Naming People

Lexical Expression of Demographic Categories in Written Standard British English

Kaisla Viljamaa

Master's Thesis

English, Degree Programme for Language Specialists

School of Languages and Translation Studies

Faculty of Humanities

University of Turku

December 2023

Master's Thesis

The 20th century has seen multiple social movements agitating for various rights; many taking the position that language creates social realities and therefore also agitating for language change. For example, this means focusing on the terminology used about people, either as individuals or groups. This work sets out to survey two samples of Standard British English, to see if the vocabulary in the samples has changed with respect to the demographic terminology used.

The research corpus consists of two sets of editorials extracted from *The Times*, both spanning the months of March and April, from the years of 1976 and 2009. The total number of editorials was 272, with the entire corpus containing approximately 169 113 words. The study employed quantitative methods. Instead of focusing on changes in individual terms, most of the study focused on the demographic categories individual terms represented. This allowed the tracking of larger trends among categories of expressions. The demographic categories found in the corpus were sex, nationality, nationhood expressions, ethnicity, geographical origin, religion, race, and a selection of unclassifiable terms. After categorisation, the extracted terms were counted and the numbers from the two samples were compared.

The study showed clear shifts in demographic terminology between the samples. Within the sex category terminology shifted away from androcentric generics and towards neutral forms. Among the other categories the terminology shifted towards nationality, religion and ethnicity, and away from the other demographic categories. Further research on larger corpora, or corpora across larger time spans, would also show changes for individual expressions and how long this process has been present.

Key words: demographic terminology, standard British English, editorials, language change, sociolinguistics, race, sex, nationality, religion, ethnicity

Table of contents

List of tables and figures

List of tables

List of figures

1   Introduction

> The world is not simply the way it is, but what we make of it through language. (Romaine 1994:29).

> There are two significant confusions around the phrase 'Standard English': that around 'standard' and that relating to 'English' (Crowley 1999:271).

In the above quote Suzanne Romaine articulates a common view, that language not only describes, but shapes reality, or at least the part of reality pertaining to the social world humans inhabit. While reality itself is not changed by the language used to describe it, certain conceptions of reality may well be vulnerable in this way. Humans express and make sense of the everyday reality they live in through concepts and categories, though this does not mean that there is no reality beyond concepts and categories, nor that this reality is fundamentally unknowable. It does mean that there is an uneasy relationship between reality, observations of reality, the conceptualisation of those observations, and finally the discourse around this conceptualisation. In short, the relationship any given concept or category has to reality at any given time is not necessarily straightforward. This is particularly the case with everyday language, where rigour may at best be merely one of many considerations. The concepts and categories we use to make sense of the world are thus, by necessity, in flux as they are reworked and refined.

Much of the 20th century can be seen as being characterised by various progressive movements agitating for legal, economical, and social rights. In the second half of the century these movements began to pay attention to the aforementioned complex relationship between reality and the language describing it, and set about correcting social ills through changing the language. As language is at the very heart of all human interaction and thus precious to many, these efforts met with both enthusiastic support and passionate resistance. The basis of the effort to change society through changing the language is essentially the Sapir-Whorf assertion that language controls how its speakers view the world, from which it is not a leap to conclude that language also affects social reality. This view, combined with the general thrust of social

movements such as the women's movement, the labour movement or movements for indigenous rights, aiming for what they consider increased fairness and decreased discrimination, has overall led to consistent calls to reform the language in a way that would remove the perceived discriminatory features and possibly insert new preferred ones. These efforts however, have not gone unopposed, with complaints about what is often called *politically correct* language being easy to find. The people advocating the changing of language to reflect their social ideals base their arguments on the previous point, whereas the critics of such attempts often appeal to established practices along with arguments of common sense and naturalness. While the arguments still rage back and forth, less attention is being paid to their effects. The *ought* is being discussed heavily, while the *is* gets relatively neglected. The focus of this work lies precisely here, in the question of whether all this sound and fury has in fact amounted to much.

At the heart of these social justice movements is the concept of people being divided by society into various demographic categories, and these categories holding various statuses. That makes the terminology surrounding these categories of particular interest to those, who wish to effect a change in these status differences through changing language. At the most basic level, this means words used to name these categories, or to signal a person's membership in a particular category.

Thus we have a situation, where the normal, continuously occurring language change is accompanied by a conscious, ideological effort to effect specific changes in language, along with a conscious opposition to it. While this is unlikely to be a unique occurrence in the history of language, the fact that it is happening before our eyes and through modern mass media, allows for a unique chance to study it. It is important to note that this project of imposed language change is an ideological one, and as such, no more natural than the opposition to it. Evaluating the efforts in question and their justifications is firmly beyond the scope of this work. The goal here is to merely observe this potential language change in action, not to pass judgement on it.

Such observation however, has so far largely been missing from the field of linguistics. The subfield Critical Discourse Analysis (henceforth CDA) commonly deals with how demographic categories are realised through language, but as it is explicitly political in its stance, it does not engage in purely exploratory mapping of phenomena as such. Therefore the change in how demographic categories are expressed lexically falls largely outside the typical scope of CDA.

As for methodology, while much of CDA does not engage in comparing corpora, either diachronically or in a single moment in time, comparative diachronic CDA studies do exist. One recent example particularly similar to this work's research design is detailed in the 2020 article by Gerlinde Mautner and Mark Learmonth. The study in question examines the change in lexis in the articles of an academic journal on business and management studies, utilising a corpus of over fifteen million words, specifically constructed for the study in question. In addition to this, the study focuses on the labels used of different referents, or as the authors put it "lexical items that represent social actors" (2020:273). The main differences between Mautner and Learmonth's research on the lexis of business and management studies and this work are that Mautner and Learmonth operate within the critical studies perspective, and with a considerably larger corpus. This means that they both expend considerable energy on analysing the data from their own political perspective, and also that the size of the corpus limited them to a subset of all available relevant lexis. This does not diminish their study as such, it merely points at the necessary trade-offs involved in formulating research questions and designing research. In contrast, this work does not take a stand on the social or psychological effects of the relevant lexis, and utilised a smaller corpus in order to fully capture the spectrum of the lexis present in the texts. Mautner and Learmonth themselves allude to these choices, when they note that "[d]ata aggregation always comes at a price" (2020:285).

In general comparing corpora is largely the purview of computational linguistics, or corpus linguistics, but there the focus is on large corpora. According to Beeching (2006:49) the typical corpus design involves a large quantity of data being collected from a sample deemed representative of the population, including both written and spoken language. While there are computational methods which allow for some open-endedness with regard to the examined lexis, such as automated word filters, the larger the corpus, the more difficult it is to embark on a fully exploratory study of its lexis. The branch of linguistics which corresponds methodologically most closely to this work, in that it regularly engages in exploratory comparative work on small corpora, is forensic linguistics. In fact many texts subjected to forensic linguistic analysis are close to the opposite extreme; e.g. ransom demands or suicide notes typically consist of fewer than 200 words and often fewer than 75 (Coulthard, Johnson and Wright 2017:9). However, due to its focus on the criminal justice system and applications therein, it does not engage in general surveys on language.

Therefore, there is a significant gap in research when it comes to how demographic categories are realised in English, and whether this has changed within recent history. This work attempts to do its part to fill this gap. Its goal is both to explore how demographic categories such as sex, religion or ethnicity, are expressed in the English lexicon, and to perform an initial survey as to how those expressions have changed over 33 years, in a specific subsection of standard English as it is expressed in the sample corpus. As similar work has not previously been done, it is by necessity a process of discovery, much like mapping uncharted territory. Much of the methodology was thus determined by the qualities of the corpus and the data emerging from it, and the work as a whole is driven by the data itself. The process consisted of first forming a corpus out of two samples of editorials of *The Times*, then reading through both samples, extracting all unambiguously demographic terminology. The found terminology was then classified according to the underlying demographic category, and in the end the quantified data from both samples was compared to see if there were noticeable differences between the two. The research questions in this work are the following. How are demographic categories expressed in the lexis of Standard British English, as it appears in newspaper editorials? Has this changed over the 33 years in question? Are there practical benefits to analysing such lexis through the lens of demographic categories, rather than looking at individual expressions alone?

The first of the following sections, chapter two, describes the starting points of this study, briefly outlining the ideological basis behind efforts to push for this language change, specifically as it pertains to the field of linguistics. Chapter three will outline the concept of demographic categories as it is used in this study and introduce the main categories here considered, with chapter four describing the research material and the methodology used. Chapters five and six will contain a report on the findings, with the former detailing those pertaining to the category of sex, and the latter focusing on the rest of the demographic categories found in the texts. Chapter seven will contain a summary of the results, as well as suggestions for future research, and chapter eight will contain the conclusion.

2   Language change and social justice ideologies

At the core of this study are two observations. The first being that within societies, people are grouped into different categories according to various demographic variables, such as religion, ethnicity or sex. This is done not only by various institutions, but by the people themselves, ranging from official statistics to self-segregating social groups based on demographic variables. These groupings are expressed linguistically through the naming of the categories and through vocabulary denoting membership in a category. Examples of this naming are such words as *Christian* or *Norwegian,* denoting religion or nationality respectively. These category structures can be anything from the complex sets of categories present in modern societies to the basic concept of "us" and "them". The more complex the society, the more complex one can expect the system of demographic categories to be as well, as the realities of a given society place specific needs on its language. For example, in a society allowing multiple organised religions to exist simultaneously, religion is a meaningful demographic category. However, in a society with a single, largely homogenous set of non-organised religious practices, it would make no sense to attempt to use religion as a basis for grouping people. This would be akin to a civilisation from an arid desert environment having little use for multiple different terms for trees.

The second core observation is that the 20th century saw a number of left-leaning social justice movements sweep through western societies. These movements aimed at anything from small changes in individual practices to large-scale changes in the fabric of society and it was arguably inevitable, that language should also become a target for reform. Many vocal criticisms of existing language practices as well as demands and suggestions for specific changes have come from sources which from a linguist's point of view can be considered lay ones. However, academic linguistics has its own branch of activist linguistics, which has taken as its task the reformation of society through language. The arguments for language change may vary in detail, but they largely maintain similar core ideas, so the academic branch of language activism, specifically the discipline of Critical Discourse Analysis (henceforth CDA), can be taken here as representative of this larger movement. This does not mean that CDA is here treated as, or assumed to be the cause of the language change under investigation. It is merely used as a proxy for the various efforts of different social movements to influence language.

By and large, Critical Theory starts from the axiomatic assumption that society is divided into groups of people, some of which unjustly hold power and privilege over others, thus following the general ethos of the 20th century social justice movements. CDA for its part maintains, a priori, that language is fundamentally a set of discourses which serve to maintain existing power and privilege structures, and that the duty of a CDA practitioner is to attempt to remedy this state of affairs. Wodak states this explicitly, as according to them a "defining feature of CDA is its concern with power as a central condition in social life, and its efforts to develop a theory of language which incorporates this as a major premise" (2001:11). This also is laid out explicitly in the preface for *Text and Practices: Readings in Critical Discourse Analysis,* where Caldas-Coulthard and Coulthard state "Critical Discourse Analysis is essentially political in intent with its practitioners acting upon the world in order to transform it and thereby help create a world where people are not discriminated against because of sex, colour, creed, age or social class" (1996:xi). The reason for taking aim at language is simply that CDA practitioners "view social practices and their linguistic realisations as inseparable" (ibid:xii). The main method for this project is, according to Fowler's explanation in the same volume, challenging and exposing misrepresentations and discrimination which are assumed to be present in discourse, as well as how power is seen to be mediated through language (1996:5).

These ideas and attitudes expressed in CDA writings can be taken as good and illustrative examples of the greater movement to reform society through reforming language, since they are largely shared by the activist movements outside of CDA itself, even if they are not articulated in exactly the same ways. As this movement has been on the ascent for several decades now, the interesting question is whether there are any signs of it having an effect in language as it exists out in the wild. This work focuses on that question, not on the claims of language's discriminatory nature themselves, but on whether any evidence can be found on whether language has in fact changed in ways preferred by the reformers. It must be here pointed out that this work is in no way an examination on the effects, or validity of CDA in particular, rather CDA is merely a convenient illustration of the ideas behind the ideological movement to effect language change. CDA would no doubt be interested in a mapping of demographic categorisation (Fowler 1996:11), but this is not a CDA work. While CDA would set out to question and problematise these categorisations, the purpose of this work is merely to discover how they manifest, and whether that has changed over a particular time frame.

3 Demographic categories as a conceptual framework

As stated above, the phenomenon this work is focused on is the lexical expression of demographic categories and how that changes over time. The main variables on which these categories are typically based are real or perceived physical qualities, ethnic or national groupings, and religion. The categories based on these variables may overlap with each other and it may not be clear which underlying demographic variable a particular named category is based on. For example, ethnicity and nationality can be very difficult to tease apart, and some religious categories may also be seen as racial ones, Jewish being a notable case. These variables, and thus also the categories they are based on, are here divided for purely practical reasons into two main groups; biological sex and everything else. The reasons for this are both that the former is expressed through a small number of specific, well-established forms which include grammatical features, whereas the latter are exclusively expressed through vocabulary, and that there has been considerably more attention paid to the former. There is thus in a manner of speaking more meat to these bones, than is the case with any other single demographic category. It should be noted that present arguments regarding the salience of biological sex as a base level demographic category expressed through language are irrelevant for the purposes of this work, seeing as the text samples under investigation predate these arguments, as well as any developments which may have followed from them.

The remaining categories are in this work grouped under the title *faction*. This is done again for practical purposes, so that there is a single, clear and simple term which covers all the other categories, but can not be confused with other concepts. The word is here used merely in the sense of a 'population group based on a particular variable', and is the author's own usage.

3.1 Sex – an introduction

Unlike languages such as German or French, English has no true grammatical gender. What is normally referred to as gender in English is in fact a set of lexical labels denoting the category of biological sex. The relevant variants in this category are female and male, with the additional possibility of treating the sex or a referent as unknown. The main question here is how these variants are expressed linguistically and whether there have been changes in this during the study period. These expressions are here divided into two groups, the androcentric generics and the asymmetrical address forms.

The androcentric generic refers to the practice of using male nouns or pronouns as a referent for groups containing both males and females. Examples of this include the practice of using the word *man* as a generic expression in place of the word *human*, or using the male third person pronoun when referring to people in general, or to a single person whose sex is either unknown or unsaid. The main argument for changing expressions identifying the sex of the referent has been that the traditional forms excluded and rendered invisible the category of female*,* while treating the category of male as the prototypical, unmarked category. Consequently, the opposition to these efforts has consistently claimed that no such universalisation of the male category is actually taking place and that the male oriented lexis indeed does truly encompass the female and that asymmetric treatment of these categories in no way signifies differences in status.

There has been some empirical research which supports the claim that these expressions render women invisible in the language, possibly even exclude them from the pool of possible referents (Doyle 1998:152). In particular androcentric generics, male terms claimed to include females, appear to have this effect (Romaine 1994:114). Spender (1980:152-153), Fasold (1990:112-113) and Romaine (1994:115) all cite studies and experiments, the results of which suggest that the inclusive male generic is not always quite as inclusive as it is supposed to be, and that people consistently link male generics with male images. According to Spender (1980:152-153), in a study by Wendy Martyna it was discovered that not only did women not think of themselves in connection with the male generics, but they also used these generics less than men. In a study by McKay and Fulkerson in 1979 cited in Fasold (1990:112-113) American students were asked if a sentence with a generic masculine form could refer to one or more females. According to Fasold (1990:113) the "[...] results are typical to what is found in this kind of research: when 'he' is used, regardless of the intent of the speaker or writer to use it in the generic sense, it will almost always be heard as excluding female referents". Romaine (1994:115) reports that in experiments where people have been asked to draw pictures to go with texts containing language which uses only male sex vocabulary, there is a strong tendency for the subjects to draw men. One of the more extreme examples used in arguing against androcentric generics is mentioned by Friederike Braun in discussing how Swiss women were denied suffrage on the basis of the constitution having been written in the grammatical masculine (1997:6-7). This was clearly a blatant political manipulation of language, but it is nevertheless used to elicit questions of whether such language has an effect on whether women are seen as people and whether people are seen as men (Braun 1997:4-7).

A second main form of sex differentiated language which has been argued against, are asymmetrical address forms. According to Suzanne Romaine, these are either feminine agentive nouns or the titles which for women express marital status, but for men are marriage neutral. Unlike with androcentric generics, the arguments against the asymmetrical usage of address forms are not that they render women invisible. Instead the claim is that they signify the differences in status between men and women, by for example making the male category the norm of which the female category is a deviant version. A typical example of this is the usage of marked titles when the referent is a woman. Whether these titles are constructed by derivation, such as *actress*, compounding, like in *woman lawyer*, or adjectival modification, such as *female doctor*, they all said to convey the idea that the woman is not quite the real thing, that she is somehow an exception to the rule of what actors, lawyers and doctors are commonly known to be. There are also claims that they are occasionally used with the deliberate purpose of belittling the achievements of women (2001:157-158).

Another type of asymmetrical address forms, which has attracted a great deal of attention is the way that the titles *Miss/Mrs* express the marital status of a woman, but the title *Mr* does not show the marital status of a man. According to Romaine this practice is objected to on the basis that it reflects the inferior status women have in society; it is seen as important to signal whether a woman is available or not, while no such distinction is deemed necessary when men are concerned. Apart from the actual form of the title, the asymmetry may also lie in the way that the titles in general are used when addressing women or referring to them. For example in sporting events, female athletes may be referred to with a title *Miss* or *Mrs* and their last name, while the male athletes would be referred to with their last names only. In addition to this, women are more likely to be addressed by their first names than men (1994:110-111).

3.2   Faction – an introduction

The most salient common feature among individual faction categories is that they are expressed exclusively through naming. This does not, however, make them easier to analyse. One of the major problems with investigating the vocabulary in this group is that the concepts of race, ethnicity, nationality, religion, etc. are anything but clear and well defined. Not only do these concepts have a muddled history, a modern illustration of the confusion is how the word *racism* has come to refer to prejudices based on decidedly non-racial concepts, such as religion. This confusion naturally causes some difficulties when it comes to linguistic analysis, even with simple quantitative questions such as the ones discussed in this work.

The basis of these demographic labels is the conceptualisation of the groups they refer to. As these conceptualisations have changed over time and new ones have emerged, the words used to label them have by necessity also changed. For example, the concept of nationality in its modern sense naturally requires the existence of nation states. Ethnicity has no such limitations, but at the same time is more nebulous, while religion as a means for identifying population groups has possibly been around as long as humans have performed worship rites. Race on the other hand is a more novel concept, and one which has morphed considerably during its existence, the famous example being the meaning of the word *white* changing its meaning through both time and space. Neither has the importance of any given category remained stable. In addition to the categories behind the words changing, the vocabulary itself has also changed. This applies both to the selection of words in use, and the connotations attached to individual words.

The push to change the part of language here designated as faction vocabulary focuses on two things, the connotations associated with particular words and the argument that groups of people should be allowed to name themselves. The first approach can be applied to larger discourses as well as individual words. The article *The genesis of racist discourse in Austria since 1989* by Ruth Wodak (1996: 107-128) is an example of taking an entire segment of discourse and analysing how it represents particular groups of people. In this article, Wodak makes the argument that the negative concepts and words associated with these groups constitute racist discourse. The movement to make the vocabulary less racist rest largely on this same argument, that certain expressions carry with them negative ideas about the people they refer to, and should therefore be replaced with neutral or even positive expressions.

Connotations of words are not stable and may over time drift from neutral to negative ones. For a language reformer this means that old terms, which have become laden with negative baggage, should be discarded in favour of other terms, at present of neutral value. It should be noted, that this may take place even in the absence of deliberate pressure to change language for ideological reasons. Over time, some words denoting population groups have become completely undesirable in what is normally considered polite company, while many other words denoting race or ethnicity have begun to be phased out. For example, the U.S. civil rights activist Martin Luther King freely used the word *negro*, but that word has since become much less acceptable. Additionally, many of the words deliberately targeted by reformists began as at least ostensibly

polite and neutral words and only became contested as the societal attitudes surrounding them changed.

The second main argument for vocabulary reform centres on the argument that people representing specific demographic groups should have the right to name themselves and that the language at large should adopt the preferred terms of the people in question (Leitner 2008:200). The important thing to note here though, is that population groupings are often ill-defined and nebulous and designations chosen by the groups themselves are not necessarily exempt from this. It is also the case that the act of naming a group may in fact end up creating the group, especially when self-definition is concerned. Similarly assigning people to this or that group is not necessarily a straightforward matter and thus juggling people between groups is possible. This means that the basis on which population naming schemes rest, is not stable, and does not necessarily follow logical rules. A good example of these groupings is the *Application for a Somerset Travel Pass* shown in appendices 5 and 6, where for example the term *Asian* includes ethnicities from the Indian subcontinent, but not the Chinese or Japanese, who are in their own group *Chinese or other ethnic group.* It is also interesting to note, that in the very same source there is a separate section for *Romany, Gypsy or Traveller*, with *Romany* and *Gypsy* as separate entries. Whether this decision was made entirely by the company or after consulting the populations who call themselves these names, is unknown.

The two possible motivations for phasing out certain names and bringing in others are the desire to avoid negative connotations and the goal to allow people in particular demographic groups to determine for themselves how they are called. For the purposes of this work, these two motivations are in practice indistinguishable, as both would result in a simple shift from one set of terms to another.

Unfortunately, while there exists a sizable body of research in psycholinguistics on the actual effects of things like androcentric generics and asymmetric address forms, there seems to be little to no such research on faction vocabulary. Plenty of linguistic scholarship argues, like Wodak, that certain expressions form racist discourse, but the experimental demonstration of the effects of faction vocabulary seems to be lacking. Similarly, a basic mapping of faction vocabulary also appears to not have taken place. While research exists on e.g. the aforementioned differences in address forms between male and female referents, a frequency mapping of faction based terminology in the style of this work, does not seem to have occurred.

3.3 Faction – race, ethnicity, nationality, religion

In everyday use, these demographic categories are not clearly defined and there is considerable intra-category, as well as inter-category confusion. Religious groupings may be used as markers for nationality, race related concepts may be used as proxies for ethnicity, and so on. Likewise terminology related to e.g. ethnicity or religion may be conflated with race, especially in discussions on racism. Van Dijk argues in favour of this approach in *Elite Discourse and Racism,* where he makes the point that "modern racism is no longer primarily racial, but also culturally based and legitimated" (1993:15). It is prudent to note however, that it is very questionable indeed whether people even make a consistent distinction between concepts such as race, ethnicity and religion in the first place, especially considering the muddled history of these concepts. This means that teasing out demographic terminology is not necessarily easy, but also that there is an opening here for research which attempts to get at the base of this terminology, i.e. the demographic categories on which the terms themselves are based.

Appiah and Gutmann (1996:69) make the point that as a biological concept, race is difficult to use for objectively classifying people into groups and no matter how it is done the resulting categories will contain within themselves almost as much genetic variation as there is within the species. They also note that while some characteristics appear to lend themselves easily for racial classification, most notably perhaps skin colour and hair, not only are they distributed unevenly in the geographical sense, they do not correlate much with other characteristics commonly associated with race (ibid:68). Over time the apparent ambiguity of race lead most biologists and physical anthropologists to abandon the entire concept. This failure to find a biological basis for race leads to the conclusion that the concept is in fact a social, economical or political category, and in social sciences it has indeed been increasingly viewed as such from the eighteenth century onward. Modern social sciences view race as a social concept, a "variable which is shaped by broader societal forces." (Rothenberg 2001:12-13)

The fluidity of the concept of race in its modern form is a very good example of intra-category confusion, and can be illustrated by the words of *white* and *black*. On the surface these might seem simple, but it takes little effort to find inconsistencies. Does *white* for example equal European-born as it does in *The UK Space, Resources Environment and the Future* (henceforth *The UK Space,* 1977:148). Does it include people from the Mediterranean area? Or what about Jewish people many of whom undoubtedly are white as far as their complexion goes? And what about *black*, does that only mean those of African origin, or does it include anyone dark

skinned? *The UK Space* from 1977 still uses the word *coloured*, whereas the 2001 Official Yearbook has moved to using *black*. One might be lead to question whether there even are clear racial groupings in the society, or whether the lexical items denoting them are merely treated as if such groupings exist.

Of the concepts here presented, *nationality* and *ethnicity* are most closely related to each other. Ideas of nationhood began to emerge in the nineteenth century, but as the political boundaries of the time did not correspond to cultural boundaries according to which nationhood was defined, it came to be associated with questions of common descent and understood as a biological unit. Here too the idea of a shared essence was crucial (Appiah and Gutmann 1996:63). Thus *nation* was not originally a political unit, but closer to the modern concept of ethnicity. The nationalist movements of the nineteenth century used cultural and linguistic characteristics to draw dividing lines between groups of people. With the emergence of nation states this term has come to a closer association with political entities than it originally had, while the increased awareness of smaller, culturally independent population groups within those states has brought the term *ethnicity* closer to what *nationhood* meant in the nineteenth century. Where race as a concept has been on the wane, such is not the case with ethnicity, if anything ethnic differences may have increased in importance as other distinguishing categories have lost some of their former importance. In the modern context, *nationality* refers to, at its simplest, to an official membership of a particular modern state.

Ethnically based lexis has one clear difference to racially based. As the latter is based on real or perceived physical characteristics, regardless of where the person is born or grows up, they remain in the racial group allocated to them by their appearance. Whereas with ethnic classifications, the way a person is grouped should in theory be based on their acculturation rather than their physique. This should make it possible for a person to become European or Asian so to speak, if they just are brought up inside the culture in question. In practice however, the terminology does not always follow this logic. If a people are immersed in a culture and grow up learning it as their own, yet are labelled with a foreign ethnic term regardless, that would seem to suggest that the label is about something more than just the ethnicity or nationality expressed by them. As stated in Abercrombie et al, many people labelled as *black, Asian* or *west-Indian* are in fact born in Britain, not immigrants or newcomers (1988:242). Similar issues may come across with the concept of nationality*,* as it is not a foregone conclusion that an immigrant who has obtained official citizenship will in reality be referred to

with the relevant nationality term. This illustrates the considerable inter-category confusion apparently inherent to these demographic categories.

The last of the main categories introduced here is religion. It overlaps and mixes with the previous ones and varies from being the dominant defining feature to making little or no difference at all. Within the past decades, religion has gained additional focus with the rise of militant religious movements as political and military forces, most famously the militant Islamist movements. As modern states emerged in Europe, they gradually moved from the one state – one religion principle to allowing for the possibility that multiple religions could exist within the same political unit. This separated the category of religion from the category of membership in a state. In the context of a modern society religion is one of the multiple sub-categories applied to people within a society, but unlike the other categories here mentioned, it also functions as an effective international demographic grouping. The thing to pay attention to here, is that terms which appear to refer to religion, may in some cases be understood as ethnic or race markers, the word *Jewish*, or *Jew* being a good example of this. However, these usages are effectively invisible unless they are somehow specified in the context. The importance of religion as a demographic divider depends heavily on the social context, being paramount in some communities, while getting largely ignored in others.

4   The material and the methods

This is a quantitative study performed on a text corpus collected specifically for this work. The work does not contain close reading of passages, but focuses purely on recording the frequency of the features in question in two samples of text, spaced some time apart. Thus this study presents a temporal perspective on the features here discussed, its goal being to discover whether there has been a change in these particular usages over time.

4.1   Research material

The research material consists of two samples of editorials from *The Times.* This newspaper was chosen as the source for research material largely due to its status as a respectable and well established newspaper with leanings more towards the conservative than to the radical. The aim was to find a source with enough authority, social standing and propriety for it to arguably present an accurate sample of what could be considered as standard English. Standard written English was chosen both because it is by its nature more stable than other varieties, and because it would reflect general attitudes on what is the proper way to use language. Another criterion for the source was that it not be easily swayed by fashions in substance or in form, thus the decision to pick a conservative rather than a radical newspaper. The extent to which *The Times* fulfils these requirements can of course be debated, but considering both the scope and the practical circumstances of this work, it was deemed sufficient.

Each sample spans the months of March and April, the first from the year 1976 and the second from 2009. The 1976 editorials were obtained from the *Times Digital Archive* available through the University of Turku, while the 2009 editorials were acquired through a direct subscription to the online edition of *The Times.* The thirty year time span was chosen in order to achieve sufficient time for linguistic differences to emerge, while the three additional years arose from unavailability of material in 1979, as *The Times* was not published at all for a considerable time during that spring due to industrial action. As the newer editorials were collected directly from the newspaper's online issues as they were published, this meant either having to postpone the sample collection, or adjusting the year of the earlier sample. The 1976 editorials were scans of the original newspaper. These scans were converted to text documents with optical character recognition software in order to allow the use of a word processor in the analysis process. Some of the original scans had quality issues ranging from slightly spotty lettering to a few instances of completely illegible text. This together with the failings of OCR software produced some

inaccuracies, which will be detailed below where relevant. No such issues arose with the 2009 sample. Where passages have been extracted from the editorials to give usage examples, the OCR inaccuracies have been edited out, as they are irrelevant to the study itself.

The samples from 1976 and 2009 consist of 118 and 154 editorials respectively, totalling 272 pieces of text altogether. The 2009 issues typically had 3 editorials each, while the 1976 issues often had fewer, sometimes only one editorial in an issue. Each editorial was identified separately with a serial number consisting of the date of publication and a letter to set apart different editorials published on the same date. The 1976 sample does not include the editorials for March the 22nd and April the 16th, as these issues were missing from the *Times Digital Archive*. The word counts of the samples were automatically generated, with 88 544 words in the 1976 sample and 80 569 words in the 2009 sample. However, since the original scans of the 1976 texts contained inaccuracies produced by the OCR process, as well as the printing quality of the original texts, the word count for the 1976 sample is not an exact number, but a close approximation. The 1976 sample is approximately 9 percent larger than the 2009 sample, but for practical reasons the samples will here be treated as being approximately the same size. In order to minimise any possible effect of this difference in sample sizes, all the numerical results were normalised, where applicable. This was done by calculating the frequency of occurrences per 100 000 words, using the multiplier 1.1294 for the 1976 sample, and 1.2412 for the 2009 sample, as shown in appendix 1. Appendix 2 shows the same data, further divided by whether given texts contained usable examples or not. In the discussion on the results the absolute numbers will be presented first, followed by the normalised numbers. This is done because the two samples are nearly comparable in size as is, and to emphasise the original numbers, as many of the terms appeared in the samples only once or twice, or not at all. Focusing on such small normalised numbers and change percentages calculated from them might give a false impression of drastic changes, when in fact the difference between samples may be only one or two occurrences.

None of the usage examples in the texts were rendered unusable by the technical issues present in the 1976 sample. Where an example usage was unclear in the text document it was always possible to ascertain the original form from the relevant scan.

4.2 Research methodology

The texts were examined for examples of expressions related to either sex or faction, such as the generic *he*, different address forms, terms like *Sudanese*, *Hindu*, *European*, *Slav*, or combined expressions such as *Xhosa Transkeians*. For expressions such as the third person pronouns this process could be automated, but for most faction terms it was necessary to read the texts and manually make note of the examples found. The texts were also divided into five categories according to their topics, to see if there was a correlation between the subject matter and the number of examples found in the texts. These categories were *international issues*, *economy or budget*, *domestic politics*, *general domestic issues*, and *art, culture and personas*, as shown in appendices 3 and 4. The examples found in the texts were recorded in tables, grouped according to the demographic category they represented. When all admissible examples had been extracted from the corpus, the number of examples in each group was calculated, and several comparisons were run on these figures, with both relative and absolute frequencies considered. The samples were compared together and also both relative and absolute frequencies of terms in different groups were noted. This data was then condensed in a number of different tables and figures, which are presented in later chapters.

The greatest practical issue with the research material with regard to the purpose of the study, was the apparent homogeneity of the populations the texts discussed. This was especially true of the category of sex, as there was a clear difference between the number of references to men and to women. This was especially the case in the 1976 sample. This was less of an issue in the faction categories, as a clear majority of all texts concerned the topic named international issues, with 105 texts, while the next most common was the topic labelled general domestic issues, with 86 texts. Editorials dealing with international issues will naturally include more references to different populations, as there will be more of a need to differentiate e.g. between people of different nationalities. Attempts to discuss the depictions of people of heterogeneous backgrounds are of course more difficult, if such people are represented poorly or not at all in the material at hand. Such under representation may reflect the general situation in the world of professionals, politics and international affairs which news are made of, especially in the editorials from 1976. This dearth of references became apparent quite early on and lead to the expansion of the time frame from one month to two months. This proved sufficient for the purposes of this research, indeed larger samples would likely have made the research unwieldy and resulted in a significantly increased workload. There was little or no correlation between

the lack of examples and the subject matter of the editorial as far as referent sex was concerned, but with faction most of the texts with no examples were in the topic category general domestic issues. This is not particularly surprising, as a host of the faction terms dealt with populations foreign to the UK. The rest of the texts were more evenly divided among the remaining topics, with economy or budget, and domestic politics categories both containing 30 editorials each, and the art, culture and personas containing the fewest texts, with only 21 editorials. This division is shown in appendices 3 and 4, with the first showing the overall distribution of topics in the editorials, while the second shows the topic distribution of editorials with and without useful examples.

Some ostensibly applicable examples had to be excluded, for they could not be established as genuine examples of sex or faction terminology. With regard to faction terms, this meant that any direct references to nations or national institutions were not included in the examples. The reason for this was that any such reference is by its nature limited to official terms of nationality, which follow the borders of political entities and are thus not subject to the same pressures as terms denoting cultural entities. Phrases such as *the British Government* or *the French army* were considered inadmissible since as long as the political entities in question exist, the terms referring to them will remain stable. Some uncertainty was presented by sentences such as *the Spanish soldiers*. In these cases the admission or inadmission of the usage was determined largely by the context and on whether it supported the notion that the word *Spanish* here referred to the nationality of individual soldiers, rather than the phrase being an euphemism for the national army. Where this could not be determined the example was left out. A subset of terms referring to official national entities was included however, due both to its emergence in the material and the way it stood out from the rest of the usages in this category. This was the practice of referring to a country through the entirety of its citizenry. In this usage the political entity was equated with its citizens, for example *the Greek* was used in a context in which it obviously referred to the official political entity of the Greek state, not individual citizens working as a group. This usage is heretofore referred as *nationhood* to separate it from *nationality,* which is used for simple references to the said category, and it will receive its own chapter in the discussion to follow. Here too there was some unclarity, when for example such phrasing was used in reference to an official delegation in such a manner that it was impossible to ascertain whether it was meant to refer to the delegation only or the nation behind it as well. Such cases were omitted from the data.

In addition to the above problematic examples, there were two groups of terms which presented a degree of unclarity, but which were nevertheless included in the study. These were the usages which combined two terms from different categories and the usages which could not clearly be classified in any of the categories or subcategories in question. The former consisted of such examples as *Muslim Lebanese* where terms of religion and nationality are combined, or *Greek Cypriots* which is a combination of ethnicity and nationality. The latter has such usages as *Maronite* or *Janjawid* which could not be established as either referring to an ethnic or religious group and which may quite likely have referred to both at the same time.

Similar issues were present with the terminology on sex, where in some cases it was difficult to decide whether something should be included as an example or not. The group that roused most questions were the androcentric generics. On several occasions it was very difficult, if not almost impossible to determine whether a specific usage was meant to be a generic form, or refer to men alone. This ambiguity does not support the claim that androcentric generics are truly generic in meaning, and it also raises questions about the usefulness of generic masculines. After all, language often tends to move away from ambiguity, and the ambiguity of generic masculines may be one of the reasons behind the persistent usage of proscribed forms (Bodine 1998). Again, the cases where it was not clear whether the form in question was generic or not, were not included in the data.

None of these issues proved insurmountable and the number of potential examples which had to be left out did not threaten the goals of the study in a meaningful manner. This was also the case with the practical issues arising, among other things, the quality of the original scans.

5 Sex-based categories in the research material

One of the most visible demographic categories in the research material was sex. The previous sections of this work detailed the different ways sex is made visible in English vocabulary and introduced the controversies that have surrounded these usages. The following is a detailed examination of how this vocabulary shows up in the corpus used for the research, with each particular subtype of expression addressed in turn.

5.1 Quantitative overview of sex categories

The general features of the sex based expressions considered in this work have been discussed above in subchapter 3.1. The following chapter will include a discussion on how these features were presented in the research material.

In the 1976 sample the portion of the texts not containing usable examples was smaller than that of useful texts, unlike in the 2009 sample where the portion of useful texts was smaller. This was true both with regard to the number of editorials in question and to their word counts. Interestingly enough, while the number of useful editorials was smaller in the 2009 sample, the absolute number of direct references to individually identified women had increased from the 14 references in the 1976 sample to 65 in 2009. This is indicated by the collated number of all address forms in the respective samples, as shown in Table 1. The normalised frequencies detailed in Table 2 show this to be a 410.2% increase. So while individual women were referred to in fewer texts, when they were referred to it was done with greater frequency than earlier. Still, while the trend within the material was clearly rising, in close to 170 000 words, there were all in all only 79 admissible examples of direct references to individual women, as seen in the sum of all address forms in Table 1. The same trend did not apply to androcentric generics, namely the generic *he* and the generic *man,* which showed a clear decline in number over time from 121 in 1976 to the mere 25 of 2009. This strong decline meant that the increase in references to women was not enough to maintain the number of examples found in the texts and the absolute number of examples for this expression category fell from 139 in the 88 500 word sample of 1976 to 100 in the 80 500 word sample of 2009, as shown in Table 1. The normalised numbers in Table 2 show this to be a 20.9% decline. The changes within individual example terms are shown in Figure 1. Of the eight terms surveyed, five increased in frequency, while the terms *Miss*, the generic *he* and the generic *man* decreased. Three of the terms were not present in the 1976 sample at all and therefore their change percentage could not be calculated. Of the remaining, the greatest

Table 1. Overview of the examples in the sex category, absolute numbers

Sex	1976	2009	Total	Change
Mrs	10	12	22	+2
Miss	3	0	3	-3
Ms	0	18	18	+18
0-address	1	35	36	+34
All address forms	**14**	**65**	**78**	**+51**
Generic *he*	92	12	104	-80
Generic *man*	33	14	47	-19
All androcentrics	**125**	**26**	**151**	**-99**
Deliberately neutral	0	9	9	+9
Feminine agentive	0	2	2	+2
TOTALS	139	102	241	-37

Table 2. Overview of the examples in the sex category, normalised numbers

Sex	Normalised 1976	Normalised 2009	Normalised total	Normalised change	Change %
Mrs	11.3	14.9	26.2	+3.6	+31.9
Miss	3.4	0.0	3.4	-3.4	-100
Ms	0.0	22.3	22.3	+22.3	-
0-address	1.1	43.4	44.6	+42.3	+3746.5
All address forms	**15.8**	**80.7**	**96.5**	**+64.9**	**+410.2**
Generic *he*	103.9	14.9	118.8	-89.0	-85.7
Generic *man*	37.3	17.4	54.6	-19.9	-53.4
All androcentrics	**141.2**	**32.3**	**173.4**	**-108.9**	**-77.1**
Deliberately neutral	0.0	11.2	11.2	+11.2	-
Feminine agentive	0.0	2.5	2.5	+2.5	-
TOTALS	157.0	126.6	283.6	-30.4	-19.4

Term		Normalised values	Change %
Mrs	'76	11.3	+31.9%
	'09	14.9	
Miss	'76	3.4	-100.0%
	'09	0.0	
Ms	'76	0.0	--
	'09	22.3	
0-address	'76	1.1	+3526.7%
	'09	41.0	
Generic *he*	'76	103.9	-85.7%
	'09	14.9	
Generic *man*	'76	37.3	-53.4%
	'09	17.4	
deliberately neutral	'76	0.0	--
	'09	11.2	
Feminine agentive	'76	0.0	--
	'09	2.5	

Figure 1.Normalised numbers of sex category examples and the percentage changes from 1976 to 2009 thereof.

percentage increase was with the zero address, which rose by 3526.7%, going from one occurrence in 1976 to 35 occurrences in 2009, as seen in Table 1. In addition to this, there was a shift in which examples dominated the samples. Figure 2 shows that in the 1976 sample, as much as 89.9% of examples were androcentric generics, whereas in the 2009 sample address forms were in the majority with 63%. That such changes are clearly visible, indicates that the sample sizes and the number of examples found were sufficient to make an analysis possible. While in the 1976 sample the generic *he* formed the majority of examples with 66.2%, in the 2009 sample no single term dominated. The most frequent one was zero address, which formed 34.3% of all examples. Of the examples in the entire corpus, these two terms formed 57.2%. This figure also shows how the diversification of the examples in the second sample, with seven different terms being much more evenly distributed than the five terms found in the 1976 sample. In the totals the androcentric generics still dominated, with 61.7% of all the examples being either generic *he* or generic *man*.
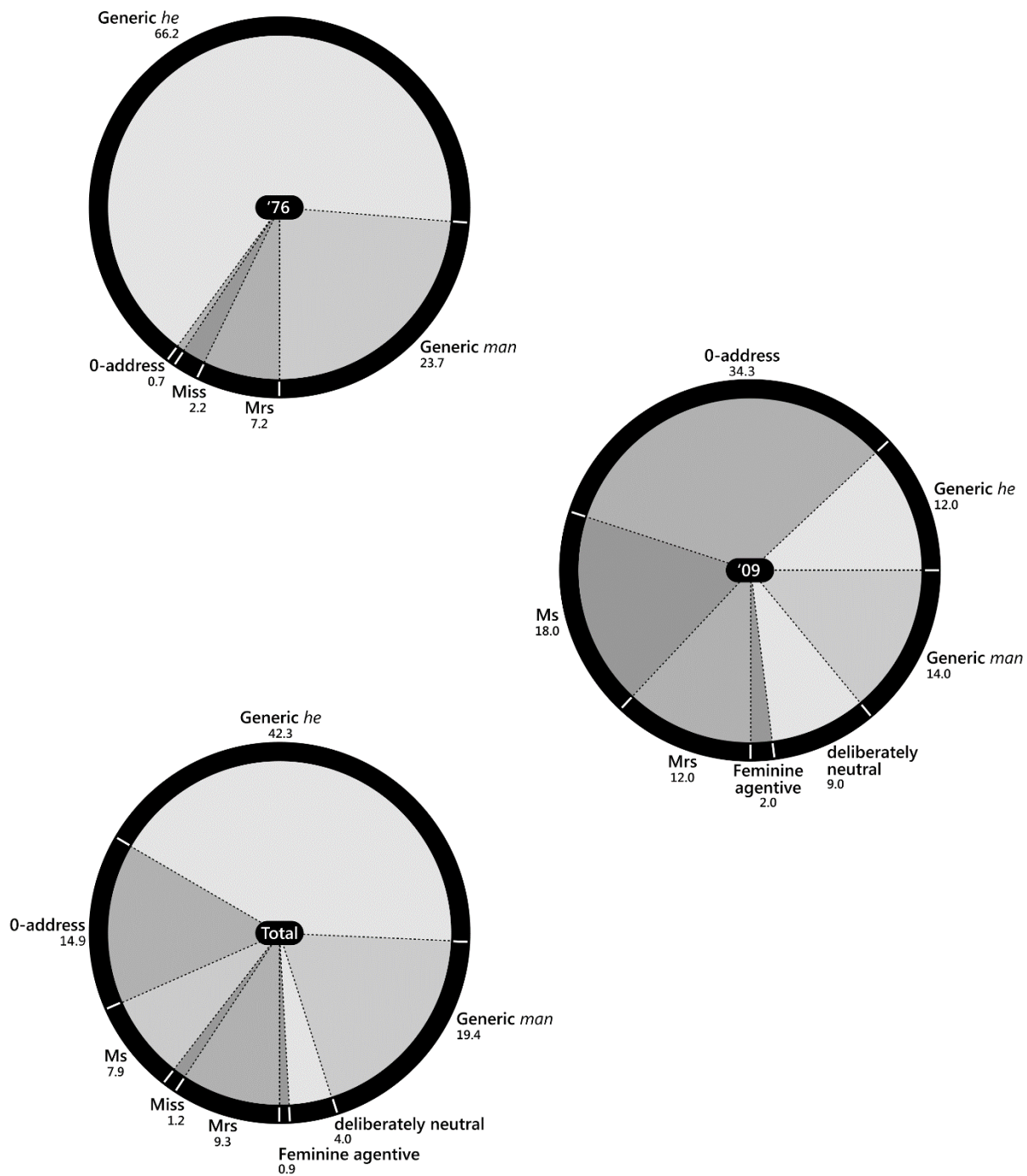
Figure 2. The percentage proportion of each type of sex related expression examples, as calculated from all examples in the 1976 sample ('76), the 2009 sample ('09) and in the material overall (Total).

5.2 *Miss/Mrs, Ms*, or none of the above?

The history of the address forms in question has been briefly outlined above. The editorials were searched for the address forms *Miss*, *Mrs.*, *Ms* and zero address. Additional address forms, such as *Dr.*, were noted separately. With this category there were fewer problematic cases, as there is considerably less room for ambiguity with address forms than there is with the generic masculines. The main issue to arise was whether to include references to foreign women. Upon examination of the editorials it became clear that in the older editorials the writers habitually used foreign language address for foreigners, for example *Mme* and *Herr*. This, albeit interesting, casts some doubt as to whether these instances can be considered truly representative of English language practices. This practice was completely absent from the 2009 editorials. Though this unbalance could conceivably skew the results, to keep the focus unequivocally on the usages native to English, address forms in other languages were not included in the final tally. Another type of address excluded from the tally were fixed formal address forms, such as *Ma'am* for The Queen. As these forms are fixed, they are of little interest in a study centring on language change.

Table 3. Absolute numbers of individual examples of address forms.

Address forms	1976	2009	Total	Change
Mrs	10	12	22	+2
Miss	3	0	3	-3
Ms	0	18	18	+18
0-address	1	35	36	+34

Term		Normalised values		Change %
Mrs	'76	11.3		+31.9%
	'09	14.9		
Miss	'76	3.4		-100.0%
	'09	0.0		
Ms	'76	0.0		--
	'09	22.3		
0-address	'76	1.1		+3945.5%
	'09		43.4	
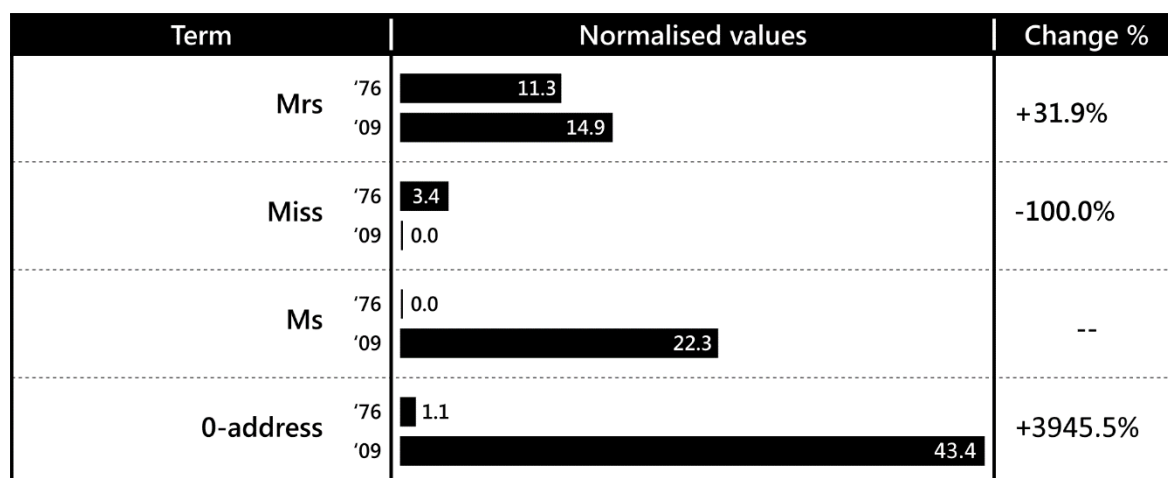
Figure 3. Normalised numbers of address form examples and the percentage changes from 1976 to 2009 thereof.

One of the greater concerns with the study was how the number of references to women would influence the numbers of different usages. Indeed, as seen in Table 3, the tally of address forms does reveal a marked increase in references to women, as address forms necessarily indicate direct references to individuals. This combined number of all address forms was 14 in the 1976 editorials with five excluded instances, whereas in the 2009 material the number was 65 with five excluded examples, the totals being shown in Table 1. The usages excluded from the 1976 sample were all references to foreign women, with one *Mme* in the editorial 170476b and several cases of referring to chairman Mao's wives with zero address. Both of these types of examples were excluded because it could not be established whether these usages followed the conventions of English or the language of the referent. The material did include both feminine and masculine foreign language address forms, suggesting that adopting the referents own culture's address forms was a common practice in the 1976 sample, but both of these are absent in the 2009 sample. In the newer set the excluded usages were mostly formal titles, with one case of *Ma'am* for the Queen, two cases of *Baroness* and one usage of *Lady,* the latter two referring to the same person. This was because being formal titles, these expressions can not be assumed to behave in the same manner as more informal address forms. In addition to these, there was one zero address in the headline of the editorial 030309a, which was dismissed on the grounds that due to the need for brevity, newspaper headlines do not necessarily follow the normal conventions of proper language. This brings the total number of included instances to 79, of which about four fifths were found in the 2009 texts.

5.2.1 *Miss* and *Mrs* against *Ms* and zero address

The absolute number of examples of address forms found in the data can be seen in Table 3. This Table shows that the numbers for both *Miss* and *Mrs* did not differ much in the two samples. *Miss* started out with only three instances in the 1976 sample, all of which were in the editorial 020476c, and was completely absent from the 2009 sample. This editorial discusses the medical ethics involved in keeping a brain dead individual on life support with specific reference to a particular patient, where the word *Miss* is used to refer to the patient in question. *Mrs* was more common than *Miss* in the older texts with ten instances, increasing slightly to twelve usages in the latter sample. *Miss* thus was found three times in the entire material, whereas *Mrs* appeared 22 times. Figure 3 shows the normalised numbers and the percentage changes seen with address forms from the 1976 sample to the 2009 sample. Here it can be seen that the instances of *Mrs* increased by 31.9%, whereas the percentage change for *Miss* is of

course a drop of 100%. Most instances of *Mrs* in the 1976 sample occurred only once per editorial, with eight editorials in total containing it. The editorial 090476a contained three examples, which refer to two different women when discussing the make up of the British government at the time. The twelve instances of *Mrs* found in the 2009 sample were divided among seven editorials, three of which only had one occurrence, while three texts had two and the editorial 020409c had three occurrences. This last editorial discusses the wives of Gordon Brown and Barack Obama. Here the women are specifically mentioned because of their position as wives of high ranking politicians. Such a context would likely provide additional motivation for using the address form *Mrs* instead of for example *Ms.* However, in order to make proper conclusions on the impact of such contexts one would need further studies with larger sample sizes and a specific focus on address forms and their contexts.

Miss, while already barely existing in the 1976 sample, is completely absent from the 2009 sample. All three admissible examples found in the entire corpus come from the editorial 020476c, where they refer to a hospital patient of whom few additional details are given. Despite this *Miss* is the second most common address form in the 1976 sample, as is seen in Table 3. It is tempting to conclude from these results that *Miss* was already on its way out in 1976 and that may indeed be the case, however it is also possible that the small number of instances was largely due to most of the women in the texts being married. This is supported by the near lack of the alternative usages in the 1976 sample, as those would have been the only option to *Miss*, *Mrs* likely not having been an acceptable replacement. Indeed the only editorial where one of the two alternatives is found, is the one with all the three instances of *Miss.* Here zero address is used the first time the woman in question is referred to, after which she is referred to as *Miss.* What is clearly visible in the Tables 1 and 2 is that the number of references to women among the examples increased considerably. The number of all address forms goes from 14 in the 1976 sample to 65 in the 2009 texts, bringing the total to 79 references which could be used as examples. In the normalised numbers this translates to an increase of 410.2%. This increase is also one contributor in the overall shift in the sex related data, as seen in Figure 2. In the 1976 sample address forms are only 10.1% of all examples, whereas in the 2009 sample the majority of examples with 63.0% are address forms. Table 3 and Figure 3 show that the same increase was channelled almost exclusively to the alternative usages *Ms* and zero address, with the former increasing by 18 instances and the latter by 34. This would suggest that both *Miss* and *Mrs* became dispreferred options over the time period in question, while *Ms* and zero address became the preferred ways of referring to women. However, if one calculates the

overall normalised counts of address forms based on marriage status from Figure 3, there is barely a change from one sample to another. For the 1976 texts this number is 14.7, while for the 2009 texts it is 14.9. Therefore the decrease of marriage based address forms is in proportion to the overall number of references to women, rather than a change in the number of occurrences.

One small caveat applies when interpreting the results however. When estimating the usage of address forms one should also consider the possibility that especially a well established public person might have expressly wished writers to use one or another. Still, that being largely impossible to ascertain either way, it remains mainly a cautionary note for the reader.

Some additional observations arise from further study of the data. The dramatic scarceness of *Miss* in the two samples becomes even more striking, when one notes that all three instances are in fact found in the same editorial, 020476c, and are in reference to the same woman, a hospital patient around whom a court case had developed. This editorial also has the only instance of zero address in the first sample, again in reference to the same woman and preceding the two cases of *Miss*. This supports the assumption that the lack of this address form in this set of texts has more to do with the status of those women who make it to the editorial page, as all other references were to women with established positions for example in politics, many of these references being to Margaret Thatcher. Most of the nine texts in the 1976 sample with usable examples only had one reference to women, with only the editorials 020476c and 090476a including four and three references respectively. In the 2009 editorials the address forms were divided more evenly across the 23 useful texts, with two editorials, 310309a and 020409c, at the lead with six references each. The examples in the older two editorials are zero address as well as *Miss* and *Mrs*, while the newer editorials contain zero address and *Ms*.

Zero address is used in the 1976 texts to some extent, but almost exclusively when referring to men, many of whom are either foreign or very famous. It was not the purpose of this study to look into how male address forms are used, so no systematic review of these was undertaken beyond ascertaining certain overall tendencies. The most salient of these was the erratic nature of references to non-British people, which lead to the exclusion of references to foreign women from the research material. On several occasions the editorials used such forms as *Herr, Monsieur* and *Señor*, however they referred to for example Russian men with a plain *Mr*. One good example is the editorial 130376b, which includes in the same sentence a reference to a German with *Herr* and immediately following that, refers to a Polish man with *Mr*. Examples

of the famous names with zero address were such familiar characters as Turner, Tito and Stalin. Such unsystematic usages were enough to cast doubt on the cases where zero address was used for a foreign woman, as it was impossible to ascertain within the scope of this study, whether those cases in the 1976 texts were reflections of English conventions or loaned language practices. No such discrepancies were evident in the 2009 set and so this reason for excluding usages was not present. There were however four address forms which used honorary titles, which are by their nature set formal titles, not given to easy linguistic changes. An example of these is the aforementioned *Ma'am* for the Queen.

The overwhelming majority of zero address usages were cases with first name and last name without an address form. Of the 34 cases only four had only the woman's last name. Two of these were references to female cricketers in the editorial 030409c and two were from the editorial 070309b, in which they referred to a woman depicted in a dramatisation. The latter editorial also contained the only professional honorific used for a woman in the entire sample, in this case *Dr*. There were two cases in the 1976 sample in which *Mrs* was used together with first name and last name, so even when a woman's full name was given, the marriage denoting title was used. Roughly half of the 21 texts in the 2009 sample containing zero address also contained other address forms. Six texts had zero address along with *Mrs,* five had it along with *Ms* and one text had all three address forms. The *Mrs* in the last of these editorials is used in reference to Sarah Palin, whereas the *Ms* refers to Angela Merkel, who while married, does not go by her current husband's last name. One might speculate that this is behind the decision to apply to her the marriage neutral address form. In general no clear pattern for the use of zero address emerged from the 2009 sample, certainly there was no implication of it being used in the same way as it was used for some very famous men in the 1976 texts. In any case it is difficult to analyse the specifics of choosing one address form or another without extensive further study. What is clear however, is that in the time period in question the ratio of address forms used has changed dramatically and that along with the greater visibility of women the more neutral address forms have markedly gained ground.

5.3   Androcentric generics

The androcentric generics consist of the masculine third person singular pronoun used to represent both male and female referents and of the word *man* used in either singular or plural form to refer to mixed sex groups. As previously pointed out, the latter can not be considered a true androcentric generic if it refers to a group which is known to consist of only males, e.g. an

all male cabinet. It is possible that the person using the phrasing would have used it even if women were present in the group referred to, but as such considerations amount to little more than guess work, these usages can not be considered genuinely androcentric. The following section will focus on these two usages as they are presented in the research material and to any findings and conclusions which might be gleaned from this.

Table 4. Absolute numbers of individual examples of androcentric generics.

Androcentrics	1976	2009	Total	Change
Generic *he*	92	12	104	-80
Generic *man*	33	14	47	-19

Term		Normalised values	Change %
Generic *he*	'76	103.9	-85.7%
	'09	14.9	
Generic *man*	'76	37.3	-53.4%
	'09	17.4	

Figure 4. Normalised numbers of androcentric generic examples and the percentage changes from 1976 to 2009 thereof.

5.3.1 The generic *he* in the research material

There is a clear drop in the numbers of androcentric generics from the 1976 data to those from 2009, as seen in Table 4. There were 104 instances of the generic *he* in the material, of which 92 were from 1976 and only 12 from 2009. In relative terms this means that there was an 85.7% drop in the number of the generic *he* from the 1976 sample to the 2009 one, as shown in Figure 4. The more numerous occurrences of the generic *he* in the 1976 sample were not evenly divided. While most texts had only one or two examples of it, two editorials, 220476a and 280476b had as many as 11 examples and the 300476b editorial had nine. Midrange was represented by one editorial with five occurrences, three editorials with six occurrences and two texts with seven occurrences. Of these, the editorial 220476a with eleven examples, the text with nine examples and the 120476b text with seven examples all used the generic *he* in reference to public officials. The first discussed the qualities required of a Prime Minister, while the second concerned legislation surrounding the ombudsman, and the third referenced a generic person in a public office. It seems likely that these usages were informed by a genuine expectation that the referent would be male, which mirrors one of the objections to the usage. The second editorial with 11 occurrences, 280476b, deals with identifying suspects in court and

here the generic refers to a witness of a crime. Interestingly the text does use the female pronoun once when rape victims are mentioned as witnesses. This is a clear indicator that unlike in the previously mentioned texts, the writer must be aware of the very real possibility of the referent being female, yet still chooses to use the androcentric generic. A similar case was found in the 020476c editorial, where the text refers first to a female hospital patient and then goes on to use the generic he when discussing hospital patients in general, as seen in example (1) below.

> (1) If the patient is in a fit state of mind, he has the right to refuse treatment, even if this may hasten his death (though the doctor's assessment of whether he is in a fit state of mind will depend on his diagnosis and his prejudices). (020476c)

In the editorials with only one generic *he* it was used as a reference to a public official in five of them, while in the other four editorials it referred to undefined individual representatives of more or less amorphous groups of people, such as the patients in long stay hospital wards in the editorial 310376b. When one looks at the distribution of these usages in the editorials containing two or three examples of the generic *he*, one finds that apart from the ones in the 100476b editorial, these usages appear either in one sentence or in the case of the 170476a editorial two almost identical sentences closely following each other. This is shown in example (2).

> (2) The Feast of the Resurrection confronts the Christian with the irreducible core of his belief, that Christ died and rose from death. That it was which the Apostles proclaimed, which excited the early Christian communities, and which has remained at the heart of Christian faith: the incarnation, death and resurrection of the Lord.
>
> The feast is also an occasion which may prompt the Christian to reflect on the nature of his belief concerning what is now to be celebrated. (170476a)

This suggests that even when there are two or three instances of this usage, these instances form in fact a single unity, instead of being genuinely individual occurrences. This being the case, the editorials where the writer systematically uses the androcentric *he* on multiple occasions appear even more unusual, and the norm appears to be to use the generic *he* relatively rarely, unsystematically and on individual occasions. It thus seems likely that already in the 1976 sample the dominating solution was to write the text in a manner which rendered a good deal of the generic *he* usage unnecessary.

A drastic contrast is offered by the 2009 sample. Here only three texts are responsible for the entirety of the 12 instances of generic *he*. Of these, the six occurrences in the 010409c editorial are similar to the references to public officials in the 1976 sample in that they refer to people commonly thought of as males, namely football players and football referees. In the two

occurrences in the 020309c editorial and the four in the 260309c editorial the case is the same as with the 1976 editorials with two or three instances. Here too the instances are either within the same sentence, or in almost identical sentences closely following one another. In both cases all the occurrences have the same referent, thus they do not stand apart as much as genuine separate instances would. If the generic *he* was rarely used systematically in the 1976 sample, in the 2009 sample it is close to non existent. In Table 4 it can be seen that the absolute number of usages dropped from 92 instances to only twelve. In 154 editorials only three used the androcentric *he* and in two of these the multiple usages can be seen as merely repetitions. The case of the footballers in the 010409c editorial is also interesting. Here the editorial satirises a group of people almost exclusively thought of as male, with the obvious intent to be discussing male players, without ever explicitly mentioning that the referents indeed are male. The examples (3) and (4) below contain all the examples found within this one editorial.

> (3) When a referee blows his whistle for a free kick or penalty he welcomes an animated debate with the transgressor about his decision. Oblige him. (010409c)

> (4) When an opponent has sustained a groin injury after a hard tackle, he will appreciate it if you assist in massaging the affected area while he waits for his team trainer to arrive on the pitch. (010409c)

While this unspoken assumption qualifies this usage as the same kind of androcentric generic as the 1976 references to a male Prime Minister, it also strongly suggests that while the androcentric *he* is in most cases unwelcome, in the few remaining instances where the culture at large sees the referent in question as almost or completely exclusively as male, the generic can still be acceptable. The 260309c editorial with four usages could also be interpreted as being in this category, since the referent here assumed to be male is the maker of cryptic crossword puzzles.

The occurrences of generic *he* were in this admittedly small sample reduced by a rather impressive 87% over the time period in question, as seen in Figure 4. The fact that three editorials alone contained roughly a third of all the occurrences from 1976 shows the usage to not have been uniformly common even in the older sample. In the 2009 sample all the examples were found in only three editorials. Judging by these results the generic *he* has experienced a rapid decline in 33 years, having been relegated to the texts of a few writers who still maintain it.

5.3.2 The generic *man* in the research material

The generic *man* also ended up with reduced 2009 numbers, but here the drop was not quite as drastic as it went from 33 in 1976 to 13 occurrences, as seen in Table 4. In the 2009 sample most examples are compounds such as *chairman, servicemen* or *statesman* and the simple generic *man* was much rarer. In addition to this, it was much harder to determine whether the usage in question was used as a true androcentric generic and thus much more case by case evaluation was needed. Some, like the above mentioned three, were easy to include. Some were easy to exclude, such as *ombudsman* in the editorial 300476c, on the basis of it being a loan word and thus not directly reflecting the development of English. As was already mentioned, the guiding principle was to include only clear examples and when in doubt, leave a particular usage out of the set in question. These unclear examples were made note of, but were not included in the final tally of the generic *man* usages. They will however be discussed below as a separate issue. In most of these unclear cases a word such as *chairman* referred to one specific man and though it is possible the writer would have used the same word for a mixed sex group, e.g. a group of former chairmen, such speculations do not amount to reliable data. In other cases the word *man* was used in a manner closely resembling an androcentric generic, but as it was unclear whether the reference was indeed exclusively to men, these were omitted as well. There were also some cases which were included due to them effectively functioning as androcentric generics, though they were outside the group of words originally intended to be under consideration. One of these was the word *manning*, which was repeatedly used when discussing workers in general in a given business or industry. As a derivative it is hardly farther removed from the original *man* than is *chairman* and its usage in the material was clear and consistent enough to merit its inclusion in the data.

The generic masculines were relatively evenly dispersed in the 1976 sample, but in the 2009 sample there was some pooling to be seen. The overall number of generic masculines in the samples was 46, of which 33 were found in the 1976 sample and 13 in the 2009 sample. The number of generic masculines ranged between one and four per text, with only one text from the 2009 sample having four occurrences. In the 1976 sample there were fifteen texts with one occurrence, six texts with two and two texts with three occurrences. In the 2009 sample however, two texts had more than half of the overall number of included examples. The texts 020309a and 280309b had three and four examples respectively, whereas all the other texts with examples in them had only one example, amounting to six occurrences altogether. In the

editorial 020309a, the writer twice used the word *manmade* to refer to global economic issues, resulting in one of the more unambiguous examples of a generic *man* found in the material. The third example from the same text was the word *man* in reference to British troops. In the editorial 280309b, the word *servicemen* was used in reference to British troops four times. Unlike the other editorial with multiple examples, this one also included the sex neutral expressions *soldiers, personnel* and even *men and women* when referring to military personnel. Examples (5) and (6) below show two of the occurrences found in this editorial.

> (5) Stress and suicide among UK servicemen is a serious issue, but also an unfashionable one. (280309b)

> (6) Service personnel who break a big toe get much the same. (280309b)

The two editorials in the 1976 sample with three instances were the texts 020376a and 060376a. Five of these six instances were straightforward usages of generic *man,* while the sixth example was the word *Ulstermen* used as a general reference to people from Ulster. Neither editorial included unambiguously sex neutral expressions, like the ones used about British troops in the above mentioned 2009 editorial.

The 1976 sample had 24 instances of a simple generic *man* being used in fifteen editorials, whereas in the 2009 sample this construction appeared only once, in the 020309a editorial. The closest to be found in the newer sample was the word *mankind,* used once in the editorial 280309a, where it was used as part of the familiar phrase *history of mankind*. Whether by chance or not, this particular word was completely missing from the older sample. Of the remaining examples, five in the 1976 editorials were the word *manning* or its derivatives *overmanning* and *under-manned*, the first two of these appearing twice and the last once. These too were missing from the new sample. The 2009 sample did however have one unambiguously generic instance of the word *manpower*. It should be pointed out here that unlike *mankind*, *manpower* did make an appearance in the 1976 sample twice, but neither of these usages were unambiguously generic so following the general guidelines set for this paper they were excluded from the total tally of masculine generics.

The remaining examples in both samples what will here be referred to as male agentives, to mirror the term female agentives. In the early sample these consisted of three cases which were *railwaymen* in the 030376b editorial, *statesman* in the 260376a editorial and a general reference to a person in the position of *chairman* in the 120476a. While the great majority of generic masculines in the 1976 sample had been usages of the simple generic *man*, in the 2009 sample

the terms denoting a profession or a position were narrowly in the majority with eight instances. Four of these were the references to *servicemen* mentioned above. Of the remaining four, three were references to *statesmanship,* while the last one was the word *policeman* used in a non gendered metaphor in the editorial 270309a.

The reduction in numbers in the generic *man* category between the two data sets is due to the almost complete disappearance of the simple generic *man.* The second subset clearly present in the 1976 sample were the word *manning* and its derivatives, which also was not to be found in the 2009 sample. The only subset which increased in the samples were the gendered words for professions or positions, but this increase does not even come close to being as dramatic as the above mentioned reduction and is in absolute numbers still relatively small, though considerable in proportion. The words *mankind* and *manpower* only appeared once each and so are not suggestive of change or lack of it. All this would suggest that the generic *man* and its derivatives are indeed clearly on their way out of polite language, even if this development may not affect all the derivatives equally.

An interesting side note to this all was revealed when the expressions excluded from the generic *man* category were examined. While the number of excluded examples in the two samples was not radically different, nineteen in 1976 and sixteen in 2009, the composition of these groups was markedly different. The examples excluded from the 1976 tally included fifteen instances of simple generic *man* or such derivatives as *manpower* and four words referred to above as male agentives. The examples excluded from the 2009 tally were however all male agentives. The usages excluded from the final numbers for this study therefore support the general trend found in the included examples, namely that while simple and easily recognised instances of generic *man* have significantly decreased in the material, male agentives did in fact increase both relatively and in absolute number. It should be noted that while the reason these words were excluded was that referring to specific individual men, they could not be strictly considered generic, they nevertheless are agentive nouns with a clearly masculine orientation in the same way as an such feminine agentives as *actress* have a feminine orientation. As such, their presence even when not used as genuine generics is not a sign of sex neutral language, but rather of a language which is still comfortable with referring to the male sex. While this abundance of male agentives was discovered while surveying generic masculines, their closest parallel would be instead the feminine agentives, which will be discussed below.

5.3.3 Neutral expressions and feminine agentives

As the focus of this part of the study was on expressions reflecting sex, sex neutral expressions were awarded a less central role. This chapter will discuss both the sex neutral expressions found in the editorials and cases where sex based language was avoided, along with any additional points of interest.

The main sex neutral expressions searched for in the material were neutral third person references and neutral agentives. These and in fact any explicitly sex neutral expressions were completely absent in the 1976 editorials, while in the 2009 editorials they made a modest appearance, totalling nine examples, as seen in Table 5. This also means that percentage change could not be calculated for these terms, as Figure 5 shows. Of these terms *he or she* was used twice, four were neutral agentives, and *men and women*, *person* and singular *they* all appeared once. The agentives were two occurrences of *police officers* and two of *salesperson*.

It should be noted here that the number of examples in either subcategory of androcentric generics alone was greater than the full number of gender neutral expressions in the entire material, as seen when comparing tables 4 and 5. The greatest difficulty in mapping sex neutrality in language in a work such as this is that one possible solution to the question of sex based language is rewording the text so as to avoid the pitfalls of biased language altogether. This solution is very effective in making the entire issue invisible and it is very difficult to attempt to assess where such measures have been taken. To accurately make such judgements one would presumably need to study the production of a specific writer over time and in different contexts to see whether a certain writing environment makes a difference in their use of language. Barring that possibility one is left with mainly idle speculation and such considerations have thus been excluded from this work. One exception to this might be editorials which due to their subject matter would lend themselves easily for the use of androcentric generics, but which for some reason or another lack them. There were three such candidates in the material, the completely neutral editorials 100376b and 260376c, and the editorial 020309c which despite of ample opportunity remained sex neutral until the second to last paragraph. The 1976 editorials deal with the laws about the age of homosexual consent and obscenity respectively, whereas the 2009 editorial is a light hearted piece about technology talking back at its user. The first two articles very judiciously refer to *people* whom the laws might effect, without explicit sex references. The last article does suggest an avoidance of androcentric generics mainly because of the one usage at the end, where a driver of a car is

referred to as a *he*, whereas lavatory goers and appliance owners had so far simply been *visitors* and *users*. These three editorials suggest that rewording is indeed taking place, but the full extent of this would be very difficult to ascertain. Examples (7) and (8) below from the editorials 260376c and 020309c respectively show both a pluralised neutral expression, in this case *people*, and a combination of a generic *he* and a second person pronoun used to directly address the reader.

> (7) The expert evidence called has usually been to the effect that certain kinds of people are helped psychologically and sexually, by reading obscene literature or seeing obscene films. (260376c)

> (8) Next time you're cursing a meandering motorist for the way he darted across your lane without signalling, expect to be challenged by your in-car computer ("His driving? What about the way you cut up that blue Mondeo at the lights?").

> As for those hectoring laptops, word-processing spellcheck programs are one thing. But what about a laptop that sniggers if you type a cliché? That would really take the biscuit, wouldn't it? (020309c)

The main conclusion to be drawn from the points here presented, is that while androcentric generics have indeed become less frequent in the material, the main method for working around them has not been explicitly sex neutral language, but rather writing the text so that the issue of sex based language is avoided in the first place. This is supported by the very small number of explicitly sex neutral expressions in the 2009 set together with the decreased number of androcentric generics in the editorials.

language change.

Table 5. Absolute numbers of individual examples of neutral expressions and feminine agentives

Neutral/Feminine agentive	1976	2009	Total	Change
deliberately neutral	0	9	9	+9
feminine agentive	0	2	2	+2

Term	Normalised values		Change %
deliberately neutral	'76	0.0	--
	'09	11.2	
Feminine agentive	'76	0.0	--
	'09	2.5	

Figure 5. Normalised numbers of neutral expression and feminine agentive examples and the percentage changes from 1976 to 2009 thereof.

Beyond neutral expressions, some additional points of interest arose from the material. One of these was the use of sex based metaphorical references to nations and peoples. There were altogether eight usages which could be interpreted as such personifications, of which three were masculine constructions using *brother* or *he*, and five were feminine. The five feminine personifications were instances where countries were referred to with the third person feminine pronoun or its inflected form. All of these personifications were from the 1976 editorials, no such expressions were used in the 2009 material.

The final two points of interest found in the texts were one direct reference to issues surrounding sex and language and one masculine generic which was in fact not of textual origin at all. In a discussion on the cohabitation rule the editorial 050376c references a commission's wish to replace the phrase "cohabiting as man and wife" with "living together as husband and wife". The editorial then further comments on this, after calling it coy, saying that it is based "on the doubtful grounds that the Anglo-Saxon is less pejorative than the Latinate". This suggest not only that already in 1976 there was some concern even in official circles about the issues of more neutral language, but also it strengthens the impression that at the time, *The Times* was not yet particularly swayed by these concerns.

The other point of interest comes up in the editorial 110309c, which discusses pedestrian traffic lights. In it we find the phrase *green man* in quotes. Naturally this phrase fell outside of the criteria set for the examples for this study, in that it is purely a descriptive phrase of an icon, since most green pedestrian signals carry a silhouette which is indeed easily recognised as male. However that symbol itself is definitely an androcentric generic, albeit a visual one and as such it presents an interesting parallel to the linguistic androcentric generics which have been discussed above.

6 Faction-based categories in the research material

The basic concept of faction categories was elaborated on above, with discussion on the coining of this umbrella term, as well as the subcategories it here encompasses. This section of the work explores how these categories were realised in the corpus, and what differences could be found between the two text samples.

6.1 Quantitative overview of faction categories

The faction terminology was examined the same way as the sex terminology. The texts were searched for admissible examples, which then were tallied up and analysed for any emerging patterns. While the sex category examples were only divided into a few different types of expressions, the faction examples were first divided into cultural categories, and then further broken down to individual expressions.

A slim majority of the texts in the 1976 sample contained useful examples, both according to word count and the number of editorials. In the 2009 sample this was reversed, with a clear majority of texts having no suitable examples. The absolute number of useful texts was higher in the 2009 sample, but the word count of these texts was lower in the new sample compared to the older. Accordingly, the amount of texts with no useful examples increased, this both according to the word count and the text count. The absolute number of all included examples dropped considerably over time, going from 337 in 1976 to 181 in 2009. In normalised numbers this would translate to 380.6 examples in 1976 and 224.7 in 2009, with a drop of 41.0% during the period in question. As will be detailed in the following chapters and as is clearly visible in Tables 6 and 7, this is reflected in the individual categories into which the examples were divided. Despite the drop, the overall number of examples in the faction categories remained higher than the number of examples in the sex category. The category of religion was the only one in which the number of examples increased during the study period, all other categories experienced a decline. Even there, the increase was only 7.3%, as is visible in Figure 6. If the percentage changes in the categories are viewed as neutral figures and the direction of these changes is disregarded for a moment, three groups emerge as the percentages appear to cluster together. As Table 7 and Figure 6 show, the categories religion and nationality changed the least, the percentage change for both hovering close to five percent. The category of ethnicity along with the group *other* both have change percentages close to 50. The three remaining categories; nationhood, geographical origin and race, all change 80% or more, with the 80,0%

of the category of geographical origin being the smallest change in these three. It is clear that while most categories diminished in the samples, the change was not even across the board.

Table 6. Overview of the examples in the faction category, absolute numbers

Faction	1976	2009	Total	Change
nationality	105	91	196	-14
nationhood	38	4	42	-34
ethnicity	54	26	80	-28
geographical origin	22	4	26	-18
religion	42	41	83	-1
race	48	3	51	-45
other	27	12	39	-15
TOTALS	336	181	517	-155

Table 7. Overview of the examples in the faction category, normalised numbers

Faction	Normalised 1976	Normalised 2009	Normalised total	Normalised change	Change %
nationality	118.6	112.9	231.5	-5.6	-4.8
nationhood	42.9	5.0	47.9	-38.0	-88.4
ethnicity	61.0	32.3	93.3	-28.7	-47.1
geographical origin	24.8	5.0	29.8	-19.9	-80.0
religion	47.4	50.9	98.3	+3.5	+7.3
race	54.2	3.7	57.9	-50.5	-93.1
other	30.5	14.9	45.4	-15.6	-51.2
TOTALS	380.6	224.7	605.3	-156.0	-41.0

Term		Normalised values	Change %
nationality	'76	118.6	-4.8%
	'09	112.9	
nationhood	'76	42.9	-88.4%
	'09	5.0	
ethnicity	'76	61.0	-47.1%
	'09	32.3	
geographical origin	'76	24.8	-80.0%
	'09	5.0	
religion	'76	47.4	+7.3%
	'09	50.9	
race	'76	54.2	-93.1%
	'09	3.7	
other	'76	30.5	-51.1%
	'09	14.9	

Figure 6. Normalised numbers of faction category examples and the percentage changes from 1976 to 2009 thereof.

Figure 7 shows the percentage of each category from the total number of examples found in the 1976 sample, the 2009 sample and in the overall tally of the examples. Of all the examples found in the samples, 38.3% were in the nationality category, the second most common category being religion with 16.2%. This was followed closely by the category of ethnicity with 15.4% of the overall number of examples. The rest of the categories all had less than 10% of examples, beginning with the 9.6% for the category of race and ending with the 4.9% for the geographical origin category. The trends shown in Tables 6 and 7 are supported by Figure 7. Only two categories grew in proportion to the overall number, those of nationality and religion. The former has 31.2% of all examples in the 1976 sample and religion has 12.5%. In the 2009 sample the nationality category has shot up to 50.3% of the examples found, with religion being the second largest category with 22.7% of examples. The categories of ethnicity and other stay relatively stable. Ethnicity moves from 16.0% of the 1976 examples to 14.4% in 2009, and the category of other falls from 8.0% to 6.6%. The strongest change shown by the visualisation is a marked relative decline in the categories of nationhood, geographical origin and race. The category of nationhood falls from 11.3% in the
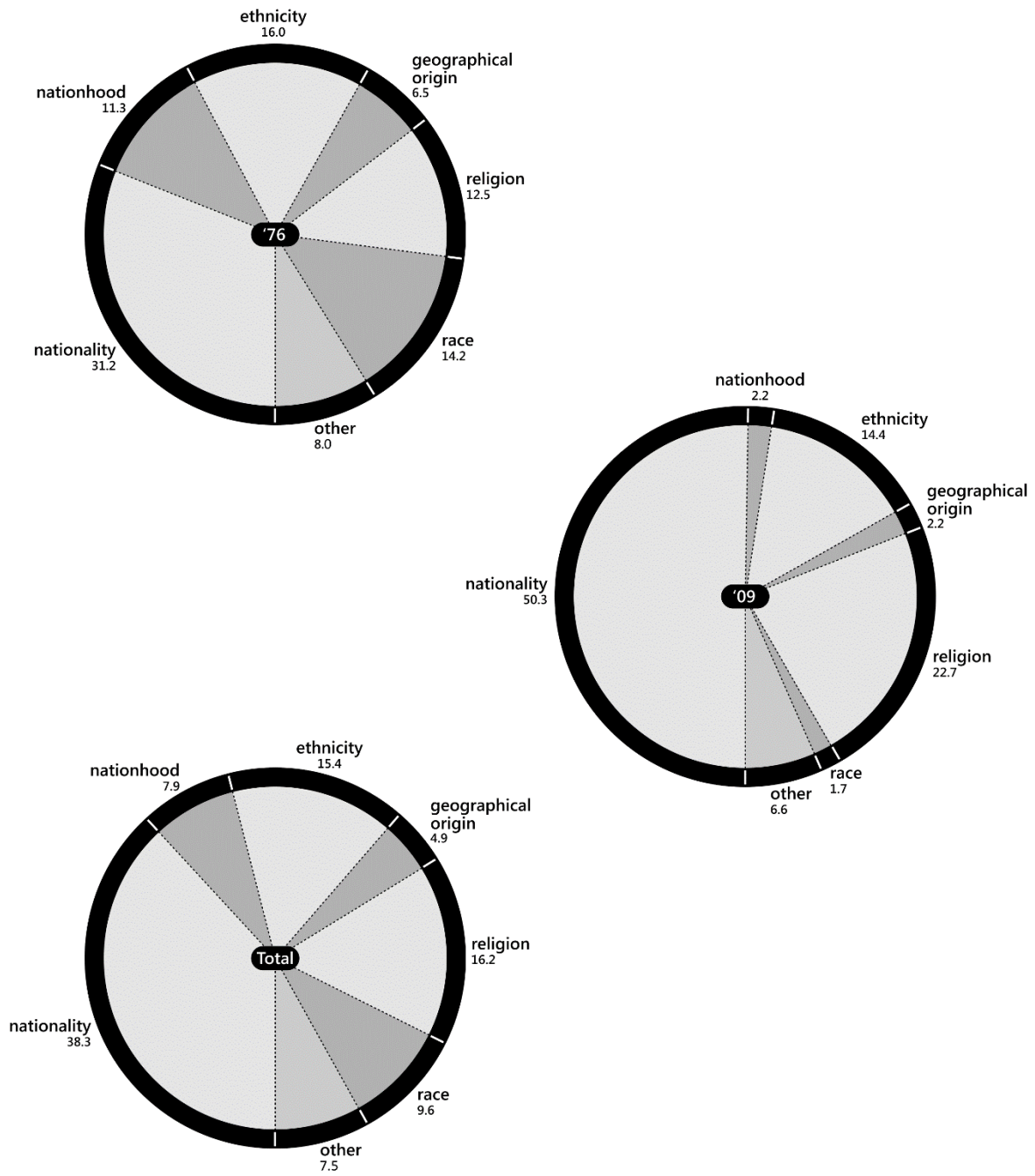
Figure 7. The percentage proportion of each type of faction examples, as calculated from all examples, in the 1976 sample ('76), the 2009 sample ('09) and in the material overall (Total).

older sample to 2.2% in the newer one, while the category of geographical origin falls from 6.5% to 2.2% of all the examples included. The greatest relative decline is seen with the category of race, which has 14.2% of the examples found in the 1976 sample, but in the 2009 sample only 1.7% of examples are in this category. Taken together with the changes in the number of examples in each category, this would suggest that some categories have indeed become less acceptable over time than others. All these changes will be further explored in the following chapters, where each category will be examined in turn.

6.2 Nationality

The first category here examined will be that of nationality. Nationality within the context of this study will refer to the identifying cultural the category of belonging to an official nation. It is here distinguished from ethnicity by its reference to an official political entity, whereas ethnicity here will be used refer to belonging to a group not immediately tied to such an entity. Therefore *British* is classified as nationality since the defining feature is the state of United Kingdom, whereas *English* is classified as an ethnicity since it is not tied to an official country nation. This division is adopted in this study for purely practical reasons and is not intended to be a definite analysis of either concept. Treating nationality and ethnicity apart has the advantage of keeping the sizes of the categories manageable and it also differentiates between population groups which are roughly equivalent to nation states and those that either have no state to call their own at all, or reside within a state which does not correspond to population boundaries. While these terms clearly appear to behave differently in the material, it is very much acknowledged here that this division is not absolute. In many cases a term would have to be estimated on a case by case basis. For example in certain passages *Greek* may refer to the Greek nationals living within the state of Greece, while in another context it may mean people of Greek ancestral or cultural origin who reside outside the borders of the said state. The former case would in this study be designated a marker of nationality, while the latter would be included in the ethnicity category. Different grammatical forms of a given nationality, ethnicity or other such expression were grouped together as varied expressions of the same term. Thus for example *Spanish* was counted together with such variations as *the Spanish, Spaniards* and *the Spanish People*, where the context indicated that the variations all referred to the same concept.

There were altogether 196 admissible examples of nationality markers in the research material. The overview Tables 6 and 7 show the amounts and change of all nationality markers over the

period in question. Of these 105 were found in the 1976 sample and 91 in the 2009 editorials. The normalised numbers were 118.6 in the first sample and 112.9 in the second one. There is approximately 4.8% drop in the number of examples in the nationality category between the 1976 and 2009 editorials, which indicates that the frequency of nationality markers within the texts has dropped slightly over time. However this drop is hardly big enough to draw any solid conclusions, as it could well be coincidental. The indication here is that the overall use of nationality markers has remained the same or experienced a slight decrease. As the Tables 6 and 7 show, in both samples nationality was the most numerous category by a generous margin. It was followed by 54 examples in the ethnicity category in the 1976 editorials and in the 2009 texts by the 41 examples in the religion category. However, in proportion to the rest of the examples found in the text, nationality markers become much more common as one moves from the 1976 sample to the 2009 sample. As seen in Figure 6, they form 31.2% of the total number of 1976 examples, but in the 2009 texts 50.3% of all the examples are in this category, with the increase being 19.1%.

In the older sample the nationality markers were found in 39 editorials, through which they were relatively evenly dispersed. The great majority of these texts had three or fewer examples, with only eight texts reaching above that. The editorial 020376b contained as many as fourteen examples. The topic of this editorial are the communist parties across Europe and their apparent efforts to distance themselves from the Soviet communist party. As such the text lists unusually many nationalities for one editorial, including for example *French* and *the French*, *Italian, British, Spanish, Yugoslavs* and *Poles.* Of these *French* and *Italian* were repeated twice and thrice, respectively. Two examples in the text illustrate well the technical difficulties posed by the scanned editorials. In addition to *Italians*, the word *Italiani* occurs once, but it is likely to be a technical error rather than a genuine usage. Same is true for the word *Czcchs*, where lower case *e* has clearly morphed into a *c*. The coexistence of *Czechs* and *Czechoslovak* in very similar contexts shows that variants of individual nationality terms were employed side by side. It is of course possible for *Czech* to refer to an ethnic group, but since the entire editorial refers only to nationalities and only utilises nationality terms, it is a reasonable assumption that *Czech* follows the same pattern. Example 9) below illustrates these multiple usages of different nationality terms as well as the two forms referring to Czechoslovakians. As noted earlier, the inaccuracies resulting from the use of OCR software have been edited out of the example passage.

(9) If the Italians and the French as well as the Yugoslavs can establish their right to choose their own form of communism there is no ideological reason why the Czechs, the Poles, the East Germans and the rest should not do the same.

There have long been close contacts between Italian communists and Czechoslovak communists who participated in the reformist regime of 1968 which was crushed by Russian tanks. (020376b)

The editorial 130376b had the second highest number of examples with seven nationality terms, all variations of either *German* or *Poles/Polish*. This text, which dwells on the post World War II settlement agreements between West Germany and Poland, unsurprisingly has variations of *German* and *Polish* as its nationality terms. The most interesting feature in this editorial however, is the interplay between *German* as a nationality term and *German* as a denoter of ethnicity. The phrase *ethnic German* is used three times to differentiate those of German origin living in Poland, from the Germans residing within the state of Germany. The difficulty posed by terms of ethnicity is in fact explicitly remarked on in this editorial in the bracketed sentence "an ethnic German is not always easily definable", though it must be noted here that the clear assumption is that such definitions are possible.

The texts 040376b, 180376b and 300376b had six examples each, with six separate terms between them. The most used term in these texts was *Italian* with eight occurrences in the editorial 180376b, followed by the four occurrences of *Rhodesian* found in the editorial 040376b and the four examples of *Italian* in the editorial 300376b. This illustrates a tendency found in the material, of example clusters containing multiple repetitions of individual terms. The eight texts with four or more examples contributed close to half of all the nationality markers in the 1976 sample, containing 52 usages, while the other 31 editorials contained 53 examples between them. The examples in the newer sample were more evenly spread. Out of 50 texts with examples in this category, 46 had three or fewer nationality markers. Of the remaining texts 140409c and 230409c had four examples, 160409b contained six and 300309a had eight examples. Here too repetition was common, as these clusters contained only seven distinct nationality terms altogether. This is manifested very well in the editorial 160409b, where the word *Afghan* accounted for all the six nationality examples. The editorial 300309a with the most occurrences touched upon the military situation in Iraq, with a short summary of events thus far. The different nationalities mentioned are *Iraqi, British,* and *US.* The latter occurs in the phrase *US civilian*, where it is taken to refer to the nationality of the civilian, rather than the US as a political entity. This text also contains the rarer form of *British,* the word

Briton. The examples (10) and (11) below illustrate the difference between these editorials, one of which repeats a single nationality term, and another which employs several of them.

> (10) the attempt to portray institutional misogyny as the heritage of patriotic Afghans. [...] Some 87 per cent of Afghan women are illiterate. [...] One in three women experiences physical, psychological or sexual violence and every 30 minutes an Afghan woman dies during childbirth. (160409b)

> (11) In six years of fighting, training and building, 179 Britons have lost their lives in Iraq. […] Jay Garner, a retired general and the first senior US civilian into Iraq after the invasion, arrived with a skeleton staff at war with itself and Washington over how and when to hand the country back to its own people. His British counterpart, Sir Hilary Synnott, was sent to Basra with half an A4 side of written instructions and orders to "play it by ear". […] It is now one for the Iraqis to answer. (300309a)

Both text samples had roughly the same number of different nationality terms, with 22 separate nationality markers in the 1976 texts and 25 in the 2009 texts. The entire material contained 38 different faction markers in the nationality category. The overwhelming majority of these terms, 29 markers, were present in only one editorial. 18 of these were only encountered once in the entire data. As seen in Table 8, these represented a geographically diverse group, as they included among others *Canadian*, *Thai*, *Irish* and *Sudanese*. Such low numbers can in no way be interpreted as clear indications of change and the appearances of individual nationalities in the material very likely reflect the topics of the day at the time of writing rather than any trends with language. In addition to this, some nationality terms were only present in one of the sets due to the rearranging of political entities, such as the unification of East and West Germany. The most common nationality marker in both sets was *British* and its permutations. Apart from *British* and *the British*, these included *Britishness, British-born* and *Britons.* The last three of these were only used in the 2009 texts, with explicit reference to British nationals. Of these *Britishness* occurred once, in the editorial 120309b, referring to the substance of a series of speeches delivered by the Prime Minister. Likewise *British-born* only occurs once in the entire corpus, in the editorial 130409b, which concerned a failed Asean (Association of South East Asian Nations) summit in Thailand and its consequences. Here the word is used for the Prime Minister Abhisit Vejjajiva, pointing out his personal background outside Thailand. Unlike these two terms, *Briton* was used seven times in the entire corpus, twice in the editorial 070309b and once per text in five editorials.

The absolute numbers of individual nationality category items and the change therein are shown in Table 8, while the relevant normalised figures, along with the change percentage between

the samples, are presented in appendix 3. The uneven presence of different nationality markers in the corpus is clearly visible in Table 8. As shown in Table 8, in the 1976 editorials there were 22 cases of the nationality marker *British*, while the 2009 sample contained 19 of them. The second most common term in both samples was *American,* which appeared 14 times in the first sample and 16 times in the second. In the 2009 texts this was accompanied by five instances of *US*, which was nevertheless counted as a separate term. Had these been merged as one marker, this would have passed *British* as the most common nationality term by two occurrences. It can not be ruled out that the intra category variation of specific nationality markers show a change in language use. However, the fact that the majority of the nationality markers were present only in one sample, makes the changes in the subject matter of the editorials, brought on by changes in policy orientation and current issues, also a plausible explanation for the intra category variation. This in connection with the very low frequencies in the corpus of many of the nationality markers, clearly evident in Table 8, suggests that attention should rather be afforded to general changes in overall frequencies. As it is, no such tendencies emerged in the detailed dissection of the nationality data and thus any conclusions must be drawn based on the nationality markers as a group rather than as individual usages. It has already been noted that this suggests that the overall frequency of terms in the nationality category in the language has remained approximately the same over time, while the proportion of nationality markers out of the entire set of faction markers grew notably, as is seen in Figure 6. The relative stability of this category as a whole is further supported by the fact that *British* and its variations were about as common in both samples, while the frequencies of other nationality terms varied greatly in the samples without there apparently being a pattern related to the structure of the language itself, as seen in Figures 8 and 9.

Table 8. Absolute numbers of individual examples of nationality terms.

Nationality	1976	2009	Total	Change
British	22	19	41	-3
Irish	0	1	1	+1
German	4	1	5	-3
East Germans	1	0	1	-1
French	10	4	14	-6
Italian	13	4	17	-9
Spanish	6	1	7	-5
Portuguese	1	0	1	-1
Poles	4	0	4	-4

Nationality	1976	2009	Total	Change
Yugoslavs	2	0	2	-2
Czechoslovak	2	0	2	-2
Russian	4	2	6	-2
Soviet	2	0	2	-2
Turkish	1	0	1	-1
Israeli	0	4	4	+4
Libyan	2	0	2	-2
Iraqi	1	4	5	+3
Iranian	1	2	3	+1
Kuwaiti	0	1	1	+1
Afghan	0	8	8	+8
Sudanese	0	1	1	+1
Somali	0	1	1	+1
Rwandan	0	1	1	+1
Eritrean	1	0	1	-1
Zimbabwean	0	1	1	+1
Mozambican	1	0	1	-1
Rhodesian	11	0	11	-11
Pakistani	0	7	7	+7
Indian	0	3	3	+3
Sri Lankan	0	1	1	+1
Chinese	1	0	1	-1
Thai	0	1	1	+1
Canadian	0	1	1	+1
US	0	5	5	+5
American	14	16	30	+2
Mexican	0	1	1	+1
Cuban	1	0	1	-1
Venezuelan	0	1	1	+1

Term		Normalised values	Change %
British	'76	24.8	-5.1%
	'09	23.6	
Irish	'76	0.0	--
	'09	1.2	
German	'76	4.5	-72.5%
	'09	1.2	
East German	'76	1.1	−100.0%
	'09	0.0	
French	'76	11.3	-56.0%
	'09	5.0	
Italian	'76	14.7	-66.2%
	'09	5.0	
Spanish	'76	6.8	-85.2%
	'09	1.0	
Portuguese	'76	1.1	-100.0%
	'09	0.0	
Poles	'76	4.5	-100.0%
	'09	0.0	
Yugoslavs	'76	2.3	-100.0%
	'09	0.0	
Czechoslovak	'76	2.3	-100.0%
	'09	0.0	
Russian	'76	4.5	-45.1%
	'09	2.5	
Soviet	'76	2.3	-100.0%
	'09	0.0	
Turkish	'76	1.1	-100.0%
	'09	0.0	
Israeli	'76	0.0	--
	'09	5.0	
Libyan	'76	2.3	-100.0%
	'09	0.0	
Iraqi	'76	1.1	+339.6%
	'09	5.0	
Iranian	'76	1.1	+119.8%
	'09	2.5	
Kuwaiti	'76	0.0	--
	'09	1.2	
Afghan	'76	0.0	--
	'09	9.9	
Sudanese	'76	0.0	--
	'09	1.2	
Somali	'76	0.0	--
	'09	1.2	

Figure 8. Normalised numbers of nationality term examples and the percentage changes from 1976 to 2009 thereof, part A.

Term		Normalised values	Change %
Rwandan	'76	0.0	--
	'09	1.2	
Eritrean	'76	1.1	-100.0%
	'09	0.0	
Zimbabwean	'76	0.0	--
	'09	1.2	
Mozambican	'76	1.1	-100.0%
	'09	0.0	
Rhodesian	'76	12.4	-100.0%
	'09	0.0	
Pakistani	'76	0.0	--
	'09	8.7	
Indian	'76	0.0	--
	'09	3.7	
Sri Lankan	'76	0.0	--
	'09	1.2	
Chinese	'76	1.1	-100.0%
	'09	0.0	
Thai	'76	0.0	--
	'09	1.2	
Canadian	'76	0.0	--
	'09	1.2	
US	'76	0.0	--
	'09	6.2	
American	'76	15.8	+25.6%
	'09	19.9	
Mexican	'76	0.0	--
	'09	1.2	
Cuban	'76	1.1	-100.0%
	'09	0.0	
Venezuelan	'76	0.0	--
	'09	1.2	

Figure 9. Normalised numbers of nationality term examples and the percentage changes from 1976 to 2009 thereof, part B.

As both Table 8 and Figures 8 and 9 show, two of the four terms with ten or more occurrences in the 1976 sample, *Italian* and *French*, dropped dramatically in frequency, while the other two, *British* and *American* either had only a slight drop or a small increase in their frequency. The number of separate nationality terms is approximately the same in both samples, with 22 in the 1976 sample and 25 in the 2009 sample. The terms *Rhodesian* and *Afghan* peaked noticeably in their respective samples, with the former only present in the 1976 editorials and the latter only in the 2009 texts. Going by Table 8, the overall impression is one of small changes in individual terms and thus a relative stability. However, upon examining the normalised figures

along with change percentages, as presented in Figures 8 and 9, it becomes apparent that the relative changes in the frequencies of individual markers fluctuate wildly both in normalised quantities and change percentages. Interestingly enough, after omitting the percentages that could not be calculated due to the start point being zero occurrences, only one nationality term is found to have a frequency change of less than 25%. With change of only -5.1% *British* is the most stable nationality marker found in the corpus. The second most stable one is *American* with a change of +25.6%. It deserves to be noted here that the variation within the term *British* itself is of some interest, however that issue falls outside of the scope of this study, if only for the fact that the data available is not sufficient to base any conclusions on. Suffice it to say, that while the older editorials only had *British* and *the British*, in the newer set these were supplemented by the above mentioned variations, which apparently had become acceptable usages during the time span in question.

6.3   Nationhood

The second faction category covered here was not something originally anticipated, rather it emerged from the data spontaneously, as certain usages were clearly noteworthy, but did not fit in to the original structure intended for the data. Originally all references to political entities were meant to be excluded, both because the focus of this study was on references to people, not artificial entities, and because they would mostly consist of official names for those entities, which are often more dependent on shifts in international and intra national politics than the language itself. There was however one type of usage which was included due to its peculiar linguistic nature. This was the practice of referring to a country by a word which denotes a group of its citizens. For example, in a passage describing the official undertakings of the French state, the said state would be referred to as *the French* instead of using *France* or *the French government*. Thus a faction term denoting nationality and explicitly referring to people would be utilised when discussing an official political body. This category was designated as *nationhood*, to separate it from the nationality category discussed above in chapter 6.2, and to express the notion that it effectively evokes the concept of nation by equating it with its citizen body. It was possible to differentiate these terms from those in the nationality category by context, as these particular expressions would not have made sense as references to groups of individuals, as is illustrated below by the examples (1) and (2). The term as it is used here should be kept apart from its usages in different fields, as it is here employed as a technical term to describe only this specific linguistic phenomenon.

There were 38 examples of the nationhood category in the 1976 samples. As seen in Table 6, this category saw the second highest drop in numbers, with only four occurrences in the newer texts, while retaining its position among the other categories in both samples and in the overall numbers. Table 7 shows the equivalent normalised figures to be 42.9 occurrences in 1976 and 5.0 in 2009, with the normalised drop of 38.0 occurrences translating to the percentage change of -88.4%. The category of nationhood declined also in proportion to other faction markers. As seen in Figure 6, it formed 11.2% of all faction markers found in the 1976 texts, but only 2.2% of all faction markers found in the 2009 sample. This decline of 9.1% is second only to the drop found in the relative proportion of race markers.

Nationhood terms were found in 14 of the 1976 editorials, while all four instances of nationhood terms in the 2009 sample came from the editorial 270409b. The text concerns nuclear disarmament and contains two instances of *the Russians* and two instances of *the Americans.* Example (12) below shows one sample sentence containing both of these terms.

> (12) If, for example, both sides cut their totals to 1,500 each, verification becomes more important, especially for the Russians, who know that the Americans could rebuild their arsenals more quickly. (270409b)

In the 1976 sample five editorials contained one or two instances, five editorials contained three occurrences, three editorials contained four instances and one editorial, 160376b, had five instances. Example (13) below is an extract from this editorial, showing the expressions *the Russians*, which occurred four times in the text, and *the Israelis*, which occurred once.

> (13) True, the Israelis failed to appreciate the significance of Arab military movements leading up to the October War in 1973, but this was one of the lessons of the October War which the West is not likely to forget so soon. The Russians, too, will have made worst-case assumptions, no doubt endowing the West with capabilities that would astonish the most optimistic general at Shape. (160376b)

As shown in Table 9, the most common term was *the Russians*, which alone accounted for 22 instances in the older sample and for both of the two instances in the newer one. On the second place was *the Americans* with four occurrences in 1976 and two in 2009, with *the Chinese* the third most common nationhood term, present four times in the 1976 sample, but absent from the 2009 sample. *The Russians* and *the Americans* were the only two nationhood terms present in the 2009 editorials, the other seven nationhood terms found in the other 1976 were absent. While the prevalence of *the Russians* in the 1976 material could be due to a specific set phrase being used as a short hand in place of *the Soviet Union*, this would not explain away the other occurrences of this category.

Table 9. Absolute numbers of individual examples of nationhood terms.

Nationhood	1976	2009	Total	Change
the British	1	0	1	-1
the French	2	0	2	-2
the Russians	22	2	24	-20
the Israelis	1	0	1	-1
the Syrians	3	0	3	-3
the South Africans	1	0	1	-1
the Chinese	4	0	4	-4
the Americans	4	2	6	-2
the Cubans	1	0	1	-1

Term		Normalised values	Change %
the British	'76	1.1	-100.0%
	'09	0.0	
the French	'76	2.3	-100.0%
	'09	0.0	
the Russians	'76	24.8	-90.0%
	'09	2.5	
the Israelis	'76	1.1	-100.0%
	'09	0.0	
the Syrians	'76	3.4	-100.0%
	'09	0.0	
the South Africans	'76	1.1	-100.0%
	'09	0.0	
the Chinese	'76	4.5	-100.0%
	'09	0.0	
the Americans	'76	4.5	-45.1%
	'09	2.5	
the Cubans	'76	1.1	-100.0%
	'09	0.0	
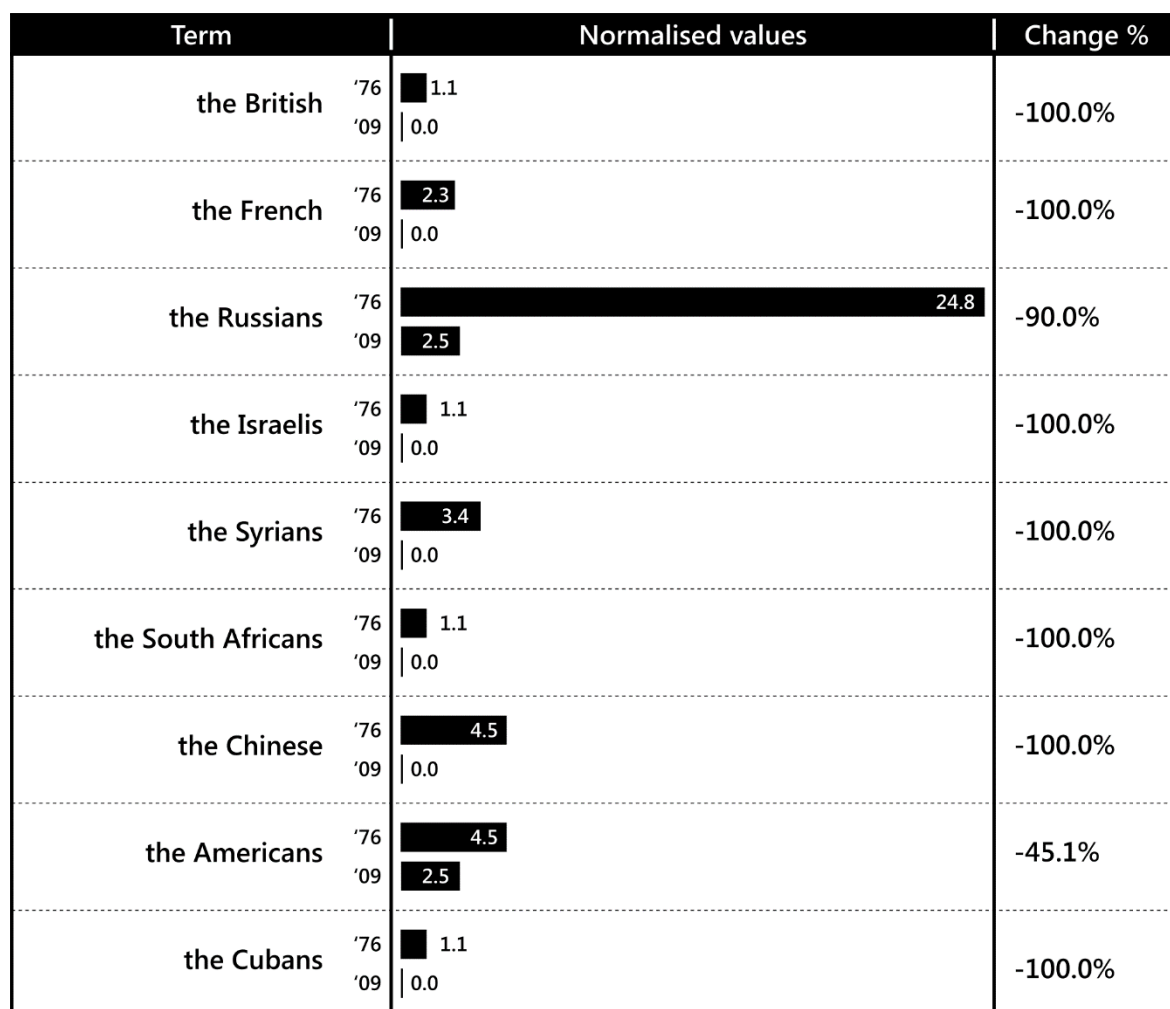
Figure 10. Normalised numbers of nationhood term examples and the percentage changes from 1976 to 2009 thereof.

It is quite clear from the material that the nationhood category has to a great extent experienced the same as the category of race. The dramatic decrease in frequency from one sample to another can not be interpreted as anything but a genuine change in language. This becomes especially clear when Figure 10 is considered, where the relative decline of separate nationhood markers is clearly visible. The smallest percentage change in this category was -45.1%, found with the term *the Americans.* While four of the eight terms were only present once in the entire sample, and thus the possibility of coincidence should not be downplayed, the fact that all the terms in this category were on a downward trend strongly suggests a wider change taking place. For one reason or another the practice of referring to an official nation with a term denoting its citizenry has in this particular sample of texts become as rare as using terms in the race or geographical origin categories. Whether these changes are related is impossible to ascertain in this study, as here only general shifts are well displayed. It does not appear far fetched to suggest such a relationship however, indeed the presence of a very clear downward trend across the different categories in addition to the shifts observed in the way the examples divide into different categories, supports the idea of a more general ongoing process.

6.4　Ethnicity

The term ethnicity is here used to refer to population categories defined by membership in cultural groups, as opposed to the defining feature being for example an official political entity, a faith group or geographical location. The political entity nation may and typically does consist of several ethnic groups and religions may similarly contain numerous ethnicities in their flocks. At times it is difficult to distinguish which defining feature a given term references, but in most cases this is clear either from the term itself or from the context. This definition too is largely adopted here for practical reasons pertaining to the thesis at hand and is not intended as a universal definition of the term. With some usages it was impossible to determine whether they referred to an ethnic group or another category, as at times the text was in no way clear as to what a given usage referred to, and any additional background information was either lacking or contradictory. It is possible that at the time of writing these terms were considered clear, but as it was beyond the means of this work to explore all the usages thoroughly, any unclear cases were left out of the examples, or if clearly faction terms, were included in the group of other or undetermined terms. Some of the terms included in the ethnicity group as examples are two word expressions, consisting of the word *ethnic* and a designation of origin, such as *ethnic*

German. The different plural forms of a specific term were analysed together as variations of a single expression.

There were 80 examples fitting the ethnicity category in the entire data, making this the third most common category overall, as shown in Table 6. It was also the most common category in the 2009 sample, with 26 examples, but in the 1976 sample ethnicity was the second most common category with 54 examples. With the normalised frequencies, the category of ethnicity went from 61.0 occurrences in 1976 to 32.3 in 2009, the percentage change being -47.1%. This was the third smallest change among the categories, with only religion and nationality changing less, as shown in Table 7. The percentage of ethnicity terms out all examples found in each sample stays relatively stable, with the drop of 1.6 from 16.0% in the 1976 sample to the 14.4% in the 2009 one being the smallest change among the categories, as seen in Figure 7. The number of individual terms in this group remained very stable, with twelve in the older sample and eleven in the new one. The overall number of individual terms in the material was 20, of which the overwhelming majority of seventeen terms were present in only one text. Ten of these terms appeared only once in the entire material, as Table 10 shows. Here too, the numbers concerning the distribution of individual terms are too small to draw conclusions on changes in the usage of these individual terms. It would be premature to assume that a change from one occurrence to two occurrences signals a change in language use, and does not for example reflect random fluctuations. When analysed as a group of terms, they do however yield some insights.

The three most common terms are all in the 1976 sample. These are, along with their variations, *Arab* which appears 20 times, *Kurd* which is found twelve times and *Palestinian*, which comes up on eight occasions. The first of these examples is present in the 2009 sample only once, in the editorial 170309b, while the second most popular one completely absent in

Table 10. Absolute numbers of individual examples of ethnicity terms.

Ethnicity	1976	2009	Total	Change
English	3	4	7	+1
Scottish	1	0	1	-1
Irish	0	1	1	+1
ethnic German	3	0	3	-3
Gallic	0	1	1	+1
Basque	2	0	2	-2
Slav	1	0	1	-1
Kurd	12	0	12	-12
Arab	20	1	21	-19
Palestinian	8	3	11	-5
ethnic African	0	1	1	+1
Afrikaner	1	0	1	-1
Xhosa	1	0	1	-1
Hutu	0	2	2	+2
Bengali	1	0	1	-1
Khmers	1	0	1	-1
Gurkha	0	6	6	+6
Pashtun	0	2	2	+2
Tamils	0	4	4	+4
Hazaras	0	1	1	+1

the 2009 editorials. The expression in the 2009 sample which comes closest to these numbers is *Gurkha*, appearing six times in the editorial 300409c. The example closest in meaning to the most popular expression in the nationality category, *English*, is only seen three times in the earlier sample and four times in the later one. *Kurds* of course refers to an ethnic group without recognised political or geographical entities and as such can not be replaced with terms from other categories. This however is not the case with *Arab*, which refers to people who form dominant or at least influential populations in several different countries, and therefore can be replaced with e.g. individual nationality terms. There is some evidence to suggest that nationality terms have replaced *Arab* to some extent, since the nationality terms *Kuwaiti, Iraqi* and *Afghan* all either appear for the first time or experience a rise in frequency

Term		Normalised values		Change %
English	'76	3.4		+46.5%
	'09	5.0		
Scottish	'76	1.1		-100.0%
	'09	0.0		
Irish	'76	0.0		--
	'09	1.2		
ethnic German	'76	3.4		-100.0%
	'09	0.0		
Gallic	'76	0.0		--
	'09	1.2		
Basque	'76	2.3		-100.0%
	'09	0.0		
Slav	'76	1.1		-100.0%
	'09	0.0		
Kurd	'76	13.6		-100.0%
	'09	0.0		
Arab	'76	22.6		-94.5%
	'09	1.2		
Palestinian	'76	9.0		-58.8%
	'09	3.7		
ethnic African	'76	0.0		--
	'09	1.2		
Afrikaner	'76	1.1		-100.0%
	'09	0.0		
Xhosa	'76	1.1		-100.0%
	'09	0.0		
Hutu	'76	0.0		--
	'09	2.5		
Bengali	'76	1.1		-100.0%
	'09	0.0		
Khmers	'76	1.1		-100.0%
	'09	0.0		
Gurkha	'76	0.0		--
	'09	7.4		
Pashtun	'76	0.0		--
	'09	2.5		
Tamils	'76	0.0		--
	'09	5.0		
Hazaras	'76	0.0		--
	'09	1.2		

Figure 11. Normalised numbers of ethnicity term examples and the percentage changes from 1976 to 2009 thereof.

in the later editorials, as can be seen in Figures 8 and 9. Again however, the numbers are hardly definitive enough to be fully conclusive. The single editorial with the most instances of *Arab,* 310376c, is one dealing with the violent situation in and around Israel. This editorial specifically makes a point of referring to Arabs as an ethnic group consisting of several religions, while referring to *Jews* as the other party. Whether the latter term is meant to refer to the ethnicity or the religion remains unclear. Such a blanket reference with the word *Arab* seems almost unthinkable in the 2009 editorials and indeed there is only one example of this term found in that entire sample. The nationality term, *Israeli*, however appears in the latter sample four times, while being completely absent in the earlier sample. These occurrences are in the editorials 030309a and 040309b, two in each text. Example (14) below from the editorial 310176c shows the word *Arab* as an explicitly ethnic category, while example (15) from the editorial 170309b contains both the single occurrence of *Arab* in the 2009 sample, as one of the occurrences of *ethnic* combined with another term, as detailed below. This latter example also contains the only occurrence in the material of the term *ethnic African*.

> (14) Arabs of all religious and political persuasions have now joined in the protest and they have carried with them the mayors of Arab towns. Something like a fixed confrontation between Israel authority and Arab resolution has thus developed of a kind that may soon expand to a point beyond police action. (310176c)

> (15) Over the past six years at least 300,000 people have been killed and 2.7 million left homeless by marauding gangs and government troops fighting ethnic African rebels in Darfur. […] So far he has had some success. Arab governments have rallied to him, including even Saudi Arabia, which is reluctant to see Iranian influence grow in Sudan after Tehran's lead in denouncing the West. (170309b)

There are two compound terms in the ethnicity category, e*thnic German* and the *ethnic African*. The previous were all in the editorial 130376b, which concerned a treaty regarding the aftermath of the Second World War. As mentioned above, the editorial itself refers to the problematic nature of the concept of ethnicity with the above quoted phrase "an ethnic German is not always easily definable", while maintaining the assumption that such a definition is possible. The term *Ethnic African* appears in the editorial 170309b, as seen in example (2) above, where it is possibly used to distinguish the said group of people from ethnic Arabs, as the single occurrence of *Arab* in the 2009 sample is here in reference to governments of other countries in the region. However, the editorial does not explicitly state which group is the term *ethnic African* is being contrasted with, so its precise meaning remains unclear. These two terms with their explicit reference to ethnicity are something of an anomaly in the material.

There is no question that the number of ethnicity markers dropped considerably over time in the texts in question. This drop was however not even across the examples. As seen in Table 9, only three terms formed the majority of ethnicity examples in the 1976 sample. *Arab*, *Kurd* and *Palestinian* appeared altogether 40 times out of a total of 54 examples, whereas the number of the three most common terms in the 2009 sample was only 14 out of 26 ethnicity terms in total. In addition to the change in numbers, none of the top three terms remained the same. The most common ethnicity markers in the later editorials were *Gurkha*, *Tamil* and *English,* with six, four and four instances respectively. The occurrences of *English* increased by 46.5%, while the other two terms were not found in the 1976 sample at all, therefore a change percentage could not be calculated for them, as seen in Figure 11. In fact all ethnicity terms with a significant presence in the 1976 editorials went through a drop in numbers when compared to the 2009 sample, and most terms were only present in one sample. This is reflected in the percentage changes shown in Figure 11, where seventeen of the twenty terms show either a non-calculable result, or a drop of 100%.

This suggests two things. First that ethnicity markers were not overly common to begin with, as they were not evenly spread across different ethnicities but rather heavily weighed on only a few, and in addition to this tended to cluster in certain editorials. Half of the twenty occurrences of the most common term *Arab* were in the editorial 310376c with yet further five in the editorial 080376b. Of the second most common term *Kurd,* all the examples were found in a single editorial 080376b and the eight occurrences of the third most frequent word *Palestinian* were in two editorials, 15037a and 140476b. Thus 35 ethnicity terms out of 54 were found in five editorials, while the entire sample from that year contained 118 editorials.

Second, it suggests that the change in numbers is largely due to these few examples being replaced with other terms. It has already been suggested here that some of them may have been replaced with nationality markers. This is partially supported by the relative increase in nationality markers shown in Figure 6, though the relative stability of ethnicity markers seen in the same figure means this is by no means a solid conclusion. In order to obtain a truly clear picture of the change in ethnicity terminology one would likely need to expand the material so as to provide more examples and perhaps also extend the research period further back in time. The latter might lend some insight as to whether the ethnicity markers have always been scarce or whether their presence in the language has been diminishing for a longer stretch of time. As things are, it appears that within the period in question the ethnicity category has decreased in

use overall, with the extent of the change being about halfway between that seen in categories which nearly disappeared and those which changed very little. The fact that the portion of ethnicity markers stayed comparatively stable across both samples suggests that terms in this category have replaced terms from the categories with the most dramatic decreases, but this also can not be taken as a solid conclusion. Any such definitive conclusion would demand a much closer analysis of the respective categories and usages, something which is beyond the scope of this study.

6.5 Geographical origin

The geographical origin category denotes faction terms based on the geographical spread of a given population segment, as opposed to denoting a cultural group, or a group grounded in a political entity. The geographical area referred to by a term in this category could be well defined such as a continent, or a more imprecise one, such as a certain portion of the hemisphere. Potential examples would be *European* or *Eastern*. These could be further defined by narrowing down the geographical reference, thus forming expressions such as *East African*. Here too the different permutations of a term are calculated as examples of one expression and the category name itself is adopted for practical reasons pertaining to this study.

This category was the least numerous one in the 1976 sample, and in the 2009 editorials it shared second to last place with the nationhood category. As seen in Table 6, the earlier editorials contained 22 unambiguous examples, where the term in question clearly referred to a geographical entity rather than for example ethnicity or race. In the new editorials there were only four such terms to be found. This category also had the least examples overall, with only 26 found in both sets of texts, while the drop in number from one sample to another was approximately mid range in absolute numbers, with 18 fewer examples in the new set, but third highest percentage wise, the change being -80.0%, as seen in Figure 5. In the 1976 sample, geographical origin is the category with the least examples, whereas in the 2009 sample it holds the second smallest percentage from the total together with nationhood, as shown in Figure 6.

This faction category is small not only in the number of examples, but also in the number of different terms that emerged from the data. Only the category race had fewer different terms overall, with three separate expressions for both categories. Table 11 shows the four terms found in this category, which were *African, European, western* and *Asian.* No other clear faction terms referring to the geographical category were present in any of the editorials.

Table 11. Absolute numbers of individual examples of geographical origin terms.

Geographical origin	1976	2009	Total	Change
European	10	3	13	-7
Western	2	0	2	-2
African	10	0	10	-10
Asian	0	1	1	+1

Term		Normalised values	Change %
European	'76	11.3	-67.0%
	'09	3.7	
Western	'76	2.3	-100.0%
	'09	0.0	
African	'76	11.3	-100.0%
	'09	0.0	
Asian	'76	0.0	--
	'09	1.2	

Figure 12. Normalised numbers of geographical origin term examples and the percentage changes from 1976 to 2009 thereof.

For example, *Asian* did emerge one other time in the material, in the editorial 020409b, but in a context where it was juxtaposed with *black* and *white,* and thus clearly was used to signify a race category, rather than a geographical one. Of these four expressions, *European* and *Asian* were found in the 2009 sample, both in only one editorial. *European* occurred three times in the editorial 020309a and the single instance of *Asian* was in the editorial 230409b. The least common term denoting geographical origin overall was *Asian*, with its single occurrence in the editorials. The ten instances each of *African* and *European* in the 1976 sample were spread roughly equally, the former found in six and the latter seven texts. There was no overlap between these texts, as no editorial had both of these terms in it and only in one text did one of them appear more than once or twice, with three repetitions of *European* in the editorial 160376a. Example (16) below from the editorial 240376c shows two of the instances of *African* found in the 1976 sample, while example (17) shows how *European* was used in the editorial 020309a.

> (16) He went along with the development of the homelands under the policy of separate development simply because it was the only way he or any other leading African could do any good to, or in any way influence, their people. […] But the

Portuguese empire has collapsed, and Africans in arms have taken over
Mozambique and Angola and are knocking on Rhodesia's gates. (240376c)

(17) And although these were floated in Berlin last week, and in Brussels yesterday,
and endorsed by Britain's European partners, there is enough resentment,
protectionist sentiment and mistrust among them still to make April's meeting
tricky. […] This also muddies the second issue: the call for America to rebuild its
frayed relations with its European allies and Nato. […] He will have little time for
European hectoring on the Middle East, Iran and repairing relations with Russia
(issues he has made his own priorities) if he cannot count on Europe to bear some
of the cost and offer supporting initiatives. (020309a)

Due to the diminutive stature of this category by every meaningful measure, there is less
information available on it than on many of the other categories. It begins small and it ends very
small, suggesting strongly that these expressions were not among popular usages to begin with.
With the third highest drop in frequency together with the starting numbers being so small
already, the category ends up one of the smallest ones. As shown in Figure 12, the only term
present in both samples, *European,* declines by 67.0%. The percentage differences shown in
Figure 12 are not in any way unusual compared to those in other faction categories, if anything
this category reflects a general trend visible throughout the study. The same is true of the drop
seen in Figure 5, where the change in this category follows the same trajectory as the other
categories except for nationality and religion. As this category is limited in size however, even
the four examples found in the latter sample should be taken with a pinch of salt, since three of
them appear in a single editorial and thus could be expressive of a slightly anomalous personal
style of the writer. Having said that, any editorial must surely fall within the newspaper's general
guidelines, so at the very least it signifies that such expressions were not at the time considered
unacceptable. It is also possible that the scarceness of terms in this category has to do with the
relative impreciseness of them, though other categories certainly also contain items which are
difficult to pin down. This category is one of the cases in which further study of samples from
different years would shed some light on whether these terms have always been scarce in this
particular variety of English, or whether they indeed have diminished in popularity over time.

6.6 Religion

The religion category refers to faction terms which denote the membership in a particular
religion. The only exception to this were the words *heathen* and *pagan*, which were nevertheless
included as they clearly are perceived religious groupings, even if they are not actual religions.
This was mostly a straightforward category, apart from some terms which can denote either a

religious or an ethnic grouping, such as *Jewish*. Only terms explicitly referring to known religions were included in this category, in ambiguous cases the category denoted by the term was determined from the context when possible. If this was not possible, but it was clear that the term was indeed a faction term, it was classified as belonging to the category of other. If the term was not clearly one denoting a faction category, it was left out of the examples. Clearly separate terms such as *Protestant* or *Christian* were not merged, for even if the groups they denote may overlap, the terms themselves are distinct. For the same reason *Roman Catholic* and *Catholic* were treated as separate terms, as were *Muslim* and *Islamic.*

The religion category was the only one to unequivocally maintain its numbers from one sample to another, with 42 examples in the 1976 sample and 41 in the 2009 one, as can be seen in Table 6. It was the second most common category overall after nationality, with slightly less than half as many examples. In the 1976 sample religion is the fourth most common category, but as Table 7 shows, due to the clear drops in the race and ethnicity categories in the 2009 sample, religion ends up as the second most common category in the 2009 texts. Figure 5 shows that religion is one of the two categories which become more common, the other being nationality. In the 1976 sample the proportion of religion terms out of all the examples is the fourth highest with 12.5%. In the 2009 sample the same figure is 22.7%, second only to the category of nationality. This is shown in Figure 6. The exceptional character of this category becomes even clearer in Figure 5, where the percentage changes calculated from the normalised values show it to be the only category to grow in size. This growth in normalised numbers of examples is not very big in itself, only a difference of 3.5, but the relative change percentage is +7.3%. While this is the second smallest change in any of the faction categories, the direction of the change does stand out. Though the change in the number of examples is small enough to suggest that this rise could be due to a coincidence, the difference between this category and almost all the others shown in Figure 5 is stark enough to point to a larger trend. Only the categories religion and nationality show any signs of retaining their frequency, all the other categories drop clearly in frequency, indicating that the phenomenon causing the latter change left these two categories untouched. This points to the expressions in the five other faction categories becoming less commonly accepted in the newer sample than in the older one, while nationality and religion terms have not gone through such a process. The fact that these two categories retain their sizes also means that the change has not been due to a simple case of resolving the issue by rendering all faction categories less visible, but that there has been an uneven shift in the use of faction terms in different categories. This is especially well illustrated in Figure 5 by

the categories of religion and race being next to each other. The first sample contains slightly more examples in the race category than either sample does in the religion category, but in the 2009 sample race terms are rarer than terms belonging to any other category, whereas religion terms have become slightly more frequent. The rise seen in Figure 6 in the percentage of religion and nationality category terms of the total number of examples further supports this. There are two possible explanations for this. Either there was a drop in faction term frequencies which left the categories of nationality and religion untouched, affecting only the other categories. The other possibility is that such a change did influence all categories, but that nationality and religion were compensated for their losses by language users opting to use terms from these categories instead of the now less accepted terms. Whether this is a case of one or the other can not be satisfactorily established within the confines of this study. The data here presented does however indicate that a linguistic shift has taken place over the thirty years in question.

Table 12. Absolute numbers of individual examples of religion terms.

Religion	1976	2009	Total	Change
Christian	17	4	21	-13
Protestant	1	2	3	+1
Anglican	0	2	2	+2
Roman Catholic	0	2	2	+2
Catholic	0	7	7	+7
Muslim	19	10	29	-9
Islamic	1	1	2	0
Sunni	0	3	3	+3
Shia	0	7	7	+7
Hindu	1	3	4	+2
Buddhist	1	0	1	-1
heathen	1	0	1	-1
pagan	1	0	1	-1

The 2009 sample contained ten distinct religion terms, while the 1976 sample had eight. Table 12 above shows the absolute numbers of the terms found in the sample texts. There were thirteen different terms altogether, of which five were present only in the 2009 sample and three only in the 1976 one. The ones only present in the 1976 sample were *heathen, pagan* and *Buddhist*, and all of these only appeared once. The ones only seen in the 2009 sample all appeared more than once, with two of them, *Shia* and *Catholic* appearing as often as seven

Term		Normalised values		Change %
Christian	'76	19.2		-74.1%
	'09	5.0		
Protestant	'76	1.1		+119.8%
	'09	2.5		
Anglican	'76	0.0		--
	'09	2.5		
Roman Catholic	'76	0.0		--
	'09	2.5		
Catholic	'76	0.0		--
	'09	8.7		
Muslim	'76	21.5		-42.2%
	'09	12.4		
Islamic	'76	1.1		+9.9%
	'09	1.2		
Sunni	'76	0.0		--
	'09	3.7		
Shia	'76	0.0		--
	'09	8.7		
Hindu	'76	1.1		+229.7%
	'09	3.7		
Buddhist	'76	1.1		-100.0%
	'09	0.0		
heathen	'76	1.1		-100.0%
	'09	0.0		
pagan	'76	1.1		-100.0%
	'09	0.0		

Figure 13. Normalised numbers of religion term examples and the percentage changes from 1976 to 2009 thereof.

times. It should be noted here that *Roman Catholic* also appeared only in the later sample, where it surfaced twice, bringing the total for *Catholic* and *Roman Catholic* to nine occurrences. The term which went through the largest change was *Christian,* which appeared a full seventeen times in the 1976 editorials, only to be seen as few as four times in the 2009 texts. The term with the second greatest change in absolute number of occurrences was *Muslim,* which in the

1976 editorials was the most numerous religion term with 19 usages. This dropped to ten in the 2009 set. Some simple calculations reveal something behind these changes however. If one takes all the terms in both sets denoting first Muslims and then Christians and adds them together, the results show a further detail. In the 1976 sample the terms *Muslim* and *Islamic* add up to 20 instances, while in the 2009 sample those two terms combined with the terms *Sunni* and *Shia* add up to 21 examples. The same is true for the terms referring to the Christian religions, with *Christian* and *Protestant* in the first editorial set adding up to 18 instances. In the second set of texts those two terms together with *Roman Catholic*, *Catholic,* and *Anglican* add up to 17 instances. This shows there has not so much been a move away from direct references to these religions, as the combined number of terms denoting the larger religion remained similar, but rather towards a more careful designation of different groups within the larger umbrella terms. There appears to be a clear preference in the 2009 material to expand terminology to more precisely denote specific subgroups rather than linguistically treating these religions as single groups. This is also shown in Figure 13, where the relative frequencies of individual terms are provided, along with change percentages. There we see *Christian* and *Muslim* decreasing most, with *Christian* falling 74.1% and *Muslim* 42.2%. If the calculation detailed above is performed with the normalised numbers of the same terms, the different terms referring to Christian faiths total 20.3 instances in 1976 and 21.1 in the 2009 sample. For the words denoting Muslims the difference is larger, with 22.6 occurrences in 1976 and 26.1 occurrences in 2009. These values give change percentages of +3.8% for the terms in the Christian group and +15.4% for those in the Muslim group. While the growth turns out to be greater for the latter group when normalised figures are considered, that does not disprove the shift suggested here. If anything, it serves as further support, since the growth in the terms denoting Muslim subgroups are responsible for this increase in the normalised numbers.

Linguistically such a shift both makes the individual subgroups of religions more visible and serves to separate them more clearly from the other denominations within the larger group. It is unlikely such a shift would ever be visible in the other categories, as by their very nature the categories of nationality, ethnicity and race do not splinter into subgroups quite as easily as the category of religion does. In fact ethnicity or race can often function as subgroups of nationality and therefore any shift similar to this one would likely be the above suggested inter category shift, rather an intracategory one. In fact there is further evidence that such a shift has taken place. The ethnicity category term *Arab* moves from 20 instances in 1976 to only one in 2009, without a corresponding increase in a similar term in that category. This supports the suggestion

that a part of that reduction was taken up by terms in religion category. This might also partly explain the stability of the religion category compared to other categories except nationality.

The 2009 examples are spread across 17 different editorials, whereas the equivalent number in the older sample is about half of that. Of the nine 1976 editorials, three texts, the editorials 10376a, 290376a and 200476a, contain the great majority of the 42 examples with 30 usages between them. The usages in the later sample were distributed in a significantly different manner, with only two editorials, 120309a and 300309a reaching five instances per editorial. The five examples in the former editorial are all occurrences of *Muslim,* whereas the second editorial contained two examples of *Sunni* and three examples of *Shia.* The heavy variation in the older texts could suggest writer specific differences with regard to terms in the religion category, as a few editorials seem to stand out so clearly from the rest with respect to the number of religion terms utilised.

The religion category contained four terms which fell outside the Christian-Muslim divide. These were *Hindu*, *Buddhist, heathen* and *pagan,* the latter three being the only terms present only in the 1976 sample. *Hindu* appeared in both samples, once in the 1976 editorials and three times in the 2009 texts. In both samples the word only occurs in a single text however, once in the editorial 200476a and three times in the editorial 210309c. *Hindu* was the only one of these terms to grow in frequency, with the change in the normalised numbers being +2.6 occurrences, translating to an increase of 229.7%. The editorial 200476a also contains the single occurrences of the words *pagan* and *heathen.* The editorial discusses the dimensions and development of the Islamic world and also includes not only the eight occurrences of *Muslim*, which is the largest number of occurrences of this term found in any of the editorials in either sample, but also more different religion terms than any other editorial collected for this study. Example (18) below from the editorial 200476a shows not only the context of the words *heathen* and *pagan*, but also shows a single sentence with two occurrences of *Muslim*.

> (18) But missionaries did follow in their wake, confident that they could compete with the established religions and bring salvation to heathen peoples. [...] The creeds of many who were once called pagan are fast disappearing because the old creeds are ill-suited to the world they are moving into. […] Muslim rebels in the southern Philippines have been getting arms and money from neighbouring Sabah, where Muslims rule, and as far afield as the redoubtable President Gaddafi. (200476a)

Two editorials, 290376a and 120309a, contained five instances of *Muslim*, the rest having three or fewer occurrences. The second largest number of individual occurrences of a specific term

was found in the editorial 150376a, where the word *Christian* was seen seven times. The editorial 200476a contained six out of the eight distinct religion terms found in the 1976 sample while six out of the total of nine 1976 texts containing religion category terms only provided one example in the religion category. In the 2009 sample none of the seventeen texts with extracted examples contained more than three separate religion category words each. Thus the editorial 200476a stands out of the material as the text with both the most distinct religion terms and as the one with most religion category occurrences overall, with thirteen religion category examples. The editorial with the next biggest number of religion terms was 150376a with ten extracted examples. The editorials reaching closest to this in the 2009 sample were 120309a and 300309a with five examples in this category each. The conclusion that follows from these details is that the relative increase in religion terms over the period in question was accompanied by a more even distribution of religion category examples both across distinct terms and through the editorials. Example (19) below, from the editorial 300309a, demonstrates the use of *Sunni* and *Shia*, and the accompanying use of two nationality terms.

> (19) Its soldiers became the footsoldiers of the Sunni insurgency. […] In the Sunni Triangle and on the streets of Basra, coalition troops died in their hundreds; Iraqi civilians in their tens of thousands. […] For the British it came in September 2007, when a secret and still unacknowledged deal with Shia militias allowed the remaining British troops in Basra Palace to withdraw to the airport without a shot being fired. (300309a)

It is clear from the material that the religion category has not diminished in the same way as most other categories. Furthermore the data shows fragmentation within the category, where expansive overall terms have been replaced by more specific subcategories, more carefully delineating one sub-faction from another. This brings increased visibility to the subcategories, while detracting from the overall terms, as instead of expressing the unity of Christendom or Islamic peoples, the texts bring out smaller, often mutually antagonistic groupings, within the larger religious communities. That the usages in this category have increased, while those in other categories have not, also suggests that some terms in other categories may have been replaced by terms in the religion category. Such shifts within and between faction categories are becoming clearly visible only after reviewing the religion category and at this point seem a very encouraging sign that the faction category approach chosen for this study has indeed proven fruitful.

6.7 Race

The most contested of the faction categories here presented is the category of race. While such categories as nationality or ethnicity tend to be questioned mostly on their division or detailed definitions, the very validity of the category of race has been forcefully questioned. Nevertheless, for the purposes studying the way faction categories are expressed in language, it remains a meaningful concept insofar as the category is expressed in the lexis. The faction category race is here taken to refer to the traditional concept of races which approximately corresponds to populations on different continents. It is not used to mean the smaller population groups, which to a great extent would overlap with modern concepts of e.g. ethnicity or nationality. Thus Jewish would not be considered to be part of the cultural category of race, but either one of ethnicity or religion. This somewhat cautious approach is adopted in order to make it possible to use the same framework of faction categories in both parts of the corpus. Here again the decision is a practical one, as the careful mapping of these categories would be a significant undertaking in its own right and thus well beyond the scope of this study. The goal is to ascertain whether, even with these rather streamlined faction categories, a shift in lexis can be observed to have taken place during the period in question. Therefore the massive labour effort needed for the detailed semantic mapping of the faction categories and the terms therein can be left out, and the study becomes feasible.

Table 13. Absolute numbers of individual examples of race terms.

Race	1976	2009	Total	Change
white	18	1	19	-17
black	30	1	31	-29
Asian	0	1	1	+1

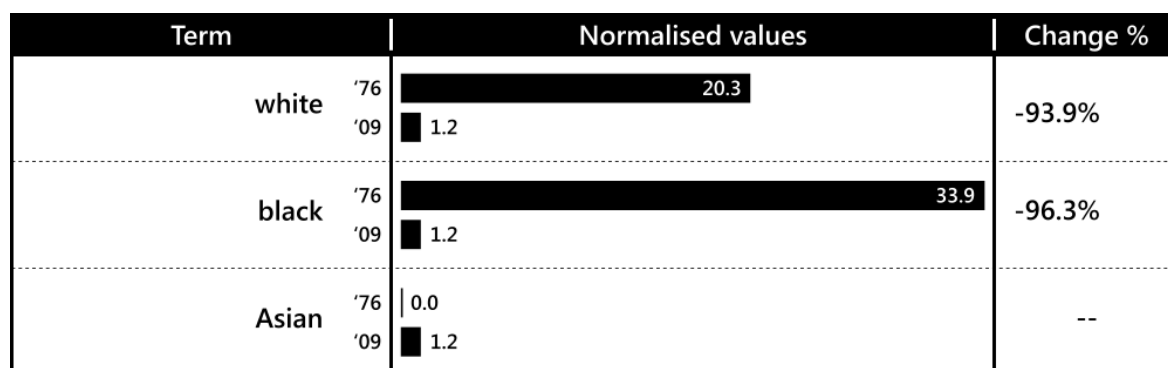| Term		Normalised values	Change %
white	'76	20.3	-93.9%
	'09	1.2	
black	'76	33.9	-96.3%
	'09	1.2	
Asian	'76	0.0	--
	'09	1.2	

Figure 14.Normalised numbers of nationality term examples and the percentage changes from 1976 to 2009 thereof.

The category of race saw a greater change in frequency than any of the other categories studied, as seen in Table 6 and Figure 5. It went from the third most common category both in absolute and normalised numbers in the older editorials to the last place in the newer editorials. With the difference of 45 instances, the drop from 48 to only three occurrences was also the greatest in absolute numbers. Figure 5 shows the normalised frequency of race category terms went from 54.2 to 3.7 occurrences, giving the highest change percentage in the faction categories, -93.1%. In overall numbers it remained the fourth most common category in both absolute and normalised numbers. A precipitous fall for this category is also visible in Figure 6, where it initially contains the third highest portion of examples in the earlier sample with its 14.2% of the total, but in the later sample holds the last place with only 1.7% of examples. This decline too is larger than with any other category. The number of individual terms in this category was however small in both samples. In the 1976 editorials there were only two separate terms present and in the 2009 editorials there were three terms as shown in Table 13. The terms found in the 1976 sample were *black* with 31 occurrences and *white* with 18 occurrences. These same terms were present in the newer text set, but only once each, and were accompanied by one usable occurrence of *Asian*. The latter could also be a term from the geographical origin category, but was used in parallel with *black* and thus it was reasonable to interpret it as referring to a category of race. In fact all the race category terms found in the 2009 sample came from the single editorial 020409b, which focused on adoption and children in care. In this text these words were specifically used in reference to the policies of child care officials to place children with adoptive parents who "reflect their own background". Example (20) below from this editorial shows that not only were the terms from a single editorial, they were from a single sentence.

> (20) It is quite wrong that black and Asian children wait on average three times longer than white children for an adoptive family, partly because of concerns that children should go to parents who reflect their own background. (020409b)

The percentage changes in individual terms and their relative frequencies can be seen in Figure 14. The changes in individual terms surpass the overall change in the race category slightly, with *white* being reduced by 93.9% and *black* by 96.3%.

In the 1976 sample the race terms were distributed across seven editorials. Most of these editorials had both terms in them, only the editorial 040376b contained *white* but not *black*. *Black* in particular seemed to emerge in clusters. In all the editorials in which it was used it appeared from four to nine times. *White* on the other hand was slightly more evenly distributed.

Three editorials, 200376a, 190476a and 280476a, contained 14 of the usages, while the remaining four editorials each had one occurrence. The example (21) below, extracted from the editorial 240376c where *black* occurred nine times, shows both the race terms in the 1976 sample.

> (21) given that Pretoria refused to create continuous entities of the scattered tribal reserves and then blandly offered to these archipelagos of African indigence in oceans of white affluence, full "independence". […] In such circumstances, the continual demand by the black lobbies abroad that South Africa must be delivered as a single entity (as "Azania") to majority black rule, has begun to look much more compelling. (240376c)

As with the geographical origin category, this group was diminutive in stature compared to the others, but the category of race did stand out as the one with the most dramatic drop in frequency. The terms *African* and *European* have quite a clear area of overlap between *black* and *white*, however since both of these pairs were clearly reduced in frequency over time, there is little real possibility of a shift in usages from one category to another. It is clear that this category in particular has gone from being relatively commonly used to an almost non existent one, even if the usages were in a relatively small number of editorials. It still would signal the acceptability of such terms in the earlier sample, even if they were not evenly applied by every editorial writer. Such a large change is a sign that in this particular language variety the vocabulary surrounding the faction categories has indeed changed in the time period studied here. It is an intriguing find on its own, but it is likely most illuminating when viewed in comparison to the other faction categories and the changes visible in them. As stated above, the uneven changes in the categories signal not only a decrease in the acceptability of faction terms in general, but that some categories have become even more undesirable, whereas other categories have been affected by this process to a much lesser extent.

6.8   Unclassifiable or other

Some usages clearly filled the criteria for a faction term, but could not definitively be classified in one category or another. These terms were combined to form a separate group, which will be discussed here. In the tables and figures this group is labelled "unclassifiable". There were two main types of faction terms, which did not fit the classification system used in this work. The first type contained usages, which combined two faction terms from different categories into one phrase. Two examples of this practice are *British Muslims* in the 2009 sample and *South African blacks* in the 1976 texts. It was not considered useful here to split these usages into their component parts, since the first faction term in them is clearly a qualifier for the second.

The other main type of unclassifiable faction terms were those for which it was unclear, which faction category they were based on. Some of these were more obscure terms, for which it was could not be established whether they denoted for example a religious or an ethnic group, such as *Maronite* or *Janjawid*. The first of these two was used twice in the 1976 sample in two different editorials, 150376a and 270376b, while the second appeared three times in the 2009 sample in the texts 050309b, 060309b and 170309b. Another type were familiar terms used in such a context that determining the underlying category became impossible. One example of this is the word *Jewish*, presented in contexts where it could have referred either to the ethnic or the religious group. This was also the only unclassifiable term which was present in both samples, as seen in Table 14, in the editorials 310376c and 1030309a. Example (22) below from the editorial 310376c shows how the term *Jewish* is used in contrast to the term *Arab* in such a manner that it is unclear whether it here refers to an ethnic, religious or perhaps national group.

> (22) This could mean making a sharp distinction between those who would argue that the Israel state must for its own security be limited in size and be confined as far as is reasonably possible to a Jewish population and those who have instinctively, if not as a matter of declared policy, looked to any opportunity of expanding the territory of the state irrespective of the number of Arabs thereby brought under Israel administration. […] It would be made credible by measures that prohibit the creeping intrusion into the Arab areas of those who intend that Jewish settlement will give an established claim to the territory as a part of Israel that cannot be given up. (310376c)

Another example of a familiar term with an unclear context was *Iranian*, which was found three times in the editorial 210309b, but only one of these occurrences could be taken to clearly refer to the nationality. With the other two occurrences it was unclear whether the reference was to people belonging to the ethnic group Iranians, or whether they were referring to people of Iranian nationality living outside Iran. Thus these two occurrences were not included in the nationality category, but were counted as unclassifiable faction terms.

Yet a few more terms denoted a clear and distinct population group which itself could not be classified without trouble. Such were the words *Cockney* and *Athenians,* the first occurring once in the editorial 240409c and the second twice in the editorial 1270409c. Neither of these are really ethnic groups, nor do they properly qualify for the category of geographical origin as it is defined for the purposes of this work. These and others in this category were included because they met the criteria set for admissible examples, and while assigning them to one of the other categories was not achievable within the set up of this study, that was mostly due to practical

problems in analysing them, rather than to a fundamental incompatibility with the categories here presented.

Table 14. Absolute numbers of individual examples of unclassifiable terms.

Unclassifiable or other	1976	2009	Total	Change
British Muslims	0	1	1	+1
Greek Cypriots	6	0	6	-6
the Iraqi Kurds	1	0	1	-1
Muslim Lebanese	2	0	2	-2
Lebanese-Christian	1	0	1	-2
Lebanese Arab	1	0	1	-1
Israeli-Palestinian	0	1	1	-1
Shia Hazaras	0	1	1	+1
White Rhodesians	3	0	3	-3
Black Rhodesians	2	0	2	-2
South African Blacks	1	0	1	-1
Xhosa Transkeians	1	0	1	-1
Overseas Chinese	4	0	4	-4
Cockney	0	1	1	+1
West Germans	1	0	1	-1
Athenians	0	2	2	+2
Jewish	3	1	4	-2
Maronite	2	0	2	-2
Janjawid	0	3	3	+3
Iranian	0	2	2	+2
the Cambodians	1	0	1	-1
the Vietnamese	1	0	1	-1

As with most categories, these terms also were less frequent in the later sample. In the 1976 editorials there were 30 instances of unclassifiable faction terms, while in the 2009 texts there were twelve, as is shown in Table 6 in chapter 6.1. They were the second least frequent type in the 1976 sample, but as the drop in numbers was not as drastic with this group as it was with some others, in the 2009 sample they were exactly in the middle on the fourth place. In the overall numbers however, they were on the shared second to last place with the nationhood

category, with 42 occurrences. In the normalised figures the placements remained the same, apart from the total number of occurrences in the corpus. Here this category was on the third to last place, narrowly surpassing the nationhood category, with the normalised number of occurrences at 48.8, compared to the 47.9 occurrences of nationhood terms, as seen in Table 7. The percentage change in this category was the fourth highest with -56.0% difference going from the 1976 sample to the 2009 one. Most of this drop was in combination terms, which went from 22 occurrences to three. The absolute number of unclassifiable terms actually rose slightly, going from five in 1976 to eight in 2009. Of all the examples found in the samples, the proportion of terms in this category changed quite little, going from 8.9% of all 1976 examples to 6.6% of all 2009 examples, as seen in Figure 6. The only category that changed less in this respect, was that of ethnicity.

The entire data contained 13 different combination terms and nine different terms of undetermined category. There was very little overlap in the terms between the samples. As stated above, the only term in this category to occur both in the 1976 and the 2009 sample was *Jewish*. One of these occurrences was the word *Jews*, which was here counted with *Jewish*, since it denotes the same population group and was unclassifiable for the same reason. The relative change in the frequency of this particular term was -63.4%, as seen in Figure 15. The most common term in this category in the 1976 sample was the combination term *Greek Cypriots*, which occurred six times in the single editorial 150476b. The four most common terms formed about half of the entire set, with 16 occurrences between them out of the total of 30. Eight of the fifteen separate terms present in the 1976 sample only appeared once. The three combined terms, *British Muslims, Israeli-Palestinian* and *Shia Hazaras,* found in the 2009 sample all occurred only once in the entire corpus. Apart from *Jewish,* the unidentifiable faction terms in the 1976 editorials all occurred only one or two times, while the five unidentifiable terms in the 2009 texts, *Cockney, Athenians, Jewish, Janjawid* and *Iranian*, were found from one to three times. Examples (2) and (3) below show the two occurrences of *Athenian* and one occurrence of *Cockney* in the corpus. Example (23) is extracted from the editorial 270409c, and (24) is from the editorial 240409c.

Term		Normalised values	Change %
British Muslims	'76	0.0	--
	'09	1.2	
Greek Cypriots	'76	6.8	-100.0%
	'09	0.0	
the Iraqi Kurds	'76	1.1	-100.0%
	'09	0.0	
Muslim Lebanese	'76	2.3	-100.0%
	'09	0.0	
Lebanese-Christian	'76	1.1	-100.0%
	'09	0.0	
Lebanese Arab	'76	1.1	-100.0%
	'09	0.0	
Israeli-Palestinian	'76	0.0	--
	'09	1.2	
Shia Hazaras	'76	0.0	--
	'09	1.2	
White Rhodesians	'76	3.4	-100.0%
	'09	0.0	
Black Rhodesians	'76	2.3	-100.0%
	'09	0.0	
South African blacks	'76	1.1	-100.0%
	'09	0.0	
Xhosa Transkeians	'76	1.1	-100.0%
	'09	0.0	
Overseas Chinese	'76	4.5	-100.0%
	'09	0.0	
Cockney	'76	0.0	--
	'09	1.2	
West Germans	'76	1.1	-100.0%
	'09	0.0	
Athenians	'76	0.0	--
	'09	2.5	
Jewish	'76	0.0	--
	'09	1.1	
Maronite	'76	2.3	-100.0%
	'09	0.0	
Janjawid	'76	0.0	--
	'09	3.7	
Iranian	'76	0.0	--
	'09	2.5	
the Cambodians	'76	1.1	-100.0%
	'09	0.0	
the Vienamese	'76	1.1	-100.0%
	'09	0.0	

Figure 15. Normalised numbers of unclassifiable term examples and the percentage changes from 1976 to 2009 thereof.

(23) Athenians invented the word to describe the illiterate corruption of the Attic dialect as spoken by those Athenian colonists who settled at Soloi (Soli) in Cilicia. (270409c)

(24) But Biggs — for all his Cockney cheekiness, his brazen taunting of the authorities and his devil-take-thehindmost audacity — is neither Butch Cassidy nor the Sundance Kid. (240409c)

Unlike in the other categories, here none of the terms clearly peaked above the others, which becomes clear when one compares Table 14 with the equivalent tables for the other categories. The occurrence for any individual term in this group was small, but this was made up by the number of different terms. The percentage changes in this category are clear, as is seen in Figure 15, but that is mostly due to the lack of overlap in the terms present in the two samples. The contents of this category demonstrate two things. First, the combined terms demonstrate that on certain occasions single categories alone have not been deemed adequate for defining population groups. This in itself is not unexpected, it stands to reason that classifying people according to a single definitive feature is unsatisfactory especially in the modern world where migration and mingling are common. But even though combination terms could potentially provide greater accuracy when discussing population groups, based on the data here presented they do not appear to be the preferred solution, in fact their numbers drop sharply from 23 in the 1976 sample to only three in the 2009 editorials, and their proportion of the overall number of examples remains quite stable. Secondly, the relative scarcity of terms which could not be fit into any of the other categories without closer analysis, suggests either that to a great extent the cultural categories here outlined are a useful tool in the attempt to analyse and conceptualise the terminology used for population groups, or even that such terminology does indeed adhere to these cultural categories to a considerable extent.

Perhaps the most fruitful way to approach this category is by using it as a way to map the limits of the approach here used and pinpointing the areas where further analysis is needed. One such case is that of the smaller local identities, such as *Cockney* or *Athenian,* which clearly need a classification but can not satisfactorily be assigned a category within the system adopted for this study. Both could conceivably be either representatives of more traditional local identities or modern population concepts emerging from the realities of modern major cities. The very existence of terms which could belong to several categories in the research material shows that mapping the vocabulary onto the underlying concepts is not an entirely straightforward

exercise. The editorials contain examples where the textual context alone is not enough to determine the classification of a specific occurrence, which shows that these categories, while mostly easily applicable, are not always explicitly denoted by linguistic means. Such confusion is at times directly alluded to, but it is more often visible in the presence of the unclassifiable terms. One interesting variant can be found in the editorial 200476a, where the writer uses the phrase *tribal peoples.* It would be tempting to assume this to be a euphemism for an expression in the race category or for small or backward population groups, but the text itself provides little support for this. Terms for tribes themselves would be classified here as belonging to the ethnicity category, but the phrase above seems to operate on almost a meta level, in a similar way as one might say for example *people of different races.*

The changes seen in this category should be viewed with caution, as the numbers are small to begin with and the differences in absolute numbers and relative frequencies are very small. The change in combined faction terms seems clear however, even without overlap of individual terms in the two sets of texts. The practice of combining separate faction terms to designate a more specific population group seems clearly more wide spread in the 1976 editorials than in the 2009 texts. This forms an interesting comparison with the fact that the 1976 sample also contained more references to the superordinate religion groups, whereas the latter saw more references to subgroups of those religions. It is possible that when such a reference was needed in the older sample, there was a greater tendency to use a combined faction term for that purpose than a specific subfaction term. It should be noted that the numbers do not obviously point this out, especially since there is no clear correlation between the shifts in the different categories. This would however be a tidy explanation for these particular shifts in the data, and there is no specific reason to rule it out as a possibility.

7 Discussion

The chosen research questions and corpus produced a clear outcome, as detailed below. The questions fit the chosen research material well enough that clear evidence of trends could be extracted. The first chapter in this section contains a general outline of the results of this work, and the second one offers some suggestions for possible future research.

7.1 Summarising the results

The study at hand yielded two types of results, those related to the chosen approach and those related to the content studied. The approach focused on using cultural categories as an aid in classifying and analysing the data provided by the sample texts. This was less applicable in the portion dedicated to sex markers in language, as that focused on different expressions within a single category. In the second half dedicated to faction bias this approach however truly bore fruit. It made it possible to easily classify faction terms into subgroups, thus making them more manageable. Through this division it was easier to locate trends and general shifts in the data, rather than just focus on changes in individual expressions. This worked well, especially in a study of this size, where the resources available severely limit the amount of material, and thus the number of examples, which can be analysed. On the practical level it made it possible to group the terms in units of manageable size. On the theoretical level separating the different categories behind the terms made it possible to more precisely evaluate the development in these usages and to break the changes down into meaningful trends. The latter estimations would have been much more difficult had all the faction terms been treated as a single group of individual expressions. Thus while the approach may seem somewhat unimportant for the part of the study dealing with sex markers, it very much proved its worth in the sections where faction terminology was analysed.

Chapter five, which details the sex marker section of the study, shows a clear shift in the material away from forms which have been criticised as showing a sex-based bias, as well as one towards the greater visibility of women in the texts. This is visible in the shift towards marriage neutral address forms and away from androcentric generics. In the light of the data presented in this study, this shift is indisputable. Most of the increase in references to women seems to be seen in the increase of marriage neutral address forms, since the form *Mrs* retained its popularity, as detailed in chapter 5.2 and shown in Figure 3. Especially for women in powerful positions being addressed with *Mrs* may be a deliberate tactical choice, intended to

portray a certain image to the people around them. It is likely that the general trend is towards marriage neutral lexis, but the form *Mrs* preserves its numbers due to a general increase in the visibility of women and the fact that some women do select it in order to, for example, communicate traditional values. The increased visibility did not however result in the increase in feminine agentives. While androcentric generics were clearly reduced, as shown in Figure 4 and deliberately neutral expressions showed moderate increase, feminine agentives were almost absent from the 2009 sample. This is clear from the normalised figures of neutral expressions and feminine agentives presented in Figure 5 in chapter 5.3.3. This suggests that the shift towards sex neutral language reduces explicit sex visibility in this respect.

A similarly clear shift was visible in the faction categories. As seen in Figure 6 in chapter 6.1, most of the categories lost popularity very clearly, with the categories of nationality and religion being the exceptions, the former decreasing by 4.8% and the latter increasing in frequency by 7.3%. As detailed in the chapters 6.3, 6.4, 6.5 and 6.7 the faction categories of nationhood, ethnicity, geographical origin, and race respectively, were all considerably reduced in numbers, while those of nationality and religion, discussed in chapters 6.2 and 6.6, remained comparably stable. The total absolute number of faction terms diminished clearly as seen in Table 6, with the relative frequency of all faction terms decreasing by 41.0%, as shown in Table 7. Figure 7 shows how the proportionate share of each category out of all examples has changed, clearly indicating that the categories of nationality and religion grow in proportion, whereas the proportions of other categories either remain stable or decrease. This indicates that while most of the faction terms have become less frequent in this language sample, those denoting nationality or religion have retained their frequency. Thus it is reasonable to assume that the concepts these categories rely on have also come under questioning, except for the above mentioned nationality and religion. It should here be noted that this is not necessarily an equal development throughout a category. This was seen in the religion category, discussed in chapter 6.6, where superordinate terms were replaced with terms denoting subgroups. While the data in the sex category shows a general movement away from sex marked language, the shifts in faction category can be interpreted as certain categories having become less acceptable to refer to, while others have been deemed largely acceptable.

Whether the shifts overviewed here become established as a new language practice will remain to be seen. It is possible for the new practices to become more prominent as they represent a high status variety to which speakers might aspire in order to capture some of that status for

themselves. If this happens the older practices will likely become stigmatised as uneducated and unrefined, even rude. In fact it is possible they are already viewed that way. The second possibility is that the new forms are rejected in the common populace as snobbery precisely due to their high status. In this case the new forms would be stigmatised as elitist and over correct usage. Whether the change is permanent or rejected, can only be ascertained through further research.

7.2   Further research

While this study answered many questions, it both left some unanswered and pointed the way to new ones. The most obvious is the one mentioned at the end of the previous chapter, that concerning the permanence of the language changes discovered in this study. A study of a similar design with material extracted from the same source after a significant time interval would give insight into whether the shift found here remains, is reversed or develops further. Similarly going back in time with the same source would likely show whether this shift is a recent phenomenon or a part of a transition taking place over a longer period of time. In addition, at this point in time this study is essentially historical in nature and it would therefore pair well with a similar study based on contemporary primary sources. Another interesting question is to which extent these changes are visible in other language varieties across societal and geographical distances. For that one would need to study the same time periods as here, but choose for example spoken language sources.

Questions which remained largely unanswered in this study were to which extent the usages in question had been replaced with neutral and invisible usages and the extent to which these changes are informed by speakers' attitudes about language questions. Due to limitations in scope and material, it was impossible to delve into either question here, but a more careful analysis of larger samples could possibly shed some light on this, as could a study which utilised interviews and tests in addition to language samples in order to glean information about language attitudes. A larger sample might also go a long way towards reducing the influence of chance and the preferences of individual writers on the results. As it is, the scarcity of many usages made it very difficult to proclaim reliable results on them. This could be remedied with methodology from computational linguistics, for example by running automatised searches in larger corpora for the expressions uncovered in the samples studied here. In general this line of enquiry would greatly benefit from a study or studies with a greater scope than this one across

the temporal, social and geographical dimensions of language. As interesting as the results of this study were in themselves, perhaps the best use for this work would be as a springboard for further exploration of how different culturally salient demographic categories are realised in language, past and present.

8 Conclusion

The study at hand shows a clear shift in the samples of Standard English here examined with regard to how demographic categories are expressed in the lexis. The sex marked usages argued to be biased gave way to neutral terms, while most faction categories were greatly reduced in the samples and only the categories religion and nationality retained their position. In addition to this, approaching these issues through the concept of cultural categories proved instrumental in acquiring these results.

This shows that a clear linguistic change has taken place in the sample of standard English chosen as the material for this study. The examined sex marked forms along with the faction categories nationhood, ethnicity, geographical origin, and race are in the process of falling out of use. This supports the claim that social pressures to use language in ways perceived to be less biased or less in favour of certain population groups are having a genuine effect on this particular sample of English. The change is far from finished however, in fact it may never be completed. The male centric linguistic practices are still present in the language and a shift from one faction category to another does not yet necessarily mean that the language is actually used in a way perceived to be inclusive. It is important to note, that there is not necessarily anything inherent in a particular demographic category, which would lead to biased language or even stigmatisation of that category.

Whether these changes in the language reflect a cultural change in the underlying cultural categories, can not be deduced from this study. While the social realities providing the push for these changes are firmly beyond the scope of this study, the results do suggest there is an underlying instability within the cultural categories these features of the language are based on. Of course this is limited to the language variety presently studied, but that does not exclude the possibility of a similar shift in the wider culture. The shift within and across the categories in this particular language sample is however clearly demonstrated here, only the larger linguistic and social dimensions remain unclear.

Standard British English is not an unchanging entity. This study has given a tantalising glimpse into a language change as it is happening before our eyes. The arguments over how much influence language has on how we perceive reality, are still taking place. Notwithstanding Tony Crowley's quip about the phrase 'standard English' containing two significant confusions, those around 'standard' and those around 'English' (1999:271), this particular high status

variety of the language known as English is clearly changing, and the change is driven not by our perception of reality, but by our perception of language itself.

References

Primary sources

The Times. 1976, March-April. Editorials.

The Times. 2009, March-April. Editorials.

Application for Somerset Travel pass. Photographed.

Secondary sources

Abercrombie, Nicholas et al. 1988. Contemporary British Society. *A New Introduction to Sociology*. Cambridge: Polity Press.

Appiah, K. Anthony and Amy Gutmann. 1996. *Color Conscious. The Political Morality of Race.* Princeton: Princeton University Press.

Beeching, Kate. 2006. Synchronic and diachronic variation: the how and why of sociolinguistic corpora. In Archer Dawn, Rayson Paul and Andrew Wilson (eds.). *Corpus Linguistics Around the World*. Boston: Brill. pp. 49-61.

Bodine, Ann. 1998. Androcentrism in prescriptive grammar: singular 'they', sex-indefinite 'he', and 'he or she'. In Cameron, Deborah (ed.). *The Feminist Critique of Language: A Reader*. London: Routledge. pp.124-138.

Braun, Friederike. 1997. Making Men out of People. In Kotthoff, Helga and Ruth Wodak (eds.). *Communicating Gender in Context.* Amsterdam: John Benjamins. pp.3-29.

Caldas-Coulthard, Carmen and Malcolm Coulthard (eds.). 1996. *Texts and Practices, Readings in Critical Discourse Analysis.* New York: Routledge.

Coulthard, Malcolm et al. 2017. *An Introduction to Forensic Linguistics. Language in Evidence.* Abingdon: Oxon.

Crowley, Tony. 1999. Curiouser and Curiouser: Falling standards in the standard English debate. In Bex Tony and Richard J. Watts (eds.). *Standard English: The widening debate*. London: Routledge. pp. 271-279.

Doyle, Margaret. 1998. Introduction to the A-Z of Non-Sexist Language. In Cameron, Deborah (ed.). *The Feminist Critique of Language: A Reader.* London: Routledge. Pp. 149-179.

Fasold, Ralph. 1990. *The Sociolinguistics of Language*. Oxford: Blackwell.

Leitner, Gerhard. [2008] 1997. The Sociolinguistics of Communication Media. In Coulmas, Florian (ed.). *The Handbook of Sociolinguistics. Malden*: Blackwell. pp. 187-204

Mautner, Gerlinde and Mark Learmonth. 2020. From administrator to CEO: Exploring changing representations of hierarchy and prestige in a diachronic corpus of academic management writing. *Discourse & Communication 14(3)*. pp. 273-293.

Office for National Statistics. 2000. *Britain 2001. The Official Yearbook of the United Kingdom.* London: Office for National Statistics.

Romaine, Suzanne. 1994. *Language in Society*. Oxford: Oxford University Press.

Romaine, Suzanne. 2001. A Corpus-based View of Gender in British and American English. In Hellinger, Marlis and Hadumod Bussmann (eds.). *Gender Across Languages: The Linguistic Representation of Women and Men Volume I*. Amsterdam: John Benjamins Publishing. pp. 153-176.

Rothenberg, Paula S (ed.). 2001. *Race, Class and Gender in the United States: an Integrated Study*. New York: Worth.

Spender, Dale. 1980. *Man Made Language*. London: Routledge.

Van Dijk, Teun A. 1993. *Elite Discourse and Racism*. Newbury Park: Sage Publications.

Wodak, Ruth. 1996. The Genesis of Racist Discourse in Austria Since 1989. In Caldas-Coulthard, Carmen and Malcolm Coulthard (eds.). *Texts and Practices, Readings in Critical Discourse Analysis.* New York: Routledge. pp. 107-128.

Wodak, Ruth. 2001. What CDA is about – a summary of its history, important concepts and its developments. In Wodak Ruth and Michael Meyer (eds.). *Methods of Critical Discourse Analysis.* London: Sage. pp. 1-13.

Appendices

Appendix 1 Sample sizes with number of texts, word counts, the change in word counts between samples, and the normalisation factors

	1976	**2009**	**Total**	**Change**	**Change %**
Number of texts	118	156	272	+36	+42.5
Word count - approximate for 1976 due to OCR and scanning inaccuracies	88 544	80 569	169 113	-7975	-9.01
Normalised word count	100 000	100 000			
Normalisation factor – per 100 000 words	1.1294	1.2412			

Appendix 2 Number of texts and word counts in the samples, according to whether the texts contained usable examples or not

	1976	2009	Total	1976 normalised	2009 normalised	Change	Change %
Sex							
Number of empty texts	58	106	164			+48	+82.8
Word count of empty texts	39 696	53 994	93 690	44 832.7	67 017.4	+22 184.7	+49.5
Number of useful texts	60	48	108			-12	-20.0
Word count of useful texts	48 848	26 575	75 423	55 168.9	32 984.9	-22 184.0	-40.2
Faction							
Number of empty texts	58	85	143			+27	+46.6
Word count of empty texts	41 692	45 452	87 144	47 086.9	56 415.0	+9328.1	+19.8
Number of useful texts	60	69	129			+9	+15.0
Word count of useful texts	46 852	35 117	81 969	52 914.6	43 587.2	-9 327.4	-17.6

Appendix 3 Topics of the editorials

	International issues	Economy or budget	Domestic politics	General domestic	Art, culture and people
1976 editorials	50	13	18	32	5
2009 editorials	55	17	12	54	16
Totals	**105**	**30**	**30**	**86**	**21**

Appendix 4 Topics of the editorials according to whether the texts contained usable examples or not

	International issues	Economy or budget	Domestic politics	General domestic	Art, culture and people
Sex					
1976 empty texts	50	13	18	32	5
1976 useful texts	55	17	12	54	16
2009 empty texts	36	13	9	35	14
2009 useful texts	19	4	3	19	2
Faction					
1976 empty texts	8	9	10	29	2
1976 useful texts	42	4	8	3	3
2009 empty texts	20	14	11	32	9
2009 useful texts	35	3	1	22	7

Appendix 5 Somerset travel pass application, showing marital status markers for women

Somerset **TRAVELPASS**

Application Form

Please complete in block capitals

Mr/Mrs/Miss/Ms*...
*Delete as appropriate

First name...

Surname ....

Date of birth ....

Permanent Address

....

....

....

Post code ....

Phone No ....

I am over 60 years old and would like a Somerset
TravelPass   ❏
I am disabled and would like a Somerset TravelPass
(If you are unsure if you qualify please contact your local
council)   ❏
(Please give further details overleaf)
I am unable to travel on my own and require my pass to
have a Companion element and I am in receipt of higher
rate mobility allowance or attendance allowance   ❏

Disability
Complete only if claiming on medical grounds

Please give brief details of your medical disability enclosing
copies of either medical registration certificates (e.g. partially
sighted, disabled), mobility component medical benefits
(e.g. disability living allowance) or something official to show
you have a medical disability in support of your application.

....

....

....

YOU MUST ENCLOSE ALL OF THE FOLLOWING PROOF -
FAILURE TO DO THIS WILL LEAD TO DELAYS
❏ Recent passport style photograph
❏ Copy of proof of age
❏ Copy of proof of residency
❏ Enclosed copy of proof of eligibility

Please write name and date of birth on reverse of photo

**I can confirm that the information given on this form is
to my knowledge true and correct. I am a permanent
resident in the County of Somerset and my council tax
is paid to one of the participating councils.**

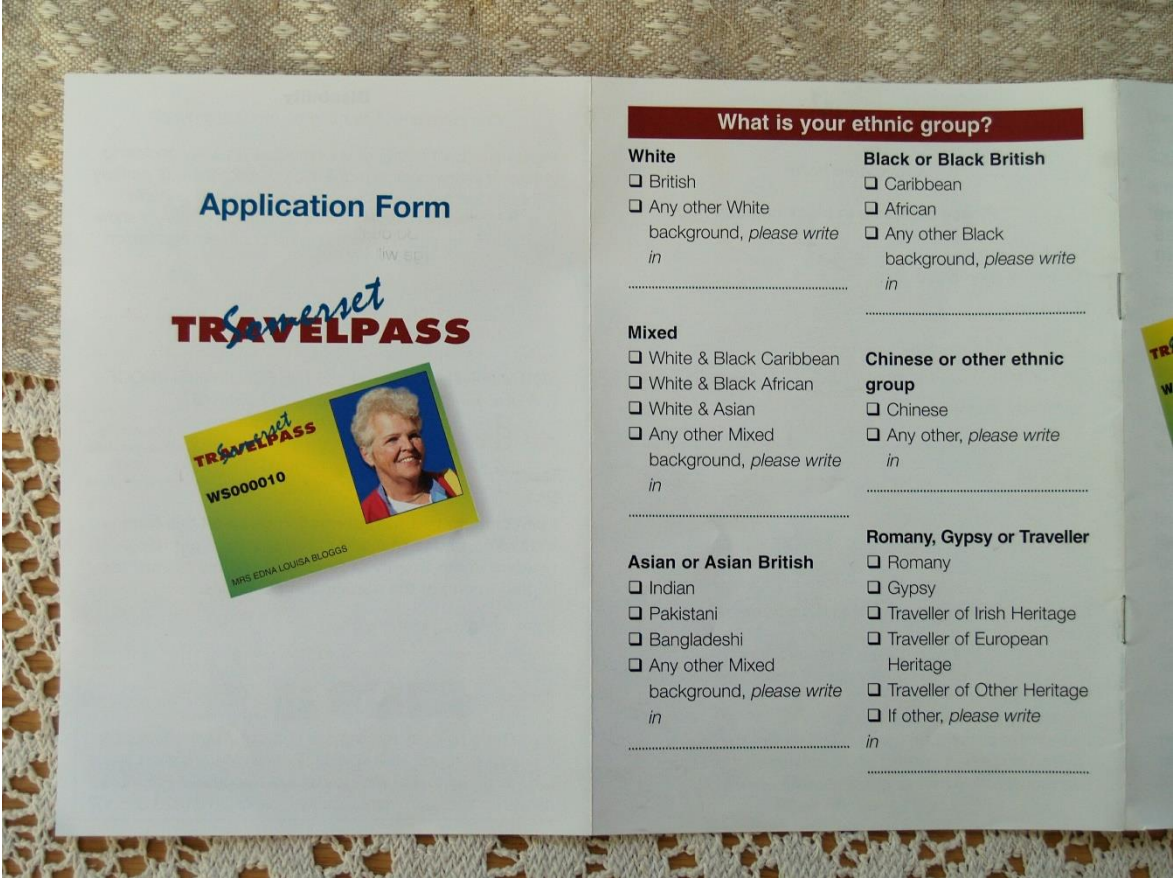Signature of applicant;
....

Lãnguãgé Lĩñé

If you have difficulty reading this publication we can supply
it in large print, on audio cassette, arrange translation or
provide a member of staff to discuss the details. For more
details contact Somerset County Council on 01823 355307

Appendix 6 Somerset travel pass application, showing the available options for "ethnic group"

Application Form

TRAVELPASS Somerset

TRAVELPASS Somerset
WS000010
MRS EDNA LOUISA BLOGGS

What is your ethnic group?

White
❑ British
❑ Any other White background, *please write in*
....

Mixed
❑ White & Black Caribbean
❑ White & Black African
❑ White & Asian
❑ Any other Mixed background, *please write in*
....

Asian or Asian British
❑ Indian
❑ Pakistani
❑ Bangladeshi
❑ Any other Mixed background, *please write in*
....

Black or Black British
❑ Caribbean
❑ African
❑ Any other Black background, *please write in*
....

Chinese or other ethnic group
❑ Chinese
❑ Any other, *please write in*
....

Romany, Gypsy or Traveller
❑ Romany
❑ Gypsy
❑ Traveller of Irish Heritage
❑ Traveller of European Heritage
❑ Traveller of Other Heritage
❑ If other, *please write in*
....

Appendix 7 Finnish language summary

Vaikuttaako kieli todellisuuteen, vai todellisuus kieleen?

Kielen suhde ympäröivään todellisuuteen on monimutkainen. Sen lisäksi, että kieli kuvailee ympäröivää todellisuutta, se monen mielestä myös osallistuu todellisuuden rakentamiseen. Fyysinen maailma itsessään ei muutu pelkästään kielen keinoin, mutta kielen kautta ihmiset jäsentävät ja ymmärtävät todellisuutta, jossa elävät. Tässä prosessissa tärkeässä roolissa ovat kategoriat ja käsitteet, jotka puolestaan ovat sosiaalisten muutosprosessien alaisia. Näillä kategorioilla ja käsitteillä jaotellaan myös ihmiset monenlaisiin ryhmiin.

1900-luvun aikana monet edistysmieliset sosiaaliset liikkeet alkoivat kiinnittää huomiota kielen rooliin sosiaalisen todellisuuden rakentumisessa ja alkoivat pyrkiä muokkaamaan tätä todellisuutta kielen kautta. Nämä pyrkimykset juontavat juurensa Sapir-Whorf-hypoteesiin, jonka mukaan kieli määrää, miten sen puhujat näkevät maailman. Niin naisasialiike, työväenliike, alkuperäiskansojen oikeuksia ajavat liikkeet, kuin monet muutkin maailman parantamista ja syrjinnän vähentämistä ajavat liikkeet ovat lisänneet tavoitteisiinsa myös kielen muuttamisen näitä asioita ajavien mielestä sopivampaan muotoon. Mutta kuten kaikki muutospyrkimykset, myös tämä toiminta on kohdannut vastustusta. Huomattavan paljon on kirjoitettu sekä asian puolesta että vastaan, mutta hyvin vähän löytyy varsinaista tietoa näiden pyrkimysten vaikutuksista. Tämä pätee myös kielitieteessä. Tämän tutkimuksen tarkoituksena onkin pieniltä osin paikata tätä aukkoa tutkimuksessa.

Edistysliikkeet perustavat pyrkimyksensä suurelta osin ajatukselle, että ihmiset jakautuvat – ja jaetaan – yhteiskunnassa eri väestöryhmiin, ja että näiden väestöryhmien asemat yhteiskunnassa poikkeavat toisistaan. Näin ollen myös liikkeiden pyrkimykset vaikuttaa kieleen ovat keskittyneet suureksi osaksi siihen, miten nämä väestöryhmät näkyvät kielessä. Kieli muuttuu jatkuvasti, mutta tässä tapauksessa kielen normaalin, luonnollisen muutoksen lisäksi kieltä on tarkoituksella pyritty muuttamaan ideologisin perustein. Tätä pyrkimystä on myös selkeästi vastustettu. Maailmanhistoriassa tämä tuskin on ainutlaatuinen tilanne, mutta tällä kertaa näitä pyrkimyksiä ja niiden vaikutuksia voidaan tarkastella niiden yhä ollessa ajankohtaisia.

Muita tutkimuksia

Kuten edellä mainittiin, näiden muutospyrkimysten vaikutuksia on tutkittu kovin vähän. Varsinaista edeltäjää tälle tutkimukselle ei löytynyt, mutta joitain yhtymäkohtia muuhun kielentutkimukseen kyllä on. Suurin osa eriarvoisiksi nähtyjä väestöryhmiä koskevaa sanastoa käsittelevästä tutkimuksesta tulee kielitieteessä kriittisen diskurssianalyysin alalta (Critical Discourse Analysis, tästä lähin CDA). Nämä tutkimukset asettuvat avoimesti kannattamaan yllämainittuja muutospyrkimyksiä, ja siten usein osallistuvatkin aktiivisesti muutosprosessiin. CDA ei tyypillisesti kartoita kielen muutosta neutraalista näkökulmasta, eikä sitä ovatko ideologiset muutospyrkimykset tuottaneet tulosta. CDA-alalla on jonkin verran tutkimusta, jossa vertaillaan keskenään eri ajoilta peräisin olevia korpuksia, eli tekstinäytekokoelmia, mutta suurin osa tutkimuksesta keskittyy analysoimaan yksittäisiä korpuksia CDA:n tarjoamin työkaluin. Käytetyt tekstinäytekokoelmat ovat myös usein sangen pieniä. Eniten korpuksia hyödynnetään tietokoneavusteisessa kielitutkimuksessa, mutta näissä tutkimuksissa käytetään tyypillisesti hyvin suuria korpuksia. Tällaisia korpuksia tutkitaan suureksi osaksi automaation keinoin, mikä tekee lähes mahdottomaksi tutkimukset, joissa tekstejä tarkasteltaessa ei ennakkoon tiedetä mitä ilmaisuja niistä tulee löytymään. Pieniä tai kohtuullisen pieniä tekstinäytekokoelmia vertaillaan toisiinsa kielitieteessä eniten forensisen kielitieteen alalla, eli töissä joissa kielitiedettä käytetään apuna rikostutkimuksessa, tai sen avulla tarkastellaan itse oikeusprosessin kulkua. Huomattavin ero tämän pro gradu-tutkimuksen ja forensisen kielitieteen tutkimusten välillä on se, että tässä työssä päämäärä on yleisluontoisempi kielen tilan kartoittaminen, kun taas forensiset kielitutkimukset pyrkivät vastaamaan hyvin rajattuihin rikostutkimuksia ja oikeusprosessia koskeviin kysymyksiin.

Peruskivet ja -käsitteet

Tämä työ on luonteeltaan kvantitatiivinen tutkimusmatka standardibrittienglannin väestöryhmiä koskevaan sanastoon ja siinä mahdollisesti tapahtuneisiin muutoksiin. Sekä instituutiot että ihmiset jakavat väeston erilaisiin kategorioihin (demographic categories), jotka kielessä tulevat esille näistä ryhmistä ja niiden jäsenistä käytettävässä sanastossa. Osa näistä kategorioista perustuu biologisille, osa kulttuurisille ominaisuuksille (demographic variable), ja on huomattavaa että joitain kategorioita voi olla hyvinkin vaikea erottaa toisistaan. Kansallisuuden ja etnisyyden välillä voi esimerkiksi olla hyvinkin vaikea tehdä eroa, ja

esimerkiksi juutalaisuus on nähty sekä uskonnollisena että rotukategoriana. Tässä työssä tarkastelun alla olevat kategoriat on jaettu puhtaasti käytännön syistä kahteen osaan, ensimmäinen osa käsittelee biologista sukupuolta ja toinen muita esiin nousseita kategorioita. Näin koska ensin mainittu koostuu englannin kielessä pienestä määrästä vakiintuneita ilmaisuja, joista osa on kieliopillisia rakenteita jälkimmäisen näkyessä puhtaasti sanastossa, ja koska ensin mainittu on saanut huomattavasti enemmän huomiota osakseen kielitieteessä. Tässä yhteydessä on mainittava, että tällä hetkellä esitetyt näkemykset biologisen sukupuolen ja varsinaisen sukupuolikategorian täydellisestä tai lähes täydellisestä erillisyydestä ovat epäolennaisia tämän tutkimuksen kannalta, jo pelkästään tutkimuksen aikakehyksenkin vuoksi. Muut kategoriat niputettiin termin *faction* alle, jälleen puhtaasti käytännön syistä. Termillä ei tässä yhteydessä tarkoiteta muuta kuin kansanryhmää, joksi se myös tässä tiivistelmässä tästä lähin käännetään.

Englannin kielen sukupuolisidonnaisesta termistöstä englanniksi keskusteltaessa käytetään tyypillisesti termiä *gender*. Kieliopissa tämä termi tarkoittaa sanan sukua, mutta englannin kielestä varsinainen kieliopillinen suku puuttuu. Piirteet, joista käytetään sanaa *gender* ovat siis itse asiassa biologista sukupuolta ilmaisevaa termistöä, jossa vaihtoehtoina ovat urospuolisuutta, naaraspuolisuutta ja tuntematonta sukupuolta ilmaisevat termit. Nämä ilmaisut voidaan jakaa kahteen ryhmään, maskuliinisiin yleistermeihin (androcentric generics) ja epäsymmetrisiin viittauksiin tai puhutteluihin (asymmetrical address forms). Maskuliiniset yleistermit ovat miespuolisiin henkilöihin viittaavia sanoja, joita käytetään myös viitatessa ryhmiin joissa on sekä mies- että naispuolisia jäseniä. Näistä yksi tunnetuimmista on sana *man* käyttö sanan *human* sijasta, ja toinen hyvin yleinen esimerkki on maskuliinisen kolmannen persoonan pronominin käyttö ihmisistä yleensä. Näiden termien vastustus perustuu käsitykseen, että ne sulkevat naispuoliset kohteet käsitteen ulkopuolelle ja tekevät heistä kielellisesti näkymättömiä. Ne, jotka eivät näistä termeistä halua luopua, luonnollisesti kiistävät tämän. Kokeellisissa tutkimuksissa on edelliselle väittämälle kuitenkin löytynyt tukea. Suzanne Romaine ja Dale Spender muun muassa viittaavat useisin tutkimuksiin, joissa koehenkilöt tulkitsivat maskuliiniset yleistermit pääasiallisesti miehiin viittaaviksi. Epäsymmetriset viittaukset ja puhuttelut ovat sanoja tai päätteitä, jotka joko merkitsevät tekijän naispuoliseksi, tai ilmaisevat aviosäädyn vain naisten kohdalla. Näitä on vastustettu, koska niiden on nähty ilmaisevan naisten ja miesten välistä statuseroa, esimerkiksi kohtelemalla miespuolisuutta oletusarvona ja naispuolisuutta poikkeuksena. Suomen kielessä vastaavia esimerkkejä olisivat *-tar* ja *-tär* -päätteet sanoissa, sekä aviosäätyä ilmaisevat *rouva* ja *neiti*.

Kansanryhmätermistössä on kyse väestöryhmien nimeämisestä, tämän termistön muuttamisen puolesta on esitetty kaksi pääargumenttia. Ensimmäinen argumentti on, että olemassaolevat termit kantavat mukanaan negatiivisia konnotaatioita ja ne tulisi sen tähden hylätä, toinen argumentti on, että väestöryhmien edustajien tulisi saada itse päättää mitä sanoja he haluavat heistä käytettävän. Konnotaatiot, eli sanoihin liittyneet sivumerkitykset, eivät ole pitkällä aikavälillä vakaita ja ne voivatkin muuttua neutraaleista negatiivisiksi ajan saatossa. Tämä voi johtaa negatiivisiksi muuttuneiden termien hylkäämiseen, vaikka ideologisia pyrkimyksiä kielen muuttamiseen ei olisikaan. Näitä termejä tarkastellessa pitää kuitenkin ottaa huomioon, että väestöryhmät eivät läheskään aina ole selkeästi rajattuja populaatioita. Läheskään aina ei ole selvää mihin kansanryhmään yksilö pitäisi sisällyttää ja ryhmän nimeäminen voi myös käytännössä johtaa uuden ryhmän syntymiseen. Kansanryhmätermistön kohdalla on valitettavan vähän tehty tutkimusta. Moni kielitieteilijä puoltaa kantaa, jonka mukaan tietyt ilmaisut ovat rasistista ja siten negatiivisesti vaikuttavaa kieltä, mutta tätä ei juurikaan ole osoitettu kokeellisesti. Samoin kansaryhmätermistöä ja sen esiintymistiheyksiä ei ole kartoitettu läheskään yhtä runsaasti kuin esimerkiksi edellämainittuja epäsymmetrisiä viittauksia.

Yksi vaikeus väestöryhmäterminologiaa tutkittaessa on, että nämä ryhmät itse eivät ole selkeästi määriteltyjä kokonaisuuksia. Uskontotermistöä saatetaan käyttää puhuttaessa kansallisuudesta, rotua sivuavia käsitteitä voidaan käyttää etnisten ryhmien sijasta ja niin edelleen. On jopa kyseenalaista missä määrin kielen käyttäjät itse tekevät selkeitä eroja näiden ryhmien välillä, varsinkin kun useilla tämän ryhmän käsitteillä on sekava historia. Tämä toisaalta tarkoittaa myös, että väestökategorioiden osuuteen paneutuva termistötutkimus voi osoittautua yllättävän mielenkiintoiseksi.

Englanninkielisessä maailmassa rodun käsite on yhä rutiininomaisessa käytössä, joskin se on suurimmaksi osaksi hylätty biologiassa ja antropologiassa. Huolimatta rotuterminologian suhteellisesta yleisyydestä, sanaston merkitys on huomattavan epäselvä. Luetaanko esimerkiksi Välimeren alueen ihmiset kuuluvaksi käsitteen *white* alle, tai tuleeko ihmisen olla afrikkalaista alkuperää ollakseen *black*, vai sopiiko termiä käyttää kenestä tahansa tummaihoisesta? Kansallisuus ja etnisyys ovat käsitteinä sukua toisilleen, mutta edellämainittu viittaa nykyään lähinnä viralliseen kansalaisuuteen, kun taas jälkimmäinen viittaa pääasiallisesti epäviralliseen kulttuuripohjaiseen ryhmään. Etnisyys ja rotu puolestaan eroavat toisistaan siten, että etnisyyden pohjautuessa kulttuuriin, moderni rotukäsite tyypillisesti pohjautuu ulkonäköön. Tätä jakoa kuitenkin monimutkaistaa se, että jossain kulttuurissa pienestä asti kasvaneesta ja sen täysin omaksuneesta ihmisestä voidaan yhä käyttää vierasta etnisyyttä ilmaisevaa termiä.

Uskonto puolestaan saattaa osua päällekkäin useankin muun kategorian kanssa ja sen merkitys saattaa sosiaalisesta kontekstista riippuen vaihdella erittäin tärkeästä lähes olemattomaan. Tämä yleinen hämmennys ei kuitenkaan tarkoita sitä etteikö termistön tarkastelu olisi hedelmällistä, varsinkin jos voidaan osoittaa että termistö voidaan jakaa sen perusteella, mihin käsitteeseen se perustuu. Tämä puolestaan mahdollistaa eri kategorioiden esiintymistiheyksien mittaamisen erilaisissa kielinäytteissä.

Materiaalit ja menetelmät

Tällaista tutkimusta ei mitä ilmeisimmin ole aiemmin tehty, ja suuri osa metodologiasta muotoutuikin käytettävissä olevia tekstikokoelmia läpi käytäessä. Tutkimusmateriaalina oli kaksi kokoelmaa *The Times*-sanomalehden pääkirjoituksia, yksi vuodelta 1976 ja toinen vuodelta 2009. Molemmissa kokoelmissa olivat mukana lehden näiden vuosien maaliskuussa ja huhtikuussa julkaisemat pääkirjoitukset. Vuoden 1976 tekstien lähteenä oli *Times Digital Archive*, jossa tarvitut numerot olivat tarjolla skannattuina versioina. Vuoden 2009 tekstit olivat *The Timesin* verkkolehdestä, joka oli saatavilla maksullisena tilauksena. Tämä sanomalehti valittiin sen korkean statuksen ja konservatiivisen luonteen vuoksi, koska näin voidaan olettaa sen sekä edustavan hyväksyttyä standardibrittienglantia, että olevan vastustuskykyinen pinnallisille muotioikuille. Vanhemmat tekstit muutettiin muokattavaan muotoon optisella tekstintunnistusohjelmalla (OCR), mikä yhdessä skannattujen tekstien vaihtelevan laadun kanssa johti epätarkkuuksiin. Uudempia tekstejä voitiin muokata sellaisenaan tekstinkäsittelyohjelmaan kopioinnin jälkeen. Ensimmäinen näyte koostui 118 pääkirjoituksesta ja toinen 154 tekstistä. Edellisessä oli arviolta 88 544 sanaa ja jälkimmäisessä 80 569 sanaa. Tuloksia käsitellessä käytettiin normalisaatiokerrointa, milloin tämä oli tarkoituksenmukaista. Teksteistä etsittiin ilmaisuja, jotka kuuluivat yksiselitteisesti johonkin edellämainituista ryhmistä. Tämä käsitti siis sukupuolitermistön ja kansanryhmätermistön. Esimerkit ryhmiteltiin tyypin perusteella taulukkoihin, niiden määrät laskettiin ja saatuja lukuja verrattiin keskenään. Sekä absoluuttiset että suhteelliset esiintymistiheydet laskettiin. Esimerkkien joukosta jätettiin pois muun muassa viralliset poliittiset termit, koska nämä ovat käytännössä vakaita yhtä kauan kuin valtiot ja muut niiden pohjalla olevat viralliset elimet. Kansanryhmätermeistä ne, joiden peruskategoriaa ei voitu osoittaa, sekä kahdesta eri kategoriaan pohjautuvasta termistä koostuvat yhdystermit laskettiin mukaan omana ryhmänään. Sukupuolitermeistä ne maskuliinisilta yleistermeiltä näyttävät termit, joiden kohdalla ei ollut

selvää viittasivatko ne sittenkin vain miehistä koostuviin ryhmiin, jätettiin tutkimuksen ulkopuolelle.

Tulokset

Sukupuolitermistöstä kartoitettiin puhuttelumuotojen *Mrs*, *Miss* ja *Ms* esiintyminen, 0-puhuttelu (0-address), yleistermit *he* ja *man*, tarkoituksellisesti sukupuolineutraalit ilmaisut sekä feminiiniset tekijätermit (feminine agentive). Laskettaessa yhteen kaikki puhuttelumuodot, mukaan lukien 0-puhuttelu jossa käytetään pelkästään henkilön nimeä, nähdään yksittäisten naisten mainintojen määrän selvästi nousseen tutkimuksen ajanjaksona. Vuoden 1976 teksteissä näitä oli 14, kun taas 2009 teksteissä vastaava luku oli 65. Suhteellisesti tämä oli 410,2 prosentin kasvu. Tässä termiryhmässä *Mrs* -puhuttelun osuus kasvoi 10 tapauksesta 12 tapaukseen, mikä oli prosenteissa +31,9. *Miss* ei ollut alun perinkään lukuisa termi, ensimmäisessä kokoelmassa se esiintyi vain kolmesti ja toisesta se oli kokonaan poissa. Sitä vastoin *Ms* -puhuttelua ei löytynyt kertaakaan ensimmäisestä kokoelmasta, kun taas toisessa kokoelmassa se esiintyi 18 kertaa. 0-puhutteluja löytyi ensimmäisestä kokoelmasta yksi ja toisesta 35 kappaletta, jolloin muutosprosentiksi tuli +3746,5. Molempien maskuliinisten yleistermien osuus putosi teksteissä. *He*-yleistermi esiintyi vuoden 1976 teksteissä 92 kertaa, mutta vain 12 kertaa vuoden 2009 teksteissä, muutosprosentti oli tällöin -85,7. *Man*-yleistermiä löytyi 33 kappaletta ensimmäisestä tekstikokoelmasta ja 14 jälkimmäisestä. Tämän termin esiintymistiheys laski 53,4%. Tarkoituksellisesti sukupuolineutraaleja ilmaisuja ja feminiinisiä tekijätermejä ei kumpaakaan löytynyt ensimmäisestä tekstikokoelmasta, toisessa oli edeltäviä 9 ja jälkimmäisiä 2 kappaletta. Sukupuolitermien painopiste myös siirtyi selvästi. Vuoden 1976 pääkirjoituksissa ylivoimaisesti yleisimmät sukupuolitermit olivat maskuliinisia yleistermejä, kun taas vuoden 2009 pääkirjoituksissa suurin osa sukupuolitermistöstä oli puhuttelumuotoja. Muutokset tässä kategoriassa ovat tasa-arvolla perusteltujen ideologisten kielenmuutospyrkimysten mukaisia. Aviosäätyneutraalit puhuttelumuodot ja yleiset tarkoituksellisen sukupuolineutraalit ilmaisut lisääntyivät ja maskuliiniset yleistermit vähenivät. Ainut tämän trendin vastainen muutos nähtiin feminiinisissä tekijätermeissä, joiden lukumäärä nousi lievästi. Absoluuttinen ero oli tämän termin kohdalla pieni, vain kaksi esiintymiskertaa, joten tämä voi olla myös sattumaa. Tätä tukee myös kokonaistuloksissa nähty selvä siirtymä pois maskuliinisista yleistermeistä. Kansanryhmätermistö kartoitettiin kuten sukupuolitermistökin, ainoa ero oli, ettei etukäteen tiedetty mitä tähän ryhmään kuuluvia termejä aineistosta löytyisi. Nämä termit jaoteltiin ensin

sen mukaan, mihin kategoriaan ne perustuivat ja sen jälkeen yksittäisiin termeihin. Sekä kategorioiden että yksittäisten termien esiintymistiheydet laskettiin. Kansanryhmätermistön kokonaismäärä putosi 337 esimerkistä 181 esimerkkiin kokoelmasta toiseen siirryttäessä, suhteellisissa määrissä tämä merkitsi 41,0 prosentin laskua. Tästä huolimatta kansanryhmätermistön esimerkkejä oli aineistossa enemmän kuin sukupuolitermistöä. Pääkirjoituksista löytyneet kansaryhmätermit voitiin jakaa seitsemään ryhmään. Nämä olivat kansallisuus, kansakuntatermit, etnisyys, maantieteellinen alkuperä, uskonto ja rotu. Seitsemäs termityyppi sai tässä työssä nimen *nationhood*. Tässä tiivistelmässä siitä käytetään sanaa kansakunta. Nämä termit rikkovat esimerkeille asetettuja kriteerejä siinä mielessä, että ne viittaavat valtioihin tai valtioiden elinten toimiin. Ne laskettiin kuitenkin mukaan, koska ne olivat selkeästi oma termityyppinsä, jonka esiintyminen ei ollut sidoksissa valtioiden virallisiin nimiin. Kansakunta tässä tapauksessa viittaa termiin jossa valtioon viitataan käyttämällä normaalisti kansalaisiin viittaavaa sanaa. Esimerkiksi Venäjän valtion tai valtion elimen toimia kommentoiva fraasi käyttäisi sanoja *the Russians*, sanojen *Russia* tai *Russian state* sijasta. Ne termit, jotka muuten selvästi sopivat asetettuihin kriteereihin, mutta joita ei voitu selkeästi jakaa mihinkään edellämainituista ryhmistä, luokiteltiin ryhmään muut termit.

Uskontoa lukuun ottamatta kaikkien kansanryhmätermien osuus laski tutkimusaineistossa. Uskonnonkin kohdalla nähty nousu oli pieni, näkyen vain suhteellisissa numeroissa. Normalisaation jälkeiset numerot osoittivat uskontotermien suhteellisen määrän nousseen 7,3%, absoluuttisissa numeroissa puolestaan näkyi yhden termin lasku tekstikokoelmien välillä. Kansallisuustermejä oli vuoden 1976 tekstinäytteissä 105 kappaletta ja vuoden 2009 teksteissä 91 kappaletta. Kaiken kaikkiaan kansallisuustermien suhteellinen määrä laski 4,8%. Kansakuntatermejä oli ensimmäisessä kokoelmassa 38 kappaletta, mutta vain neljä toisessa, suhteellisesti tämä oli 88,4 prosentin lasku. Etnisyyteen perustuvien termien määrä putosi 54 termistä 26 termiin, näiden termien kohdalla lasku oli 47,1%. Maantieteelliseen alkuperään perustuvien termien määrä laski 80%, 22 termistä 4 termiin. Rotupohjaisten termien määrä oli ensimmäisessä näytteessä 48, mutta toisessa vain 3. Lasku tässä ryhmässä oli 93,1%. Muut tai luokittelemattomat termit vähenivät myös, vuoden 1976 pääkirjoituksissa niitä oli 27 kappaletta, mutta vuoden 2009 teksteissä vain 12. Näiden termien suhteellinen määrä putosi 51,2%. Kaikista kansanryhmätermeistä suurimman ryhmän muodostivat ensimmäisessä tekstikokoelmassa kansallisuustermit, joita oli 31,2% kaikista esimerkeistä. Seuraavaksi suurin ryhmä olivat etnisyystermit 16,0 prosentilla ja rotutermit, joita oli 14,2% termien koko määrästä. Vuoden 2009 pääkirjoituksissa sen sijaan näkyy selkeä muutos aiempaan näytteeseen verrattuna. 50,3% kaikista kansanryhmätermeistä on nyt kansallisuustermejä, toiseksi suurin

ryhmä ovat uskontotermit, joita on 22,7%. Rotutermit sen sijaan ovat tämän tekstikokoelman pienin ryhmä, niitä on vain 1,7% kaikista esimerkeistä. Tutkimusmateriaalissa nähdään siis selvä siirtymä pois rotu- ja etnisyysterminologiasta kansakuntatermien ja rotutermien vähentyessä huomattavasti. Näiden tilalle ovat numeroista päätelleen tulleet lähinnä kansallisuustermit ja uskontotermit. Tämä muutos on myös ideologisten kielenmuutospyrkimysten mukainen siinä mielessä, että siirtymä on tapahtunut kohti neutraalimmaksi koettuja termejä, virallista kansallisuutta ja monikansallista uskontoa.

Kategorioiden sisällä muutokset olivat sen sijaan epätasaisia. Joka kategoriassa osa termeistä väheni ja osa lisääntyi. Esimerkiksi uskontotermistössä tämä muutos näkyi ensimmäisessä tekstikokoelmassa suosittujen termien vähenemisenä ja uusien termien ilmestymisenä myöhemmässä tekstikokoelmassa. On kuitenkin otettava huomioon, että näissä kategorioissa monet yksittäiset termit olivat tiukasti sidoksissa tekstien aihepiireihin ja absoluuttiset numerot saattoivat olla hyvinkin pieniä. Näin ollen yhdenkin sopivaa aihetta käsittelevän pääkirjoituksen lisääminen tai poistaminen aineistosta voisi vaikuttaa tuloksiin.

Päätelmät ja seuraavat askeleet

Tutkimuksen tulokset ovat kahdenlaiset. Ensinnäkin tutkimukseen valittu sangen kokeellinen lähestymistapa, jossa väestöryhmiä koskevaa terminologiaa käsiteltiin sen edustamien kategorioiden kautta, tuotti tulosta. Vertailemalla tästä näkökulmasta kahta eri tekstinäytekokoelmaa voitiin näiden kokoelmien edustamassa standardibrittienglannissa nähdä selviä siirtymiä. Sen sijaan, että olisi paneuduttu vain yksittäisten termien määriin, tässä tutkimuksessa voitiin kartoittaa aineistossa näkyviä trendejä ja yleisiä muutoksia. Toisekseen, näissä trendeissä nähdyt siirtymät tukevat olettamusta, että ideologiset pyrkimykset muuttaa englannin kieltä tasa-arvoisemmaksi nähtyyn suuntaan, ovat vaikuttaneet itse kieleen. Tämä oli hyvin selvää varsinkin sukupuolitermistössä, jossa maskuliiniset yleistermit vähenivät huomattavasti tutkimuksen ajanjaksona.

Tulokset osoittavat, että tutkimuksen kohteena olevissa standardibrittienglannin näytteissä tapahtui selkeä muutos. Tämä muutos ei kuitenkaan ole täydellinen, eikä välttämättä koskaan tule täydellinen olemaankaan. Standardikieli ei ole muutoksille immuuni ja tässä tutkimuksessa nähtiin yksi esimerkki lähimenneisyydessä tapahtuneesta standardikielen muutoksesta. Kielen vaikutuksesta todellisuuteen kiistellään yhä, mutta tätä kielen muutosta ajavat eteenpäin käsitykset kielestä, eivät käsitykset todellisuudesta.

Tämä tutkimus oli luonteeltaan suurelta osin tuntemattoman maaston kartoitusta, sekä metodeiltaan että sisällöltään. Näin ollen se on hedelmällinen lähtökohta lisätutkimukselle ja lisäkysymyksille. Onko tässä tutkimuksessa havaittu kielen muutos pysyvä? Tätä voisi selvittää toistamalla tutkimuksen metodologian käyttäen lähteenä saman sanomalehden myöhempiä numeroita. Onko tämä muutos uusi trendi, vai onko kieli ollut siirtymässä tähän suuntaan jo aiemmin? Tähänkin kysymykseen voisi vastata toistamalla metodologian, mutta käyttämällä tällä kertaa vanhempaa aineistoa. Entä näkyykö tämä sama muutos toisissa englannin kielen varianteissa? Esimerkiksi puhekielessä, tai amerikanenglannissa. Laajoista kielinäytekokoelmista voitaisiin kartoittaa tämän tutkimuksen löytämiä termejä tietokoneavusteisen kielitieteen keinoin. Suurentamalla tekstinäytekokoelmien kokoa voitaisiin huomattavasti vähentää sattuman osuutta tuloksissa ja vahvistaa – tai kumota – tämän tutkimuksen esille tuomia siirtymiä.

Tuntemattoman kartoitus on sellaisenaan palkitsevaa ja mielenkiintoista, mutta eniten siitä hyötyvät tulevat kulkijat, joiden ei tarvitse itse etsiä samoja polkuja. Tämä työ toi itsessään tuloksia, mutta moninkertaisesti siitä tulee hyötymään se, joka sen kartakseen ottaa ja jatkaa entistä pidemmälle.