

Organise or Organize?

**The Development, Use and Recognition of Verbal Endings *-ise* and *-ize* in
Contemporary British English**

Johanna Mäki
University of Tampere
School of Language, Translation and Literary Studies
English Philology
Pro Gradu Thesis
May 2015

Tampereen yliopisto
Englantilainen filologia
Kieli-, käännös- ja kirjallisuustieteiden yksikkö

MÄKI, JOHANNA: Organise or Organize? The Development, Use and Recognition of Verbal Endings *-ise* and *-ize* in Contemporary British English

Pro gradu -tutkielma, 86 sivua + 3 liitesivua
Toukokuu 2015

Tämä pro gradu -tutkielma käsittelee englannin kielen verbipäätteitä *-ise* ja *-ize* nykybrittienglannissa, jossa molempia kirjoitusasuja käytetään rinnakkain, mutta jossa kumpikaan muoto ei ole vakiintunut ainoaksi, toisin kuin esimerkiksi amerikanenglannissa. Tästä syystä variaatiota päätteiden käytössä esiintyy paljon, ja myös asenteet niitä kohtaan vaihtelevat suuresti tilanteesta riippuen.

Tutkielman tarkoituksena on selvittää korpusten ja muiden lähteiden avulla miten kahta rinnakkaista verbipäätettä käytetään nykybrittienglannissa, missä tekstilajeissa, ja onko niiden käytössä tai käyttäjissä tapahtunut muutoksia viime vuosina ja vuosikymmeninä. Päätteiden käyttöä tutkitaan myös kirjoittajien sukupuolen mukaan sekä iän perusteella. Tutkimuksen pääpaino on korpusten aikarajoitteista johtuen 1900-luvun lopusta 2000-luvun alkuun, mutta kirjallisia lähteitä tutkitaan myös 1900-luvun alusta lähtien.

Tutkielmani kahdessa ensimmäisessä osassa tarkastelen korpuslingvistiikan keinoja kielentutkimuksessa ja korpuksiin liittyviä rajoitteita, sekä kerron aineistostani ja sen käyttötavoista. Seuraavassa osassa käsittelen lyhyesti englanninkielisen kirjoittamisen historiaa, kirjoitusasujen vakiintumista sekä tutkittavien verbipäätteiden ja esimerkkisanojen etymologiaa. Lisäksi tutkin miten verbipäätteisiin *-ise* ja *-ize* suhtaudutaan sanakirjoissa, kielenoppaissa ja kustannuslalla. Myös uutismedian ja eräiden virallisten tahojen suhtautuminen näihin kahteen kirjoitusmuotoon otetaan huomioon. Viimeiseksi käyn läpi korpusaineistoa useasta eri näkökulmasta ja vertailen kirjoitusasujen esiintymistä esimerkiksi eri tekstilajeissa ja tekstityypeissä.

Korpusaineistostani selviää, että *-ise* on yleisempi kirjoitusmuoto nykybrittienglannissa kuin *-ize*, ja joitakin muutoksia niiden käytössä on nähtävissä eri vuosikymmenten välillä. Verbipäätteiden käytössä on paikoin huomattavia eroja eri tekstilajien sisällä. Joitakin kiinnostavia eroja löytyy myös miesten ja naisten tavassa käyttää päätteitä, mutta ikävertailussa ei löytynyt suurta variaatiota ryhmien välillä.

Avainsanat: verbipäätteet, brittienglanti, ortografia, korpustutkimus, *-ise*, *-ize*

1 INTRODUCTION	1
2 CORPUS LINGUISTICS	4
2.1 SOCIOLINGUISTIC APPROACHES	5
2.2 PREVIOUS STUDIES	5
3 DATA AND METHODS.....	8
3.1 THE CORPORA STUDIED	9
3.1.1 <i>The British National Corpus</i>	9
3.1.2 <i>The British Academic Written English Corpus</i>	10
3.1.3 <i>Limitations to the Corpora</i>	11
4 VERBAL ENDINGS -ISE AND -IZE: ORIGIN, DEVELOPMENT, USAGE AND DEBATE	13
4.1 PHONOLOGY AND ETYMOLOGY.....	13
4.2 STANDARDISING ENGLISH SPELLING: A HISTORICAL OVERVIEW	15
4.3 SPELLING REFORMS	17
4.4 INTERNATIONAL VARIETIES OF ENGLISH	20
4.5 -ISE AND -IZE ACCORDING TO DICTIONARIES, USAGE GUIDES, PRINTERS AND PUBLISHERS	22
4.5.1 <i>Dictionaries</i>	22
4.5.2 <i>Usage Guides</i>	24
4.5.3 <i>Printers and Publishers</i>	28
4.6 THE BRITISH NEWS MEDIA	31
4.7 INFLUENCE OF THE PUBLIC SECTOR	32
4.8 INFLUENCE OF SPELL CHECKERS AND OTHER SOFTWARE	33
4.9 ATTITUDES TOWARDS THE VARIATION	34
5 CORPUS STUDY: -ISE AND -IZE ACCORDING TO THE BNC AND BAWE	35
5.1 THE BNC.....	35
5.1.1 <i>Unrestricted Search</i>	37
5.1.2 <i>Diachronic Comparison</i>	40
5.1.3 <i>Comparing Authors by Age</i>	45
5.1.4 <i>Comparing Authors by Gender</i>	49
5.1.5 <i>Comparing Mediums</i>	52
5.1.6 <i>Comparing Text Types</i>	57
5.1.7 <i>Comparing Domains</i>	61
5.2 BEYOND THE BNC: THE BAWE CORPUS.....	66
5.2.1 <i>Unrestricted Search</i>	66
5.2.2 <i>Comparing Disciplines</i>	67
5.2.3 <i>Comparing Authors by Gender</i>	69
5.3 DISCUSSION OF THE CORPUS FINDINGS	71
6 CONCLUSION.....	77
7 REFERENCES.....	83
APPENDICES	87

1 Introduction

Languages are constantly evolving, and English is no exception. Even though the spelling system is the most fully standardised part of the language as it shows the least amount of national variation between different English-speaking parts of the world (Nevalainen and Tieken-Boon Van Ostade 2006, 271), English is still a very irregular language. It has been estimated that as much as one in every five words in the language today has variants in spelling, capitalisation and hyphenation (Crystal 2013, 23).

[Our] modern standard spelling arises from the fixing, in the eighteenth century, of printing conventions that had grown up and developed during the centuries between Caxton and Johnson. Most of our words have one form, or spelling, and one form only. There is no doubt about it. But some – and they form an interesting group – may be spelt in two, or (rarely) three, different ways. This is because printers and dictionaries themselves disagree (Vallins 1965, 150).

Should one *emphasise* or *emphasize*, *organise* or *organize*? Is one or the other more correct? There are two competing spelling choices in English language verbs like *emphasise/emphasize*, *organise/organize* and *realise/realize*. Neither spelling is false, and both are widely used, but there is great variation in their usage. Depending on where the writer comes from, there can be great differences in standards and conventions: the spelling is fixed in American English where *-ize* has become the standard, but in British and Australian English either spelling is correct (Fritz 2010, 258). The ratio in British English has been said to be roughly 3:2 in favour of the *-ise* spelling, whereas in Australian English *-ise* appears to be more widely used with a 3:1 ratio (Peters 2004, 298). Why is there is so much variation in British English regarding the use of *-ise* and *-ize*, and have there been changes in the usage of the two different forms? Is one or the other gaining more popularity in British English, or shall the issue remain unresolved? This level of variation between different international varieties of English is not uncommon, but the fact that usage among speakers of one variety of the language is so varied is an interesting phenomenon and begs for closer inspection.

The potential of text corpora has not been fully explored in this matter, at least from the point of view of British English, and it would be interesting to look into the phenomenon more

closely, especially since there seems to be a widespread misconception in Britain that the *-ize* spelling in British English is incorrect, an Americanism, and that is why *-ise* should be preferred (e.g. Dale 2013, Horne 2012, Oxford Dictionaries 2011). The purpose of this thesis is to examine, with the help of corpora, usage guides, dictionaries and several other sources, how and where the verbal endings *-ise* and *-ize* are used in British English, how authors of different ages and genders use them, and what changes, if any, have occurred in their usage over time. The corpora used in this study are the British National Corpus and the British Academic Written English corpus. Due to restrictions in the corpora used, the main focus of this investigation will be between the 1960s and the first decade of the twenty first century, but some attention will also be given to early twentieth century developments.

The reason for choosing the *-ise/-ize* divide as the point of discussion is simple. Previous studies have concentrated more on other orthographical features dividing British and American English, such as the suffix *-our/-or* (in *colour* and *honour*) and the double consonant in words like British English *travelling* and American English *traveling*. The spellings of the aforementioned features are so fixed in the orthography of British English that there is barely any variation in their usage. However, although in American English *-ize* has long been the only accepted form (Fritz 2010, 258), except in some special cases that only have one possible spelling, in British English the coexistence of *-ise* and *-ize* appears to have continued for several decades and even centuries, up until the present day. Even if some attention has been given to this phenomenon, and although various text corpora are readily available, no exhaustive study has been made to this day.

My research questions are as follows: what do linguistic text corpora reveal about the use of the suffixes *-ise* and *-ize* in British English? Are there differences in their usage between different domains, mediums or text types? Do men and women or people of different ages use them differently? Have there been any developments in their usage over time?

The structure of this thesis is as follows: Chapter 2 explains some of the general principles of corpus linguistics. Previous studies on the use of *-ise* and *-ize* will also be discussed in this

chapter. Chapter 3 will explain the data and methods used in this study. Chapter 4 focuses on the history of English language spelling and the etymology of the verbal endings *-ise* and *-ize* as well as the example words chosen for this study. Chapter 4 will also take a closer look at the differences in the usage of the two spellings between three major varieties of English, and examines how *-ise* and *-ize* are treated in dictionaries, usage guides and by printers, publishers and the news media. In addition, the role of the public sector regarding the spelling choices will also be discussed. Finally, in Chapter 5, the corpora will be studied from several different viewpoints, and the findings will be compared with each other. The results of the findings will be then be discussed at the end of the chapter.

My research will show that according to the corpora studied, *-ise* is the more common variant of the two spellings in contemporary British English, and that the frequencies of usage differ significantly between different domains, mediums and text types. There are also some interesting differences between male and female authors. Some changes were detected when texts from several time periods were compared with each other, but no significant variation was found between groups when comparing authors by age.

2 Corpus Linguistics

Text corpora are invaluable tools for linguists. A corpus is “a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis” (Tognini-Bonelli 2001, 2). Corpora are supposed to be composed of natural language, that is, they should not contain material created specifically for linguistic analysis (Stubbs 2004, 111). A balanced corpus, one that aims to give a comprehensive sample of a given language at a given time, “must represent variables of demography, style, and topic, and must include texts which are spoken and written, casual and formal, fiction and non-fiction, which vary in level (e.g. popular and technical), age of audience (e.g. children or adults), and sex and geographical origin of author, which illustrate a wide range of subject fields (e.g. natural and social sciences, commerce, and leisure)” (Stubbs 2004, 112).

A public corpus, like the BNC, is a collection of such data, available to linguists “either as an identifiable whole or from easily accessible materials” (Bauer 2004, 99). One important benefit of using public corpora is replicability; it is possible for any two researchers to reach the same results if they use exactly the same search methods. Also, by using corpora, linguistic phenomena can be analysed numerically (Bauer 2004, 102-3).

Not even balanced corpora can easily contain enough data from all possible points of view and the types of sources it holds, therefore making it difficult to create large enough subcorpora within one to study all specific categories in detail. However, a balanced corpus can at least give some kind of a foundation from which to mine information regarding variation within these categories (Aston and Burnard 1998, 24).

No corpus is by itself a perfectly objective source, and whoever accesses a corpus must keep in mind that there can be ambiguities in the findings. A certain amount of criticism towards corpus data is required. How one uses and interprets the material is important, and cross referencing is vital. The representativeness and credibility of any findings depends very much on what kind of sources the compilers of the corpus have used as their raw data. As Bauer (2004, 103) points out,

there is no guarantee that the samples are representative of all texts produced in the same variety of English at the same time. However, if different corpora provide similar results, it is more likely that they have been comparatively representative (Bauer 2004, 103).

In addition, if a corpus, or a part of it, is not constructed exactly like one would need or want, it is still possible to make generalisations based on the results found (Meyer 2002, 121). As samples within a corpus do not always represent all the variants of a given language or the usages within, it is often necessary to find information outside the corpus as well (Meyer 2002, 124).

2.1 Sociolinguistic Approaches

The primary focus in sociolinguistics is to study how variables such as age, gender and social class affect the way in which people use language (Meyer 2002, 18). Sociolinguists attempt to find out, for example, what linguistic differences and similarities there are between or within groups of people, and how social variables, like those listed above, impact on language use (Baker 2010, 2).

The social variables that are relevant for this study are gender and age, as those are the two that are possible to examine in the primary corpus used in this thesis. Gender comparison is also possible in the case of the secondary corpus. Although the purpose of this thesis is not to compare and contrast the differences between writers of different ages or genders in too much detail, they will be looked into to some extent, as much as is possible with the help of the corpora used.

2.2 Previous Studies

Although the unfixed nature of the verbal endings *-ise* and *-ize* has been noted and discussed in countless dictionaries, usage and style guides as well as linguistic histories and other works in the field of linguistics, the co-existence of the two forms has resulted in few detailed corpus analyses. Most discussion on the subject has focused on the differences in style between British and American English, but Australian English is also interesting from the point of view of the discussion at hand.

An insight into British versus American English usage has been provided by, for example, Shin'ichiro Ishikawa (2011). In his study he found, among other things, that while *-ize* has long been a standard variety in American English, no such standardisation has happened in British English. According to his corpus findings, *-ise* is used almost exclusively in newspapers and ephemera, but in books *-ize* is more used than *-ise* (Ishikawa 2011, 395)

The *-ise/-ize* variation seems particularly well researched in Australian English. Some detailed, corpus based investigations into the use of *-ise* and *-ize* have been conducted by, for example, Clemens Fritz (2010) and Pam Peters (2007). Fritz compared eighteenth and nineteenth century corpus data with several modern text sources and concluded that the use of *-ise* has increased over time, and that there is variation, for example, between its use among people of different social classes (Fritz 2010, 260). Peters, on the other hand, found that although the *-ise* variant is by far the most used in Australia, some regional variation exists (Peters 2007, 431).

A brief but interesting article by Aronson (2001, 1173) found that in medical texts *-ize* is clearly the most common spelling of the two variants. In his Medline search with *digitalise* and *digitalize*, he found 154 instances (17%) of the *-ise* spelling and 778 instances (83%) of the *-ize* spelling.

Number (%) of occurrences of <i>digitalise</i> or <i>digitalize</i> in bioscience papers				
Form	Total	Source of articles		
		US/Canada	UK	Rest of world
<i>Digital-ise, -ised, -ises, -ising, -isation</i>	154	4 (2)	30 (21)	120 (76)
<i>Digital-ize, -ized, -izes, -izing, -ization</i>	778	248 (32)	56 (7)	472 (61)

Table 1 *Digitalise* and *digitalize* in bioscience papers according to a Medline search.

Much of this can be explained by the number of articles from North America. It is not clear whether Aronson's numbers are instances in separate articles or occurrences of words in the entire database, including multiple hits within one text, but the findings are very interesting regardless. Even though *-ize* appears to be much more common than *-ise* overall, even in Britain, the results do indicate that

-ise is not an entirely unfamiliar spelling in the field of medical science, neither in Britain nor elsewhere.

3 Data and Methods

In order to study the changes in British English orthography regarding the suffixes *-ise* and *-ize*, the following two corpora were consulted: the *British National Corpus* (BNC) and the *British Academic Written English Corpus* (BAWE). The BNC was chosen as the primary corpus because of its size and representativeness of real language. It is by far the more versatile one of the two, and therefore it is studied in more detail. The BAWE is considerably more limited in terms of representativeness and search methods, but it gives useful data about more recent developments.

When comparing results from different corpora, they should ideally be of similar size, or at least the frequencies would need to be normalised to reflect the size of the corpus, otherwise the results would be distorted (Meyer 2002, 126). However, since I will compare and contrast frequencies of two competing spellings of the same words first *within* one corpus and only then compare the results to those of the other corpus, the fact that the lengths of the corpora used in this study are dissimilar should not cause concern. The sizes of the two corpora studied are not, in fact, too contradictory when looking at the number of samples they are made of: the BNC contains 3140 text samples and the BAWE 2897 samples. However, since the BNC consists of several kinds of different types of texts whereas BAWE contains only academic texts, the findings are not directly comparable with each other.

Since it was not possible to search the corpora in a way that would have included all those verbs that can have both the *-ise* and *-ize* construction in British English and still retain all the refined search methods that the corpora allow, I have chosen to study the ten most frequent verbs in the BNC that can have both spellings, in order of frequency in the database. They are *realise*, *recognise*, *organise*, *emphasise*, *criticise*, *characterise*, *specialise*, *summarise*, *apologise* and *minimise* (see Table 4 in Chapter 5). These were the ten most common examples of both spellings, and the same ten verbs will also be the words studied in the secondary corpus.

In order to get a more detailed picture of the suffixes *-ise* and *-ize*, and how and where they are used in British English, several other sources were also consulted. The policies and opinions of

dictionaries, usage guides and the news media were looked into. In addition, style guides created by printers, publishers and some important public sector operators were also included. The sources listed under each section of discussion are not meant to be exhaustive, but are there to provide some influential or interesting examples for each category.

3.1 The Corpora Studied

3.1.1 The British National Corpus

The British National Corpus (BNC) is a database of 100 million words from ca. 4000 samples of text and spoken language, collected from various kinds of sources to represent contemporary British English (Aston and Burnard 1998, 5). However, since the corpus was collected between 1991 and 1994, and it consists of data from 1960 to 1993, rather than forming a sample of the latest forms of language today, it is more of a time capsule of British English in the late twentieth century when the corpus was compiled.

The written part of the BNC, which will be used in this study, consists of some 87 million words and makes almost 90 per cent of the corpus. The corpus aims to represent various types of language and it is compiled of 3140¹ individual text samples from, for example, newspapers both local and national, a wide range of periodicals, academic and non-academic books and publications, university essays and personal correspondence (Burnard 2009), and therefore it can be classified as a balanced corpus instead of a register-specific or a dialect-specific one (Aston and Burnard 1998, 5).

The BNC was chosen as the main source of data for this thesis for its impressive size and representativeness of real language. The BNC “aims to represent the universe of contemporary British English” (Aston and Burnard 1998, 5), and the samples are from a “wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century”

¹ The web interface of the BNC used in this study contains 3140 texts in total, whereas Burnard (2000, see Appendix 1) refers to 3144 texts.

(Burnard 2009). The BNC is also a corpus that was easily available, and its exhaustive search methods make it possible to find detailed information on the chosen subject.

For the purposes of this thesis the corpus was consulted from various points of view. In order to get a detailed picture of the distribution of *-ise* and *-ize* in contemporary British English, queries were performed based on the date of publication and the age and gender of the author, as well as according to the medium, domain and text type of the samples. The results of these queries are presented numerically, as numbers of texts with word matches and as percentages calculated from these numbers. The results are then compared between the different groupings within each section, both as individual words and as a larger grouping of words.

In order to find out how the two spellings are truly distributed within the corpus, simple word counts were abandoned and only the number of individual texts in which any or all of the example words and one or both of their spellings were found were taken into consideration. This approach was taken so that the matches found would not just represent how often the two spellings appear in the whole of the corpus but, rather, how many different texts and therefore different individuals or institutions used one or the other spelling.

3.1.2 The British Academic Written English Corpus

In order to get a fresh perspective to the research questions, a more recent sample of written British English was needed for comparison. The British Academic Written English corpus, or BAWE, is a collection of student assignments collected from the universities of Warwick, Reading and Oxford Brookes between 2004 and 2007. The 6.5 million word corpus is made of 2897 samples of text, 2761 of which are assignments. They were written by 1039 students at various levels of study, and they are of four disciplinary areas: arts and humanities, life sciences, physical science and social sciences (Heuboeck, Holmes and Nesi 2010, 6).

Compared to the BNC, the BAWE is very limited in size and range. However, it does represent one aspect of the language of young Britons in the 2000s, albeit restricted to a formal and academic context. Unlike the BNC, the web interface of BAWE that was used for this study did not

offer the possibility to limit the search results to individual texts in which the words and spellings were found, and so the findings are presented as numbers of word matches in the corpus. Percentages were then calculated from these matches.

3.1.3 Limitations to the Corpora

One of the most pressing challenges of the BNC for this particular study is its age: as stated earlier, it was compiled in the early 1990s. Even the most recent data in the corpus is from twenty years ago, from 1993 – before the Internet became a global phenomenon, before the emergence of text messaging and the resulted, abbreviated text message language of the youth, and so on. In the past twenty years much has changed in the way people use language, and how they become exposed to written language, especially written English. This change has, perhaps, been more radical since the completion of the BNC than in the twenty years prior because the Internet has broken the physical boundaries of geography. In the twenty first century, being exposed to international and non-native forms of English, both formal and informal in style, is an everyday occurrence for a large number of people.

Although the BNC contains samples of various kinds of sources and it aims to represent modern British English from many points of view, it does not, however, contain an equal portion of samples from each different type of text or by all different types of authors. In many cases, detailed information regarding the authors of each text sample in the corpus is not known, and therefore comparing the corpus findings by the age or the gender of the writer is often difficult because of the limited size of these subcorpora. For example, people aged 24 or younger are seriously underrepresented in the corpus, mainly because they do not produce the kinds of texts that were collected for the corpus, like press reportages and technical reports (Mayer 2002, 49).

There are also some errors in the BNC: some texts or parts of text appear in the corpus more than once (Aston and Burnard 1998, 39). In addition, as the texts added into the corpus were not proof-read at any stage, any errors in the original source, including misspellings, will also be found in the corpus (Aston and Burnard 1998, 37).

The BAWE, on the other hand, is a highly specialised corpus as it contains only student assignments from universities. The texts are of a very specific genre, and the spelling choices made by the writers may have been heavily influenced by guidelines provided by the educational establishments themselves. It must also be noted that since just 1039 writers created the 2897 text samples that form the corpus, there are multiple samples created by one person. However, no such estimates are known for the BNC, and many of the text samples within could also have been written by one individual.

4 Verbal Endings *-ise* and *-ize*: Origin, Development, Usage and Debate

Variation in spelling is common, even in contemporary English, as languages are never fully fixed. The written form of the English language has developed quite organically over time as there has been very little interference by official authorities, at least in the case of British and American English (Venezky 1999, 6). However, the written form and its conventions have been affected and moulded by individual authors, printers and their dialects and, for historical reasons, other languages, especially Latin and French.

4.1 Phonology and Etymology

While one might think that the most logical letter to represent the sound /z/ in the English language would be the letter *z*, it is, in fact, *s* which is the most common spelling of the sound (Treiman 1993, 136). The plural marker, for example, is always *-s*, regardless of the pronunciation. However, the choice between *s* and *z* has, in some cases, remained unfixed until quite recently. As late as in the nineteenth century, using *s* or *z* could be a matter of personal preference for many authors and other educated individuals, in other words, masters of their native tongue:

Surprize rather than *surprise* was used by Georg Eliot and Walter Scott; Michael Faraday (the pioneering English chemist and physicist) selected *fuze* rather than *fuse*. Darwin embarked on a *cruise* rather than *cruiſe* in his voyage on the *Beagle*. *Cozy* was the preferred form of Queen Victoria and of the novelist (and politician) Benjamin Disraeli [...]. Dorothy Wordsworth preferred *cozie* while Dickens used *cosey* (Mugglestone 2006, 280).

In Old English the letter *z* was used only in loan words with the value [ts], but when orthographic developments in French spread to the English language, *z* came to replace *s* in some native words like *freeze* (Middle English *fresen*) (Pyles 1971, 69).

For word such as *organise/organize*, *realise/realize* and *recognise/recognize*, what really is the correct spelling, *-ise* or *-ize*? Contrary to common belief among speakers of British English (e.g. Dale 2013, Horne 2012), *-ize* is not an Americanism but, in *most* cases, the etymologically ‘correct’ form. The verbal endings *-ise* and *-ize* derive from the Greek *-izein* (-ιζειν), which came to the English language from French *-iser* via Late Latin *-izāre* (*The Oxford English Dictionary* 1989).

However, even though the suffix *-ize* itself may be of Greek origin, not all verbs ending in *-ise* in English derive from Greek. Some words have come into the language directly from French, and in some cases the *-ise* construction of a verb is not a suffix, but rather a part of a larger word element like *-mise*. The *-s* can also be a part of the stem of the word, like in the case of *televise* (*Oxford Dictionary of English* 2003, 922). This is why there are several words that can only be spelled with *s* both in British and American English: *advertise*, *advise*, *apprise*, *chastise*, *circumcise*, *comprise*, *compromise*, *demise*, *despise*, *devise*, *disfranchise*, *disguise*, *enfranchise*, *enterprise*, *excise*, *exercise*, *franchise*, *improvise*, *incise*, *premise*, *revise*, *supervise*, *surmise*, *surprise*, *televise* (Fowler 1965, 314; Peters 2004, 298). In contrast, there is only one *-ise/-ize* verb longer than one syllable that must always be spelled with *z*, and that is *capsize* (Peters 2004, 298). Therefore choosing one spelling over another could make it easier to avoid making spelling mistakes in the exceptions listed above. However, the rules regarding these exceptions may be changing: for example *Webster's New Twentieth Century Dictionary* (1983) gives *advertize*, *apprize* and *comprize* as acceptable spellings in American English. The *-ise/-ize* construction is very productive in verb formation and new words are constantly added to the lexicon, which is one reason why the question of the dual spelling remains topical.

The earliest known verb using the suffix *-ize* in English is *baptize*, which was first recorded in the late thirteenth century (Burchfield 1996, 422). The ten verbs studied in this thesis, *realise*, *recognise*, *organise*, *emphasise*, *criticise*, *characterise*, *specialise*, *summarise*, *apologise* and *minimise*, have all entered English at different points in time, and they will now be listed in order of appearance in the English language, according to examples in *The Oxford English Dictionary* (*OED*).

The oldest of the ten, according to the *OED*, is *organise*, which entered the language in 1425 as *organize*. The first example of an alternate spelling is from the 1500s in the form of *organyse*. *Recognise* is another early example, first seen in 1456 as *racwnnis* and, in a more recognisable form, as *recognise* in 1534. Contrary to many of the other words, in the case of

recognise the *-ise* spelling is the most common in all the text examples given in the *OED*, which may indicate that its French origin has influenced its usage more than the other words studied.

Characterise is first found in 1581 as *characterize*, and it derives partly from Greek via Latin *characterizare*, and partly from adding *-ize* to the noun *character*. The first *-ise* form listed in the dictionary is from 1594. *Apologise* was formed by adding the verbal ending to the noun *apology*. The first recorded sighting according to the *OED* is from 1609 as *apologise*.

The origin of *realise* is most likely French (from *réaliser*), and it is first seen in 1611 as *realize*. The first *s*-spelling listed is from 1755. *Specialise* is another French loan, from *spécialiser*. The first example given in the *-ize* form is from 1613 and the first *-ise* form from 1616.

Criticise was borrowed from Latin *criticus* and is first seen in 1649 as *criticize*. The first *s*-spelling appears more than a hundred years later in 1790.

The newest words of the ten, *minimise*, *emphasise* and *summarise*, were all coined in the nineteenth century. These are examples of words formed by adding the suffix to an existing word rather than by adopting the whole word from another language, as was the case with most of the older words. *Minimise* (from *minimum*), was first seen in 1825 as *minimize*, and the first *s*-spelling listed is from 1884. The first example for *emphasise* is from 1828 as *emphasize*, and the first *s*-spelling is from the 1860s. *Summarise* (from *summary*) appears in 1871, and most of the earliest example of the word use the *-ise* spelling.

4.2 Standardising English Spelling: A Historical Overview

The norms and conventions of written English varied greatly up until Early Modern English. English ceased to be a written language for centuries after the Norman Conquest in 1066 when French took over its place. Latin continued to be the language of choice in the fields of religion and education, reducing English to a second rate language spoken by commoners (Nevalainen and Tieken-Boon van Ostade 2006, 272). Even before this demotion in importance, English was divided into dialects. The long period of neglect deepened the differences between these dialects, and when the language was once again used in a written form, local variations were noticeable both in

spelling and in grammar (Fritz 2010, 228). Writers had their own individual styles, and their dialects influenced their texts heavily.

The fifteenth century is usually seen as the time of standardising English spelling (Smith 2006, 133). Although there was still more variation in different possible spellings than in Modern English, most of the dialectal varieties had been dropped from use and printers as well as some early spelling reformers and scholars had set their own prescriptive norms of standard English (Smith 2006, 134; 136). The reforms continued through the sixteenth century and into the seventeenth century as a large number of Latin, Greek and French loan words were coming into English (Carney 1994, 467). In the sixteenth century, with the revival of interest in learning, there was a growing interest in English grammar and orthography. That, together with the increased amount of printing, helped in fixing some rules of English spelling (Venezky 1970, 18). However, it was usually the learned alternative of a spelling that was preferred, rather than a previously existing one that would have been closer to pronunciation (Scholfield 1994, 63).

In the eighteenth century, uniformity in spelling was spreading from books and other more public mediums into letters, diaries and other private writing. This was made possible by a more widespread access to education (Fritz 2010, 229). This process of standardisation was still not complete, however, and a desire for correctness in spelling and writing resulted in plans to produce a standard dictionary. This task was undertaken by Samuel Johnson, who aimed to ‘fix’ and ‘standardise’ English (Davis 1999, 80), and his dictionary, published in 1755, became so influential that the spellings preferred by Johnson’s dictionary came to be accepted as the standard in England (Trask 1994, 34), even if Johnson himself was not completely consistent with his spelling choices (Clemens 2010, 265).

The spelling reforms since Johnson and other more current developments will be discussed in the following section (4.3).

4.3 Spelling Reforms

Unlike the French and their L'Académie française,² neither Britons nor Americans have a language academy or any other official language authority, nor have they ever had one (Venezky 1999, 6).³ Individual authors, lexicographers, grammarians, printers and publishers have been the pioneers in developing a more uniform written form of English. Where British English has kept many of its imported or even archaic spellings, American spelling usually tends to follow the principles of simplicity and derivational uniformity, making such pairs of words as *defense-defensive* and *offense-offensive* when the British would spell *defence* but *defensive*, and *offence* but *offensive*. However, occasionally it is the American spelling that violates this principle of derivational uniformity: unlike British English *analyse* and *paralyse*, American English *analyze* and *paralyze* do not correspond with the nouns *analysis* and *paralysis* (Gramley and Pätzold 2004, 280).

The American preference for *z* is, however, consistent in its attempt to imitate pronunciation. There have been many attempts to simplify English spelling, and the modern American standards have much to thank for Noah Webster who is responsible of, for example, using *-er* instead of *-re* in *theatre* etc., and for simplifying the suffix *-our* to *-or* in *behaviour*, *colour* and the like (Pyles 1971, 266-7). In American English, using *z* instead of *s* even in places where it has not traditionally been, such as the *-yze* in *analyze* and *paralyze* (British English *-yse*), is an extension of the aim for orthographic simplicity and an attempt to reflect pronunciation (Gramley and Pätzold 2004, 280).

Despite the long history of changes and reforms, there are still many questions and debates over the state of the English language. English is a world language: there is no one single English language but several international forms, not to mention all the different regional varieties. Spelling conventions may vary greatly from country to country, and some argue for global uniformity. Foster

² L'Académie française is the official authority in France who makes recommendations on usage, vocabulary and grammar of the French language. The Académie carries no legal power, but it has a high status and its dictionaries are considered official.

³ The U.S. Board on Geographic names, founded in 1890, has set standards or preferences for spelling place names in the U.S., but apart from that there has not been another language institution with such power in the United States (Venezky 1999, 6).

(1968, 255), for example, suggests that since the ending *-or* already exists in British English, in words such as *tailor* and *actor*, it would be perfectly natural to discard the British *-our* in words like *colour* and *honour*, and apply the shorter form *-or* everywhere, just like in American English. “This minor reform would incidentally benefit the conscientious Englishman trying to decide whether he should write ‘Pearl Harbour’ to please himself or ‘Pearl Harbor’ as a compliment to the American spelling” (Foster 1968, 255).

It is true that such seemingly small differences in the different varieties of English around the world are, perhaps, unnecessary, and often result in confusion or in the need to produce several versions of the same text; one for the British, one for the American market, et cetera. Several publishing houses and periodicals, especially those operating in the academic field, recommend using *-ize* instead of *-ise* whenever possible (see 4.5 for more), either to avoid confusion or to appeal to a wider audience. However, the differences between global varieties of English go beyond the *-ise/-ize* debate and even orthography in general. Some differences are rooted in vocabulary and grammar, and in comparison to them, the choice between *-ise* and *-ize* seems insignificant.

There have been some active campaigns during the twentieth century to simplify British English spelling and to adopt some of the American spelling standards. The English Spelling Society, formerly known as Simplified Spelling Society, was founded in 1908 by philanthropists and educational reformers, and it aims to improve literacy and to bring attention to issues regarding spelling (The English Spelling Society 2015). Upward (1997b, 30-32), in a paper for The English Spelling Society, lists some of the most obvious advantages of a single set of spelling standards, and claims that bringing spelling closer to pronunciation would benefit both native speakers of English and learners of the language. More uniform standards and simplified spelling would also reduce costs to publishers, since they would not need to produce separate editions in different English speaking countries (1997b, 30-32). In Upward’s opinion, letters should be used to represent speech-sounds so that writers are not forced to check dictionaries or style guides to be able to spell

correctly. He believes that every step towards a more predictable sound-symbol correspondence is an improvement (Upward 1997a, 13-20).

If campaigns such as this gain enough attention, perhaps *-ize* will eventually override *-ise* as the standard also in British English. However, not everyone believes that proposals made by these reformists will have any effect, and, according to Scholfield (1994, 65), most linguists agree that the current spelling system is not as bad as it seems. Indeed, as Katamba (2005, 217) points out, The English Spelling Society and its American equivalent, Spelling Reform Association of America, have made little progress in their hundred or so years of existence because they have tried to introduce alternative spelling systems. Of course the question of *-ise* versus *-ize* is hardly analogous with changing the whole spelling system of English, and the question regarding the suffixes could be a topic that is already discussed outside these associations. Nevertheless, the suggestions made by these groups of reformers seem to not get heard.

In addition, “it seems that writing is less of an automatic activity than speech and that spelling is more of a conscious choice, so that change is more vigorously resisted in this domain than elsewhere in the language” (Foster 1968, 256). If, during the past few decades, the use of *-ise* has increased to a point that it is now the more common variant in British English, surely it would take a few generations of writers for it to disappear in favour of the *-ize* spelling, and even in that case there would first need to be an official acknowledgement of the issue, and a willingness to implement changes regarding it. Without a general agreement among all or most of the entities that could have some direct or indirect influence in this matter, that is, official entities like the government, counties and councils, leading publishers, printers and word processing software manufacturers, et cetera, it would be difficult to imagine a drastic change happening on its own over time. However, the arguments for spelling reforms do contain some important points: a closer correlation between spelling and sound would benefit language learners and decrease the level of illiteracy, and simplifications could also save money as the writing process would become more efficient (Katamba 2005, 214-215).

4.4 International Varieties of English

One could assume that American influences of the *written* form have been marginal in British (popular) culture before the emergence of the Internet. However, an awareness of the differences between British and American English grew after the 1930s with the increasing popularity of films (McArthur 2006, 375) and the twentieth century, especially since the Second World War, was dominated by American English in many fields. New vocabulary and new meanings of words have, for example, spread unnoticed into British English via the British press who used material from American news agencies (Foster 1968, 38; Quirk 1972, 29). The twentieth century saw American English becoming increasingly fashionable and gaining prestige, especially among younger Britons (Quirk 1972, 25).

In light of the present rather one-sided cultural exchange between the United States and Great Britain it would be easy to assume that American influences, also in orthography, are spreading rapidly into British English. It would also seem plausible that younger generations would tolerate and use Americanised forms more often than older generations, since youth culture on both sides of the Atlantic derives very much from the same foundations. It has been shown that young adults are prone to adjust their vernacular to that of the mainstream society in order to gain acceptance (Bailey 2004, 324), and so it might be likely that young British writers would to some extent adopt the spelling norms that nearly everyone outside Britain and the British Commonwealth follow.

In mainland Europe British English has continued to be the model among teachers and other conservative users of the language, at least up until the end of the twentieth century. However, younger speakers have tended to turn more towards American English (McArthur 2006, 375). American English usage and slang has become very easy to adopt, and the continuous and ever increasing exposure to Americanisms will no doubt continue to have profound consequences in Britain and elsewhere. Indeed, Foster believes that the impact of American English is “the greatest

single influence shaping [British English] today” (1968, 14-15). Nearly fifty years later, the comment still carries weight.

However, as noted by Graddol (1997, 57), British English is far from unimportant compared to American English, even on a global scale, as most areas where English is spoken as a second language still have an orientation towards British English. British publishers are among some of the largest in the world, and even some American companies have been found to use British English in order to gain acceptance in some parts of the world (Graddol 1997, 57). Indeed, adapting one’s spelling to match international varieties of English for commercial purposes is not an unknown concept in the world of business. For example Scott (2004, 153-5) lists some of the differences between British and American English from the point of view of business communication, and discusses whether or not changing the style of writing is worth the effort:

To influence groups positively and to build and maintain their loyalty in the fickle marketplace, business communications such as promotional materials and product user manuals should conform to customers’ needs and expectations, including culturally acceptable spelling; otherwise, the customers are alienated. ... Thus, in at least a number of circumstances, accommodating for spelling differences is prudent business practice that has potential to keep customers satisfied and to affect positively the bottom line. (Scott 2004, 162)

The question of *-ise* versus *-ize*, however, is not as straightforward in this respect since, as Scott himself acknowledges, both of the spellings are used in British English (Scott 2004, 158).

In Australian English both variants are used, but the *-ise* form is far more common than *-ize*. The *Macquarie Dictionary*, the ‘national dictionary’ of Australia, considers both spellings as acceptable, but gives priority to the *-ise* form (Delbridge 2001, 305). This, according to Delbridge, is mainly because *-ise* has for several decades been recommended in the Style Manual published by the Australian Government Publishing Service (Delbridge 2001, 307). The *-ise* spelling is also the standard in Australia’s press today (Fritz 2010, 258). However, some variation is found within Australia, for example in education departments in different parts of the country (Peters 2007, 431).

4.5 *-ise* and *-ize* According to Dictionaries, Usage Guides, Printers and Publishers

[H]ow can one set of facts have more authority than another? It seems very unlikely that Oxford University Press, say, would accuse its rivals of presenting a mere catalogue of errors or deliberate lies. Covertly, then, a claim to (greater) authority must rest on values rather than facts. Oxford is, in essence, presenting itself as Coke to other dictionaries' Pepsi ('the real Thing') – a matter of image, not substance (Cameron 1995, 50).

Dictionaries and usage guides carry much weight in shaping the public opinion of what is 'correct'. The general public considers dictionaries, grammars and handbooks as authorities, and often look for strong opinions instead of a descriptive approach. "If, for example, lexicographers (dictionary-makers) attempt to remove all traces of value-judgment from their work and refuse to label particular usages (such as *ain't*) as 'colloquial' and others as 'slang', there is likely to be a public outcry" (Milroy and Milroy 1999, 4). However, dictionaries and usage guides are much less prescriptivist than they used to be in the past. Also printers and publishers have much influence, but this is probably more covert from the point of view of the general public who mainly consume these texts as opposed to creating them.

The attitudes towards the co-existence of *-ise* and *-ize* have been very colourful in the past century or so, and this section will focus on the way the two spellings have been treated in dictionaries, usage guides and by printers and publishers from early twentieth century to the present.

4.5.1 *Dictionaries*

Dictionaries can be very influential, especially where the non-academic public is concerned. In order to see whether both spelling conventions are recognised in different dictionaries, several works from different publishers were consulted. Entries for the ten most common verbs with the dual ending, according to the BNC, were searched in the dictionaries (see Table 4 in Chapter 5).

In all the dictionaries studied, priority is given to the *-ize* spelling of the words. In some cases this may be the only spelling given, but most dictionaries give the alternate *-ise* spelling alongside the actual entry. The main entry for the suffix itself is always *-ize*, but most dictionaries give *-ise* a short entry of its own, either just to redirect the readers to the main entry or to note that *-ise* is a

British (and Australian) variant. Or, in the case of some dictionaries, the separate entry is for the French derived ending for words that can only end in *-ise*.

In *Collins Dictionary of the English Language* (1986) both variants are listed under the *-ize* spelling of the words as equals without further explanation: e.g. “*criticize or criticise*”. The suffix *-ise* gets recognised as a variant, but the main entry is under *-ize*. The same approach is used in later dictionaries from the same publisher, *Collins Cobuild English Dictionary* (1995) and *Collins Cobuild Advanced Learner’s Dictionary* (2006). Although the *-ise* spelling of each word is also given in its full form in the newer dictionaries, that is, spelled out in full, it is given less of a prominent place within the entry and is labelled as a British variant: “*apologize*, in [British English] also *apologise*”. In the 2011 edition the two spellings appear once again side by side as full words under the *-ize* spelling.

In *Longman Dictionary of the English Language* (1984) the alternate spelling is given within the main entry of each word, though not as a full word but as a suffix following the *-ize* spelling: “*apologize, -ise*”. The suffix itself gets multiple entries: *-ise* is first listed on its own as the British variant of *-ize*, and the main entry lists both *-ize* and *-ise* side by side. Some discussion on the usage of the two spellings is given in the main entry. A newer edition of the dictionary from 1995 is similar. *Longman Dictionary of Contemporary English* (2003), however, differs from the two slightly: although both variants are given under the *-ize* version of the words and *-ise* gets a brief entry of its own like before, there is no further discussion on the usage of the two spellings and it is only noted that *-ise* is also used in British English. Also *Cambridge International Dictionary of English* (1995) and its third edition from 2008, *Cambridge Advanced Learner’s Dictionary*, give both spellings under one entry, stating that the *-ise* form is British (and Australian).

In the *OED* (1989), which consists of multiple volumes, only the *-ize* spelling of each word is given in each entry, though the texts samples within may contain either spelling. However, under the entry of the suffix itself the alternative spelling is given. *The Concise Oxford Dictionary of Current English* (1964) and *Oxford Advanced Learner’s Dictionary of Current English* (1989), both

concise, one volume dictionaries, on the other hand, give the variant in the main entry of each word (*apologize*, *-ise*), as does the newer edition of the latter from 2010. *The Oxford Dictionary of English* (2003), a corpus-based dictionary, on the other hand, takes an even more diplomatic approach and lists the two possible spellings in the same entry as equal forms: e.g. “*criticize* ALSO *criticise*”.

Descriptivism seems like the most popular approach today, and although all one volume dictionaries that were consulted give both spellings under the main entry, either as whole words or as a suffix following the main entry, the duality of the spelling is sometimes overlooked and left without explanation. In these cases it is usually only stated that in British (and Australian) English the *-ise* spelling is also used. However, most of the dictionaries do discuss the issue, and may even give a brief etymology of the suffixes.

4.5.2 Usage Guides

The opinions of language and usage guides and their compilers seem to have long fluctuated between the two choices, and for various reasons. While most guides today are descriptive in nature, and advise that whichever spelling one chooses of the two, it should be used consistently (e.g. Greenbaum and Whitcut 1988, 392), in the past writers of these guides have been more strongly in favour of one or the other choice.

One of the most cited and consulted guides to English usage since the early twentieth century has probably been Fowler’s *Dictionary of Modern English Usage* from 1926. Fowler and his colleagues Ernest Gowers and Eric Partridge are among the authorities who several members of the public have turned to in their questions of what is good English (Bex 1999, 91). Even if their work may not have had a significant impact on academic debate on the subject, they have been very influential in shaping the public perception of what is considered standard and ‘correct’ (Bex 1999, 91).

The numerous reprints and editions of Fowler's *Dictionary of Modern English Usage*,⁴ for example, show that it has had a wide appeal since its first publication in the 1920s up until the present day (Bex 1999, 93). Fowler himself argued for choosing *-ize* for etymological and phonological reasons, but perhaps his view was also influenced by the fact that his publisher was Oxford University Press (Peters 2007, 431). Although many of Fowler's contemporaries, at least in the form of printers in Britain, often opt for the simpler *-ise* spelling (Fowler 1926, 306), to him, *-ize* has more prestige: "the *OED* of the Oxford University Press, the *Encyclopaedia Britannica* of the Cambridge University Press, *The Times*, & American usage, in all of which *-ize* is the accepted form, carry authority enough to outweigh superior numbers" (Fowler 1926, 306). He acknowledges the fact that there are many exceptions to be memorised if keeping with the *-ize* spelling, and a number of verbs ending in *-ise*, such as *advertise*, *devise* and *surprise*, have nothing to do with the Greek suffix *-izein*, the etymological reason for choosing *-ize*. "The difficulty in remembering which these *-ise* verbs are is in fact the only reason for making *-ise* universal, & the sacrifice of significance to ease does not seem justified" (Fowler 1926, 306). Printers' preference for *-ise* was expressed in a more colourful way by Ernest Gowers in a revised edition from 1965: "Most English printers, taking their cue from Kent in *King Lear*, 'Thou whoreson zed! Thou unnecessary letter!', follow the French practice of changing *-ize* to *-ise*" (Fowler 1965, 314).

Eric Partridge, who wrote his *Usage and Abusage* in 1947, has great admiration for Fowler's pedantry and agrees that where two choices are given, *-ize* should be used (Partridge 1957, 162). *An A. B. C. of English Usage* by Treble and Vallins, on the other hand, balances between the two spellings. While the guide prescribes the use of *-ise* to its readers on the grounds of simplicity, in the book itself *-ize* is used in many verbs because "the Oxford University Press, together with many other printers, prefers the *-ize* in those verbs whose etymology demands it" (Treble and Vallins 1936, 107). It should be noted that *An A. B. C. of English Usage* was published by Oxford

⁴ Reprints or new editions of the dictionary were published in 1930, 1934, 1952, 1958, 1959, 1965 (revised by Ernest Gowers), 1966, 1968, 1977, 1978, 1982, 1983, 1984, 1988, 1944 and 1996 (revised by Robert Burchfield) (Bex 1999, 93). A pocket version, edited by Robert Allen, was published in 2003, with a second edition in 2008 (Fowler 2009, xxv).

University Press, but whether the choice of the authors to use *-ize* was made independently or only for the benefit of the publisher is not clear. “In ordinary writing the point is of little or no importance; in writing for print one is justified in leaving the decision to the printer, who settles the matter according to the rules of his house” (Treble and Vallins 1936, 107).

The suggestion in *An A. B. C. of English Usage* that *-ise* should be used is echoed and referred to in Gowers’ *The Complete Plain Words* (Gowers 1962, 235), which was first published in 1948 as *Plain Words*. Although Gowers’ guide has not become quite as iconic as Fowler’s *Dictionary of Modern English Usage*, it has also gone through several editions and reprints (Bex 1999, 100). The *Plain Words* guides were created by the request of Her Majesty’s Treasury, initially meant for those working in civil service, though later popular with the general public as well (Preston 2014), which means that Gowers’ advice must have carried much weight throughout the country. A guide was much needed because the British civil service and the number of official documents created expanded greatly as a result of the Second World War, and the new personnel responsible for these documents were less well educated in literacy skills than their predecessors (Bex 1999, 102).

Vallins, who had already expressed his opinion in the 1920s in *An A. B. C. of English Usage*, which he co-wrote with Treble, continues to promote the *-ise* spelling in the 1950s with *Good English: How to Write It* (1951) and *Better English* (1955). In the former he introduces the topic of *-ise* and believes that the issue could be solved altogether, if only writers would stand their ground. According to Vallins, to an average member of the public the question regarding the spellings is of no interest (Vallins 1951, 242).

Pedants and printers keep alive a distinction between *-ise* and *-ize* as verb endings. No one knows why. The ordinary man does not care a brass farthing, and uses *-ise* for them all. If those who write for publications would only stick to their guns and defy the tyranny of the influential Publishing Houses, they would soon bring about a minor but useful spelling reform. An artificial distinction based on an etymological subtlety that cannot be known to the ordinary man is an unnecessary archaism, and ought to be abolished forthwith in the interest of everybody – including printers (Vallins 1951, 242).

In *Better English* he goes on to make further suggestions for spelling reforms. He believes that were printers not reluctant to allow *-ise*, it would quickly become the norm on its own accord (Vallins 1955, 109).

In the 1960s and 1970s the general attitude seems to have been in favour of the *-ise* spelling, at least when looking into usage guides from that period. Wood, in his *Current English Usage* is of the opinion that keeping with the etymological *-ize* confuses writers unnecessarily. “[T]he ordinary writer of English, who cannot be expected to know the derivation of every verb he uses, is puzzled by the system and has constantly to consult a dictionary” (Wood 1965, 128). However, in his view, if a text is to be published, the choice of the spelling should be left to the hands of the printer (Wood 1965, 128).

Copperud’s American *A Dictionary of Usage and Style* (1964) labels the *-ise* ending as characteristically British. *Cassell’s New Spelling Dictionary* (1976) and *The Pergamon Dictionary of Perfect Spelling* (1978), which are both aimed at a British audience, only list the *-ise* spelling, although the latter mentions the alternative in its preface. In fact, whereas the former simply chooses to omit the question entirely, one could argue that by introducing the topic while not explaining the background or the etymology of the issue at hand, *The Pergamon Dictionary of Perfect Spelling* implies that the *-ise* spelling would be more acceptable in British English than the *-ize* spelling: “[w]here alternative spellings exist these have mostly been omitted. In the case of words ending in *-ise*, *-isation*, the *-ize* and *-ization* versions have not been given (nor have they been given as mis-spelt versions since they cannot be counted as such)” (Maxwell 1978, 7).

By the end of the 1970s attempts had been made to adopt the *-ise* ending in all possible cases, though this does not apply to all usage guides from that period. *Cassell’s New Spelling Dictionary*, however, only gives the *-ise* form for all this kind of verbs (Firnberg and Firnberg 1976). Despite the resistance of scholars who saw it as a simplification, it became accepted and even “recommended as a means of avoiding error” (Phythian 1979, 71-72) since so many verbs ending in *-ise* can only be spelled with *s*, both in British and in American English (see 4.1). When

using *-ize* one must constantly keep in mind the numerous exceptions to the rule. Vallins raises the same question:

The natural answer would be to spell them all in *-ise* or *-ize*; but a queer conservatism, mainly on the part of printers, supported by the OED, forces us back on an etymological distinction which few of us are capable of making, offhand at any rate. The pundits say that words derived from the Greek suffix *-izein* should be spelt with the *-ize* ending; the others in *-ise* (Vallins 1965, 35).

At present, usage guides tend to not prescribe either choice and only advise their readers to be consistent in their spelling (e.g. Swan 2005, 550; Burt 2002, 104) or to avoid unnecessary new coinages like *tenderise* (Amis 1997, 113).

Although one would think that in the age of the Internet, when so much knowledge is just a click away, usage guides, at least in the printed form, would have become obsolete. However, even today such guides are being printed, reprinted and purchased. For example, new editions and prints are still made from classic works such as Fowler's *A Dictionary of Modern English Usage*, which has seen three revised editions just in the twenty-first century: in 2003 by Robert Allen, in 2009 by David Crystal and in 2015 by Jeremy Butterfield. Languages and the rules within still seem to be relevant topics.

4.5.3 Printers and Publishers

Regardless of the personal style of the author, external influence has often been very important in the matter of *-ise* versus *-ize*. In the past, most writers would have had no say in the matter as the choice was eventually in the hands of the printer (Vallins 1965, 152). Today, authors are usually asked to apply to their texts the 'house rules' of the publication they are writing for. For the most part these rules are flexible and allow some choice, as long as the authors are consistent with their chosen style. Some publishers, however, advise that specific rules should be followed.

According to Gramley and Pätzold (2004, 281), the decisive factor in the *-ise/-ize* divide has, in fact, been publishers' preference for *z*. Since several large houses operate simultaneously in Britain and in the United States, it seems natural to encourage uniformity in orthographic choices.

In fact, it is cost effective since there would be no need for separate editions on either side of the Atlantic, although the *-ise/-ize* question is hardly the only orthographic feature dividing British and American English.

Publishers have also been known to change their policies over time, and it seems that in the question regarding *-ise* and *-ize* they have been quite unresolved. Some influential publishers, such as the Oxford University Press, have used the z-spelling consistently for several decades, and still do. However, many publishers in Britain now use *-ise* instead of *-ize* (Burchfield 1996, 422), or leave the choice to the author. Cambridge University Press, for example, having first changed their stance from *-ize* to *-ise* (Gowers 1962, 235; Burchfield 1996, 422), now advise their authors to use either, as long as they are used consistently (Cambridge University Press 2015a). Also, contrary to their current advice, numerous Routledge publications from the 1990s use the *-ise* spelling. Some of these policies may be based on pure principle or etymology or they are, perhaps, followed in order to continue a long-standing in-house tradition, but at least Routledge's *Instructions for Authors* says *-ize* should be used "for the benefit of the US market" (Taylor & Francis 2001, 18).

If one looks at the policies of printers and publishing houses from a larger perspective, it seems that the stance of publishers may have been even more complicated during the past few decades, or even the past century, than one might think. Table 2 below, collected by Walker (2001, 101-102), demonstrates how the recommended spellings have varied from publisher to publisher, and decade to decade, starting from 1895.

Neill & Co (1895)	-ise
Chiswick Press (1913)	-
Spottiswoode, Ballantyne & Co. Ltd (1926)	-ise
London School of Printing (1947)	-ize
Penguin (1947)	-
Labour Party (1948)	-ize
T. Nelson & Sons (1948)	-ise
Curwen Press (1950)	-
Lund Humphries (1950)	-ise
Tillotsons (1952)	-ise
Jonathan Cape (1960)	-ise
Longman (1964)	-ise
Tillotsons (1965)	-
Staples (1966)	-ize
HMSO ⁵ (1970)	-
Penguin (1972)	-ize
Balding & Mansell (1972)	-ize
Curwen Press (1973)	-ize
Routledge (1974)	-ize
Monotype (1974)	-ise
BS 5261 ⁶ (1974)	-
Butcher (1975)	-
MHRA ⁷ (1978)	-ize
CBE ⁸ (1994)	both

Table 2 Conventions recommended in printers' and publishers' style manuals published in the UK.

The selection of printers and publishers in the table below is far from a conclusive one, and the information provided is most likely outdated, but the results are very revealing nonetheless, and demonstrate how the spellings have been distributed in the past.

Both *-ise* and *-ize* appear on the list eight times, one publisher accepts either spelling, and seven manuals dismiss the question altogether. In conclusion, it seems that the question regarding the choice between of *-ise* and *-ize* has been a complicated one for publishers and printers for quite some time.

⁵ Her Majesty's Stationery Office.

⁶ British Standards Institution.

⁷ Modern Humanities Research Association.

⁸ Council of Biology Editors.

4.6 The British News Media

It is difficult to tell just how much newspapers and other media can have an influence in the development and change of a language. Milroy and Milroy (1999, 25) state that “although radio, film and television may not have had much influence on everyday speech, they are amongst the many influences that promote a consciousness of the standard and maintain its position.” It is, indeed, probable that the conventions and forms the public is regularly being exposed to will inevitably influence their perception of what is ‘correct’ or ‘proper’ usage.

The current position of the majority of British newspapers and other news media is to use *-ise* instead of *-ize*. For example, *The Guardian and Observer style guide* advises that *-ise* should be used (The Guardian 2015), and so does the Telegraph style book (The Telegraph 2008) as well as that of BBC News (BBC Academy 2015).

A search through The British Newspaper Archive, which currently contains around three million pages of newspaper content published in Britain between 1700 and 1999, reveals that the *-ise* spelling has, indeed, been the dominant one in British newspapers for a considerable long time. Table 3 below shows that for two of the three words searched in the database, the shift in style happened as early as in the late eighteenth century and early nineteenth century. According to the data, in the case of *realise* the shift occurred a little later: between 1800 and 1849 the *-ize* spelling was very much the norm with usage at 87 per cent, but in the following fifty-year period the tables have turned and *-ise* is the more popular style at 86 per cent.

	Number of articles with word matches											
	realise				recognise				organise			
Date	ise	ize	ise %	ize %	ise	ize	ise %	ize %	ise	ize	ise %	ize %
1700-1749	18	8	69 %	31 %	20	85	19 %	81 %	0	0	0 %	0 %
1750-1799	271	1858	13 %	87 %	505	2314	18 %	82 %	19	211	8 %	92 %
1800-1849	21586	149048	13 %	87 %	201502	69802	74 %	26 %	12963	8990	59 %	41 %
1850-1899	1934300	306793	86 %	14 %	1560262	156637	91 %	9 %	126006	22972	85 %	15 %
1900-1949	1876338	151328	93 %	7 %	1068591	74105	94 %	6 %	137490	11599	92 %	8 %
1950-1999	46040	9403	83 %	17 %	25552	5014	84 %	16 %	5422	1077	83 %	17 %

Table 3 The distribution of *-ise* and *-ize* in the British Newspaper Archive.

As can be seen in Table 3, not all papers had abandoned the *-ize* spelling by the latter half of the twentieth century, as it is still found in 16 or 17 per cent of the cases in the years between 1950 and 1999. One good example is *The Times*, a prestigious British newspaper that, until quite recently used to be among the models and language authorities who preferred the *-ize* spelling where it was possible. However, in the 1990s there was a sudden change of strategy as the paper adopted the *-ise* spelling, which it is also using today. Richard Dixon, former Chief Review Editor for *The Times*, explains that up until the early 1990s the policy of the paper was to use *-ize* wherever it was etymologically correct. The 1992 *Style Guide* of *The Times*, however, opted for simplicity and instructed writers to avoid *-ize* and *-ization* (Dixon 2004).

4.7 Influence of the Public Sector

Most British authorities and the educational system today seem to favour *-ise* instead of *-ize*, although direct policies for the use of either are not easy to find for all institutions. The British government has chosen to use the *-ise* spelling, as stated in their Digital Service style guide under Americanisms. “Use the ‘ise’ rather than ‘ize’ suffix, eg organise not organize (this isn’t actually an Americanism but is often seen as such)” (Government Digital Service 2015). This is also true for the Department for Education who use *-ise* in throughout their documents. Although no official guidelines are given to teachers or other educators regarding the spelling of *-ise* and *-ize* in the National Curriculum in England, for example, the documents themselves use the *-ise* spelling (e.g. Department for Education 2013).

The style chosen in school textbooks will most likely influence children learning to read and write just as much as the one used by their teachers. One of the leading textbook publishers in Britain, Schofield & Sims, answers the question regarding the spelling choice to those using their teaching material thus: “Many verbs can be spelt with either *ise* or *ize*. ... Rather than learning by heart the words that must use *ise*, it can be easier to simply use the *ise* spelling for all words. In US spelling, however, *ize* is the standard form” (Schofield & Sims 2015).

Official documents for the European Union that are translated into English also only use *-ise* (European Commission Directorate-General for Translation 2011). The National Health Service (NHS) is another considerable public service entity also advocating the use of *-ise* (and *-isation*), except for proper nouns (Bolton 2008, 10).

Dale (2013) notes that “large, influential organizations such as London Transport invariably used the ‘*-ize*’ form in posters and other public communications”. However, the current position of London Transport is to use *-ise*, with the exception of proper nouns like company names (Transport for London 2015).

Although the list of establishments and authorities above is far from a conclusive one, they nevertheless form a formidable group. These examples show that many British public service operators do indeed favour the use of *-ise*, and it is very possible that the preference has become deeply rooted in society, at least in recent years.

4.8 Influence of Spell Checkers and Other Software

Since the dawn of word processing software for computers and lately, smart phones, choices made by those who create them have influenced the way people write and also, perhaps, what they perceive as correct usage. Automatic spell checkers and other word processing software can have a profound impact on written language because they are so widely used (Hogg and Denison 2006, 33), and often their suggestions are accepted without criticism by their users. What these spell checkers suggest may easily become to be believed as the norm.

In the case of *-ise* and *-ize*, the problem with word processing software and their spellcheckers is at least twofold. Some of them incorrectly ‘correct’ *-ize* to *-ise* when British English is chosen as the language (e.g. Ask Different 2011), and may thereby change people’s perception of what is correct. Others accept either spelling, as is the case with current versions of Microsoft Word, for example, but problems arise because they do not then mark either style as incorrect even when used in the same document. Therefore a piece of text may contain both spellings even if the author has meant to be consistent in their choice.

4.9 Attitudes Towards the Variation

Arguments for and against the two spellings have occasionally been rather heated. Although it seems that many authorities now say that both *-ise* and *-ize* are acceptable, some have been reluctant to acknowledge *-ise* as a serious alternative. “But the suffix itself, whatever the element to which it is added, is in its origin the Gr[reek] *-ιζειν*, L[atin] *-izāre*; and, as the pronunciation is also with *z*, there is no reason why in English the special French spelling should be followed, in opposition to that which is at once etymological and phonetic” (*The OED*, 1989). Those who use the *-ise* spelling throughout may have even been labelled as ‘lazy spellers’ who cannot be bothered to learn all the exceptions to the spelling rules (AskOxford 2008).

On the other hand, some have wondered if there is any sense in holding on to complicated etymological differentiations, which may not always be clear to the general public (Vallins 1965, 35) or, in fact, anyone without extensive knowledge in classical languages or etymology.

As stated earlier, there is a widespread misconception in Britain that the *-ize* spelling is an Americanism and that is why *-ise* should be used (e.g. Horne 2012, Oxford Dictionaries 2011). The origin of this belief may lie in the pervasiveness that the *-ise* spelling has in society, and could also indicate that the educational system has long favoured *-ise*. People have become accustomed to it, and since it differs from the American standard, *-ize* has become ‘foreign’.

Choosing *-ise* over *-ize* is, perhaps, an easy way to differentiate oneself from Americans and to underline one’s Britishness, even if the reason behind the choice is false. “What is particularly strong in Britain ... is the passion to preserve our language from Americanisms” (Whitcut 1985, 160). Indeed, it has also been suggested that some people not only seem to hate something if it is American, but also assume something is American because they hate it (Crystal 1981, 37-39). Choosing one style over another could therefore have become a means of expressing one’s identity and nationality. Whether or not this attempt is successful, or based on the right reasons, is another question entirely.

5 Corpus Study: *-ise* and *-ize* According to the BNC and BAWE

5.1 The BNC

In order to study the distribution of *-ise* and *-ize* in British English verbs in the BNC, the ten most frequently used verbs that can take either ending were searched in the database. As explained earlier in Chapter 2, this was done because it was not possible to search the corpus in a way that would have included all those verbs that can be spelled both with *-ise* and *-ize* and exclude those that cannot, and still retain all the search methods that the corpus allows. A lemma query revealed that the most common verbs in the BNC that can be spelled with either *-ise* or *-ize* are *realise*, *recognise*, *organise*, *emphasise*, *criticise*, *characterise*, *specialise*, *summarise*, *apologise* and *minimise*. These were the ten most frequent examples for both spellings, and this makes comparisons between two spellings of one word easier.

The BNC consists of 4000 individual samples of language, 10 per cent of which are of spoken language. While the remaining 3140 texts form a fair number of samples for a corpus, the findings may not be very representative when breaking the data into smaller and more detailed portions based on all the search methods the corpus allows. This needs to be taken into account when analysing the corpus findings. For example, for the most part the age of the authors is unknown: only 623 of the 3140 texts in the BNC are categorised based on the age of the writer (see Appendix 2), which adds up to just 20 per cent of all samples. In a significant majority of cases this information is not specified, which means that the source material for an age comparison is very limited.

As can be seen in the lemma query results of Table 4 below, the usage of *-ise* and *-ize* in the BNC data seems to vary slightly depending on the word. Interestingly, at least according to this table, the *-ise* spelling is particularly strong in the case of *specialise*, *apologise* and *minimise*, whereas for *criticise* and *characterise* the two spellings are more equally distributed.

Lemma	Words		Percentages	
	ise	ize	ise %	ize %
realise	9015	4565	66 %	34 %
recognise	8799	5388	62 %	38 %
organise	5360	2919	65 %	35 %
emphasise	2918	1907	60 %	40 %
criticise	1984	1584	56 %	44 %
characterise	1375	1264	52 %	48 %
specialise	1145	440	72 %	28 %
summarise	1122	665	63 %	37 %
apologise	1010	365	73 %	27 %
minimise	977	481	67 %	33 %

Table 4 Lemmata tagged as verb ending in *-ise* and *-ize* in the written component of the BNC.

This calls for more thorough investigation. Since the percentages listed in Table 4 vary depending on the word in question, does it mean that the suffixes are not mutually interchangeable in British English, or that it is more acceptable to use one or the other depending on the word? Indeed, there seems to be noticeable variation in usage when comparing different domains, mediums and text types, as will be explained later in sections 5.1.5, 5.1.6 and 5.1.7. This can explain some of the differences between the results when comparing the ten words with each other. Overall, the *s*-spelling seems to be the more popular one for these ten verbs, the average being 64 per cent.

The lemma query above, however, is a rather crude source of data. Although the percentages of usage between the ten example words seem comparable with each other, it would be unwise to rely on the number of individual occurrences of words in the entire written part of the corpus. One text could contain more examples of the words studied than another. Thus the spelling choice in one text source, or made by one person, if multiplied, could tilt the scales in a specific direction, thus distorting the accuracy of the numbers. It could easily make it seem like their choice is the more common one throughout, even though the result would only represent one individual.

In order to see how the spellings are distributed between different texts and, ultimately, between individual authors, as opposed to comparing word frequencies in the entire database, the verbs were analysed based on the number of separate texts in which they were found. These

numbers will better represent individual authors or text samples rather than merely the frequency of certain words within the entire written part of the corpus. The results of these queries will be presented in the following sections, starting with a general look into how the example words and their two spellings are distributed in the entire written part of the corpus. The other six sections will examine the use of *-ise* and *-ize* from several specific points of view, and they are time, age and gender of the author, and the medium, text type and domain of text. To ensure that every possible occurrence of the words was taken into account, all the inflectional forms of the ten verbs were included in the searches; the 3rd person singular *-s*, the past participle *-ed*, and the progressive *-ing*.

5.1.1 Unrestricted Search

In the entire written part of the BNC, the *-ise* spelling is found in 2351 different texts (68 per cent) and *-ize* in 1091 texts (32 per cent), when all the ten example words are combined in one search. The highlighted column in Table 5 below, *true percentages* refers to the results that include overlapping spellings: one text can contain either one or both of the two variants. This means that percentages calculated by simply adding the texts containing *-ise* to the ones with *-ize* can only demonstrate how the two spellings relate to each other. In order to see how the two are truly distributed within the texts, percentages need to be calculated from the number of texts containing *either* spelling. This method of presenting percentages will be also be used in all the following chapters discussing the findings in the BNC.

All ten words combined in one search								
Number of texts containing					True percentages		Overlap	
ise	ize	either	ise %	ize %	ise %	ize %	Texts	%
2351	1091	2868	68 %	32 %	82 %	38 %	574	20 %

Table 5 Texts containing *-ise* or *-ize* in the BNC, an overview.

As can be seen, this is far from an inconsequential distinction: according to these results, *-ise* can be found in 82 per cent of all texts containing either of the spellings, and *-ize* in 38 per cent.

The comparison also reveals that overlap in the spellings is very common indeed: from the 2868 texts where either of the two spellings were found, 574 contained both, which means that 20

per cent of all texts with word matches had two competing spellings of one or several of the words studied. Some of the overlap in the numbers may be explained by author error and it could simply be a case of misspelling, but there could be other reasons as well, especially since this phenomenon seems very common. An author may have chosen the *-ise* spelling for one word and *-ize* for another, perhaps as a conscious choice, perhaps subconsciously. Other possible explanations for the cases of overlap, besides author error or some other mistake, could include the possibility that a text sample in the database contains content from multiple authors, for example in the form of a direct quote from another source. Indeed, a closer inspection of the results of some of the corpus queries revealed that some text samples do contain quotations from outside sources.

When looking at the results word for word, it becomes clear that although this overlap of two spellings is far less ordinary than when all the ten example words are combined in a single search, it is not an uncommon phenomenon. Table 6 below shows that 2 to 8 per cent of texts contain both spellings of one particular word. It seems that the more common the word is, the more overlap occurs.

All ten words combined in one search					
	Number of texts containing			Overlap of spellings	
	ise	ize	either	texts	%
realise	1598	752	2182	168	8 %
recognise	1707	838	2359	186	8 %
organise	1418	689	1965	142	7 %
emphasise	1000	529	1431	98	7 %
criticise	779	426	1131	74	7 %
characterise	530	372	854	48	6 %
specialise	849	377	1181	45	4 %
summarise	469	263	713	19	3 %
apologise	461	190	639	12	2 %
minimise	492	255	721	26	4 %
all combined	2351	1091	2868	574	20 %

Table 6 Texts containing *-ise*, *-ize* or either in the BNC.

The unrestricted word search in the written part of the corpus, when presented in percentages, reveals that the results are very different depending on the viewpoint. As with Table 5 with the overall results, if one calculates the percentages by adding the occurrences of the two

spellings together, the results do not completely equate with the percentages calculated from the number of texts where *either* of the spellings were found. These *true percentages*, adjusted for overlapping spellings, which are highlighted in Table 7 below, show that *-ise* is even more popular within the texts than indicated by the other two columns.

All ten words combined in one search				
	ise	ize	ise	ize
	%	%	true %	true %
realise	68 %	32 %	73 %	34 %
recognise	67 %	33 %	72 %	36 %
organise	67 %	33 %	72 %	35 %
emphasise	65 %	35 %	70 %	37 %
criticise	65 %	35 %	69 %	38 %
characterise	59 %	41 %	62 %	44 %
specialise	69 %	31 %	72 %	32 %
summarise	64 %	36 %	66 %	37 %
apologise	71 %	29 %	72 %	30 %
minimise	66 %	34 %	68 %	35 %
all combined	68 %	32 %	82 %	38 %

Table 7 The frequency of *-ise* and *-ize* in the BNC, presented in percentages.

Not all of the ten verbs selected for this study seem to behave in the same way, and there is some variation in usage between them. In the case of *characterise*, for example, the difference between the frequencies of the two spellings is not as dramatic as with most of the other words: the *-ize* spelling is the strongest of the ten, at 41 per cent against the 59 of the *-ise* spelling. The rest of the words behave in a more uniform way, at least according to these overall findings.

To summarise, according to data gathered from an unrestricted word search in the written part of the BNC, it seems that the *-ise* spelling is used in 59 to 71 per cent of the cases, depending on the word in question, and when cases of overlap of the two spellings is taken into consideration, the *-ise* spelling is even more common. When all the ten words are combined in one search, the results are 68 per cent for the *-ise* spelling against 32 per cent for *-ize*. If calculated from the number of texts where either of the two spellings were found, the percentages for the combined word search are even higher: *-ise* was found in 82 per cent of all texts with matches and *-ize* in 38 per cent of the texts.

5.1.2 Diachronic Comparison

The BNC consists of texts published or written in several different periods in time, and in order to see if the corpus could be used to demonstrate how *-ise* and *-ize* have been used over time, findings from these time periods were compared with each other. The data in the BNC is divided into three specific time periods, and they are 1960 to 1974, 1975 to 1984, and 1985 to 1993. The oldest two periods, however, are not as well represented in the corpus as the latest one. The earliest time category contains only 46 texts in total, and all of these contained one or two of the spellings. The second period is made up of 155 texts, out of which 147 include examples of the words studied. The third group has 2777 texts, 2573 of which are relevant for this study.

The number of texts where the two spellings were found are listed in Table 8 below, which shows how *-ise* and *-ize* are distributed among the texts where matches were found. All three time periods seem to have some overlap in the spellings found, that is, one text contained both spellings. At word level the overlap percentages range between 2 and 13 per cent, and when all the words are combined in one search, from 13 to 20 per cent.

Time	1960-1974					1975-1984					1985-1993				
	Texts containing			Overlap		Texts containing			Overlap		Texts containing			Overlap	
	ise	ize	either	texts	%	ise	ize	either	texts	%	ise	ize	either	texts	%
realise	12	30	39	3	8 %	74	56	119	11	9 %	1484	658	1991	151	8 %
recognise	11	34	44	1	2 %	82	55	128	9	7 %	1553	739	2120	172	8 %
organise	7	19	25	1	4 %	71	48	112	7	6 %	1319	607	1799	127	7 %
emphasise	4	15	19	0	0 %	53	39	87	5	6 %	907	466	1285	88	7 %
criticise	4	5	9	0	0 %	29	25	53	1	2 %	729	389	1046	72	7 %
characterise	4	5	8	1	13 %	39	22	59	2	3 %	476	337	771	42	5 %
specialise	3	10	12	1	8 %	44	26	70	0	0 %	784	333	1075	42	4 %
summarise	2	7	9	0	0 %	36	22	57	1	2 %	390	228	601	17	3 %
apologise	4	15	19	0	0 %	11	10	21	0	0 %	444	165	597	12	2 %
minimise	1	3	4	0	0 %	21	14	35	0	0 %	462	231	668	25	4 %
all combined	15	37	46	6	13 %	108	67	147	28	19 %	2132	968	2573	527	20 %

Table 8 Diachronic comparison of the BNC findings.

While on the whole it seems that at word level it is less common to find two competing spellings of one word within the same text, some overlap does occur even when looking at individual words. The columns highlighted in Table 8 show that between 1960 and 1974, there are a

few instances of overlap in the spellings in half of the words: *realise*, *recognise*, *organise*, *characterise* and *specialise*. The overlap is only minor, possibly due to the small number of source texts in the category. In the time frame, between 1975 and 1984, instances where one text contains two spellings of the same word is slightly more common, but still quite moderate. All but three words, *specialise*, *apologise* and *minimise*, have some overlap in the findings. Between 1984 and 1993, instances of overlap are very noticeable, ranging from 12 and 172 cases, depending on the popularity of the words in question.

As with the findings in section 5.1.1, it seems that the more frequently a word is used, the more overlap occurs. The results are fairly similar in all of the three time periods, although in the oldest category this appears to be less frequent than in the latter two. However, since the first period is so poorly represented in the corpus, the results of this calculation are not to be generalised. The latter two periods give much more convincing numbers for each word in question, so it is possible to make some conclusions based on the results they give. In the question of overlap in the two spellings, not much seems to separate these two periods.

In order to compare and contrast the distribution of *-ise* and *-ize* between the time periods more easily, the frequencies of the two spellings are presented as percentages in Table 9 below. As in section 5.1.1, the first two columns under each time frame show how the two spellings relate to one another, both at word level and as a whole, whereas the highlighted columns give the true percentages that include overlap in the spellings.

When comparing data from a combined search including all of the ten example words with the results of the individual words, there are noticeable differences in the frequencies. If looking at all the ten example words combined, between 1960 and 1974 the *-ize* spelling is the more popular one as it is used in 71 per cent of the cases. It appears that between 1975 and 1984 this has changed radically, and *-ise* has become the more common spelling, dropping the frequency of *-ize* to 38 per cent. This trend continues, and in the third and final time frame, between 1985 and 1993, the roles of the spellings are reversed: *-ise* has taken the top place, and the frequency of *-ize* is down to 31

per cent. Similar patterns emerge when the example words are studied in more detail. Not all words behave in the same way, however, and there are noticeable differences between them.

Time	1960-1974				1975-1984				1985-1993			
	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %	%	%	true %	true %
realise	29 %	71 %	31 %	77 %	57 %	43 %	62 %	47 %	69 %	31 %	75 %	33 %
recognise	24 %	76 %	25 %	77 %	60 %	40 %	64 %	43 %	68 %	32 %	73 %	35 %
organise	27 %	73 %	28 %	76 %	60 %	40 %	63 %	43 %	68 %	32 %	73 %	34 %
emphasise	21 %	79 %	21 %	79 %	58 %	42 %	61 %	45 %	66 %	34 %	71 %	36 %
criticise	44 %	56 %	44 %	56 %	54 %	46 %	55 %	47 %	65 %	35 %	70 %	37 %
characterise	44 %	56 %	50 %	63 %	64 %	36 %	66 %	37 %	59 %	41 %	62 %	44 %
specialise	23 %	77 %	25 %	83 %	63 %	37 %	63 %	37 %	70 %	30 %	73 %	31 %
summarise	22 %	78 %	22 %	78 %	62 %	38 %	63 %	39 %	63 %	37 %	65 %	38 %
apologise	21 %	79 %	21 %	79 %	52 %	48 %	52 %	48 %	73 %	27 %	74 %	28 %
minimise	25 %	75 %	25 %	75 %	60 %	40 %	60 %	40 %	67 %	33 %	69 %	35 %
all combined	29 %	71 %	33 %	80 %	62 %	38 %	73 %	46 %	69 %	31 %	83 %	38 %

Table 9 Diachronic comparison of BNC findings, presented in percentages.

When looking at the words individually, according to Table 9, between 1960 and 1974 the frequencies vary between 71 and 79 per cent in favour of the *-ize* spelling (or between 75 and 83 per cent when overlap in spellings is taken into account), and between 21 and 27 per cent in favour of the *-ise* spelling, apart from two noticeable exceptions. What stands out is how *criticise* and *characterise* behave: contrary to the other words studied, in the case of these two words the *-ise* spellings are already quite common during the period between 1960 and 1974, ranging between 44 and 50 per cent in popularity, depending on whether one looks at the highlighted or the non-highlighted percentages. The number of texts in which these words were found, however, are rather small compared to the others. Table 8 shows that in this category, only nine texts in total had occurrences of either *criticise* or *criticize*, and eight contained either *characterise* or *characterize*. Overall, the period between 1960 and 1974 is the least represented in the whole of the BNC, and therefore the results must be handled with great care.

Comparing the results of the first and third time frames shows that, according to the BNC, considerable changes in usage seem to have happened over time, and between 1975 and 1984 these changes are beginning to show. The frequencies are turned slightly in favour of the *-ise* spelling,

ranging from 52 per cent to 64 per cent at word level. When all the words are combined in one search, the *-ise* spelling is at 62 per cent and the *-ize* spelling at 38 per cent. During this middle period, the words that are most unlike the others are *criticise* and *apologise* and *emphasise*, and in the case of these words *-ize* is still almost as popular as *-ise*.

The third and final time period, between 1985 and 1993, is one of great changes in the *-ise/-ize* divide, at least according to the data above. By that time the popularity of *-ize* had declined to the level *-ise* had between 1960 and 1974. *Characterise* seems to be the word that resists the change the most, as the *-ize* spelling is still used in 41 per cent of all cases.

It appears that while in the case of spelling overlap there seems to be little variation between the different time periods in the BNC data, there are considerable and very interesting changes in the overall frequencies of the spellings. Although the three time periods are not equally well represented in the corpus, the results do indicate that very dramatic changes in the usage of the two spellings happened somewhere between 1960 and 1993. However, since the sample texts in the oldest time category in the BNC are entirely made of fiction, the results of the period between 1960 and 1975 can only explain the situation in the publishing world and, particularly, books. Informative texts were selected only from 1975 onwards (Aston and Burnard 1998, 30). If we then leave the earliest period out of the comparison, the differences between the remaining two periods diminish. Some changes can, however, be seen. Between 1975 and 1984 *-ise* is the choice in 62 per cent of the cases, whereas in the latter period the percentage is up to 69. If overlap in the spellings is taken into consideration, the frequency of *-ise* shows similar growth in time: in the middle time period *-ise* is found in 73 per cent of all texts with word matches, and in the latest period this percentage is up to 83.

When the findings of the three time periods are divided by text type, the bias in the source texts becomes less apparent and the time periods become easier to compare with each other. As can be seen in Table 10 below, the corpus material in the earliest period does indeed consist mostly of *fiction and verse*, and newspaper texts are limited to the latest period.

Time	1960-1974			1975-1984			1985-1993		
	Texts with matches			Texts with matches			Texts with matches		
Text type:	ise	ize	either	ise	ize	either	ise	ize	either
academic prose	1	6	6	30	21	40	251	218	360
fiction and verse	10	24	32	15	19	31	211	171	351
non-academic prose and biography	2	5	5	36	19	46	528	333	665
newspapers	0	0	0	0	0	0	390	10	390
other published written material	2	1	2	11	5	14	557	185	607
unpublished written material	0	1	1	16	3	16	195	51	200

Table 10 Diachronic comparison of BNC findings by text type.

The percentages in Table 11 below show that the use of the two spellings changes in time within all text types that are represented. The greatest differences are, again, between the earliest (1960 to 1974) and the middle period (1975 to 1984): while *-ize* is by far the most popular spelling in *fiction and verse* between 1960 and 1974 (71 per cent), the percentages drop to 56 per cent in the next time period and even lower in the final period, where *-ize* is used in only 45 per cent of the cases. In the case of the earliest time period the rest of the text types are so poorly represented that not much can be deducted from the results. The remaining two time periods, however, although not equally proportioned, are easier to compare in further detail.

Time	1960-1974				1975-1984				1985-1993			
	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize
Text type:	%	%	true %	true %	%	%	true %	true %	%	%	true %	true %
academic prose	14 %	86 %	17 %	100 %	59 %	41 %	75 %	53 %	54 %	46 %	70 %	61 %
fiction and verse	29 %	71 %	31 %	75 %	44 %	56 %	48 %	61 %	55 %	45 %	60 %	49 %
non-academic prose and biography	29 %	71 %	40 %	100 %	65 %	35 %	78 %	41 %	61 %	39 %	79 %	50 %
newspapers	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	98 %	3 %	100 %	3 %
other published written material	67 %	33 %	100 %	50 %	69 %	31 %	79 %	36 %	75 %	25 %	92 %	30 %
unpublished written material	0 %	100 %	0 %	100 %	84 %	16 %	100 %	19 %	79 %	21 %	98 %	26 %

Table 11 Diachronic comparison of BNC findings by text type, presented in percentages.

It seems that although overall, based on the numbers seen earlier in Table 9, the use of *-ise* increases slightly by the latest time period, this is not true for all text types. Although within *fiction and verse* and *other published written material* the use of *-ise* is on the increase even after 1984, the rest of the categories show a slight step backwards, with the exception of *newspapers*, a text type that is only represented in the latest time period, which is probably one of the main reasons why the

use of *-ise* seems to have increased between these two periods when looking at the overall results. For example, when between 1975 and 1984 *academic prose* contains noticeably more *-ise* than *-ize* (59 against 41 per cent), the gap in the usage of the two spellings diminishes slightly by the final time period: between 1985 and 1993, just 54 per cent of the texts favour *-ise*. Similar drops in the popularity of *-ise*, or increases in the popularity of *-ize*, can be found in *non-academic prose and biography* and *unpublished written material*.

5.1.3 Comparing Authors by Age

Authors in the BNC data have been divided into six groups based on their age, but the age of the author is known in only 25 per cent of the texts in the corpus (see Appendix 1). In addition, the sizes of the groups vary greatly, and some of them are very poorly represented. This makes comparing the use of the two spellings by the age of the author quite difficult. Of the total of 3140 individual text samples in the BNC, only three (0.1 per cent) were written by authors aged 0 to 14, and nineteen (0.6 per cent) by authors aged 15 to 24. Because of this, the youngest two age groups are left out of the comparison as it would be misleading to make any conclusions based on the results. However, a summary of the findings is presented in Table 12 below.

Age	0-14					15-24				
	Texts containing			Percentages		Texts containing			Percentages	
	ise	ize	either	ise %	ize %	ise	ize	either	ise %	ize %
realise	3	0	3	100 %	0 %	14	6	18	70 %	30 %
recognise	1	1	1	50 %	50 %	14	7	17	67 %	33 %
organise	0	0	0	0 %	0 %	8	6	13	57 %	43 %
emphasise	1	1	1	50 %	50 %	8	2	9	80 %	20 %
criticise	1	0	1	100 %	0 %	8	5	10	62 %	38 %
characterise	0	0	0	0 %	0 %	5	2	7	71 %	29 %
specialise	0	0	0	0 %	0 %	3	0	3	100 %	0 %
summarise	0	0	0	0 %	0 %	3	0	3	100 %	0 %
apologise	1	0	1	100 %	0 %	4	0	4	100 %	0 %
minimise	0	0	0	0 %	0 %	1	0	1	100 %	0 %
all combined	3	1	3	100 %	33 %	16	12	19	57 %	43 %

Table 12 Comparing authors of different ages in the BNC, part 1/2.

Table 12 shows how the spellings were distributed among these texts. Although the numbers presented cannot be considered representative, especially so in the case of the youngest

group, it would appear that among the 15- to 24-year olds the choice is often *-ise*, even though the percentages given below disguise this. The numbers corrected for overlap in the spellings reveal that from the nineteen texts that contained either of the spellings of any of the words 84 per cent contained *-ise* and 63 per cent contained *-ize*.

The other four groups, on the other hand, while not equal compared to each other, have a much healthier number of occurrences of texts in total and of the spellings studied. Table 13 below shows the distribution of the spellings as numbers of texts.

Age	25-34			35-44			45-59			60+		
	Texts containing			Texts containing			Texts containing			Texts containing		
	ise	ize	either	ise	ize	either	ise	ize	either	ise	ize	either
realise	36	30	57	93	90	165	93	94	175	69	65	125
recognise	30	28	58	81	97	167	98	96	183	71	73	132
organise	21	23	42	52	73	124	77	85	153	52	52	98
emphasise	13	13	26	47	51	95	64	56	115	37	43	76
criticise	13	12	25	35	39	71	38	44	81	30	32	58
characterise	10	9	18	31	33	61	30	39	68	13	22	32
specialise	8	18	26	34	41	74	38	36	72	28	29	54
summarise	6	10	16	28	18	45	23	24	45	15	20	35
apologise	18	10	27	28	28	56	28	23	51	25	25	49
minimise	6	6	12	19	23	40	20	28	46	10	13	23
all combined	39	34	61	108	110	179	120	116	198	90	85	138

Table 13 Comparing authors of different ages in the BNC, part 2/2.

Overlap in spellings does occur in all of the remaining four age groups, as can be seen in Table 14 below, but there are some differences to be found between them. For example, in a combined search of all the words studied, the results comply with those in the general results presented in section 5.1.1, and overlap is somewhere between 19 and 22 per cent, depending on the age group.

Age	25-34		35-44		45-59		60+	
	Overlap		Overlap		Overlap		Overlap	
	Texts	%	Texts	%	Texts	%	Texts	%
realise	9	16 %	18	11 %	12	7 %	9	7 %
recognise	0	0 %	11	7 %	11	6 %	12	9 %
organise	2	5 %	1	1 %	9	6 %	6	6 %
emphasise	0	0 %	3	3 %	5	4 %	4	5 %
criticise	0	0 %	3	4 %	1	1 %	4	7 %
characterise	1	6 %	3	5 %	1	1 %	3	9 %
specialise	0	0 %	1	1 %	2	3 %	3	6 %
summarise	0	0 %	1	2 %	2	4 %	0	0 %
apologise	1	4 %	0	0 %	0	0 %	1	2 %
minimise	0	0 %	2	5 %	2	4 %	0	0 %
all combined	12	20 %	39	22 %	38	19 %	37	27 %

Table 14 Comparing overlap in spellings by the age of author in the BNC.

However, one group stands out: according to the BNC, authors aged 60 or over seem to be more likely to use two competing spellings in one text than the other age groups, the overall percentage being as high as 27. Does it indicate that older authors are more prone to make mistakes, or that have they been used to different spelling standards than the other authors? That is, are they accustomed to use the *-ise* spelling for some words and *-ize* for others? It is possible, but such a question is impossible to answer without inspecting every text sample containing the spellings. When looking at the ten words individually, all four groups are within the range of the general results of section 5.1.1, and the more frequent a word is in the database, the more overlap occurs. In this respect, authors aged 60 or over do not drastically differ from the other age groups.

When comparing the frequencies of the spellings found in percentages, there is barely any variation in usage between the remaining four groups, and it appears that the two spellings are quite equally used in all four when all the words are combined. Table 15 below shows that *-ise* and *-ize* are both used in roughly 50 per cent of the cases when the ten words are combined in one search.

Age	25-34				35-44			
	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %
realise	55 %	45 %	63 %	53 %	51 %	49 %	56 %	55 %
recognise	52 %	48 %	52 %	48 %	46 %	54 %	49 %	58 %
organise	48 %	52 %	50 %	55 %	42 %	58 %	42 %	59 %
emphasise	50 %	50 %	50 %	50 %	48 %	52 %	49 %	54 %
criticise	52 %	48 %	52 %	48 %	47 %	53 %	49 %	55 %
characterise	53 %	47 %	56 %	50 %	48 %	52 %	51 %	54 %
specialise	36 %	64 %	31 %	69 %	47 %	53 %	46 %	55 %
summarise	38 %	63 %	38 %	63 %	61 %	39 %	62 %	40 %
apologise	64 %	36 %	67 %	37 %	50 %	50 %	50 %	50 %
minimise	50 %	50 %	50 %	50 %	45 %	55 %	48 %	58 %
all combined	53 %	47 %	64 %	56 %	50 %	50 %	60 %	61 %

Age	45-59				60+			
	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %
realise	50 %	50 %	53 %	54 %	51 %	49 %	55 %	52 %
recognise	51 %	49 %	54 %	52 %	49 %	51 %	54 %	55 %
organise	46 %	54 %	50 %	56 %	50 %	50 %	53 %	53 %
emphasise	53 %	47 %	56 %	49 %	46 %	54 %	49 %	57 %
criticise	46 %	54 %	47 %	54 %	48 %	52 %	52 %	55 %
characterise	43 %	57 %	44 %	57 %	37 %	63 %	41 %	69 %
specialise	44 %	56 %	53 %	50 %	63 %	37 %	52 %	54 %
summarise	49 %	51 %	51 %	53 %	43 %	57 %	43 %	57 %
apologise	55 %	45 %	55 %	45 %	50 %	50 %	51 %	51 %
minimise	42 %	58 %	43 %	61 %	43 %	57 %	43 %	57 %
all combined	51 %	49 %	61 %	59 %	51 %	49 %	65 %	62 %

Table 15 Age comparison in the BNC in percentages.

Some differences can be found in usage between the words, although, interestingly, the words that differ from the rest are usually different in every age group. For example, in the case of the 25- to 34-year-olds, *summarise* is very often spelled with *-ize* (63 per cent). However, among the 34- to 44-year-olds it is completely the opposite: the preferred choice is *-ise*, which is used in 61 per cent of the cases. On the other hand, the 25- to 34-year-olds spell *apologise* mostly with *-ise* (64 per cent), whereas in the other three groups the two spellings are more equally distributed, and roughly half of the authors use *-ise* and the other half *-ize*.

As the results stand, not much can be said about differences between age groups as regards to the question at hand. Apart from the two youngest groups that were left out of the comparisons

for insufficient data, the two spellings are used quite similarly in all the groups. Only at word level some differences can be detected, and they are mostly only minor. It is quite surprising that there is so little variation between the four oldest age groups, and that the two spellings are so equally represented in all of them. This is in contrast with the findings in the other comparisons so far, and also with the overall results in the whole of the corpus.

5.1.4 Comparing Authors by Gender

Gender representation in the written part of the BNC corpus is not equal. Of the 1568 texts that have been identified by the gender of the author, 58 per cent were written by male authors, 26 per cent by female authors, and 15 per cent by authors of both genders. On the other hand, in a vast number of cases the gender of the author remains unknown: in half of the texts in the BNC the gender of the author has not been specified (see Appendix 1). However, despite this limitation to the corpus data, and in spite of the gender bias within the gender specified texts, these three categories are much better represented than the age groups discussed in 5.1.3. A total of 920 texts are attributed to male authors, 414 to female authors and 234 to both genders, and it is therefore possible to make some conclusions based on the information gathered.

Table 16 below gives an overview of the findings as numbers of texts containing one or two of the spellings and also shows how the two spellings overlap within the texts. According to the data, there seems to be a noticeable difference between the three gender groups when looking at how the two spellings overlap in the text samples. Those written by male authors contain more instances of overlap, both at word level and in general. In the case of male authors, when all the ten words are combined in one search, the overlap percentage is as high as 27 when in the corpus overall it is somewhere closer to 20 per cent, as explained in 5.1.1.

Gender	Male					Female					Mixed				
	Texts containing			Overlap		Texts containing			Overlap		Texts containing			Overlap	
	ise	ize	either	Texts	%	ise	ize	either	Texts	%	ise	ize	either	Texts	%
realise	413	409	740	82	11 %	236	154	363	27	7 %	159	24	179	4	2 %
recognise	454	453	822	85	10 %	223	161	366	18	5 %	148	28	167	9	5 %
organise	330	356	631	55	9 %	151	122	260	13	5 %	148	27	166	9	5 %
emphasise	270	284	525	29	6 %	125	76	192	9	5 %	94	15	108	1	1 %
criticise	198	217	378	37	10 %	85	78	160	3	2 %	96	14	109	1	1 %
characterise	172	200	347	25	7 %	51	39	89	1	1 %	49	7	56	0	0 %
specialise	186	197	372	11	3 %	59	60	113	6	5 %	107	14	118	3	3 %
summarise	151	151	295	7	2 %	33	24	56	1	2 %	40	9	49	0	0 %
apologise	109	90	195	4	2 %	100	60	159	1	1 %	53	3	56	0	0 %
minimise	125	130	243	12	5 %	37	28	63	2	3 %	47	8	55	0	0 %
all combined	595	534	886	243	27 %	267	195	398	64	16 %	193	41	209	25	12 %

Table 16 Usage according to the gender of the author as texts with matches in the BNC.

Also at word level the overlap of spellings in texts written by male authors is quite high. As Table 16 above shows, the percentages range between 2 and 11, depending on the word. Three words stand out: *realise* is spelled in both ways in 11 per cent of the texts where matches were found, *recognise* and *criticise* in 10 per cent, and *organise* in 9. The results for female and mixed gender authors are more moderate and more in line with the results in previous chapters, the highest percentage for a word being 7. Does this mean that male authors are less concerned about consistency in spelling? It is possible, but it would require a thorough investigation of the text samples to find the reason behind this anomaly in the findings. A high frequency of quotations from external sources could also explain the phenomenon.

Interesting differences between the three gender groups can also be found when comparing the use of the suffixes in percentages. It appears that male authors do not seem to favour either spelling, either at word level or in general. The percentages are quite close to 50 per cent for both spellings in all cases, even when overlap in the spellings is taken into consideration. This can be seen in Table 17 below.

Gender	Male				Female				Mixed			
	Texts containing				Texts containing				Texts containing			
	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %	%	%	true %	true %
realise	50 %	50 %	56 %	55 %	61 %	39 %	65 %	42 %	87 %	13 %	89 %	13 %
recognise	50 %	50 %	55 %	55 %	58 %	42 %	61 %	44 %	84 %	16 %	89 %	17 %
organise	48 %	52 %	52 %	56 %	55 %	45 %	58 %	47 %	85 %	15 %	89 %	16 %
emphasise	49 %	51 %	51 %	54 %	62 %	38 %	65 %	40 %	86 %	14 %	87 %	14 %
criticise	48 %	52 %	52 %	57 %	52 %	48 %	53 %	49 %	87 %	13 %	88 %	13 %
characterise	46 %	54 %	50 %	58 %	57 %	43 %	57 %	44 %	88 %	13 %	88 %	13 %
specialise	49 %	51 %	50 %	53 %	50 %	50 %	52 %	53 %	88 %	12 %	91 %	12 %
summarise	50 %	50 %	51 %	51 %	58 %	42 %	59 %	43 %	82 %	18 %	82 %	18 %
apologise	55 %	45 %	56 %	46 %	63 %	38 %	63 %	38 %	95 %	5 %	95 %	5 %
minimise	49 %	51 %	51 %	53 %	57 %	43 %	59 %	44 %	85 %	15 %	85 %	15 %
all combined	53 %	47 %	67 %	60 %	58 %	42 %	67 %	49 %	82 %	18 %	92 %	20 %

Table 17 Usage according to the gender of the author in the BNC, presented in percentages.

On the other hand, according to the numbers above, female authors are slightly more likely to choose *-ise* than *-ize*. The difference between male and female authors is not a dramatic one, but the percentages do differ to some extent. The third group, however, is strikingly different from the other two. Depending on whether one looks at the basic or the true percentages, *-ise* is used in around 80 to 90 per cent of the cases.

One explanation to the differences between the gender groups could be that the texts written by each group represent different types of mediums and domains. As will be explained in sections 5.1.5, 5.1.6 and 5.1.7, the frequencies of the two spellings appear to differ greatly depending on the medium, text type and domain of the text. Perhaps the texts that have been credited to mixed gender authors are mostly from sources where *-ise* is the preferred style. Indeed, when taking a closer look at some of the different categories where samples credited to mixed gender authors are more numerous than those written by male or female authors, this explanation suddenly becomes very plausible. It appears that out of the 209 texts in total that are attributed to mixed gender authors and that contain either spelling of the words studied, 78 per cent are from periodicals, and 85 per cent of those used *-ise*. According to the findings in section 5.1.5, periodicals contain more *-ise* spellings than any other medium in the corpus: in a combined search

of all the ten verbs, an overwhelming 98 per cent of texts containing either of the spellings had occurrences of *-ise*.

Rather than the medium of text determining the style of writing, which is often the case in the form of house-styles or other predetermined rules, it may well be that the gender of the author can sometimes influence the style within a medium. When taking a closer look at male authors in the corpus, it seems that the near 1:1 ratios in usage as seen in Table 17 is, in fact, true for male authors even in environments where *-ise* is usually the norm. For example, when looking at different domains of text, in the field of leisure where, according to the findings in 5.1.7, the *-ise* spelling overrides *-ize* very noticeably (80 per cent versus 20 per cent), male authors are more conservative and do not favour one spelling by much: a moderate 54 per cent of the texts use *-ise*. Also in periodicals, which contains a very high number of *-ise* spellings overall, as will be explained in section 5.1.5, the results for male authors within this medium are not higher than 59 per cent in favour of the *-ise* spelling, when overall in this category *-ise* is used in over 80 per cent of the cases. This seems to suggest that male authors, regardless of the context, keep closer to the 1:1 ratio in the two spellings than the other two gender groups.

A more detailed study of the gender groups within the corpus might reveal more about these behavioural patterns, but therein lies the problem: since in the vast majority of the texts in the corpus the gender of the author is unknown, the number of texts credited by gender is very small, and even more so when breaking the results into smaller portions by medium or domain of the texts. It is also difficult to know for certain if and how the spelling choices made by male and female authors can have an influence in the results of the different domains and text types. Without knowing the genders of the uncategorised authors in the corpus it is impossible to make further conclusions.

5.1.5 Comparing Mediums

From the different types of text mediums included in the written English part of the BNC, *books* make up the largest part with 1411 texts, which is 45 per cent of the data in the corpus. 38 per cent

of the samples are from *periodicals* (see Appendix 1). The remaining three groups, *miscellaneous published*, *miscellaneous unpublished* and *to-be-spoken* are much smaller in size, but still contain enough texts that some conclusions can be made based on the results found.

The numbers of texts in which the example words were found in the largest two groups are listed below in Table 18, which gives the data for *books* and *periodicals*. Overlap in spellings is more common in *books* than in *periodicals*, but nothing out of the ordinary compared to the overall results in the corpus, as listed in section 5.1.1. However, as a medium, *books* would most likely contain highly edited content and should therefore contain fewer spelling mistakes than the other groups. In this respect the percentages of overlap for this medium are relatively high.

Medium	Books					Periodicals				
	Texts containing			Overlap		Texts containing			Overlap	
	ise	ize	either	texts	%	ise	ize	either	texts	%
realise	687	650	1215	122	10 %	666	77	713	30	4 %
recognise	716	680	1288	108	8 %	706	117	771	52	7 %
organise	514	544	984	74	8 %	630	94	687	37	5 %
emphasise	435	420	812	43	5 %	438	89	483	44	9 %
criticise	301	322	600	23	4 %	409	98	458	49	11 %
characterise	264	294	524	34	6 %	212	67	272	7	3 %
specialise	282	300	562	20	4 %	400	51	439	12	3 %
summarise	231	223	445	9	2 %	162	32	188	6	3 %
apologise	209	162	366	5	1 %	198	26	217	7	3 %
minimise	174	187	354	7	2 %	226	58	270	14	5 %
all combined	883	793	1365	311	23 %	1050	199	1072	177	17 %

Table 18 Texts with matches in different text mediums in the BNC, part 1/2.

As can be seen in Table 19 below, *miscellaneous published* and *unpublished* texts, though fewer in number than books and periodicals, are fairly well represented as well, and although *to-be-spoken* is the smallest of the categories, the results found differ from the other four so much that it is not excluded from the comparisons at this stage.

Medium	Miscellaneous published					Miscellaneous unpublished					To-be-spoken				
	Texts containing			Overlap		Texts containing			Overlap		Texts containing			Overlap	
	ise	ize	either	texts	%	ise	ize	either	texts	%	ise	ize	either	texts	%
realise	90	8	92	6	7 %	125	17	132	10	8 %	30	0	30	0	0 %
recognise	102	14	108	8	7 %	151	24	160	15	9 %	32	3	32	3	9 %
organise	107	19	114	12	11 %	139	30	152	17	11 %	28	2	28	2	7 %
emphasise	49	9	53	5	9 %	70	11	75	6	8 %	8	0	8	0	0 %
criticise	20	3	23	0	0 %	21	3	22	2	9 %	28	0	28	0	0 %
characterise	34	8	37	5	14 %	20	3	21	2	10 %	0	0	0	0	0 %
specialise	66	15	73	8	11 %	83	8	88	3	3 %	18	3	19	2	11 %
summarise	26	4	28	2	7 %	47	4	49	2	4 %	3	0	3	0	0 %
apologise	15	0	15	0	0 %	24	2	26	0	0 %	15	0	15	0	0 %
minimise	37	5	41	1	2 %	49	5	50	4	8 %	6	0	6	0	0 %
all combined	168	37	173	32	18 %	216	55	224	47	21 %	34	7	34	7	21 %

Table 19 Texts with matches in different text mediums in the BNC, part 2/2.

The overlap patterns seem to vary slightly from medium to medium and from word to word. Whereas overall, the most overlap occurs in *books* (23 per cent) and the least in *periodicals* (17 per cent), differences can be found at word level. *Miscellaneous published* seems to contain quite a lot of overlap within one word, and so does *miscellaneous unpublished*. In this respect it does not seem to matter whether the texts are published and therefore possibly edited or whether they are not meant to be published at all. However, since *miscellaneous unpublished* as well as *to-be-spoken* contain texts not meant to be published, it would be unlikely that they contain large quantities of quoted material from other sources, thus including two contrasting spellings. At least some of this overlap in the spellings, then, especially when looking at individual words, is likely to be the result of misspellings or stylistic inconsistency.

Comparing the findings of the word queries in percentages reveals big differences between the text mediums, as exemplified by Tables 20 and 21. The contrast between the largest group, *books*, with *periodicals* and *miscellaneous published* is particularly striking. In *books* *-ise* and *-ize* are used almost equally often in all cases. Only *apologise* challenges this pattern a little: 56 per cent of texts are in favour of the *-ise* spelling. *Periodicals* and *miscellaneous published*, on the other hand, are decidedly in favour of the *-ise* spelling: results of the combined word searches show that

-ise is the preferred choice in 84 per cent of the texts in the former category and in 82 per cent of the texts in the latter. Even more striking numbers can be found in the highlighted columns. When cases of overlap within the texts are taken into consideration, *-ise* is even more frequent: within *periodicals* it can be found in 98 per cent of all the texts with word matches and *-ize* in 19 per cent. In *miscellaneous published*, *-ise* is found in 97 per cent of all texts with matches and *-ize* in 21 per cent.

Medium	Book				Periodical				Miscellaneous published			
	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %	%	%	true %	true %
realise	51 %	49 %	57 %	53 %	90 %	10 %	93 %	11 %	92 %	8 %	98 %	9 %
recognise	51 %	49 %	56 %	53 %	86 %	14 %	92 %	15 %	88 %	12 %	94 %	13 %
organise	49 %	51 %	52 %	55 %	87 %	13 %	92 %	14 %	85 %	15 %	94 %	17 %
emphasise	51 %	49 %	54 %	52 %	83 %	17 %	91 %	18 %	84 %	16 %	92 %	17 %
criticise	48 %	52 %	50 %	54 %	81 %	19 %	89 %	21 %	87 %	13 %	87 %	13 %
characterise	47 %	53 %	50 %	56 %	76 %	24 %	78 %	25 %	81 %	19 %	92 %	22 %
specialise	48 %	52 %	50 %	53 %	89 %	11 %	91 %	12 %	81 %	19 %	90 %	21 %
summarise	51 %	49 %	52 %	50 %	84 %	16 %	86 %	17 %	87 %	13 %	93 %	14 %
apologise	56 %	44 %	57 %	44 %	88 %	12 %	91 %	12 %	100 %	0 %	100 %	0 %
minimise	48 %	52 %	49 %	53 %	80 %	20 %	84 %	21 %	88 %	12 %	90 %	12 %
all combined	53 %	47 %	65 %	58 %	84 %	16 %	98 %	19 %	82 %	18 %	97 %	21 %

Table 20 Comparing mediums in the BNC, part 1/2.

When looking at the results word for word, some variation can be seen between them and between the three mediums. However, surprisingly, *characterise* is the word with the highest portion of *-ize* within all three mediums, although in the case of *periodicals* and *miscellaneous published* the percentages are still very low at 24 and 19 per cent, respectively. On the whole, the percentages of the individual words are in line with the results of the combined word search.

What these three mediums, *books*, *periodicals* and *published written material*, have in common is that the texts were all meant to be published, and it seems that there are, indeed, great differences between them in style. Since the results between *books* and the other two categories are so considerable, it is very likely that at least some of the authors have followed some rules or customs regarding the spellings, and thus influenced the results. As discussed in Chapter 4, printers, publishers, newspapers and the like who publish printed material often have preferences or rules

regarding the use of *-ise* and *-ize*. Even though the choice is often left to the author, many have strong preferences towards one or the other style and these may be enforce on the authors.

Since *periodicals* are one of the main sources of texts in the corpus, the spelling choices made within have a considerable influence in the whole of the corpus. However, as *books* form the largest group of sources, and therein the results are more even between the two spellings, the numbers provided by *periodicals* cannot completely overwhelm the overall results.

Although the remaining two categories, *miscellaneous unpublished* and *to-be-spoken*, are much smaller in size than the largest two, they provide an interesting angle into the question of usage. The texts in these two categories were not meant to be published, which makes them more indicative of personal spelling choices made by their authors. While *miscellaneous unpublished* consists of, among other things, letters, memos and essays, *to-be-spoken* is comprised of, for example, television news scripts and church sermons.

Medium	Miscellaneous unpublished				To-be-spoken			
	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %
realise	88 %	12 %	95 %	13 %	100 %	0 %	95 %	13 %
recognise	86 %	14 %	94 %	15 %	91 %	9 %	94 %	15 %
organise	82 %	18 %	91 %	20 %	93 %	7 %	91 %	20 %
emphasise	86 %	14 %	93 %	15 %	100 %	0 %	93 %	15 %
criticise	88 %	13 %	95 %	14 %	100 %	0 %	95 %	14 %
characterise	87 %	13 %	95 %	14 %	0 %	0 %	95 %	14 %
specialise	91 %	9 %	94 %	9 %	86 %	14 %	94 %	9 %
summarise	92 %	8 %	96 %	8 %	100 %	0 %	96 %	8 %
apologise	92 %	8 %	92 %	8 %	100 %	0 %	92 %	8 %
minimise	91 %	9 %	98 %	10 %	100 %	0 %	98 %	10 %
all combined	80 %	20 %	96 %	25 %	83 %	17 %	96 %	25 %

Table 21 Comparing mediums in the BNC, part 2/2.

The percentages in Table 21 above show that the results for *to-be-spoken* are almost exclusively in favour of the *-ise* spelling. However, as was seen in Table 19, this is the smallest of the five categories with just 34 texts with word matches. There are only 35 texts in total in this category in the BNC, and on closer inspection, all the texts containing *-ize* appear to be from the

same news channel. If we then disregard the results of *to-be-spoken* for lack of sufficient source material, what remains is *miscellaneous unpublished*.

As seen in Table 21, in *miscellaneous unpublished* the percentages are strongly in favour of one spelling: 80 per cent of *-ise*, 20 per cent of *-ize* in the combined word search. The true percentages, as the highlighted columns show, are even higher: 96 per cent of all texts containing either of the spellings of the example words contained *-ise*, compared to the 25 per cent of *-ize*. Since *miscellaneous unpublished* consists of texts that most likely represent the free choice of the author, as opposed to *books* and *periodicals*, at least, because they are the ones that are less likely to have been influenced by external parties and their spelling conventions, it could be argued that this medium is the one that is closest to the average British English speaker. However, even if this were true, the voices of these authors are barely visible in the scope of the whole corpus. Since the category of *miscellaneous unpublished* is underrepresented in the corpus compared to the largest two, *books* and *periodicals*, which together make up 83 of all the texts in the corpus, the choices made by these authors are fairly insignificant within the corpus data. Whatever the results in this category are, they make such a small contribution to the whole that the results within *books* and *periodicals* outweigh them.

5.1.6 Comparing Text Types

All of the texts in the written part of the corpus have been divided into six categories based on their text type. Of the texts that contained either of the two spellings discussed, *academic prose* contained 470 of them (16 per cent), *fiction and verse* 415 (14 per cent), *non-academic prose and biography* 726 texts (25 per cent), *newspapers* 391 texts (14 per cent), *other published written material* 645 texts (22 per cent) and *unpublished written material* 221 texts (8 per cent). Unlike with the different domains in the previous section, no one category seems to completely overwhelm the others in size, although *unpublished written material* is smaller than the other groups. The findings are listed in numbers of texts in tables 22 and 23 below, together with the cases of overlap of the spellings within the texts.

Text type	Academic prose					Fiction and verse					Non-academic prose and biography				
	Texts containing			Overlap		Texts containing			Overlap		Texts containing			Overlap	
	ise	ize	either	Texts	%	ise	ize	either	Texts	%	ise	ize	either	Texts	%
realise	186	174	325	35	11 %	224	195	398	21	5 %	384	270	587	67	11 %
recognise	265	210	425	50	12 %	206	191	393	4	1 %	417	295	632	80	13 %
organise	154	181	307	28	9 %	128	114	240	2	1 %	330	267	540	57	11 %
emphasise	198	186	359	25	7 %	97	63	160	0	0 %	274	207	430	51	12 %
criticise	132	136	254	14	6 %	59	38	97	0	0 %	188	182	343	27	8 %
characterise	150	167	295	22	7 %	15	15	30	0	0 %	192	161	331	22	7 %
specialise	104	110	206	8	4 %	36	46	82	0	0 %	229	152	364	17	5 %
summarise	165	123	279	9	3 %	13	8	21	0	0 %	145	109	247	7	3 %
apologise	20	12	31	1	3 %	132	107	237	2	1 %	78	55	126	7	6 %
minimise	102	94	190	6	3 %	14	9	23	0	0 %	132	106	233	5	2 %
all combined	346	254	470	130	28 %	236	215	415	36	9 %	572	363	726	209	29 %

Table 22 Comparing text types in numbers in the BNC, part 1/2.

The highest instances of overlap in the spellings can be found in *academic prose* and *non-academic prose and biography*, 28 and 29 per cent, respectively, when all the ten words are taken into account. At least in the case of *academic prose* this is not a surprising result, as quotations from external sources are to be expected in these types of texts. *Other published written material* and *unpublished written material* are next with 22 and 23 per cent, respectively.

Text type	Newspapers					Other published written material					Unpublished written material				
	Texts containing			Overlap		Texts containing			Overlap		Texts containing			Overlap	
	ise	ize	either	Texts	%	ise	ize	either	Texts	%	ise	ize	either	Texts	%
realise	258	2	259	1	0 %	413	97	475	35	7 %	133	14	138	9	7 %
recognise	232	5	235	2	1 %	432	112	513	31	6 %	155	25	161	19	12 %
organise	237	1	238	0	0 %	422	96	483	35	7 %	147	30	157	20	13 %
emphasise	129	0	129	0	0 %	235	60	279	16	6 %	67	13	74	6	8 %
criticise	184	5	185	4	2 %	171	62	206	27	13 %	45	3	46	2	4 %
characterise	50	0	50	0	0 %	106	25	129	2	2 %	17	4	19	2	11 %
specialise	125	0	125	0	0 %	267	61	313	15	5 %	88	8	91	5	5 %
summarise	29	0	29	0	0 %	81	18	98	1	1 %	36	5	39	2	5 %
apologise	91	0	91	0	0 %	103	15	116	2	2 %	37	1	38	0	0 %
minimise	45	0	45	0	0 %	153	42	183	12	7 %	46	4	47	3	6 %
all combined	391	10	391	10	3 %	592	192	645	139	22 %	214	57	221	50	23 %

Table 23 Comparing text types in numbers in the BNC, part 2/2.

The lowest result for overlap in the spellings is found in *newspapers*, where just 10 texts out of 391 with word matches (3 per cent) contained both spellings of any of the words studied.

This is perhaps not because of less spelling mistakes or other errors within the newspaper material, but due to the fact that the *-ize* spelling is practically non-existent in this category: only 10 texts out of the 391 containing any of the words studied had occurrences of *-ize*. In fact, this is the lowest rate of overlap in all of the different categories studied in Chapter 5. *Fiction and verse* is another category where overlap of the two spellings is noticeably rare, just 9 per cent of all texts containing any of the ten example words. At word level the occurrences are even harder to find. Here, the reason for this low number of occurrences cannot be explained by the exclusive use of either form, since they are almost equally distributed: *-ise* was found in 236 texts and *-ize* in 215. The explanation must lie elsewhere, for example in publishers' interest for uniformity in style.

A comparison of the results of the word queries, when presented in percentages, reveals great differences between the six text types. Tables 24 and 25 below show that while *academic prose* and *fiction and verse* both have a relatively high number of occurrences of *-ize* compared to the other categories (42 and 48 per cent, respectively), the opposite is true for the other groups.

Text type	Academic prose				Fiction and verse				Non-academic prose and biography			
	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %	%	%	true %	true %
realise	52 %	48 %	57 %	54 %	53 %	47 %	56 %	49 %	59 %	41 %	65 %	46 %
recognise	56 %	44 %	62 %	49 %	52 %	48 %	52 %	49 %	59 %	41 %	66 %	47 %
organise	46 %	54 %	50 %	59 %	53 %	47 %	53 %	48 %	55 %	45 %	61 %	49 %
emphasise	52 %	48 %	55 %	52 %	61 %	39 %	61 %	39 %	57 %	43 %	64 %	48 %
criticise	49 %	51 %	52 %	54 %	61 %	39 %	61 %	39 %	51 %	49 %	55 %	53 %
characterise	47 %	53 %	51 %	57 %	50 %	50 %	50 %	50 %	54 %	46 %	58 %	49 %
specialise	49 %	51 %	50 %	53 %	44 %	56 %	44 %	56 %	60 %	40 %	63 %	42 %
summarise	57 %	43 %	59 %	44 %	62 %	38 %	62 %	38 %	57 %	43 %	59 %	44 %
apologise	63 %	38 %	65 %	39 %	55 %	45 %	56 %	45 %	59 %	41 %	62 %	44 %
minimise	52 %	48 %	54 %	49 %	61 %	39 %	61 %	39 %	55 %	45 %	57 %	45 %
all combined	58 %	42 %	74 %	54 %	52 %	48 %	57 %	52 %	61 %	39 %	79 %	50 %

Table 24 Distribution of *-ise* and *-ize* by text type in the BNC presented in percentages, part 1/2.

However, even with academic prose the power structure changes when looking at the true percentages. In all the texts where the words and their competing spellings were found, *-ise* appeared in 74 per cent and *-ize* in 54 per cent of these texts. With *fiction and verse* the percentages corrected for overlap in the spellings are more moderate: 57 per cent of the texts contained *-ise* and

52 per cent *-ize*, so even here they are almost equally used. It should be noted that the category of *fiction and verse* is entirely made up of books, and as it was established in the section 5.1.5, in books the two spellings are very equally distributed. *Fiction and verse* is not the only text type found within that medium, though, and it represents a third of all the texts categorised as books in the BNC.

The groups where *-ise* is more clearly the chosen style are *newspapers*, *other published written material* and *unpublished written material*, as seen in the percentages listed in Table 25 below. As already seen in Table 23, there are barely any instances of *-ise* in the newspaper texts in the corpus. The results confirm what was established in the discussion in section 4.6, which is that British newspapers today use *-ise* almost exclusively.

The remaining two categories, *other published written material* and *unpublished written material*, are quite alike in terms of findings when compared to each other. Table 25 below shows that the *-ise* spelling is the more popular one in both groups, and it can be found in 76 and 79 per cent of the cases. When overlap is taken into consideration, the numbers in favour of *-ise* are 92 and 97 per cent, respectively.

Text type	Newspapers				Other published written material				Unpublished written material			
	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %	%	%	true %	true %
realise	99 %	1 %	100 %	1 %	81 %	19 %	87 %	20 %	90 %	10 %	96 %	10 %
recognise	98 %	2 %	99 %	2 %	79 %	21 %	84 %	22 %	86 %	14 %	96 %	16 %
organise	100 %	0 %	100 %	0 %	81 %	19 %	87 %	20 %	83 %	17 %	94 %	19 %
emphasise	100 %	0 %	100 %	0 %	80 %	20 %	84 %	22 %	84 %	16 %	91 %	18 %
criticise	97 %	3 %	99 %	3 %	73 %	27 %	83 %	30 %	94 %	6 %	98 %	7 %
characterise	100 %	0 %	100 %	0 %	81 %	19 %	82 %	19 %	81 %	19 %	89 %	21 %
specialise	100 %	0 %	100 %	0 %	81 %	19 %	85 %	19 %	92 %	8 %	97 %	9 %
summarise	100 %	0 %	100 %	0 %	82 %	18 %	83 %	18 %	88 %	12 %	92 %	13 %
apologise	100 %	0 %	100 %	0 %	87 %	13 %	89 %	13 %	97 %	3 %	97 %	3 %
minimise	100 %	0 %	100 %	0 %	78 %	22 %	84 %	23 %	92 %	8 %	98 %	9 %
all combined	98 %	2 %	100 %	3 %	76 %	24 %	92 %	30 %	79 %	21 %	97 %	26 %

Table 25 Distribution of *-ise* and *-ize* by text type in the BNC presented in percentages, part 2/2.

The results in the category of *unpublished written material* are of particular interest as they are texts that, out of these six groups, are most likely written without considering style guides or

other external influences. The results are therefore indicative of the choices of the general public, if not directly, then at least more than the results of the other groups.

To summarise the findings in this section, the text types that contained the lowest frequencies of the *-ise* spelling and the highest frequencies of *-ize* are *academic prose* and *fiction and verse*, where the two spellings are almost equally distributed. This is not surprising, since the medium of most of the texts within these two groups is *books*, where, as seen in the previous section, also contains almost equal numbers of both spellings. *Non-academic prose and biography* is slightly more likely to contain *-ise*, and within *other published written material* and *unpublished written material* *-ise* is even more popular. The s-spelling is most often found in *newspapers* where it is almost exclusively used.

5.1.7 Comparing Domains

In addition to comparing usage based on age, time, gender, text type and medium, the data in the BNC can also be categorised by text domain. From the nine different categories given, the ones that are best represented by volume in the written part of the BNC are *social science* (17 per cent with 526 texts), *world affairs* (15 per cent with 483 texts), *imaginative prose* (15 per cent with 476 texts), *leisure* (14 per cent with 437 texts) and *applied science* (12 per cent with 370 texts). The other categories are *commerce and finance* (9 per cent with 295 texts), *arts* (8 per cent with 261 texts), *natural and pure sciences* and *belief and thought* (both 5 per cent each with 146 texts) (see Appendix 2). The category of *imaginative prose* consists of texts which are fictional, literary or otherwise creative (Burnard 2000).

Out of the nine different domains, finding two competing spellings within one text seems to be most common in the fields of *social science* (28 per cent), *world affairs* (26 per cent), *arts* (23 per cent), *belief and thought* (23 per cent) and *applied sciences* (21 per cent), as listed in Table 26 below. At word level the percentages within some of these groups are also quite noticeable, sometimes as high as 21 per cent in the case of one word, as appears with *criticise* in *applied sciences*. As the percentage is so high, it is unlikely that all the cases of overlap could be explained

by misspellings alone. A large number of quotations from external sources or specific spelling conventions used within this group may explain some of them.

Domain	Social science		World affairs		Arts		Belief and thought		Applied science	
	Overlap		Overlap		Overlap		Overlap		Overlap	
	Texts	%	Texts	%	Texts	%	Texts	%	Texts	%
realise	39	11 %	33	9 %	23	12 %	12	12 %	12	6 %
recognise	56	13 %	51	13 %	14	8 %	11	10 %	14	6 %
organise	36	10 %	39	10 %	15	9 %	6	6 %	16	9 %
emphasise	30	9 %	39	15 %	6	5 %	3	5 %	6	5 %
criticise	10	4 %	31	11 %	6	6 %	0	0 %	21	21 %
characterise	23	11 %	9	5 %	6	6 %	2	4 %	0	0 %
specialise	14	6 %	3	2 %	5	5 %	0	0 %	8	5 %
summarise	8	3 %	0	0 %	0	0 %	1	3 %	1	1 %
apologise	2	5 %	5	4 %	1	2 %	0	0 %	0	0 %
minimise	6	4 %	3	3 %	1	2 %	0	0 %	11	10 %
all combined	133	28 %	120	26 %	55	23 %	30	23 %	69	21 %

Table 26 Overlap of competing spellings across texts in different domains in the BNC, part 1/2.

Within *social science* and *world affairs* it is not uncommon to find two spellings of the same word in one text, as the numbers above show. Quite a few of the words are spelled in two ways within one text, and the percentages of this overlap are particularly high among the four or five of the most common words, ranging between 9 and 15 per cent. The reason behind the exceptional numbers could be similar than in the case of *applied sciences*.

The next four groups, *commerce and finance*, *natural and pure sciences*, *leisure* and *imaginative prose* contain considerably less overlap, as can be seen in Table 27 below. Even at word level instances of overlap are not as common as with the five other groups. These numbers are more in line with the results of the unrestricted word search in 5.1.1.

Domain	Commerce and finance		Natural and pure sciences		Leisure		Imaginative prose	
realise recognise organise emphasise criticise characterise specialise summarise apologise minimise	Overlap		Overlap		Overlap		Overlap	
	Texts	%	Texts	%	Texts	%	Texts	%
	10	5 %	3	4 %	14	5 %	22	5 %
	13	6 %	8	8 %	15	5 %	4	1 %
	13	7 %	1	1 %	14	5 %	2	1 %
	7	5 %	2	3 %	5	3 %	0	0 %
	4	4 %	0	0 %	2	2 %	0	0 %
	4	6 %	4	6 %	0	0 %	0	0 %
	9	6 %	0	0 %	6	3 %	0	0 %
	4	4 %	4	6 %	1	3 %	0	0 %
	1	3 %	0	0 %	1	1 %	2	1 %
	1	1 %	2	3 %	2	2 %	0	0 %
all combined	49	18 %	22	17 %	58	15 %	38	9 %

Table 27 Overlap of competing spellings across texts in different domains in the BNC, part 2/2.

The frequencies of the two spellings do seem to vary to a great extent when comparing the text domain categories in the BNC with each other in detail. While some domains seem to follow the general pattern seen throughout this study, of roughly 60 to 70 per cent in favour of *-ise* against 40 to 30 per cent for the *-ize* spelling, there are a few that break this pattern in a very dramatic fashion. The results for *social science* and *world affairs*, as listed below in Table 28, appear to be rather ordinary, though some differences can be seen in usage between words and the overall results between these two categories.

Domain	Social science				World affairs				Imaginative prose			
realise recognise organise emphasise criticise characterise specialise summarise apologise minimise	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %	%	%	true %	true %
	68 %	32 %	75 %	36 %	62 %	38 %	68 %	41 %	55 %	45 %	58 %	48 %
	68 %	32 %	78 %	36 %	60 %	40 %	68 %	45 %	53 %	47 %	53 %	48 %
	63 %	37 %	70 %	41 %	60 %	40 %	66 %	44 %	54 %	46 %	54 %	47 %
	62 %	38 %	68 %	42 %	60 %	40 %	68 %	46 %	61 %	39 %	61 %	39 %
	59 %	41 %	61 %	43 %	58 %	42 %	65 %	46 %	61 %	39 %	61 %	39 %
	53 %	47 %	59 %	52 %	46 %	54 %	48 %	56 %	50 %	50 %	50 %	50 %
	63 %	37 %	66 %	40 %	57 %	43 %	58 %	44 %	45 %	55 %	45 %	55 %
	66 %	34 %	68 %	36 %	49 %	51 %	49 %	51 %	62 %	38 %	62 %	38 %
	82 %	18 %	86 %	19 %	75 %	25 %	78 %	26 %	55 %	45 %	56 %	45 %
	61 %	39 %	63 %	41 %	48 %	52 %	49 %	54 %	61 %	39 %	61 %	39 %
all combined	68 %	32 %	87 %	41 %	64 %	36 %	81 %	45 %	53 %	47 %	58 %	51 %

Table 28 Comparing domains in percentages in the BNC, part 1/3.

Imaginative prose, on the other hand, seems to be the one field where usage of the two spellings is nearing 50 per cent, both in general and at word level. Even when overlap in the spellings is taken into account and the percentages are calculated from the number of texts where either of the spellings were found, the results are fairly moderate: 58 per cent of the texts had occurrences of *-ise* and 51 of *-ize*. As 96 per cent of all texts in the corpus categorised as *imaginative prose* are found in books which, according to the findings in 5.1.5, is the medium where *-ise* and *-ize* were used almost equally, it is no surprise that the results of *imaginative prose* coincide with those found in books.

The results for *belief and thought* and *natural and pure sciences*, as listed in Table 29 below, are fairly similar with those of *social science* and *world affairs*, and seem to follow the same ‘universal’ pattern of the BNC data. In *arts* the percentages are slightly higher in favour of *-ise*, 70 per cent against the 30 of the *-ize* spelling.

Domain	Arts				Belief and thought				Natural and pure sciences			
	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %	%	%	true %	true %
realise	69 %	31 %	77 %	35 %	63 %	38 %	70 %	42 %	59 %	41 %	62 %	42 %
recognise	68 %	32 %	74 %	34 %	67 %	33 %	73 %	37 %	56 %	44 %	61 %	47 %
organise	64 %	36 %	70 %	40 %	67 %	33 %	71 %	35 %	58 %	42 %	59 %	43 %
emphasise	66 %	34 %	70 %	35 %	59 %	41 %	62 %	42 %	55 %	45 %	56 %	46 %
criticise	65 %	35 %	69 %	37 %	57 %	43 %	57 %	43 %	61 %	39 %	61 %	39 %
characterise	67 %	33 %	71 %	35 %	66 %	34 %	69 %	35 %	53 %	47 %	56 %	50 %
specialise	75 %	25 %	79 %	26 %	57 %	43 %	57 %	43 %	54 %	46 %	54 %	46 %
summarise	69 %	31 %	69 %	31 %	68 %	32 %	70 %	33 %	53 %	47 %	56 %	50 %
apologise	79 %	21 %	80 %	22 %	69 %	31 %	69 %	31 %	57 %	43 %	57 %	43 %
minimise	65 %	35 %	66 %	36 %	47 %	53 %	47 %	53 %	65 %	35 %	67 %	36 %
all combined	70 %	30 %	86 %	37 %	66 %	34 %	82 %	42 %	61 %	39 %	71 %	46 %

Table 29 Comparing domains in percentages in the BNC, part 2/3.

While the results of these two tables above follow a certain trend, apart from *imaginative prose*, the remaining three domains, *leisure*, *applied science* and *commerce and finance* make a very distinct grouping. Table 30 below shows that out of the nine categories, *leisure*, *applied science* and *commerce and finance* are the ones that use *-ise* the most.

Domain	Leisure				Applied science				Commerce and finance			
	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize	ise	ize
	%	%	true %	true %	%	%	true %	true %	%	%	true %	true %
realise	85 %	15 %	89 %	16 %	75 %	25 %	79 %	27 %	80 %	20 %	85 %	21 %
recognise	84 %	16 %	88 %	17 %	73 %	27 %	77 %	28 %	79 %	21 %	83 %	23 %
organise	88 %	12 %	92 %	13 %	73 %	27 %	80 %	29 %	78 %	22 %	84 %	24 %
emphasise	86 %	14 %	89 %	14 %	66 %	34 %	69 %	36 %	77 %	23 %	80 %	24 %
criticise	92 %	8 %	93 %	8 %	59 %	41 %	71 %	50 %	77 %	23 %	80 %	24 %
characterise	84 %	16 %	84 %	16 %	74 %	26 %	74 %	26 %	68 %	32 %	72 %	33 %
specialise	87 %	13 %	90 %	13 %	78 %	22 %	82 %	23 %	79 %	21 %	84 %	22 %
summarise	89 %	11 %	91 %	11 %	68 %	32 %	68 %	33 %	69 %	31 %	71 %	33 %
apologise	88 %	12 %	89 %	12 %	75 %	25 %	75 %	25 %	91 %	9 %	94 %	9 %
minimise	90 %	10 %	92 %	10 %	67 %	33 %	74 %	36 %	76 %	24 %	77 %	24 %
all combined	80 %	20 %	92 %	22 %	73 %	27 %	89 %	32 %	77 %	23 %	91 %	27 %

Table 30 Comparing domains in percentages in the BNC, part 3/3.

According to the percentages listed in Table 30, the *-ize* spelling is in a very distinct minority within these categories, compared to the other domains, especially so when looking at the combined results but in most cases also at word level. In *leisure* the percentages are the highest, 80 against 20 per cent in favour of the *-ise* spelling when the two spellings are compared directly with each other. When the percentages are calculated to include overlap of two spellings within one text, we find that the results go even higher for the *-ise* spelling. In the fields of *leisure*, *applied science* and *commerce and finance*, 89 to 92 per cent of texts containing either spelling include *-ise*. The *-ize* spelling, in contrast, is found in 22 to 32 percent of all texts with matches.

What makes these categories, particularly *leisure*, so different from the other domains? Why is one spelling so overwhelming in the findings? *Leisure* does seem like a field that would invite a more relaxed style and, perhaps, locally aimed material rather than text for an academic or an international audience. A search through the corpus reveals that *leisure* is mostly made of *newspapers* (20 per cent), *other published written material* (61 per cent) and *unpublished written material* (16 per cent). British newspapers, as has been shown, use *-ise* almost exclusively, but since the portion of newspaper texts within this domain is not higher than 20 per cent, they alone cannot explain the results.

Judging by the high frequencies of the *-ise* spelling in the texts samples for *commerce and finance*, the texts could hardly contain material intended for international markets. It would be unlikely that these texts would use region-specific spellings unless the texts were meant for British readers.

5.2 Beyond the BNC: The BAWE Corpus

Although the BAWE corpus is very different both in size and scope when compared to the BNC, and it represents only a very specific type of language and writers, it is nevertheless a very interesting source to compare with the BNC findings. The results of the word queries in BAWE are presented as word matches in the corpus, from which percentages were calculated.

5.2.1 Unrestricted Search

A search through the whole corpus shows that *-ise* is the more common spelling according to the texts samples in BAWE, the average percentages being 71 in favour of *-ise* and 29 in favour of *-ize*. However, some variation occurs, and not all the ten words behave in the same way. As can be seen in Table 31 below, the words that are more common in the corpus tend to have *-ise* spellings more often than those that are used less.

	Words			Percentages	
	ise	ize	total	ise %	ize %
realise	550	173	723	76 %	24 %
recognise	1024	406	1430	72 %	28 %
organise	330	160	490	67 %	33 %
emphasise	674	258	932	72 %	28 %
criticise	327	98	425	77 %	23 %
characterise	345	155	500	69 %	31 %
specialise	63	29	92	68 %	32 %
summarise	234	131	365	64 %	36 %
apologise	7	4	11	64 %	36 %
minimise	349	151	500	70 %	30 %
all	3903	1565	5468	71 %	29 %

Table 31 The distribution of *-ise* and *-ize* in BAWE.

All the words that appear in the corpus 490 times or less have slightly higher numbers of the *-ize* spelling, the results ranging between 33 and 36 per cents, with the exception of *criticise*, which is only found 425 times but has the highest number of *-ise* spellings of all the words at 77 per cent.

5.2.2 Comparing Disciplines

The source material in the corpus is divided into four disciplines: *arts and humanities*, *life sciences*, *physical science* and *social sciences*, and these are all fairly equally represented, the number of source texts contained in each group ranging between 640 and 802 (Heuboeck, Holmes and Nesi 2010, 5). *Arts and humanities* consists of assignments in archaeology, classics, comparative American studies, English, history, linguistics, philosophy. The category of *life sciences* consists of agriculture, biological sciences, food sciences, health, medicine and psychology. *Physical science* contains papers and assignments in architecture, chemistry, computer science, cybernetics and electronics, engineering, mathematics, meteorology, physics and planning. The last group, *social sciences*, is made of anthropology, business, economics, law, politics, publishing and sociology (Heuboeck, Holmes and Nesi 2010, 5-6).

The results of the word searches show that *-ise* is unarguably the more common spelling in all the four disciplines, but differences between the groups do exist, as exemplified by tables 32 and 33 below.

Discipline	Arts and humanities					Life sciences				
	Words			Percentages		Words			Percentages	
	ise	ize	total	ise %	ize %	ise	ize	total	ise %	ize %
realise	218	54	272	80 %	20 %	99	10	109	91 %	9 %
recognise	293	81	374	78 %	22 %	212	41	253	84 %	16 %
organise	96	38	134	72 %	28 %	81	19	100	81 %	19 %
emphasise	346	79	425	81 %	19 %	58	17	75	77 %	23 %
criticise	129	34	163	79 %	21 %	45	4	49	92 %	8 %
characterise	128	36	164	78 %	22 %	66	38	104	63 %	37 %
specialise	7	2	9	78 %	22 %	14	5	19	74 %	26 %
summarise	76	37	113	67 %	33 %	62	25	87	71 %	29 %
apologise	2	0	2	100 %	0 %	3	0	3	100 %	0 %
minimise	12	8	20	60 %	40 %	113	54	167	68 %	32 %
all	1307	369	1676	78 %	22 %	753	213	966	78 %	22 %

Table 32 Comparing disciplinary groups in BAWE, part 1/2.

The two groups where the instances of *-ise* are the highest are *arts and humanities* and *life sciences*. In both groups the overall frequency of *-ise* is 78 per cent. Variation is great between the words, more so in *life sciences* than in *arts and humanities*, but this can be partly explained by the smaller number of occurrences of words in the former category. Whereas arts and humanities had 1676 word hits, life sciences only had 966.

The results for the next two groups, *physical science* and *social sciences* differ somewhat from those of the other two categories. The combined results in Table 33 below reveal that in these two disciplines *-ise* is not as overwhelmingly popular, and the percentages are slightly lower at 66 per cent for *physical science* and 65 per cent for *social sciences*. Again, differences between the words are at times great, but as with the other two disciplines, some of this can be explained by the low number of hits in the subcorpora.

Discipline	Physical science					Social sciences				
	Words			Percentages		Words			Percentages	
	ise	ize	total	ise %	ize %	ise	ize	total	ise %	ize %
realise	83	45	128	65 %	35 %	150	64	214	70 %	30 %
recognise	85	51	136	63 %	38 %	434	233	667	65 %	35 %
organise	26	18	44	59 %	41 %	127	85	212	60 %	40 %
emphasise	25	16	41	61 %	39 %	245	146	391	63 %	37 %
criticise	9	1	10	90 %	10 %	144	59	203	71 %	29 %
characterise	34	19	53	64 %	36 %	117	62	179	65 %	35 %
specialise	16	7	23	70 %	30 %	26	15	41	63 %	37 %
summarise	42	24	66	64 %	36 %	54	45	99	55 %	45 %
apologise	0	0	0	0 %	0 %	2	4	6	33 %	67 %
minimise	130	50	180	72 %	28 %	94	39	133	71 %	29 %
all	450	231	681	66 %	34 %	1393	752	2145	65 %	35 %

Table 33 Comparing disciplinary groups in BAWE, part 2/2.

To summarise, the four disciplines in BAWE behave differently, and two competing pairs are formed: whereas in *arts and humanities* and *life sciences* the *-ise* spelling is noticeably more popular at 78 per cent, the latter two groups, *physical science* and *social sciences* show lower results at 66 and 65 per cent. Both groups differ from the overall results of Table 31 in 5.2.1 where the frequency of *-ise* was 71 per cent.

It seems that, according to BAWE, the *-ize* spelling is more used in *physical science* and *social sciences* than in the other two groups. Some of the differences between the groups may be explained by style preferences within these disciplines, but perhaps the fields within these groups are in favour of a more international style of writing than the other two groups. As listed earlier, *physical science* and *social sciences* contain texts from, for example, electronics, mathematics, engineering, business, law and politics. Perhaps these fields are slightly less likely to favour a decidedly British spelling than the ones in the other two groups.

5.2.3 Comparing Authors by Gender

When comparing texts written by male and female authors, the results are very similar for both genders. As Table 34 below shows, male writers chose *-ise* in 71 per cent of the cases, whereas the same result for female writers is 72 per cent.

Gender	Male					Female				
	Words			Percentages		Words			Percentages	
	ise	ize	total	ise %	ize %	ise	ize	total	ise %	ize %
realise	226	72	298	76 %	24 %	324	101	425	76 %	24 %
recognise	338	119	457	74 %	26 %	686	287	973	71 %	29 %
organise	110	44	154	71 %	29 %	220	116	336	65 %	35 %
emphasise	224	104	328	68 %	32 %	450	154	604	75 %	25 %
criticise	95	39	134	71 %	29 %	232	59	291	80 %	20 %
characterise	127	68	195	65 %	35 %	218	87	305	71 %	29 %
specialise	30	13	43	70 %	30 %	33	16	49	67 %	33 %
summarise	113	59	172	66 %	34 %	121	72	193	63 %	37 %
apologise	2	0	2	100 %	0 %	5	4	9	56 %	44 %
minimise	165	71	236	70 %	30 %	184	80	264	70 %	30 %
all	1430	589	2019	71 %	29 %	2473	976	3449	72 %	28 %

Table 34 Comparing authors by gender in BAWE.

The only variation one can find is between different words, but as the instances of hits differ from word to word, some of the variation can be explained by the low number of hits as well as by the inaccuracy of the search method.

In order to see if the discipline of the texts has any effect in the results between male and female writers, the four disciplinary groups were compared based on the gender of the author. Since this closer inspection means that the word hits diminish drastically, and results for individual words

may be very small indeed, only the total number of each spelling of all the ten words combined is given.

By comparing the results within each disciplinary group, differences are beginning to form between male and female writers. Table 35 below shows that in *arts and humanities* as well as in *life sciences* female writers are more likely to use the *-ise* spelling than male writers. Interestingly, the results are similar in both groups: male writers use the *-ise* spelling in 71 or 72 per cent of the cases and female writers in 80 or 81 per cent of the cases.

Arts and humanities									
Male					Female				
Words			Percentages		Words			Percentages	
ise	ize	total	ise %	ize %	ise	ize	total	ise %	ize %
411	157	568	72 %	28 %	896	212	1108	81 %	19 %

Life sciences									
Male					Female				
Words			Percentages		Words			Percentages	
ise	ize	total	ise %	ize %	ise	ize	total	ise %	ize %
146	59	205	71 %	29 %	607	154	761	80 %	20 %

Table 35 Gender comparison by discipline in BAWE, part 1/2.

When looking at *physical science* and *social sciences*, however, the results are quite the opposite. As Table 36 exemplifies, although *-ise* is still the more popular spelling for both genders in both groups, female writers use *-ise* much less than male writers, the percentages dropping as low as 56 per cent in *physical science* and 62 in *social sciences*.

Physical science									
Male					Female				
Words			Percentages		Words			Percentages	
ise	ize	total	ise %	ize %	ise	ize	total	ise %	ize %
341	146	487	70 %	30 %	109	85	194	56 %	44 %

Social sciences									
Male					Female				
Words			Percentages		Words			Percentages	
ise	ize	total	ise %	ize %	ise	ize	total	ise %	ize %
532	227	759	70 %	30 %	861	525	1386	62 %	38 %

Table 36 Gender comparison by discipline in BAWE, part 2/2.

Male writers seem more consistent with their style, at least according to the findings in the comparisons above. The results for all four disciplinary groups are exactly or very close to 70 per cent in favour of *-ise*, whereas female writers seem to change their style from discipline to discipline. Could this mean that male writers are more prone to keep a spelling choice they have become accustomed to, and are female writers more open to changing their style according to style preferences or rules within a genre or discipline?

As it looks like female authors are better represented in the corpus than male authors, at least according to the word searches performed, the preferences of these two gender groups affects the overall results in the corpus. However, the results seem to cancel each other out, and while in some disciplines female writers are more in favour of *-ise* than in others, the overall results that include all four disciplines are still very similar for both genders.

5.3 Discussion of the Corpus Findings

The findings in the two corpora show that the matter of *-ise* versus *-ize* is quite complicated, and variation occurs between many different types of texts and writers. The results of the unrestricted word searches in both corpora gave similar results: in the BNC *-ise* was the more popular spelling with usage at 68 per cent, and as many as 82 per cent of all texts with matches contained *-ise*. In BAWE the overall results were 71 per cent in favour of *-ise*.

The diachronic comparison in the BNC revealed that changes seem to have happened in usage over time. The most dramatic difference was between the earliest time period (1960-1974) and the two later periods (1975-1984 and 1985-1993). Since the oldest time period contained only works of fiction and it is, in general, a much smaller subcorpus, it is not directly comparable with the other two periods. The results for the remaining two periods show some moderate changes: although *-ise* is the more popular spelling in both time frames, its frequency is slightly higher during the latter period. However, when the findings of all three time periods were organised based on text type, making the results more comparable, it was revealed that even within text types

changes in usage could be found over time. Within *fiction and verse* *-ize* used to be the most common choice by far between 1960 and 1974, but towards the end of the century its popularity had declined dramatically, and between 1985 and 1993 it was down to 45 per cent from the earlier 71 per cent.

When these results are compared with those found in the BAWE corpus, the overall trend seems to continue. Although BAWE is not directly comparable with the BNC because it is a specialised corpus containing student assignments, the results do indicate that *-ise* is the more common spelling of the two in British English also in the first decade of the twenty first century.

Age comparison in the BNC revealed very little, possibly because the youngest two groups were so poorly represented. The four groups that contained enough source material to make conclusions on differed very little from each other. Among authors of different ages the two spellings were used quite equally. It is a pity that the BNC is so limited and unbalanced when it comes to comparing authors of different ages. The lack of sufficient source texts for the youngest two age groups means that it is impossible to see developments in usage based on age. It would have been useful to see if the results of younger authors had differed from those of the other groups. Such information might have made it possible to predict usage trends beyond the scope of the BNC. However, what the spelling choices of young British English speakers at the time the corpus was compiled were will remain a mystery. It is entirely possible that they, too, would have followed the patterns of the four older age groups where the two spellings were almost equally represented in the corpus data.

Results found in the two corpora suggest that there is some variation to be found between how male and female authors use the two suffixes. At first glance the results look similar in both corpora: in the BNC data both genders are slightly more in favour of *-ise*, but the differences between the two genders do not seem dramatic. In the case of BAWE, there is no variation at all between male and female authors if looking at the results gathered from the whole corpus.

However, when the corpora are studied in more detail, it is revealed that there is, indeed, at least one very significant difference between male and female authors. It appears that, according to the corpus findings, female authors are more flexible in their spelling choices and change their style according to the surroundings. Male authors, on the other hand, seem less likely to adapt their spelling and are more prone to retain the style they have become accustomed to, even if in general the spelling convention within the text type, medium, domain or discipline is in contrast with that chosen style. Some of the variation in the findings can, perhaps, be explained by differences in the backgrounds of the authors, for example, from the point of view of education. Also, since only half of the texts in the BNC have been categorised based on the gender of the author, and of those that are, the majority were written by men, the results found in the corpus can only be indicative of gender differences. However, since these behavioural patterns were seen in both of the two corpora, it could be argued that female authors are, indeed, more willing to adjust their spellings in different environments, whereas male authors are more conservative in their choices.

Comparing different mediums in the BNC revealed that *books* and *periodicals*, which together make up most of the corpus data (45 and 38 per cent, respectively), are the decisive counterparts in the question of *-ise* versus *-ize* in the corpus findings: while within books the two spellings are used in equal proportion, in periodicals the *-ise* spelling is the chosen style in 84 per cent of the cases, and as high as in 98 per cent if overlap in the two spellings is taken into account. This battle between these two groups is, ultimately, what is seen in the results of the overall corpus, when all text types, domains and mediums are included. Were there more books in the corpus, the overall results would most likely be less in favour of either spelling. In contrast, if the corpus data leaned more towards *periodicals* or contained less books than it does, the *-ise* spelling would probably appear to be even more popular in the overall results.

The medium that most resembles texts written by an average person is *miscellaneous unpublished*, because this content is less likely to have been influenced by stylistic rules or house styles than the texts in *books* or *periodicals* because these texts were not meant to be printed or

published. Within this medium *-ise* is found in 96 per cent of all texts containing either of the spellings. If overlap in the spellings is disregarded, the percentages are 80 in favour of *-ise* and 20 in favour of *-ize*.

Comparing the text types in the BNC revealed similar patterns than the mediums: the highest proportions of *-ize* were found in *academic prose* and *fiction and verse*, which together make up around 50 per cent of the texts in the medium of *books* in the corpus. The results are therefore not surprising. *Newspapers* were the ones that contained most examples of *-ise*, and occurrences of *-ize* were nearly non-existent in them. The rest of the text types were somewhere between these two extremes. Within both *other published written material* and *unpublished written material* *-ise* is the more common spelling, ranging between 76 and 79 per cent in frequency. The latter group is interesting from the point of view of unedited text: as with *miscellaneous unpublished* in the medium comparison, of all the text type categories in the BNC, *unpublished written material* is closest to representing the text of an average person, someone for whom writing is not a profession, at least not in the same way as literary authors and journalists, for example.

The final comparisons were done between text domains (or disciplines, as was the case in BAWE), and this is where the results of the two corpora are most comparable with each other, as some of the categories are similar, although not exactly the same nor do they contain exactly the same kind of source texts.

What can be seen in the BNC data is that *-ize* is found most often within *imaginative prose*, as there the two spellings are used almost in equal portions. The domain with the most occurrences of *-ise* is *leisure*, where the frequency of *-ise* is 80 per cent. Of all the texts categorised as *leisure* that contained either spelling of any of the words, 92 per cent contained examples of *-ise*.

The differences between the four disciplinary groups in BAWE are not as dramatic, but some variation occurs. The highest proportions of *-ise* were found in *arts and humanities* and *life sciences*, where 78 per cent of the words used the *-ise* spelling and 22 per cent the *-ize* spelling. The results were the same in both categories. In the other two categories, *physical science* and *social*

sciences, *-ise* is still the more common spelling, but not by as much: 66 per cent and 65 per cent of the matches were *-ise*, respectively.

Although the BNC is a much older corpus than BAWE, there are significant similarities in the findings when looking at the domains in the former and the disciplines in the latter. If one compares the category of *art* in the BNC with *arts and humanities* in BAWE, for example, it can be seen that while *art* in the BNC contains a slightly lower proportion of *-ise* than *arts and humanities* in BAWE (70 per cent versus 78 per cent, respectively), the results are not too dissimilar, especially when compared to the general results in each corpus: the proportion of *-ise* is higher in these categories than in each corpora overall. The category of *social sciences*, on the other hand, contains a similar portion of the *-ise* spelling in both corpora: 68 per cent in the BNC and 65 per cent in BAWE.

In summary, the biggest differences in usage in the BNC can be found between certain mediums (*books* versus *periodicals*), text types (*fiction and verse* versus *newspapers*) and domains (*imaginative prose* versus *leisure*). Some of these overlap: although *books* contain texts of various types, *fiction and verse* is a text type that can only be found in books. *Imaginative prose* is also something that is almost exclusively found in books in the corpus. The results of these overlapping types, *books*, *fiction and verse* and *imaginative prose* are the same: *-ise* and *-ize* are quite equal with each other in terms of frequency.

In the end, it seems that the greatest reasons for the overall results in the BNC are due to two large, very contrasting groups, *books* and *periodicals*, which together make 83 per cent of the whole of the corpus, and there seems to be a difference in adopted style within both of these. In *books* both spellings are used almost equally, whereas in *periodicals* the *-ise* spelling is the overwhelming winner: 98 of all texts in this category that had occurrences of either spelling contained *-ise*.

While the two corpora are very different in terms of the contents and the type of authors they represent, the results nevertheless indicate that the two spellings have long been in competition

with each other and continue to be so. Comparing the results in the BNC with those found in BAWE reveals that the power structure between the two spellings has not changed dramatically since the compilation of the BNC, and both forms are still widely used.

It could very well be that outside of the confinement of academic context the occurrences of *-ise* would be even higher in the twenty first century, but without a balanced corpus containing several different types of texts from the same period as BAWE it is impossible to make further approximations. While in the BNC *academic prose* contained higher percentages of *-ize* and lower percentages of *-ise* than all of the other text types apart from *books*, the category is hardly comparable with BAWE, as the BNC texts are mainly from published works: 68 per cent of the texts labelled as *academic prose* were found in books and a further 31 in periodicals. The texts that BAWE consists of may be academic, but of a different kind: the student essays and other assignments were not written in order to be published, and are therefore only subject to the style and spelling preferences of their universities or of the schools and the staff within.

The decision to compare the use of the two verbal endings in the BNC as instances in separate texts with matches instead of simple word counts in the database was done so that the results would better represent the spelling choices made by different authors. Although this may very well have been the better choice of the two, it did not eliminate problems in the corpus findings. As was seen in many of the comparisons, in some categories the corpus texts have been gathered from a very select group of samples, sometimes from just a few different sources. Also, since some groups of texts or authors in the corpus are better represented than others, the spelling choices made within those categories influence the results heavily.

6 Conclusion

The findings in the two corpora studied reveal that the suffixes *-ise* and *-ize* do not appear in the same quantities in different environments, but that there are significant differences to be found when comparing various types of texts and authors. Noticeable variation can be found, for example, when comparing written language by text type, domain or medium. The differences were greater within the BNC, but not because of the age of the corpus but mainly because the corpus is larger and it contains text samples from a much varied selection of sources than BAWE. The texts in BAWE were also analysed in a more restricted manner than the ones in the BNC, due to limitations in the search methods.

According to both corpora, *-ise* is the more popular spelling in British English overall, at least within the types of texts they represent, with around 70 per cent of the texts in favour of *-ise* and around 30 per cent in favour of *-ize*. However, in the BNC data it was possible to see that many texts contained not just one but both of the spellings. One word could be spelled with *-ise* and another with *-ize*, but it was also very common to find two spellings of the same word within one text. If this is taken into consideration, in the BNC *-ise* can be found in 82 per cent of all the texts with word matches, and *-ize* is found in 38 per cent of the texts.

The overlap in the spellings could be explained by a number of reasons. Some of it could be due to author error or negligence, or perhaps a specific spelling is often chosen for one word and another spelling for another word, either by choice, habit or mistake. The corpus material also revealed that some of the source texts contained quotations from external texts, which is another plausible explanation for overlapping spellings within one text.

Noticeable differences in usage can be found when comparing different text types, mediums and domains in the BNC, and this was also true for the four disciplines in the BAWE corpus. The most dramatic differences are within the BNC, and they are between certain mediums (*books* versus *periodicals*), text types (*fiction and verse* versus *newspapers*) and domains (*imaginative prose* versus *leisure*). As *fiction and verse* and *imaginative prose* all both only or

mostly found within books, it is no surprise that also the results of these three groupings are similar: the two spellings are used in almost 1:1 ratio in all of them. Within the contrasting groupings, *periodicals*, *newspapers* and *leisure*, on the other hand, the *-ise* spelling is much more common than the overall results in the BNC could indicate: in periodicals the *-ise* spelling was favoured in 84 per cent of the cases, in *leisure* the percentage is 80, and the highest ratio of *-ise* can be found in *newspapers*, where it is practically the only spelling used. This result is directly related to the findings in section 4.6, where it was concluded that British newspapers and other news media have long favoured the use of *-ise*.

The four different disciplinary groups in the BAWE corpus also revealed some variation between them, but the differences are much smaller than what was found in the BNC. *Arts and humanities* and *life sciences* contained the highest numbers of the *-ise* spelling (78 per cent), and *physical science* and *social sciences* contained the lowest numbers (66 and 65 per cent, respectively). The results found in the two corpora in this respect are not in contradiction with the findings in Chapter 4, where it was evident that the policies of publishers and printers are often in favour of the *-ize* spelling, at least more often than in the case of some other types of authorities like the public sector, for example.

The question regarding the usage of the spellings between authors of different ages and genders is an interesting one, and the BNC could answer that at least partly. No major differences could be found within the age groups within the BNC. However, because of an insufficient number of source texts found in the corpus for the youngest two age groups, the comparisons performed were not exhaustive. The two corpora revealed that male and female authors seem to approach the question of *-ise* versus *-ize* in very different ways. While female writers can be seen adapting their style according to the type of text in question, male writers tend to keep to their chosen style even when other texts within that medium, domain or text type are clearly in favour of the other spelling.

Without knowing the social status and educational background of each of the male and female authors studied, it is impossible to say whether some social factors could explain the

differences. However, since the findings in both of the corpora studied show similar patterns, which indicate that there are some great differences between male and female authors, the results found here could provide a starting point to further study.

According to the corpora, some change seems to have happened in usage over time. There are challenges both within the BNC and when comparing the two corpora with each other but, overall, some increase in the use of *-ise* can be seen in the BNC data. When compared with the results of the findings in BAWE, which was collected some 20 years after the completion of the BNC, the percentages of usage seem similar. This indicates that the use of *-ise* has not diminished in British English since the 1990s when the BNC was published. However, as the BAWE findings only represent university assignments, the results could be very different if more types of texts were studied from the same period. As was explained in Chapter 4, many influential public sector operators today favour *-ise*, and their influence, together with that of newspapers and other media could mean that the public might have become more accustomed to *-ise* than *-ize* since the completion of the BNC.

Since most of the data in the written part of the BNC consists of published texts, and therefore they have most likely been through the process of copy editing, the choices made by the authors do not necessarily represent their personal styles. A corpus of different leanings might have yielded very different results. The texts in both of the corpora studied contain a high number of texts which may have been written in concordance with specific, pre-existing rules and style regulations regarding the spellings. Neither corpus represents the voice of the average speaker of British English, but those of a select few. In the case of BAWE the texts are all university assignments, and so the authors represent only a small portion of their contemporaries. Even in the case of the BNC, which intends to give a cross-section of British English at the time of its collection, is mainly comprised of texts written by either those who write in a professional capacity (newspapers, fiction and verse, books), academics (academic prose, sciences) or other types of writers that do not necessarily represent the average person. However, the texts in the BNC are

often the types of texts that the average person will be exposed to, like newspaper articles, for example, and in that sense the results are relevant even from the point of view of the general public.

Despite the faults and limitations of the BNC, there are few contemporary British English corpora readily available that are as varied and balanced. In order to find information of more current trends regarding the use of the suffixes, newer sources of data would be needed, preferably ones with stronger leanings into unedited texts by members of the public and authors of different ages. Some vast corpora do exist that might reveal more about the use of the two spellings discussed, but only a select few have access to them. The Cambridge English Corpus, for example, is a multi-billion word corpus of British and American English that contains both written and spoken language and claims to be the largest of its kind in the world. Unfortunately, it is only available to authors and editors of Cambridge University Press as well as researchers and students of the University of Cambridge (Cambridge University Press 2015b).

The dictionaries, usage guides and style manuals that were consulted in Chapter 4 show that attitudes towards the two spellings and advice given on their usage have changed over time. Although some strong preferences can still be found, and style manuals do often prescribe one form or the other, in general, dictionaries, usage guides and style guides have become less prescriptive of late. However, have these authorities begun to consider *-ise* as a serious alternative because it is used so frequently, or has it become more frequent because the spelling is now ‘allowed’ by them? The answer is more likely to be closer to the former than the latter. For example, many modern dictionaries have been compiled with the help of linguistic corpora. “As the use of the ‘s’ caught on, dictionaries had to follow the trend, and are now quoting it as ‘correct’. It is arguable, but dictionaries do not necessarily promote correct usage, but follow prevailing practice (which is then taken as correct, creating a spiral of decline)” (Dale 2013).

It seems that, despite the efforts of some language authorities, for example *The Oxford English Dictionary* or certain publishers, the *-ize* form has not become to replace *-ise*, especially in non-academic contexts. Quite the contrary, it seems to have lost ground significantly between the

1960s and the 1990s. And, as was seen in the BAWE findings, between 2004 and 2007 *-ise* is still the more common spelling, even in an academic context, although the source texts are student assignments rather than published works.

On the Internet American spellings are more common than British ones, partly because the Internet originated in the U.S., and partly because American spellings are shorter and therefore more economic (Crystal 2001, 88). It would therefore be very possible that changes in British English, also in spelling, are happening as a result of the increasing exposure to American conventions. On the other hand, speakers and writers of British English might feel the need to ‘protect’ their national conventions and avoid everything they consider American influence. In addition, the *-ise* form seems to be the heavily preferred form in British society today, so it may very well survive even in an increasingly globalised world.

As was noted earlier, many speakers of British English falsely believe that *-ize* is an Americanism. This could be explained by the lack of exposure to the *-ize* variant in everyday situations: large quantities of the written material that the general public sees uses mostly *-ise*. According to the findings in Chapter 4, this is true for most newspapers and also the public sector, to a large extent. In addition, as some spell checkers and other such software seem to automatically ‘correct’ *-ize* to *-ise* when British English is chosen as the language (Ask Different 2011), it is easy for someone not familiar with etymology to assume that *-ize* is foreign influence. Even if usage guides are still somewhat popular, only a certain part of the population is interested in ‘correct’ usage and consults dictionaries or usage guides. Probably the majority go by the standards they have been exposed to in the educational system and in everyday life.

It is difficult to predict whether the *-ise* spelling, a decidedly British (and Australian) form although also used elsewhere, will survive American English influence and the international quest for uniformity. It might very well be that in this globalised world a simple orthographic feature may become an underlined statement, a means of expressing one’s origin or nationality. On the other hand, there are several other orthographical features that separate British English from American

English, and the issue of *-ise* versus *-ize* may very well be forgotten at some point. The position taken by the educational system and the public sector, the media, dictionaries, language guides and publishers will all have their consequences.

If spelling differences are usually separating British and American English from one another, it seems that the battle between *-ise* and *-ize* is, in fact, dividing Britain. “Is it not about time we established a uniform spelling for verbs with these endings?” (Wood 1962, 128), a cry that has been repeated for decades by spelling reformists, compilers of language guides, et cetera and, last but not least, most likely by countless members of the public. To no avail, it seems. According to the corpus findings, at present, as Burchfield (1996, 422) so eloquently phrases, “[t]he matter remains delicately balanced but unresolved”. Perhaps the peaceful co-existence of the two spellings will persist, as it has done so far.

7 References

Corpora:

British Academic Written English Corpus (BAWE)

The British National Corpus (BNC)

Dictionaries:

Cambridge Advanced Learner's Dictionary (2008) 3rd edn. Cambridge: Cambridge University Press.

Cambridge International Dictionary of English (1995) Cambridge: Cambridge University Press.

Collins Cobuild Advanced Learner's English Dictionary (2006) 4th edn. London: HarperCollins.

Collins Dictionary of the English Language (1986) 2nd edn. London: Collins.

Collins Cobuild English Dictionary (2011) 11th edn. Glasgow: HarperCollins.

Collins Cobuild English Dictionary (1995) London: HarperCollins.

The Concise Oxford Dictionary of Current English (1964) 5th edn. Oxford: Oxford University Press.

Longman Dictionary of Contemporary English (2003) 4th edn. London: Longman.

Longman Dictionary of the English Language (1995) London: Viking.

Longman Dictionary of the English Language (1984) Harlow: Longman.

Oxford Advanced Learner's Dictionary of Current English (1989) 4th edn. Oxford: Oxford University Press.

The Oxford English Dictionary (1989). 2nd edn. Oxford: Clarendon Press.

Oxford Dictionary of English (2003) 2nd edn. Oxford: Oxford University Press.

Webster's New Twentieth Century Dictionary of the English Language (1983). New York: Simon and Schuster.

Works Cited:

Aronson, J. (2001) "When I use a word: -ize right." *British Medical Journal* 323, p. 1173.

Aston, G. and Burnard, L. (1998) *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Amis, K. (1997) *The King's English: A Guide to Modern Usage*. London: HarperCollins.

Bailey, G. (2004) "Real and Apparent Time." In J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes (eds), *The Handbook of Language Variation and Change*. Paperback edn. Oxford: Blackwell, pp. 312-332.

Baker, P. (2010) *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Bauer, L. (2004) "Inferring Variation and Change from Public Corpora." In J. K. Chambers, Peter Trudgill and Natalie Schilling-Estes (eds), *The Handbook of Language Variation and Change*. Oxford: Blackwell, pp. 97-114.

Bex, T. (1999) "Representations of English in Twentieth-Century Britain: Fowler, Gowers and Partridge." In Tony Bex and Richard J. Watts (eds), *Standard English: The Widening Debate*. London: Routledge, pp. 89-108.

Burchfield, R. W. (ed.) (1996) *The New Fowler's Modern English Dictionary*. Oxford: Oxford University Press.

Burt, A. (2002) *The A to Z of Correct English*. 2nd edn. Oxford: How to Books.

Cameron, D. (1995) *Verbal Hygiene*. London: Routledge.

Carney, E. (1994) *A Survey of English Spelling*. London: Routledge.

- Copperud, R. H. (1964) *A Dictionary of Usage and Style*. New York: Hawthorn Books.
- Crystal, D. (2013) "Did Johnson Fix English Spelling?" *The New Rambler*, 2012-13, pp. 21-23.
- Crystal, D. (2001) *Language and the Internet*. Cambridge: Cambridge University Press.
- Crystal, D. (1981) "Language on the air – has it degenerated?" *The Listener*, 9 July 1981, pp. 37-39.
- Davis, H. (1999) "Typography, Lexicography and 'Standard English.'" In Tony Bex and Richard J. Watts (eds), *Standard English: The Widening Debate*. London: Routledge, pp. 69-88.
- Delbridge, A. (2001) "Lexicography and National Identity: The Australian Experience." In David Blair (ed.) *English in Australia*. Philadelphia: John Benjamins, pp. 303-316.
- Firnberg, L. B. and Firnberg, D. (1976) *Cassell's New Spelling Dictionary*. London: Cassell.
- Foster, B. (1968) *The Changing English Language*. London: Macmillan.
- Fowler, H. W. (1926) *A Dictionary of Modern English Usage*. Oxford: Clarendon Press.
- Fowler, H. W. (1965) *A Dictionary of Modern English Usage*. 2nd edn, revised by Ernest Gowers. Oxford: Oxford University Press.
- Fritz, C. (2010) "A Short History of Australian Spelling." *Australian Journal of Linguistics*, Vol. 30 (2), pp. 227-281.
- Gowers, E. (1962) *The Complete Plain Words*. Harmondsworth: Penguin Books Ltd.
- Gramley, S. and Pätzold, K.-M. (2004) *A Survey of Modern English*. 2nd edn. London: Routledge.
- Greenbaum, S. and Whitcut, J. (1988) *Longman Guide to English Usage*. Harlow: Longman.
- Hogg, R. and Denison, D. (2006) "Overview." In Richard Hogg and David Denison (eds), *A History of the English Language*. Cambridge: Cambridge University Press, pp. 1-42.
- Ishikawa, S. (2011) "Duality in the Spelling of English Verb Suffixes -ize and -ise: A Corpus-based Study." *International Proceedings of Economics Development and Research*, Vol. 26, pp. 390-396.
- Katamba, F. (2005) *English Words: Structure, History, Usage*. 2nd edn. London: Routledge.
- Maxwell, C. (1978) *The Pergamon Dictionary of Perfect Spelling*. 2nd edn. Exeter: Wheaton.
- McArthur, T. (2006) "English World-Wide in the Twentieth Century." In Lynda Mugglestone (ed.), *The Oxford History of English*. Oxford: Oxford University Press, pp. 360-393.
- Meyer, C. (2002) *English Corpus Linguistics*. New York: Cambridge University Press.
- Milroy, J. and Milroy, L. (1999) *Authority in Language: Investigating Standard English*. 3rd edn. London: Routledge.
- Mugglestone, L. (2006) "English in the Nineteenth Century." In Lynda Mugglestone (ed.), *The Oxford History of English*. Oxford: Oxford University Press, pp. 274-304.
- Nevalainen, T. and Tieken-Boon van Ostade, I. (2006) "Standardisation." In Richard Hogg and David Denison (eds.), *A History of the English Language*. Cambridge: Cambridge University Press, pp. 271-311.
- Partridge, E. (1957) *Usage and Abusage: A Guide to Good English*. 5th edn. London: Hamish Hamilton.
- Peters, P. (2007) *The Cambridge Guide to Australian English Usage*. Melbourne: Cambridge University Press.
- Peters, P. (2004) *The Cambridge Guide to English Usage*. Cambridge: Cambridge University Press.
- Phythian, B. A. (1979) *A Concise Dictionary of Correct English*. London: Hodder and Stoughton.
- Pyles, T. (1971) *The Origins and Development of the English Language*. 2nd edn. New York: Harcourt Brace Jovanovich.
- Quirk, R. (1972) *The English Language and Images of Matter*. London: Oxford University Press.
- Scholfield, P. J. (1994) "Writing and Spelling: The View from Linguistics." pp. 51-71. Chichester: John Wiley & Sons.
- Scott (2004) "American and British Business-Related Spelling Differences." *Business Communication Quarterly*, June 2004, pp. 153-167.
- Smith, J. J. (2006) "From Middle to Early Modern English." In Lynda Mugglestone (ed.), *The Oxford History of English*. Oxford: Oxford University Press, pp. 120-146.
- Stubbs, M. (2004) "Language Corpora." In Alan Davies and Catherine Elder (eds.), *The Handbook of Applied Linguistics*. Oxford: Blackwell, pp. 106-132.

- Swan, M. (2005) *Practical English Usage. International Student's Edition*. Oxford: Oxford University Press.
- Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Trask, R. L. (1994) *Language Change*. London: Routledge.
- Treble, H. A. and Vallins, G. H. (1936) *An A. B. C. of English Usage*. Oxford: Clarendon Press.
- Treiman, R. (1993) *Beginning to Spell: A Study of First-Grade Children*. Cary: Oxford University Press.
- Upward, C. (1997a) "The Potential of Stylguides as Vehicles for Spelling Reform with a case-study of The Times English Style and Usage Guide." *Journal of the Simplified Spelling Society*, 21, 1997/1, pp. 13-20.
- Upward, C. (1997b) "American Spellings for British Schools?" *Journal of the Simplified Spelling Society*, 21, 1997/1, pp. 30-32.
- Vallins, G. H. (1965) *Spelling*. London: Andre Deutsch.
- Vallins, G. H. (1955) *Better English*. London: Andre Deutsch.
- Vallins, G. H. (1951) *Good English: How to Write It*. London: Pan Books.
- Venezky, R. (1999) *The American Way of Spelling: The Structure and Origins of American English Orthography*. New York: Guilford Press.
- Venezky, R. (1970) *The Structure of English Orthography*. The Hague: Mouton.
- Walker, S. (2001) *Typography and Language in Everyday Life: Prescriptions and Practices*. Harlow: Pearson Education Limited.
- Whitcut, J. (1985) "English, My English?" In Sidney Greenbaum (ed.), *The English Language Today*. Oxford: Pergamon Press.
- Wood, F. T. (1962) *Current English Usage: A Concise Dictionary*. New York: St Martin's Press.

Internet Sources:

- Ask Different (2011) <http://apple.stackexchange.com/questions/21429/uk-spelling-dictionary-teach-os-x-all-ize-spellings> [Accessed 25 March 2015]
- AskOxford (2008) "Are spellings like 'privatize' and 'organize' Americanisms?" <http://www.askoxford.com/asktheexperts/faq/aboutspelling/ize?view=uk> [Accessed 10 March 2008]
- BBC Academy (2015) "BBC News style guide." <http://www.bbc.co.uk/academy/journalism/news-style-guide/article/art20131010112740749> [Accessed 30 April 2015]
- Bolton, B. (2008) "NHS Choices Editorial Style Guide." <http://www.nhs.uk/aboutNHSChoices/aboutnhschoices/Aboutus/Documents/NHS-Choices-Editorial-Style-Guide-v3.pdf> [Accessed 30 April 2015]
- The British Newspaper Archive (2015) <http://www.britishnewspaperarchive.co.uk/> [Accessed 10 April 2015]
- Burnard, L. (2000) *Reference Guide for the British National Corpus (World Edition)*. Published for the British National Corpus Consortium by the Humanities Computing Unit at Oxford University Computing Services. <http://www.natcorp.ox.ac.uk/docs/userManual/> [Accessed 24 April 2008]
- Burnard, L. (2009) <http://www.natcorp.ox.ac.uk/corpus/index.xml> [Accessed 12 March 2013]
- Cambridge University Press (2015a) <https://authornet.cambridge.org/information/productionguide/hss/text.asp> [Accessed 11 March 2015]
- Cambridge University Press (2015b) <http://www.cambridge.org/gb/cambridgeenglish/about-cambridge-english/cambridge-english-corpus> [Accessed 25 April 2015]
- Dale, I. (2013) "A Question of English". <http://www.iaindale.com/posts/2013/04/23/a-question-of-english> [Accessed 26 April 2013]

- Department for Education (2013) "The national curriculum in England: Key stages 1 and 2 framework document." <https://www.gov.uk/government/publications/national-curriculum-in-england-primary-curriculum> [Accessed 30 April 2015]
- Dixon, R. (2004) "Questions Answered." <http://www.timesonline.co.uk/tol/comment/article993612.ece> [Accessed 3rd April 2009]
- The English Spelling Society (2015) <http://spellingsociety.org/about-us> [Accessed 26 March 2015]
- European Commission Directorate-General for Translation (2011) "English Style Guide." http://ec.europa.eu/translation/english/guidelines/documents/styleguide_english_dgt_en.pdf [Accessed 25 April 2015]
- Government Digital Service (2015) "Style Guide." <https://www.gov.uk/guidance/style-guide/a-to-z-of-gov-uk-style> [Accessed 25 March 2015]
- Graddol, D. (1997) "The Future of English: A Guide to Forecasting the Popularity of the English Language in the 21st Century." <http://www.teachingenglish.org.uk/sites/teacheng/files/learning-elt-future.pdf> [Accessed 25 March 2015]
- The Guardian (2015) "Guardian and Observer Style Guide". <http://www.theguardian.com/guardian-observer-style-guide-i> [Accessed 20 April 2015]
- Heuboeck, A, Holmes, J. and Nesi, H. (2010) "The BAWE Corpus Manual." <http://www.reading.ac.uk/internal/appling/bawe/BAWE.documentation.pdf> [Accessed 8 February 2013]
- Horne, M. (2012) "The '-ize' have it!". <http://www.metadyne.co.uk/ize.html> [Accessed 7 March 2013]
- Oxford Dictionaries (2011) "-ise or -ize?" <http://blog.oxforddictionaries.com/2011/03/ize-or-ise/> [Accessed 7 March 2013]
- Preston, J. (2014) "Speak plainly: are we losing the war against jargon?" <http://www.telegraph.co.uk/culture/books/10710840/Speak-plainly-are-we-losing-the-war-against-jargon.html> [Accessed 30 April 2015]
- Schofield & Sims (2015) "Schofield & Sims Progress Papers in English: Alternative spellings." <http://www.schofieldandsims.co.uk/downloads/ProgressPapersinEnglish/alternative-spellings.pdf> [Accessed 30 April 2015]
- Taylor & Francis (2001) "Instructions for Authors." <http://www.contemporarysecuritypolicy.org/assets/Routledge%20instructions%20for%20authors.pdf> [Accessed 11 March 2015]
- Transport for London (2015) "Editorial Style Guide." <https://www.tfl.gov.uk/info-for/suppliers-and-contractors/digital-design-toolkit/editorial-style-guide> [Accessed 30 April 2015]

APPENDICES

Appendix 1 The BNC in numbers. Modified from Burnard (2000).

AUTHOR DOMICILE

	texts	%	word units	%	sentence units	%
Author domicile						
Unknown	2273	72,30	56750777	65,02	3144578	62,71
UK and Ireland	843	26,81	29570097	33,88	1812550	36,14
Commonwealth	12	0,38	407076	0,46	25765	0,51
Continental Europe	6	0,19	232275	0,26	12469	0,24
USA	8	0,25	243177	0,27	15677	0,31
Elsewhere	2	0,06	74803	0,08	2936	0,05

WRITTEN MEDIUM

	texts	%	word units	%	sentence units	%
Medium						
Book	1414	44,97	49891770	57,16	2895652	57,75
Periodical	1208	38,42	28356005	32,48	1487725	29,67
Published miscellanea	238	7,57	4197450	4,80	288004	5,74
Unpublished miscellanea	249	7,92	3508500	4,01	222438	4,43
To-be-spoken	35	1,11	1324480	1,51	120153	2,39

AUTHOR AGE GROUP

	texts	%	word units	%	sentence units	%
Author age						
Unknown	2519	80,12	65457159	74,99	3707600	73,94
0-14	3	0,10	59071	0,06	3447	0,06
15-24	19	0,60	537251	0,61	29862	0,59
25-34	67	2,13	2286936	2,62	163079	3,25
35-44	191	6,08	6660606	7,63	410324	8,18
45-59	205	6,52	7157985	8,20	410717	8,19
60+	140	4,45	5119197	5,86	288943	5,76

AUTHOR SEX

	texts	%	word units	%	sentence units	%
Author sex						
Unknown	1573	50,03	35825335	41,04	1970482	39,29
Male	922	29,33	30434132	34,87	1675236	33,41
Female	415	13,20	14480939	16,59	972106	19,38
Mixed	234	7,44	6537799	7,49	396148	7,90

PLACE OF PUBLICATION

	texts	%	word units	%	sentence units	%
Region						
Unknown	690	21,95	14583761	16,70	790465	15,76
UK (unspecific)	264	8,40	7124424	8,16	383046	7,63
Ireland	37	1,18	567046	0,64	31825	0,63
UK (North)	192	6,11	3778114	4,32	230008	4,58
UK (Midlands)	93	2,96	2622554	3,00	192379	3,83
UK (South)	1854	58,97	58066891	66,53	3365045	67,11
United States	14	0,45	535415	0,61	21204	0,42

Appendix 2 Suffixes *-ise* and *-ize* in the BNC, a summary of findings.

	Number of texts in total	Number of texts with matches		
		ise	ize	either
All matches	3140	2351	1091	2868
Publication date:				
1960-1974	46	15	37	46
1975-1984	155	108	67	147
1985-1993	2777	2132	968	2573
All	2978	2255	1072	2766
Age of Author:				
0-14	3	3	1	3
15-24	19	16	12	19
25-34	66	39	34	61
35-44	191	108	110	179
45-59	205	120	116	198
60+	139	90	85	138
All	623	376	358	598
Sex of Author:				
Male	920	595	534	886
Female	414	267	195	398
Mixed	234	193	41	209
All	1568	1055	770	1493
Domicile of Author:				
UK and Ireland	841	535	465	809
Commonwealth	12	6	9	11
Continental Europe	6	2	5	6
USA	8	4	5	8
Elsewhere	2	1	1	2
All	869	548	485	836
Medium of Text:				
Book	1411	883	793	1365
Periodical	1207	1050	199	1072
Miscellaneous published	238	168	37	173
Miscellaneous unpublished	249	216	55	224
To-be-spoken	35	34	7	34
All	3140	2351	1091	2868

Derived text type:

Academic prose	497	346	254	470
Fiction and verse	452	236	215	415
Non-academic prose and biography	744	572	363	726
Newspapers	486	391	10	391
Other published written material	710	592	192	645
Unpublished written material	251	214	57	221
All	3140	2351	1091	2868

Text Domain:

Imaginative prose	476	251	219	432
Natural and pure sciences	146	92	59	129
Applied science	370	292	106	329
Social science	526	419	196	482
World affairs	483	372	207	459
Commerce and finance	295	249	73	273
Arts	261	205	88	238
Belief and thought	146	106	54	130
Leisure	437	365	89	396
All	3140	2351	1091	2868