

The Development and Use of Gender Language in Contemporary English – A Corpus Linguistic Analysis

Prepared for the Committee on Bible Translation by:

*Collins Dictionaries, a department of Collins Language Division
Westerhill Road, Bishopbriggs, Glasgow*

*Foreword to the report by Professor Douglas Moo,
Chair of the Committee on Bible Translation*

Prior to the update of the New International Version of the Bible (NIV) for 2011, all previous Bible translation efforts have been hampered by the lack of accurate, statistically significant data on the state of spoken and written English at a given time in its history. Beyond appealing to traditional style guides, all that translators and stylists have been able to do is rely on their own experiences and others' anecdotal evidence, resulting in arguments such as, "I never see anybody writing such-and-such," or "I always hear such-and-such," or "Sometimes I read one thing but other times something else."

As part of the review of gender language promised at the announcement of the latest update to the NIV on September 1, 2009, the Committee on Bible Translation sought to remove some of this subjectivity by enlisting the help of experts. The committee initiated a relationship with Collins Dictionaries to use the Collins Bank of English, one of the world's foremost English language research tools, to conduct a major new study of changes in gender language. The Bank of English is part of Collins' corpus holdings, which contain more than 4.4 billion words drawn from text publications and spoken word recordings from all over the world.

Working with some of the world's leading experts in computational linguistics and using cutting-edge techniques developed specifically for this project, the committee gained an authoritative, and hitherto unavailable, perspective on the contemporary use of gender language – including terms for the human race and subgroups of the human race, pronoun selections following various words and phrases, the use of "man" as a singular generic and the use of "father(s)" and "forefather(s)" as compared to "ancestor(s)." The project tracked usage and acceptability for each locution over a twenty-year period and also analyzed similarities and differences across different registers and varieties of English: for example, UK English, US English, written English, spoken English, and even the English used in a wide variety of evangelical books, sermons and internet sites.

Research of this type is just one tool in the hands of translators, and, of course, it has no bearing on the challenge of preserving transparency to the form and structure of the original text. But, since its first publication in 1978, the NIV has always aimed not only to offer transparency to the original documents but also to express the unchanging truths of the Bible in forms of language that modern English speakers find natural and easy to comprehend. And this is where a tool like the Bank of English comes into its own.

The pages that follow provide an insight into the wealth of information that emerged from this program of research and the methods that were employed. We hope it will be of interest to scholars and lay people alike as they familiarize themselves with the updated text of the NIV.

Professor Douglas Moo, Chair of the Committee on Bible Translation, September 2010.

Contents

1. Introduction	4
2. Executive summary of findings	5
2.1. Generic pronouns and determiners	5
2.2. Mankind, man and synonyms	6
2.3. Forefather, ancestor and father	8
3. Background to the study	9
3.1. The corpus	9
3.2. Anaphora resolution grammar	9
3.3. Project plan and methodology	11
4. Detailed description of findings	15
4.1. Generic pronouns	15
4.1.1. General English – written	15
4.1.2. General English – spoken	19
4.1.3. US English – written	24
4.1.4. US English – spoken	28
4.1.5. Evangelical English	32
4.1.6. Other observations	36
4.1.7. Commentary/interpretation	36
4.2. Mankind, man and synonyms	38
4.2.1. General English – written	38
4.2.2. General English – spoken	40
4.2.3. US English – written	42
4.2.4. US English – spoken	44
4.2.5. Evangelical English	46
4.2.6. Follow-on pronoun use with man, mankind and humanity	48
4.3. Forefather, ancestor and father	49
4.3.1. General English – written	49
4.3.2. General English – spoken	51
4.3.3. US English – written	52
4.3.4. US English – spoken	53
4.3.5. Evangelical English	54
4.3.6. A note on '(fore)mothers'	55

1. Introduction

This report presents the results of a significant original study of gender language in English since 1990. Three areas of usage are described:

- A. **generic pronouns and determiners:** the types of pronouns and determiners that are used to refer to indefinite pronouns (such as *someone*, *everybody* and *one*) and non-gender-specific nouns (such as *a person*, *each child* and *any teacher*):
 - i. masculine (*he*, *his*, *himself*, etc.);
 - ii. feminine (*she*, *her*, *herself*, etc.);
 - iii. plural/gender-neutral (*they*, *them*, *one*, *themselves*, etc.);
 - iv. alternative forms (*s/he*, *him or her*, *his/her*, etc.)
- B. **mankind, man and synonyms:** the use of the terms *man*, *mankind*, *humankind*, *humanity*, *humans*, *human beings*, *the human race* and *people* in the sense 'the human species' or 'humans collectively'.
- C. **forefather, ancestor and father:** the use of the terms *forefather(s)*, *ancestor(s)* and *father(s)* in the sense 'a person/people from whom one is descended' or 'the founder(s) of a movement/nation etc.'

The study was undertaken using parts of Collins' 4.4 billion word corpus holdings and was facilitated by state-of-the-art computational tools described in section 3. The study draws from balanced sub-corpora of general written English, general spoken English, US written English and US spoken English, as well as an additional custom-built corpus of Evangelical English assembled from a wide variety of evangelical books, sermons and internet sites. As part of the study, the research team also sought insights into southern American English, assembling a substantial sub-corpus using extracts from the Houston Chronicle. On analysis, however, it was found that the Houston Chronicle data did not differ in any material way from the broader holdings in the US spoken sub-corpus. For this reason, and also because the sub-corpus was so narrowly based, it was not used by the committee in its work. It has therefore been omitted from this report.

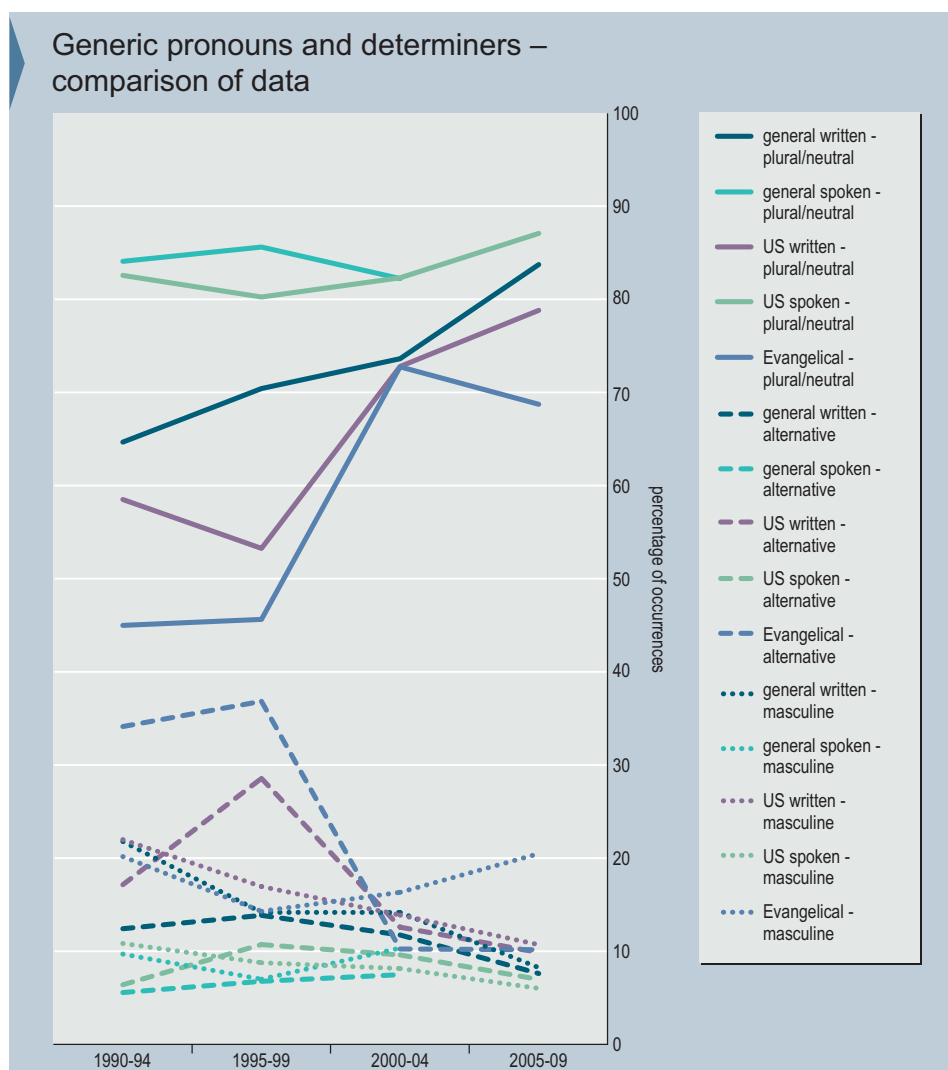
2. Executive summary of findings

2.1 Generic pronouns and determiners

The chart below compares the proportions of generic masculine, plural/neutral and alternative pronouns and determiners in general written English, general spoken English, US written English, US spoken English and Evangelical English, since 1990. (Percentages of feminine generics are not shown in this chart as they are consistently low; they are given in the detailed findings in section 4.)

It is evident that, in all the varieties of English analyzed, plural/neutral pronouns and determiners account for the majority of usages. Diachronically, the pattern is that plural/neutral pronouns and determiners have become more frequent since 1990; this increase is particularly marked in general and US written English and in Evangelical English. There has been a slight decrease in gender-neutral uses since 2005 in Evangelical English, and a corresponding increase in masculine uses, but gender-neutral pronouns and determiners are still over three times more frequent than masculine ones in current Evangelical English.

The other main pattern is the decrease in alternative forms such as 'him or her', particularly in Evangelical English, and also in US written English. These were quite frequent in the 90s in Evangelical English, particularly following non-gender-specific nouns (*a person, every child*, etc.). However, in current Evangelical English they account for only 10% of occurrences.



2.2 Mankind, man and synonyms

Because of the large amount of data for this part of the study, the following charts summarize two key areas:

- A. synchronically, the relative frequency of *man*, *mankind*, *humankind*, *humanity*, *the human race*, *human beings*, *humans* and *people* in the most recent slice of each corpus (05-09) (general spoken English is not included here since there is no 05-09 section for this corpus);
- B. diachronically, the frequency of *man* and *mankind* since 1990 in all the corpora.

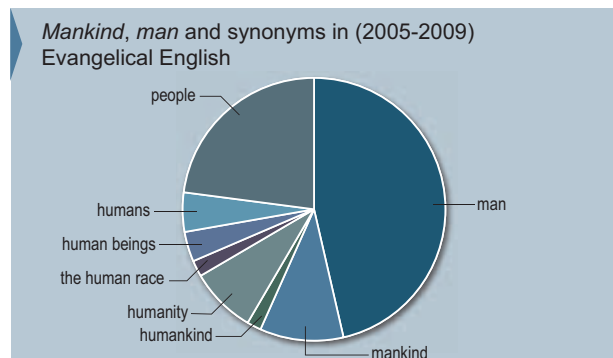
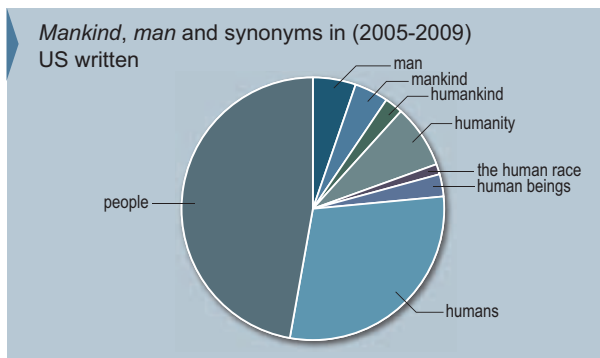
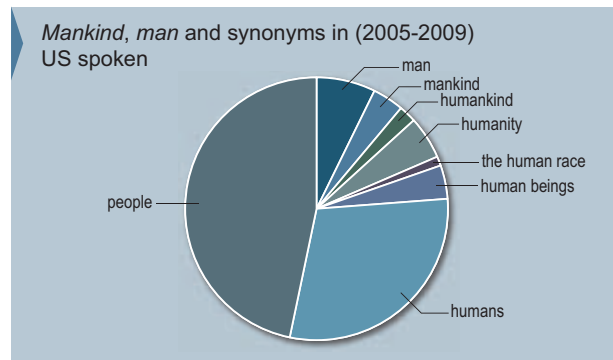
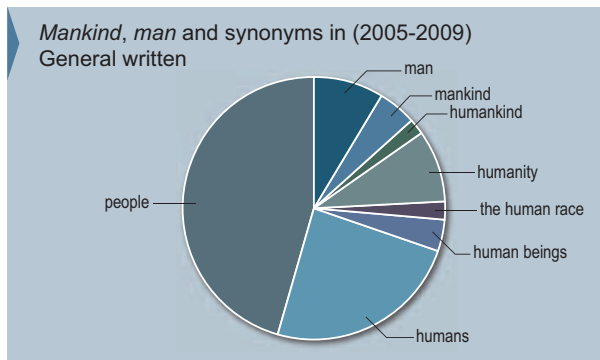
The changes in patterns of frequency of the other synonyms are shown in the detailed findings.

Synchronic pattern

It is evident that in all the corpora except Evangelical English, *people* is by far the most frequent synonym, followed by *humans*. *People* and *humans*, however, are much looser synonyms for the human species as a whole, and it is often difficult to determine whether they refer to all humans, or to a smaller subset. Although examples of *people* and *humans* referring to specific groups were carefully excluded from the study (see section 3.3, “the analysis”) it became clear that, even after screening, citations involving *people* and *humans* were often ambiguous in this respect. Thus, in the following analysis, figures excluding *people* and *humans* – and focusing instead on the more precise *man*, *mankind*, *humankind*, *humanity*, *the human race* and *human beings* as alternatives that naturally occupy this more formal register – have also been provided.

Of these alternatives, *man*, *humanity* and *mankind* are the most frequent synonyms in the general written English, general spoken English, US written English and US spoken English corpora. Of all the synonyms, *man* accounts for between 4.0% and 8.7% of citations, *humanity* accounts for between 5.0% and 9.0% of citations, and *mankind* accounts for between 3.5% and 4.8% of citations depending on the corpus analyzed. *Humankind* is very infrequent, accounting for approximately 1% or less of all usages. *Human beings* and *the human race* are also infrequent.

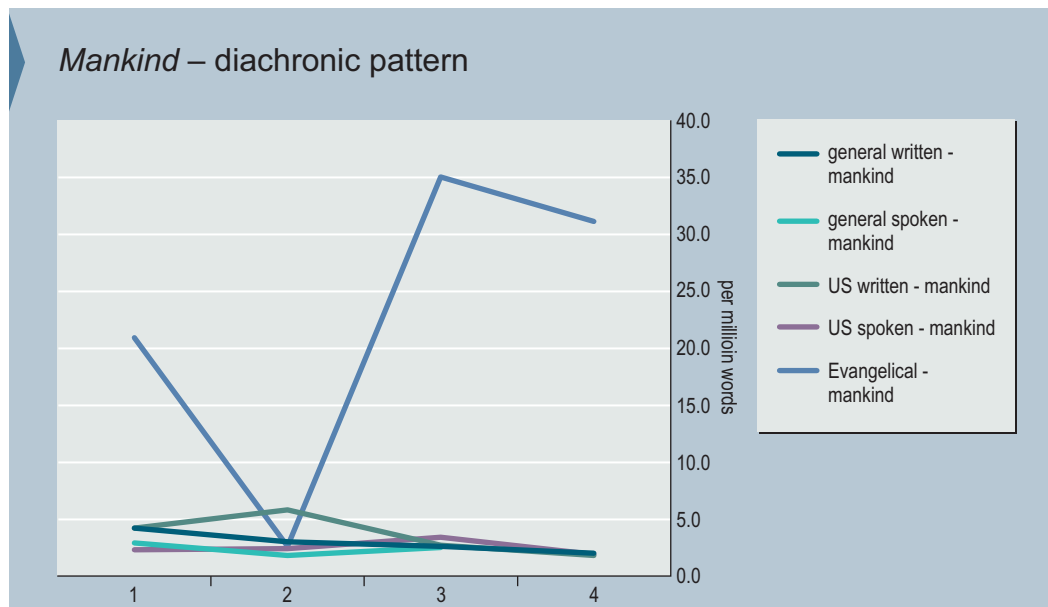
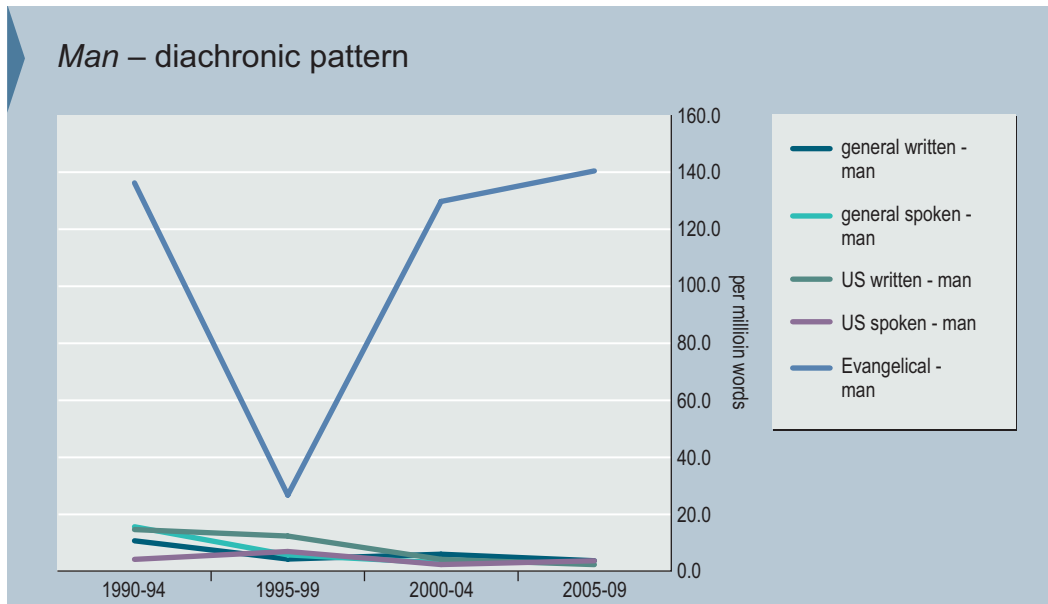
As the chart for Evangelical English shows, the distribution of synonyms is markedly different in Evangelical English, where *man* is the most frequent, accounting for nearly half of occurrences. *People* accounts for about a quarter, while *mankind* and *humanity* each make up approximately 10% of the overall usages. *Humans* is much less frequent than in the other corpora. *Humankind*, *human beings* and *the human race* are, as in the other corpora, relatively infrequent.



Diachronic pattern for man and mankind

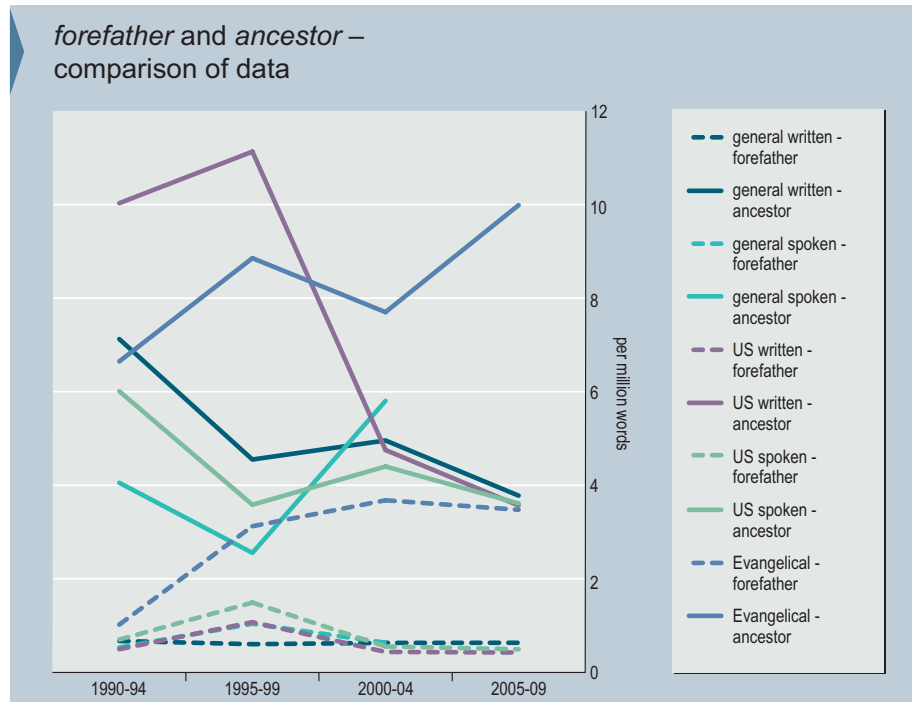
The charts below show the diachronic frequency of *man* and *mankind* in each corpus. In all the corpora except Evangelical English, *man* and *mankind* have become slightly less frequent (with some fluctuations) since 1990. The frequencies of both *man* and *mankind* have tapered off to a very similar level in all the corpora: approximately 3 per million words for *man*, and approximately 2 per million words for *mankind*.

Again, the pattern in the Evangelical corpus is quite different. Firstly, the frequency with which all of the synonyms tracked in this part of the study appear in the Evangelical corpus is markedly higher than it is in the other corpora. This fact is most likely due to the nature of the subject matter addressed in Evangelical books and sermons. Secondly, there was a significant dip in the frequency of both *man* and *mankind* in 95-99, followed by an increase since the turn of the century. Possible reasons for this dip are discussed in more detail in section 4 below.

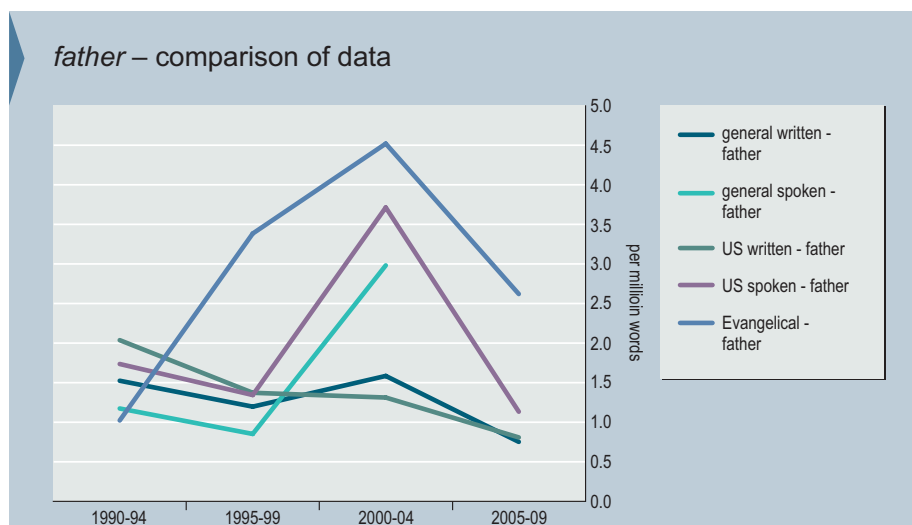


2.3 Forefather, ancestor and father

The chart below shows the frequency of *forefather* and *ancestor* in each corpus since 1990. The frequencies have fluctuated, but it is evident that *ancestor* is significantly more frequent than *forefather* in each corpus and each period. The frequency of *forefather* is higher in Evangelical English than in the other corpora, but still much less frequent than *ancestor* in Evangelical English.



The frequency of *father* since 1990 is shown separately in the chart below. It is useful to show *father* separately from *forefather* and *ancestor* because the usage of *father* is slightly different: most of the citations are of the form ‘the father(s) of’, meaning ‘the founder(s) of’ (as in ‘The ancient Greek philosopher Anaximenes has been called **the father of science**’). The chart shows that the use of *father* in this sense has fluctuated since 1990; it became particularly frequent in 00-05 (especially in Evangelical English, US spoken English and general spoken English), but has since decreased in frequency.



3. Background to the study

3.1. The corpus

In this section, the databases, or 'corpora', of English texts which form the basis of the study are described. These were created from selected parts of Collins' 4.4 billion word corpus holdings, and supplemented and refined in order to ensure breadth and representativeness¹. Five corpora were created:

1. *General written English*,

This corpus contains nearly **2 billion words**, covering a variety of written genres (including newspapers, magazines, fiction, reports, and journals) from a variety of English-speaking countries (including the United States, the United Kingdom, Canada, Australia and South Africa).

2. *General spoken English*

This **115 million word** corpus contains transcriptions of radio and television programmes from US, Canadian and UK sources, as well as transcriptions of recorded conversational speech.

3. *US written English*

This contains the **380 million words** of US written English extracted from the General written corpus. It includes samples from American newspapers and magazines such as the *Wall Street Journal*, *Time Magazine*, the *Washington Times*, *USA Today*, as well as numerous books and other written material.

4. *US spoken English*

This contains the **72 million words** of US spoken English extracted from the General spoken corpus. It includes transcriptions of *Voice of America*, *PBS NewsHour* and *National Public Radio*.

5. *Evangelical English (all written)*

This custom-built corpus contains approximately **113 million words** of texts from a wide variety of evangelical books, sermons and internet sites.

Each corpus was divided into four 5-year periods: 1990-1994, 1995-1999, 2000-2004 and 2005-2009. These were made as equal as possible, although the varying availability of materials from different periods means that some slices are smaller than others. However, since all the analysis is based on proportions of types of usage, or frequency per million words, these variations do not affect results.

3.2. Anaphora resolution grammar

The analysis of generic pronouns was aided considerably by the development of an anaphora resolution grammar which was designed to track the relationship between pronouns/determiners and antecedents in sentences such as:

- A. '... we live in a tolerant society that protects the right of **any human being** to believe whatever **he or she** wants to believe...'
- B. 'Don't hesitate to listen to **somebody's** conversation. Listen to the words **they** use and the way **they** phrase them and maybe you can pick up an idea.'

¹ Throughout this report word-counts include numbers and punctuation marks as well, i.e. correspond to what are known in linguistics as 'tokens'.

The methods in developing this grammar, and the resulting improvements to the study, are now summarized.

Methods

1200 sentences were manually tagged as follows:

A. anaphoric and generic. For example -

'I'm still hoping **someone** will change **their** mind and sort it out.'

- where 'someone' is a person of unspecified gender, and 'their' refers to 'someone'.

B. anaphoric but not generic, i.e. a specific person is referred to. For example -

'But it is a very moving story because really she was **someone** who had a very tragic life, who looked and looked for love and **she** found it with Nehru.'

- where 'she' refers to 'someone', but 'she' is a known person in the story.

C. not anaphoric. For example:

'Tirin's letters were very funny and would have convinced **anyone** that they referred to real emotions and events, but the physics in **them** was dubious.'

- where 'them' refers to 'letters' rather than 'anyone'.

For the purposes of the study, only examples which are both anaphoric and generic (i) are relevant. However, the anaphora resolution grammar was only expected to filter out examples which are not anaphoric (iii); it was not designed to filter out those which are anaphoric but not generic (ii). These were later removed by manual analysis.

Half of the tagged sentences were used as development data, which were analyzed to identify relevant patterns which could be applied to the grammar. The other half were used to test the final version of the grammar after development had finished.

One aspect of the grammar development was the addition of semantic tags to nouns, so that nouns denoting a person of unspecified gender (e.g. *person*, *human*, *teacher*, *doctor*, *sinner*, etc.) could be identified and distinguished from nouns denoting males (e.g. *man*, *husband*, *father*, etc.), females (*woman*, *sister*, *nun*, etc.), and inanimate concepts or objects (*time*, *hand*, *table*, etc.) These tags were based on automatically generated lists from the lexical database WordNet, and were manually refined. This retagging of nouns was invaluable to the study, as it meant that one could automatically search for citations with, for example, *each* followed by a person of unspecified gender, and find all the instances of *each teacher*, *each person*, *each manager*, and so on. Without this tool it would have been impossible to distinguish such citations from irrelevant (and frequent) ones such as *each day* and *each time*. These tags were also used as part of the grammar as it was found that citations with nouns denoting males or females in close proximity to words such as 'someone' were usually not generic.

Other aspects of the grammar included the identification of the optimum number of words between pronoun/determiner and antecedent, and the elimination of citations with proper nouns or male/female person nouns within a given span.

Resulting improvements

Most importantly, the anaphora resolution grammar yields a higher proportion of positive (relevant) examples than would have been possible without it. Samples which were manually analyzed in the feasibility study for the project were compared with samples using the grammar: the latter retrieved over double the proportion of relevant examples, and also allowed searches across a broader span of words. The results were also checked to ensure that the grammar did not in any way bias the results. Furthermore, the grammar was developed so that it would improve the interface for researchers, and make the analysis easier and quicker.

3.3. Project plan and methodology

Preparatory stage

As described in 3.1 and 3.2, the preparatory stage of the project involved creating the corpus and designing the anaphora resolution grammar. The parameters of each area of the study were also defined:

A. generic pronouns and determiners

The following 17 pronouns, determiners and nouns – *someone, somebody, everyone, everybody, anyone, anybody, no-one/no one, nobody, each, any, every, no, whoever, one, person, human* and *individual* – were searched where one of the following pronouns/determiners, or a combination of them, appeared within 10 words – *he, she, his, her, him, hers, himself, herself, they, them, their, theirs, themselves, theirselves, theirsself, themself, one, one's, oneself*.

B. *mankind, man* and synonyms

Various printed and online thesauri were examined, and the following set of synonyms meaning ‘the human race’ or ‘humans collectively’ was identified: *man, mankind, humankind, the human race, humanity, humans, human beings* and *people*². Since several of these words have more than one meaning, corpus searches were refined in order to find only instances which were not preceded by an adjective (*happy, lovely*, etc.), determiner (*a, the, some*, etc.), possessive pronoun (*my, their*, etc.), participle (*condemned, dying*, etc.) or noun modifier (*family, business*, etc.). This eliminated irrelevant citations such as:

‘I’d always been one of **those people** who regard themselves as inside life...’

‘... so thrilled to celebrate the joy of their **common humanity**.’

‘He’s a **charming man**, of course, but totally incompetent.’

‘He was also a devoted **family man**.’

However, there is one determiner – *all* – which often occurs in relevant citations, e.g.

‘So we cannot long for their salvation as we long for the redemption of **all humanity**.’

Therefore each search was supplemented with an additional search for ‘all’ + target word. The same search parameters were applied to all the synonyms to ensure consistency of results.

C. *forefather, ancestor* and *father*

Because *forefather* and *ancestor* are predominantly used in the target sense ‘ancestor’ (including figurative extensions) or ‘founder’, special searches did not need to be developed for these. *Father*, on the other hand, is of course predominantly used in the sense ‘male parent’ (and, in the Evangelical corpus, in the sense ‘God’ or ‘priest’). A sample of citations with *father* was analyzed to determine the most useful search which would capture as many relevant examples as possible within a manageable sample size. It was found that almost all relevant citations were of the form ‘our/the...fathers’ and ‘the father(s) of...’, e.g.

‘He protested that his preaching agreed with that of **the fathers**, and said that he did not object to certain traditions...’

‘Or have we forsaken our birthrite, the perennial wisdom of **our fathers**, for a mess of modern pottage?’

‘The ancient Greek philosopher Anaximenes has been called **the father of science**.’

² Initially we planned to include *personkind* and *Homo Sapiens*. However, *personkind* appears only six times in all of the corpora, while *Homo Sapiens* is only ever used in a scientific sense and not as a genuine synonym of *mankind*.

Once the searches had been agreed upon, preliminary samples of citations for all three areas of usage were manually analyzed in order to determine how relevant citations were to be distinguished from irrelevant ones, and how researchers would be asked to identify these. Briefs were written and researchers were tested, trained and monitored in order to ensure the highest standard of analysis.

Creation of slices of data for analysis

The corpora were divided into the following 19 slices for analysis:

- | | |
|--------------------------------------|-----------------------|
| 1. General written 90-94 | 11. US written 05-09 |
| 2. General written 95-99 | 12. US spoken 90-94 |
| 3. General written 00-04 | 13. US spoken 95-99 |
| 4. General written 05-09 | 14. US spoken 00-04 |
| 5. General spoken 90-94 | 15. US spoken 05-09 |
| 6. General spoken 95-99 | 16. Evangelical 90-94 |
| 7. General spoken 00-04 ³ | 17. Evangelical 95-99 |
| 8. US written 90-94 | 18. Evangelical 00-04 |
| 9. US written 95-99 | 19. Evangelical 05-09 |
| 10. US written 00-04 | |

The analysis

A. Pronouns and determiners

Each of the 17 indefinite pronouns, determiners and nouns (*someone, somebody, everyone, everybody, anyone, anybody, no-one/no one, nobody, each, any, every, no, whoever, one, person, human* and *individual*) was searched in each slice of each corpus. Where the results were fewer than 100, all citations were analyzed; where there were more, random samples of 100 were taken. Each sample was saved as a text file, and researchers analyzed and tagged every citation. The researchers were asked to identify relevant generic citations and to tag these according to whether the pronoun/determiner used was:

- Masculine (*he, his, him* or *himself*);
- Feminine (*she, her, hers* or *herself*);
- Alternative with masculine first (*he* or *she, his* or *her, him* or *her, his* or *hers, himself* or *herself, he/she, him/herself, himself/herself, him/her, his/hers* or *his/her*);
- Alternative with feminine first (*she* or *he, her* or *his, her* or *him, hers* or *his, herself* or *himself, she/he, her/himself, herself/himself, her/him, hers/his* or *her/his*);
- Plural/gender-neutral (*they, them, their, theirs, themselves, theirselves, theirself, themself, one, one's, oneself*);
- the form *s/he*.

Researchers were asked to identify as relevant only those cases which were felt to be genuine instances of generic pronouns/determiners. That is, citations such as the following were excluded:

'If it weren't for the fact that your hatchet had been taken it would be a good guess that **someone** broke in, looking for what **he** could pick up.'

'So, he believes, is the one [rule] that debar **anyone** who has played first-class cricket as a home player in **his** native land in the previous 12 months.'

3 Note that there is no slice for General spoken English 05-09. All the material from this slice is US English, therefore the same as US spoken 05-09.

'... mean steal a baby to give it away don't be crazy why did they **whoever** left it on our doorstep maybe **she**'ll come back for it like Carmella did and for six months we...'

Violent crime is often associated with men; most cricketers are men; and most people who leave babies on doorsteps are women. Therefore, these citations were not considered to be genuine examples of generic masculine or feminine pronouns/determiners, but rather cases where the writer/speaker had a person of a specific gender in mind. On the other hand, citations such as following were tagged as relevant:

'We should teach people to be kind to those we consider a little "odd" in our community—in fact, to be kind to everyone and let **no one** consider **himself** an "outsider", which is the first sign of trouble.'

'However, if room sharing is necessary, do your best to encourage compromise, if necessary working out some kind of timetable to enable **each child** to have the room to **herself** for a specific period.'

These cases *are* genuine generics: we know that not only men are outsiders, and that not all children who share rooms are girls. Although it is of course impossible to know for certain what a writer/speaker was thinking, and there were many ambiguous examples, this use of real-world knowledge was essential to obtaining results which would reflect the use of genuine generics.

In addition, researchers were asked to identify citations where a neutral pronoun was used even though the gender was known, as in:

'Nobody, it seemed, wanted to stand out. Finally, **someone** put **their** hand up and Sally gave **him** the floor.'

'If an **individual** is pregnant, **they** need to do everything they can to get early and ongoing prenatal care.'

These citations appear to be evidence of the spread of plural *they* even when it is not being used in a deliberately non-sexist way. This usage is discussed in section 4.1.6 below.

Researchers were also asked to highlight any citations which they felt were not representative of the language under investigation, such as historical or biblical quotations, or translated phrases. These were checked and excluded from the final figures.

Once the files had been tagged, results were counted and entered into spreadsheets, and, for each slice, the percentage of masculine, feminine, alternative and gender-neutral occurrences was calculated. The results of these calculations are presented and discussed in section 4. Early in the analysis, researchers were given the same file to analyze so that consistency could be measured. There was a high level – 92% – of consistency between researchers. Areas of inconsistency were addressed; these were mostly related to ambiguous cases where it was not clear whether or not a citation referred to a person of a known gender.

B. *Mankind, man* and synonyms

Each of the 23 slices of the corpora was searched for *man, mankind, humankind, the human race, humanity, humans, human beings* and *people*. In some cases, where the number of results was relatively low (100 or fewer), all citations were analyzed. In other cases, random samples were taken. Sample size varied depending on the word. Samples of 100 (plus the additional results for 'all' + target word) were sufficient for *mankind, humankind, humans, human beings, the human race* and *humanity*, which contained fairly high proportions of relevant examples. Higher sample sizes of between 200 and 300 were necessary for *man* and *people*, as these contained a lot of irrelevant citations.

Samples were given to researchers, who were asked to identify all instances of the relevant sense – ‘the human race’ or ‘humans collectively’ – and to tag as negative all instances of irrelevant senses such as ‘particular group of people’ (in *people*, *human beings* and *humans*), ‘quality of being human’ (in *humanity*) and ‘male person’ (in *man*). They were also asked to identify historical and biblical quotations, idioms (such as *crime against humanity* and *man’s best friend*), and proper nouns (such as the organization ‘Habitat for Humanity’). These were not included in the final figures.

Once the citations were analyzed, figures for relevant citations were entered into a spreadsheet and frequency per million words was calculated based on sample size and the size of the corpus slice⁴. By analyzing frequency per million words rather than raw figures it was ensured that comparisons could be made irrespective of the varying sizes of the corpus slices. Percentages of overall occurrences were also calculated.

C. *Forefather, ancestor and father*

Either full results or (if more than 100) random samples of 100 citations of *forefather* and *ancestor* and 200 citations of *father* were analyzed. Researchers were asked to identify all citations with the sense ‘a person/people from whom one is descended’ (including the figurative sense ‘a person from whom one is spiritually/intellectually descended’) or ‘the founder of a movement/nation etc.’, and to mark all other senses – such as ‘male parent’ in *father* and ‘early type of animal or plant’ in *ancestor* – as negative. Historical and biblical quotations, idioms and proper nouns were also identified and excluded. Citations with ‘forefathers and (fore)mothers’ and ‘fathers and mothers’ were also excluded from the final count and treated separately. As with *man* and its synonyms, frequencies per million words, as well as percentages of overall occurrences, were calculated and compared.

4 For example, if a random sample of 200 citations with *man* was extracted from a total of 500 citations, and 115 citations were found to be relevant, the extrapolated total of relevant citations was $115 \cdot (500/200) = 287.5$. If this result was from a slice of the corpus containing 16,000,000 words, then the frequency per million words was calculated as $(287.5/16,000,000) \cdot 1,000,000 = 18$ per million words.

4. Detailed description of findings

4.1 Generic pronouns

For each of the six corpora, charts showing percentages of masculine, feminine, alternative and plural/gender-neutral forms are given for each of the following:

- overall pattern – all the data combined;
- distributive pronouns (*everyone, everybody, anyone, anybody, no(-)one, nobody, whoever*);
- non-distributive pronouns (*someone, somebody* and *one*);
- determiners + nouns (*each, every, no* and *any* + person noun);
- the nouns *person, human* and *individual*.

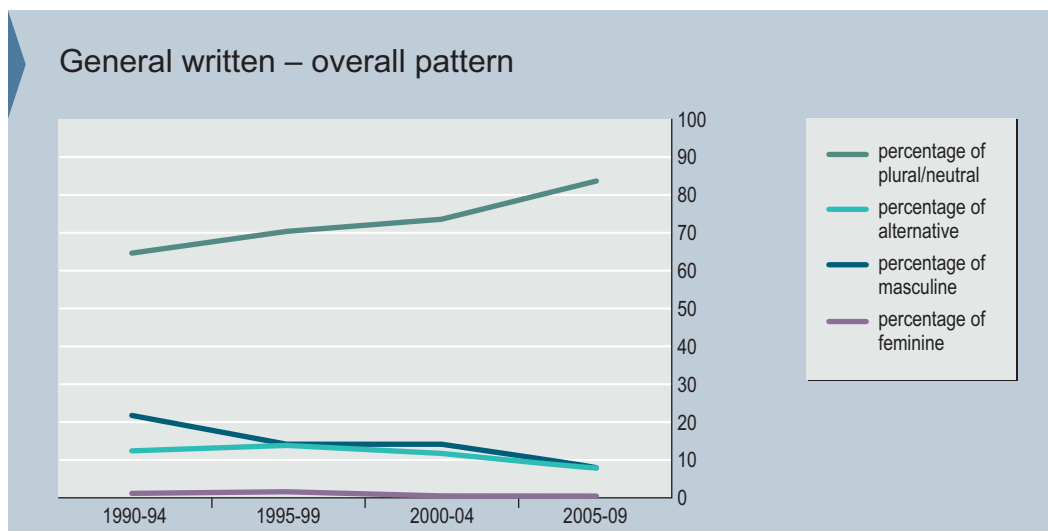
In addition, summative charts showing overall patterns of plural, masculine and alternative pronouns and determiners are given.

It should, of course, be borne in mind that the overall patterns are based on the largest sets of data; the more finely-grained analyses of different types are based on smaller subsections of data and are therefore more subject to fluctuation and variation.

4.1.1 General English – written

Overall

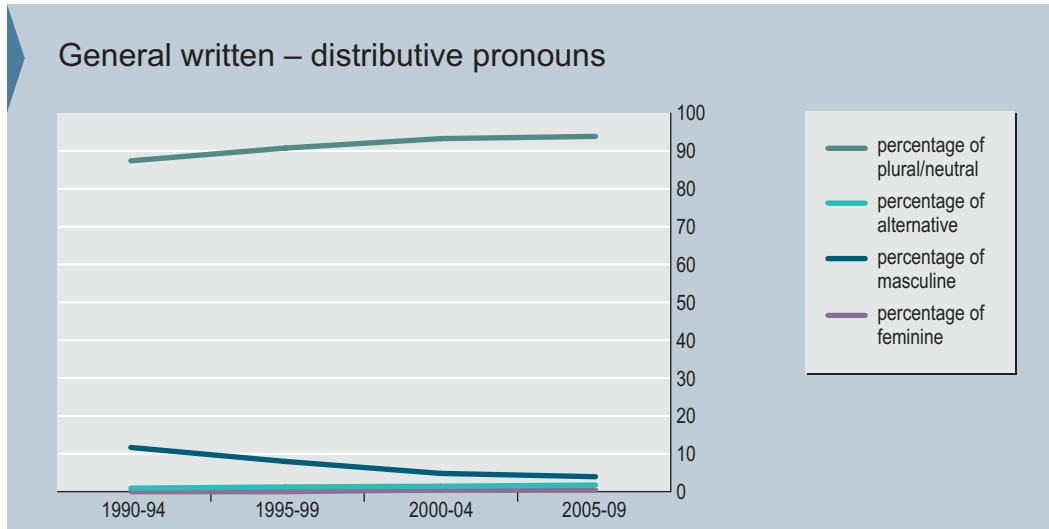
In general written English, the overall pattern is that plural/neutral pronouns and determiners have steadily increased in frequency, while masculine forms have decreased.



Distributive pronouns

Plural/neutral forms are particularly frequent when referring to distributive pronouns, e.g.

'Surely, as long as we are talking about words and not actions, **everyone** should be free to say what **they** please.' (2006)

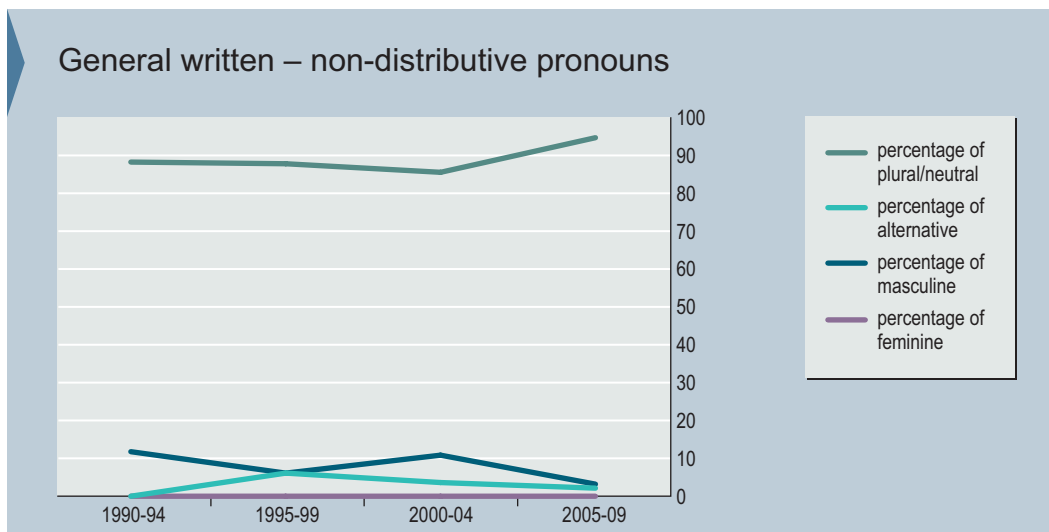


Non-distributive pronouns

Plural/neutral forms have also become more frequent with non-distributive pronouns, e.g.

'As soon as **someone** realises **they** can't pay their bills they should sit down and do their sums.'

(2006)



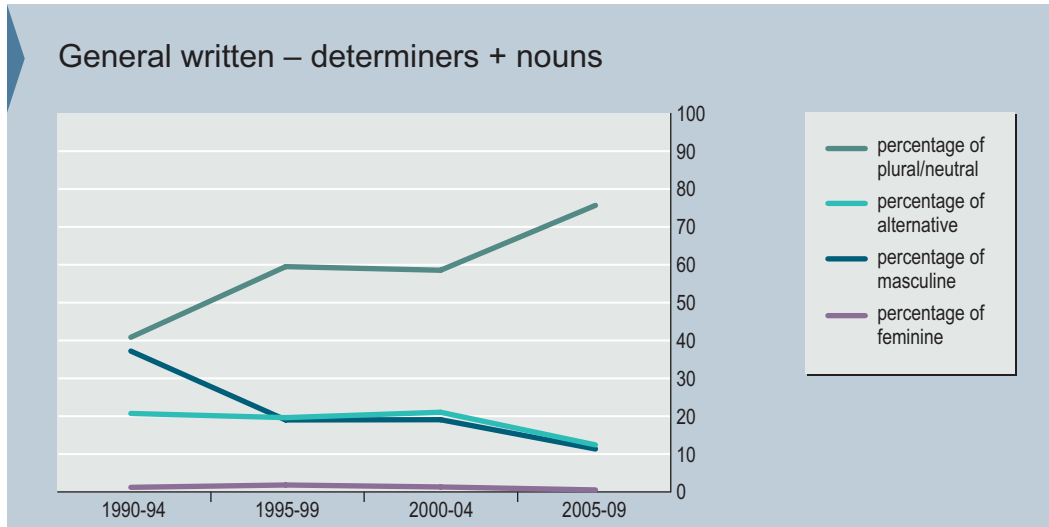
Determiners + Nouns

The pattern with non-gender-specific person nouns has changed significantly since 1990. In the 90-94 slice, plural/neutral and masculine forms were almost equally common, and examples such as the following accounted for almost 30% of citations:

'Faced with a ballot list of several candidates **each elector** is required to list **his** order of preferences.'
(1993)

Alternative forms have become much less frequent, and in current usage plural/neutral forms are much more frequent, e.g.

'Why should **any worker** have to accept violence in **their** workplace?' (2008)



Person, human and individual

A similar pattern is evident in the data for *person*, *human* and *individual*. In the 90-94 and 95-99 slices, alternative and masculine forms are quite frequent, e.g.

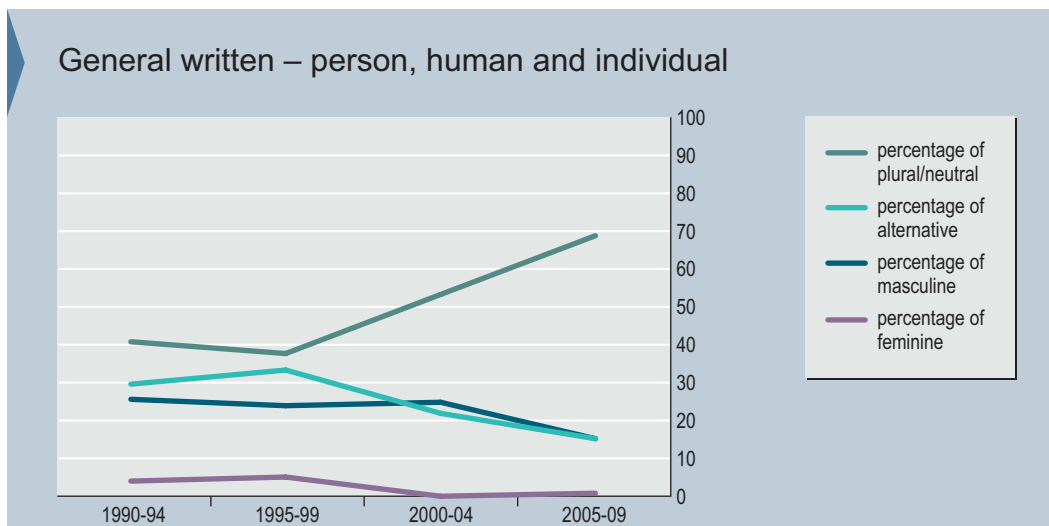
'... the struggle of a **human being** to adapt and grow and to communicate **his or her** needs to the outside world...' (1990)

'First, when a **person** accepts unconditional responsibility, **he** denies himself the privilege of "complaining" and "finding faults."' (1996)

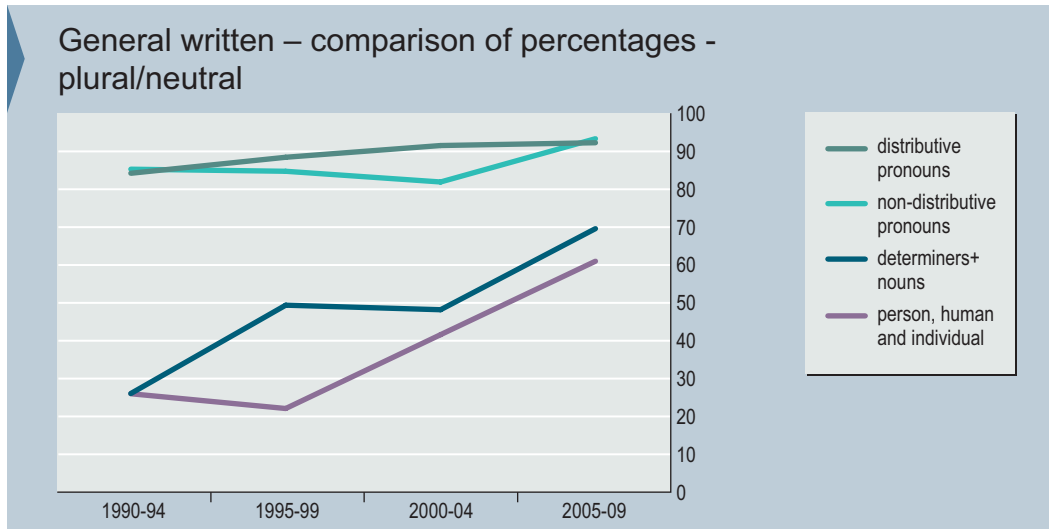
By 05-09, the plural form is much more frequent in such constructions, e.g.

'If you can identify an **individual** who metabolises nicotine faster you can treat **them** more effectively.' (2006)

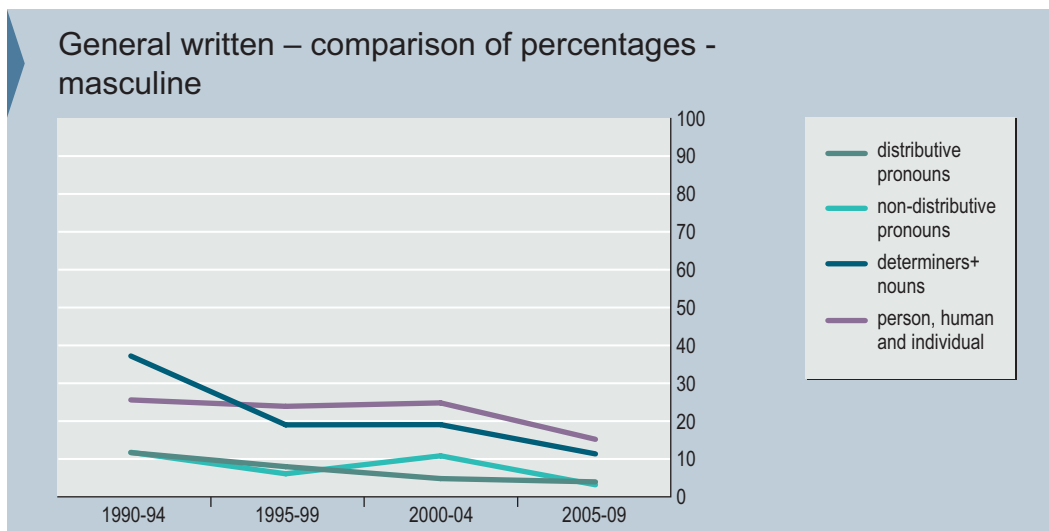
Comparison



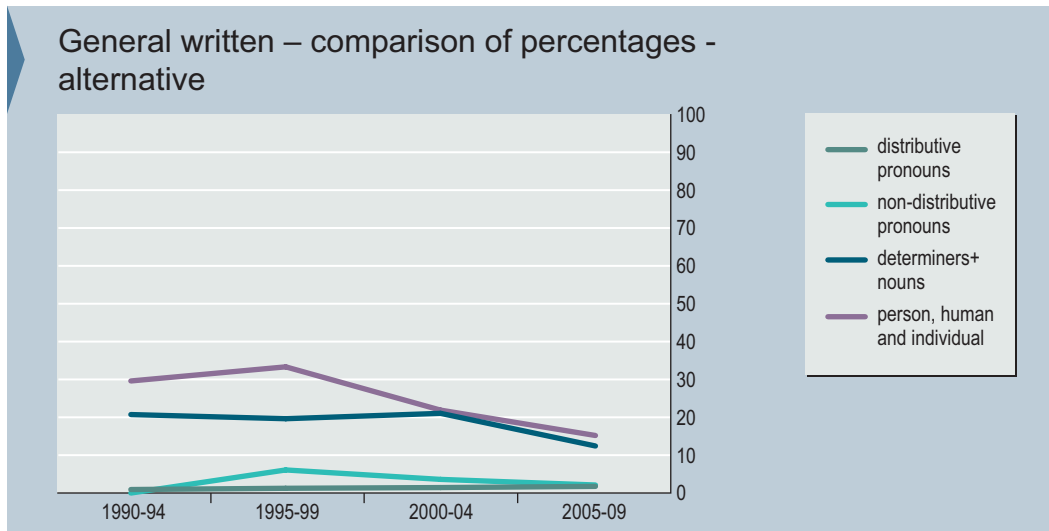
The charts below summarize the trends for plural/neutral, masculine and alternative forms, depending on whether they refer to distributive or non-distributive pronouns, to determiners + nouns, or to the nouns *person*, *human* and *individual*. It is evident that plural/neutral forms have increased in all these types; they are particularly frequent following distributive and non-distributive pronouns; and their use has risen dramatically when following a determiner + noun and the nouns *person*, *human* and *individual*.



Masculine pronouns have become less frequent with all types, but especially when following a determiner + noun and the nouns *person*, *human* and *individual*.



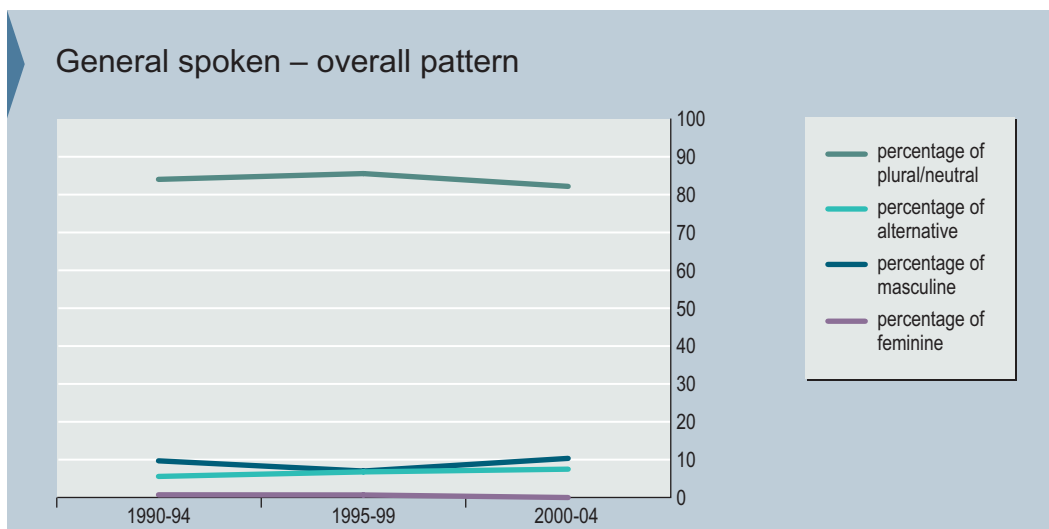
Alternative pronouns are very infrequently used to refer to distributive and non-distributive pronouns. They have become less frequent when following a determiner + noun or the nouns *person*, *human* and *individual*.



4.1.2 General English – spoken

Overall

In general spoken English, the overall pattern is that plural/neutral forms are by far the most frequent in all periods, accounting for over 80% of occurrences. The slight dip in frequency in 00-04 can be explained by variations in source materials. Whereas a lot of the 90s spoken material is from casual conversations which were recorded and transcribed when the corpus was first built, the 00s data are mainly from transcriptions of television and radio interviews, which are slightly more formal in register.

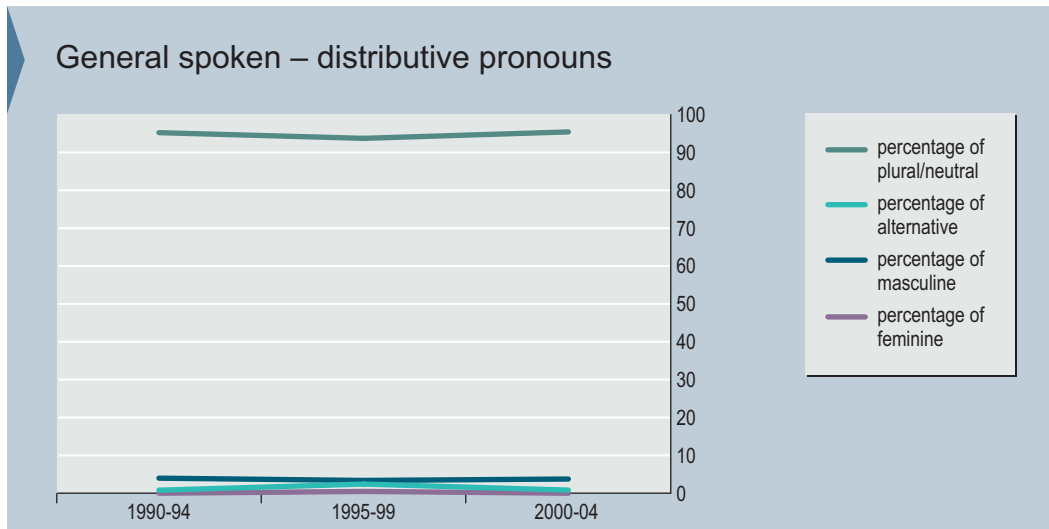


Distributive pronouns

In all periods, the vast majority – approximately 95% – of the citations with distributive pronouns use plural/neutral forms, e.g.

'Yeah. Mm... has **anybody** else had that experience where **they**'ve—they've done something **they** haven't really wanted to do 'cos of their friends?' (1995)

'I was horrified. I mean **nobody** likes **their** child to be hit and especially a child that has special needs.' (2002)

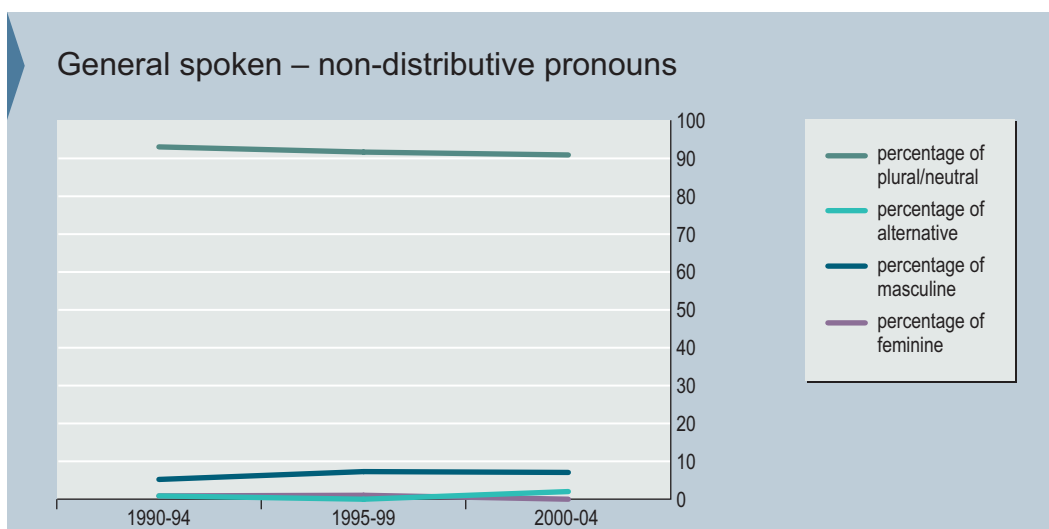


Non-distributive pronouns

In all periods, over 90% of citations with non-distributive pronouns use plural/neutral forms, e.g.

'And also it depends on how **somebody's** wearing it. Like if **they**'ve got jewellery or a brooch or a scarf or a hat or something...' (1995)

'But relationships in terms of being able to talk with **someone**, in government and understand where **they**'re coming from on a particular issue...' (2003)



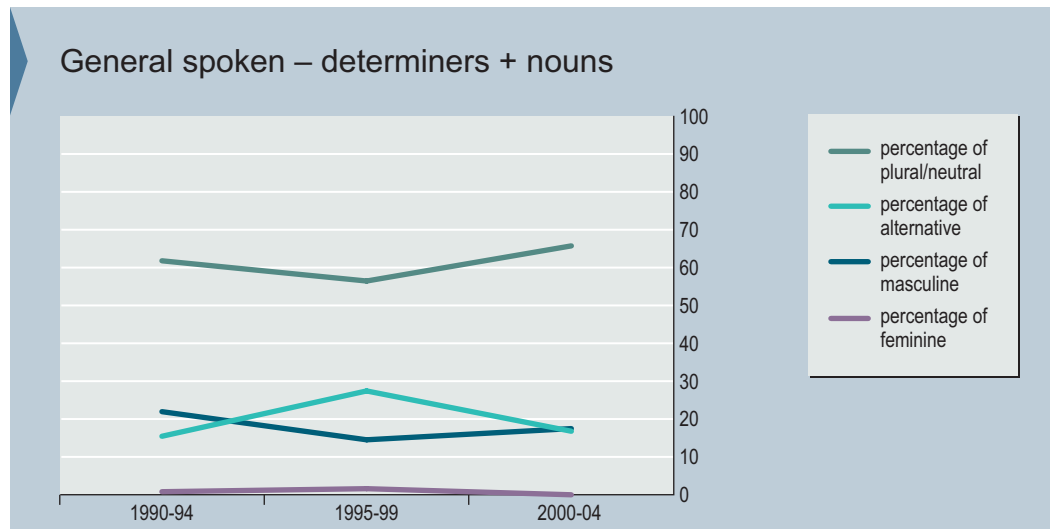
Determiners + nouns

In 95-99 there was an increase in alternative forms used with determiners + nouns, e.g.

'**Any citizen** who wants to educate **himself or herself** has plenty of sources from which to do so.'
(1996)

Notably, though, these are all from transcriptions of radio and TV; none are from informal conversations. In 00-04, plural/neutral forms became more frequent even in radio and TV transcriptions, e.g.

'And the way it will work is, **every member** was mailed a ballot, and **they** can either vote by returning that ballot by mail, or voting on the Internet...'
(2003)



Person, human and individual

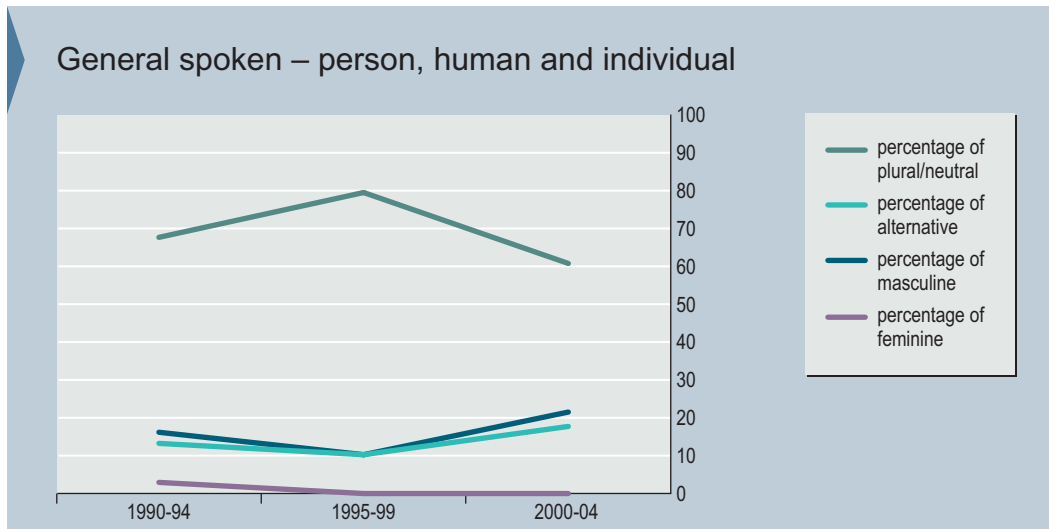
An interesting pattern is evident with the nouns *person*, *human* and *individual*, where plural/neutral forms became relatively less frequent in 00-04, while masculine and alternative forms became slightly more frequent, e.g.

'But we don't have to outlaw common sense to hold the view that if a **person** chooses or **his** family chooses for **him** to grow up in Afghanistan, if that's where his life is, then the Canadian citizenship could be considered only a matter of form.'
(2003)

'What that means for taxpayers? If an **individual** is earning 50 000 a year, **he or she** would save about a thousand dollars.'
(2004)

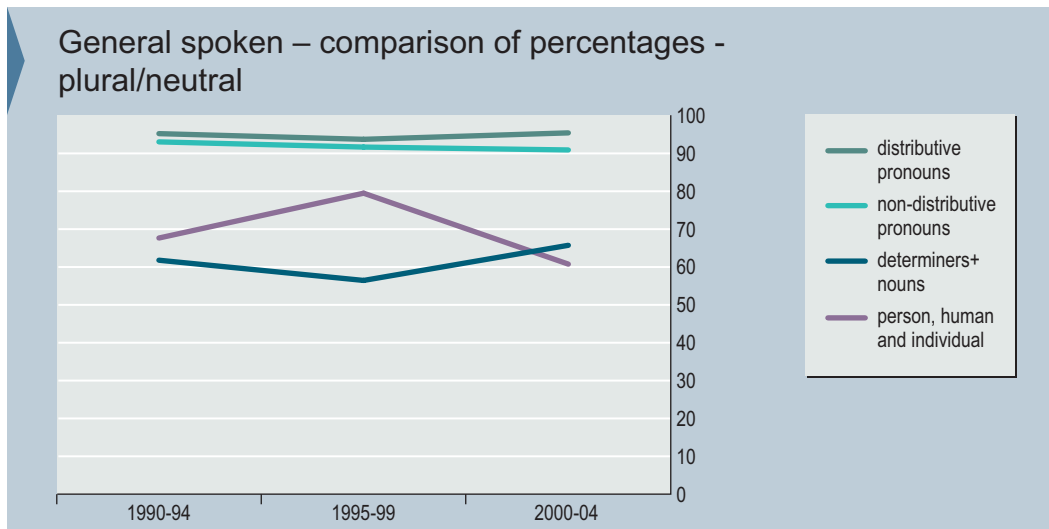
However, plural forms still account for over 60% of citations in the 00-04 slice, with examples such as:

'You know, that's what ... a responsible **person** would do with **their** household budget if they wanted to go on a vacation.' (2002)

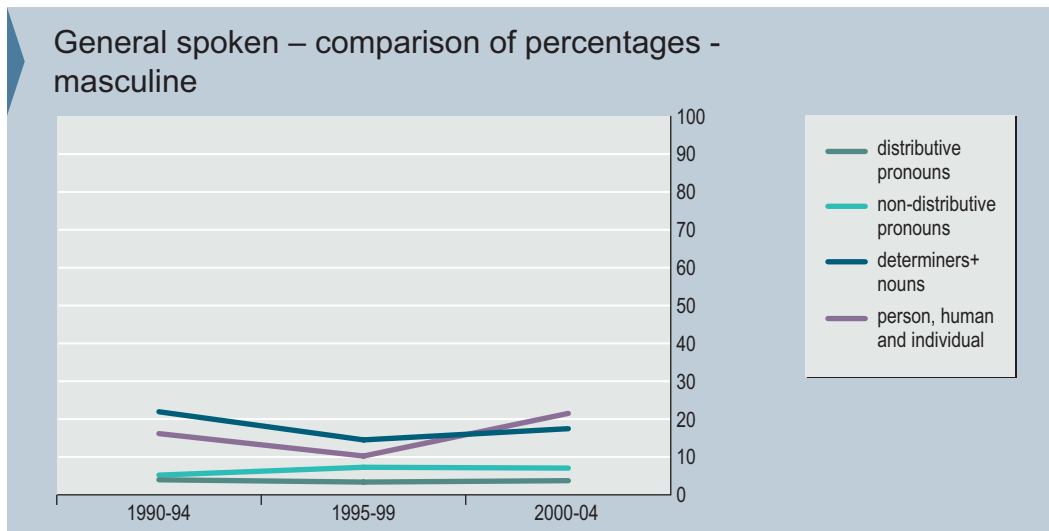


Comparison

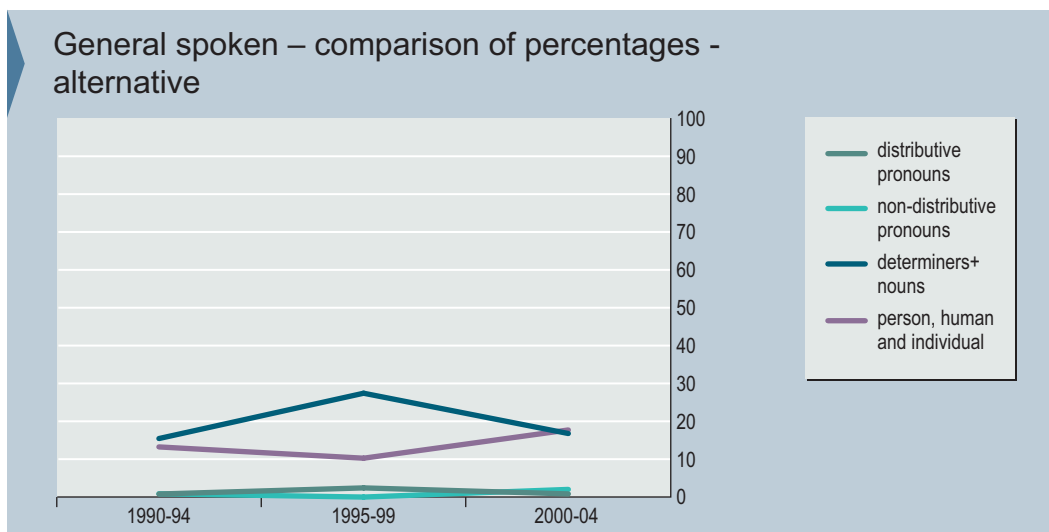
Plural/neutral forms have been consistently frequent when referring to distributive and non-distributive pronouns, while their frequency has fluctuated with determiners + nouns and the nouns *person*, *human* and *individual*.



For all types, masculine pronouns and determiners have been low in all periods (around 20% or lower).



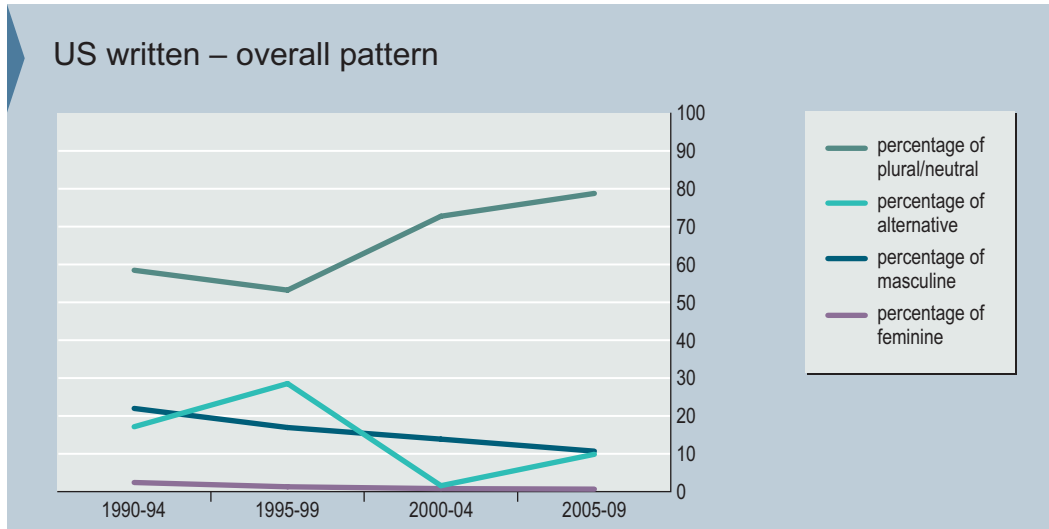
Alternative forms became more frequent in 95-99 when referring to determiners + nouns, but have since decreased in frequency.



4.1.3 US English – written

Overall

Overall, the percentage of masculine pronouns and determiners has decreased over time, while the percentage of plural/neutral has increased. There was a slight increase in alternative ‘his or her’ types in the 95-99 slice, and a subsequent decrease after 00. In current US written English, plural/neutral pronouns and determiners account for 80% of usages; masculine 10%; and alternative 10%. Feminine generics are very infrequent in all periods.

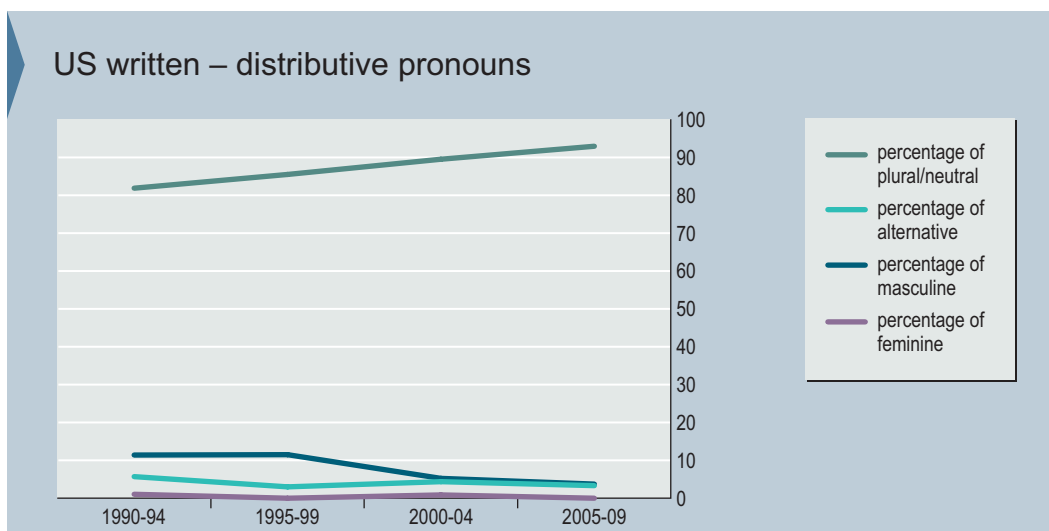


Distributive pronouns

Plural/neutral pronouns and determiners are particularly frequent with distributive antecedents, and have become more frequent since 1990. In current US written English, the vast majority (over 90%) of these constructions use the plural form, e.g.

‘I tell that joke to **everybody** when I first meet **them**.’ (2008)

‘**Anybody** who wants to join this new group, we invite **them** to do so.’ (2008)

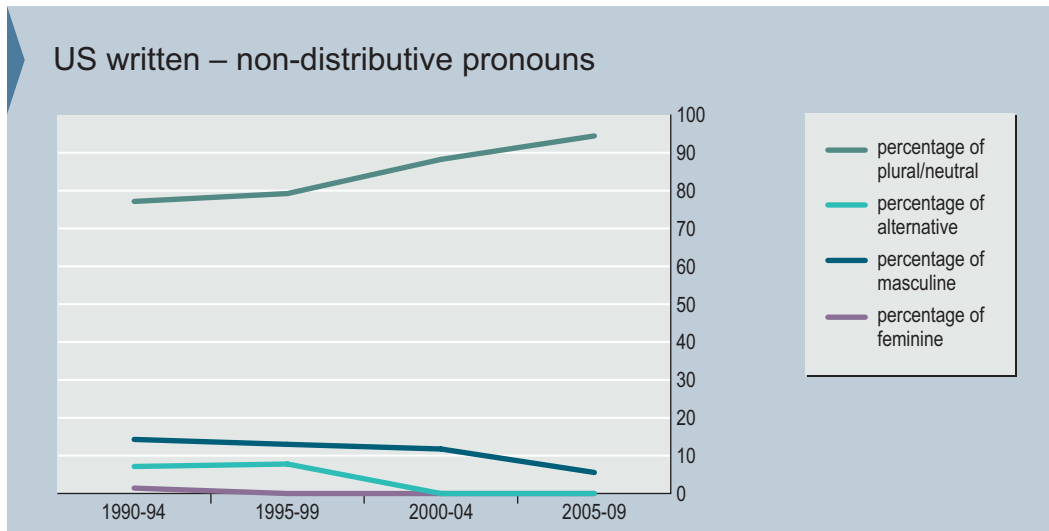


Non-distributive pronouns

Non-distributive pronouns also tend to be referred to by plural/neutral pronouns and determiners in all periods. Again, in current US written English, these account for over 90%. Examples include:

'If you want to foster growth in **somebody**, you have to accept **them** first.' (2005)

'Whenever **someone** has a personal failing, at what point have **they** become a hypocrite?' (2008)



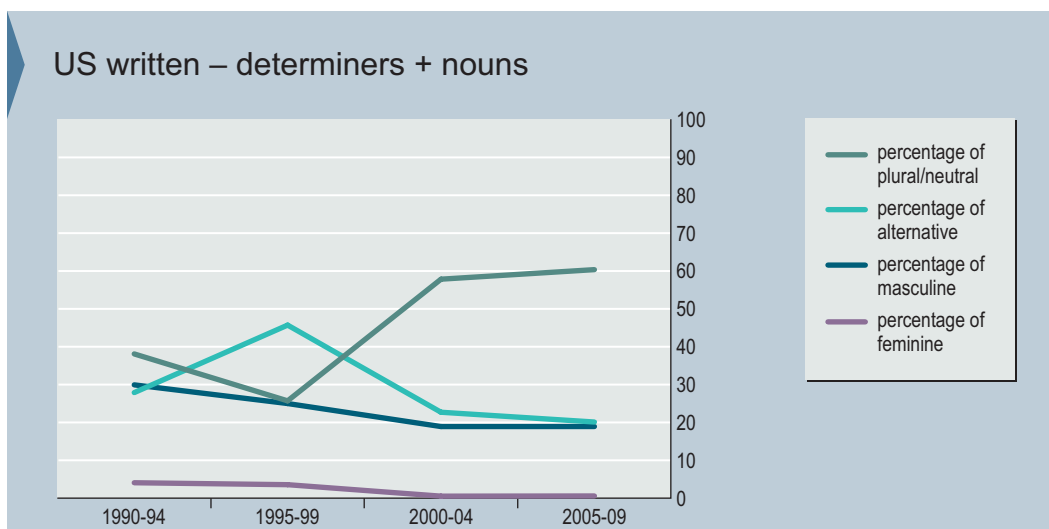
Determiners + nouns

The pattern with determiners + nouns is more varied. In 90-94, usage was divided almost equally between plural/neutral, masculine and alternative types. In 95-99, the alternative 'him or her' type became more frequent, and there was a corresponding decline in plural and masculine generics. Thus we find examples such as:

'**Any adult** will seem to be unintelligent if **he or she** develops skills that are not valued by the larger society.' (1995)

But in recent years, plural/neutral usages such as the following are more common:

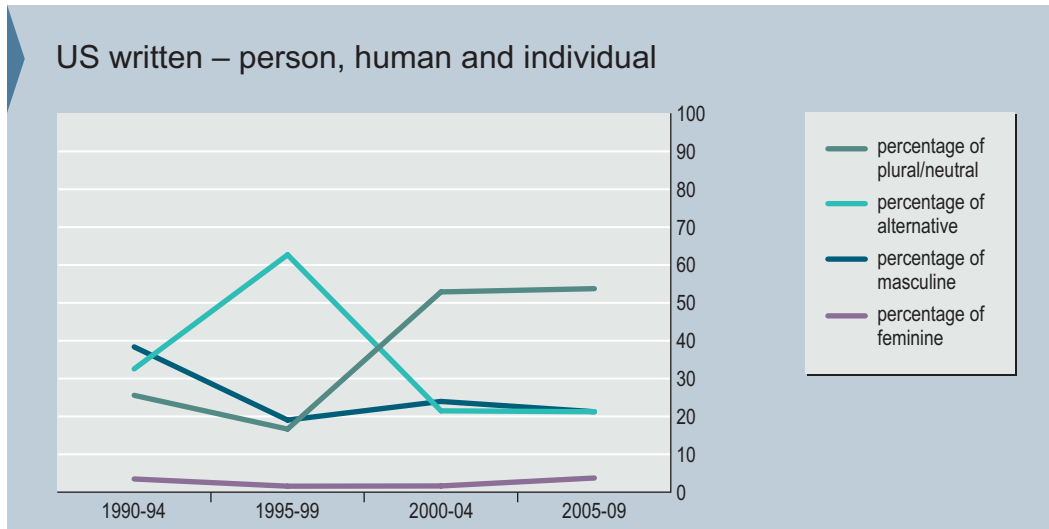
'**Each winner** received a plaque with **their** award title.' (2008)



Person, human and individual

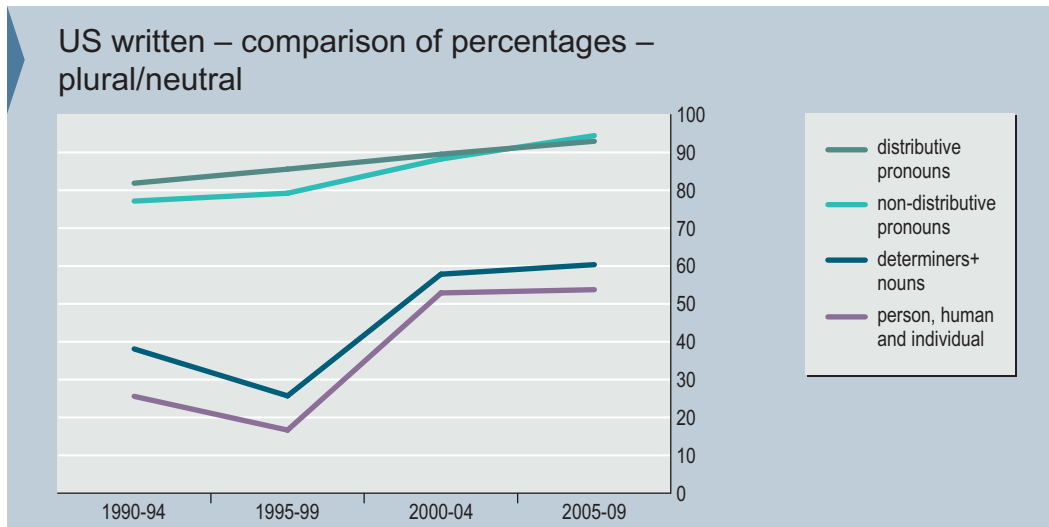
This pattern is even more pronounced with the nouns *person*, *human* and *individual*: 95-99 saw a surge of alternative 'him or her' usages, which have subsequently declined and been replaced by plural/neutral forms, such as:

'There was no way to separate out the parts of a **person**. That was to deny **their** individuality.' (2004)

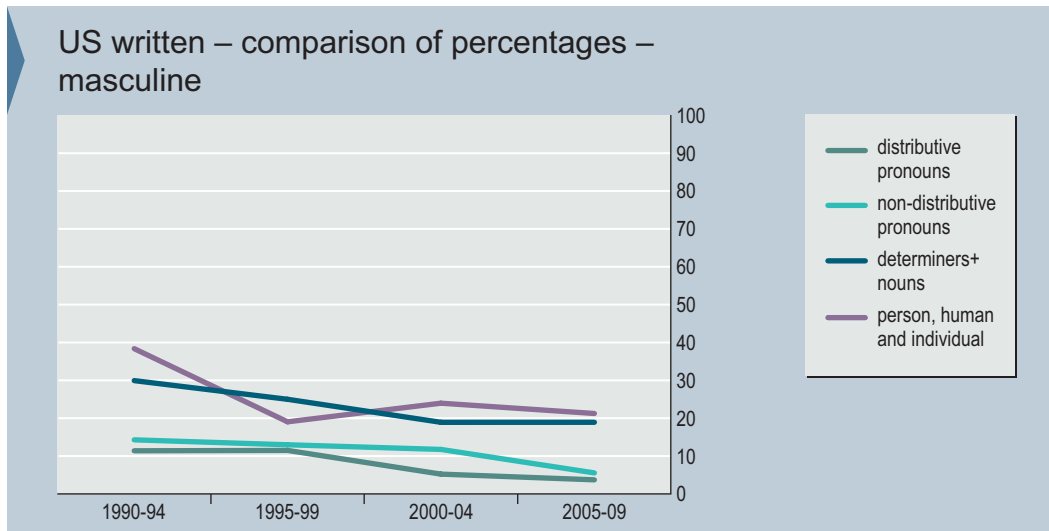


Comparison

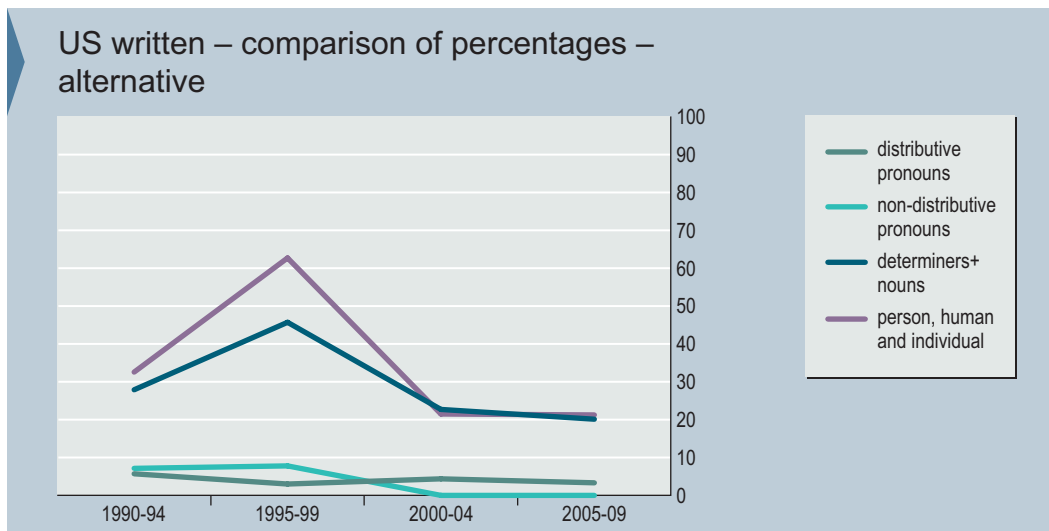
As with general written English, plural/neutral forms have increased in all types. They are particularly high following distributive and non-distributive pronouns, and their use has increased dramatically when following a determiner + noun and the nouns *person*, *human* and *individual*.



Masculine pronouns and determiners have decreased in all types.



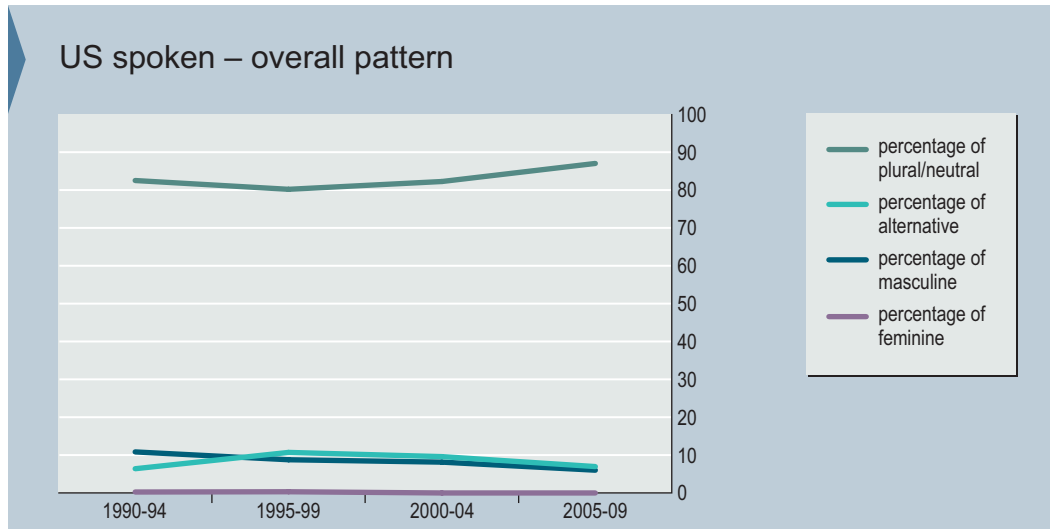
Alternative forms are infrequently used to refer to distributive and non-distributive pronouns. In 95-99, they were used more frequently following determiners + nouns and the nouns *person*, *human* and *individual*, but these usages have since become less frequent.



4.1.4 US English – spoken

Overall

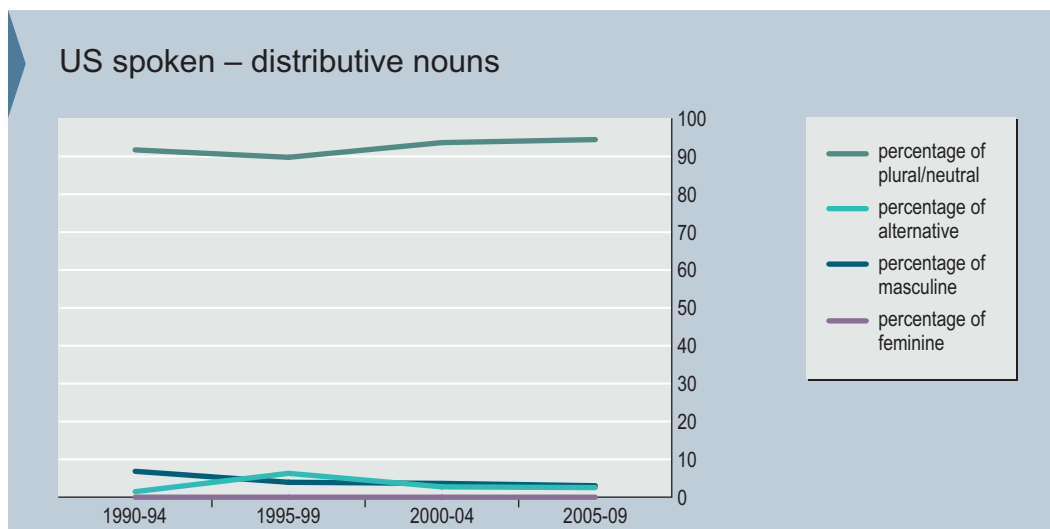
Overall, plural/neutral pronouns and determiners have consistently been the most frequent type, accounting for between 80% and 90% of usages.



Distributive pronouns

Plural/neutral pronouns and determiners are used particularly frequently when referring to distributive pronouns, e.g.

'I mean, **anybody** can run for public office if **they** want to.' (1999)

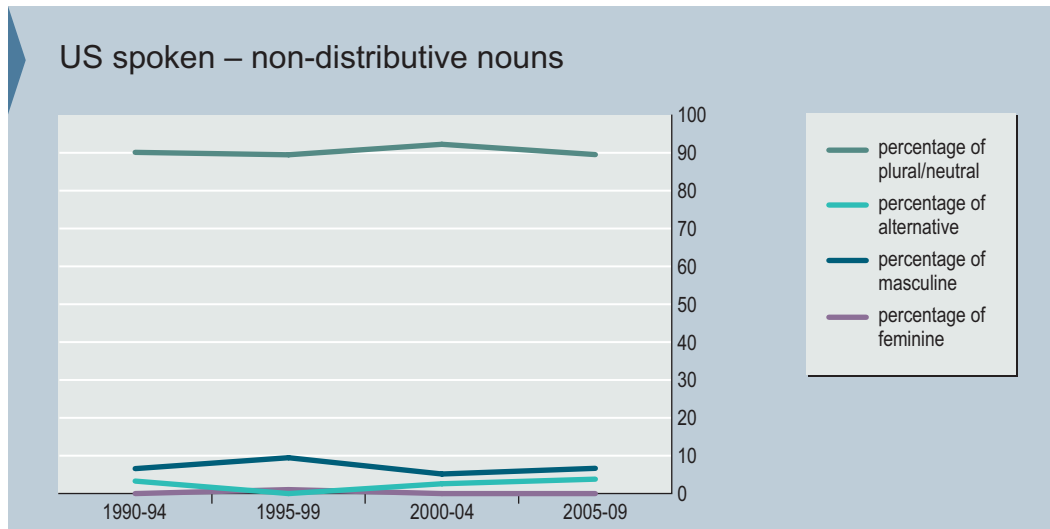


Non-distributive pronouns

Plural/neutral pronouns and determiners are also very frequent when referring to non-distributive pronouns, e.g.

'I mean, by way of example, you know, if **someone** comes into the emergency room, you know, and **they**'ve had a heart attack...' (2005)

'And **one** can be very choosy and get what **they** want, rather than settling for whatever they could afford...' (2006)



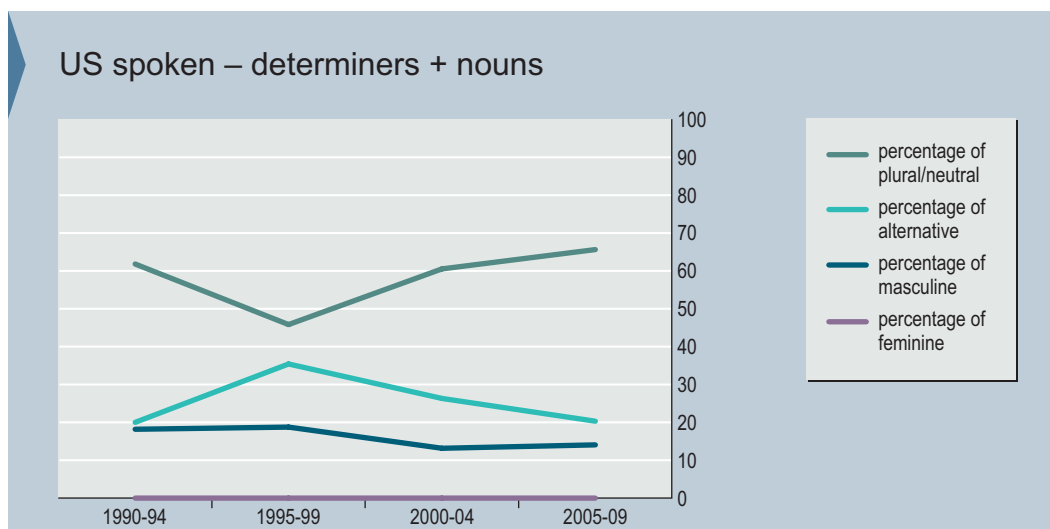
Determiners + nouns

The pattern with determiners + nouns is more balanced between plural/neutral, alternative and masculine forms, although the former are still the most frequent in all periods. Again, there was an increase in alternative 'him or her' types in 95-99, e.g.

'I am really uncomfortable about the way multiculturalism is administered while I'm totally in favor of its aims: to show **every child** how precious **he or she** is, no matter where **he/ she** came from.' (1996)

– followed by a decrease, and an increase in plural forms such as:

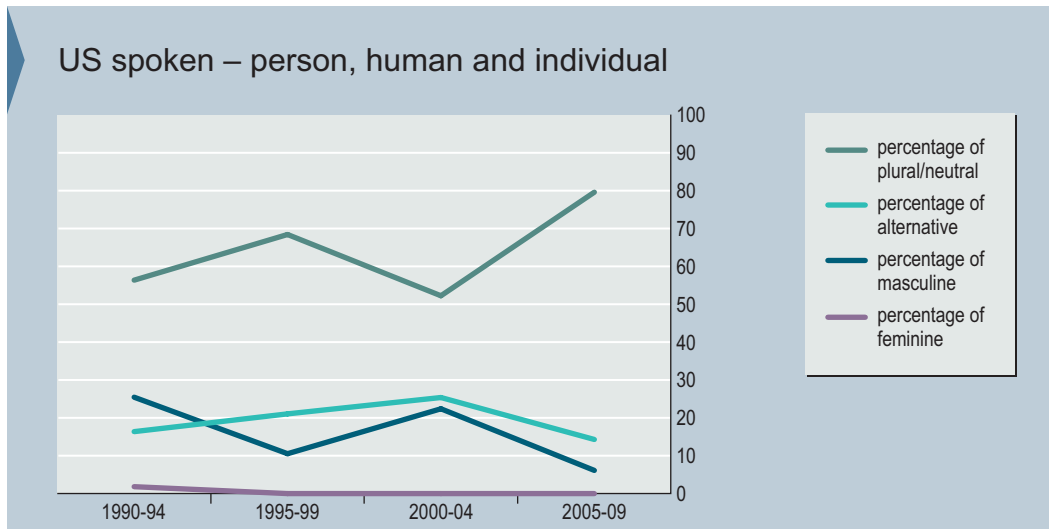
'Republicans believe that **every parent** must be able to take **their** sick child to the closest emergency room.' (2000)



Person, human and individual

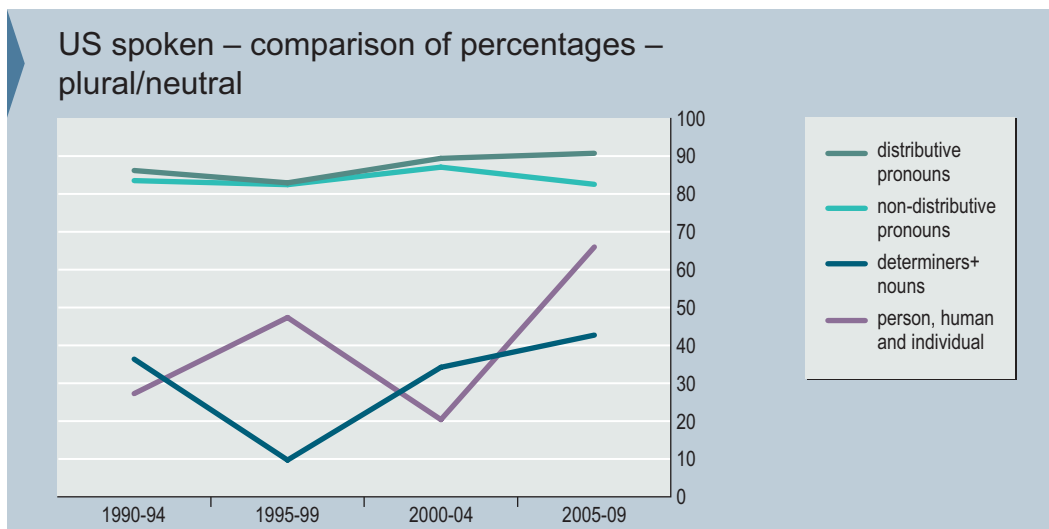
There was a slight increase in masculine generics in 00-04 in the US spoken data for *person*, *human* and *individual*, followed by a decrease in 05-09. In current US spoken English, plural forms such as the following account for 80% of occurrences:

'The retention piece allows an **individual** to transfer a portion of **their** benefit or all of **their** benefit at different points in their career....' (2008)

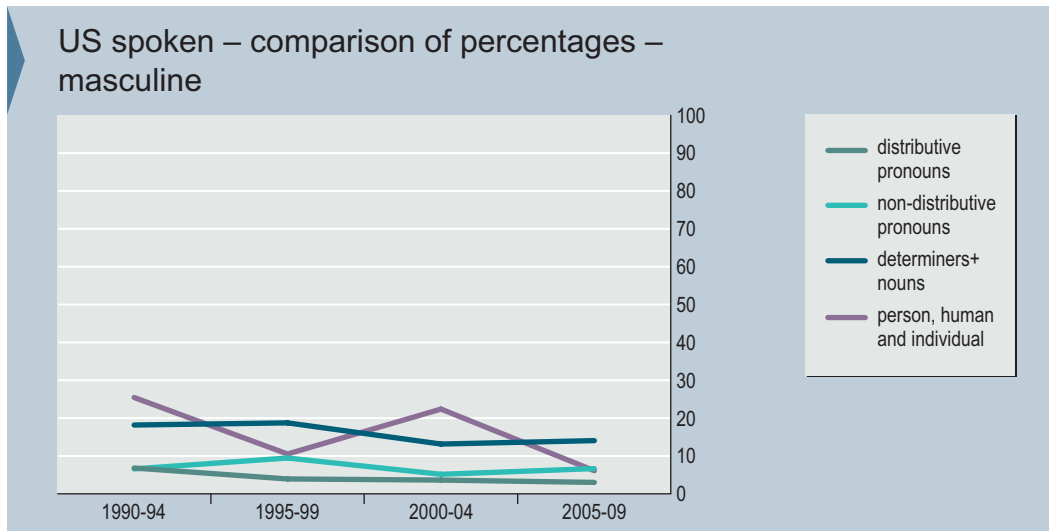


Comparison

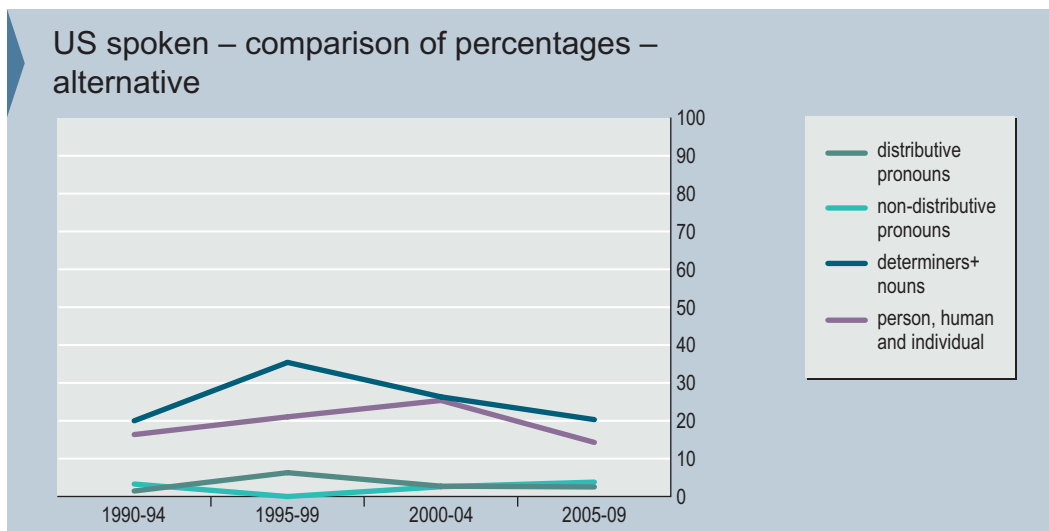
Overall, plural/neutral forms are particularly frequent with distributive and non-distributive pronouns. With determiners + nouns and the nouns *person*, *human* and *individual*, they have fluctuated but become increasingly frequent since 2005.



Masculine forms have been infrequent overall, particularly since 2005.



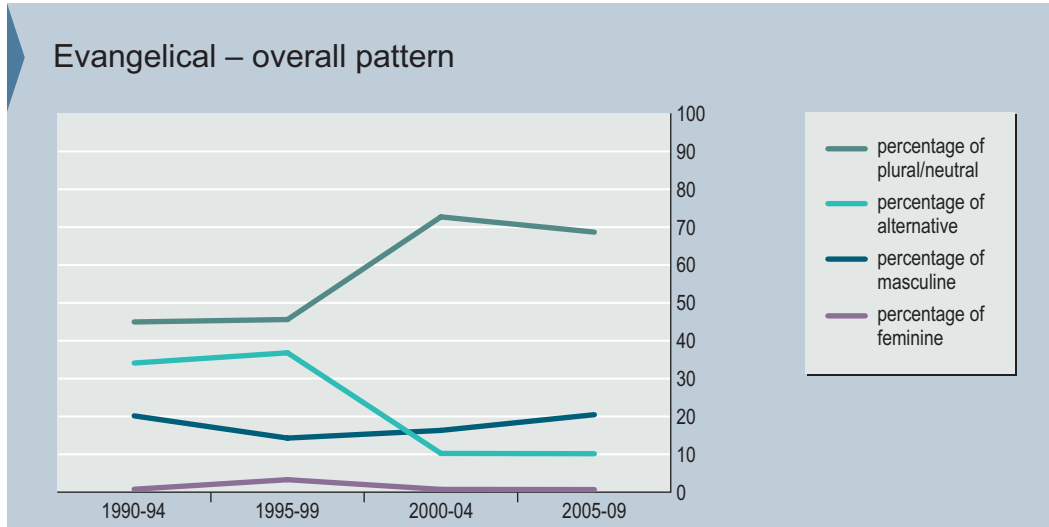
Alternative forms are very infrequent with distributive and non-distributive pronouns, in all periods. They were quite frequently used (over 30% of occurrences) with determiners + nouns in 95-99, but have since decreased in frequency.



4.1.5 Evangelical English

Overall

In 90-94 and 95-99, usage was quite evenly split between plural/neutral and alternative pronouns and determiners. Since the turn of the century, plural/neutral types have become much more frequent.



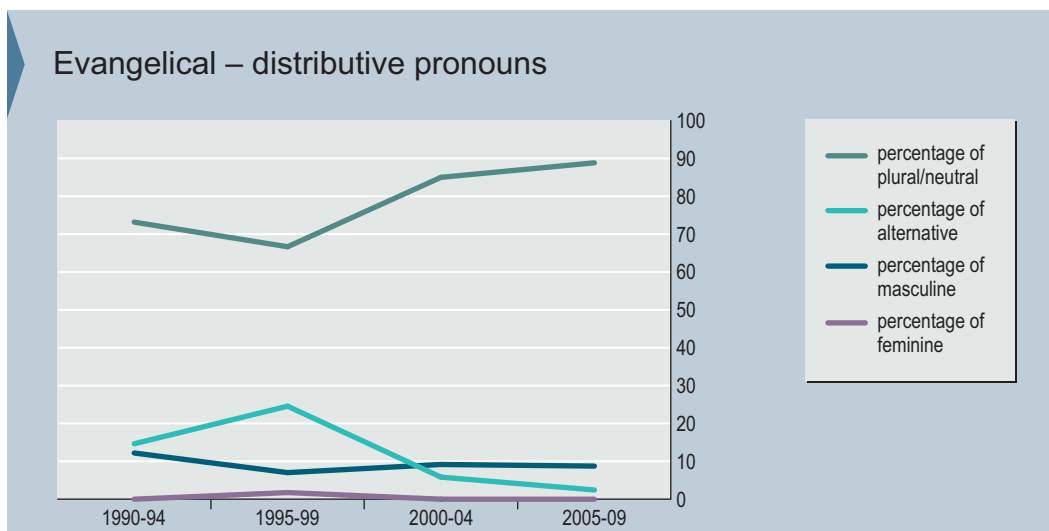
Distributive pronouns

In 95-99, there was an increase in alternative forms following distributive pronouns, e.g.

'Nobody in my congregation felt **he or she** could do that'. (1997)

However, plural/neutral forms have always been more frequent, and have become more frequent since 2005, e.g.

'If you preach in vague, sentimental generalities, you may not inspire **anybody**, but you probably aren't going to offend **them** either.' (2007)



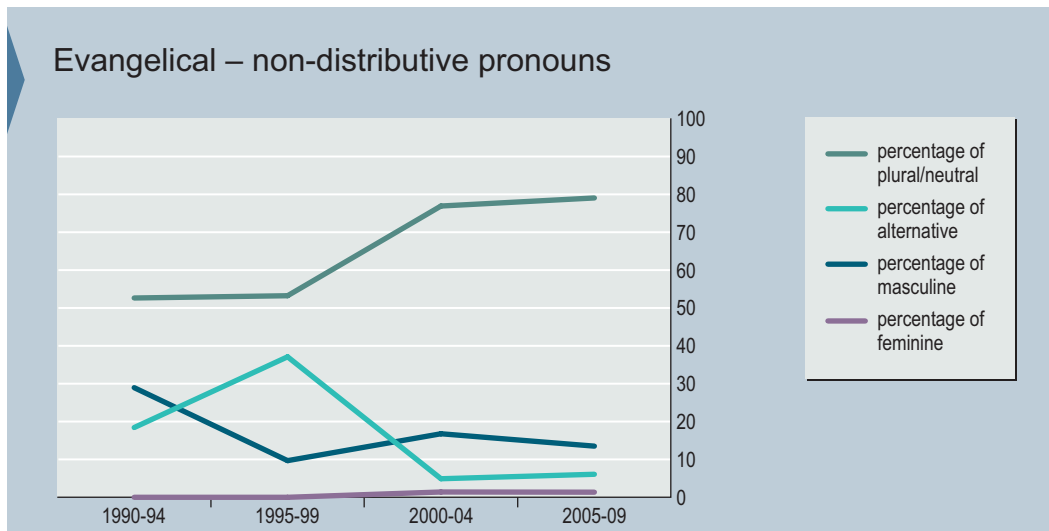
Non-distributive pronouns

Much the same pattern is evident in non-distributive pronoun usage. There was a peak in alternative forms in 95-99, e.g.

'I might test **somebody's** commitment by putting **him or her** on a task force.' (1996)

– followed by a decrease, and an increase in plural/neutral forms such as:

'**Someone** on fire may believe that the swimming pool will save **them**, but they are not saved until they dive into the pool.' (2005)



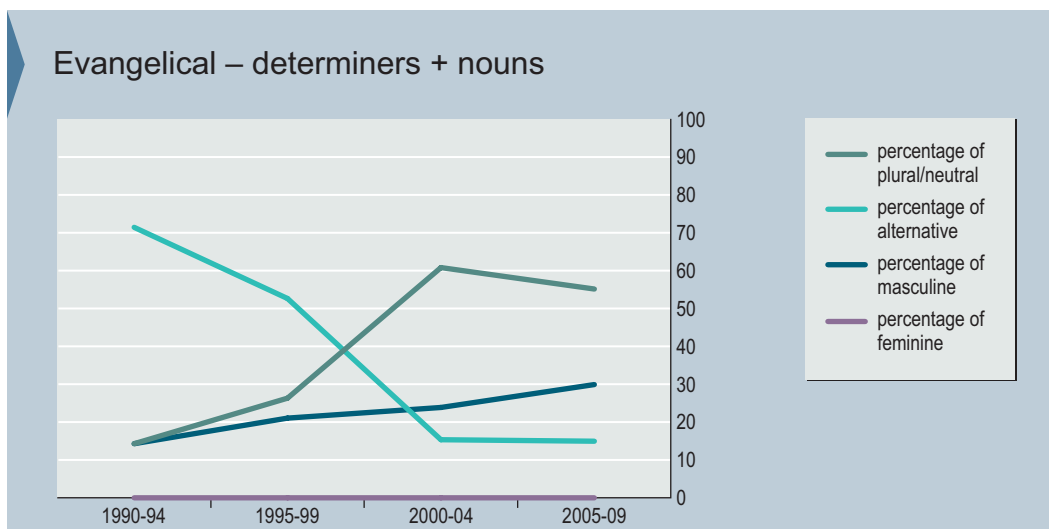
Determiners + nouns

The data for the determiners + nouns show a marked decline in alternative 'him or her' usages since 1990, and a corresponding increase in plural/neutral usages. There has also been an increase in masculine usages, which in 05-09 account for 30% of occurrences, e.g.

'I don't see how **any serious believer**, whichever side **he** takes, can be cheered by schism.' (2008)

However, plural forms still account for over half of occurrences, e.g.

'Shield **each student** from any dangers in **their** home...'. (2009)



Person, human and individual

A similar pattern can be seen in the results for *person, human* and *individual*. In the 90-94 slice alternative forms account for almost 60% of occurrences, e.g.

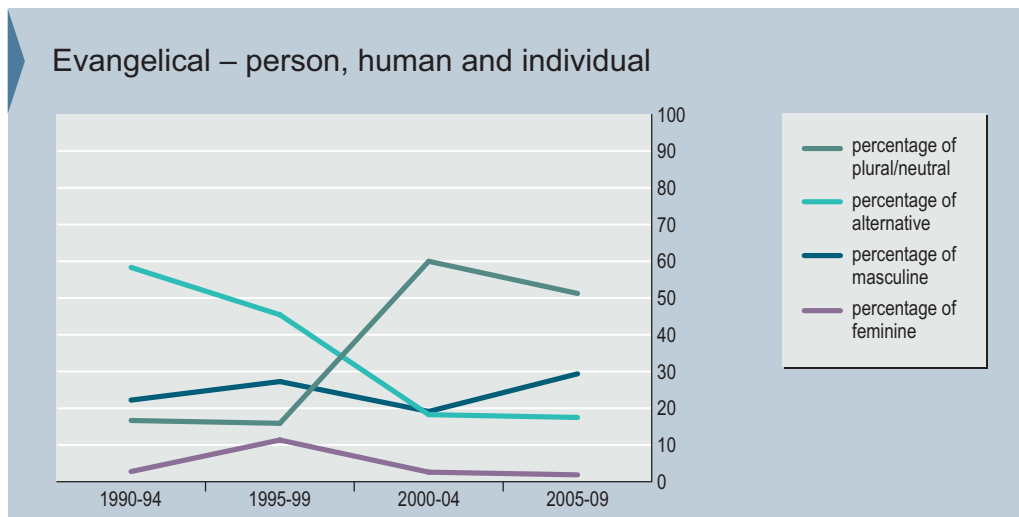
'In the Bible a **person's** name is a description of **his or her** character.' (1994)

Many of these are from Grudem's *Systematic Theology* (1994), which consistently uses alternative forms (see 4.1.7 below). Subsequently, there is a marked decline in alternative pronouns and determiners and a corresponding increase in plural/neutral ones. Again, there has been an increase in masculine generics since 2000, and in the 05-09 slice these account for almost 30% of occurrences, with examples such as:

'A **person** cannot ignore the past but **he** can choose his future.' (2005)

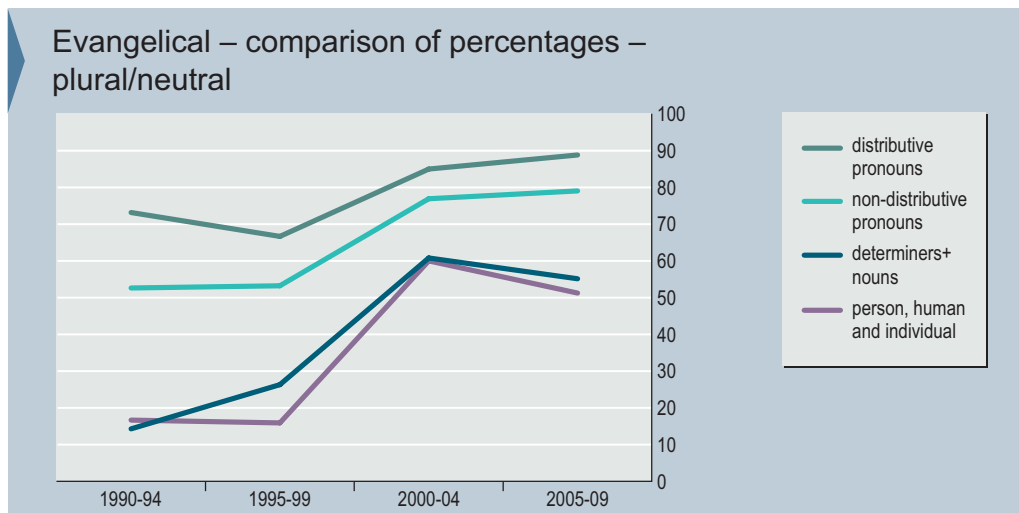
However, plural/neutral forms are still the most frequent, e.g.

'An enlightened **individual** will recognize that **their** individual good is only attainable in the context of a community.' (2006)

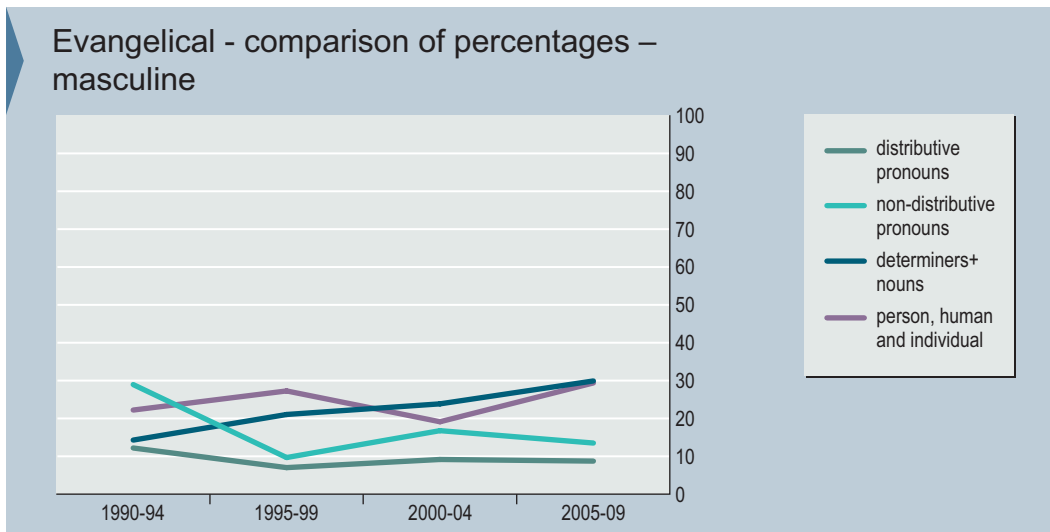


Comparison

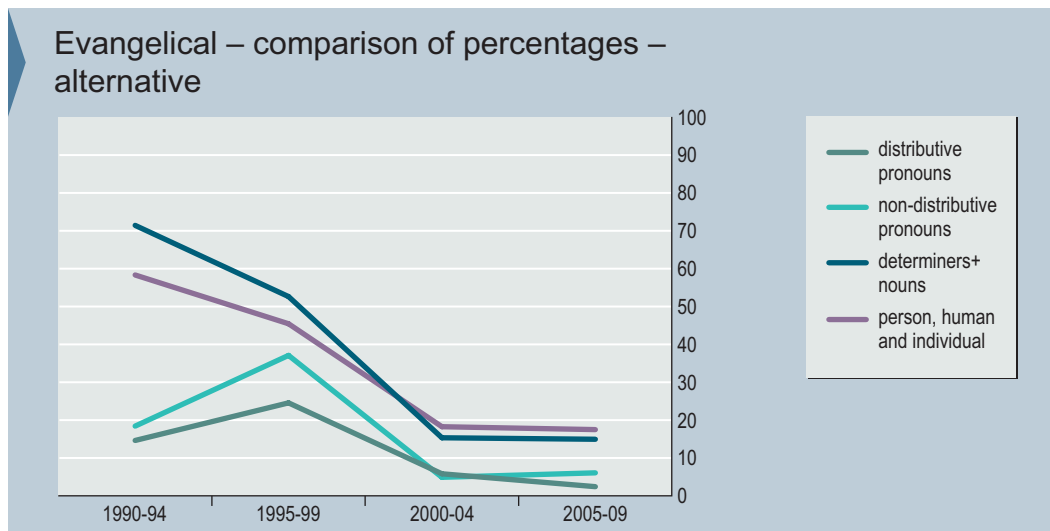
Overall, plural/neutral forms have become more frequent following distributive and non-distributive pronouns. There was also an increase in this type following determiners + nouns and the nouns *person, human* and *individual* up until 00-04, but a slight decrease since 2005.



Masculine pronouns and determiners have become less frequent following distributive and non-distributive pronouns. They have become slightly more frequent following determiners + nouns and the nouns *person*, *human* and *individual*, but are still less frequent than plural/neutral forms.



Alternative forms have become significantly less frequent in all types of usage.



4.1.6 *Other observations*

Neutral pronouns/determiners used when gender is known

As noted in section 3 above, researchers were asked to identify cases where neutral *they*, *their* etc. were used even though the gender of the antecedent was known. Unfortunately, this can often be difficult to recognise without context, and no clear pattern of the development of this usage has emerged. However, it is worth noting that it occurs in all slices of the corpus, and it suggests the prevalence of the plural form even when it is not required for reasons of political correctness. The following are some typical examples:

'The same things will be going on again before the Western Bulldogs game. **Each player** will be feeling different emotions and analysing **their** different roles.' (General written, 2000) [The Western Bulldogs are an all-male Australian football team.]

'And if **someone** has an abnormal mammography, it does not mean **they** have breast cancer'. (General spoken, 2002)

'Ask the young mothers and **no one** will say **they** regret having their baby.' (US spoken, 1992)

'If **anybody** had a right to be proud of **their** accomplishments, it was Paul.' (Evangelical, 2003)

'I talked to **somebody** else in line, and **they** said it would be many, many hours.' (US written, 2004) [In this citation, even though the reader does not know the gender of 'somebody', the writer presumably does.]

Shift in pronoun/determiner

Another fairly frequent occurrence is where the writer or speaker seems to change his or her mind about which pronoun/determiner to use. Examples include:

'Historically, some independent candidates have had difficulty finding **someone** willing to risk **his** standing within the established party system by breaking ranks and throwing **their** lot in with a candidate who has little chance of winning.' (US written, 1992)

'How can **one** talk about multi-ethnic ministry when **one** can't even embrace **his** own culture?' (Evangelical, 2006)

There are also conscious attempts at gender-neutrality, where the writer/speaker alternates between masculine and feminine forms in succeeding sentences, e.g.

'**No little child** is tainted with sin. **He** is not responsible for the mistakes of his parents. **She** is not somehow predestined for failure because of her parent's irresponsibility.' (Evangelical, 2008)

'If **an individual** goes five days upon **his** return without drinking soda, she said, that helps. If **she** spends one evening or two walking after dinner, that's a start.' (US written, 2008)

S/he

One final point worth noting is that not a single citation with 's/he' was found in any of the samples: it is clearly a very infrequent form.

4.1.7 *Commentary/interpretation*

Some possible reasons for the trends presented above are now set out:

- In several of the corpora, there is a peak in alternative 'him or her' types in the mid-90s. This may have been a result of heightened sensitivity to the gender-language debate in this period, and heightened concern with political correctness. Since the turn of the century, this type has become much less frequent, perhaps because of a more relaxed attitude towards politically correct language.

- The predominance of ‘him or her’ types in the early Evangelical material can be explained in part by the distribution of source materials (many of these usages are from Grudem’s *Systematic Theology* (1994)), but might also indicate a particular concern with gender language in the Evangelical corpus materials in this period. (This will be discussed further in relation to the *mankind* data below.)
- It is possible that the slight increase in masculine pronouns in recent Evangelical English was motivated by a reaction to TNIV and a conscious attempt *not* to use gender-neutral forms.
- Alternative ‘him or her’ types are very rarely used with distributive and non-distributive pronouns, and the evidence suggests that these pronouns are considered by speakers and writers to be plural. We found many examples such as ‘I think **everybody** led by the government **are** doing **their** best efforts’ (US spoken, 2004) where the use of the plural verb indicates that ‘everybody’ is thought of as plural. This supposition is also supported by the fact that *they* is often used even when the gender is known.

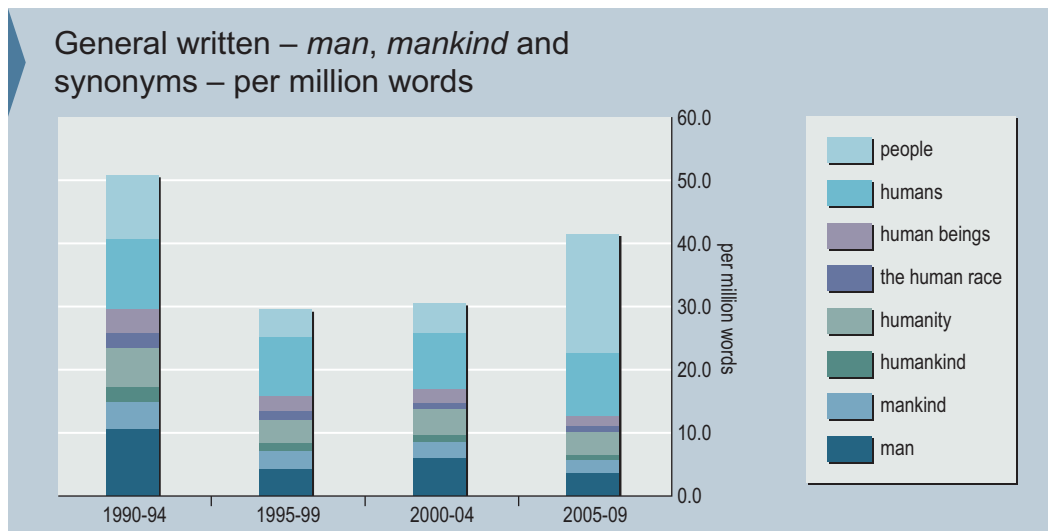
4.2 Mankind, man and synonyms

For each slice of data, three charts are shown:

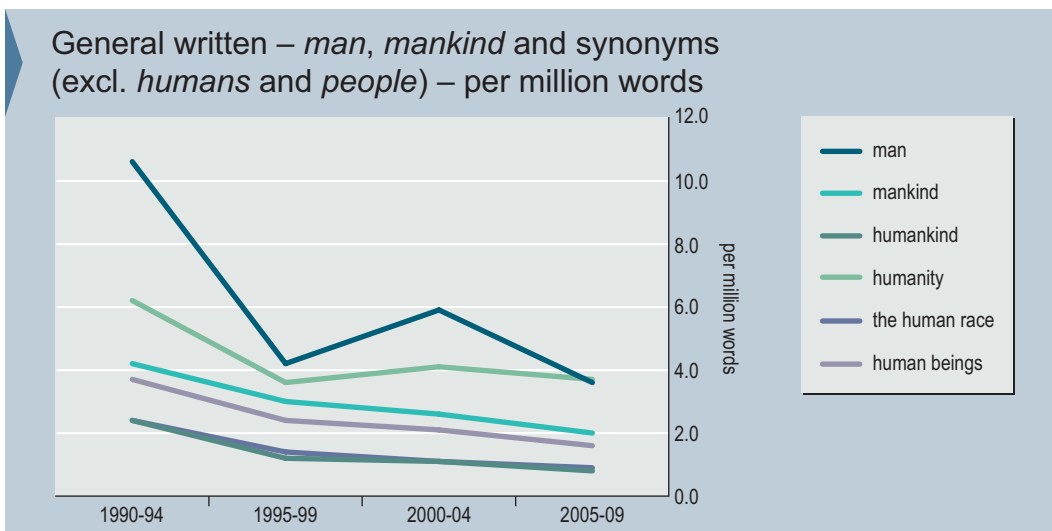
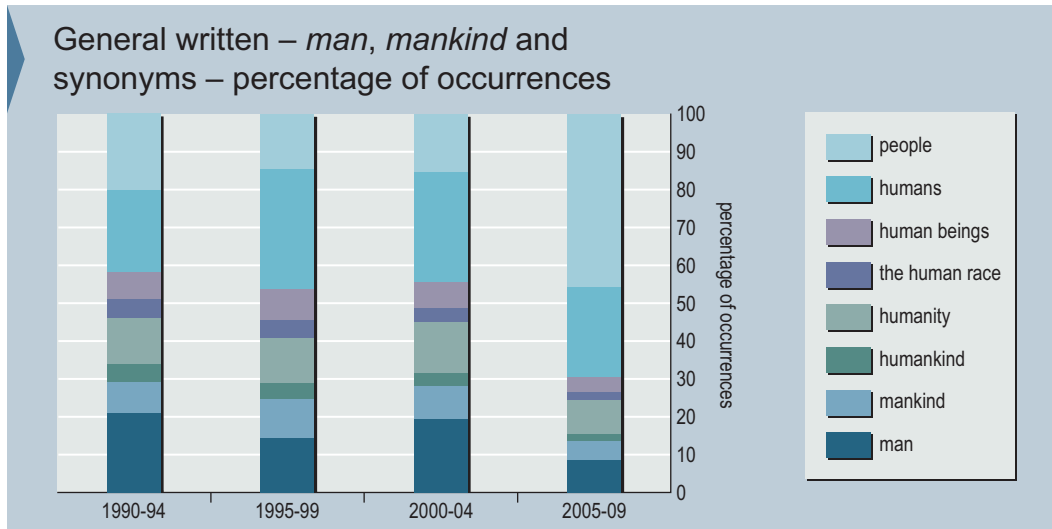
- A. The absolute frequencies of *man*, *mankind*, *humankind*, *humanity*, *the human race*, *human beings*, *humans* and *people*, per million words.
- B. The relative frequencies of these eight words/phrases; that is, the number of occurrences of each as a percentage of the total occurrences. This is useful because in some cases variation in subject matter means that there are differences in the overall frequency of these words/phrases in a particular slice.
- C. The absolute frequencies of *man*, *mankind*, *humankind*, *humanity*, *the human race* and *human beings*, per million words. It is useful to have a chart that focuses on these and excludes *humans* and *people*. *Humans* and *people* are looser synonyms; it is often difficult to determine whether they refer to all humans, or to a smaller subset, and they tend to be used in a different register from the more formal *man*, *mankind* and so on. Furthermore, the figures for *people* and *humans* can only be approximate, since these words are so frequent that the figures are necessarily based on relatively small samples of their overall occurrences⁵.

4.2.1 General English – written

The charts below show that the most frequent words used to express the sense ‘humans collectively’ or ‘the human species’ in general written English are *humans*, *people*, and *man*, with diachronic fluctuations.



⁵ For example, *people* occurs (without a preceding determiner, adjective etc.) 71342 times in the 05-09 part of the Evangelical corpus. Even quite a large sample of this total – 250 citations, which would take a researcher approximately 2 hours to analyze – is only 0.003 of the total. It would have been impossible to analyze all citations with *people*, so the resulting figures are necessarily approximate.



Humans has been consistently frequent, in citations such as:

‘Of the 135 species of snake found in Australia, only 10 are considered deadly to **humans**.’ (1999)

‘The deepest musical note ever generated in space—a B flat one million billion times deeper than can be heard by **humans**.’ (2003)

‘Collecting is instinctive to **humans**, declares Paige West, who is the curator of the very significant West Collection’. (2008)

The frequency of *people* has varied, but it is used very frequently in current general written English, in citations such as:

‘... pursue and protect the common good where all **people** are treated fairly for a just society...’ (2006)

‘Historian, television presenter and Shakespeare biographer Michael Wood said the “dumbing-down” trend was damaging **people**’s knowledge of the past.’ (2004)

Man was particularly frequent in the early 90s, in citations such as:

'If, therefore, the environmental hypothesis on the origins of inequalities does not gain or lose from the scientific research on the origins of **man**, there is little substance in it to sustain much speculation.' (1990)

'...the germination of all lifeforms on Earth, including plants, animals and **man**.' (1991)

Man has decreased in frequency but is still one of the most frequent synonyms (after *people* and *humans*) in current general written English, in citations such as:

'**Man**'s aggression also explains what went on much further away, when **man** first colonised Australia and New Zealand.' (2002)

Humanity is, in the 05-09 slice, the other most frequent synonym, occurring approximately the same number of times as *man*. Examples include:

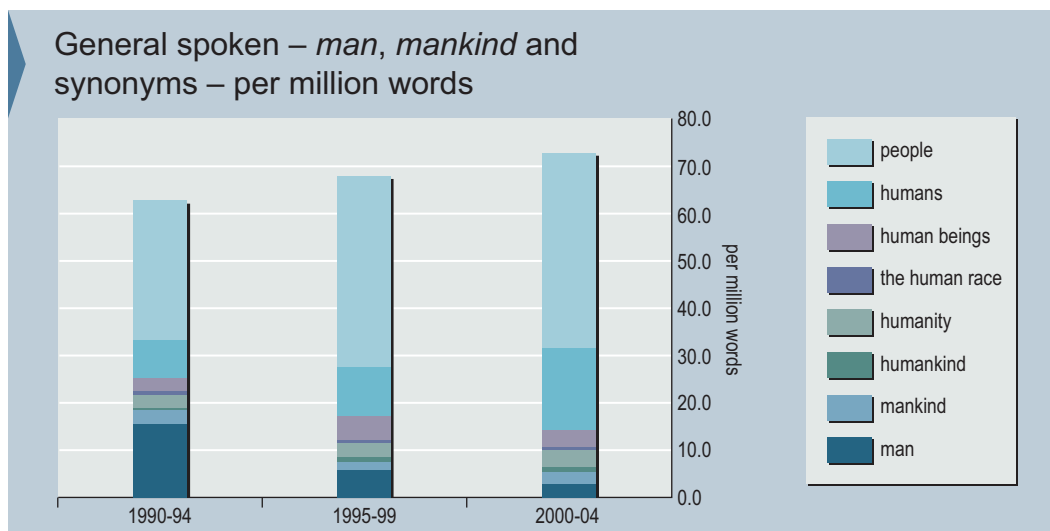
'The spiritual implications of this fear epidemic run to contracting the soul and destroying **humanity**'s ability to engage positively in the world.' (2003)

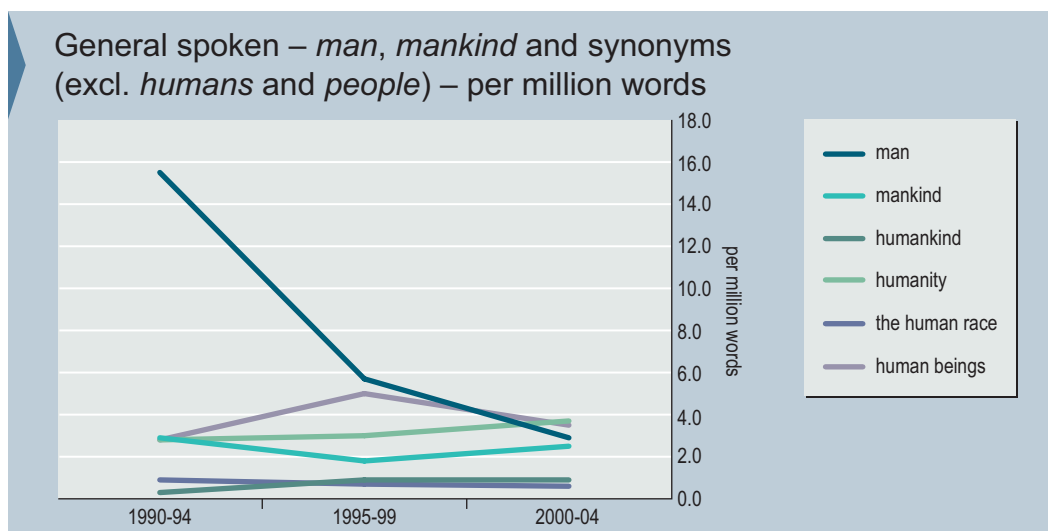
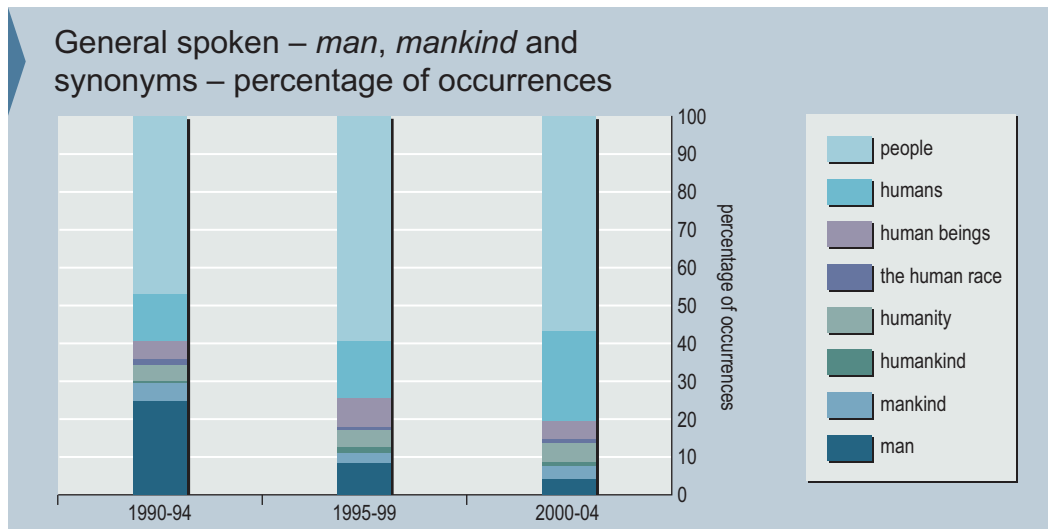
'... these 120 pages ... should be required reading for all **humanity**.' (2006)

Mankind, *humankind*, *the human race* and *human beings* have all become slightly less frequent since 1990.

4.2.2 General English – spoken

The charts below show that, in general spoken English, *people* is the most frequently used word to express the sense 'humans collectively' or 'the human species'. In the 90-94 slice, *man* was the second most frequent synonym, but *man* has become gradually less frequent while *humans* has increased in frequency.





Typical citations with *people* are:

'I think that there is a lot of reflection about **people's** past, **people's** heritage, **people's** history...' (1992)

'But, increasingly, **people** are becoming aware of the vulnerabilities of agricultural resources.' (2003)

Man was very frequent in the 90-94 slice, in citations such as:

'I sort of think the evidence of er a lot of very warm years in the er Mm. the last decade is pretty damn convincing that we have in fact er warmed Mm. it up a bit by **man**.' (1994)

'So I think there is a potential here if we can show that we can in fact reverse the lesions, that would be important beneficially to treatment of the disease in **man**.' (1990)

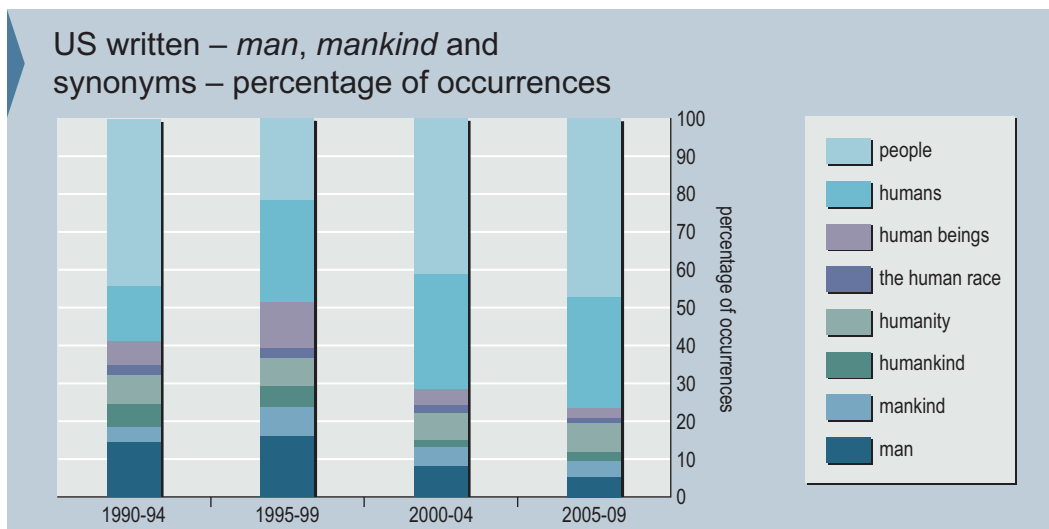
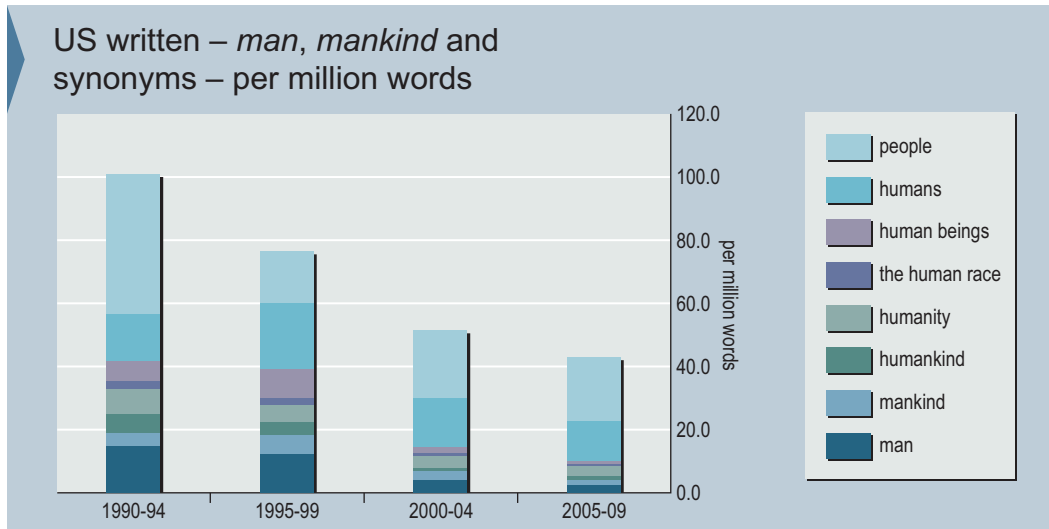
However, the frequency of *man* has decreased dramatically, and seems to have been replaced by *humans*, used in examples such as:

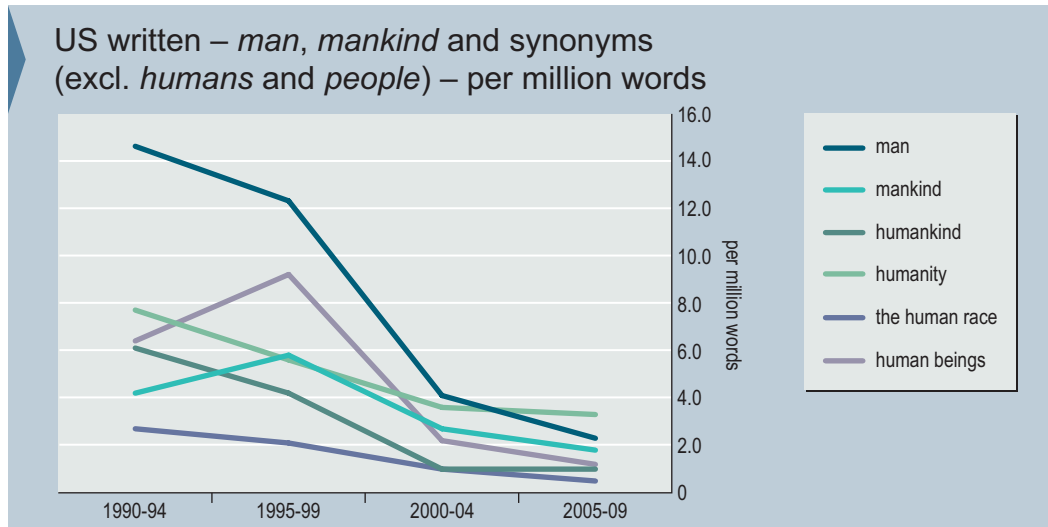
'...and possibly something that, well, we can get to the bottom of and find that difference between chimps and **humans**.' (1999)

'Everyone's a critic. And while this is certainly true of **humans**, we're not the only species on the planet with picky preferences.' (2002)

4.2.3 US English – written

The charts below show that, in all periods, *people* and, to a lesser extent, *humans* are the most frequently used synonyms of *man/mankind*.





Examples include:

'... research that co-mingles human and animal tissue as vital to ensuring that experimental drugs and new tissue replacement therapies are safe for **people**.' (2005)

'It's time to stop the carnage and redirect our efforts into making the world safer for all **people**.' (2005)

'Learning through imitation is central to the mental development of many species, **humans** included'. (2002)

'Insect life and culture may be puzzling to **humans**, but it isn't as disgusting as the kids would sometimes like the old folks to think.' (2008)

It is striking that all the synonyms have decreased in frequency since 1990. This is partially due to the contents of the corpus (which aims to be balanced in text-type and region but cannot always be balanced in terms of subject-matter). Many of the citations with *man*, *mankind* etc. are from several books published in the 90s which discuss topics such as evolution and extra-terrestrial life, and therefore have a high proportion of words for the human species⁶.

The chart below shows the figures for 00-04 and 05-09 only. It is evident that *man*, *mankind*, *the human race* and *human beings* have decreased in frequency, while *humankind* has remained stable but infrequent. *Humanity* is, in current US written English, the most frequent synonym (after *people* and *humans*). Examples include:

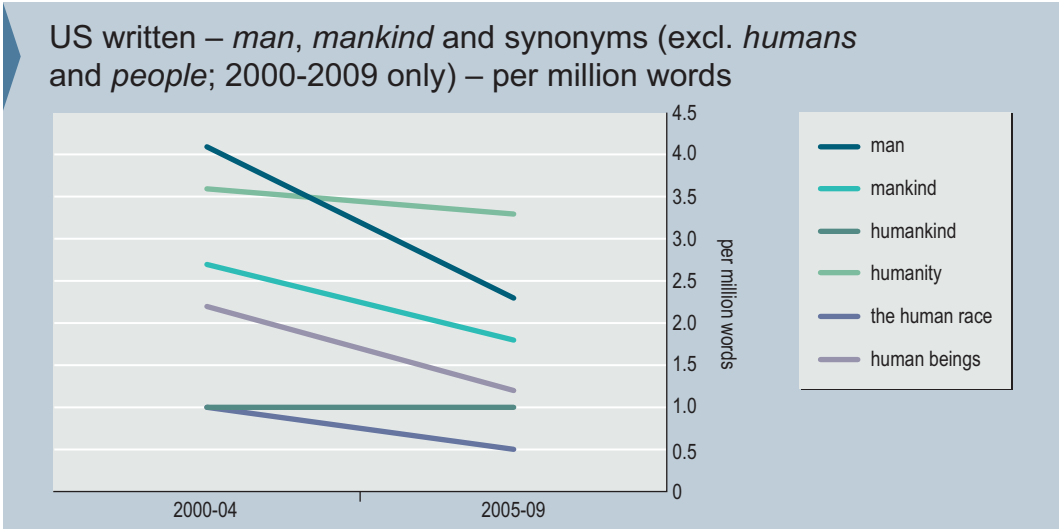
'The modern age is often seen as an awakening of reason from its slumbers, **humanity**'s enlightenment after an age of darkness.' (2005)

'A treaty that takes effect this month could benefit one quarter of **humanity**.' (2008)

'Religion marked a spiritual advance for **humanity** because in its monotheistic forms it implied ethical universalism.' (2008)

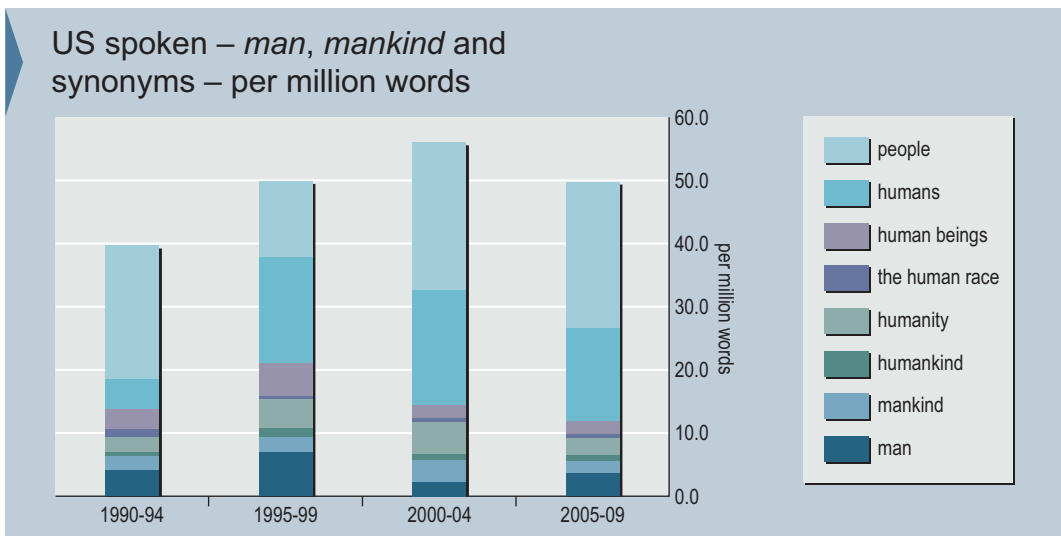
'My fellow pilgrims represented all **humanity**'s faces and cultures, black, white, rich, poor.' (2009)

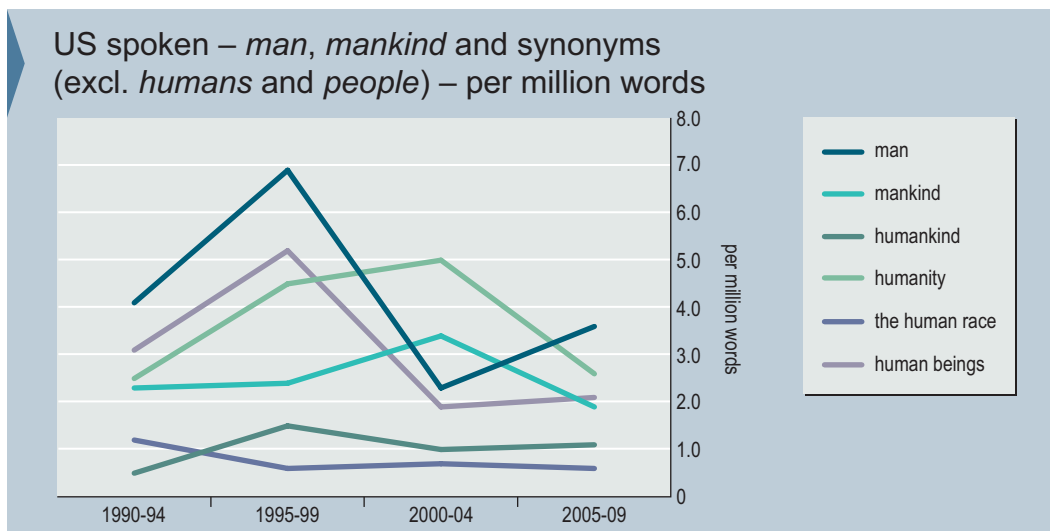
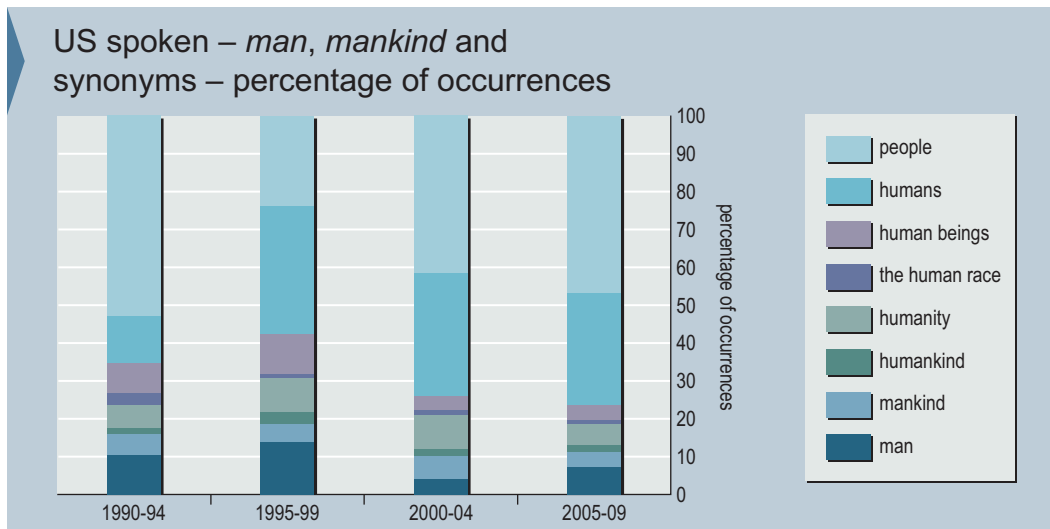
⁶ For example, Wills (1993) *The Runaway Brain: the Evolution of Human Uniqueness*; Kaufman (1994) *No Turning Back: Dismantling the Fantasies of Environmental Thinking*; Davies (1995) *Are We Alone? Philosophical Implications of the Discovery of Extraterrestrial Life*.



4.2.4 US English – spoken

The charts below show that, as in the other corpora, *people* and *humans* are the most frequent synonyms in each period. The frequencies of *man* and *humanity* have fluctuated, and in current US spoken English *man* is slightly more frequent than *humanity* in this sense. Except for a slight increase in 00-04, the frequency of *mankind* has remained quite consistent. *Human beings* has become less frequent, while *humankind* and *the human race* are relatively infrequent in each period.





Examples with *people* and *humans* include:

'Tonight concern and respect for all **people**, all races and creeds and faiths and beliefs, blossomed and bloomed.' (1991)

'So I'm saying, if you can give away your money, why don't you invest to do good to **people** and get back your investment...' (2006)

'...these organisms changed over millions of years to produce all the different kinds of plants and animals, including **humans**.' (2004)

'At least, I believe climate change is a real problem and that **humans** are contributing to it and we need to deal with it.' (2008)

Examples with *man* include:

'... he wound up being the chief architect of the Southern secessionist movement, and of course, one of the most awful wars in **man**'s history is what followed.' (1992)

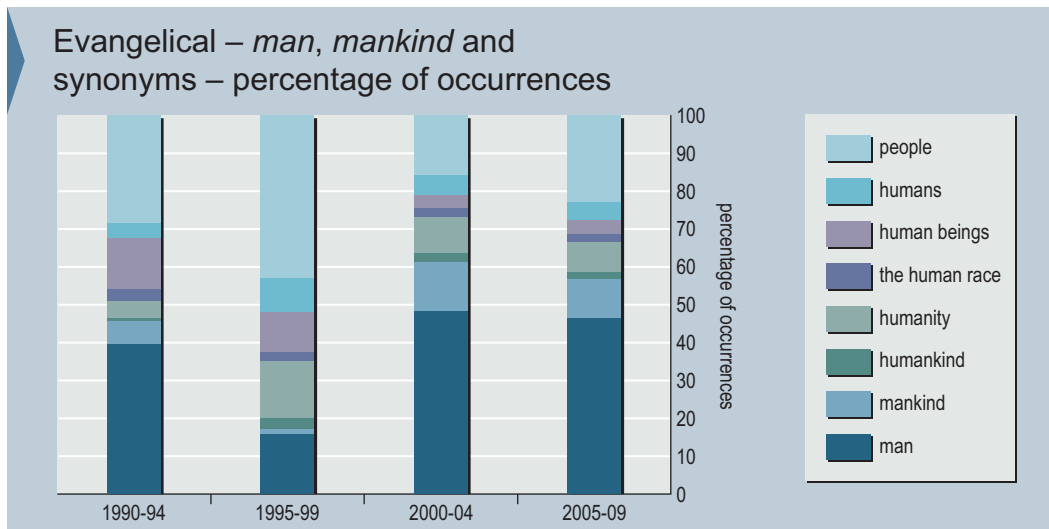
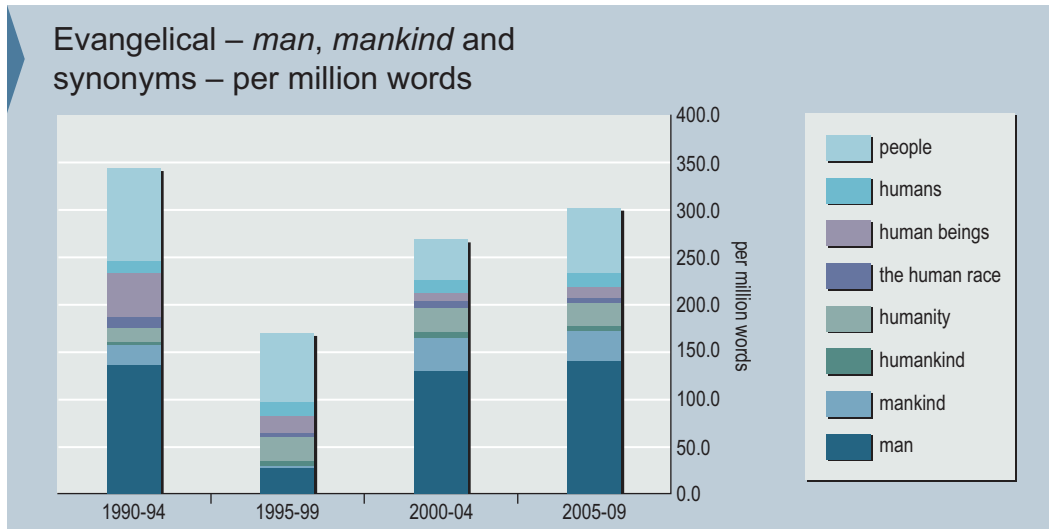
'It's a lot of work to undo the damage that **man** has caused through many years of really ignorance of what they were doing to the environment. (1997)

'We have to redouble our efforts to do what we can to help, not just the victims of the force of nature, but also the victims of the failure of **man**...' (2005)

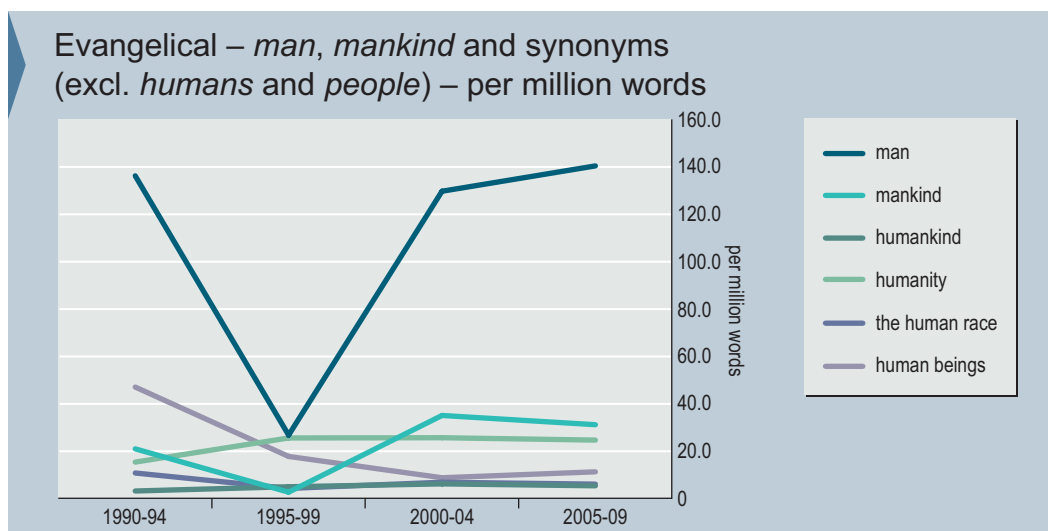
'The Bush administration has consistently said more research needs to be done on why the Earth is heating up, particularly **man's** role in that.' (2009)

4.2.5 Evangelical English

The charts below show that, in contrast with the US written data, *man* is used very frequently in the Evangelical corpus, even compared to *people*. There is a striking dip in the overall frequency of words for 'man/mankind' in the 95-99 slice.



The dip in the 95-99 slice is particularly marked in *man*:



This appears to be related to the contents of this slice. It seems that in this period, sensitivity to the gender language question was at its height, and many of the citations in our samples were *about* gender language, and were therefore not counted as relevant citations. Examples include:

'In the ancient world it was common to say "**man**" or "he" when speaking of all people.' (1997)

'And for some evangelicals, such gender-specific terms as **man** (as a name for the human race) and brothers (as a way of addressing the church) carry with them a theologically significant aspect of human sexuality.' (1997)

It is likely that such sensitivity led to an avoidance of *man* – and *mankind*, which also declined in frequency in 95-99, although to a lesser extent. To some extent *people* was used instead of these words, although it does not fully account for the overall decrease in these synonyms in this period. It may be that generic words for the human race were avoided altogether, or that circumlocutions were used instead; this must remain a topic for future research.

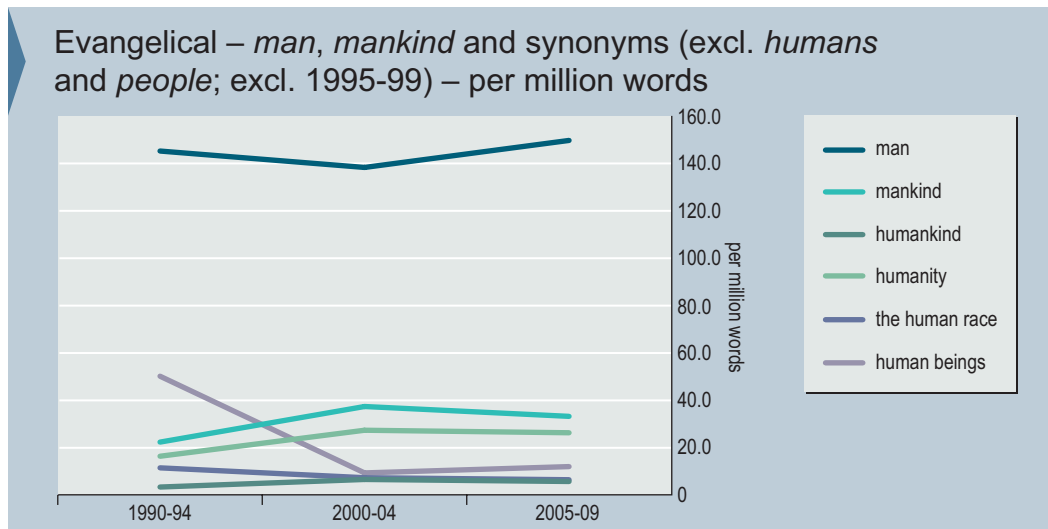
The chart below shows the results with the 95-99 slice excluded. It is evident that, with the exception of the 95-99 slice, synonyms for *man* and *mankind* have remained quite stable. (The only exception is *human beings*, which has become less frequent since 2000.) *Man* is very frequent – examples include:

'Some have thought that the image of God consists in **man's** intellectual ability.' (1994)

'**Man's** concept of peace is based solely upon outside forces.' (2005)

'Do you know what makes **man** the most suffering of all creatures?' (2008)

'Through its [Scripture's] revelation of God's dealings with **man** in the past, we know God keeps His promises.' (2009)



Mankind is, after *man* and *people*, the most frequent synonym, with examples such as:

'Such sanctification cripples **mankind's** ability to face today's dilemmas.' (2005)

'We first see it in Genesis 3 in the fall of **mankind**.' (2009)

Humanity is slightly less frequent, and is found in examples such as:

'God created the world with **humanity** in mind, set us at its center, continued to love us despite our failings, and even sent the Son on a rescue mission.' (2003)

'Francis Bacon called for a change in **humanity's** relationship with the natural world.' (2008)

Humankind and *the human race* are very infrequent in all periods.

4.2.6 Follow-on pronoun use with *man*, *mankind* and *humanity*

When combined with the results outlined under Section 4.1 regarding generic pronoun usage, the data presented above in Sections 4.2.1 – 4.2.5 regarding the use of *mankind*, *man* and their synonyms poses the question, "what happens in cases when *mankind*, *man* and their synonyms are referred to by a follow-on pronoun or determiner?"

The research team conducted a limited study to address this question, examining the frequencies of generic references to *man*, *mankind* and *humanity* when referred to anaphorically by the following pronouns and determiners: *he* (including *he*, *him*, *his* and *himself*); *they* (including *they*, *them*, *their*, *theirs*, *themselves*, *theirselves*, *theirself* and *themself*); and *it* (including *it*, *its* and *itself*). Examples include:

'Clinical ecology shows us how to restore the balance between **man** and **his** environment.'

'When the Almighty himself condescends to address **mankind** in **their** own language...'

'In the so called age of enlightenment and knowledge, **humanity** has managed to lose **its** footing in truth.'

The study was undertaken using the US written English, US spoken English and Evangelical English corpora, covering the period 1990-2009.

In all periods and parts of the corpora, the most frequent of the combinations analyzed was *man* followed by *he*.

In the US written and US spoken corpora, the main rival to *man* followed by *he* is *humanity* followed by *it*, which seems to be the most acceptable option for those wishing to avoid generic *he*. A particularly frequent phrase is

'humanity itself' (for example '... other species, with the exception of humanity itself').

Man, *mankind* and *humanity* are rarely referred to by *they* in the US corpora, even though gender-neutral *they* is the most frequent form used to refer to indefinite pronouns and determiners such as *someone* and *each* (as shown in 4.1 above). Indeed, where anaphoric pronouns and determiners are involved, there may be a tendency to avoid these words altogether and to choose synonyms such as *people* and *humans*, which are both grammatically and conceptually plural.

In the Evangelical corpus, there is a slight increase in the number of citations with *man* and *mankind* followed by *they* in the 00-04 and 05-09 slices. However, *man* followed by *he* is still by far the most frequent combination in all periods.

4.3 Forefather, ancestor and father

For each slice of data, two charts are shown:

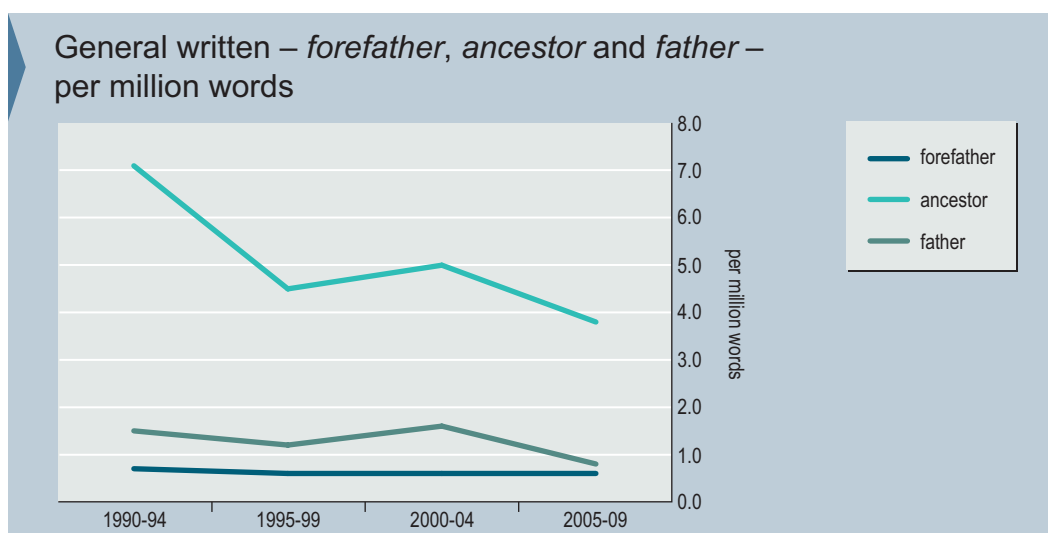
- A. The frequencies of *forefather*, *ancestor* and *father*, per million words;
- B. The relative frequency of *forefather* compared with *ancestor*. It is useful to compare *forefather* and *ancestor*, which are very close synonyms, and exclude *father*, which tends to be used in the slightly different sense 'founder'.

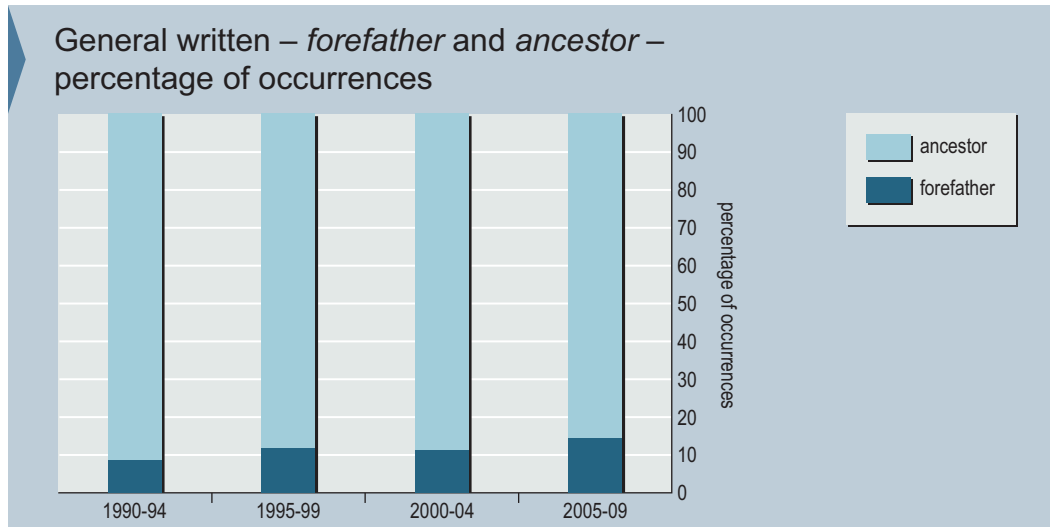
As with *mankind* and synonyms, there are variations in frequencies per million words because of variations in subject-matter: that is, some slices contain more texts which discuss ancestors, and thus have higher frequencies.

A note on *foremother* is given in 4.3.6.

4.3.1. General English – written

As the charts below show, *ancestor* has consistently been more frequent than both *forefather* and *father* in general written English.





Typical examples with *ancestor* are:

'The Duchess's **ancestor**, the great Russian writer, Aleksandr Pushkin...' (2005)

'I am always overwhelmed by the thought that our **ancestors** saw these ancient buildings as theirs in hope and in pain.' (2006)

Forefather tends to be used in the plural, and is often used in more formal or archaic language, e.g.

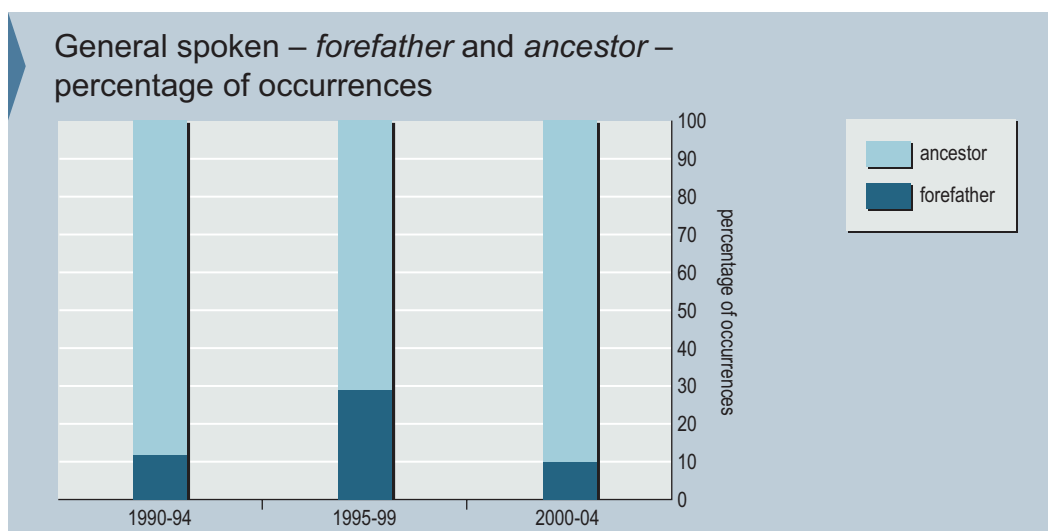
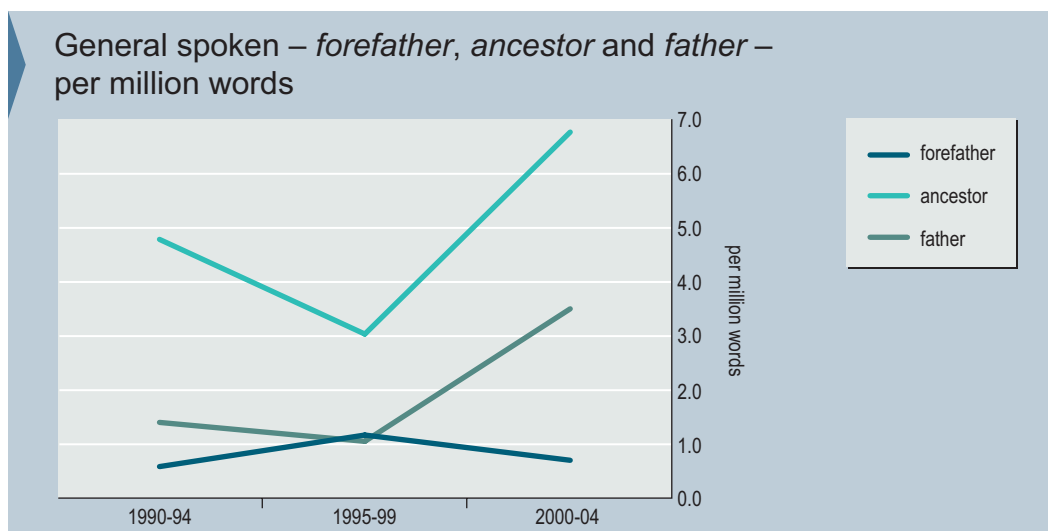
'...pass them on to other tried men from our clan who may keep alive the traditions of our **forefathers** until another chief be born.' (2002)

Most of the occurrences of *father* are of the sense 'founder', e.g.

'In Britain, Cromwell is considered by many to be the **father** of modern democracy.' (2009)

4.3.2. General English – spoken

Ancestor is more frequent than both *forefather* and *father* in general spoken English. The overall frequency per million words of all the relevant words increased in the 00-04 slice. This may be because a lot of these materials are from transcribed radio and television programmes rather than informal conversations (see 4.1.2 above), and the former are, perhaps, more likely to include discussions of ancestors. By examining relative proportions of occurrences, in the second chart, we can see that *ancestor* has consistently been the most frequent. *Forefather* became slightly more frequent in the 95-99 slice, but subsequently decreased in frequency.



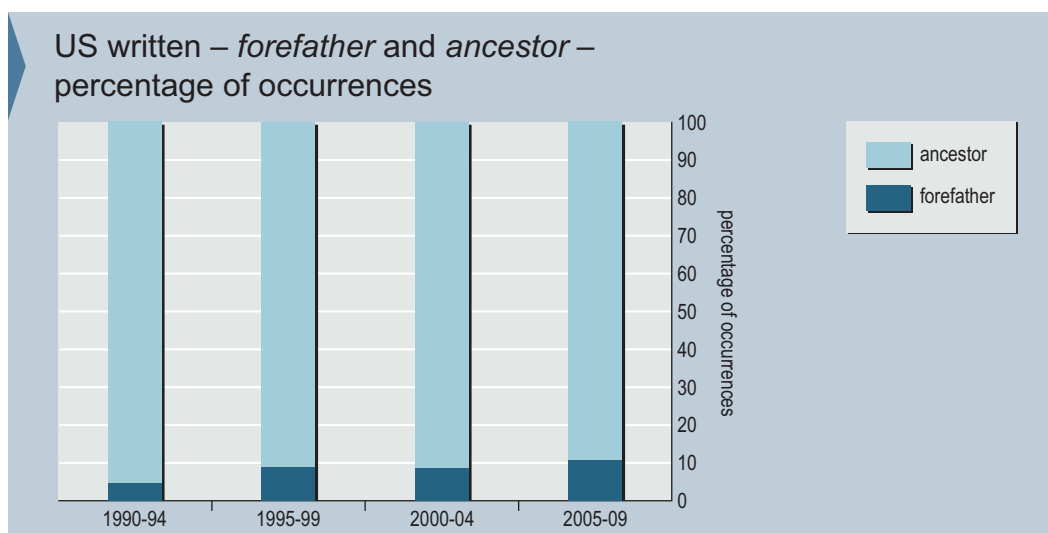
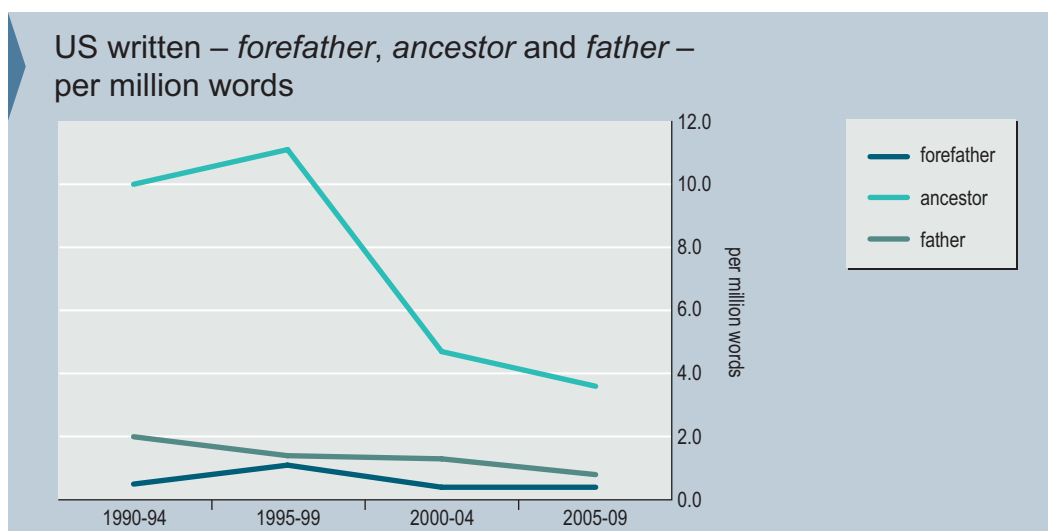
The increase in the use of *father* in 00-04 is largely owing to citations with *the father(s) of* meaning ‘the founder(s) of’, e.g.

‘The **fathers** of the deal forecast a brave new world of technology for what they dubbed the Internet century.’ (2000)

‘Cheryl Sim worked tirelessly for Tommy Douglas, the **father** of Medicare in Canada.’ (2004)

4.3.3. US English – written

As the charts below show, *ancestor* has consistently been far more frequent than either *forefather* or *father* in the sense ‘ancestor’ or ‘founder’.



In the 90-94 and 95-99 slices, there were more citations for all types due to the presence of several books about human evolution. However, by looking at percentages of overall occurrences we can see that the word *ancestor* is most frequently chosen to express this sense, with examples such as:

‘Our European **ancestors** settled here for like reasons.’ (1995)

‘When my **ancestor**, George, Duke of Brunswick-Luneburg, founded the House of Hanover in 1635...’ (2008)

Forefather accounts for fewer than 1 per million words, and fewer than 12% of occurrences, in every slice. Most citations are in the plural form, and many refer to the Bible, e.g.

‘Together they form the holy triumvirate of biblical **forefathers**, the patriarchs, from the Greek words patria, meaning family or clan, and arche, meaning ruler.’ (2001)

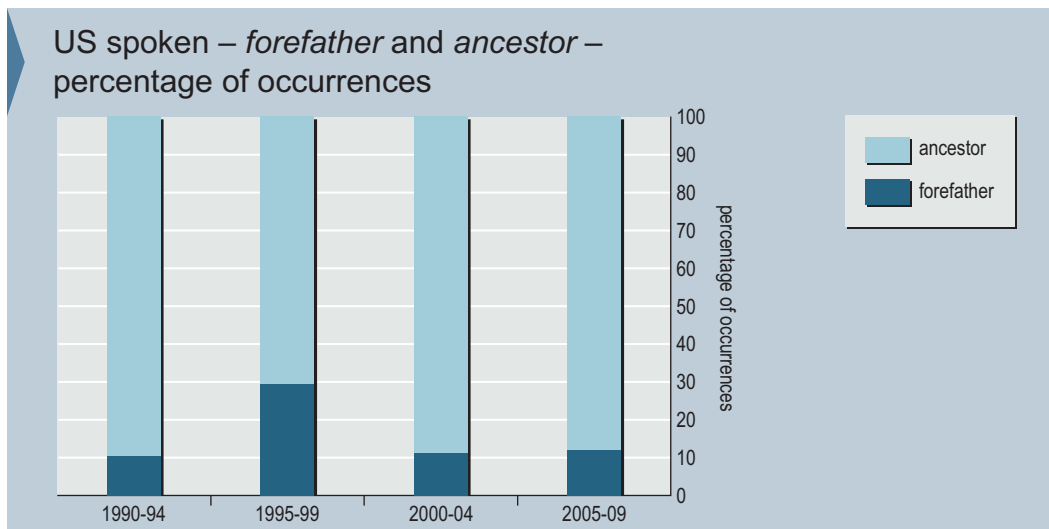
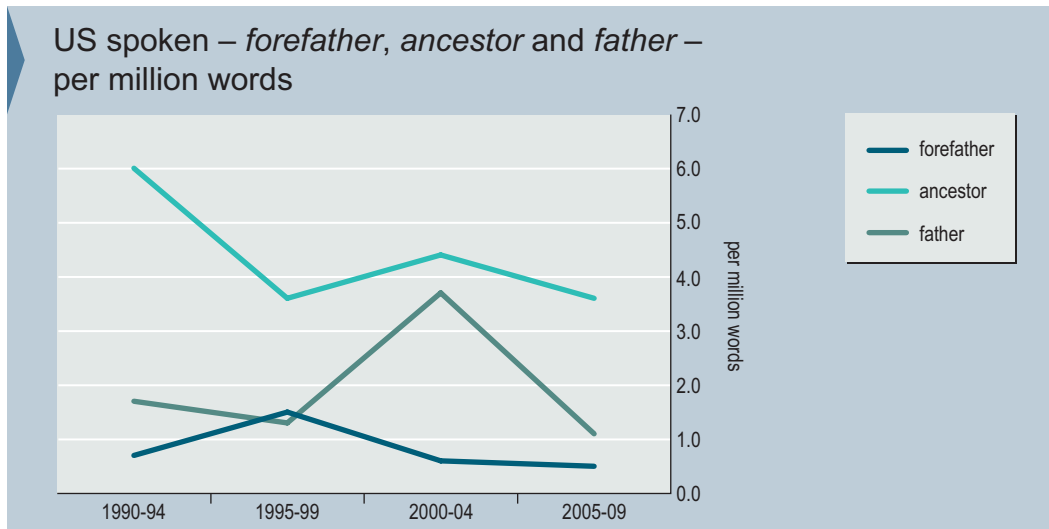
Almost all the citations with *father* are of the sense 'founder, originator', e.g.

'Hugo Grotius, generally regarded as the **father** of international law...' (1995)

'The story of the **father** of Western civilization begins with the absence of birth, a listless despair.'
(2002)

4.3.4. US English – spoken

The charts below show that *ancestor* has been the most frequent usage in all periods. *Forefather* became slightly more frequent in 95-99 but has since become less so. There was a surge in the frequency of *father* in 00-04, and a subsequent decline.



Typical examples with *ancestor* are:

'Today many Penobscot Indians live in the same area where their **ancestors** lived.' (2005)

'Personally, I like to walk the ground where my **ancestors** were, and I feel good about that.' (2007)

As in the other corpora, *forefathers* is usually in the plural. It is often used to refer to the founding fathers of America, as in:

'The reality is that obviously all of us, as American citizens, enjoy a wonderful right in this country that our **forefathers** gave us to exercise the freedom of elections.' (2000)

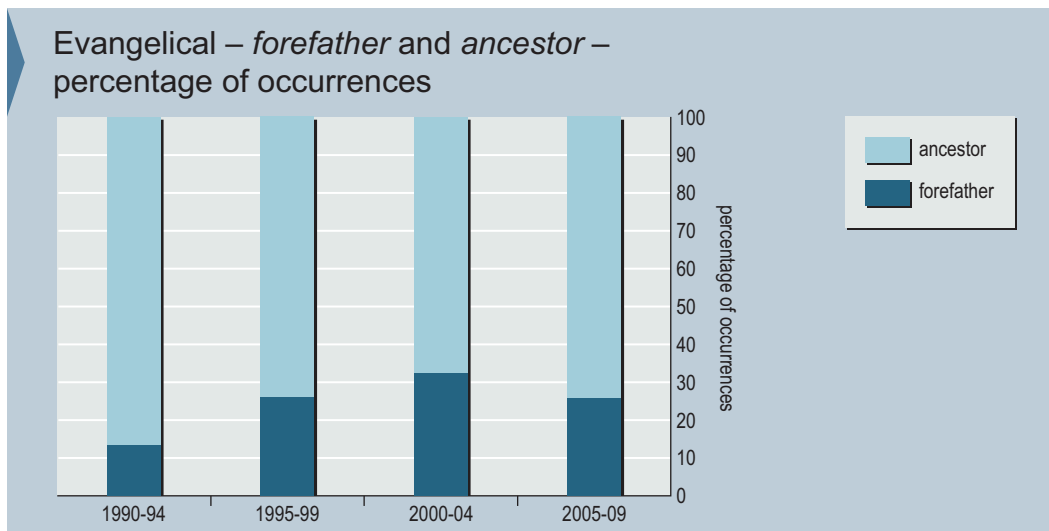
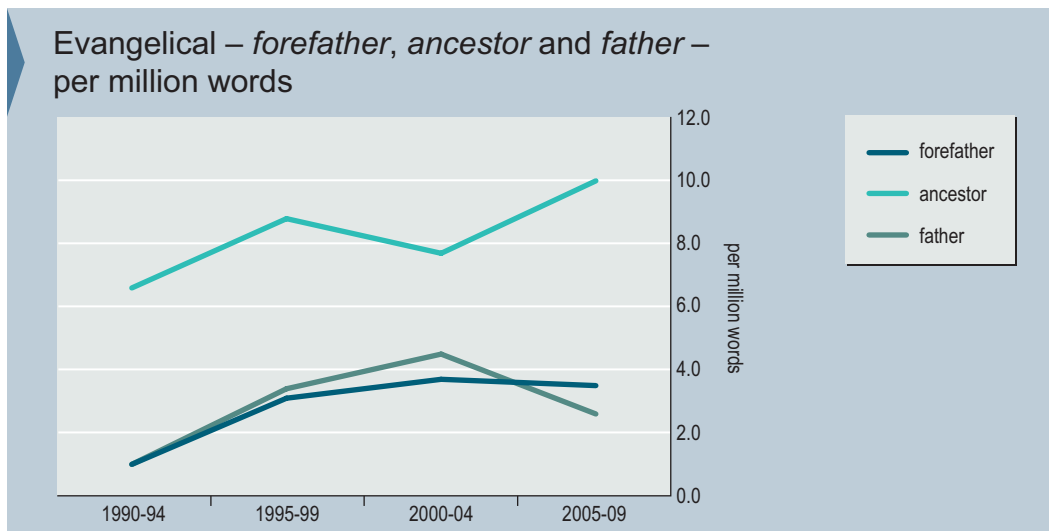
Again, the surge in the frequency of *father* in 00-04 is due to the increased popularity of the phrase 'the father of' in this period, in citations such as:

'We look at Teddy Roosevelt today as a significant figure in conservation history, the **father** of American conservation.' (2001)

'Some say Milan Hodza was the Slovak equivalent of George Washington, the first U-S president Americans refer to as the **father** of his country.' (2002)

4.3.5. Evangelical English

It is evident from the charts below that *ancestor* has consistently been more frequent than *forefather* and *father*.



Forefather is more frequent in the Evangelical corpus than in the other corpora, in citations such as:

'We know, as our **forefathers** did not, that this is simple physiology.' (2004)

'Today, few Americans are aware of the spiritual epidemic that wiped out the land of our Christian **forefathers**.' (2009)

However, *ancestor* is still more frequent in all periods, in citations such as:

'We are all sinners because of our **ancestor**, Adam, which means that we are sinners by nature.' (2003)

'And also the faith that we have is a continuation of the faith and aspirations of our spiritual **ancestors**.' (2009)

There was a slight decrease in the frequency of *ancestor* in 00-04 and a corresponding increase in *father*, but the pattern subsequently reversed. *Father* appears frequently in biblical quotations and phrases (such as 'Abraham is the father of all those who believe' and Satan is 'the father of lies') and in set phrases (such as *founding fathers* and *Church fathers*), but less frequently in natural current usage in the sense 'forefather' or 'founder', although there are some exceptions, such as

'God brought us from there in order to lead us in and give us the land that He swore to our **fathers**.' (2008)

As in the other corpora, the majority of the relevant citations with *father* are in the form 'the father of', e.g.

'Abraham, who was the **father** of the Jewish nation...' (2003)

'Jonathan Edwards became the **father** of the Great Awakening because he knew his purpose.' (2009)

4.3.6. A note on '(fore)mothers'

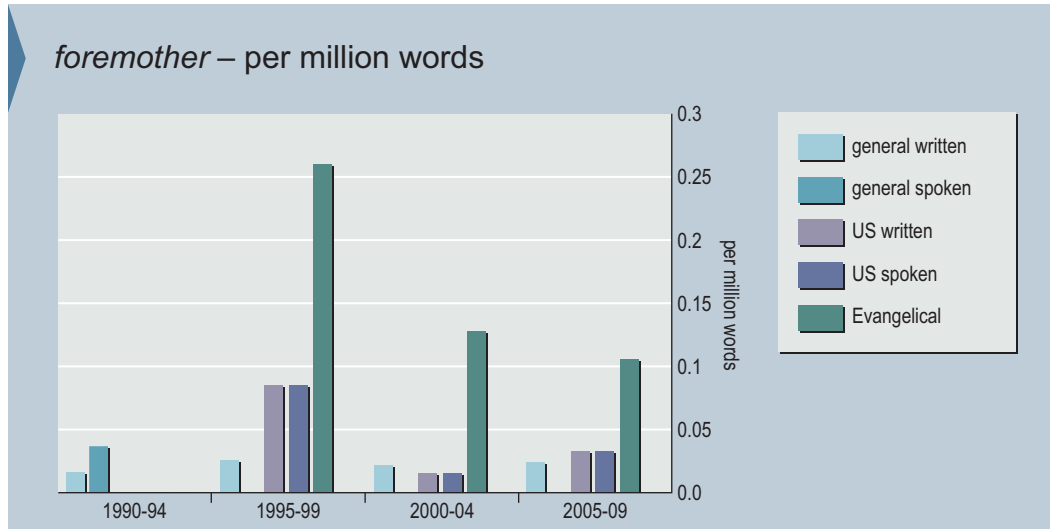
Occasional attempts at gender-neutrality can be seen in the use of *(fore)fathers* and *(fore)mothers*, e.g.

'... we pay great tribute to our **fathers** and **mothers** for their courage in hewing out of the wilderness the making of a great province and a great country...' (General written, 1996)

'She had something else in her—the blood of her **forefathers, and foremothers**, too...' (US written, 2003)

'... we all stand on the shoulders of **our forefathers (and mothers)**.' (Evangelical, 2009)

While it would have been too labour-intensive to search additionally for *mother* in this sense, the uses of *foremother* were infrequent enough for us to check them. The chart below shows the number of occurrences per million words in each corpus. Although the frequencies are too low to make a clear case, two trends are suggested. The first is that *foremother* is used more frequently in the Evangelical corpus than in the other corpora. This is not surprising, given that forefathers and ancestors are, in general, mentioned more often in the Evangelical corpus (see 4.3.5 above). The second trend is that in 95-99 *foremother* was used slightly more frequently; this would be in keeping with the tendency towards more politically correct language in this period.



On the other hand, one interesting usage was found in the US spoken corpus which highlights the fact that, for at least one speaker, *forefather* is not thought of as a markedly male form:

'And I look at the way that I cook my Thanksgiving dinner, and yes, it pays great homage to **my forefathers, especially to my grandmother**, who taught me to prepare that meal'. (US spoken, 1999)