method and the approaches applied. Furthermore, it comprises the classification of the material. In Chapter 5, the two corpora used in this study, the BNC1994DS and the Spoken BNC2014, are presented and compared. Chapter 6 gives a

presentation of the study's results and analyses for part one and two, which is followed by a discussion of the findings in Chapter 7. Lastly, Chapter 8 provides concluding remarks as well as some limitations of the study and suggestions for further research.

2 Previous studies

In the following chapter, a selection of previous studies relevant for the topic of this thesis will be introduced. First, previous studies about the Americanisation of English varieties will be presented, covering both Inner, Outer and Expanding Circle varieties. Further, section 2.2 is concerned with previous research on the impact of demographic parameters on the use of language, focusing on the variables of age, gender, and class.

2.1 Previous research on the Americanisation of English varieties

2.1.1 The Americanisation of English: Gonçalves et al. 2018

In their study, Gonçalves et al. (2018) investigate how varieties of English have evolved both across space and time by using two large datasets derived from Twitter and the Google book database. Their focus is on standard written English and its diachronic evolution, as well as on the synchronic spatial variation of colloquial language of microblogging messages. The aim is to investigate how vocabulary and spelling vary across countries and how the prevalence of either British or American English in these countries has evolved over the last two centuries. Their analysis shows that both vocabulary and spelling conventions have shifted from British to American, especially in the time following the Second World War, the end of the Cold War and the advent of the Internet. They found that an American influence is undeniable in Expanding Circle territories, as well as in Outer Circle and Inner Circle varieties. Only in countries where British influence has been dominant, such as the UK, Ireland and India, do British forms in both spelling and vocabulary still dominate to a certain extent. However, some ex-colonies such as South Africa, New Zealand and Australia clearly drift towards the use of American vocabulary.

2.1.2 Corpora and culture: Leech and Fallon 1992

Leech and Fallon (1992) examine vocabulary differences between American and British English using the Brown Corpus and LOB. Their study is based on Hofland and Johansson's (1982) parallel word frequency lists for the two varieties. Leech and Fallon's (1992) goal is to conduct a more systematic study of the vocabulary differences and, thereby, find evidence of cultural differences between the two varieties in Brown and LOB. Applying a corpus-based approach, regarding the two corpora as reliable sources of "comparative information about varied social, political, and cultural aspects" (Leech and Fallon 1992, 29) of the two varieties, they also state three limitations of the method:

a. The corpora are restricted to written language.
b. The corpora are restricted to one historical period (1961).
c. The size of the corpora. Each corpus contains only a million words.

Leech and Fallon (1992) divide their findings into either linguistic (spelling and lexical choice) or non-linguistic contrasts (proper nouns and 'other'). Moreover, focusing on the 'other' non-linguistic items only, they subdivide this category into fifteen domains reflecting cultural differences. Their results show, though cautious and tentative of generalising, that the corpora produce evidence of cultural contrast: US culture in 1961 appears to be masculine and driven by enterprise, while British culture appears to be more temporizing and concerned with family life (Leech and Fallon 1992, 44–45).

2.1.3 Pre-study: Morgner 2020

In the pre-study for this master thesis, Morgner (2020) investigates the ongoing process of Americanisation of contemporary English, considering vocabulary, in three Inner Circle varieties: British English, Irish English and New Zealand English. The study's hypothesis, based on the study conducted by Gonçalves et al. (2018) presented in section 2.1.1 predicts that even Inner Circle varieties show signs of Americanisation, especially in oral language. By using the *International Corpus of English* (ICE), differences between spoken and written speech and differences between semantic domains were examined. Morgner's analysis shows that all three varieties display a higher frequency of Americanisms in spoken language than in written language, which could be due to the dichotomy between a media presence dominated by American English and formal education subscribing to British English. Moreover, it was found that the semantic domains of 'Transport', 'Food and household' and 'Legal terms' were influenced by American vocabulary the most.

2.1.4 The Englishes of Kenya, Singapore, and Trinidad and Tobago: Hänsel and Deuber 2013

In their study, Hänsel and Deuber (2013) analyse the spelling and vocabulary of three newspapers published in the former British colonies Kenya, Singapore, and Trinidad and Tobago. Their aim is to learn whether the degree of Americanisation in postcolonial Englishes is somehow connected to the countries' participation in economic and social globalisation. Their analysis reveals that British English spelling is preferred in all three varieties, whereas American English is found to have stronger influence on vocabulary

depending on the semantic domain. Moreover, Hänsel and Deuber (2013) find the following for the three countries:

a. Kenya being the least globalized country among the three also exhibits the smallest number of Americanisms in the data. American cultural and social influence through for example the internet, television and music is largely restricted to urban areas, allowing mostly the young residents to be exposed to American English.
b. The data for Singapore show the use of both British and American forms. As the country is highly globalized, an increasing number of Americanisms can be observed. However, there are still strong ties to British English as well.
c. Both American and British spelling and vocabulary are found in the data collected for Trinidad and Tobago. However, the distribution depends on the semantic domain the word belongs to. The country has strong ties to the US, which is reflected in both the economic and cultural Americanization there.

In conclusion, Hänsel and Deuber (2013) find that American English influence is due to a number of reasons, such as the exposure to Americanisms and American cultural exports. Moreover, globalization appears to be a driving force for importing Americanisms into other varieties of English.

2.1.5 New Zealand English: Vine 1999

In her study, Vine (1999) examines the lexical shift in New Zealand English, which has its origins in British English, from British to American words, and investigates the assumption that New Zealanders avoid terms that they know are of American origin. Vine (1999) bases her discussion on data gathered in a survey exploring New Zealanders' perceptions and attitudes towards American terms, from which she draws conclusions about their usage of these terms. Her results show that neither the speakers' perception nor their attitude indicates an avoidance of American terms or an intolerance towards them. This positive attitude towards and choice of American terms was especially found among the younger participants. Moreover, Vine (1999) suggests that certain semantic areas, such as the entertainment industry, are more influenced by the US and, thus, more dominated by American terms. Lastly, as none of the lexical pairs examined in her study can be said to be strict synonyms, Vine (1999) concludes that speakers of New Zealand English adopt American terms out of a need for new terms rather than replacing one British term with its American equivalent.

Therefore, she argues that the adoption of American English in New Zealand is not a sign of lexical shift but rather of lexical borrowing.

2.1.6 Nigerian English: Igboanusi 2003

Igboanusi (2003) examines the influence of American English in Nigeria and investigates both the knowledge, usage and attitudes among speakers of Nigerian English (NE). In accordance with other research, such as Vine (1999), he implies that the influence of American English on other varieties of English is a worldwide trend and not only restricted to Nigerian English. As the acceptance for American English in native-speaker communities around the world grows and the use of Americanisms becomes a popular choice among new generations, the American variety of English gains influence and attracts attention both among native and non-native speakers. In his study, Igboanusi (2003) found that among other things cable television, movies, music, exposure through travels, and the availability of printed materials in American English contributed to the increased number of Americanisms in Nigerian English. Focusing mainly on the lexical level, his analysis shows that everyday vocabulary is influenced by American English the most and that the use of American English vocabulary seems to steadily increase. However, speakers of Nigerian English do not adopt Americanisms because they are American; they are rather influenced by the huge influx of American English without being aware of the terms' origins. Vine (1999), presented in section 2.1.5, observes the same for speakers of New Zealand English. Finally, Igboanusi (2003) concludes that the attraction towards and the acceptance of American English in Nigerian English, in addition to the unrestricted influx of American culture and language, will lead to Nigerian English being American-dominated.

2.1.7 Australian English: Korhonen 2020

Korhonen (2020) investigates linguistic features in Australian English that have been influenced by American English, as well as the public's reactions and comments regarding these features. Her findings show that Australians' attitudes towards the Americanisation of Australian English are mostly negative both among older speakers and younger speakers. Moreover, the participants experience the largest influence of American English in the areas of vocabulary and spelling, followed by pronunciation. Television is regarded as the main reason for this Americanisation of Australian English.

Korhonen (2020) also examines some of the American English linguistic features that are currently used in Australia. Starting with the area of vocabulary, where American English input is most prominent, she finds that a number of Americanisms have been transferred to

Australian English. Many of these Americanisms, however, become part of Australian usage and are not considered American anymore after a while. So rather than randomly adopting new words, Australians accept Americanisms into Australian English only when these, for example, fill a gap. Especially due to the advent of the Internet, American vocabulary has become part of Australians' everyday life, and the Internet will uphold the influx of Americanisms in future. Considering vocabulary, it is often the younger speakers that readily adopt Americanisms. As for the areas of spelling, pronunciation and grammar, some patterns that are perceived as American are being used in Australian English. However, none of these indicate a large-scale Americanisation of Australian English.

Korhonen (2020) concludes, then, that even though American English influences certain linguistic features, it is not replacing traditional AusE usage. Rather, Americanisms are adopted to fill a gap, especially in informal speech, and become quickly nativized, thereby loosing "their American flavour" (Korhonen, 2020, 200).

2.1.8 Concluding remarks

In this thesis, the studies introduced above will be used as reference points regarding both results and method. Some of the studies focus both on written and spoken language. In this thesis, however, the focus will be narrowed to oral speech and vocabulary in order to provide a more comprehensive outline of this area, as few studies carry out their research on naturally occurring spoken data. In addition, both the diachronic and synchronic dimension of the use of Americanisms will be analysed in this thesis. This will be an addition to earlier research, as only Gonçalves et al. (2018) examine the evolution of this phenomenon over time. Moreover, each of the studies introduced examines at least one variety of English, mostly Inner and Outer Circle varieties, whereas only two (Gonçalves et al. 2018, Morgner 2020) also include the Inner Circle variety of British English. The present study focuses solely on English spoken in Great Britain. Arguably, it is meaningful to compare data from the previous studies investigating other Inner and Outer Circle varieties, especially postcolonial varieties, with data found for British English, as one might expect some similar traits of development in these varieties. Furthermore, some studies (Vine 1999; Hänsel and Deuber 2013; Morgner 2020) investigate the influence of American English on certain semantic spheres. In this study, I will examine the items' semantic domains even more thoroughly by assigning each item to one semantic category, thereby, revealing a clearer picture of which domains are influenced the most by American English.

Regarding method, most of the presented studies draw on collected speech material

such as newspapers, radio, television and corpora, to investigate the American English influence on a respective variety. The present study applies a corpus-based approach inspired by earlier research (Leech and Fallon 1992; Gonçalves et al. 2018; Morgner 2020) to investigate cultural change and variation. As limitations of earlier studies are considered, larger corpora containing spoken data are used, and corpora from two time periods are compared, this study will enhance and broaden the field of Americanisation.

2.2 Previous research on the impact of demographic parameters on the use of language

2.2.1 Rayson et al. 1997

In their study, Rayson et al. (1997) examine the spoken part of the *British National Corpus* (BNC) considering the speakers' and respondents' gender, age and social group in order to investigate the relationship between those social factors and the participants' use of language. Their results for gender, age and social group variation show among other things the following:

a. Gender variation: There are distinct differences between the speech of the two sexes. However, this variation is not absolute and rather a matter of frequency.
b. Age variation: British speakers under 35 seem to follow American English conventions in certain lexical choices, whereas older speakers seem to prefer predominantly British vocabulary, for example in the choice of swearwords.
c. Social group variation: Speakers of the "upper" group (A/B/C1) tend to use less third-person pronouns and taboo words than speakers among the C2/D/E group.

In conclusion, Rayson et al. (1997) find significant differences in the use of language along all three social dimensions and, thus, they show how the BNC can be applied for studying social differentiation in the use of vocabulary.

2.2.2 Schmid 2003

Schmid (2003) seeks to demonstrate how corpora can be used to investigate sociolinguistic issues, inspired by earlier work carried out by Leech and Fallon (1992, see section 2.1.2). He does so by focusing on differences in speech between men and women and examines whether men and women "live in different cultures" (Schmid 2003, 3). Schmid (2003) uses two subcorpora, speech by males and females respectively, taken from the spoken part of the BNC. In addition, he investigates both the domains of conversational behaviour and different

semantic fields, complementing Rayson et al.'s study (1997). These semantic fields are expected to be either dominated by men, such as car and traffic, sports and work, or women, such as clothing, home, food and drinks. Schmid (2003) finds that most words are used with different frequencies by men and women and that scores in most domains conform to stereotypical uses of words by males and females. He assumes that these frequency scores reveal more than just differences in language use, and argues, in accordance with Leech and Fallon (1992), that corpora can be linked to culture and, thus, reflect cultural differences and trends about the two sexes.

2.2.3 Concluding remarks

The second part of this study is concerned with demographic variables influencing the use of Americanisms in spoken British English. Both Rayson et al. (1997) and Schmid (2003) apply a corpus-based approach in their research using the spoken part of the BNC1994. This thesis adopts the same method to examine social differentiation, however, using the newer BNC2014. More specifically, I gather data from the BNC2014 core subcorpus, a smaller and more balanced corpus, to achieve higher representativeness across the demographic categories (see Chapter 5). Following Rayson et al. (1997), this study investigates the three demographic variables age, gender and class separately. In addition, I will investigate items in relation to their semantic domain, similarly to what Schmid (2003) does. However, previous research in this area does not cover the influence of demographic variables on the use of Americanisms in British English. Therefore, the present study provides new data on this aspect of Americanisation.

3 Theoretical framework

This chapter provides an overview of theory relevant for the issue of this thesis, starting with an outline of the nature of language change and variation, focusing on lexis, in 3.1. Section 3.2 offers a short overview of how British and American English developed historically and, eventually, diverged from each other. Moreover, contemporary language change and the trend of Americanisation in British English are explored in 3.3. Lastly, we take a closer look at the terms *sociolinguistics*, *variation*, and *change* in 3.4.

3.1 Language change and variation

In order to understand how British (BrE) and American English (AmE) came to be different, it is essential to look at reasons why languages in general vary and change. There are a number of reasons for the development of linguistic variation within a language: spatial and temporal divergence, social mobility, social contact, urbanization, among others. Language change, on the other hand, is usually brought about by either internal (endogenous) or external (exogenous) factors but can also be explained with social factors. Variation within a language is typically found in spoken texts, as written language usually conforms to some sort of formal standard. Variation can be found at all levels of language, such as syntax, phonetics, or morphology (Melchers et al. 2019, 11–14). For the purpose of this thesis, the focus will be on variation within the lexicon (vocabulary), which will be outlined in section 3.1.1.

3.1.1 Variation in lexis

All varieties of English, including BrE and AmE, share a common set of vocabulary based on "a common body of knowledge and a common set of texts" (Melchers et al. 2019, 22). Nonetheless, there is considerable variation in lexis within and between varieties of one language, especially when considering content words. In order to investigate whether BrE has become more like AmE, it is useful to understand in which ways lexis can vary. To characterize this kind of variation we need to regard the two dimensions of form and meaning and the types of variation that arise. The categories, descriptions and examples are adopted from Melchers et al. (2019, 24–26).

1. **Localism**: words that are mainly used in a particular variety, such as Nigerian *foofoo* (pounded yam) and South Asian *crore* (ten million, often as a unit of value).
2. **Local but widely and unambiguously used in General English**: words that are of local origin but are widely known in other varieties, such as Australian *kangaroo.*
3. **Cross-variety synonymy**: words that have different forms for the same concept in different variations, such as *thumb back* in AmE and *drawing pin* in BrE. Words such as these are also called 'heteronyms' (Görlach 1995, in Melchers et al. 2019, 25).
4. **Cross-variety polysemy/homonymy**: words that have the same form but refer to different concepts in different varieties, such as *robin* which refers to different birds in different regions.
5. **Both synonymy and homonymy/polysemy**: words that can be both polysemous and synonymous at the same time, such as *pants* which is synonymous with *trousers*, and at the same time polysemous, with the meanings 'a pair of trousers' and 'underwear'.

Most of the items investigated in this study belong to Melcher et al.'s (2019) category three. Such words (AmE/BrE) are, for example, *truck/lorry*, *eggplant/aubergine*, *garbage/rubbish*, and *dessert/pudding*. Most lexical differences between AmE and BrE fall under this category (Trudgill and Hannah 2017). Moreover, the study covers words belonging to category five, such as *subway/underground* and *hood/bonnet*.

Lastly, it is necessary to have a look at how variation in vocabulary between varieties develops. The source of variation can either be intralingual or interlingual. Intralingual means that changes happen within a language. In English, for example, there existed two words describing one concept, such as the pair *autumn* (BrE) and *fall* (AmE). At the time when the two varieties separated, they happened to accept different words as the unmarked form ("fall, n.2", OED Online 2021). Other examples of this kind that will be investigated in this thesis are the word pairs *railroad* (AmE) and *railway* (BrE), and *sidewalk* (AmE) and *pavement* (BrE). The other type of source of variation between varieties is interlingual, i.e. changes that happen between languages, and involves borrowing where one language or variation borrows a variant from another one and adopts it (Melchers et al. 2019, 26–27).

3.2 The divergence of British and American English

Starting in Elizabethan times around 1600 with the arrival of the first settlers on the North American continent, BrE and AmE began to develop individually, and thus diverged, being

influenced by different cultures, communities, and events (Strevens 1978, 27). In the following, the development of the varieties will be investigated more closely, with a focus on lexical divergence.

BrE has primarily changed as a consequence of borrowings and coinings. It has been influenced greatly by other languages, coining new words and borrowing terms from Latin and Greek, but also from foreign countries within the British Empire (Strevens 1978, 29–30). As Strevens (1978, 29–31) points out, English in general has always had the habit of borrowing items from different sources as long as they are considered useful. More recently, AmE has also become a source of borrowing for BrE. This issue will be discussed further in section 3.3.1.

AmE, on the other hand, emerging as a new variety of English, has primarily developed through contact with other cultures and foreign language communities. After the first settlers had arrived, new words had to be found for places, plants and animals that did not exist in Britain and which therefore needed naming. The biggest source of such new words were American Indian languages, enriching AmE with words such as *chipmunk*, *igloo*, *kayak* and *tepee*, to mention a few. Other foreign languages that influenced AmE were French, Spanish, and Dutch, as these nations gradually colonised different parts of the continent. Such borrowed words that found their way into AmE are, for example, *pumpkin* from French, *ranch* from Spanish and *Santa Claus* from Dutch. Another source of influence were African languages, but despite the huge numbers of African slaves, only a few words of African origin are found in AmE, such as *okra* (a plant) and *voodoo* (a religion). Lastly, more recent immigration left its mark on AmE both culturally and linguistically, which is demonstrated in, for example, Italian (*pizza*, *spaghetti*) and German (*hamburger*, *semester*) borrowings (Strevens 1978, 31–35). The second source of AmE development originates in the growth of the modern American community. The establishment of new political institutions and governmental systems, but also new inventions and innovations, created a number of new words and expressions. Many of these coinages have also spread to other varieties of English, making AmE a leading force for distributing new words throughout the world (Strevens 1978, 40–42).

3.3 Contemporary language change: The Americanisation of British English?

3.3.1 Americanisation

At the time of the first settlements in North America, it was BrE that influenced the new American variety. However, since the US declared independence (Strevens 1978, 42) and the first dictionary for American English was published by Noah Webster in 1828 (Gonçalves et al. 2018, 10), AmE has developed its own character throughout the centuries. At the present day, the American variety has lost its 'colonial' status and has acquired equal status with the British one. It even seems that the direction of influence has changed in that AmE has become a source of, for example, new lexical items for BrE and other varieties (Strevens, 1978, 42). In the following, this development and the term *Americanisation* will be examined more closely. First, the term will be considered in its wider sense, and then it will be approached in regard to lexis by looking at a number of definitions.

Starting out with an elaboration of Americanisation in its non-linguistic sense, Bell and Bell (1998) define the term as the process of assimilation by immigrants to the US concerning "language, citizenship, customs and ideals" (Hill 1919, 612). Moreover, Americanisation started to spread to other nations, initially in the form of values and ideals such as democracy and liberalism, and later by the extension of America's military, economic, industrial and cultural influence throughout the world, especially after the Second World War (Bell and Bell 1998, 2–3). The determined effort of the US not only to promote their ideas and values across their own borders, but also to expand their global market and generate a demand for US developed goods and technologies, resulted in a projection of the American lifestyle abroad, and to the US becoming a global superpower. Acknowledging America's large impact on certain areas such as military affairs, global economy and commercial culture, Bell and Bell (1998) also highlight that not all ongoing internationalisation is equal to Americanisation (Bell and Bell 1998, 4–5). Economic power elsewhere and cultural change in general throughout the world needs to be considered as important factors contributing to globalisation as well (Desmond 2012, 3). Moreover, the term *Americanisation* is often used in the sense of "an assumption concerning the origins of a cultural example (language, dress, food) which may or may not be accurate" (Bell and Bell 1998, 5) and often carries connotations of rejection and dislike of assumed 'Americanisms' (Desmond 2012, 3).

If we zoom in on language, Americanisation is one of several possible forces of diachronic change in BrE (McEnery and Hardie 2012, 102). However, most of these forces or

trends, such as grammaticalisation and democratisation[1], are relevant for describing changes in grammar or discourse rather than lexis. Moreover, some trends, such as colloquialisation, a tendency of written speech to adopt more speech-like features (McEnery and Hardie 2012, 101), are more concerned with changes in written language than spoken language. Therefore, in accordance with the aim of this thesis, the focus will be on the sociolinguistic trend of Americanisation with emphasis on the lexis of spoken BrE, putting aside other possible forces behind lexical change in Britain. In the following, four definitions of Americanisation will be presented, starting off with the simplest one by McEnery and Hardie (2012):

1. Americanisation is "a tendency for British English to follow changes in American English" (McEnery and Hardie 2012, 101).
2. Trudgill and Hannah (2017, 94–95) explain lexical Americanisation as the continuous traffic of AmE words into other varieties of English caused by the influx of American culture through film and music, thereby not only providing new items, but even replacing local words such as *wireless* (BrE) with *radio* (AmE).
3. Melchers et al. (2019, 208–209) describe the process of Americanisation as the influence of AmE on other English varieties due to its high prestige caused by America's continuous power and media presence. They highlight, however, that this influence is least pronounced in Inner Circle varieties. Nonetheless, even the lexis of these varieties shows signs of steady AmE influence.
4. Lastly, Trudgill (2002, 148–149) defines the lexical Americanisation of English as the "homogenisation in the direction of North American usage" (148). He explains this contemporary trend with US domination in media and, thereby, constant exposure to American words which are being picked up by speakers of other varieties. He concludes by suggesting that a homogenisation of English indeed is taking place slowly but steadily, at least at the level of lexis.

In contrast, especially to Trudgill (2002), Leech et al. (2009, 258–259) claim that even though AmE is spreading to other varieties of English, this will not lead to a homogenisation of English. Rather, they observe that AmE forms are being adopted, but function differently in different sociolinguistic environments. Not denying the strong influence of AmE on other

[1] Democratisation is defined as "the social tendency for people to avoid 'unequal and face-threatening' ways of expressing a particular meaning; that is, a move away from explicit linguistic marking of social power relations" (McEnery and Hardie 2012, 102).

varieties, they describe a development shaped by both globalisation and diversity.

To sum up and extract the essence of the definitions above, Americanisation, a contemporary sociolinguistic trend, is described as the influence of AmE on other English varieties. This influence is ascribed to America's continuing power in the domains of especially media, economics and politics, becoming most visible on the level of lexis. Nonetheless, it is important to note that all language change cannot be explained by Americanisation, nor does an adoption of AmE items mean that local features are entirely replaced and disappear.

3.3.2 What is an Americanism?

As this thesis aims to investigate Americanisms in contemporary spoken BrE, it is necessary to close this section with a definition of the term *Americanism*. The Oxford English Dictionary (OED) defines the term as such: "A word, phrase, or other use of language characteristic of, peculiar to, or originating from the United States" ("Americanism, n.", OED Online 2021). As early as in 1781, J. Witherspoon (Pennsylvania Journal 1781, quoted in OED Online 2021) says the following about Americanisms in BrE:

> The first class I call Americanisms, by which I understand an use of phrases or terms, or a construction of sentences, even among persons of rank and education, different from the use of the same terms or phrases, or the construction of similar sentences, in Great Britain.

However, as Tottie (2002, 94) points out, many items that could be considered 'Americanisms' do not remain typical of AmE for very long nowadays, as they become adopted by other varieties of English more quickly than before. In this thesis, a number of Americanisms will be studied, and the aim is to find out whether there has been an increase of Americanisms in spoken BrE over the last twenty years. As the results will show, Tottie's (2002) notion about the short lifespan of Americanisms these days is arguably true for some of the items investigated.

3.4 Sociolinguistics: Variation and change

This thesis aims at investigating whether BrE has become more like AmE on the level of lexis and is concerned with both diachronic and synchronic variation in spoken BrE due to the sociolinguistic trend of Americanisation. The second part of this thesis specifically

focuses on variation in the use of lexical Americanisms related to demographic variables such as age, gender and class. In this section, a short introduction to relevant aspects of sociolinguistics will be presented.

Starting with a definition, Labov (1976, 183) defines *sociolinguistics* as "linguistic research which focuses upon language in use within the speech community". There are several different fields within sociolinguistics, but what they all have in common is the focus on the social aspects of language and on what ways the identity of a speaker influences their use of language. For the aim of this thesis, it is especially relevant to see how demographic variables such as age, gender and class determine the way in which people use language (Baker 2010, 2).

Moreover, in regard to the focus on both the diachronic and synchronic dimension of Americanisation, the terms *variation* and *change* need to be elaborated on briefly. In this thesis, variation will be considered both in terms of demographic variables and variation over time, i.e. diachronic variation. Baker (2010) also uses the term *change* for the latter type of variation and reserves the term *variation* for synchronic differences. The first part of this study is concerned with the type of variation where one is interested in possible changes over a timespan in a particular population or location (Baker 2010, 4–5). More precisely, in regard to this thesis, the question is whether there has been an increase of Americanisms in the spoken BrE lexicon over the past twenty years, and in which semantic domains the increase is most pronounced. Thus, the focus is on diachronic change in terms of lexis.

The second part of this study focuses on the characteristics of the speakers investigated, rather than on change in time, i.e. synchronic variation. It is concerned with variation occurring among different "speakers who exist at the same point in time, though they may not share the same identity variables or live in the same location" (Baker 2010, 4). In this thesis, the identity variables examined are age, gender and class. Moreover, the characteristics of the words investigated, in this case the semantic domain of the words, will be analysed. The topic of studying synchronic and diachronic variation with the help of corpus linguistics will be elaborated on further in Chapter 4.

Lastly, a word on the relationship between change and variation. Baker (2010, 5) states that a particular linguistic feature, such as Americanisms, needs to be used by someone in the first place before it eventually spreads to the rest of the population. Thereby, variation usually precedes change. However, not all variation leads to change. Part of a population may start to use a particular linguistic feature, but this does not necessarily mean that over time

other groups of the population will adopt this feature. He also mentions that some linguistic forms may be absorbed by a larger part of the population, while other features "operate as ways of differentiating between social or demographic groups" (Baker 2010, 5). This study's aim is to investigate whether the use of Americanisms has increased over time, which semantic domains are influenced the most, and whether it is certain social and demographic groups within the population that use Americanisms.

4. Method

In order to investigate the use of Americanisms in spoken British English, data have been collected using large corpora. In the following, corpus linguistics including relevant methods will be briefly introduced, followed by a more careful presentation of the applied approaches. Moreover, the extraction and classification of the data collected for this study will be outlined.

4.1 Corpus linguistics

McEnery and Hardie (2012, i) define corpus linguistics as "the study of language data on a large scale". This language data, then, is what constitutes what we call a *corpus*: "a (usually) very large collection of naturally occurring language, stored as computer files" (Baker 2010, 6). In short, corpus linguistics deals with the computer-aided analysis of large amounts of naturalistic language data. Moreover, there are other considerations that need to be taken into account when working with corpora, such as the issues of representativeness, balance, and sampling, which will be elaborated upon in Chapter 5.

To examine those large collections of data, different tools or methods need to be applied. In the following, the two tools used in this study, namely frequency and concordance, will be explained. In addition, an elaboration of the two approaches applied for examining the research questions of the thesis is provided.

4.1.1 Corpus tools: Frequency and concordance

The first tool, frequency, is the most basic tool in corpus linguistics and simply refers to the number of occurrences of a particular item in a corpus. This can be useful for detecting sociocultural differences, for instance (Gries 2017, 12–14). However, the frequency of an item alone reveals nothing of great meaning unless it is compared to something. For example, knowing that the lexical item *cookie* is said 18 times by speakers aged 0–39 in the BNC does not tell us much about BrE in general. However, if we also find that speakers of the age 40–99 say *cookie* five times, there is some basis for a comparative analysis which might reveal something about sociolinguistic differences. What is important to mention here is that raw frequencies can only be compared if we deal with the same amount of data of the two groups or corpora compared. Therefore, it is meaningful to rather calculate and compare standardized frequencies per million words (Baker 2010, 19–20).

The second method, the concordance, is a tool that displays every instance of the

investigated item in its wider linguistic context within the corpus. This process is also referred to as Key Word In Context (KWIC) (Gries 2017, 18). For this study, concordances are especially important in that they open up the possibility of checking whether "words have the meanings or uses that we claim they do" (Baker 2010, 21). This is essential for answering the thesis' research questions about language change and variation: The analysis of the data collected, and the claims made about language based on this data rest on the items' frequencies. Consequently, it is crucial that the frequency count is correct, which, then, is dependent on the analysis of the items' concordances (Baker 2010, 21). The limitation of this comprehensive and explicit tool is that concordances need to be analysed manually (Gries 2017, 19), which is the reason why some word pairs (AmE/BrE) such as *bus/coach* and *pants/trousers* were excluded from the study, as it is too time-consuming to check the meaning of, for example, 695 concordance lines for {bus/N} by hand.

4.1.2 Diachronic change: A corpus-based comparative approach to the lexis of British English

The first part of this thesis concerns how the Americanisation of spoken BrE has changed over time. Studies of diachronic change like the present one have often been conducted by applying a comparative corpora approach and have usually investigated written data due to the unavailability of large enough spoken corpora. At the time of Baker (2010), for example, no equivalent large spoken corpus which could be compared to the original BNC was available (McEnery and Hardie 2012, 118; Baker 2010, 57). However, the sought-after equivalent in form of the BNC2014 came in 2017 and is used in this study.

The two corpora investigated in this study, namely the BNC1994DS and the Spoken BNC2014, are used to collect data. These corpora share the same design, or sampling frame and can therefore be said to be comparable (more on the corpora in Chapter 5). According to McEnery and Hardie (2012, 20), a comparable corpus is "a corpus containing components that are collected using the same sampling method". This refers, for example, to the size of the corpus, the approach of collecting data, proportions, types of texts, and other parameters (McEnery and Hardie 2012, 20). For RQ1, whether there has been an increase of Americanisms in the spoken BrE lexicon over the past twenty years, a sufficient amount of quantitative data collected from two different time periods is needed in order to verify any claims made about the increase of Americanism. Moreover, the type of data from the two time periods needs to be equivalent, which is the case for the BNC1994DS and the Spoken BNC2014, in order to be able to make any valid comparisons. Therefore, it makes sense to

apply a comparative corpus approach which enables us to investigate diachronic change in the last twenty years (Hundt 2001, 739–741).

Furthermore, corpus linguistics is used as a method in this study. The present study is corpus-based, meaning that corpora are used in order to validate and exemplify the thesis' research questions, which are all based on previous studies and theories about the Americanisation of English. In other words, the corpora are not approached "from a completely naïve stance" (Baker 2010, 8), which would be the case for corpus-driven studies in which corpus evidence is used to formulate hypotheses rather than explore already established ones (Tognini-Bonelli 2001, 11; McEnery and Hardie 2012, 6).

The limitation of using corpora for this type of research question is that corpora can only provide the researcher with quantitative data. Quantitative data can be very useful in finding an answer to whether there has been change considering frequency. However, this type of data does not tell us much about other factors, such as speakers' attitudes, which cannot be measured in numbers, but which may (partially) cause the change we observe (Leech et al. 2009, 257).

4.1.3 Synchronic variation: Sociolinguistics and corpus linguistics combined – a variationist approach

Both parts of this thesis focus on the lexical Americanisation of spoken BrE and therefore study a sociolinguistic phenomenon. Part one focuses on diachronic change, as mentioned above (see 4.1.2). Part two on the other hand, focuses on variation in contemporary language rather than on change over time and aims at investigating the individual speaker and variation across different demographic variables (McEnery and Hardie 2012, 94). Compared to part one, corpus data from only the Spoken BNC2014 and, thus, from the same time period will be compared. In short, I study synchronic variation applying a variationist approach.

The variationist approach is one way of exploring sociolinguistic issues, especially if demographic variation across different populations is considered, as in part two. This approach, in the tradition of Labov (1969, 1972), seeks to examine language use by looking at specific linguistic variables, such as lexis or grammar, and by categorizing the language users according to demographic variables such as gender, age, and class. In other words, the focus is on the individual speaker and the variation across different demographic categories. Often, several categories in combination, such as age *and* gender, are taken into account (Baker 2010, 31; McEnery and Hardie 2012, 94). However, due to the great number of items investigated, this study looks at the three variables of age, gender, and class individually.

Corpus linguistics has not played a huge part in exploring issues within variationist sociolinguistics since corpus linguistics usually tends to focus on variation at the level of text rather than at the level of the speaker. However, with the advent of larger spoken corpora which include speaker metadata, as the spoken parts of the BNC1994 and the Spoken BNC2014 do, focus on variation among individual speakers in corpus linguistics grows (McEnery and Hardie 2012, 116). There are several reasons why corpus linguistics can be useful in exploring sociolinguistic issues. First, the two methods overlap in terms of some basic principles. One of them is that "[b]oth approaches entail the collection and analysis of naturally occurring language data […]. Both place a great deal of emphasis on language-in-use or social context" (Baker 2010, 8). Moreover, both corpus linguistics and sociolinguistics make use of quantitative data in order to explore similarities and differences between different populations, sample data so as to be able to generalise to a wider population, study variation and change, consider different linguistic features such as lexis and grammar, and aim at explaining the results of their research.

The advantage of using corpus linguistics to aid in the study of sociolinguistic issues is, then, that it provides sufficient amounts of quantitative data (Baker 2010, 8–9), and "combines the large numbers of real participants used in elicitation studies with more varied samples of language from each participant, as used in recorded data" (Baker 2010, 33). Nevertheless, Baker (2010) also notes that some sociolinguists view corpus linguistics as being too focused on quantitative data and admits that "classifications based on concepts like ethnicity, social class, sexuality and sex *can* be problematic, resulting in over-simplifications, stereotyping or reinforcing prejudice" (Baker 2010, 9). Moreover, he advises to exercise caution in focusing too much on differences at the cost of similarities between different speakers (Baker 2010, 56). In accordance with that, Oakes and Farrow (2007, 12–13) recommend being careful in regarding differences in data as equalling cultural differences. Another issue that Vine (1999) addresses is that low frequency items, such as most of the investigated items in this study are, produce either none or few hits in corpora, and might be dispersed unevenly between one or few speakers. In addition, low frequency items may appear in different contexts and might, therefore, not always correspond to the meanings and uses that we expect them to have (Vine 1999, 15).

4.2 Extraction of the material

The words included in this study, the Americanisms, were chosen on the basis of a number of previous studies introduced in Chapter 2, as well as from the online dictionary *Lexico*, which is a collaboration between Oxford University Press and Dictionary.com (Lexico.com 2021). The aim was to pick AmE words that correspond to an "equivalent term" (Algeo 1986, 272) in BrE. In cases where items are used quite commonly in both varieties, such as *dessert* or *living room*, the words were checked both in the Cambridge Online Dictionary and the Merriam-Webster Online Dictionary, as well as in representative corpora of the two varieties, i.e. the BNC1994 and COCA (*Corpus of Contemporary American English*), to make sure that they truly represent AmE and do not carry a different meaning in BrE. Furthermore, the number of items had to be limited. This was done by eliminating words that were unlikely to produce more than a few, if any, hits, such as *sideboard* compared to the more commonly used word *closet*. The limitation of this procedure is that the exclusion of other possible items has a direct influence on the overall results of this study: If other words were investigated, the results might look different.

For part one of the study, the data from the BNC1994DS and the Spoken BNC2014 were retrieved using both the tools of frequency and concordance. Search strings such as {cookie/N} and {elevator/N} were used in the two corpora to find the words only as nouns, but in all forms, for example *cookie*, *cookies*, *cookie's*, and *cookies'*. The output of the searches, the concordance lines, were sorted by hand in order to eliminate instances that did not correspond to the intended meaning and use. For instance, only hits for AmE *truck* corresponding to BrE *lorry* were included in the study, meaning that other non-AmE uses of *truck* in BrE, such as *monster truck*, *forklift truck* or *tow truck*, had to be weeded out manually. Each lexical item was searched for in both corpora and sorted according to this procedure. In addition, I categorized the investigated lexical items into semantic domains (see section 4.3.2) and counted the frequency for each domain for both corpora in order to be able to say something about 1) which domains exhibit the strongest influence of Americanisation and 2) which domains have changed the most over the last twenty years.

For part two, data were extracted only from the core subcorpus of the Spoken BNC2014 (see section 5.3). The procedure for extracting data was the same as for part one, but in addition, speaker metadata for each instance had to be extracted manually. This means that in addition to sorting the words according to their intended meaning and use, I registered the speaker's age, gender and class for each relevant concordance and counted the frequency

within each of the demographic groups. Moreover, I eliminated concordances where a speaker appeared several times for the same item. For instance, there are 88 matches for *apartment* in the core set of the Spoken BNC2014. However, only 42 of these matches are relevant for the frequency count since the same speakers appear several times. Moreover, as in part one, the lexical items were sorted into semantic domains and frequencies for each demographic category were counted. This enables an overview of the relation between demographic variables and semantic domain of the Americanism.

4.3 Classification of the material

4.3.1 Demographic categories

4.3.1.1 Age

Speakers in the Spoken BNC2014 are categorised according to two schemes: One corresponding to the age categorisation of the Spoken BNC1994 for the purpose of comparison, and the other more fine-grained for the purpose of "more sophisticated apparent-time analysis of the new data" (Love et al. 2017, 330). For this study, only the latter scheme with its ten age groups (0–10, 11–18, 19–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89, 90–99) will be applied (Love et al. 2017, 330). However, it should be noted that these ten age groups are collapsed into three larger cohorts. Firstly, this was done for the sake of simplicity, and, secondly, to increase the amount of data and the number of hits for Americanisms for each group in order to be able to draw some generalizing conclusions about potential trends in different age groups. Hence, the following three age groups emerge:

Table 1. Number of words for age groups in the core set of the Spoken BNC2014

Age	**No. words**
0–29	2,315,122
30–59	2,892,801
60–99	961,373

4.3.1.2 Gender

In the present study, data are categorized as either 'female' or 'male' regarding gender. This classification is based on the Spoken BNC2014's self-report forms.

Table 2. Number of words for gender categories in the core set of the Spoken BNC2014

Gender	no. words
f	3,611,155
m	2,558,141

4.3.1.3 Class: Social Grade

In the Spoken BNC2014, speakers are categorized according to their socio-economic status based on their occupation. The categories used in the corpus are adopted from the National Readership Survey's Social Grade demographic classification system (Table 3) (Love et al. 2018, 27).

Table 3. National Readership Survey Social Grade classifications

Code	Description
A	Higher managerial, administrative and professional
B	Intermediate managerial, administrative and professional
C1	Supervisory, clerical and junior managerial, administrative and professional
C2	Skilled manual workers
D	Semi-skilled and unskilled manual workers
E	State pensioners, casual and lowest grade workers, unemployed with state benefits only

Source: National Readership Survey 2014, in Love et al. 2018, 27.

In this thesis, these six categories are collapsed into two large cohorts, following Rayson et al.'s (1997) example: an "upper" category (A, B, C1) and a "lower" category (C2, D, E). The reasons for this are the same as for the division for the age groups (see section 4.3.1.1).

Table 4. Number of words for social grade categories in the core set of the Spoken BNC2014

Class	no. words
A/B/C1	3,090,422
C2/D/E	2,778,780

4.3.2 Semantic domains

In addition to classifying the investigated items in relation to their speakers' demographic variables age, gender, and class, the words are sorted into semantic domains which are inspired by and adopted from previous research such as Baker 2017, Gonçalves et al. 2018 and Korhonen 2020. All words investigated belong to one of the following semantic domains:

1. Food and cooking: *French fries, candy, cookie, eggplant, zucchini*, etc.
2. Housing and household: *apartment, elevator, garbage, living room, flashlight,* etc.
3. Clothing and accessories: *sneakers, sweater, diaper*, etc.
4. Transport: *highway/freeway, truck, gas station, subway, airplane,* etc.
5. Commerce: *drug store, shopping cart*
6. Entertainment, sports, and leisure: *vacation, the movies, Santa Claus,* etc.
7. Electronics: *cell phone, flashlight, phone booth*
8. Occupation titles: *attorney, mortician, janitor*
9. Education: *math, semester, eraser*
10. Miscellaneous: *homey, ladybug, zip code*

5 Material

This chapter starts out with a short overview on corpus design in terms of representativeness, balance and sampling in general. The two corpora used in this study, the BNC1994 and the Spoken BNC2014, will be presented regarding their content, data collection and design. Lastly, a word on the comparability of the two corpora is provided.

5.1 Corpus design: Representativeness, balance and sampling

Before the two corpora used in this thesis are presented in detail, some issues relating to corpus design and construction in general need to be discussed first. The data for this thesis were gathered from parts of two corpora, the BNC1994 and the BNC2014. Both these corpora are so-called *balanced* or *sample* corpora, meaning that they contain data which are meant to represent the language of a specific time period. The data in such corpora is collected based on a particular sampling frame with the intention to achieve *balance* and *representativeness* of the language we want to investigate, the so-called *population*. To attain a balanced set of data for a specific population, for example spoken British English in the 1990s, we would first need to decide on the range of speakers we would like to sample and then collect data evenly from across that range. To attain representativeness, we would in addition need to consider the relative proportions of the collected data (McEnery and Hardie, 2012, 8–9). A corpus is representative if the findings for this corpus hold true for the whole population. In other words, the results of any study conducted on a representative corpus can be generalised for the whole language that the corpus represents (Leech 2007, 135). It should be noted, however, that although balance and representativeness are basic requirements on which a corpus is built, these are difficult to achieve fully and, therefore, are rather "matters of degree" (McEnery and Hardie, 2012, 10). Nonetheless, Sinclair (2004) emphasises that any decision regarding corpus design and sampling must be guided by the principles of representativeness and balance.

5.2 British National Corpus 1994

The BNC1994, the *British National Corpus 1994*, is a monolingual, synchronic and general corpus: It is a collection of contemporary written and spoken British English, consisting of 100 million words gathered between 1991 and 1994 across a wide range of sources. The written part, which makes up 90% of the corpus, comprises data from for example

newspapers, academic books, fiction, essays and more. The spoken part contains orthographic transcriptions of both informal conversations by speakers categorized according to demographic categories such as gender, age, class etc. and context-governed speech (Burnard 2015). For the purpose of this thesis, only the demographically sampled spoken component of the BNC1994, the BNC1994DS, will be used. The BNC1994DS consists of around four million words of conversational speech. Speakers of British English across the United Kingdom were sampled according to the demographic parameters of gender, age, social group and region so as to achieve representativeness. This approach is called demographic sampling. In total, 124 speakers aged 15 or older were recruited. In addition, data from speakers aged 16 or younger were collected as part of the COLT Teenager Language Project.

Despite the wish for total representativeness, compromises between what is theoretically optimal and what is practically feasible had to be made (Burnard 2007). Therefore, the BNC1994DS does not contain equal amounts of data for each age group, class or gender. There is, for example, markedly less material of speakers aged 0–14 (6.72%) compared to the other age groups. Regarding the parameter of class, there is almost the double amount of data for AB respondents (32.39%) compared to DE respondents (15.25%). Moreover, female speakers (58.94%) stand for a significantly larger part of the Spoken BNC1994 compared to men (40.65%). The context-governed content of the Spoken BNC1994 is meant to complement the corpus with other spoken text types that do not appear as often in spontaneous speech (Burnard 2007). However, considering the need of comparability with the BNC2014, this component will be disregarded for this study.

5.3 Spoken British National Corpus 2014

The BNC2014, the *British National Corpus 2014*, is the follow-up corpus to the original *British National Corpus 1994* (BNC1994). It has been developed on the assumption that there is a need for an updated corpus of conversational British English, as the original BNC from 1994 is not suitable anymore for being a proxy for present-day British English (Love et al., 2017, 321–323). The BNC2014 consists of both a spoken and a written component. The written component is currently under development and, therefore, only the spoken part of the BNC2014 will be used in this thesis. The Spoken BNC2014 is a collection of orthographically transcribed present-day spoken British English which was gathered between 2012 and 2016, and it consists of around 11.5 million words. The material was collected in

informal settings and features 668 native speakers from the United Kingdom. The speakers are classified according to the following demographic categories: age, accent/dialect, highest qualification, gender and class (Love et al. 2017; Love et al. 2018).

Regarding data collection and corpus design, the Spoken BNC2014 only contains data gathered from informal contexts, excluding so-called 'context-governed data' which make up about 60% of the Spoken BNC1994 (Love et al., 2017, 324). Similar to the BNC1994, an 'opportunistic approach' was adopted to collect data that is available as well as try to maintain a balance across the demographic categories (Love et al., 2017, 327). Even if the data collection is more balanced than the BNC1994, the Spoken BNC2014 is nonetheless "not a properly balanced corpus if taken as a whole" (Love et al., 2017, 327). Therefore, for part two of this thesis, the Spoken BNC2014's core subcorpus, a smaller and more balanced corpus containing 6,169,296 words, will be used in order to achieve higher representativeness across the demographic categories gender, age, and socio-economic status.

5.4 Comparability

According to Leech (2007, 141–142), two corpora are comparable if they have equal designs, but the sampled data's origin differs either temporarily or locally. In other words, the data collected for comparable corpora might for example vary according to the country of origin or the period of time in which the texts were published or recorded. Comparing two or more comparable corpora, then, entitles the researcher to assume that differences in linguistic frequency between the corpora is due to the variability in space or time rather than to variability within one or the other corpus (Leech, 2007, 142). The BNC1994 and the BNC2014 are such comparable corpora; they share the same design but differ in the data's temporal provenance of the data.

As mentioned earlier, only the BNC1994DS, which covers about 40% of the whole Spoken BNC1994, will be used, as only this part is comparable to the Spoken BNC2014, which does not include context-governed data (Love et al., 2017, 324). That is, only data occurring in informal conversations by respondents categorized according to demographic categories will be compared. The advantage of comparing these two corpora is that it allows for investigating possible diachronic change over the last twenty years, utilizing the BNC1994 as a historical corpus and, thereby, creating "a useful resource for exploring recent change in spoken English" (Love et al., 2017, 324).

6 Results and analysis

The following chapter comprises the presentation of the study's findings followed by an analysis of these. The chapter is divided into two parts of analysis according to the thesis' bipartite focus. Section 6.1 presents the results of the diachronic study, which intends to answer the question whether there has been an increase in Americanisms in the spoken British English lexicon in the last twenty years. Section 6.2 presents the findings for the synchronic part, which aims at providing an answer to who uses Americanisms and whether the demographic variables gender, age and class have an effect on the use of Americanisms. In addition, each section displays the findings for the different semantic domains investigated. The aim is to investigate whether there are differences in the use of Americanisms between the different semantic domains.

6.1 Diachronic change: Has there been an increase in Americanisms in the spoken British English lexicon in the last twenty years?

This section provides the results and a quantitative analysis for part one of this study. First, the overall findings of Americanisms for the BNC1994DS and the Spoken BNC2014 will be presented. Thereafter, the results for each semantic domain will be shown. The quantitative analysis exhibits the differences between the two corpora and gives indications of differences in the two periods. Moreover, differences between the semantic domains reveal something about which areas are influenced the least and the most by American English.

6.1.1 The overall diachronic change

Table 5 shows the overall instances as well as the relative frequencies of Americanisms for the BNC1994DS and the Spoken BNC2014. In addition, the table gives the percentage of change between the two corpora. The percentage of change provides the change between two positive numbers in percentage, which can be either an increase or decrease (Pawlik and Czernia, 2020). In this case, the relative frequencies of Americanisms per million words in the BNC1994DS and the Spoken BNC2014 will be compared. The 'old' value, i.e. the relative frequency for the BNC1994DS, will serve as the reference point by which the difference between the 'new' value and the 'old' value is divided and multiplied by 100. The percentage difference is calculated by using the following formula (Pawlik and Czernia, 2020):

$$\frac{|\text{SpokenBNC2014}freq - \text{BNC1994DS}freq|}{\text{BNC1994DS}freq} \times 100$$

The final column presents the log-likelihood value for the change between the two corpora, which gives us an indication of the significance of the difference. The higher the score is, the more statistically significant the difference is between the frequencies of the two corpora. A value greater than 6.63 tells us that there is a 1% chance that the difference is due to a sampling error. If the value is above 3.84, the chance is 5% (Baker 2010, 63).

Table 5. Change in use of Americanisms. Relative frequencies per 1 million words (raw frequencies)

	BNC1994DS	**Spoken BNC2014**	**Overall change (%)**	**Log-likelihood**
Number of hits – all items	43.87 (220)	115.47 (1319)	163.2	219.63
Number of hits – significant items only	34.49 (173)	99.54 (1137)	188.6	215.85

The first row of data is based on all 72 items investigated in the study. The results show that there are 43.87 Americanisms per million words in the BNC1994DS, whereas there are 115.47 Americanisms in the BNC2014. This is an increase of 163.2%. This difference is statistically significant as well, with a log-likelihood score of 219.63. If we restrict our attention to the twenty items that proved to be significant (see table 7 in section 6.1.2), the second row shows that there were 34.49 AmE items per million words in the BNC1994DS, while almost three times as many Americanisms were found in the Spoken BNC2014. This is an increase of 188.6% over the last twenty years. In sum, the data show that there has been a significant increase in the use of Americanisms between 1994 and 2014, with nearly three times the number of Americanisms found in the Spoken BNC2014.

6.1.2 Semantic domains

The results for the second question of part one, whether there are differences between different semantic domains, are shown in Table 6. For each of the ten domains, the raw and relative frequency are given for both corpora. Moreover, the percentage of change between

the two corpora, corresponding to two different time periods, in addition to the value of significance for the change, is presented.

Table 6. Change in use of Americanisms according to semantic domains. Relative frequencies per 1 million words (raw frequencies)

	BNC1994DS	**Spoken BNC2014**	**Percentage change (%)**	**Log-likelihood**
Food and cooking	5.98 (30)	37.12 (424)	520.7	158.88
Housing and household	23.73 (119)	45.17 (516)	90.4	45.46
Clothing and accessories	4.59 (23)	2.19 (25)	-52.3	6.35
Transport	4.79 (24)	11.99 (137)	150.3	21.12
Commerce	0 (0)	0.18 (2)	-[2]	1.46
Entertainment, leisure and sports	2.59 (13)	5.34 (61)	106.2	6.48
Electronics	0 (0)	0.44 (5)	-	3.64
Occupation titles	0 (0)	1.31 (15)	-	10.92
Education	2.19 (11)	10.86 (124)	395.9	40.14
Miscellaneous	0 (0)	0.26 (3)	-	2.18

The data illustrate that all semantic domains, except 'Clothing and accessories', show an increase in the Spoken BNC2014. However, due to the low number of items in some categories, not all change is significant. This applies to the domains of 'Commerce', 'Electronic', and 'Miscellaneous' in particular. Moreover, the domains of 'Occupation titles', and 'Education' show significance, but they consist of only three items each, which makes the results less reliable for these domains. Considering these limitations, the remaining semantic domains that show a significant increase over the last twenty years are 'Housing and household', 'Food and cooking', 'Transport', and 'Entertainment, leisure and sports'.

As mentioned in section 6.1.1, this study investigates 72 Americanisms of which twenty prove to be statistically significant regarding change. Table 7 presents the frequency

[2] When the old value is zero, a percentage change cannot be calculated as one cannot divide by zero.

analysis of these twenty significant items sorted according to their semantic domains and provides raw and relative frequencies for both corpora as well as showing the change in percentage and its associated value of significance. A full overview of all items investigated can be found in the appendix.

Table 7. Change in use of significant items classified according to their semantic domain. Relative frequencies per 1 million words (raw frequencies)

		BNC1994 DS	**Spoken BNC2014**	**Percentage change (%)**	**Log-likelihood**
Food and cooking	*candy*	0.8 (4)	5.78 (66)	622.5	26.88
	cookie	1.39 (7)	8.49 (97)	510.8	35.93
	dessert	3.19 (16)	20.4 (233)	539.5	88.81
	scallion, green onion	0 (0)	0.61 (7)	-	5.1
	takeout	0 (0)	0.53 (6)	-	4.37
Housing and household	*apartment*	4.59 (23)	16.63 (190)	262.3	47.11
	closet	0.6 (3)	2.1 (24)	250	5.76
	living room	8.18 (41)	15.15 (173)	85.2	14.19
	mailbox	0 (0)	1.58 (18)	-	13.1
	to vacuum	1.99 (10)	0.7 (8)	-64.8	4.84
	vacuum cleaner	3.59 (18)	1.49 (17)	-58.5	6.62
Clothing and accessories	*pantyhose*	0.6 (3)	0 (0)	-100	7.12
	sweater	2.99 (15)	0.96 (11)	-67.9	8.2
Transport	*airplane*	0 (0)	0.96 (11)	-	8.01
	subway	0.2 (1)	2.19 (25)	995	12.09
	truck	4.59 (23)	7.79 (89)	69.7	5.66
Entertainment, leisure and sports	*vacation*	0.2 (1)	3.06 (35)	1,430	18.71
Occupation titles	*attorney*	0 (0)	0.53 (6)	-	4.37
	janitor	0 (0)	0.7 (8)	-	5.82
Education	*semester*	1.6 (8)	9.89 (113)	518.1	42.33

Most of the items in Table 7 demonstrate an increase between 1994 and 2014. Especially the domains ‘Food and cooking’, ‘Housing and household’ and ‘Transport’ exhibit a slightly higher number of significant items than other domains, which could indicate that these areas are influenced the most by AmE. However, since these categories contain the highest number

of Americanisms investigated in total (see Appendix), it can be expected that the number of significant items in these categories is also higher than for other domains with fewer items in total. The largest difference between the two corpora is found in the domain of 'Entertainment, leisure and sports' with a percentage change of 1,430% for the word *vacation*. This corresponds to one item found in the BNC1994DS and 35 items found in the Spoken BNC2014. Interestingly, a couple of Americanisms, namely *sweater* and *pantyhose* in the domain 'Clothing and accessories', and *to vacuum* and *vacuum cleaner* in the domain 'Housing and household', show a decrease (represented by the - sign) in the Spoken BNC2014. This indicates not only that words belonging to these domains are less influenced by AmE, but also that the popularity of the BrE version for these items seems to increase. This goes against the general trend of increased use of Americanisms in BrE shown in Table 5. Yet, it is important to mention that both the choice of words (see section 4.2) and the choice of classification into these specific semantic domains might influence the overall results of this study. A different classification and distribution of the Americanisms might have produced a different result.

6.2 Synchronic variation: Who uses Americanisms? Do the demographic variables age, gender, and class have an effect on the use of Americanisms in spoken language?

In the following section, the results and a quantitative analysis of part two of this study will be presented. The section is divided into three subparts corresponding to the three demographic variables that were investigated: Gender (6.2.1), age (6.2.2), and class (6.2.3). For each variable, the overall findings of Americanisms in the core subcorpus of the Spoken BNC2014 (see section 5.3) will be presented. Thereafter, the results for each semantic domain are shown. The quantitative analysis exhibits potential differences between the different groups within each variable, and thereby indicates whether certain demographic groups use Americanisms more than others. Moreover, differences between the semantic domains point to which domains are influenced the most by AmE, as well as how the use of Americanisms in certain semantic domains is distributed between the groups within each variable.

6.2.1 Gender

Table 8 shows the raw frequency of Americanisms as well as the number of hits per million words for male and female speakers in the core set of the Spoken BNC2014. In addition, the

value of percentage difference between the two gender groups is presented. The percentage difference (in %) provides the difference between two positive numbers in percentage and, thus, allows us to compare two different values (Díez and Czernia, 2020). In this case, the relative frequencies of Americanisms per million words for male and female speakers in the core set of the Spoken BNC2014 will be compared. The reference point by which the absolute difference between the two numbers is divided is the average of the two numbers, since the order does not matter, in contrast to calculating the percentage change in part one of this study (see 6.1.1). The percentage difference is calculated by using the following formula (Díez and Czernia, 2020):

$$\frac{|A-B|}{\left[\frac{(A+B)}{2}\right]} \times 100$$

The final column displays the log-likelihood value of significance for the difference between the groups. Differently to part one of this study, it is not meaningful to look at the significant items in addition to all items, as only very few items are characterized by significance. This is due to the small size of the core set, which is divided into even smaller subcorpora due to the categorization into the different demographic groups. Therefore, only the total number of Americanisms found will be presented.

Table 8. Use of Americanisms according to gender. Relative frequencies per 1 million words (raw frequencies)

	male	**female**	**Percentage difference (%)**	**Log-likelihood**
Number of hits	58.25 (149)	45.97 (166)	23.6	4.37

Table 8 is based on the 72 Americanisms investigated, and shows that there is not an overwhelming difference between the number found for male and female speakers. For male speakers, 149 Americanisms were found. This corresponds to 58.25 items per million words. For female speakers, 166 Americanisms, which corresponds to 45.97 items per million words, were found. This is a difference of 23.6%, which is also significant with a log-likelihood value of 4.37.

Proceeding to the different semantic domains, Table 9 presents the number of hits per

domain for both male and female speakers, in addition to the percentage difference and log-likelihood value for each domain. The table is based on all 72 Americanisms, which are distributed among ten domains.

Table 9. Use of Americanisms by gender according to semantic domains. Relative frequencies per 1 million words (raw frequencies)

Domains (number of items)	male	female	Percentage difference (%)	Log-likelihood
Food and cooking (15)	15.25 (39)	13.57 (49)	11.7	0.29
Housing and household (18)	23.45 (60)	31.32 (77)	28.7	0.30
Clothing and accessories (8)	1.17 (3)	1.94 (7)	49.5	0.56
Transport (11)	10.16 (26)	3.32 (12)	101.5	11.23
Commerce (2)	0 (0)	0 (0)	-	0.00
Entertainment, leisure and sports (6)	3.13 (8)	1.94 (7)	46.9	0.85
Electronics (3)	0.39 (1)	0 (0)	-	0.69
Occupation titles (3)	0.78 (2)	0.83 (3)	6.2	0.00
Education (3)	2.74 (7)	2.77 (10)	1.1	0.00
Miscellaneous (3)	0.39 (1)	0 (0)	-	0.69

The results reveal that the top-three areas with most hits for both male and female speakers are ‘Food and cooking’, ‘Housing and household’ and ‘Transport.’ Yet again, as it is these domains that contain the highest number of Americanisms, it is not unexpected that the frequency is higher than for other domains with a smaller number of items. For example, the domain ‘Food and cooking’ includes fifteen Americanisms out of a total of 72, whereas only six words belong to the domain ‘Entertainment, leisure and sports’ (see also Appendix). It is striking, however, that the fourth largest domain ‘Clothing and accessories’, which contains eight items, produces comparatively few hits. This indicates that words belonging to this domain are less influenced by AmE. Moreover, the data show that the use of Americanisms belonging to the domains of ‘Food and cooking’, ‘Transport’, ‘Entertainment, leisure and sports’, ‘Electronics’ and ‘Miscellaneous’ is dominated by male speakers, whereas female speakers dominate the use of AmE words belonging to the categories of ‘Housing and

household', 'Clothing and accessories', 'Occupation titles', and 'Education'. However, the only difference between male and female speakers found to be significant is in the domain of 'Transport' with a difference of 101.5% and a log-likelihood value of 11.23. This result implies that male speakers use Americanisms associated with 'Transport' about three times as much as female speakers do.

6.2.2 Age

Proceeding to the demographic variable of age, Table 10 shows the number of hits in the core set for each of the three age groups. As there are more than two categories for this variable, it was not possible to calculate the value of significance, as the log-likelihood test only works for the comparison of two sets of data. Nonetheless, it was possible and meaningful to calculate the different values of percentage difference between the age groups, which are presented in Table 11 (see page 39).

Table 10. Use of Americanisms according to age. Relative frequencies per 1 million words (raw frequencies)

	Group 1: 0–29	**Group 2: 30–59**	**Group 3: 60–99**
Number of hits	68.68 (159)	40.45 (117)	36.41 (35)

The data in Table 10 show that it is speakers in age group 0–29 that have the highest frequency of AmE items, with 68.68 hits per million words in the core set of the Spoken BNC2014. For speakers aged 30–59, 40.45 items per million words were found, while the search for speakers aged 60–99 produced 36.41 hits per million words. This confirms the hypothesis, based on findings from earlier studies such as Igboanusi (2003), Korhonen (2020), Rayson et al. (1997) and Vine (1999) (see Chapter 2), that younger speakers are more prone to use Americanisms than older speakers. Figure 1 further illustrates the distribution of Americanisms between the different age groups.

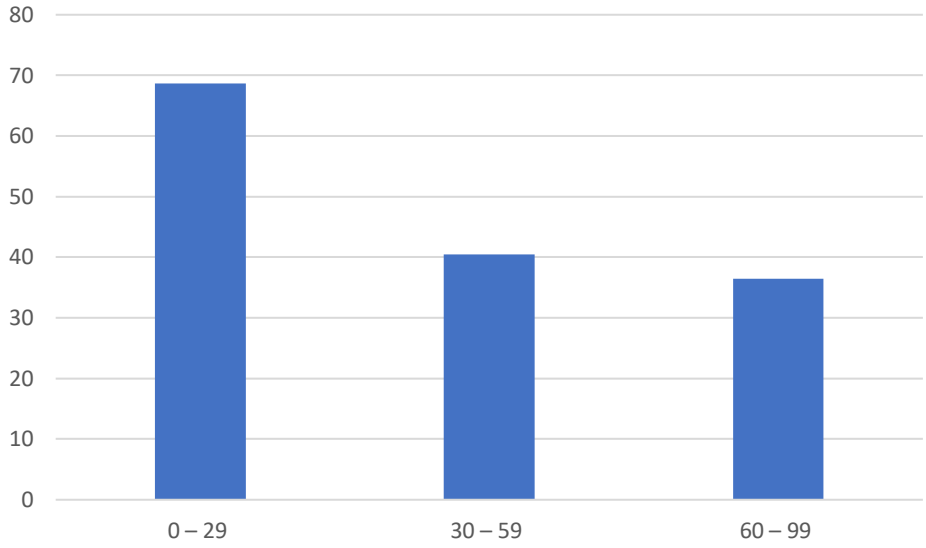

Figure 1. Use of Americanisms according to age groups. Relative frequency per million words

Table 11. Percentage difference between the different age groups

	Between Group 1 and Group 2	**Between Group 2 and Group 3**	**Between Group 1 and Group 3**
Percentage difference	51.7	10.5	61.4

The results presented in Table 11 show that the difference between age group 0–29 and age group 30–59 is 51.7%, and 61.4% between speakers in age group 0–29 and speakers aged 60–99. However, the difference between speakers from age group 30–59 and 60–99 is merely 10.5%. This implies that the difference in use of Americanisms is greatest between younger and older speakers, whereas the difference between middle-aged speakers and older speakers is less pronounced.

Table 12. Use of Americanisms by age according to semantic domains. Relative frequencies per 1 million words (raw frequencies)

Domains (number of items)	0–29	30–59	60–99
Food and cooking (15)	19.87 (46)	11.06 (32)	5.2 (5)
Housing and household (18)	27.64 (64)	19.01 (55)	16.64 (16)
Clothing and accessories (8)	3.02 (7)	0.35 (1)	2.08 (2)
Transport (11)	8.21 (19)	3.46 (10)	7.28 (7)
Commerce (2)	0 (0)	0 (0)	0 (0)
Entertainment, leisure and sports (6)	4.32 (10)	2.42 (7)	0 (0)
Electronics (3)	0 (0)	0.35 (1)	0 (0)
Occupation titles (3)	0.86 (2)	1.04 (3)	0 (0)
Education (3)	3.46 (8)	2.42 (7)	2.08 (2)
Miscellaneous (3)	0.43 (1)	0 (0)	(0)

Concerning the distribution of Americanisms by semantic domains, Table 12 presents the number of hits per domain for all three age groups. The results indicate that for all age groups the use of Americanisms is most common in the domains of 'Housing and household', 'Food and cooking', and 'Transport'. As mentioned before, it is these categories which contain the highest number of AmE items. Therefore, it can be expected that these domains display a higher frequency of Americanisms than other domains (see also section 6.2.1). The domain of 'Housing and household' exhibits the greatest frequency of Americanisms for all three age groups. For speakers aged 0–29 and 30–59, the second most influenced domain is 'Food and cooking' followed by the domain 'Transport'. For speakers in age group 3, however, it is the domain of 'Transport' that is in second place, followed by 'Food and cooking'. These results suggest that the language of older speakers is more influenced by AmE in the domain of 'Transport' than 'Food and cooking'. The difference between the two is small, though, compared to the gap in frequency for the two other age groups.

6.2.3 Class

Lastly, we will consider the data for the demographic variable of class. Table 13 enables an overview of the number of hits found for the "upper" (A, B, C1) and "lower" (C2, D, E) categories. Furthermore, as we only compare two sets of data, the table also displays the percentage difference and value of log-likelihood for this difference.

Table 13. Use of Americanisms according to class. Relative frequencies per 1 million words (raw frequencies)

	A/B/C1	**C2/D/E**	**Percentage difference (%)**	**Log-likelihood**
Number of hits	41.74 (129)	61.9 (172)	38.9	11.58

The frequency analysis of the findings demonstrates that speakers belonging to class A, B and C1 use less Americanisms than speakers belonging to class C2, D and E. For the "upper" category, 41.74 AmE items per million words were found, whereas 61.9 AmE items were found for the "lower" category. This is equivalent to a percentage difference of 38.9%. Further, the value of difference proves to be highly significant as the log-likelihood value is 11.58. The results confirm the hypothesis about speakers belonging to class C2, D and E appearing inclined to use Americanisms more often than speakers of the "upper" group.

Table 14 displays the frequencies of AmE items for each domain investigated and gives an overview of differences between the two class groups as well as the value of significance for these differences.

Table 14. Use of Americanisms by class according to semantic domains. Relative frequencies per 1 million words (raw frequencies)

Domain (number of items)	A/B/C1	C2/D/E	Percentage difference (%)	Log-likelihood
Food and cooking (15)	10.68 (33)	19.79 (55)	59.8	8.15
Housing and household (18)	21.03 (65)	22.31 (62)	5.9	0.11
Clothing and accessories (8)	1.29 (4)	1.8 (5)	33	0.24
Transport (11)	3.88 (12)	8.64 (24)	76	5.45
Commerce (2)	0 (0)	0 (0)	-	0.00
Entertainment, leisure and sports (6)	1.94 (6)	3.6 (10)	59.9	1.48
Electronics (3)	0.32 (1)	0 (0)	-	0.41
Occupation titles (3)	0.32 (1)	1.08 (3)	108.6	1.27
Education (3)	2.27 (7)	3.6 (10)	45.3	0.9
Miscellaneous (3)	0 (0)	1.08 (3)	-	1.62

The data show that, as expected, the top three domains exhibiting most Americanisms are ‘Housing and household’, ‘Food and cooking’ and ‘Transport’ in this order for both groups. The biggest differences between the “upper” and “lower categories” are in the domains of ‘Occupation titles’ with 108.6%, though with very few examples, ‘Transport’ with 76%, ‘Entertainment, leisure and sports’ with 59.9%, and ‘Food and cooking’ with a difference of 59.8%. However, only the differences in the domains of ‘Food and cooking’ and ‘Transport’ prove to be significant. Moreover, speakers belonging to classes C2, D and E use Americanisms more often than speakers of classes A, B and C1 in eight out of ten semantic domains. Lastly, the smallest difference between the two groups is found in the category of ‘Housing and Household’ with a percentage difference of merely 5.9%. However, this difference does not prove to be significant with a log-likelihood value of 0.11.

7. Discussion

This chapter, the thesis' research questions will be discussed in light of the study's findings, previous research introduced in Chapter 2 and relevant theory (see Chapter 3). The aim of this study was to investigate both diachronic and synchronic variation in the lexicon of spoken BrE. The first part of this thesis examined diachronic change in BrE over the last twenty years and aimed at answering the following research questions:

> **RQ1**: Has there been an increase in Americanisms in the spoken British English lexicon in the last twenty years? If yes, what causes this increase?
>
> **RQ2:** Are there diachronic differences in frequency between different semantic domains?

The second part of this study sought to investigate the synchronic and sociocultural aspect of the Americanisation of spoken BrE. The research questions were as follows:

> **RQ3:** Who uses Americanisms? Do the demographic variables age, gender, and class have an effect on the use of Americanisms in spoken language?
>
> **RQ4:** Are there synchronic differences in frequency between different semantic domains?

7.1 RQ1: Diachronic change

The findings for the first research question reveal that there has indeed been a significant increase in the use of Americanisms in the last twenty years. If we consider the investigated items that proved to be significant, the frequency of Americanisms in the Spoken BNC2014 was found to be almost three times as high as in the BNC1994DS. This corresponds to an increase of 188.6% from 1994 to 2014 and confirms the hypothesis that spoken BrE shows signs of increased Americanisation over the last two decades. In general, these results support previous research in that AmE influences other varieties of English in Expanding Circle, Outer Circle, and Inner Circle varieties (Vine 1999; Igboanusi 2003; Hänsel and Deuber 2013; Gonçalves et al. 2018; Korhonen 2020; Morgner 2020). However, very few studies examine the diachronic and synchronic aspect of Americanisation simultaneously. An exception is Gonçalves et al.'s study (2018), which focuses on both aspects: They examine vocabulary and spelling of written language both spatially by using a corpus of geolocated

tweets, and temporally over the last two centuries by analysing the Google Books dataset (Gonçalves et al. 2018, 1–3). Especially interesting for RQ1 are their findings for the last two decades, which correspond approximately to the time span investigated in the present study. Gonçalves et al. (2018) find that there is a clear shift from BrE to AmE vocabulary and spelling conventions, which becomes "especially visible in the decades following WWII and the fall of the Berlin Wall" (Gonçalves et al. 2018, 12). This study's results, then, do not only support Gonçalves et al.'s (2018) conclusion about diachronic evolution, but also expand it by contributing affirmative results for spoken language regarding BrE.

Previous research about other Inner Circle (Vine 1999; Gonçalves et al. 2018; Korhonen 2020; Morgner 2020) and Outer Circle varieties (Igboanusi 2003; Hänsel and Deuber 2013; Gonçalves et al. 2018) show that AmE has a strong influence on the English language. Most of these studies investigate written language and find that it is the domain of vocabulary where the influx of AmE is most prominent, compared to for example the domain of grammar. In addition, the pre-study for this thesis (Morgner 2020) examines vocabulary not only in written speech, but also in spoken language. The results show that the frequency of Americanisms is highest in oral language. Moreover, Morgner (2020) finds that even Inner Circle varieties, in this case BrE, Irish English, and New Zealand English, are influenced by the world-wide trend of Americanisation. The results of the present study, then, support and extend the findings of the pre-study by specifically investigating BrE and confirming that spoken BrE indeed is being Americanised to a certain extent.

Looking at the findings of this study, it becomes clear that there has been an increase in the use of Americanisms in BrE. However, based on the quantitative analysis of the data in this study, it is difficult to state with certainty what the nature of and reasons for this change might be. Whether the observed change is the start of a slow shift from BrE to AmE lexis (Trudgill 2002, 148–149) or whether speakers of BrE merely borrow AmE items to fill a gap and nativize certain items instead of replacing their own words as is the case in New Zealand English (Vine 1999), Nigerian English (Igboanusi 2003) and Australian English (Korhonen 2020), further research will have to investigate in the future.

As for the reasons for the increase of Americanisms in spoken BrE, especially considering the diachronic dimension, Gonçalves et al. (2018) observe a rise in the use of AmE items after two major historical events leaving "the US as the only superpower" (Gonçalves et al. 2018, 12): WWII and the fall of the Berlin Wall. The data collected for this thesis, representing the historical periods of 1994 and 2014, covering the decades after the end of the Cold War and the German reunification, agree with Gonçalves et al.'s (2018)

findings and support their theory of cause. Other reasons for the rising influence of AmE, at least in other Inner and Outer Circle varieties of English, are the constant influx of American cultural exports such as music, television, and books (Igboanusi 2003; Hänsel and Deuber 2013; Korhonen 2020), the economic and social globalisation (Hänsel and Deuber 2013), and the advent of the Internet by the end of the 20th century (Gonçalves et al. 2018; Korhonen 2020). It can be argued that at least some of those reasons, especially those linked to the influx of American culture through media, are also valid for BrE.

7.2 RQ3: Synchronic variation

In this section the results for the synchronic and sociocultural part of this study, which were presented in section 6.2, will be discussed. The findings for the various semantic domains will, however, be discussed in section 7.3 together with the results of the diachronic part in order to be able to draw a complete picture of the differences between different domains.

To answer the question about who uses Americanisms most, the results indicate, roughly speaking, that it is male speakers, speakers aged 0–29 and speakers belonging to classes C2, D and E who use Americanisms the most. Nonetheless, the use of AmE items was found in all groups along the three social dimensions. Starting with gender, the findings of this study do not reveal an overwhelming difference in the use of Americanisms between male and female speakers. Nevertheless, a higher frequency of AmE items was found in the speech of males. The relation between gender and the use of language has been investigated in previous research (Rayson et al. 1997; Schmid 2003), and differences between male and female speakers have been observed. However, as none of these studies focuses on the use of Americanisms, it is difficult to make any clear comparisons between their results and the findings in this thesis. In his study, Schmid (2003, 19–20) suggests that the observed differences between the two sexes can be linked to culture, and thus reflect cultural differences, such as interests, hobbies and worries, and trends about male and female speakers. The findings from the present study could therefore indicate that the discrepancy in the use of Americanisms between male and female speakers is due to cultural differences, which in turn affect the use of AmE items in spoken BrE.

Moving on to the second demographic variable, age, the study's results show that it is the youngest speakers, aged 0–29, that use Americanisms the most, followed by speakers in group 30–59 and older speakers between 60–99. Almost twice as many AmE items per million words were found for age group 0–29 compared to age group 60–99. These findings

support the hypothesis that younger speakers are the most inclined to adopt AmE items in their language. This hypothesis is based on studies on other varieties of English. Here, researchers found that it is the younger generations that have a more positive attitude towards the use of AmE words (Vine 1999; Igboanusi 2003) and adopt Americanisms more readily, especially concerning vocabulary (Rayson et al. 1997; Korhonen 2020). Even more specifically, Rayson et al. (1997, 9) find that it is the conversational language of British speakers under 35 that "appears to be following the lead of American English", which is confirmed by findings in the present study. Therefore, it can be claimed that the demographic variable age indeed has an effect on the use of Americanisms in British oral speech.

Lastly, the demographic variable of class is considered. Results suggest that speakers who belong to the "upper" group, i.e. classes A, B and C1, use considerably fewer Americanisms in conversational language than do speakers belonging to classes C2, D and E. The difference was found to be 38.9%. This confirms the hypothesis that speakers from "lower" classes are less concerned with using predominantly British words and, therefore, are more inclined to adopt Americanisms. However, as none of the studies introduced in Chapter 2 address the issue of the use of Americanisms in relation to class, it is difficult to draw any generalising comparisons. Rayson et al. (1997, 10) find that speakers belonging to the "lower" group use more informal language, for example greater use of third-person pronouns, swearwords and colloquial variants like *yeah* and *bloke*. If we were to classify Americanisms as informal language, it could be argued that the findings of the present study support the results of Rayson et al. (1997). In any case, the data in this thesis confirm that there are differences in oral speech between speakers of the two groups, and hence it may be argued that the demographic variable of class has an effect on the use of Americanisms in spoken BrE.

7.3 RQ2 and RQ4: Semantic domains

In this final part of the discussion, the findings for the different semantic domains, both diachronic and synchronic, will be discussed. Results for the diachronic part show that all semantic domains which were investigated show an increase in the frequency of Americanisms, except the domain 'Clothing and accessories'. One could interpret this as a sign of strength of the British fashion industry since none of the words in this domain shows a significant increase in the last twenty years. The domains of 'Housing and household', 'Food and cooking' and 'Transport' exhibit the highest frequency of Americanisms in both

corpora, as well as the highest number of significant items. In fact, the synchronic part reveals as well that it is these three areas that produce most hits for all groups within each demographic variable. Researchers such as Trudgill and Hannah (2017), Gonçalves et al. (2018) and Morgner (2020), who investigate the Americanisation of English and its influence on different domains, observe that the cultural influence and media presence of the US is eventually "accompanied by the American linguistic variety, which ends up affecting (global) English" (Gonçalves et al. 2018, 12). In regard to the three domains 'Housing and household', 'Food and cooking' and 'Transport', the high frequency of Americanisms indicates that these areas are influenced the most by this American influx. Moreover, Vine (1999, 17) suggests that lexical items that are used more regularly and in different contexts are more likely to change. This could apply to the investigated items in the three domains.

Hänsel and Deuber (2013, 350) note that the distribution of Americanisms depends on the semantic domain and observe that especially newer domains such as the domain 'Car' and other modern inventions such as the *cell phone* are more influenced by the AmE variety, at least in Trinidad and Tobago. Results from this study support these findings to some degree if we assume that the vocabulary field 'Car' includes such items as were investigated in the study's domain 'Transport'. Furthermore, in the pre-study for this thesis, Morgner (2020, 9) finds that the domains of 'Transport', 'Food and household' and 'Legal terms' are most influenced by AmE in the Inner Circle varieties Irish English, BrE and New Zealand English. This is largely confirmed by the findings in the present study, except for the domain of 'Legal terms', which is not a domain in the present study. Interestingly, the domain of 'Electronics' seems to be affected to a lesser degree than one might expect considering America's influence in this industry. However, considering the low number of items investigated for this domain, this finding is less reliable than for categories with a larger number of items. Moreover, it could be that many items belonging to the domain 'Electronics' are no longer considered typically American, which is why they were not included in this study, such as the word *radio* which is entirely adopted in BrE.

Regarding the demographic variables of gender, age and class, the domains with the highest frequency of Americanisms for all groups are 'Housing and Household', 'Food and cooking' and 'Transport'. For gender, the results show that male speakers use more Americanisms in the domains of 'Food and cooking', 'Transport', 'Entertainment, leisure, and sports', 'Electronics' and 'Miscellaneous', whereas a higher frequency of AmE items for female speakers was found in the domains of 'Housing and household', 'Clothing and accessories', 'Occupation titles', and 'Education'. These findings match to some extent with

Schmid's (2003, 18) conclusion that the frequency scores for most domains are "in line with widespread stereotypes about favourite female and male topics". The exception in this study is the category 'Food and cooking' which shows more hits for male speakers. Perhaps this is a consequence of the fact that chefs are often male. In sum, it can be argued that these gender differences reflect trends about the use of Americanisms for women and men in certain domains. Furthermore, findings for the three different age groups reveal that the frequency of Americanisms varies between the groups (see section 7.2), but that the domains in which most Americanisms are found are the same for the three. In other words, it is the frequency of AmE items rather than the domain the item belongs to that distinguishes the speakers of different ages. The same accounts for the variable of class, where speakers belonging to classes C2, D and E use more Americanisms than speakers belonging to classes A, B and C1 in eight out of ten domains investigated. The biggest significant differences in frequency were found in the domains of 'Transport' and 'Food and cooking'.

8 Concluding remarks

The aim of this study was to examine the diachronic and synchronic Americanisation of spoken BrE. To investigate whether there has been an increase in Americanisms in the spoken British English lexicon in the last twenty years, a corpus-based comparative approach has been applied to compare the BNC1994DS and the Spoken BNC2014. Furthermore, a variationist approach has been adopted to study synchronic variation and the effect of the demographic variables of gender, age and class on the use of AmE items.

The first part of the analysis revealed that spoken BrE indeed shows signs of increased Americanisation over the last twenty years, which confirms the hypothesis proposed at the beginning of this thesis (see section 1.1.1). These findings resonate with Gonçalves et al.'s study (2018), which served as an inspiration and a point of departure for this work. Furthermore, the quantitative analysis of the demographic variables showed that, generally speaking, it is men, speakers aged 0–29 and speakers belonging to classes C2, D and E that demonstrate a higher frequency of AmE items in their spoken language. The results, then, indicate that gender, age and class seem to have an effect on the use of Americanisms. This confirms previous research in the cases where one can draw comparisons between the different studies. Finally, regarding RQ2 and 4, findings revealed that nine out of ten semantic domains show an increase over the last two decades. Especially the categories of 'Housing and household', 'Food and cooking' and 'Transport' exhibited high frequencies, both in the diachronic and synchronic part. Moreover, it was found that differences between the various groups of speakers is rather a matter of frequency than of the semantic domain the item belongs to. Possible reasons for the Americanisation of conversational BrE were discussed and can be summarized as America's continuing influence on culture and its strong media presence, leading to AmE words finding their way into other varieties of English, including BrE. However, no absolute and comprehensive answer to why the lexis of BrE is changing can be found by considering corpora data only. The following section sums up limitations of this study and presents suggestions for further research on the topic of Americanisation.

8.1 Limitations and suggestions for further research

This study has provided further insight into the Americanisation of spoken BrE, both on a diachronic and synchronic level, by collecting data from two large corpora. Quantitative data

show whether there are changes in frequency; however, they do not reveal anything about other factors which might cause the change that has been observed. Therefore, it might be meaningful to combine the method of corpora with a qualitative method such as self-report or interview. Results about frequency together with findings about the speakers' attitudes and perception towards the use of Americanisms (see Vine 1999; Igboanusi 2003; Korhonen 2020) would allow researchers to draw a more comprehensive picture of the Americanisation of conversational BrE. Especially considering the interpretation of the results for the synchronic part of this study, a number of issues such as over-simplified conclusions, overfocusing on differences at the cost of similarities, and regarding differences in data as equalling cultural differences could be tackled by a combination of quantitative and qualitative methods.

Another problem of using corpora is that low frequency items, such as some of those that were investigated in this thesis, often produce none or just a few hits in corpora. In this study, I tried to minimize this limitation by using rather large corpora. Nonetheless, further research could use even larger collections, e.g. the GloWbE, to investigate informal BrE. Furthermore, as the choice of items and domains directly influence the study's results, it would be interesting to see whether a change of items or domains would have an impact on the overall results.

Lastly, a suggestion for further research is to include other speaker metadata and investigate different demographic variables such as highest qualification, dialect, location or foreign languages spoken, as well as look at several variables in combination in order to examine the influence of demographic variables on the use of Americanisms even more thoroughly.

Appendix

Table 15. Complete list of American English items

Semantic domain	AmE item
Food and cooking	appetizer
	to broil
	candy
	cookie
	cotton candy
	dessert
	eggplant
	fish stick
	French fries
	heavy cream
	popsicle
	scallion, green onion
	soy, soy bean
	takeout
	zucchini
Housing and household	apartment
	baby carriage
	couch
	closet
	crib
	elevator
	garbage
	garbage can
	living room
	mailbox
	pacifier
	pitcher
	plastic wrap
	stovetop
	stroller
	to vacuum
	vacuum cleaner
	washcloth
Clothing and accessories	diaper
	overalls
	pantyhose

	parka
	rubbers, rubber boots
	sneakers
	sweater
	tuxedo
Transport	airplane
	driver's licence
	expressway, freeway
	gas station
	hood
	railroad
	sidewalk
	subway
	traffic circle
	truck
	trunk
Commerce	carousel
	drug store
	the movies
Entertainment, leisure and sports	Santa Claus
	shopping cart
	sled
	soccer
	vacation
Electronics	cell phone
	flashlight
	phone booth
Occupation titles	attorney
	janitor
	mortician
Education	eraser
	math
	semester
Miscellaneous	homey
	ladybug
	zip code

Reference list

Primary sources

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/; accessed via https://cqpweb.lancs.ac.uk/bncxmlweb/.

Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina, and and Tony McEnery. 2017. "The Spoken BNC2014: Designing and Building a Spoken Corpus of Everyday Conversations. *International Journal of Corpus Linguistics* 22, no. 3 (November): 319–344. https://doi-org.ezproxy.uio.no/10.1075/ijcl.22.3.02lov; accessed via https://cqpweb.lancs.ac.uk/bnc2014spoken/.

Tools and dictionaries

Cambridge English Dictionary (online). 2021. https://dictionary.cambridge.org/ (last accessed June 11, 2021).

Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

Lexico.com. 2021. Oxford University Press https://www.lexico.com/ (last accessed June 11, 2021).

Merriam-Webster (online). 2021. https://www.merriam-webster.com (last accessed May 8, 2021).

Oxford English Dictionary (online). 2021. Oxford University Press. https://www-oed-com.ezproxy.uio.no/ (last accessed June 14, 2021).

Rayson, Paul. n.d. "Log-likelihood and effect size calculator." Lancaster University, University Centre for Computing Corpus Research on Language. Accessed June 16, 2021. http://ucrel.lancs.ac.uk/llwizard.html.

Secondary sources

Algeo, John. 1986. "The Two Streams: British and American English." In *Journal of English Linguistics* 19, no. 2: 269-84. https://journals-sagepub-com.ezproxy.uio.no/doi/pdf/10.1177/007542428601900208.

Baker, Paul. 2010. *Sociolinguistics and Corpus Linguistics*. Edinburgh Sociolinguistics. Edinburgh: Edinburgh University Press.

Baker, Paul. 2017. *American and British English: Divided by a Common Language?* Cambridge: Cambridge University Press.

Bell, Philip, and Roger Bell. 1998. *Americanization and Australia*. Sydney: University of New South Wales Press.

Burnard, Lou. 2007. "Reference Guide for the British National Corpus (XML Edition)." Published for the British National Corpus Consortium by the Research Technologies Service at Oxford University Computing Services, February 2007. http://www.natcorp.ox.ac.uk/docs/URG/.

Burnard, Lou. 2015. "What is the BNC?" Oxford Text Archive, IT Services, University of Oxford. http://www.natcorp.ox.ac.uk/corpus/index.xml.

Desmond, Jane C. 2012. "Americanization." In *Encyclopedia of Global Studies*, edited by Anheier, Helmut K., and Mark Juergensmeyer, 63-64. Thousand Oaks, CA: SAGE Publications, Inc., 2012. http://dx.doi.org.ezproxy.uio.no/10.4135/9781452218557.n18.

Díez, Álvaro, and Dominik Czernia. 2020. “Percentage Difference Calculator.” Omni Calculator. https://www.omnicalculator.com/math/percentage-difference.
Gonçalves, Bruno, Lucia Loureiro-Porto, Jose J. Ramasco, and David Sanchez. 2018. "Mapping the Americanization of English in Space and Time." *PLoS ONE* 13, no. 5 (May): e0197741. https://doi.org/10.1371/journal.pone.0197741.
Görlach, Manfred. 1995. “Heteronymy in international English.” In *More Englishes*, edited by Manfred Görlach, 93–123. Amsterdam: Benjamins.
Gries, Stefan Th. 2017. “The Four Central Corpus-Linguistic Methods.” In *Quantitative Corpus Linguistics with R. A Practical Introduction*, 8–20. 2nd ed. New York/London: Routledge.
Hänsel, Eva C., and Dagmar Deuber. 2013. "Globalization, Postcolonial Englishes, and the English Language Press in Kenya, Singapore, and Trinidad and Tobago." *World Englishes* 32, no. 3 (August): 338-57. https://doi-org.ezproxy.uio.no/10.1111/weng.12035.
Hill, Howard C. 1919. “The Americanisation movement.” *American Journal of Sociology* 24, no. 6 (May): 609–42. http://www.jstor.org/stable/2764116.
Hofland, Knut, and Stig Johansson. 1982. *Word Frequencies in British and American English*. Bergen: Norwegian Computing Centre for the Humanities.
Hundt, Marianne. 2001. “Grammatical variation in national varieties of English – The corpus-based approach.” *Revue Belge De Philologie Et D'histoire* 79, no. 3 (March): 737-756. https://doi.org/10.3406/rbph.2001.4545.
Igboanusi, Herbert. 2003. "Knowledge, Use, and Attitudes Towards Americanisms in Nigerian English." *World Englishes* 22, no. 4 (December): 599-604. https://doi-org.ezproxy.uio.no/10.1111/j.1467-971X.2003.00326.x.
Korhonen, Minna. 2020. “American influences on Australian English.” In *Australian English Reimagined: Structure, Features and Developments*, edited by Louisa Willoughby, and Howard Manns, 193–209. London: Routledge.
Labov, William. 1969. “Contraction, deletion, and inherent variability of the English copula.” *Language* 45, no. 4 (December): 715–62. https://doi-org.ezproxy.uio.no/10.2307/412333.
Labov, William. 1972. *Language in the Inner City: Studies in Black English Vernacular*. Philadelphia: University of Pennsylvania Press.
Labov, William. 1976. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press. (Orig. pub. 1972.)
Leech, Geoffrey, and Roger Fallon. 1992. “Computer corpora – What do they tell us about culture?” *ICAME Journal* 16, no. April: 29–50. http://icame.uib.no/archives/No_16_ICAME_Journal_index.pdf.
Leech, Geoffrey. 2007. “New resources, or just better old ones?” In *Language and Computers: Studies in Practical Linguistics*, edited by Christian Mair, Charles F. Meyer, and Nelleke Osstdijk. Vol. 59, *Corpus Linguistics and the Web*, edited by Marianne Hundt, Nadja Nesselhauf, and Carolin Biewer, 133–149. Amsterdam: Rodopi.
Leech, Geoffrey, Marianne Hundt, Christian Mair and Nicholas Smith. 2009. *Change in Contemporary English: A Grammatical Study*. Studies in English Language. Cambridge: Cambridge University Press.
Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina, and Tony McEnery. 2017. “The Spoken BNC2014: Designing and Building a Spoken Corpus of Everyday Conversations.” *International Journal of Corpus Linguistics* 22, no. 3 (November): 319–344. https://doi-org.ezproxy.uio.no/10.1075/ijcl.22.3.02lov.

Love, Robbie, Abi Hawtin, and Andrew Hardie. 2018. *The British National Corpus 2014: User Manual and Reference Guide (version 1.1)*. Lancaster: ESRC Centre for Corpus Approaches to Social Science.
McEnery, Tony and Andrew Hardie. 2012. *Corpus linguistics: method, theory and practice.* Cambridge: Cambridge University Press.
Melchers, Gunnel, Philip Shaw, and Peter Sundkvist. 2019. *World Englishes*. 3rd ed. London: Routledge.
Mencken, Henry L. 1930. "What America is Doing to Your Language". *Daily Express*.
Morgner, Lisa-Marie. 2020. "The lexical Americanisation of three Inner Circle varieties." Term paper, University of Oslo.
National Readership Survey. n.d. "Social Grade." Accessed June 13, 2021. http://www.nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade/.
Oakes, Michael, and Malcolm Farrow. 2007. "Use of the chi-squared test to examine vocabulary differences in English-language corpora representing seven different countries." *Literary and Linguistic Computing* 22, no. 1 (April): 85–99. https://doi-org.ezproxy.uio.no/10.1093/llc/fql044.
Pawlik, Kacper, and Dominik Czernia. 2020. "Percentage Change Calculator." Omni Calculator. https://www.omnicalculator.com/math/percentage-change.
Rayson, Paul, Geoffrey Leech, and Mary Hodges. 1997. "Social Differentiation in the Use of English Vocabulary: Some Analyses of the Conversational Component of the British National Corpus." *International Journal of Corpus Linguistics* 2, no. 1 (January): 133–152. https://doi.org/10.1075/ijcl.2.1.07ray.
Schmid, Hans-Jörgen. 2003. "Do women and men really live in different cultures? Evidence from the BNC." In *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech.* Łódź Studies in Language, edited by Andrew Wilson, Paul Rayson, and Tony McEnery, vol. 8. Frankfurt Am Main: Peter Lang.
Sinclair, John. 2004. "Corpus and Text – Basic Principles." In *Developing Linguistic Corpora: a Guide to Good Practice*, edited by Martin Wynne, Chapter 1. ADHS Literature, Languages and Linguistics. https://users.ox.ac.uk/~martinw/dlc/index.htm.
Strevens, Peter. 1978. *British and American English*. London: Cassell. (Orig. pub. 1972.)
Tognini-Bonelli, Elena. 2001. *Studies in Corpus Linguistics*. Vol. 6, *Corpus Linguistics at Work*. Amsterdam: Benjamins.
Tottie, Gunnel. 2002. *An Introduction to American English*. The Language Library. Oxford: Blackwell.
Trudgill, Peter. 2002. *Sociolinguistic Variation and Change*. Edinburgh: Edinburgh University Press.
Trudgill, Peter, and Jean Hannah. 2017. *International English: A guide to varieties of English around the world.* 6th ed. London: Routledge, Taylor & Francis Group.
Vine, Bernadette. 1999. "Americanisms in the New Zealand English lexicon." *World Englishes* 18, no. 1 (January): 13–22. https://doi-org.ezproxy.uio.no/10.1111/1467-971X.00118.
Witherspoon, John. 1781. *Pennsylvania Journal*, May 9, 1781.