

**Changes in Subject-Verb Agreement with Collective Nouns in  
British English from the 18<sup>th</sup> Century to the Present Day**

Henri Annala  
University of Tampere  
School of Modern Languages and Translation Studies  
English Philology  
Pro Gradu Thesis  
Spring 2008

Tampereen yliopisto  
Kieli- ja käännöstieteiden laitos  
Englantilainen filologia

ANNALA, HENRI: Changes in Subject-Verb Agreement with Collective Nouns in British English from the 18<sup>th</sup> Century to the Present Day

Pro gradu -tutkielma, 89 sivua + liitteet (6 sivua)  
Kevät 2008

---

Tässä pro gradu -tutkielmassa selvitetään kollektiivisubstantiivien ja verbien välistä kongruenssia brittienglannissa sekä sen muutoksia viime vuosisatojen aikana. Koska kollektiivisubstantiivit voidaan brittienglannissa käsitteellistää sekä yksiköllisiksi että monikollisiksi, tutkimuksen pääpainona on kollektiivisubstantiiveja seuraavien verbien luku ja siinä ilmenevät muutokset. Tutkimus kohdistuu yhdeksään eri kollektiivisubstantiiviin: *government, army, tribe, family, crowd, gang, staff, crew* ja *clergy*. Valintaperusteina käytetään kollektiivisubstantiivien mahdollisimman tasaista jakautumista sekä yksikkö/monikkopreferenssin että frekvenssin näkökulmasta. Tutkielmassa pyritään selvittämään onko yksikkö/monikkopreferenssi suhteellisen pysyvä kollektiivisubstantiivien ominaisuus, onko siinä tapahtunut merkittäviä muutoksia viime vuosisatojen aikana, ja mitkä tekijät ovat vaikuttaneet mahdollisiin muutoksiin. Lisäksi tutkielmassa spekuloidaan tulevaisuuden kehitysnäkymillä.

Tutkielmassa käytetään pääosin kahta korpusa. The Extended Version of the Corpus of Late Modern English Texts (CLMETEV) sisältää vanhempia englanninkielisiä tekstejä, kun taas The British National Corpus (BNC) edustaa nykyenglantia. BNC:n niitä osia, jotka koostuvat litteroidusta puheesta tai ennen vuotta 1985 julkaistuista teksteistä ei oteta tutkielmassa huomioon, toisaalta mahdollisimman suuren korpuksen välisen vastaavuuden ja toisaalta mahdollisimman havaittavien erojen saavuttamiseksi.

Teoriaosuudessa käsitellään kriittisesti eri grammatikkojen ja lingvistien näkemyksiä kollektiivisubstantiiveista, verbien luvusta, subjektin ja verbin välisestä kongruenssista, ja jo tunnetuista tekijöistä jotka vaikuttavat kollektiivisubstantiivien seuraavan verbin lukuun. Lisäksi esitellään kaksi lingvististä käsitettä, tunnusmerkisyys (*markedness*) ja leksikaalinen diffuusio, jotka liittyvät keskeisesti kielen muutosprosesseihin.

Tutkielmassa osoitetaan, että yksikkö/monikkopreferenssi on melko pysyvä kollektiivisubstantiivien ominaisuus, mutta löydetään myös selviä merkkejä erilaisista muutosprosesseista. Ne kollektiivisubstantiivit, jotka valitsevat useimmiten yksiköllisen verbin näyttävät ajautuvan yhä enemmän kohti yksiköllistä statusta. Sen sijaan ne kollektiivisubstantiivit, jotka valitsevat useimmiten monikollisen verbin, näyttävät ajautuvan yhä enemmän kohti monikollista statusta yleisen käsityksen vastaisesti. Vaihtelevat kollektiivisubstantiivit puolestaan pysyvät vaihtelevina liikkuen kuitenkin hiljalleen joko yksikön tai monikon suuntaan. Tutkielmassa osoitetaan myös, että kollektiivisubstantiivien seuraavan verbin aikamuoto vaikuttaa verbin lukuun, ja tämä vaikutus on puolestaan kytköksissä kollektiivisubstantiivien muutoksen suuntaan yksikkö/monikkoakselilla: imperfektissä olevat verbit seuraavat muutosta hitaammin kuin preesensissä olevat.

Avainsanat: kollektiivisubstantiivi, kongruenssi, korpuslingvistiikka, subjekti, verbi

## Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Corpus Linguistics</b>	<b>3</b>
2.1	The CLMETEV .....	4
2.2	The BNC .....	5
2.3	The LOB and the FLOB .....	7
<b>3</b>	<b>Theoretical Background</b>	<b>9</b>
3.1	Collective nouns .....	10
3.2	Number .....	15
3.2.1	Singular nouns ending in -s .....	16
3.2.2	Plural nouns not ending in -s .....	16
3.3	Subject-verb agreement .....	18
3.3.1	Grammatical agreement .....	18
3.3.2	Notional agreement .....	19
3.3.3	Proximal agreement .....	20
3.4	Reasons for variation .....	23
3.4.1	Differences according to region .....	23
3.4.2	Differences in style .....	24
3.4.3	Differences in context .....	24
3.4.4	Differences in point of view .....	27
3.5	Language change .....	30
3.5.1	Markedness .....	30
3.5.2	Lexical diffusion .....	33
<b>4</b>	<b>Corpus Analysis</b>	<b>35</b>
4.1	Irrelevant and problematic tokens .....	37
4.1.1	Irrelevant tokens .....	38
4.1.2	Problematic tokens considered relevant .....	41
4.2	Results from the CLMETEV .....	46
4.2.1	The ‘mostly singular’ group in the CLMETEV .....	47
4.2.2	The ‘variable’ group in the CLMETEV .....	50
4.2.3	The ‘mostly plural’ group in the CLMETEV .....	52
4.3	Results from the BNC .....	53
4.3.1	The ‘mostly singular’ group in the BNC .....	56
4.3.2	The ‘variable’ group in the BNC .....	57
4.3.3	The ‘mostly plural’ group in the BNC .....	58
<b>5</b>	<b>Comparison of the Results</b>	<b>60</b>
5.1	Comparison of the ‘mostly singular’ group .....	61
5.1.1	<i>Government</i> followed by the present and past tense forms of <i>be</i> .....	63
5.1.2	<i>Army</i> followed by the present and past tense forms of <i>be</i> .....	65
5.1.3	<i>Tribe</i> followed by the present and past tense forms of <i>be</i> .....	67
5.2	Comparison of the ‘variable’ group .....	68
5.2.1	<i>Family</i> followed by the present and past tense forms of <i>be</i> .....	70
5.2.2	<i>Crowd</i> followed by the present and past tense forms of <i>be</i> .....	71
5.2.3	<i>Gang</i> followed by the present and past tense forms of <i>be</i> .....	72

5.3	Comparison of the ‘mostly plural’ group .....	74
5.3.1	<i>Staff</i> followed by the present and past tense forms of <i>be</i> .....	76
5.3.2	<i>Crew</i> followed by the present and past tense forms of <i>be</i> .....	78
5.3.3	<i>Clergy</i> followed by the present and past tense forms of <i>be</i> .....	79
<b>6</b>	<b>Discussion</b>	<b>80</b>
6.1	Changes in the collective nouns’ number preference.....	80
6.2	Changes in the tense proportions of the verbs following collectives.....	82
6.3	Changes in the number preference from the viewpoint of frequency .....	85
<b>7</b>	<b>Conclusion</b>	<b>88</b>
	<b>Primary Sources</b>	<b>90</b>
	<b>Secondary Sources</b>	<b>90</b>
	<b>Appendix</b>	<b>95</b>

## 1 Introduction

Consider the following examples from the British National Corpus (the BNC):

- [1] The *crowd* was shouting and gesticulating, parents lifting children on shoulders to keep them out of the crush, fists punching the air, workmen's tools being waved like weapons. (GWF 3289)
- [2] 'The camera *crew* were shaking with laughter,' recalled Gavin Birkett. (HRF 1667)

According to the traditional view that most of the grammarians also embrace (e.g. Biber et al. 1999, 188; Carter and McCarthy 2006, 347; *Collins Cobuild English Grammar* 1990, 16; Poutsma 1904, 283; Quirk et al. 1985, 316, 758), the first example highlights the unit and the second one the individuals. The difference in the point of view is often taken to be the ultimate reason behind the variation between singular and plural verb forms following collective nouns, although it does not always seem to be a sufficient explanation. While it is true that grammars provide us with some reasons behind this variation, some recent studies (e.g. Levin 2006, 322; Depraetere 2003, 85) have pointed out that the choice between the singular and the plural could first and foremost be caused by collective nouns' internal word-specific preferences.

The present thesis focuses on investigating this relatively unexplored niche in the field of English grammar. More specifically, the fundamental research questions of the present study are defined as follows:

- (1) Do collective nouns possess an internal, relatively stable preference for singular/plural verbs?
- (2) If they do, has this number preference changed significantly during the past centuries, how has it changed, and which factors have affected the observed change?
- (3) What kind of developments are to be expected in the future as regards collective nouns' number preference?

Since it is nowadays a fairly commonly acknowledged fact that using electronic corpora is the most reliable and efficient way to study linguistic change (see for example Mair 2002,

18), corpora were used in this study as well. For present-day British English, the 100-million-word British National Corpus (the BNC) was used, and for tokens of older British English, the Extended Version of the Corpus of Late Modern English Texts (the CLMETEV; also known as the Leuven corpus) containing approximately 15 million words was used.

First, corpus linguistics and the corpora used in this study are presented. Then, the theoretical foundation of the thesis will be laid by discussing the views of various grammarians and linguists. After that, the focus will shift to the corpus analysis and the comparison of the results. Finally, the results are critically examined and the thesis will be summarised by a conclusion where the research questions set in this section are answered.

## 2 Corpus Linguistics

Kennedy (1998, 1) defines a corpus in a fairly broad manner as “a body of written text or transcribed speech which can serve as a basis for linguistic analysis and description”, and adds (ibid., 3) that corpora “attempt to be representatives of a language as a whole”. This definition is aptly elaborated by Bauer (2004, 98-101), who goes through various properties of a corpus by using polar oppositions: it can be either public (easily available to all) or restricted; electronic (allowing searches) or paper-based (e.g. the Bible, the *OED*); and it can consist of textual material or word-lists (e.g. dictionaries). In addition, corpora are usually fragmented in order to guarantee their representativeness and comparability – i.e. a predetermined number of words from various texts are included, in some cases ranging from instruction manuals to modern poetry. The use of corpora differs from that of other textual material in that fragmented corpora are read for discovering underlying patterns, tendencies, and repeated elements in language, whereas coherent texts (e.g. books and newspaper articles) are read for learning new information or for entertainment (Tognini-Bonelli 2001, 55).

In the 21<sup>st</sup> century, the use of electronic corpora in linguistic research has become increasingly common as well as motivated because of their massive increase both in size and representativeness. Mair (2002, 18) notes that the earlier studies of grammatical change, mostly based on anecdotic observation, have not been able to provide accurate information on the rather slow developments in grammar, whereas electronic corpora have proven to be especially useful when studying diachronic change. Token frequency is one of the most powerful tools that can be used in studying linguistic change with the help of corpora (e.g. Krug 2003, 7). By using frequency, even the ongoing and incomplete grammatical processes can be relatively easily traced in corpus data. Hence, the use of corpora in the present study was logical and justified.

Finally, although electronic corpora have significantly benefited linguistic research, it is an inescapable fact that the researcher is still responsible for the critical analysis of the data (*BNCweb Manual* 2002):

It is important to stress that [corpus searches produce] only raw data – a meaningful interpretation of this data remains the task of the researcher. [A corpus] can't replace human intuition – but it can relieve the careful scholar of a lot of tedious work.

The corpora used in the present study are public and electronic in nature, and consist of fragmented textual material. The following subsections will shortly present those corpora.

## 2.1 The CLMETEV

The Corpus of Late Modern English Texts (the CLMET) consists of untagged Late Modern English literary texts drawn from the *Project Gutenberg* and the *Oxford Text Archive* (De Smet 2005, 69-70). The extended version of the corpus (the CLMETEV) includes additional texts from these two text repositories, and it also contains material from the *Victorian Women Writers project*. The CLMETEV has been divided into three parts according to the year of publishing for each text. The following table displays the structure of each part:

<b>The CLMETEV</b>	<b>Years</b>	<b>Number of words</b>
Part I	1710-1780	3,040,000
Part II	1780-1850	5,720,000
Part III	1850-1920	6,250,000
<b>Total</b>	1710-1920	15,010,000

Table 1: The composition of the CLMETEV

As the table above shows, each of the three parts covers a time-span of 70 years. The oldest part, ranging from 1710 to 1780, is also the smallest with roughly 3 million words, and



the most recent part (1850-1920) is the largest with over 6 million words. According to De Smet (2005, 70-71), four principles have been followed in compiling the corpus:

- (1) The texts included within one sub-period of the CLMETEV are written by authors born with a correspondingly restricted time-span.
- (2) All authors are British native speakers of English.
- (3) Any one author can only contribute a restricted amount of text to the corpus.
- (4) Variation in terms of text genre and authorial social background has been ensured.

While the greatest advantages of the CLMETEV are its fairly large size and manipulability, its structure is somewhat biased in terms of sociolinguistics, genre and register (ibid., 78). However, since the purposes of the present study were not sociolinguistic or genre-dependent, the design of the CLMETEV was quite adequate and its fairly large size was certainly of use.

In order to keep the number of analysed tokens as high as possible and to gain statistically significant results, all three parts of the CLMETEV were analysed as one in the present study (for further information, see section 4 below).

## **2.2 The BNC**

The British National Corpus (BNC) consists of roughly 100 million words of British English, and the 4,054 texts constituting it have been published between 1960 and 1993. Unlike the CLMETEV, it is a tagged corpus, which means that search strings can contain parts of speech. The BNC also includes transcribed speech, and its design is multi-faceted with several subcategories and domains. The following table summarises the structure of the BNC (*BNC User Manual* 2005, 2):

<b>Text type</b>	<b>Number of texts</b>	<b>Number of words</b>
Spoken demographic	153	4,300,000
Spoken context-governed	757	6,280,000
All Spoken	910	10,580,000
Written books and periodicals	2,688	80,490,000
Written-to-be-spoken	35	1,350,000
Written miscellaneous	421	7,550,000
All Written	3,144	89,390,000

Table 2: The composition of the BNC

As the purpose of the present study was to compare the results obtained from the CLMETEV with the BNC results, and because the CLMETEV includes only written material, the spoken parts of the BNC were excluded. In addition, only the texts published between 1985 and 1993 were included in the present study in order to make the temporal gap between the two corpora as wide as possible, hence making the potential differences between the corpora even more discernable (see also section 4 below). The table below displays the structure of the written part of the BNC (*BNC User Manual 2005, 5*):

<b>Domain</b>	<b>Number of texts</b>	<b>Number of words</b>
Applied science	370	7,100,000
Arts	261	6,520,000
Belief and thought	146	3,010,000
Commerce and finance	295	7,260,000
Imaginative	477	16,380,000
Leisure	438	12,190,000
Natural and pure science	146	3,780,000
Social science	527	13,910,000
World affairs	484	17,130,000

Table 3: The domains of the written part of the BNC

Although the BNC has been carefully compiled, and in a way it could be regarded as “a microcosm of current British English” (Aston and Burnard 1998, 29), it is by no means perfect in terms of representativeness. However, for the purposes of the present study and for comparison with the CLMETEV results, the BNC – or more specifically, the BNC subcorpus consisting of written texts published after 1985 – is more than sufficient.

In the present study, the BNC was mostly accessed through the *Sketch Engine* (also known as the *Word Sketch Engine*) interface, which enables the use of tag search strings through the CQL search. The *Sketch Engine* was originally developed for generating word sketches, which are defined as “one-page automatic, corpus-based summaries of a word’s grammatical and collocational behavior” by Kilgarriff et al. (2004, 105). In addition, it has plenty of other useful features which the other interfaces do not have, and therefore it was selected for the present study.

### **2.3 The LOB and the FLOB**

In addition to the CLMETEV and the BNC, the Lancaster-Oslo/Bergen Corpus (LOB) and its later counterpart the Freiburg-Lancaster-Oslo/Bergen Corpus (FLOB) were used to draw additional data if needed (e.g. see section 5.3 below).

The LOB consists of one million words of British English texts from 1961. The texts are divided into 15 categories, and they are carefully selected to represent British English in its whole diversity. The categories are listed in the table below (each sample contains approximately 2,000 words; Johansson et al. 1986):

<b>Text categories</b>	<b>Number of samples</b>
Press: reportage	44
Press: editorial	27
Press: reviews	17
Religion	17
Skills, trades and hobbies	38
Popular lore	44
Belles lettres, biography, essays	77
Miscellaneous (government documents, foundation reports, industry reports, college catalogue, industry house organ)	30
Learned and scientific writings	80
General fiction	29
Mystery and detective fiction	24
Science fiction	6
Adventure and western fiction	29
Romance and love story	29
Humour	9
<b>Total</b>	<b>500</b>

Table 4: The composition of the LOB and the FLOB

The FLOB is exactly similarly constructed as the LOB – only the texts in it are published 30 years later, in 1991. Therefore, the corpora are extremely suitable for diachronic comparisons, although being admittedly small in size compared to the 15-million-word CLMETEV or the 100-million-word BNC.

### 3 Theoretical Background

Let us begin with a language-philosophical question that is embedded deep in the nature of collective nouns. The speakers of English are in a rather special position as regards the use of noun phrases, because the language forces them to express whether they perceive the referent as a countable collection of individuals or as an uncountable entity. According to Downing and Locke (1992, 420), this semantic contrast made on the cognitive plane is “a feature that for English speakers is salient in their experience of ‘things’”, mainly because it is needed for successful communication. Downing and Locke go on by saying that when semantic contrasts become realised in a language, they are often matched by linguistic contrasts (lexical, morphological, syntactical, or intonational). The distinction between countable and uncountable noun phrases is of course morphological and syntactical in nature.

Interestingly enough, Downing and Locke claim that “in languages which do not grammaticalise the distinction, we may presume that either it is not perceived because it is not needed for successful communication, or that it is not expressed because the language does not have the resources for doing so”. Although it is without doubt true that linguistic contrasts reflect semantic perceptions, it does not necessarily entail that the speakers of languages which do not distinguish between countability and uncountability do not perceive that characteristic in a given referent. Downing and Locke might here refer to Whorf’s (2000, 138) famous claim from 1938, according to which Eskimos must conceive of snow differently than English speakers, because Eskimo lexicon has dozens of words distinguishing different types of snow, whereas English has only one. Whorf’s claim has later received loads of well-founded criticism. For instance, Pullum (1991, 170) points out that it could be argued that English does in fact have different words for various forms of snow (*slush, sleet, powder, granular, blizzard, avalanche*, etc.), and Pinker (1994, 61) spoofs the circularity of Whorf’s

claim: “[They] speak differently so they must think differently. How do we know that they think differently? Just listen to the way they speak!”

While it is certainly true that language shapes the way we think to some extent (e.g. Casasanto et al. 2004, 575), it is not exactly unequivocal to draw direct conclusions about how people think by observing their language usage. Therefore, instead of claiming that a certain distinction “is not perceived because it is not needed for successful communication” (Downing and Locke 1992, 420), a more cautious approach is adopted in this study, acknowledging the possibility that although not reflected in the actual language, a certain distinction might still be fully perceived. Collective nouns, denoting at times a countable collection of individuals and at times an uncountable entity, are in the very core of this language-philosophical question (cf. examples [1] and [2] in section 1 above).

The following is a summary of what several distinguished grammarians and linguists say about subject-verb agreement with collective nouns and about differences in the ways singular and plural agreement are used. The sources consulted in this thesis agree on most of the questions regarding collective nouns and subject-verb agreement in English, but there are also some more problematic areas where there were differing points of view. In order to match the theoretic perspective with the diachronic aspect of the present study, the selection of grammars ranges from Sweet’s grammar, published in 1891, to that of Leech and Svartvik’s from 2002.

### **3.1 Collective nouns**

Collective nouns differ from ordinary nouns in that they can be followed by either singular or plural verbs when they are in the singular, and consequently they can take either singular *it* and relative *which* or plural *they* and relative *who* as pronoun coreferents. Quirk et al. (1985, 316) give the following examples:

[3] The *government has* voted and *it has* announced the decision.

[4] The *government have* voted and *they have* announced the decision.

In other words, collectives are at the same time singular and plural nouns. However, collective nouns can also have plural forms constructed by adding the plural inflection *-s* (e.g. *governments*). These forms can naturally take only plural verbs and *they* and *who* as pronoun coreferents. Sweet (1891, 55) defines collective nouns in a very concrete fashion in his old grammar: “[c]ollective nouns express a number of things collected together so that they may be regarded as a single object [. . .] *Crowd* means a number of human beings so close together that at a distance they seem to form a solid mass.” Jespersen’s (1924, 195) definition is rather practical as well: “[collective nouns are] words which denote a unit made up of several things or beings which may be counted separately”.

According to Fowler (1996, 157), there are approximately 200 collective nouns in English, if names of animals (*deer, grouse, sheep*, etc.), terms for groups of animals (*a pride of lions, a gaggle of geese*, etc.), and the names of institutions (*BBC, Marks & Spencer, Tottenham Hotspur*, etc.) are excluded. However, grammarians have been fairly unanimous about the exact boundaries of the term *collective noun*, largely because the concept of *collection* in itself is rather vague (Svensson 1998, 140-141). I will here present two entirely different approaches to the troublesome task of classifying collective nouns.

Fowler’s (1965, 95) classification is mainly based on semantic and pragmatic perceptions.

He divides collectives into seven subgroups:

- (1) Nouns denoting a whole made up of similar parts (e.g. *committee, crew, firm*)
- (2) Nouns that make no plural form but are used as both singular and plural (e.g. *deer, trout, sheep*; or in a few cases, as plural only: e.g. *cattle*)
- (3) Nouns that make plural forms in the normal way but whose singular form may also be used as a plural, sometimes with little if any difference of meaning (e.g. *cannon, duck, fish*; more often with some special implication: e.g. *hair, straw*, and *timber* are used for the mass; *hairs, straws*, and *timbers* for the particular)
- (4) Names of materials used for a collection of things made from them (e.g. *china, linen, silver*)

- (5) Words of number or amount that when used after definite or indefinite numerals have the singular instead of the plural form (e.g. *six brace of grouse, a few hundredweight of coal, dozen, hundred, pound*)
- (6) Abstract singulars used instead of concrete plurals (e.g. *accommodation* (= rooms or lodgings), *kindling* (= pieces of wood), *pottery* (=earthenware articles))
- (7) Nouns denoting substances of indefinite quantity (e.g. *butter, water*)

Although the division is undeniably clever, it evokes some questions. For instance, nouns in group (7) are by many grammarians perceived as mass nouns, not collectives, since they cannot form regular plural forms. Sweet (1891, 56) motivates their perception as mass nouns as follows:

*Iron* means not only the nails and the hammer I may have in my hands at this present moment, but all the iron in the universe, whatever may be the form or quantity of each portion of it. Material words thus make us think more of the attributes they suggest than of the thing itself. Thus iron makes us think of hardness, weight, liability to rust, etc., associated together in a substance of indefinite form.

Depraetere (2002, 90) employs a totally different, morphologically and syntactically oriented classification in her study. The following table illustrates her classification of collective nouns:



Collective noun that:	can be used as a count noun, i.e. base compatible with plural marker <i>-s</i>	can be used as an unmarked count noun with a cardinal number	can be used as an unmarked count noun with a plural demonstrative determiner	can be used as an unmarked count noun with quantifier <i>some</i>
<b>I</b> <i>government, family, crowd</i>	+	–	–	–
<b>II</b> <i>buffalo</i>	+	+	+	+
<b>III</b> <i>crew, people</i>	+	+	+	+
<b>IV</b> <i>deer, cattle clergy, staff</i>	–	+	+	+
<b>V</b> <i>folk, youth</i>	+	–	+	+
Class I:	<i>the governments</i>	<i>*four government</i>	<i>*these government</i>	<i>*some /səm/ government</i>
Class II:	<i>the buffaloes</i>	<i>four buffalo</i>	<i>these buffalo</i>	<i>some buffalo</i>
Class III:	<i>the crews</i>	<i>four crew</i>	<i>these crew</i>	<i>some crew</i>
Class IV:	<i>*the staffs</i>	<i>four staff</i>	<i>these staff</i>	<i>some staff</i>
Class V:	<i>my folks</i>	<i>*four folk</i>	<i>these folk</i>	<i>some folk</i>

Table 5: The morphological and syntactical classification of collective nouns according to Depraetere

The reason why there are two separate classes (II and III) with the same characteristics is that there is a subtle distinction between them: whereas the plural *-s* in class II refers to a ‘set’, in class III it refers to a ‘set of sets’ (Depraetere 2002, 89). In addition, the names of certain animals (class II) have certain restrictions concerning the use of the plural marker *-s*: in the context of hunting, an unmarked plural is used, whereas the plural *-s* is usually used in a more general context.

Although Depraetere’s (2002, 91) division based on the acceptability judgments of native speakers and corpus findings has many advantages, it is not entirely problem-free either, however. For example, although Depraetere believes *staffs* to be an ungrammatical construction, as many as 134 instances of it can be found in the BNC.

These two attempts to classify collective nouns satisfactorily show that the task is rather challenging. In the present study, a fairly broad view is adopted with respect to the exact boundaries of collective nouns. Having said that, the research questions of the present study

(see section 1 above) required certain limitations when selecting the collective nouns for the study (discussed in greater detail in section 4 below).

Collective nouns typically refer to a group of people that are somehow connected with one another (e.g. *committee*, *tribe*), as Leech and Svartvik (2002, 40) point out, and therefore they are also labelled as “group nouns”. However, they can sometimes refer to a group of animals as well – Jespersen (1961, 94) correctly notes that collective nouns can be used not only to refer to a group of people but also to denote a group of living beings:

[5] The whole *fox family* were lying in the sun.

In addition, collective nouns can refer to inanimate objects (e.g. “the *set* [of tools] lies on the table”). However, these kinds of collectives cannot be followed by the plural verb construction according to Jespersen (1961, 94).

Persson (1989, 179) divides collective nouns into three subgroups: the “specific” collectives (e.g. *army*), the “generic” collectives (e.g. *the aristocracy*), and the “unique” collectives (e.g. *the Kremlin*). The “generic” and “unique” collectives are usually preceded by the definite article.

According to Biber et al.’s (1999, 188) corpus findings, singular agreement is the preferred pattern in actual use nowadays. However, they also found out that plural agreement seems to occur almost as often as singular with some collective nouns (e.g. *crew* and *family*), and with a few collectives (e.g. *staff*) plural agreement is the predominant choice. Moreover, some grammarians seem to have slightly different opinions on which of the two is the more preferred or recommended pattern in actual use. This topic will be discussed in greater detail in section 3.4.4 below.

## 3.2 Number

According to Huddleston (1988, 87), number is “an inherent property of the lexical item”. Therefore, it is not merely a separate constituent that can be attached to a word for instance by using a suffix. Words that deviate from the customary way of indicating plurality prove this (e.g. *sheep*, *news*). The notion of number is visible in various parts of speech, including nouns, pronouns, verbs, and numerals. However, as number is essentially a category of nouns and pronouns (Bache and Davidsen-Nielsen 1997, 239) and because of the present study focuses on collective nouns, the following discussion will mainly concentrate on realisations of number from the viewpoint of nouns.

There is a two-term contrast in the number system of present-day English: singular denoting ‘one’ and plural denoting ‘more than one’ (Quirk et al. 1985, 297). Although some grammarians consider the contrast simple enough (e.g. Swan 1980, 427), Fowler (1926, 388) agonises over the numerous difficulties connected to the concept of number in the first edition of *A Dictionary of Modern English Usage*:

Several kinds of mistake are common, & various doubts arise, involving the question of number. With some of them pure grammar is competent to deal; in others accommodations between grammar & sense are necessary or usual or debatable; rarely a supposed concession to sense issues in nonsense.

It is an interesting curiosity that about forty years later, in the second edition of his work, Fowler (1965, 400-401) repeats the whole quotation word by word, apart from one minor exception: the word *rarely* in the last line has been substituted by *occasionally*. It can be speculated whether this small but very meaningful change reflects developments in the tendencies of actual language use or in the linguistic climate of the early 20<sup>th</sup> century.

English is a peculiar language in that the same inflection *-s* is regularly used for denoting plurality with nouns, but singularity with verbs (Quirk et al. 1985, 756). However, there are

some exceptions to that rule as regards nouns. The most important deviations are briefly discussed below.

### 3.2.1 Singular nouns ending in -s

There are a number of singular nouns in English that would at first sight seem to be in the plural because they have the typical plural marker -s. However, they are singular in nature and therefore used with singular verbs (cf. Huddleston's (1988, 87) statement in section 3.2 above). Singular nouns ending in -s include fields of study (e.g. *linguistics*, *mathematics*, *politics*), games and sports (e.g. *billiards*, *checkers*, *darts*), and diseases (e.g. *measles*, *mumps*, *aids*; Swan 1980, 430; Thomson and Martinet 1960, 10; etc.). In addition, there are several fairly frequent nouns outside the aforementioned categories (such as *news*, *means*, and *series*) that select singular verbs despite the plural inflection -s (Quirk et al. 1985, 299).

However, the situation is not as straightforward as it might look: Peters (2004, 24, 26) points out that if some of these nouns refer to "particular objects or instances (and are no longer names)", plural agreement is used (e.g. "Measles are breaking out all over her face."). In addition, some grammarians (e.g. Sweet 1898, 48; Huddleston and Pullum 2002, 181) state that agreement with these nouns is somewhat variable – the following verb may sometimes be in the singular, sometimes in the plural.

### 3.2.2 Plural nouns not ending in -s

Apart from collectives, there are also other nouns that indicate plurality without having the plural marker -s. These include names of animals (e.g. *sheep*, *deer*, *trout*), some nationality nouns (e.g. *the English*, *the Swedish*, *the Dutch*), and quantitative nouns (e.g. *pound*, *foot*, *gross*; Quirk et al. 1985, 307-313; Leech and Svartvik 2002, 319-320; etc.). Again, the list

above is not comprehensive, of course: there are plenty of other fairly frequent nouns not ending in *-s* (such as *aircraft, cattle, people*; Thomson and Martinet 1960, 10; Quirk et al. 1985, 756; etc.).

Although some names of animals, such as *sheep* and *deer*, are formally plural, i.e. they do not have a separate plural form (*\*sheeps, \*deers*), there are also names of animals that have a regular plural form (e.g. *trouts, antelopes*) but can still be used without the plural marker *-s* in the plural (e.g. *trout, antelope*). Hence, it would be just as acceptable to say “we were hunting antelopes” as “we were hunting antelope”. According to Sweet (1898, 44-45), this usage applies only to wild animals “hunted for their usefulness to man”, but not to those killed for other purposes, such as vermin. Therefore, *duck* in “I used to shoot duck” would in fact imply that the ducks were wild (ibid., 44). Some foreign plurals are also variously treated as singular or plural: for instance, both alternatives are equally common with *data*, whereas with *media*, plural agreement is more frequent (Biber et al. 1999, 181).

The borderline between plural nouns not ending in *-s* and collective nouns is all but clear. As is the case with most grammatical concepts, there is no universally accepted definition of a collective noun (cf. section 3.1 above), and therefore different grammarians consider different words collectives. For instance, *Collins Cobuild English Grammar* (1990, 17) quite surprisingly lists *bacteria* as “a common collective noun”, whereas it is usually considered a foreign plural, and Zandvoort (1966, 100) and Schibsbye (1970, 102) count names of wild animals as collectives (see section 3.1 above), whereas Quirk et al. (1985, 307) define them as unmarked plural nouns (see section 3.2.1 above). However, since the present study does not endeavour to define the exact boundaries of the grammatical concept *collective noun*, the matter is not discussed further here.

### 3.3 Subject-verb agreement

Quirk et al. (1985, 755) define agreement (also known as *concord*) as “the relationship between two grammatical units such that one of them displays a particular feature (eg plurality) that accords with a displayed (or semantically implicit) feature in the other”. According to Dekeyser (1987, 255), the most important type of agreement in English is number agreement between the subject and the verb. However, it should be noted that subjects and verbs can also agree in person (Bache and Davidsen-Nielsen 1997, 238). There are three types of agreement in English, all of which are briefly discussed below.

#### 3.3.1 Grammatical agreement

If the verb matches the subject in number, the relationship between them is called *grammatical agreement* – also termed *formal concord* (Dekeyser 1987, 256) or *syntactic agreement* (Corbett 2000, 187). This type of agreement is overwhelmingly normal in English language (e.g. Fowler 1996, 34; Huddleston and Pullum 2002, 499; Quirk et al. 1985, 757).

There are some restrictions concerning the use of subject-verb agreement. According to Biber et al. (1999, 180), there cannot be subject-verb agreement with modal auxiliaries, non-finite and imperative verbs, and subjunctives, because these kinds of verb forms stay unchanged regardless of the subject’s number and person – hence, only verbs in the indicative mood can agree with the subject. In addition, Huddleston and Pullum (2002, 499) point out that subject-verb agreement is restricted to verbs in the present tense, since it is the only tense that displays variation in number and person (with the exception of *be*, which has marked forms even in the past tense).

Although the agreement of 3<sup>rd</sup> person number between the subject and the verb is the most important type of agreement in English (Quirk et al. 1985, 755), it is apparently violated most

often as well, since the fact that singular collective nouns have a tendency to sometimes select plural verbs is one of the most common overrides of the agreement rule (Huddleston and Pullum 2002, 501). This kind of violation against the 3<sup>rd</sup> person number agreement rule is often called *notional agreement*.

### 3.3.2 Notional agreement

According to Biber et al. (1999, 187-188), there are three major types of notional agreement: (1) names, titles and quotations (e.g. *The United States, The New York Times*), (2) measure expressions (e.g. *two pounds, eighteen years*), and (3) collective nouns. In addition, Fowler (1996, 531) ascribes the use of singular agreement with nouns ending in *-s* (discussed in section 3.2.2 above) to notional agreement. Regardless, only the third group is relevant in view of the purposes of the present study, and therefore no further space is devoted to discussing the other instances of notional agreement.

When used in connection with collective nouns, the term *notional agreement* – also known as *referential concord* (Dekeyser 1987, 256) or *semantic agreement* (Corbett 2000, 187) – means that the verb agrees with the plural notion embedded in the collective noun rather than with the singular form it has (Leech and Svartvik 2002, 274): in other words, there is an agreement mismatch (Corbett 2006, 159). Quirk et al. (1985, 757) point out that because traditional prescriptive teaching has insisted on grammatical agreement, even native speakers of English are often uncertain whether the use of plural verbs with singular collective nouns is grammatical or not. Although this prescriptivism may have affected the choice between notional and grammatical agreement significantly in the course of history, the extent of its effects remains an unanswered question for the time being.

Even though the use of plural verbs with singular collective nouns is nowadays totally accepted and grammatical, this has not always been the case. Jespersen (1961, 94) considers

the use of plural verbs with singular collective nouns as a universal example of how people tend to forget the rules of grammar: “[T]here is in all languages and at all times a tendency to forget the fact that collectives are grammatically singular, and we often find plural constructions”. Furthermore, he adds that this tendency may even be stronger in English than in most other languages, because so few verb forms in English show variation in number (see also section 3.1 above). However, the way Jespersen begins the sentence cited above reveals that he is fully aware that this tendency is very common and unavoidable.

Notional agreement is closely connected to the way the referent is perceived in its context. It is found only with collective nouns denoting living beings – as opposed to inanimate objects (Jespersen 1961, 94; see also section 3.1 above). In addition, according to Swan (1980, 428), notional agreement is generally used when a collective noun denoting a group is thought of as individuals. However, Leech and Svartvik (2002, 274) point out that the distinction is not always easy to perceive (cf. example sentences [1] and [2] in section 1 above).

### 3.3.3 Proximal agreement

The term proximal agreement – sometimes also referred to as *attraction* (Leech and Svartvik 2002, 275; Quirk et al. 1985, 757) – signals by definition “agreement with the number of the nearest noun” (e.g. Dekeyser 1987, 256; Fowler 1996, 36; Peters 2004, 23) or with a closely preceding noun phrase (Quirk et al. 1985, 757). Quirk et al. offer an example:

[6] *No one* except his own supporters *agree* with him.

The subject [*n*]o *one except his own supporters* in the example sentence above is clearly singular because of its singular head [*n*]o *one*. However, the plural verb is triggered by the closely preceding plural noun *supporters*.



As in example [6] above, proximal agreement often works together with notional agreement (“only his own supporters agree with him”). Some common head nouns, such as *number* and *majority*, are more likely to trigger the principle of proximity than others, because they convey the idea of plurality despite their singular form, as Leech and Svartvik (2002, 275) note. Sometimes even a coordinated noun phrase may influence the choice of the verb (Quirk et al. 1985, 757):

[7] A good *knowledge* of English, Russian, and French *are* required for this position.

Distance plays a significant role in proximal agreement: the further away from the verb the head noun is, the more likely it is that an intervening noun may influence the number of the verb (Jespersen 1961, 94). Quirk et al. (1985, 756) also note that the principle of proximity is more likely to effect a change from the singular to the plural than vice versa, probably because the plural is the unmarked verb form (for further discussion on markedness, see section 3.5.1 below).

Proximal agreement is one of those phenomena that have existed quite a long time in the English language without being accepted as part of its “official” grammar. Fowler (1926, 389) describes proximal agreement quite vividly in the first edition of *A Dictionary of Modern English Usage*, published in 1926:

Some writers are as easily drawn off the scent as young hounds. They start with a sg subject; before they reach the verb, a pl noun attached to an *of* or some other similar distraction happens to cross, and off they go in the pl; or vice versa. This is a matter of carelessness or inexperience only, and needs no discussion; but it is so common as to call for a few illustrations: “This argument for Mr. Macmillan’s attending the Assembly does not alter the irksome fact that he, in common with President Eisenhower, are dancing to Mr. Khrushchev’s tune.”

The attitude to proximity principle has not significantly changed from Fowler’s early views: Quirk et al. (1985, 757) condemn the use of proximal agreement “in defiance of the other principles” (probably meaning formal and notional agreement) as erroneous; Biber et al. (2002, 236) together with Greenbaum and Quirk (1990, 215) associate proximal agreement

mainly with unplanned discourse and say that it should be corrected if noticed in writing (cf. section 3.4.2 on style); Huddleston and Pullum (2002, 501) refer to an instance of proximal agreement as “a processing error”; and Reid (1991, 203) uses the term “Speaker Error” to refer to manifestations of proximal agreement.

On the other hand, there are also some less strict views: Huddleston and Pullum (2002, 501) weaken their aforementioned statement later by saying that although there clearly is no rule permitting the use of proximal agreement in the English grammar, there are some cases where proximity is “a relevant factor in more acceptable departures from simple agreement”. They give an example where the head of the noun phrase is *number*, one of the nouns Leech and Svartvik (2002, 275) list above among those which are relatively likely to trigger proximal agreement:

[8] A *number* of special units *are* available for patients requiring hospitalisation.

In addition, Huddleston and Pullum (2002, 501) contrast the views of Biber et al. (2002, 236) and Greenbaum and Quirk (1990, 215) by saying that proximal agreement is not uncommon even in written texts (cf. section 3.4.2 below).

There are some special cases where the proximity principle is often in use. Firstly, as Peters (2004, 26) notes, when proximal agreement is reinforced by notional agreement, it “prevails in many kinds of writing”, but on its own it is usually played down. Secondly, proximal agreement is usually resorted to when noun phrases of different numbers are involved, or when the complex subject combines different persons. Dekeyser (1987, 259) provides us with examples:

[9] Either *the teacher or the students are* to be blamed.

[10] Either *John or I have* to go.

Although Dekeyser does not directly mention it, it is fairly reasonable to assume that proximal agreement in these kinds of sentences is also more likely to take effect from the singular to the plural than vice versa.

### **3.4 Reasons for variation**

Fowler aptly describes the variation in subject-verb agreement with collective nouns (1926, 390; 1965, 402):

In general it may be said that while there are always a better & a worse in the matter, there are seldom a right & a wrong, & any attempt to elaborate rules would be waste labour.

Although Fowler must be credited for recognising that the choice between the singular and the plural can often be dictated by stylistic factors rather than strict grammatical rules already in the 1920s, investigating the patterns of actual language use and thereby unveiling potential regularities is not considered “waste labour” in the present study.

There are a number of commonly offered explanations to the question why there is sometimes a singular verb following a collective noun and sometimes a plural one. The following subsections will briefly illustrate the reasons various grammarians and linguists offer for this variation.

#### **3.4.1 Differences according to region**

One of the explanations for singular/plural variation is that the plural construction is more typical of British English, whereas it is far less common in American English (Leech and Svartvik 2002, 40). Moreover, Biber et al. (1999, 188) note that nearly all collective nouns occur at least occasionally with the plural construction in British English. This regional variation is rather exceptional, because it is one of the few grammatical differences between British and American English, as Quirk et al. (1985, 19) point out. However, the regional factor is irrelevant for the purposes of the present study, since it focuses only on British English.

### 3.4.2 Differences in style

The second explanation for the variation in subject-verb agreement with collective nouns is difference in style. The plural is more common in informal speech and the singular in formal writing (e.g. Svensson 1998, 139; Huddleston and Pullum 2002, 502). This difference is recognised by only a couple of grammarians, and Quirk et al. (1985, 758) show their uncertainty about this variation by saying that there is “probably” a difference in style. However, Huddleston and Pullum (2002, 502) emphasise that although the plural construction is more common in informal style, it is still a fully grammatical construction in English.

### 3.4.3 Differences in context

The third reason behind the number variation identified by grammarians and linguists is contextual differences (e.g. Westergren-Axelsson 1999, 294). Some elements in the context can strongly affect or even wholly determine the number of the verb following a collective noun.

First of all, the semantic characteristics of certain verbs following collective nouns require singular agreement. Huddleston and Pullum (2002, 502) say that the plural is not possible if the verb is applicable to the group as a whole but not to the individual members, and Biber et al. (1999, 189) give an example:

[11] The *committee comprises/consists of/has* eight members.

In fact, this phenomenon is not only restricted to the properties of the following verb: in case of *has* in example [11] above, it is actually the complement *eight members* that require the singular (see also section 3.4.4 below).

Secondly, the choice of personal pronoun may affect the number of the verb through internal coherence of the sentence (own examples):

[12] The *government itself* was responsible for the new law.

[13] The *government themselves* were responsible for the new law.

However, as Quirk et al. (1985, 759) point out, sometimes the verb and the pronoun do not agree in number:

[14] The *committee has* not yet decided how *they* should react to the Governor's letter.

Thirdly, some determiners or quantifiers preceding the collective noun affect the choice between the singular and the plural. First of all, there are determiners that require either a singular or a plural head according to several grammarians. The determiners requiring a singular head include *much* and *every*, whereas *several*, *few*, *both*, and the cardinal numerals require a plural head (Huddleston 1988, 88). In addition, Biber et al. (1999, 184) add that *more* and *most* prefer plural agreement as well, and according to Huddleston and Pullum (2002, 502), plural agreement is of questionable acceptability if a collective noun is preceded by *one*, *a*, *an*, or *another*:

[15] One *committee*, appointed last year, *has/?have* not yet met.

Moreover, Poutsma (1904, 283) and Jespersen (1961, 99) note that the singular is the ordinary construction if a collective noun is preceded by *the whole*.

In addition to the aforementioned determiners that seem to dictate the number of the verb in a rather straightforward fashion, there are also some determiners that divide grammarians' opinions. As regards *all*, the grammarians seem to have surprisingly heterogeneous views. Poutsma (1904, 283) claims the singular construction to be the norm with *all*; surprisingly, Biber et al. contradict Poutsma's claim in pointing out that the plural is on one hand "the rule" with *all* (*Longman Grammar of Spoken and Written English*, 1999, 184) and on the other hand that it "can take either sg or pl concord, according to whether [it has] sg or pl reference" (*Longman Student Grammar of Spoken and Written English*, 2002, 234); and finally,

Huddleston and Pullum (2002, 502) state that the plural construction is “virtually required” in connection with *all*:

[16] The *class have/?has* now *all* received certificates of merit.

Grammarians seem to be just as unanimous on the number of the verb with *some*, *many* and *each*: Biber et al. (1999, 184; 2002, 234) claim on one hand plural agreement to be the rule with *some*, and on the other hand both alternatives to be acceptable. According to Huddleston (1988, 88) and Biber et al. (1999, 184), *many* requires the plural, while Jespersen (1961, 106-107) aptly notes that “difficulties often arise in connexion with the [word] *many*”. According to the first edition of Fowler’s *A Dictionary of Modern English Usage* (1926, 391) *each* is “all singular, that is undisputed”; Biber et al. (1999, 184) and Dekeyser (1987, 258) agree by stating that the singular is the normal choice with *each*; and finally, Fowler (1996, 35) weakens his claim in the third edition of his work by noting that sometimes “contextual considerations lead to the use of a plural verb” with *each*. It could be speculated whether the diachronic change that is visible in Fowler’s comments has happened in the English language itself or in the attitudes from prescriptivism to descriptivism (cf. section 3.2 above).

As regards *any*, *none* and *no*, Biber et al. (2002, 234) and Fowler (1996, 35) state that they can be followed by either a singular or a plural verb. This is confirmed by Leech and Svartvik (2002, 276), who point out that the aforementioned indefinite expressions of amount often cause agreement problems. Biber et al. (1999, 184) also mention that the singular was once considered correct with these expressions, but there is apparently “little sign of such a deliberate preference these days”. However, this tradition is obviously still recognised at least by Swan (1980, 429), who associates the singular construction after *none* with formal style and the plural construction with informal style (see also section 3.4.2 above). Dekeyser (1987, 258) adds that “formal discord is very common with *none* even in ‘careful’ English”. Consider Fowler’s (1996, 35) examples:

[17] [*N*]one of her *features* is particularly striking.

[18] *None* of our fundamental *problems* have been solved.

Moreover, there is a group of expressions sometimes preceding collective nouns that have been defined as *number-transparent* in several grammars. According to Huddleston (1988, 87-88), these expressions “allow the number of the NP embedded within the *of*-phrase complement to flow through”, and therefore they cannot affect the choice between the singular and the plural. These expressions include *number of*, *rest of*, *remainder of* and *proportion of* (ibid., 88). The following examples are from Carter and McCarthy (2006, 347):

[19] The *rest of the European Union* now *seems* ready to reluctantly bow to British pressure that justice and home affairs remain primarily a matter for intergovernmental cooperation. (the subject treated as a unified body)

[20] The *rest of the cast* are okay as well but no one really stands out, and the special effects are merely okay, too. (the subject treated as composed of separate individuals)

Finally, two noun phrases coordinated by *and* require a plural verb, but there is variation with *or/nor*: Biber et al. (1999, 183) note that the verb is often in the singular, but could occasionally be in the plural as well, especially if one of the coordinated noun phrases is in the plural or bears a notion of plurality (e.g. collective nouns). Consider the following examples by Biber et al.:

[21] Neither geological evidence *nor physical theory* supports this conclusion. (both noun phrases in the singular)

[22] Neither Lorillard *nor the researchers* who studied the workers *were* aware of any research on smokers of the Kent cigarettes. (one of the noun phrases in the plural)

### 3.4.4 Differences in point of view

The fourth and most commonly offered explanation for the variation between the singular and the plural constructions is that the two forms convey two slightly different meanings: the singular is used when the focus is on the group as an undivided body, and the plural when the

focus is on the group as a collection of individuals (e.g. Visser 1963, 62). Consider the following examples by Quirk et al. (1985, 758):

[23] The *audience was* very large.

[24] The *audience were* really attentive all the time.

Indeed, the audience is clearly thought of as a whole in [23], whereas the individuals making up the audience are in focus in [24]: in other words, there is grammatical agreement in [23] (see also section 3.3.1 above) and notional agreement in [24] (see also section 3.3.2 above). Surprisingly many grammarians even go as far as saying that the variation between the singular and the plural with collective nouns is solely caused by the differences in point of view – for instance, Chalker (1990, 138) plainly claims that “*family* is a collective noun that can take singular or plural verb according to the meaning”, and gives two examples:

[25] The *family is* the basis of most societies all over the world.

[26] My *family have* always lived in the East End of London.

This oversimplification might – at least hopefully – be caused by the desire to make the English grammar as general and easily approachable as possible.

As mentioned in section 3.4.3 above, certain verbs require singular agreement. Moreover, the use of the singular with collective nouns extends far beyond those verbs: in fact, the singular “is the predominant choice with a wide range of verbs” (Biber et al. 1999, 189; see also section 2.1 above). Hence, it happens that the singular is sometimes used even if the focus is clearly on the individuals (Poutsma 1904, 286):

[27] The great *crowd was* most enthusiastic.

[28] We must have a good goose to give to this gentleman in place of the one which your *family is* now devouring.

On the other hand, Poutsma (1904, 286) also lists some cases where the plural is used despite the focus on the group as a whole:

[29] The *army of the Queen mean* to besiege us.



[30] The *Committee* chosen to examine Dr. Cook's records *have* decided against the explorer's claim.

Poutsma's example sentence [30] above illustrates well the ambiguous nature of collective nouns: it is debatable whether the focus is on the committee as a whole or as a collection of individuals, i.e. did the committee make the decision jointly or separately? Both interpretations would seem to be plausible in the above context, although the focus does seem to be slightly more on the committee as a whole, particularly because of the predication involving a decision: that is what committees are for. However, there are also clear-cut cases where only one of the forms is practically possible:

[31] The *committee are* having their lunch.

[32] The *committee has* been abolished.

Nevertheless, Poutsma (1904, 283) counts the aforementioned vacillation (the use of the singular even if the focus is on the individuals and vice versa) as lack of thinking: "it is but natural that writers and speakers do not always pause to think whether the communication given applies to the individuals of which a collection is composed jointly or separately". Leech and Svartvik (2002, 274) agree by saying that it is often difficult to distinguish between the two meanings, which is also the case in Poutsma's example sentence [30] above. This claim is further corroborated by Biber et al. (1999, 189), who give an example where the number of the verb following the collective noun *team* shifts within the same context (cf. section 3.4.3 above):

[33] Most of those seasons have involved a struggle against relegation and, while they have been lucky to survive some, this *team are* good enough to stay up [. . .] Indeed, he says: "This *team is* good enough to stay up but whether we will I don't know."

Rather surprisingly, Poutsma (1904, 284) also notes that although the choice between the singular and the plural is at least partly a matter of personal taste, "attentive reading will show that on the whole the plural construction is more in favour than the singular". This contrasts sharply with what Biber et al. (1999, 189) say above. However, the two contrasting views

seem to make perfect sense in the light of Bauer's study on linguistic change in British English in the 20<sup>th</sup> century. The key finding of this study regarding collective nouns is that there appeared to be a certain shift away from the plural constructions following collective nouns (in other words: notional agreement; see also section 3.3.2 above) in *The Times* corpus during the 20<sup>th</sup> century (Bauer 1994, 65). The fact that this shift was perceived in the data taken from *The Times* corpus, which consists of 20<sup>th</sup> century editorials written in formal British English, makes the finding even more significant, since it is a fairly generally acknowledged fact that formal written language is most resistant to linguistic change. Therefore, it could be possible to postulate that the shift from notional agreement to grammatical in the 20<sup>th</sup> century was of a universal character, involving all types of spoken and written English. However, these kinds of conclusions should not be drawn too hastily, especially if there is no research to confirm them.

### **3.5 Language change**

In this section, two concepts relevant to the study of language change, especially from the collective nouns' angle, are discussed. In the literature, these two concepts are referred to as *markedness* and *lexical diffusion*. As for both of these terms, frequency is an important factor, since it is often claimed to be the most accurate tool for perceiving slow linguistic change (Levin 2006, 325).

#### **3.5.1 Markedness**

The concept of markedness originates from the Prague School of phonology, but has later been adopted to describe other aspects of linguistics as well. Markedness is based on polar oppositions, which are aplenty in the realm of grammar: positive and negative, masculine and

feminine, present and past, and singular and plural, just to name a few (Battistella 1990, 1). The principle of markedness was first created by Roman Jakobson and Nikolai Trubetzkoy, and later developed by Noam Chomsky, among others. According to this principle, terms of polar oppositions, such as the grammatical pairs mentioned above, are not merely opposites; instead, “they show an evaluative nonequivalence that is imposed on all oppositions” (ibid., 1). What this statement basically means is that whenever there is a binary choice in grammar, the two options are never precisely equal: an element of evaluation is always present. This is caused by the fact that one of the two choices is always perceived as more general or simpler than the other – the simpler pole is the normal, unmarked term whereas the more complex pole is the special, marked term (Battistella 1996, 10). Hence, the unmarked poles are often associated with a certain notion of impartiality or neutrality (Cruse 1986, 247).

Let us illustrate the concept of markedness by two examples from different areas of language use. From the semantic angle, *tall* is the unmarked pole of the pair *tall–short*, since it is the one that is used in a neutral context: “How tall is she?” sounds perfectly normal, whereas “How short is she?” implies that the speaker expects the person to be short (Levin 2006, 326). Similarly, grammatical contrasts are influenced by markedness. For instance, singular nouns are unmarked, whereas plural nouns are marked because an overt mark (or “an extra morpheme”; Moravcsik 1988, 93) is added to them: hence *dog* is the unmarked form and *dogs* is the marked one.

The concept of markedness is closely connected to grammaticalisation: when a linguistic item is affected by grammaticalisation, it automatically becomes more unmarked, as Lehmann (1989, 176) points out. Despite the obvious similarities between the two concepts, they differ in many aspects as well: for instance, markedness is directly based on oppositions whereas grammaticalisation is not. In addition, grammaticalisation could be described as operating on a more general plane compared to markedness – in other words, becoming the unmarked

alternative is often part of a grammaticalisation process. However, both of these phenomena can be researched by observing changes in textual frequency.

In view of markedness and diachronic change, several linguists have suggested that language changes tend to go towards the unmarked alternative (e.g. Levin 2006, 326). Markedness becomes relevant as regards changes in subject-verb agreement with collective nouns in that the singular is the unmarked pole for nouns and verbs in English according to Battistella (1990, 4). If combined with Bauer's (1994, 65) results discussed in section 3.4.4 above, the principle of markedness seems to entail that the number preference of collective nouns would potentially on the whole be on the move towards the singular, which is the unmarked pole. Battistella (1990, 66) introduces five criteria used to determine whether the singular or the plural is the unmarked form in English:

- (1) The unmarked term is semantically indeterminate, which means that its meaning is broader and includes that of the marked term.
- (2) The unmarked category often serves as the prototype in an opposition.
- (3) The unmarked category often displays a greater freedom of distribution or use. This can be interpreted in three different ways: (1) as occurring in a greater number of contexts, or (2) as there being certain contexts in which the contrast between the marked and the unmarked term is neutralized in favour of the unmarked one, or (3) as corpus frequency, which is brought about by factors (1) and (2).
- (4) The unmarked member tends to exhibit a larger number of subcategorical distinctions. For instance, the number of case distinctions in a language is usually greater in the (unmarked) singular than in the (marked) plural.
- (5) The unmarked term is usually signalled by formal simplicity.

Levin (2006, 327) aptly points out that in view of verbs in English, (4) and (5) are not applicable, unlike the other three criteria. However, he also states in line with Battistella's (1990, 4) views that the remaining three criteria are enough to draw a conclusion that the singular is the unmarked alternative as regards verbs following collective nouns.

In view of point (3) above, Greenberg (1966, 65) connects the concept of markedness with frequency by suggesting that since corpus frequency usually correlates with the other criteria,

it could be used alone to measure markedness. However, using frequency as the sole indicator of markedness inevitably causes some problems. For example, if we follow Andersen's (2001, 17) discussion on *markedness shift*, using corpus frequency as the only measurement for markedness would entail that as soon as a collective noun is followed by a plural verb more often than by a singular verb, the plural becomes the unmarked alternative for the verbs following that particular collective noun, and therefore the markedness value could practically shift according to minor changes in usage. In addition, the same verbs would be given different markedness values depending on the context: *walks* would be unmarked after *a dog*, but marked after *crew*. Levin (2006, 326) notes that there are also cases where frequency does not correlate with markedness.

Thus, Battistella's (1990, 4) and Levin's (2006, 327) aforementioned views on the unmarkedness of singular verbs in English are adopted in the present study, and it is assumed that the markedness value does not change according to linguistic context. This view is corroborated by Biber et al. (1999, 189), who state that the singular is the predominant choice after collective nouns, and therefore, following Battistella's (1996, 10) definition of markedness discussed above, it is logically the unmarked pole. The latter part of the present study also aims to shed light on whether markedness significantly affects the changes in the number preference of collective nouns – a view indirectly supported for example by Bauer (1994, 65).

### **3.5.2 Lexical diffusion**

As with markedness discussed above, the concept of lexical diffusion has also been mostly applied in studies of phonology, but as Levin (2006, 329) points out, it can well be adapted to morphological and syntactical studies as well. Lexical diffusion is a certain type of language change, where individual items are affected one after another, until all the relevant items have

been transformed by the process (Phillips 1984, 320). Typically, the velocity of this change forms an S-curve: the first few items are affected relatively slowly; then, as more and more items become affected, the speed of the change increases and the steep part of the curve is created; and eventually, the rest of the items are affected slowly, often leaving some less relevant items unchanged (Levin 2006, 329). However, since language change is always a relatively slow process, it must be emphasised that even the rapid phase of the change could last for centuries in some cases.

Phillips (2001, 134) connects the concept of lexical diffusion with frequency by suggesting that while certain types of sound change affect high-frequency items first, low-frequency items are affected first by analogical change. Therefore, it would seem to be reasonable to assume that low-frequency collectives in the present study would serve as good indicators of possible future developments, since they do not have as entrenched agreement patterns as high-frequency collectives (Langacker 1987, 59; Levin 2006, 329).

Interestingly, Hopper and Traugott (2003, 155) state that the “analogical spread of one allomorph at the expense of others is aided by the sheer textual frequency of the successful allomorph”. In view of subject-verb agreement with collective nouns, this statement would seem to suggest that since singular agreement has become predominant with collective nouns, the proportion of singular verbs following collectives is expected to be on the increase. This prediction goes hand in hand with the principle of markedness discussed in section 3.5.1 above: both theories would seem to suggest that singular agreement would be on the increase with all collective nouns. The theories are also corroborated by Bauer’s (1994, 65) findings, but not by for instance Levin’s (2006, 321) results. These clashing results will be subjected to critical empirical examination in the following sections.

## 4 Corpus Analysis

Four restrictions were used in analysing the corpora, in order to – as much as possible – eliminate the conditions that are known to cause variation in the number of the verb following a collective noun (discussed in section 3.4 above).

Firstly, the regional effect (see section 3.4.1 above) was eliminated in restricting the present study to corpora consisting of British English only.

Secondly, to get rid of the effect that stylistic factors have on the number of the verb (see section 3.4.2 above), the spoken part of the BNC was excluded from the present study. This was also done in order to increase the correspondence between the corpora, since the CLMETEV consists of written texts only. In addition, the BNC searches were restricted to texts published between 1985 and 1993 (i.e. the most recent ones in the BNC) in order to make the possible variation in the results between past and present-day British English even more discernable. The large number of tokens in the BNC and the fact that nearly all of the texts (containing over 90 million words) included in the BNC are published between 1985 and 1993 enabled employing this restriction without a considerable risk of causing a bias in the results.

Thirdly, the effect of context on the results (see section 3.4.3 above) was minimised by excluding all the tokens influenced by some element in the context that made variation in number impossible (discussed in greater detail in section 4.1.2 below). However, the tokens were excluded only if both the grammars consulted in this study and the corpus data agreed on the effect of the contextual element.

Fourthly, the effect of the proximity principle (see section 3.3.3 above) was eliminated by taking into account only the tokens where the collective noun was directly followed by a verb in both corpora.

To gain a fairly complete picture of the ongoing and past linguistic changes in the area of subject-verb agreement with collective nouns, I decided to study the frequencies of nine different collective nouns. In order to avoid causing a bias in the results, I selected the collective nouns so that they are evenly divided into the following three groups: high-frequency, mid-frequency, and low-frequency collectives. Thus, each group includes three collective nouns. In addition, after completing a small pilot study, I was able to select the collectives so that each group includes one collective that mostly prefers singular agreement in present-day British English, one that has no clear preference, and one that mostly prefers plural agreement. Hence, there are also three collectives in the ‘mostly singular’, ‘variable’, and ‘mostly plural’ groups each.

Moreover, some limitations were applied to the collectives selected. Firstly, only the collective nouns that display at least some variation in singular/plural contrast were chosen, and secondly, only “specific” collectives representing human beings were selected (see sections 3.1 and 3.3.1 above for further information). The nine collective nouns selected for the present study and their characteristics are shown in the table below:

<b>collective</b>	<b>prefers sg</b>	<b>no pref</b>	<b>prefers pl</b>	<b>high-freq</b>	<b>mid-freq</b>	<b>low-freq</b>
<i>government</i>	x			x		
<i>army</i>	x				x	
<i>tribe</i>	x					x
<i>family</i>		x		x		
<i>crowd</i>		x			x	
<i>gang</i>		x				x
<i>staff</i>			x	x		
<i>crew</i>			x		x	
<i>clergy</i>			x			x

Table 6: The selected nine collective nouns divided into categories according to their characteristics



The selection of the preference for singular/plural as the first defining characteristic was motivated by Levin's (2006, 321) recent study, suggesting that those collective nouns that prefer plural apparently continue to move towards plural agreement, those that have no preference remain variable, and those collectives that are usually followed by a singular verb continue to move towards singular agreement. Levin's findings go against the results of some other studies, according to which the singular would be on the increase among all collective nouns (e.g. Bauer 1994, 65). This study aims to shed light on this apparent discrepancy.

The reason for selecting frequency as the other defining characteristic was that it has often been argued that low-frequency items are usually the first to be affected by analogical change (e.g. Hooper 1976). Therefore, the low-frequency collectives (and also possibly the mid-frequency collectives) may well give us a glimpse of possible future developments.

#### **4.1 Irrelevant and problematic tokens**

The raw data obtained by the corpus searches included plenty of irrelevant tokens that had to be manually removed. In addition, some of the tokens were considered problematic, because according to some grammarians they allow variation in the following verb, and according to others they do not. If instances of variation were encountered in the data analysis, they were considered relevant regarding the aims of the present study, and therefore justifiably included in the study. The following sections will shed some light on the principles of exclusion and inclusion in the present study by discussing the corpus findings and matching them against grammars.

### 4.1.1 Irrelevant tokens

All the irrelevant tokens were removed from the data. Firstly, there were occurrences where the referent of the verb is another noun or pronoun, or the collective noun is part of a larger referent (e.g. a co-ordinated subject). Therefore, there is logically no agreement between the collective noun and the following verb:

- [34] Perhaps fittingly, a *member* of the Joiner family *is* actively involved with the Trust, the formation of which was long overdue. (ANC 1279)
- [35] What *we* as a family *were* unanimous about was that he adopted no pose of being the distinguished writer and even gave the impression that he much preferred not to be treated as such. (H9X 718)
- [36] The small number of married women who pursued an active public life between the wars continued to assume that *home and family were* part of their natural responsibilities and solved the problem [. . .] (GUW 1397)

Secondly, the BNC is unfortunately all but free from tagging errors, which cause unnecessary distortion in results obtained by tag searches. Therefore, the verb following the collective noun sometimes turned out to be a bare infinitive (not a plural), which entailed that the collective noun is functioning as the object – or possibly subject – of the head verb, which could be found earlier in the sentence. Naturally, there is no agreement between the collective noun and the bare infinitive:

- [37] He's been telling them at the public house that you'll *help* his noble family *become* great again. (GW8 2108)
- [38] Only after the dissolution of the Malaysian Federation, *did* the Malaysian government *take* over control of monetary policy [. . .] (EEF 335)

Sometimes the verb following a collective was a past-tense verb incorrectly tagged as a plural present-tense verb. However, the tense is revealed by the context:

- [39] In September the Ivoirien *government cut* the price it paid cocoa farmers for the second time since July [. . .] (HL1 3163)

Tagging errors also manifested themselves in sentences where a collective noun is followed by a noun, an adverb, or a preposition that looks like a verb:

- [40] You can't have *family snaps* when there's just the two of you [. . .] (AE0 3228)

[41] For non-professional *staff further* training was likely to be envisaged as internal courses, or City and Guilds/BEC courses. (BNM 396)

[42] What are the *staff like* – how many are there? (GXJ 3687)

It was also because of tagging errors that there were some genitives among the results. These look similar to the contracted form of *is* or *has* on the surface and are therefore among the search results:

[43] The *family's* home in Stockton appeared deserted with all the curtains drawn and no-one was answering the door. (K2Y 43)

[44] A few minutes spent now completing your Enrolment Form could make a world of difference to the rest of your life and your *family's*, if one day a serious accident happens to you. (AYR 162)

Yet another result of tagging errors removed from the data was the manifestation of tokens followed by a verb in subjunctive mood:

[45] Crowds in Ramtha demanded that the *government ban* the passage of food to Saudi Arabia in retaliation for a Saudi ban on trade with Jordan. (HL4 2325)

[46] The report *recommended* that all British *staff be* put on temporary contracts. (K1G 3466)

Thirdly, tokens where the verb directly following the collective noun belongs to a different clause were excluded from the present study:

[47] If you want away from the *crowd try* the upstairs restaurant. (ECS 1061)

Fourthly, some of the collective nouns chosen for this study have more than one meaning, and oftentimes this additional meaning bears no notion of collectivity. Hence, tokens such as the following were removed from the data (*staff* here signifying *a wand*):

[48] On the roll of a 1 or 2 the *staff has* run out of energy and will not work for the remainder of the battle. (CN1 2340)

In addition to the fairly straightforward cases mentioned above, the raw data contained plenty of tokens that were problematic. As Depraetere (2003, 96) points out, collective nouns may be preceded by determiners or quantifiers that require a certain kind of verb (for example a singular one), and therefore the collective noun that is situated between the two will automatically be interpreted as singular as well. However, my aim was to include as much

tokens as possible, and therefore I kept all the tokens displaying variation between the singular and the plural in my data, unlike Depraetere in her study. These problematic tokens included in the study are discussed in section 4.1.2 below.

There were some patterns initiated by certain determiners and quantifiers in the data that cause the number of the following verb to remain invariable, and they were accordingly removed. First of all, the tokens preceded by *much*, *much of*, or *every* were disregarded:

[49] The UN knows that *much of* the Cambodian *army has* not been paid for some time [. . .] (CR8 759)

[50] *Every crowd is* made up of individuals. (C8K 222)

The exclusion of these tokens is further corroborated by Huddleston (1988, 88), who lists *much* and *every* as determiners requiring a singular head (see section 3.4.3 above).

Secondly, the tokens preceded by *many of*, *several of*, *few*, or *both* were removed from the data:

[51] Yet *many of* the *crowd were* kept interested by the possibility of victory, which is, I suppose, what derbies are about. (A1N 351)

[52] *Several of* the office *staff were* staring out of their windows at her [. . .] (JXY 747)

[53] It seems to me that very *few staff are* addressing themselves to the kinds of things [. . .] (GUR 1436)

[54] If *both crew are* slain the bolt thrower can not shoot. (CM1 1711)

According to Huddleston (1988, 88), these determiners require a plural head, and the data from the BNC seems to confirm that (see section 3.4.3 above).

Thirdly, the tokens preceded by cardinal numerals (with or without and *of*-phrase) were disregarded:

[55] *2 crew were* killed and many locals felt they'd had a lucky escape. (K1G 2008)

[56] *Three of* the *staff have* been killed since April 1992. (K5D 6995)

Again, cardinal numerals are listed among the determiners requiring a plural verb by Huddleston (1988, 88; see also section 3.4.3 above), and the corpus data corroborate the

claim. Surprisingly enough, a couple of tokens with a pattern *cardinal numeral + collective noun + singular verb* were found in the data as well:

[57] *Ten crew is* about the right to race the boat, though more weight in a breeze - if allowed by the rules - is an advantage. (G37 2040)

[58] A sellout *40,000 crowd was* expected to welcome the Wigan ace for tomorrow's home game against Balmain. (HJ4 823)

However, the fact that these two examples have a singular verb does not prove that there is real variation within this particular pattern. In [57] above, it can be argued that *ten crew* is seen as a singular unit or idea, and therefore the verb must also be in the singular. The quantity, which is a semantically singular concept, is in focus here: it is not the *crew* who are *about the right to race the boat*, it is the *number* of the crew. Therefore, [*t*]*en* does not define *crew* here – it is *crew* that specifies [*t*]*en*. Huddleston (1988, 88) gives a similar example with a plural noun: “Eight cups isn’t enough.” As to example [58], *40,000* could be interpreted as a kind of an adjective in a similar fashion as *sellout*, because the noun phrase is introduced by the indefinite article *a*.

#### 4.1.2 Problematic tokens considered relevant

First of all, despite the rather strict views of some grammarians (discussed in section 3.4.3 above), an instance of a collective noun preceded by *a* and followed by a plural verb was found in the data:

[59] But this is the house which *a young family have* been invited to make their home this Christmas. (K27 2143)

Hence, the tokens preceded by an indefinite article were considered relevant and therefore included in the study.

The tokens preceded by *the whole* or *all/all of* were also included in the study, because there is clear variation in the number of the following verb in the corpus data. Consider the following examples:

- [60] *The whole family is* over the moon with his improvements. (CBF 7825)
- [61] *The whole family are* coming, it's all been arranged for ages - and I'm not there! (H8F 3932)
- [62] *All the local army was* ranged up on either side [. . .] (H0A 557)
- [63] *All staff are* expected to help entertain key clients, socialise constructively with suppliers and trading partners, and carry out any on the spot research projects deemed necessary by our department heads. (HGN 197)

In view of [61] above, it is rather surprising that Poutsma (1904, 283) associates *the whole* with the singular, and Jespersen (1961, 99) even calls it “an indication of the sg construction” (see also section 3.4.3 above). Jespersen admits that it is not impossible for a phrase initiated by *the whole* to take a plural verb and gives even a few examples, but according to the BNC data, it seems to be followed by a plural verb far more often than one would be led to think. However, the fact that the more recent grammars do not seem to take a very radical stance on this matter forces one to consider whether there has been a change in subject-verb agreement as regards *the whole*. With respect to *all/all of*, the grammarians' opinions seem to vary greatly, as pointed out in section 3.4.3 above. The obvious discrepancy between these views is corroborated by the BNC data (see [62] and [63] above), and therefore the tokens preceded by *all/all of* were also included in the present study.

The tokens preceded by *some/some of*, *many*, or *each* also display some variation in number in the BNC data:

- [64] *Some of the army was* camped outside the walls, but only a few sentries were stirring. (FSE 1736)
- [65] Fear is one of the Dark Lord's most powerful weapons yet *some of his army are* defeated by it. (KAY 462)
- [66] *Many a Dark Elf army has* been lost in Tiranoc [. . .] (CM1 286)
- [67] *Many staff have* worked well beyond the call of duty [. . .] (GXJ 1432)
- [68] Trained to conform to the million tiny rules that *each family makes* for itself. (AC4 2633)

Since the grammarians' unanimous views on these three quantifiers (see section 3.4.3 above for a more detailed discussion) are mirrored in the corpus results, the tokens were included in the study.

The tokens preceded by *any/any of*, *none of*, or *no* are also followed by both singular and plural verbs in the data, although for some of the examples one has to turn to the CLMETEV:

- [69] Things to be provided, when *any great family is* going into the Country, for a Summer. (Bradley (1732): The Country Housewife and Lady's Director)
- [70] If *any crew are* slain the cannon's speed is reduced proportionally. (CN1 878)
- [71] *None of the crew was* injured but the fire engine was scratched and dented. (K97 16723)
- [72] *None of my family have* ever attempted to run this house with women only. (Ward (1894): Marcella (Vol. 1))
- [73] [. . .] *no additional staff is* required. (HTK 859)
- [74] [. . .] one was cancelled because *no crew were* available due to a clerical error in the rostering. (HHW 14620)

As already pointed out in section 3.4.3 above, also the grammarians agree on the number variation with these determiners. Together with the evidence from the corpus data, including the tokens preceded by *any/any of*, *none of*, or *no* was well grounded.

The number-transparent expressions discussed in section 3.4.3 above were logically included in the study, since they do not affect the number of the complement (i.e. the collective noun). Interestingly, some of these expressions display variation in number in the corpora, while others do not. However, this must be attributed to the fact that there are only a few tokens of each pattern in the data:

- [75] *The rest of the army is* just softies compared to us. (CMC 612)
- [76] While the *rest of the tribe were* sleeping, She-Who-Is-Alone walked quietly to the hill where the Great Spirits had spoken and offered her beloved doll as a sacrifice. (G25 421)
- [77] In the decade 1420-30, a sizeable *proportion of* Charles VII's *army was* composed of Scots who fought at both Cravant and Verneuil [. . .] (EDF 1267)
- [78] On the revenue side it meant such devices as changing the weighting of the staff mix so that a greater *proportion of* unqualified *staff were* employed than might ideally have been desired. (CS7 1317)

- [79] *A number of the crew are* Belgian and we therefore have a Belgian chef on board so that we can offer traditional Belgian fare said Mike. (K9N 82)
- [80] And the *remainder of the family are* on this side. (Hardy (1873): A Pair of Blue Eyes)

There are also expressions in the corpus data that are clearly related to the aforementioned ones but not mentioned by Huddleston, Carter and McCarthy, or any other consulted grammarian. They include *fraction of*, *part of*, *percentage of*, *half/half of*, and various proportions such as *four-fifths of* and *12 percent of*. Tokens preceded by these expressions were also included in the study. Consider the examples below:

- [81] What *fraction of* the Smith *family are* males. (FEH 656)
- [82] *Part of* each functional *staff was* dedicated to the new businesses [. . .] (FA8 541)
- [83] [. . .] and a high *percentage of* the female *staff were* part-time workers. (B2S 619)
- [84] *Half* your *army wants* to hang back and shoot, the other half wants to get stuck in as quickly as possible. (CM1 1619)
- [85] ‘Over *half* my *crew are* dead’, Finnan said. (GWF 2932)
- [86] But the Turks for their part point out, when you visit them, that something like *four-fifths of* the undiminished Greek *army is* concentrated in Thrace [. . .] (CRB 3740)
- [87] The results revealed that *12 percent of* *staff think* Newsline content is very interesting and 71 percent fairly interesting. (GX9 295)

Even though not displaying any variation in the corpora, the tokens preceded by *more* and *most/most of* were also included in the study, because Biber et al. (1999, 184) juxtapose them with *many* in that they prefer plural agreement but can also be found with singular agreement at times (see section 3.4.3 above). The non-existence of the variation in the data must again be due to the small number of tokens in the pattern *more/most/most of* + *collective noun* + *verb*. The tokens preceded by *most/most of* adhere to Biber et al.’s aforementioned juxtaposition, but the two tokens preceded by *more* are followed by a singular verb:

- [88] *Most of* the *crew were* killed and we came back very short handed. (J1X 191)
- [89] *Most* *staff are* untrained and there is as yet little evidence that training makes much difference to the quality of care. (CGD 1680)



[90] More funds were required generally for longer opening hours; *more paid staff was* wanted as was more training. (B0N 766)

[91] [. . .] and as our programme production increased *more staff was* added, including the versatile and ever helpful announcer Dick Halhed [. . .] (B11 838)

Finally, the tokens coordinated by *or/nor* in the data display variation in number. Biber et al. (1999, 183) confirm this finding by noting that the number of the following verb varies in constructions coordinated by *or/nor*, especially if one of the coordinated noun phrases bears a notion of plurality (see also section 3.4.3 above). Therefore, the tokens where a collective noun followed *or* or *nor* were considered relevant (see also section 3.4.3 above):

[92] Moreover, there may well be reluctance or bad feeling when the operative *or gang is* obliged to undertake work without bonus. (FYS 418)

[93] However, as many overseas assignments fail because the wife *or family are* unable to cope with the new way of life and culture [. . .] (CHS 1365)

The problematic tokens that were considered relevant and thereby included in the study are summarised in the following table:

<b>prefers sg</b>	<b>no preference</b>	<b>prefers pl</b>	<b>number-transparent</b>
<i>a/an</i>	<i>the whole</i>	<i>some/some of</i>	<i>number of</i>
<i>each</i>	<i>all/all of</i>	<i>many</i>	<i>part of</i>
<i>more</i>	<i>no</i>	<i>none of</i>	<i>rest of</i>
	<i>or/nor</i>	<i>any/any of</i>	<i>remainder of</i>
		<i>most/most of</i>	<i>proportion of</i>
			<i>fraction of</i>
			<i>percentage of</i>
			<i>half/half of</i>
			<i>four-fifths of</i>
			<i>12 percent of</i>

Table 7: The problematic tokens considered relevant in categories

## 4.2 Results from the CLMETEV

For the purposes of this study, the token frequencies in the CLMETEV will be discussed jointly for the whole corpus. Although it would be really interesting to see the developments in subject-verb agreement with collective nouns between the different subparts of the CLMETEV (1710-1780, 1780-1850, and 1850-1920; see also section 2.1 above), the number of relevant tokens appeared to be too small to yield any statistically significant results separately, something which is obviously caused by the relatively small size of the subparts (3.0, 5.7, and 6.3 million words, respectively). However, when the frequencies are put together, we have a fairly considerable number of instances that can well be compared with the BNC results.

Searches were made using the collective noun as the search term. Since the CLMETEV is not a tagged corpus (for further information, see section 2.1 above) and in order to avoid causing any bias, the results were analysed manually using the *AntConc* freeware program. All the present tense verbs and past tense forms of *be* directly following the nine collective nouns defined in section 4 above were taken into account and the irrelevant tokens were removed (see section 4.1.1 above). The frequencies of the relevant tokens in the whole *CLMETEV* are displayed in the following table:

<b>collective noun</b>	<b>singular verbs</b>	<b>plural verbs</b>	<b>sg %</b>
<i>government</i>	154	21	88.0%
<i>family</i>	91	99	47.9%
<i>staff</i>	0	2	0.0%
<i>army</i>	122	13	90.4%
<i>crowd</i>	34	8	81.0%
<i>crew</i>	9	19	32.1%
<i>tribe</i>	6	6	50.0%
<i>gang</i>	3	3	50.0%
<i>clergy</i>	3	24	11.1%

Table 8: The frequencies of the collective nouns followed by singular or plural verbs in the CLMETEV (1710-1920)

The ‘sg %’ (singular percentage) column in table 8 above represents the percentage of tokens followed by a singular verb. The plural percentage for each word can easily be calculated by subtracting the singular percentage from 100%: the plural percentage for *government* is 12.0%, etc.

For the sake of clarity, the detailed analysis of the CLMETEV results is divided into subsections according to the division based on the number preference of the collectives (explained in detail in section 4 above; see also table 6).

#### **4.2.1 The ‘mostly singular’ group in the CLMETEV**

The frequencies of the tokens grouped in the ‘mostly singular’ category followed by singular or plural verbs are presented in the table below. The abbreviations ‘pr sg’ and ‘pr pl’ stand for ‘present singular’ and ‘present plural’, respectively:

<b>collective</b>	pr sg	was	<b>sg total</b>	pr pl	were	<b>pl total</b>	tokens total	<b>sg %</b>
<i>government</i>	125	29	<b>154</b>	9	12	<b>21</b>	175	<b>88.0%</b>
<i>army</i>	59	63	<b>122</b>	4	9	<b>13</b>	135	<b>90.4%</b>
<i>tribe</i>	4	2	<b>6</b>	2	4	<b>6</b>	12	<b>50.0%</b>
<b>total</b>	188	94	<b>282</b>	15	25	<b>40</b>	322	<b>87.6%</b>

Table 9: The frequencies of the ‘mostly singular’ collective nouns followed by singular or plural verbs in the CLMETEV (1710-1920)

As noted in section 4 above, the design of the ‘mostly singular’, ‘variable’, and ‘mostly plural’ categories was based on the pilot study on present-day British English. Table 9 above shows that the preference for the singular or the plural seems to be a relatively steady characteristic of collective nouns, at least for those preferring singular agreement – *government* and *army* are most often followed by a singular verb in the CLMETEV, and with *tribe* the plural is equally possible:

- [94] It is well known that *government is* supported by opinion [. . .] (Johnson (1740-1): Parliamentary Debates (Vol. 1))
- [95] The French *army seems* on the verge of universal mutiny. (Carlyle (1837): The French Revolution)
- [96] A savage *tribe resembles* a herd of gregarious beasts [. . .] (Bagehot (1869): Physics and Politics)

With the present results in mind, it is rather surprising that Poutsma (1904, 284) lists *government* among the collective nouns that are usually followed by a plural verb. This peculiarity is most likely due to the fact that linguistic research was largely based on anecdotal observation in the beginning of the 20<sup>th</sup> century, and the scope and objectivity of observation as a research method can be questioned (see also section 2 above).

In addition, the three ‘mostly singular’ collective nouns are also followed by plural verbs in the corpus. With *tribe*, a plural verb seems to be even as common as a singular verb, as already pointed out. This exception is perhaps caused by the fact that *tribe* is the low-frequency item in the ‘mostly singular’ group, and the number of relevant tokens in the corpus

is therefore rather small (N=12). Moreover, some of the tokens are also followed by a plural verb even when the focus is quite clearly on the entity (c.f. section 3.4.4 above):

[97] By the discipline of their armies the *government were* triumphant. (Churchill (1899): The River War)

[98] [. . .] but the king of Sweden's whole *army were* come up with him. (Haywood (1744): The Fortunate Foundlings)

[99] [. . .] his *tribe were* a blot upon the face of the earth [. . .] (Churchill (1899): The River War)

It is important to keep in mind that although *their* in [97] above refers to a subject in the previous sentence, a plural pronoun may sometimes refer to a collective noun in the same sentence and thereby influence the number of the verb following the collective, as already pointed out in section 3.4.3 above:

[100] The *army were* now dependent for *their* existence on the partly finished railway [. . .] (Churchill (1899): The River War)

However, it should also be noted that this is not always the case. The presence of a plural pronoun referring to a collective noun in the same sentence does not necessarily entail that the verb that follows is in the plural, as the following example with *crowd* shows:

[101] The *crowd was* dense in the street, and *their* progress was slow [. . .] (Brebner (1910): The Brown Mask)

Interestingly, the number of tokens followed by *were* is greater than the number of those followed by all present tense plural verbs together (*are, have, do, say, etc.*) with all three collectives. This could be caused by the fact that unlike the BNC, the CLMETEV consists of narrative texts only, and the past tense is most characteristically used in narrative style. On the other hand, *government* is followed by present tense singular verbs (*is, has, does, says, etc.*) much more often than by *was*, whereas *army* prefers *was* to present tense verbs. *Tribe* also favours the present tense, although the small number of tokens sheds some uncertainty on the results concerning that particular collective. One possible but not fully satisfactory explanation for the fact that *were* is encountered so much more often than the present tense

plural verbs could be the *were*-subjunctive (Leech and Svartvik 2002, 147) that manifests itself in *if*-clauses (for further discussion, see section 5.1.1 below):

[102] [. . .] *if a good government were* thus placed, I should think it decidedly better that the agreements of the administration with foreign powers should be submitted to Parliament. (Bagehot (1867): The English Constitution)

#### 4.2.2 The ‘variable’ group in the CLMETEV

The frequencies of the ‘variable’ collective nouns followed by a singular or plural verb can be seen in the table below:

<b>collective</b>	pr sg	<i>was</i>	<b>sg total</b>	pr pl	<i>were</i>	<b>pl total</b>	tokens total	<b>sg %</b>
<i>family</i>	58	33	<b>91</b>	46	53	<b>99</b>	190	<b>47.9%</b>
<i>crowd</i>	11	23	<b>34</b>	1	7	<b>8</b>	42	<b>81.0%</b>
<i>gang</i>	2	1	<b>3</b>	2	1	<b>3</b>	6	<b>50.0%</b>
<b>total</b>	71	57	<b>128</b>	49	61	<b>110</b>	238	<b>53.8%</b>

Table 10: The frequencies of the ‘variable’ collective nouns followed by singular or plural verbs in the CLMETEV (1710-1920)

Table 10 above shows how evenly the tokens with singular and plural agreement are divided for *family* and *gang*. On the other hand, the mid-frequency collective *crowd* seems to favour singular agreement quite clearly. Consider the following examples from the CLMETEV:

[103] And when the *family goes* away, there are only about three servants to preach to when I get there. (Hardy (1873): A Pair of Blue Eyes)

[104] As a man’s *family go* on muttering in his maturity incorrect phrases derived from a just observation of his early youth [. . .] (Bagehot (1867): The English Constitution)

[105] [. . .] and then the *crowd laughs* vehemently, and invents nicknames for them on the spur of the moment [. . .] (Hughes (1857): Tom Brown’s School Days)

[106] The *crowd cheer*, and begin to chaff Joe, who turns up his nose and swaggers across to the sticks. (Hughes (1857): Tom Brown’s School Days)

[107] You must continue to act under his Direction, for the moment we break loose from him, our *Gang* is ruin'd. (Gay (1728): The Beggar's Opera)

[108] I should think Westall is right, and that the *gang* have got hold of him. (Ward (1894): Marcella (Vol. 1))

Interestingly, the oldest (1710-1780) tokens with *family* prefer quite decidedly singular agreement. This could be due to the fact that many of the oldest tokens were used to refer to *family* in the wide sense, including grandparents, uncles, aunts, cousins, etc. Hence, *family* was often seen more as a large, impersonal unit, and the singular was used, because the individuals are possibly harder to conceptualise in a large group (Poutsma 1904, 283). Poutsma even goes as far as claiming that the size of the referred body determines the number of the verb, at least to some extent (ibid., 283-284): the singular is used with a large body of persons (e.g. *army*, *community*), and the plural with a small body of persons (e.g. *family*, *government*). Levin (2006, 334) confirms that the size of the group does sometimes affect the selection between the singular and the plural: he found out that the probability of singular agreement increases along with the size of the group with the collectives *duo*, *trio*, *quartet* and *quintet*. However, Depraetere (2003, 114) says that Poutsma's claim needs to be refuted, or at least weakened. According to her study, the size of the group does not decide the number of the verb as straightforwardly as Poutsma presents. Depraetere bases her claim on the fact that most collectives prefer singular agreement "irrespective of the number of the individuals they are composed of", and gives plenty of examples: for instance, *people* admittedly refers to a larger group than *organisation*, yet the former is a 'mostly plural' collective and the latter a 'mostly singular' one in present-day British English.

As with the 'mostly singular' group (see section 4.2.1 above), the 'variable' collectives seem also to be followed more often by *were* than by all present tense plural verbs in the CLMETEV. In addition, again in line with the 'mostly singular' group, there is more variation with singular tokens: *family* prefers present tense singular verbs to *was*, whereas *crowd*

prefers *was* to present tense verbs. With *gang*, there seems to be no clear preference, but the number of relevant tokens is too small to draw any valid conclusions.

### 4.2.3 The ‘mostly plural’ group in the CLMETEV

The following table shows the frequencies of the ‘mostly plural’ collectives followed by singular or plural verbs in the CLMETEV:

<b>collective</b>	pr sg	<i>was</i>	<b>sg total</b>	pr pl	<i>were</i>	<b>pl total</b>	tokens total	<b>sg %</b>
<i>staff</i>	0	0	<b>0</b>	2	0	<b>2</b>	2	<b>0.0%</b>
<i>crew</i>	4	5	<b>9</b>	3	16	<b>19</b>	28	<b>32.1%</b>
<i>clergy</i>	3	0	<b>3</b>	14	10	<b>24</b>	27	<b>11.1%</b>
<b>total</b>	7	5	<b>12</b>	19	26	<b>45</b>	57	<b>21.1%</b>

Table 11: The frequencies of the ‘mostly plural’ collective nouns followed by singular or plural verbs in the CLMETEV (1710-1920)

All three collectives in the ‘mostly plural’ category select plural verbs more often than singular verbs in the CLMETEV:

[109] His *Staff* go different ways. (Carlyle (1837): The French Revolution)

[110] [. . .] and upon his refusing to return other way, all the *Crew* were by Arm’d force taken out of the Boat [. . .] (Cook (1768-71): Captain Cook’s Journal)

[111] The *Clergy* have means and material: means, of number, organization, social weight; a material, at lowest, of public ignorance, known to be the mother of devotion. (Carlyle (1837): The French Revolution)

Surprisingly, the high-frequency collective *staff* has only two relevant instances where it is followed by a singular or a plural verb in the whole CLMETEV, besides a couple of irrelevant tokens signifying *stick*, *rod*, or *wand*:

[112] Each *staff* was borne by a knight, of whom sixteen were in attendance to relieve one another when fatigued. (Ainsworth (1843): Windsor Castle)

[113] The flag *staff* was naked; the royal standard that had flapped in the wind last night was gone. (Hope (1898): Rupert of Hentzau)



The reason behind the very small number of tokens can be found in the *Oxford English Dictionary (OED)*: the first citation of *staff* in the sense of “a body of persons employed, under the direction of a manager or chief, in the work of an establishment or the execution of some undertaking (e.g. a newspaper, hospital, government survey, school, etc.)” is from 1837, in fact from the same Carlyle’s work as example [109] above. Thus, only the latest texts in the CLMETEV could possibly include any relevant tokens, and presumably the sense has been rather rare at first, slowly making its way to a fairly frequent word in the English language that it is nowadays.

Moreover, it should be noted that the *OED* also has a separate entry for the word *staff* when used in military contexts, i.e. “a body of officers appointed to assist a general, or other commanding officer, in the control of an army, brigade, regiment, etc., or in performing special duties (as the *medical staff*)”. The instances of this sense were also included in the present study, as both of the senses denote a group of people working for someone or something, and are therefore considered collective nouns. However, the military sense is rather rare and only a couple of clear instances of it were encountered in the data, one of them being example [109] above. According to the *OED*, the earliest instance of this sense dates back to 1700.

Although there are very few instances of *staff* followed by a singular or a plural verb in the CLMETEV, it is still justifiable to have *staff* as part of the present study, because it is one of the most frequent, if not the most frequent plural-preferring collective noun in present-day British English (further discussed in section 4.3.3 below). It is also mentioned by a number of grammarians as a typical example of a collective noun with a plural preference (e.g. Biber et al. 1999, 188).

Turning to the other two collectives in the ‘mostly plural’ group, *crew* and *clergy* have also some instances where they are followed by a singular verb, even though they clearly prefer

the plural construction – and again these tokens include instances where the individuals are in the focus, or at least it is rather difficult to tell whether the sense conveyed by the choice of number indicates the group as a whole or the individuals constituting it (c.f. sections 3.4.3 and 4.2.1 above):

[114] The ship's *crew* was beat to quarters, and within one minute a boat was to the rescue. (Bacon (1902): The Dominion of the Air)

[115] Enough, and *Clergy* has strength, the *Clergy* has craft and indignation. (Carlyle (1837): The French Revolution)

Again, as with the 'mostly singular' and 'variable' groups (see sections 4.2.1 and 4.2.2 above), the 'mostly plural' collective nouns are followed by *were* relatively often in the corpus: *crew* is followed by *were* 16 times, whereas it precedes present tense plural verbs only 3 times. In the singular, the figures are much more evenly divided: *crew* selects *was* 5 times and present tense singular verbs 4 times. With *clergy*, the distribution is more even, although *were* is still selected fairly often.

### 4.3 Results from the BNC

The *Sketch Engine* interface was used in the analysis of the BNC data because it has the CQL search function, with which it is possible to search for collective nouns followed by all present tense verbs, singular and plural separately. I ran distinct CQL searches for each collective noun, first for the tokens followed by a singular verb and then for those followed by a plural verb. The search strings were as follows: [word="xxx"] [tag="VBZ|VHZ|VDZ|VVZ"] for receiving the tokens followed by singular verbs, and [word="xxx"] [tag="VBB|VHB|VDB|VVB"] for receiving the tokens followed by plural verbs (*xxx* signifying the collective noun). However, it should be noted that the past-tense verbs *was* and *were* were not included in these searches. Therefore, I ran separate searches for

collectives followed by *was* and *were* in order to complete the results. The search strings for the past tense verbs were: [word="xxx"] [word="was"] and [word="xxx"] [word="were"].

As with the CLMETEV, I also analysed the BNC results manually and removed the irrelevant tokens (for a detailed description, see section 3.1.1 above). If the total number of tokens exceeded 300 in the present tense searches, I took a sample of 300, analysed it, and calculated the total number of relevant instances by using that sample. For instance, the search for *family* followed by a present tense singular verb returned 948 tokens. I took a random sample of 300 tokens and analysed them, after which the number of relevant tokens was reduced to 197. Then I took into account the total number of tokens in the BNC by multiplying the number of the relevant tokens in the sample (N=197) by the number of all tokens in the BNC (N=948), and dividing the sum by the total number of the sample (N=300). Thus, the result was 623, which represents the total number of relevant tokens in the BNC. Correspondingly, if the total number of tokens was more than 100 in the searches for *was* and *were*, I took a sample of 100 tokens. The frequencies of the relevant tokens in the written part of the BNC can be seen in the following table:

<b>collective noun</b>	<b>singular verbs</b>	<b>plural verbs</b>	<b>sg %</b>
<i>government</i>	3,370	64	98.1%
<i>family</i>	885	583	60.3%
<i>staff</i>	73	1,306	5.3%
<i>army</i>	313	21	93.7%
<i>crowd</i>	131	51	72.0%
<i>crew</i>	78	201	28.0%
<i>tribe</i>	26	3	89.7%
<i>gang</i>	26	19	57.8%
<i>clergy</i>	1	82	1.2%

Table 12: The frequencies of the collective nouns followed by singular or plural verbs in the written part of the BNC (1985-1993)

As with the CLMETEV tokens in section 4.2 above, the detailed discussion of the results from the BNC is also divided into subsections according to the preference groups (see section 4 and table 6).

### 4.3.1 The ‘mostly singular’ group in the BNC

The following table presents the frequencies of the ‘mostly singular’ collective nouns followed by singular or plural verbs in the BNC:

<b>collective</b>	<b>pr sg</b>	<b>was</b>	<b>sg total</b>	<b>pr pl</b>	<b>were</b>	<b>pl total</b>	<b>tokens total</b>	<b>sg %</b>
<i>government</i>	2,609	761	<b>3,370</b>	45	19	<b>64</b>	3,434	<b>98.1</b>
<i>army</i>	191	122	<b>313</b>	13	8	<b>21</b>	334	<b>93.7</b>
<i>tribe</i>	17	9	<b>26</b>	2	1	<b>3</b>	29	<b>89.7</b>
<b>total</b>	2,817	892	<b>3,709</b>	60	28	<b>88</b>	3,797	<b>97.7</b>

Table 13: The frequencies of the ‘mostly singular’ collective nouns followed by singular or plural verbs in the written part of the BNC (1985-1993)

Some clear tendencies can be seen in table 13 above. As predicted by my aforementioned pilot study, *government*, *army* and *tribe* strongly prefer the singular in the BNC – in fact, over 90 percent of the relevant tokens are followed by a singular verb, for the high-frequency *government*’s part even over 98 percent:

[116] The US *government* *is* considering selling off all the satellites to a private company. (B7M 376)

[117] By now the Emperor’s *army* *was* approaching from the south, led by one of his most able generals, Marshal Otto Blucher. (CMC 237)

[118] The Red Cloud Goblin *tribe* *attacks* Karak Azgul and are repulsed. (CMC 453)

Having said that, the three collective nouns are also sometimes followed by plural verbs – even in contexts where the focus quite clearly seems to be on the collective entity, not on the individuals (see also sections 3.4.3, 4.2.1, and 4.2.3 above):

[119] Why the *government try* to stop us? (FEP 580)

[120] Recently the *army have* purchased new, heavier US flak jackets [. . .]  
(K34 420)

Finally, it is noteworthy that 35 out of all 191 instances of *army* followed by present tense singular verbs (i.e. over 18% of them) are from one and the same book (Priestley's *Warhammer Armies: Orcs & Goblins*) and they are quite repetitive – something that probably slightly affects the results in favour of the singular constructions:

[121] If your *army includes* up to five Trolls these form a single unit. (CMC 2197)

[122] If your *army includes* six to ten Trolls these may be organised into either one unit or into two as near as possible equally sized units. (CMC 2199)

[123] If your *army includes* at least one Mob of Orcs (including Big'uns) it may also include any number of Orc boar chariots. (CMC 2237)

### 4.3.2 The 'variable' group in the BNC

The frequencies of the 'variable' collectives followed by singular or plural verbs in the BNC are shown in table 14 below:

<b>collective</b>	pr sg	<i>was</i>	<b>sg total</b>	pr pl	<i>were</i>	<b>pl total</b>	tokens total	<b>sg %</b>
<i>family</i>	609	276	<b>885</b>	343	240	<b>583</b>	1,468	<b>60.3%</b>
<i>crowd</i>	56	75	<b>131</b>	24	27	<b>51</b>	182	<b>72.0%</b>
<i>gang</i>	19	7	<b>26</b>	10	9	<b>19</b>	45	<b>57.8%</b>
<b>total</b>	686	358	<b>1,044</b>	377	276	<b>653</b>	1,697	<b>61.5%</b>

Table 14: The frequencies of the 'variable' collective nouns followed by singular or plural verbs in the written part of the BNC (1985-1993)

As expected, *family*, *crowd* and *gang* are quite variable – they occur relatively commonly with both singular and plural verbs:

[124] They say her *family is* being held hostage to her good behaviour. (A0D 1682)

[125] The whole *family are* coming, it's all been arranged for ages – and I'm not there! (H8F 3932)

[126] He then cleverly pauses, to see what the *crowd* *thinks* of his speech so far. (KA1 1897)

[127] The *crowd* *go* delirious, the band respond with the entire concept. (HWX 1436)

[128] An armed *gang* *has* stolen jewels worth more than five million pounds from London's Hatton Garden. (K1U 313)

[129] It's believed the same *gang* *have* raided at least three homes [. . .] (K1B 2038)

However, there are also some differences in the number preference of the 'variable' group: *crowd* appears to select singular verbs somewhat more often than *family* and *gang*.

### 4.3.3 The 'mostly plural' group in the BNC

Table 15 below displays the frequencies of the 'mostly plural' collective nouns followed by singular or plural verbs in the BNC:

<b>collective</b>	pr sg	was	<b>sg total</b>	pr pl	were	<b>pl total</b>	tokens total	<b>sg %</b>
<i>staff</i>	47	26	<b>73</b>	1,045	261	<b>1,306</b>	1,379	<b>5.3%</b>
<i>crew</i>	46	32	<b>78</b>	106	95	<b>201</b>	279	<b>28.0%</b>
<i>clergy</i>	1	0	<b>1</b>	34	48	<b>82</b>	83	<b>1.2%</b>
<b>total</b>	94	58	<b>152</b>	1,185	404	<b>1,589</b>	1,741	<b>8.7%</b>

Table 15: The frequencies of the 'mostly plural' collective nouns followed by singular or plural verbs in the written part of the BNC (1985-1993)

Again in line with my small pilot study, *staff*, *crew* and *clergy* are most often followed by a plural verb in the BNC:

[130] The *staff* *come* from varied intellectual backgrounds [. . .] (HTD 2074)

[131] The train *crew* *were* already collecting in the locker room [. . .] (BP9 81)

[132] I think the *clergy* *wonder* what to do with him at times. (H8B 796)

As was the case with the 'mostly singular' group, the agreement patterns of the plural-prefering collectives seem to vary as well – there are also tokens that select singular

agreement in the BNC, even when the focus is clearly on the individuals, not on the entity (see also sections 3.4.3, 4.2.1, 4.2.3 and 4.3.1 above):

[133] Their *staff is* busy compiling lists of works of art [ . . . ] (EBS 425)

[134] Seals around the hatches provide a physical barrier, and air for breathing is filtered, so that the *crew does* not have to wear protective clothing. (B7C 173)

[135] Exactly what a celibate *clergy needs*. (AJW 32)

It is true that some grammarians (e.g. Quirk et al. 1985, 756) consider *clergy* an unmarked plural noun together with *people*, *police*, etc. (see also section 3.2.2 below). This may be caused by the fact Sweet (1898, 46) points out in his old grammar: there exists “distinct individual nouns by the side of collective ones” for these words (i.e. *person* for *people*, *policeman* for *police*, and *clergyman* for *clergy*). Although this classification might be plausible in some cases, the BNC results for *clergy* also include some singular instances, such as [135] above – not to mention the fact that over 10% of the relevant tokens of *clergy* are followed by singular verbs in the CLMETEV (see section 4.2.3 above). In addition, it also has the plural form *clergies*, formed using the plural marker *-s*, and it is fairly common. Therefore, it is justifiable to consider *clergy* a collective noun.

## 5 Comparison of the Results

The results from the CLMETEV and the BNC discussed in the previous sections revealed several interesting developments. In this section, these results will be compared side by side. Again, the analysis is divided into subsections according to the preference groups defined in section 4 above. In addition, the results for the present and past tense forms of the verb *be* are discussed separately for each collective in order to investigate whether the tense of the following verb bears any correlation with the number preference of the collective noun. Since the CLMETEV consists of narrative texts which are characteristically written in the past tense, relatively many past tense tokens were found. However, due to the fact that low-frequency items are involved in the study, the results for their part have to be considered with some caution (especially as regards the smaller CLMETEV).

The statistical significance for the changes in corpus results was calculated by using the test for the difference between two proportions or percentages (for further information see Leppälä (2007, 63; in Finnish) or Morris and Lobsenz (1998, 6)). In order to increase the reliability of the results, the levels of confidence were set at  $p \leq 0.01$  for significant and at  $p \leq 0.001$  for highly significant results, instead of the rather loose  $p \leq 0.05$ , which has been used in some linguistic studies. However, it should be noted that if a given result does not turn out to be statistically significant at the level of confidence employed in the present study, it does not automatically entail that it is not important. Therefore, it is strongly advisable to observe the exact  $p$  values for each result as well.

For the sake of clarity, the comparisons are presented with the help of column graphs in this section; see the appendix for exact raw frequencies, singular percentages, and  $p$  values for each graph.



## 5.1 Comparison of the ‘mostly singular’ group

The following figure shows the changes in the frequencies of the ‘mostly singular’ collectives between the two corpora. The dark columns represent the percentages of singular verb constructions selected by the three collective nouns (indicated by ‘sg %’ in the tables above). Correspondingly, the light columns stacked above the dark ones represent the percentages of plural verb constructions. The adjacent columns represent the CLMETEV and the BNC in that order from left to right:

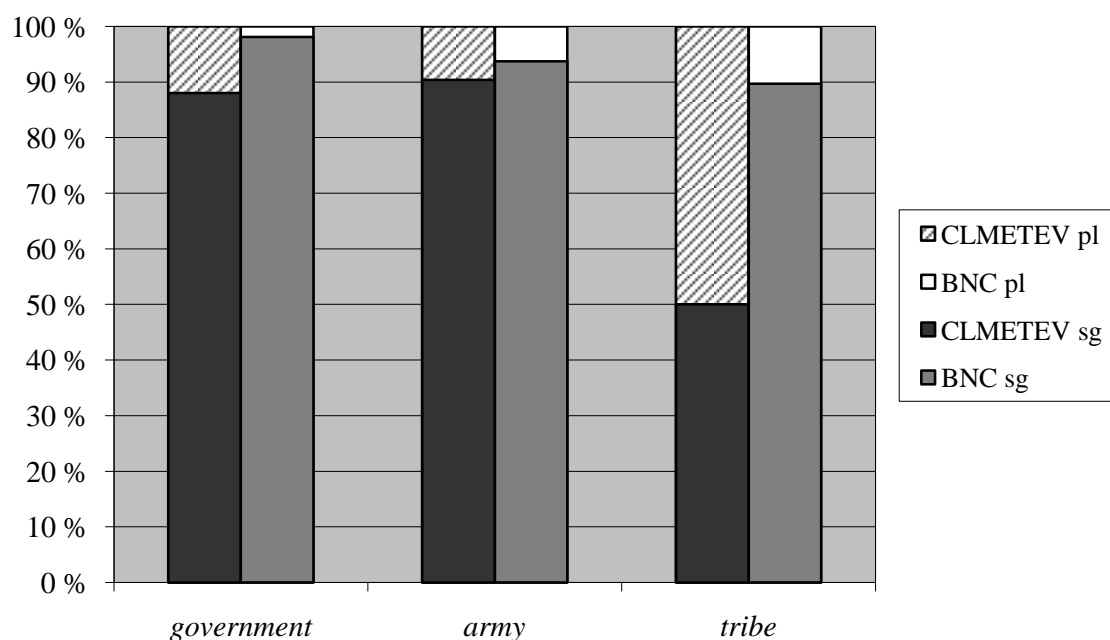


Figure 1: The changes in the frequencies of the ‘mostly singular’ collectives followed by singular and plural verbs in the CLMETEV and in the BNC

As the figure above clearly shows, the collectives that prefer singular agreement in the CLMETEV prefer it even more in the BNC, and correspondingly the number of plural constructions following them has decreased between the two corpora. The increase for *government* is about 10 percentage points, which turns out to be highly significant ( $p \leq 0.001$ )

because the number of relevant tokens is fairly large in both corpora – 175 in the CLMETEV and as much as 3,434 in the BNC (see also the appendix). Although the singular percentage for *army* has not increased as much as for *government*, it has still increased over 3 percentage points between the corpora. However, the change is not significant at  $p \leq 0.01$ . The huge increase for *tribe* (almost 40 percentage points) could be partly explained by the fact that it is by definition the low-frequency item in the ‘mostly singular’ group (see section 4 above), and therefore relatively few relevant tokens can be encountered in the corpora – only 12 in the CLMETEV and 29 in the BNC. Supposedly the increase would not be quite as great if the corpora and hence also the samples were larger; yet the frequencies clearly indicate that the preference for singular agreement would seem to be on the increase with *tribe* – and in fact, the change between the two corpora is significant ( $p \leq 0.01$ ). It is also worth noting that if the categories based on number preference used in this paper would have been formed on the basis of the CLMETEV frequencies instead of the BNC frequencies, *tribe* would have belonged to the ‘variable’ group, not to the ‘mostly singular’ group.

The following example from the BNC quite aptly demonstrates how strong the preference for singular verb construction after both *government* and *army* seems to be in present-day British English:

[136] [. . .] the economic cost of disruption to daily life in Britain will continue to rise as long as the British *government and its army continues* to occupy part of Ireland [. . .] (HLE 1974)

Even though the subject of the sentence is a coordinated one (i.e. *government and its army*), the verb is in the singular. Although the example above clearly does not adhere to the well-known rule on number with coordinate constructions stated by most grammarians (e.g. Quirk et al. 1985, 760) and could therefore be considered deviant or even erroneous, it may be an instance of the proximity principle in action (for further discussion, see section 3.3.3 above). However, if it were an example of proximal agreement, it would be a very unusual one, for at

least two reasons. Firstly, the proximity principle does not usually effect a change from the plural to the singular (Leech and Svartvik 2002, 275), and secondly, distance usually plays an important role in proximal agreement; i.e. the principle of proximity is more likely to be triggered if there is a considerable distance between the verb and its actual head (Jespersen 1961, 94), which is obviously not the case here. One possible interpretation could also be that the speaker views *government* and *its army* as some kind of a close entity, and therefore uses a singular verb. If this was the case, [136] would be an instance of notional agreement (see section 3.3.2 above for more information), not proximal agreement. All in all, regardless whether there is proximal agreement in the example above, it is an unusual construction, and of questionable acceptability.

### **5.1.1 *Government* followed by the present and past tense forms of *be***

The following figure displays the developments in the frequencies of present and past tenses of *be* following *government* and compares them in both corpora. The dark columns represent the singular forms *is* and *was*, and the light ones the plural forms *are* and *were*:

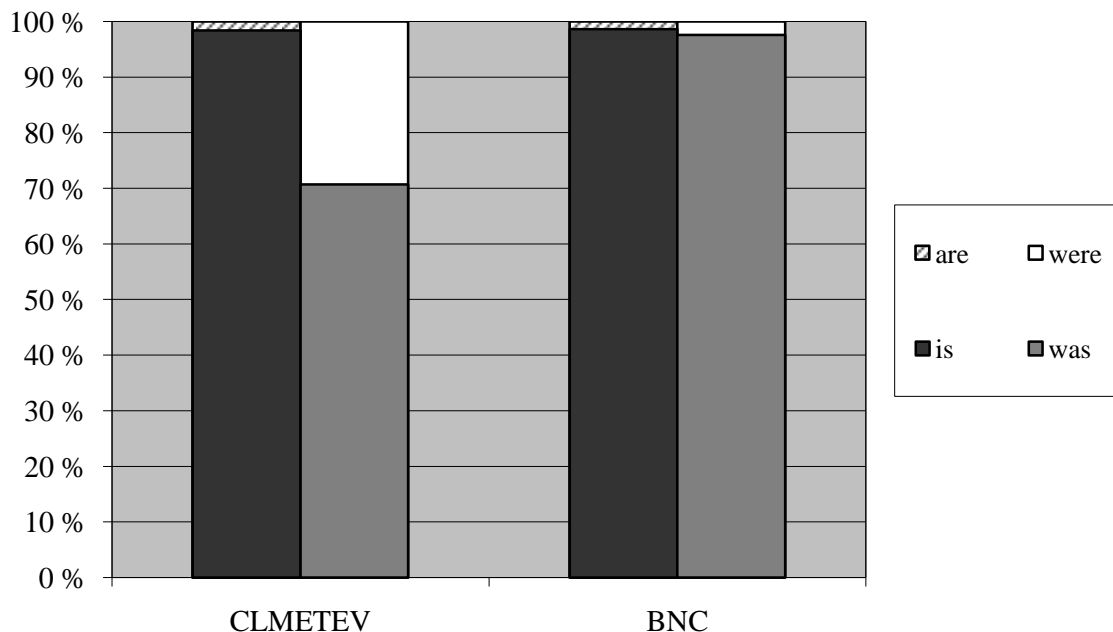


Figure 2: *Government* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

As figure 2 above clearly shows, the proportions of the present tense verbs *is* and *are* have remained almost exactly the same between the corpora: in the CLMETEV, as much as 98.4% of the present tense instances of the verb *be* are in the singular, whereas the corresponding percentage in the BNC is 98.6%. Not surprisingly, the change in the present tense results is not significant at  $p \leq 0.01$ .

As regards the past tense tokens, it is rather surprising to find that almost every third of them following *government* in the CLMETEV is an instance of the plural *were*. However, in the BNC the proportion of the past tense singular *was* is nearly at the level of the present tense singular *is*, and the change is also highly significant at  $p \leq 0.001$  (for exact figures, see the appendix). The rise is made even more significant if we take into account that as many as 10 out of the total of 19 past tense plural tokens in the BNC are instances of hypothetical past tense manifested in *if*-clauses (Leech and Svartvik 2002, 111, 146), whereas the

corresponding figure in the CLMETEV is only 3 out of the total of 12 tokens. Consider the examples from the BNC:

[137] If the *Government were* concerned to arrive at a proper system, they would accept that fairness was the answer. (HHX 12199)

[138] If the *Government were* likely to meet the full cost, the proposal would be unnecessary. (A1F 136)

[139] The CPRS paper argued that if the *Government were* serious about reducing public expenditure, the time had come to consider some radical solutions. (ABU 989)

If these cases of hypothetical past tense were excluded from the study, the proportion of past tense plural verbs would be even smaller in the BNC results, and hence the rise in the singular percentage of past tense tokens would be even more significant. It is worth noting that with the exception of *government*, cases of hypothetical past tense are rare in the data: there are only one case with *army*, one with *tribe*, and three with *family*. Therefore, it is safe to state that cases of hypothetical past tense do not affect the results of the present study significantly.

### 5.1.2 *Army* followed by the present and past tense forms of *be*

The figure below displays the singular and plural instances of the verb *be* following *army* in the CLMETEV and in the BNC:

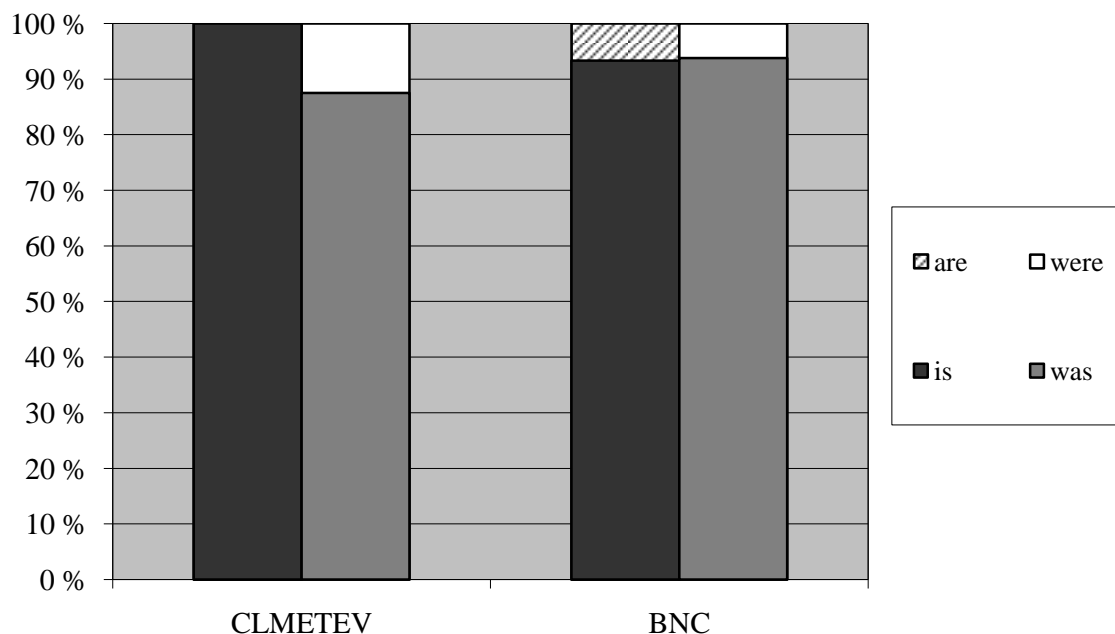


Figure 3: *Army* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

The columns for *army* in the figure above bear a great resemblance to those in figure 2 in section 5.1.1 above. As with *government*, there are relatively more instances of *were* – or relatively less instances of *was* – following the collective in the CLMETEV than in the BNC, although the increase of 6 percentage points is not just as great as the 25 percentage points found with *government*. However, the proportions for present tense tokens have changed to the opposite direction between the corpora: there are no instances of *army* followed by *are* in the CLMETEV, whereas almost 7 percents of the present tense tokens in the BNC are instances of *are*.

As the CLMETEV includes plenty of war narratives (Churchill's *The River War*, Freeman's *William the Conqueror*, Carlyle's *The French Revolution*, Gibbon's *Decline and Fall of the Roman Empire*, etc.), the overall amount of relevant instances of *army* is relatively high in the CLMETEV, especially as regards past tense tokens (N=72):

[140] The Pannonian *army was* at this time commanded by Septimius Severus, a native of Africa [. . .] (Gibbon (1776): *Decline and Fall of the Roman Empire* (Vol. 1))

[141] The whole French *army was* slain, scattered, or taken prisoners. (Freeman (1888): *William the Conqueror*)

[142] Nejami's *army was* recruited almost entirely from these sources. (Churchill (1899): *The River War*)

However, neither the change perceived in the present tense forms of *be* following *army* nor the results for the past tense is statistically significant at  $p \leq 0.01$ .

### 5.1.3 *Tribe* followed by the present and past tense forms of *be*

The following figure presents the proportions of present and past tense instances of the verb *be* following *tribe* in the two corpora:

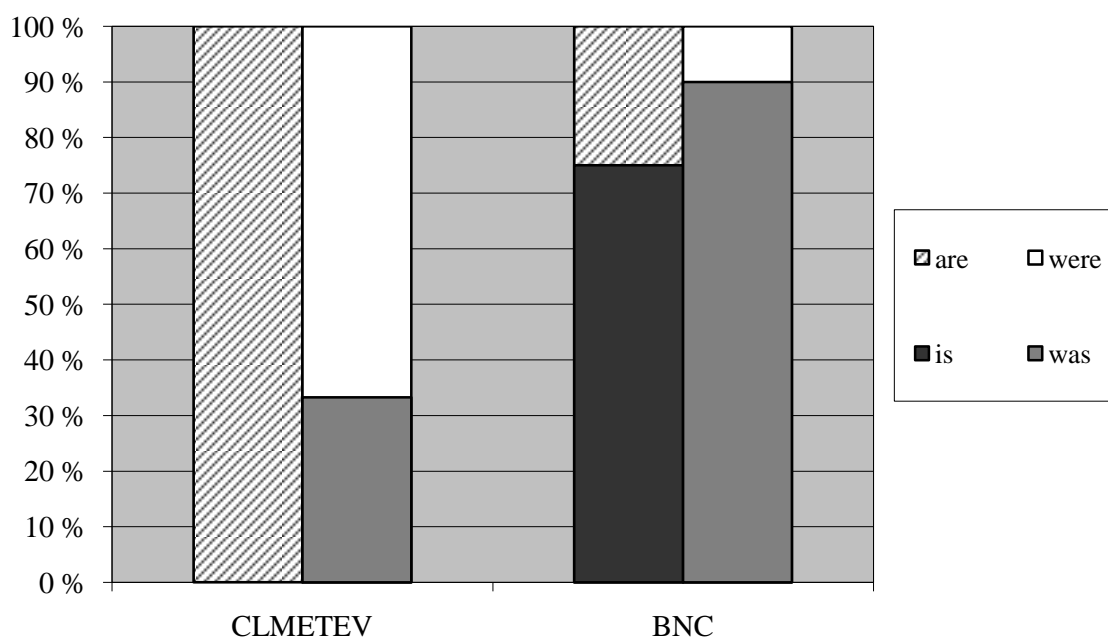


Figure 4: *Tribe* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

As mentioned in section 5 above, the number of verbs found following the low-frequency collective nouns is in many cases rather small, and therefore the results have to be considered with certain caution. As for *tribe*, there are only one present tense and 6 past tense instances of *be* in the CLMETEV. Therefore, the results for the BNC (the number of present tense tokens being 8 and past tense tokens 10) probably mirror real language usage at least somewhat better. Despite the rather small number of tokens, the change visible in the figure above is quite dramatic: *tribe* has substantially shifted its preference towards the singular in both present and past tense (see also section 5.1 above).

Because of the low number of relevant present tense tokens in the corpora (1 in the CLMETEV and 8 in the BNC), the present tense results are not statistically significant at  $p \leq 0.01$ . However, as the number of past tense tokens in the corpora is somewhat greater (6 in the CLMETEV and 10 in the BNC), the results are in fact statistically significant at  $p \leq 0.01$  (see the appendix for exact figures). Therefore, it seems that just like with the two other ‘mostly singular’ collective nouns (see sections 5.1.1 and 5.1.2 above), the proportion of the past tense singular tokens has increased at the expense of the plural ones with *tribe* as well. Having said that, the increase would most likely be slightly smaller than the 55 percentage points displayed in figure 4 above, if the corpora and samples were greater in size.

## 5.2 Comparison of the ‘variable’ group

The figure below shows how the frequencies of the ‘variable’ collective nouns followed by singular and plural verbs have changed between the two corpora:



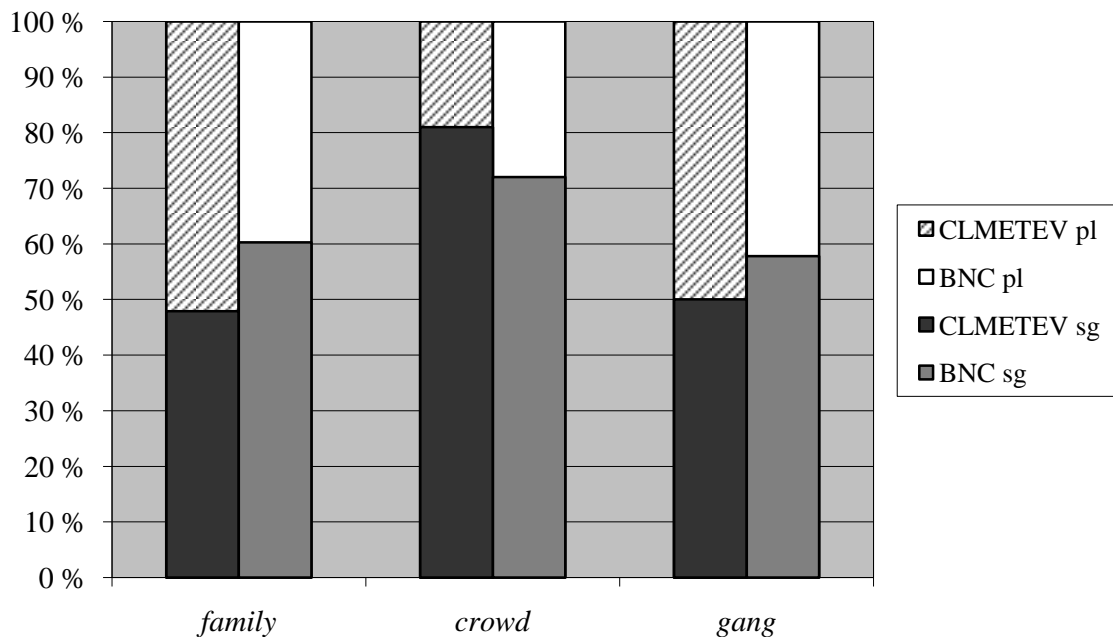


Figure 5: The changes in the frequencies of the ‘variable’ collectives followed by singular and plural verbs in the CLMETEV and in the BNC

As the figure above shows, *family* and *gang* select somewhat more often singular verbs in the BNC than in the CLMETEV. However, while over 80 percent of the tokens with *crowd* are in the singular in the CLMETEV, only 72 percent of them are in the singular in the BNC. It is rather interesting that even though *crowd* could have been categorised as ‘mostly singular’ in the CLMETEV, it has moved towards the ‘variable’ status according to the data from the BNC. While at the same time *family* and *gang* – both of them genuinely variable in the CLMETEV (ca. 50 percent both singular and plural tokens) – have moved towards singular preference, it is fairly logical to state that changes in number agreement with collective nouns cannot be predicted only by looking at data from one single point in time. Instead, in order to make accurate future predictions, diachronic data from several periods of time should be compared, as in the present study.

Whereas the changes perceived in the number preference of *crowd* and *gang* are not statistically significant ( $p \leq 0.01$ ), the developments in the tokens following *family* are highly

significant ( $p \leq 0.001$ ) because of the large number of relevant tokens (190 in the CLMETEV and as much as 1,468 in the BNC; see also the appendix).

### 5.2.1 *Family* followed by the present and past tense forms of *be*

The following figure displays the frequencies of the collective noun *family* followed by present and past tense tokens of *be* in both corpora:

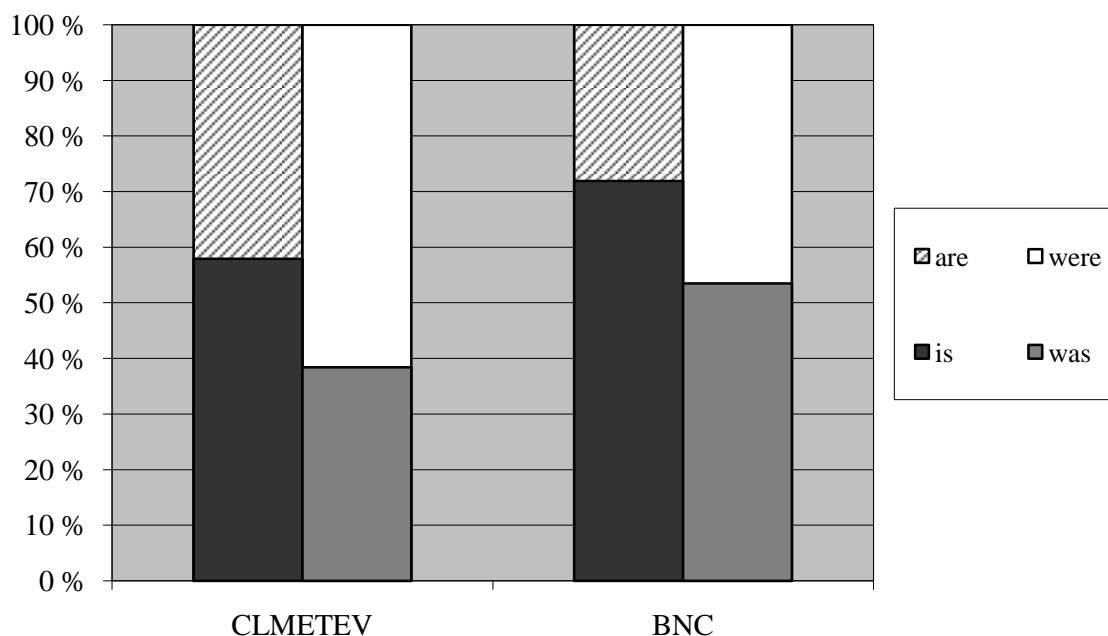


Figure 6: *Family* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

The columns in the figure above obviously reflect the overall change in the number preference of *family* observed in section 5.2 above. However, it is rather surprising that the division between the present tense singular and plural tokens of *be* following *family* differ by almost 20 percentage points from the division between the past tense singular and plural instances of *be*. Moreover, the finding is further confirmed by the fact that the difference is

clearly present in both corpora. It therefore seems that *family* tends to select singular verbs more easily in the present tense than in the past tense, and correspondingly the plural is relatively more often resorted to in the past tense than in the present tense.

The change in the present tense tokens is not significant at  $p \leq 0.01$  (although it is at  $p \leq 0.02$ ; see also the appendix), but the development observed in the past tense is ( $p \leq 0.01$ ).

### 5.2.2 *Crowd* followed by the present and past tense forms of *be*

The frequencies of the present and past tense tokens of *be* following *crowd* in the CLMETEV and in the BNC are presented in the figure below:

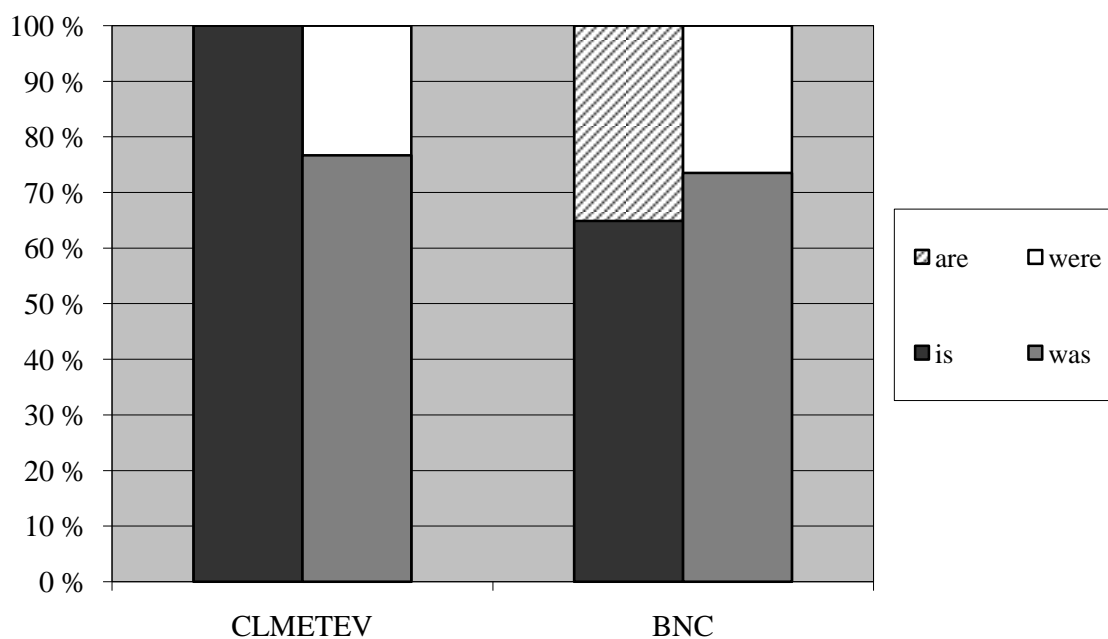


Figure 7: *Crowd* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

Again, figure 7 above seems to mirror *crowd*'s overall change towards plural agreement discussed in section 5.2 above. However, it should be noted that there are only 4 relevant

present tense instances of the verb *be* following *crowd* in the CLMETEV, and therefore the 100% preference for singular agreement shown in figure 7 above does not probably convey the whole picture truthfully. Instead, the number of past tense tokens in the CLMETEV is 30, and therefore the results for their part should be more reliable, although neither of the results is statistically significant at  $p \leq 0.01$ . Interestingly enough, the singular percentage for the past tense tokens is almost 10 percentage points greater than for the present tense tokens in the BNC. If we keep in mind the overall change of *crowd*'s number preference towards plural agreement discussed in section 5.2 above, the development visible in figure 7 above would seem to suggest that the past tense instances of *be* follow the overall change somewhat more slowly than the present tense instances.

### **5.2.3 *Gang* followed by the present and past tense forms of *be***

The frequencies of *gang* followed by present and past tense forms of *be* in the two corpora are presented in the figure below:

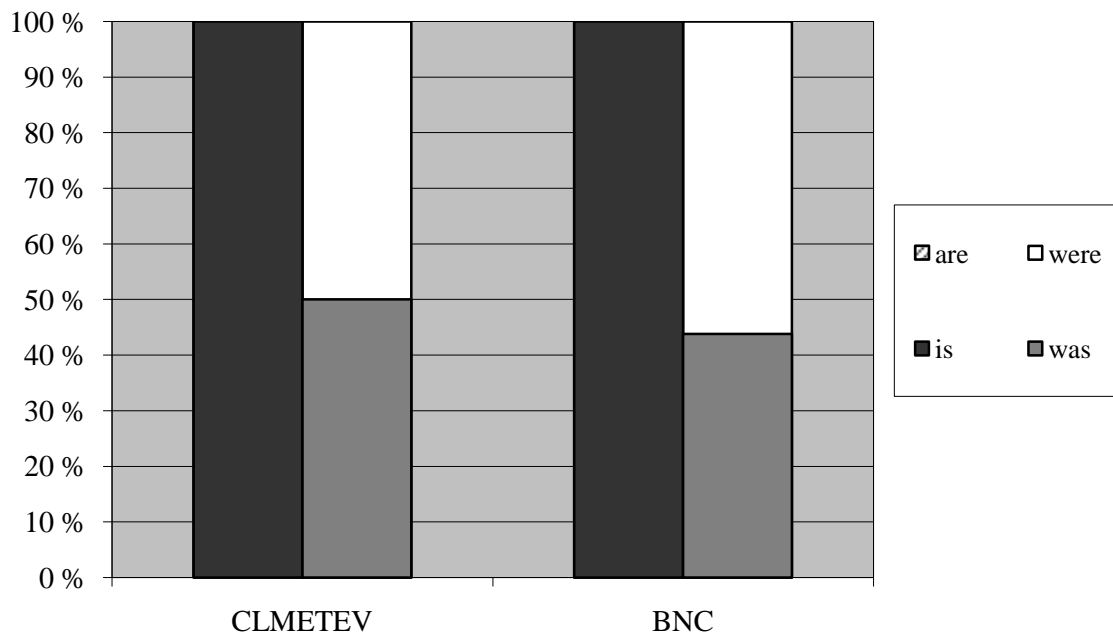


Figure 8: *Gang* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

As with the low-frequency collective *tribe* in section 5.1.3 above, the number of relevant tokens following *gang* is also rather low in both corpora: there are altogether only 4 tokens in the CLMETEV and 25 in the BNC. Although the numbers are admittedly small and therefore the results are not statistically significant ( $p \leq 0.01$ ), they might still offer a hint as to what kind of change has taken place in the time-span defined by the two corpora. For instance, it is rather interesting that the relevant present tense tokens in the BNC include 9 tokens of *gang* followed by *is* and not a single one followed by *are*, whereas the corresponding figures for the relevant past tense tokens are 7 and 9, respectively. These figures would seem to suggest that there is a certain difference as regards the number preference of *gang* in various tenses – in the present tense it has a strong tendency to select singular verbs, whereas plural verbs are almost equally possible in the past tense. However, larger corpora with a greater number of relevant tokens would be required in order to draw any decisive conclusions as regards *gang* and other low-frequency collectives.

### 5.3 Comparison of the ‘mostly plural’ group

The following figure displays the changes in singular/plural preference of the three ‘mostly plural’ collectives followed by singular and plural verbs in the two corpora:

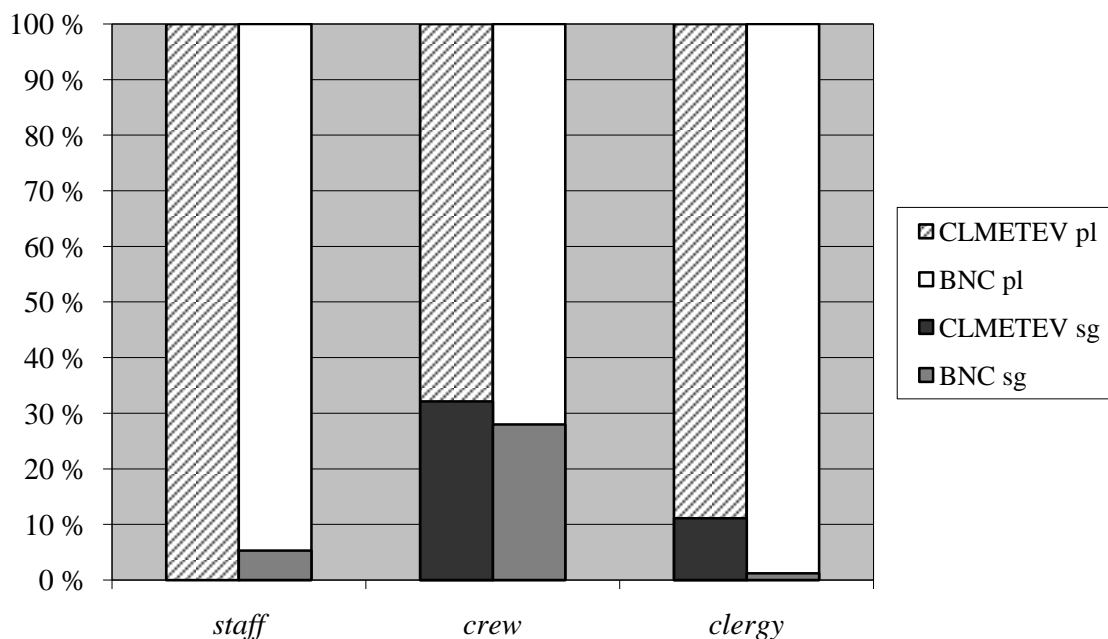


Figure 9: The changes in the frequencies of the ‘mostly plural’ collectives followed by singular and plural verbs in the CLMETEV and in the BNC

As already mentioned in section 4.2.3 above, there are only two relevant instances of *staff* in the CLMETEV, and therefore a comparison with the BNC results is unfortunately not possible for *staff*'s part. As for *crew* and *clergy*, the preference seems to have moved further towards plural agreement between the two corpora: the difference between the results is over 4 percentage points for *crew* and 10 percentage points for *clergy*. The measured change in the number preference of *clergy* is statistically significant at  $p \leq 0.01$ , whereas the change for *staff* and *crew* is not ( $p \leq 0.01$ ; for further information, see the appendix).

To complete the picture for *staff*'s part, a couple of searches were run using the Lancaster-Oslo/Bergen corpus (LOB) and the Freiburg-Lancaster-Oslo/Bergen corpus (FLOB) (for further information, see section 2.3 above). The data were analysed using the *WordSmith Tools* software program.

The results from the LOB and the FLOB seem clearly to corroborate the overall direction of the change observed in 'mostly plural' group between the CLMETEV results and the BNC results. In the LOB, there are altogether 8 instances of *staff* directly followed by a present tense verb or a past tense *was/were*. 3 out of these tokens are in the singular, whereas 5 are in the plural.

However, in the FLOB, there are altogether 10 relevant instances of *staff*, out of which only 1 is in the singular and 9 in the plural. Although the sample sizes are rather small due to the small size of the corpora and the perceived change is not statistically significant ( $p \leq 0.01$ ), the difference is interesting, especially because the time-span between the corpora is only 30 years – compared to the fact that the changes in the frequencies of *crew* and *clergy* visible in figure 7 above have happened in 70-300 years' time. The following chart graphically illustrates the change in the frequency of *staff* followed by singular and plural verbs in the LOB and the FLOB:

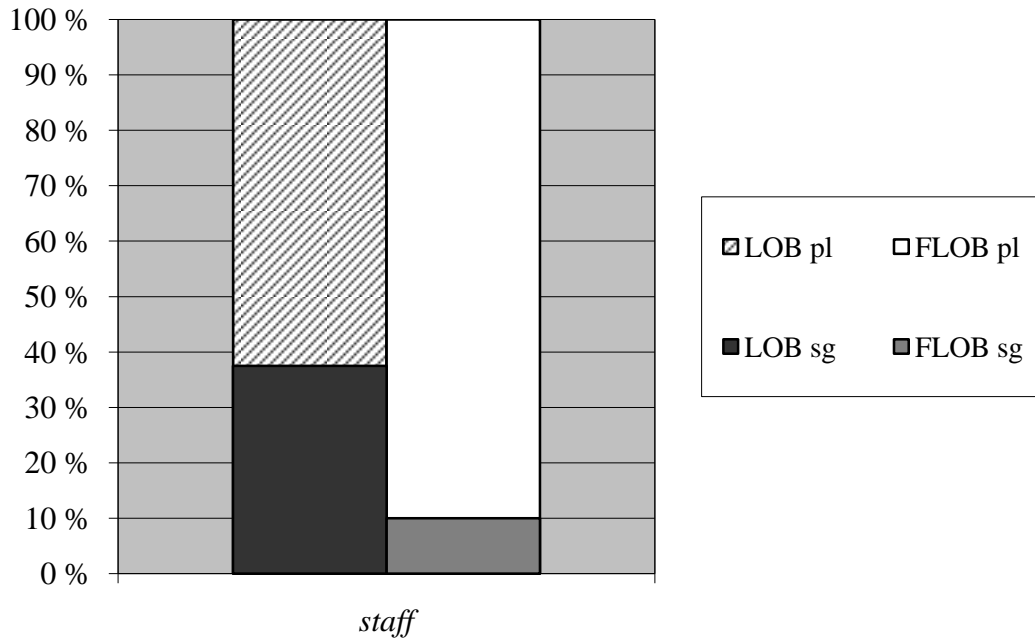


Figure 10: The changes in the frequencies of *staff* followed by singular and plural verbs in the LOB and in the FLOB

### 5.3.1 *Staff* followed by the present and past tense forms of *be*

The figure below shows the frequencies of the present and past tense verbs following *staff* in the CLMETEV and in the BNC:



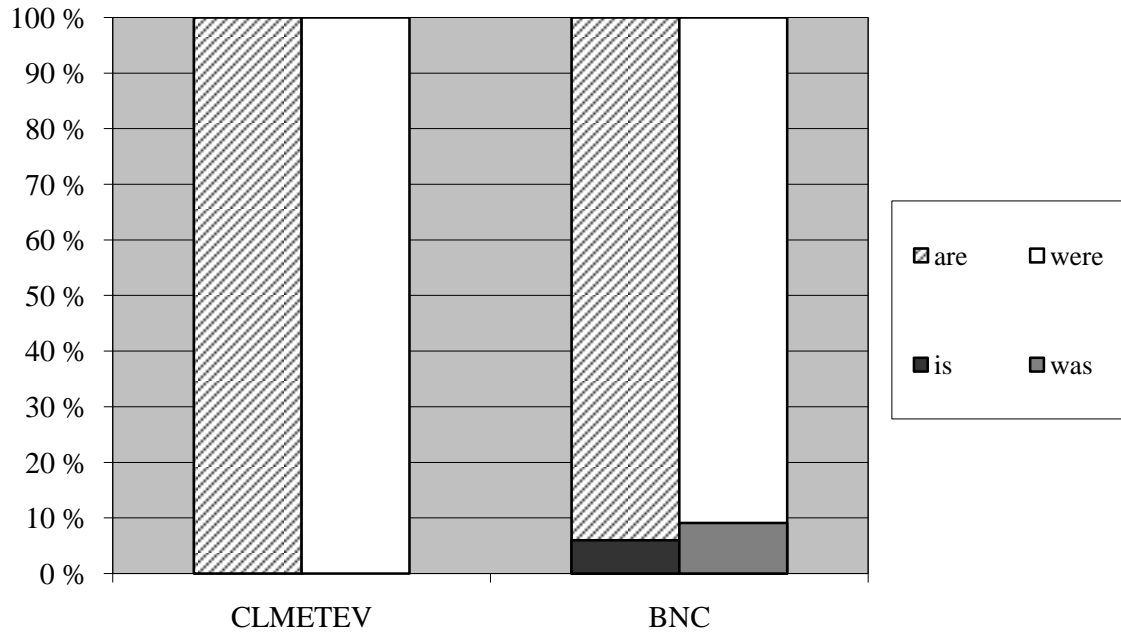


Figure 11: *Staff* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

As is evident from the figure above, the results for *staff* from the two corpora cannot be compared due to the aforementioned lack of relevant tokens in the CLMETEV (see sections 4.2.3 and 5.3 above), and therefore the results are not statistically significant ( $p \leq 0.01$ ) either. However, although the CLMETEV results have to be ignored, the 1,379 relevant tokens from the BNC offer a reliable starting point for speculations and interpretations.

The overall singular percentage for the verbs following *staff* in the BNC is 5.3%, as mentioned in section 4.3 above. The singular percentages for both the present and past tense tokens of the verb *be* are above the overall level: 6% and 9.1%, respectively. Although the differences are rather small, they should not be ignored because of the fairly large number of relevant tokens. The difference of 3.1 percentage points between the singular percentage of present and past tense tokens would seem to suggest that in the past tense, *staff* is relatively more often followed by a singular form of *be* than in the present tense.

### 5.3.2 *Crew* followed by the present and past tense forms of *be*

The following figure compares the frequencies of the present and past tense forms of the verb *be* following the collective noun *crew* in the two corpora:

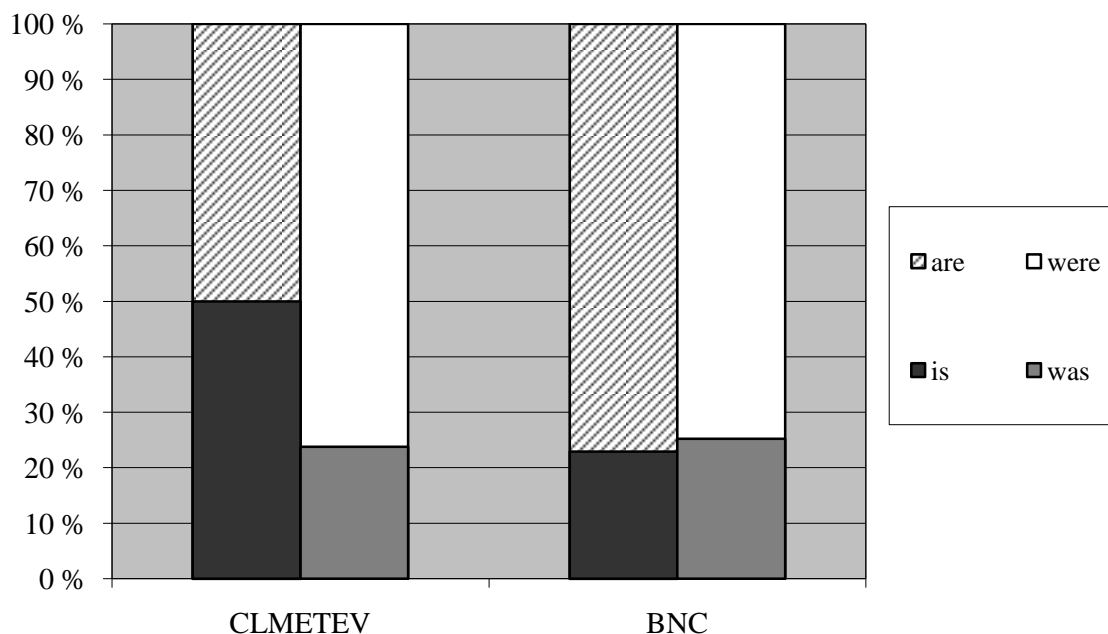


Figure 12: *Crew* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

The seemingly dramatic decrease in the singular percentage of the present tense forms of *be* between the corpora can mostly be explained by the fact that there are only 4 present tense forms of *be* following *crew* in the CLMETEV, whereas the corresponding figure for the past tense forms is 21. The proportions of *was* and *were* seem to have remained relatively unchanged between the corpora, and in the BNC, the proportions of *is* and *are* are only a few percentage points closer to plural agreement than their past tense counterparts. Neither the present tense nor the past tense results are statistically significant at  $p \leq 0.01$ .

### 5.3.3 *Clergy* followed by the present and past tense forms of *be*

The figure below displays the frequencies of the present and past tense forms of *be* following *clergy* in the CLMETEV and in the BNC:

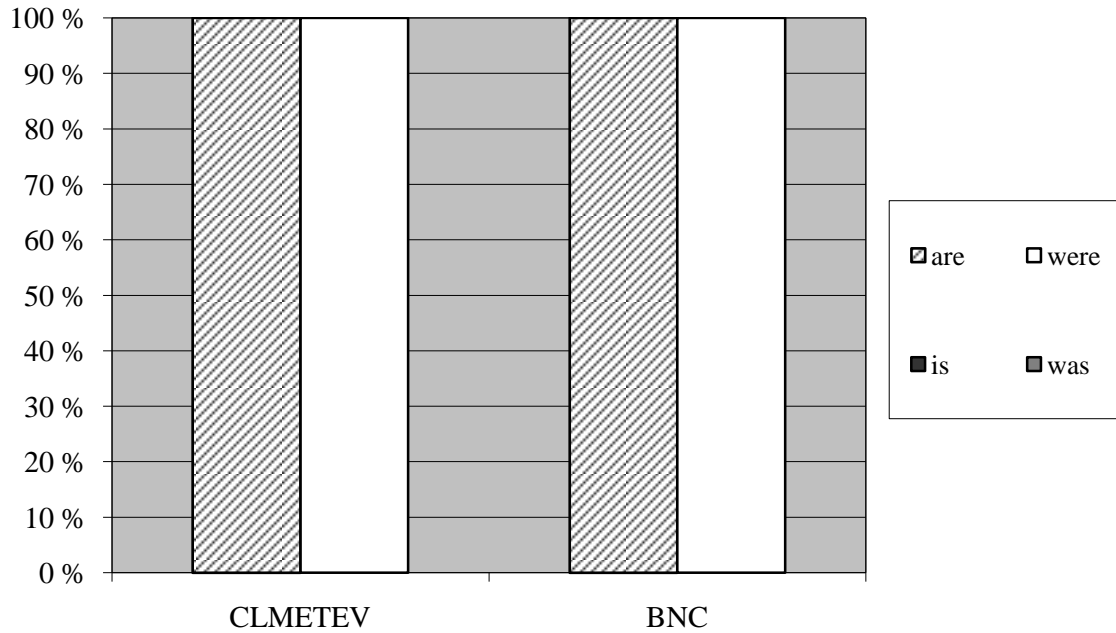


Figure 13: *Clergy* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

As figure 13 above clearly shows, *clergy*, being the low-frequency item in the ‘mostly plural’ group (see section 3 above), is followed by only plural forms of *be* in both corpora. Unfortunately, all the singular verbs following *clergy* are other than forms of *be*, and therefore any comparison between the corpora is impossible with this particular collective. Needless to say, the results are not statistically significant ( $p \leq 0.01$ ).

## 6 Discussion

For clarity's sake, the discussion of the CLMETEV and the BNC results is divided into subsections according to the type of change in question. The discussion begins by the changes observed in the collective nouns' number preference, goes on to analysing these developments as regards the tense of the following verb, and finally proceeds to the changes in the collective nouns' number preference from the viewpoint of frequency groups.

### 6.1 Changes in the collective nouns' number preference

First of all, the corpus results show that collective nouns do indeed possess distinct number preferences that are relatively stable, yet not unchangeable. However, perhaps the most apparent result emerging from the comparison between the CLMETEV and the BNC data is that the three 'mostly singular' collective nouns (*government*, *army*, and *tribe*) have moved further towards singular agreement during the time-span defined by the CLMETEV and the BNC (see section 5.1 above), and correspondingly the three 'mostly plural' collective nouns (*staff*, *crew*, and *clergy*) have moved further towards plural agreement (see section 5.3 above). The direction of change with the 'variable' collective nouns (*family*, *crowd*, and *gang*) seems to vary: *family* and *gang* have moved towards singular agreement, whereas *crowd* has moved towards plural agreement (see section 5.2 above). In addition, each preference group includes at least one collective noun displaying a statistically significant change ( $p \leq 0.01$ ) in number preference. Therefore, the results of the present study should be fairly reliable.

Interestingly, the potential slight number preference of the 'variable' collectives in the CLMETEV does not seem to affect the direction of change: although as much as 80% of the verbs following *crowd* are in the singular in the CLMETEV, *crowd*'s preference has moved towards the plural in the BNC – and although more than 50% of the relevant instances of

*family* are followed by plural verbs in the CLMETEV, its preference has moved towards the singular in the BNC. The differences between the three groups are very clear, which is typical of lexical diffusion (see section 3.5.2 above).

These results seem to corroborate Levin's (2006, 321, 333) findings obtained by comparing the 1990 and 2000 CD-ROM editions of *The Independent*: he found that the choice between a singular and a plural verb is largely motivated by the individual preferences of collective nouns, not necessarily by the semantics of the following verb, as is commonly believed. He also found that even in a rather short ten years' time, those collectives that preferred singular verbs continued to move further towards singular agreement and those preferring plural verbs continued to move further towards plural agreement. According to his results, variable collectives remained variable. However, the rather short period of time which the corpora used in his study represent may be a restrictive factor in noticing a steady change, especially if it is slow in nature, as he himself also points out (*ibid.*, 332-333). Considering the present results obtained from the CLMETEV and the BNC, it is plausible to suggest that variable collective nouns could also be slowly moving either towards singular or plural agreement, depending on their internal word-specific preferences. Of course, this is merely speculation, and further research needs to be conducted before any final conclusions could be drawn.

The results of the present study apparently challenge Bauer's (1994, 65) claim that singular agreement would be on the increase with all collective nouns. Moreover, although the principle of markedness discussed in section 3.5.1 above seems to be valid to describe the behaviour of most collective nouns, the present results show that it does not override the effect of the word-specific number preference of each individual token, since some of the collective nouns have developed a clear and increasing preference for the plural. Therefore, instead of the expected change where one alternative would become more and more unmarked

at the expense of the other, two simultaneous drifts towards the opposite poles were observed. In addition to this, the change documented in the present study also displays two features that are typical of a grammaticalisation process: a loss in choice or variability, and, closely connected to the first feature, a loss of meaning distinctions between the opposite poles – here, the singular and the plural (Lehmann 1995, 127, 138; see also section 3.5.1 above).

Thus, the central observation as regards markedness and grammaticalisation seems to be that instead of slowly changing towards one unmarked alternative (and perhaps eventually becoming fully grammaticalised), the agreement patterns of collective nouns are becoming increasingly divided, and therefore it would be justifiable to talk about two unmarked alternatives instead of one as regards the number preference of collective nouns.

## **6.2 Changes in the tense proportions of the verbs following collectives**

The comparison of the corpus results also revealed interesting aspects about the effect that the tense of the following verb seems to have on the choice of number. Most English speakers intuitively know that the semantic attributes of the verb following a collective noun indirectly influence the choice of number; meaning that the semantics of the verb often seem to fall into place quite satisfactorily as regards the aspectual distinction between an undivided body and a collection of individuals (discussed already in section 3.4.4 above). If we take *family* as an example, the verbs following it that most often occur in the plural in the BNC include for example *live, say, do, know, get, feel, and enjoy*. They are all relatively concrete in meaning and normally used in connection with individual personal agents, whereas verbs that most often occur in the singular – for example *need* and *own* – are more abstract and describe a state, not an action. Therefore, in order to exclude the effects of semantic factors as much as possible, the effect of different tenses of the same verb *be* were investigated in the present study. Two significant developments were found in the analysis of the results.

Firstly, it turned out that the tense of the following verb does indeed affect the collectives' choice of number. With several collective nouns, the proportions of the singular and plural forms of *be* are surprisingly different in the present and past tense. For example, almost 60% of the present tense forms of *be* following *family* in the CLMETEV are in the singular, whereas the corresponding figure in the past tense is less than 40% (see section 5.2.1 above). The difference is repeated in the BNC as well.

Secondly, the differences between the singular percentages of present and past tense forms of *be* are clearly connected to the direction of change in the overall agreement patterns of the collectives. To summarise the corpus findings on the effect of tense, the past tense forms of *be* seem to be more resistant to changes in number preference than the present tense forms. In order to prove this statement, let us discuss each preference group separately.

With the 'mostly singular' collectives *government* and *army*, the singular percentage of the past tense forms of *be* is clearly lower than that of the present tense forms in the CLMETEV, but in the BNC the difference has greatly diminished (see sections 5.1.1 and 5.1.2 above). In addition, it should be noted that the change observed with the past tense tokens following *government* is also statistically highly significant at  $p \leq 0.001$ .

As regards the 'variable' collectives, *family*, whose number preference has shifted towards singular agreement between the corpora, displays a fairly similar tendency as *government* and *army*: the proportion of the present tense forms of *be* is clearly more bent towards singular agreement than that of the past tense forms in both corpora (see section 5.2.1 above), but in the BNC, the singular percentages of both the present and past tense forms of *be* have greatly increased. Again, the change in the singular percentage of past tense tokens is statistically significant at  $p \leq 0.01$ . However, with *crowd*, whose number preference has changed towards the plural between the CLMETEV and the BNC, the proportions of the present and past tense forms of *be* have developed exactly the other way round compared to *government* and *army*:

in the CLMETEV, the singular percentage is higher with the present tense forms of *be* than with the past tense tokens, but in the BNC the situation is reversed (see section 5.2.2 above).

Finally, the singular percentages of the present and past tense forms of *be* following the ‘mostly plural’ collective *crew* reflect to a large extent the results received by analysing the ‘variable’ collective *crowd*: in the CLMETEV, the singular percentage of present tense tokens is much greater than that of past tense tokens, but in the BNC, there is a decrease of almost 30 percentage points. At the same time, the singular percentage of past tense tokens remains relatively unchanged (see section 5.2.3 above).

Thus, it seems that the past tense forms of *be* are indeed more resistant to change than the present tense forms: the singular percentage of the past tense forms of *be* is clearly lower than that of the present tense forms for the collective nouns that are on the move towards singular agreement (i.e. *government*, *army*, and *family*), and correspondingly, the singular percentage of the past tense forms of *be* is clearly higher than that of the present tense forms for the collectives on the move towards plural agreement (i.e. *crowd* and *crew*). Interestingly, the BNC results also seem to indicate that when the frequency of singular or plural verbs following a collective noun exceeds a certain limit, the singular percentages of both the present and past tense forms of *be* eventually tend to reach an equal state, in spite of the aforementioned resistance the past tense forms seem to possess. According to the frequencies of *government* and *army* in the BNC, this limit would lie somewhere around 93-99% (expressed by means of singular percentage) for collectives with singular preference, and presumably around 1-7% for collectives with plural preference. In short, the corpus findings suggest that the past tense forms of *be* are resistant to change until the preceding collective noun has reached a certain, rather high level of constancy in its number preference.



Since the effect of tense on the collective nouns' agreement patterns has not been studied earlier, these interesting results would seem to call for further investigation and additional tests with larger corpora that enable even longer time-span.

### 6.3 Changes in the number preference from the viewpoint of frequency

Turning to changes between the corpora from the frequency groups' (defined in section 4 above) viewpoint, the following table summarises the changes in the number preference (more precisely, the changes in the singular percentage – see section 4.2 above for definition) of the collective nouns in each frequency group:

<b>high-freq</b>	change (sg %)	<b>mid-freq</b>	change (sg %)	<b>low-freq</b>	change (sg %)
<i>government</i>	10.1%	<i>army</i>	3.3%	<i>tribe</i>	39.7%
<i>family</i>	12.4%	<i>crowd</i>	-9.0%	<i>gang</i>	7.8%
<i>staff</i>	5.3%	<i>crew</i>	-4.1%	<i>clergy</i>	-9.9%

Table 16: The changes in the singular percentage of the collective nouns in each frequency group between the CLMETEV and the BNC

Before drawing any conclusions, there are two things to keep in mind. Firstly, the results for *staff* have to be ignored because of the lack of relevant tokens in the CLMETEV (see section 4.2.3 above), and secondly, the additional reasons influencing the changes connected to some of the collectives (suggested in sections 4.4.1, 4.4.2, and 4.4.3 above) may cause a slight bias in the results. However, it is fairly safe to assume that the general tendencies indicated by the figures in the table above represent real language use accurately enough.

The figures in table 16 seem to suggest that the change in the number preference of *tribe* and *clergy* – the 'end-of-the-scale' low-frequency collectives – has been greater than with the mid-frequency or the high-frequency collectives in their preference groups. Interestingly, the

situation is quite the opposite with the ‘variable’ collectives: the change with the low-frequency collective *gang* has been smaller than with the mid-frequency *crowd*, and the high-frequency *family* in fact displays the greatest change (12.4%) in the ‘variable’ category. Keeping in mind the aforementioned Hooper’s (1976) statement (see section 3.5.2 above on lexical diffusion) that low-frequency items are usually affected by analogical change before others, in the light of the present results it is possible to make certain predictions on the likely future developments as regards the number preference of collective nouns.

The relatively great changes in the singular/plural preference of the ‘mostly singular’ and the ‘mostly plural’ low-frequency collectives (i.e. *tribe* and *clergy*, respectively) would seem to entail that those collective nouns that today prefer singular verbs will increasingly continue to move towards singular agreement in the future, and consequently those collectives that currently have a preference for the plural will increasingly continue to move towards plural agreement. These results seem to be in line with the results already obtained in the present study by comparing the selected collectives in the CLMETEV and in the BNC from the viewpoint of preference groups (discussed earlier in section 6.1 above), and they also corroborate Levin’s (2006, 321, 333) findings. On the other hand, the fact that the number preference of the high-frequency collective *family* has changed the most and the number preference of the low-frequency *gang* has changed the least in the ‘variable’ group would seem to suggest that the rate with which the number preference of ‘variable’ collectives has changed will slow down in the future, or even change direction. This observation would seem to lend support to Levin’s (2006, 321) aforementioned results, according to which variable collectives remain variable.

In view of lexical diffusion, it seems that since the number preference of the low-frequency collective nouns in the ‘mostly singular’ and ‘mostly plural’ groups is changing at a relatively high rate, they are still in the middle of a fairly fast phase of the process (the steep

part of the S-curve; see section 3.5.2 above for further discussion). Therefore, it could be expected that the speed of change will at some point slow down and eventually stop as the process reaches the final stage of entrenchment. Moreover, Hopper and Traugott's (2003, 155) aforementioned statement implying that the spread of one allomorph at the expense of others is aided by the sheer frequency of the successful allomorph does not seem to accord perfectly with subject-verb agreement and collective nouns. This can be stated for two reasons: firstly, if we postulated that the singular is the successful allomorph, the corpus findings would challenge the statement, since several collectives in the corpora are on their way towards plural agreement; secondly, even if we postulated that the singular is the successful allomorph for the collectives preferring singular agreement and the plural for the collectives preferring plural agreement, the corpus results include tokens such as *crowd* (see section 5.2 above), which has a relatively clear preference for singular in the CLMETEV, but in the BNC the preference has changed towards the plural. On the other hand, cases like this are relatively few – for the majority of ‘mostly singular’ and ‘mostly plural’ collectives, Hopper and Traugott's statement seems to be valid.

## 7 Conclusion

The present thesis has empirically shown that although factors such as region, style, point of view, and syntactic context affect the choice between singular and plural after a collective noun to some extent, the basis for that choice seems to be the relatively stable and highly individual word-specific preferences that are embedded in the very nature of collective nouns. While the corpus results prove that this internal preference is indeed quite steady, there are also clear signs of slow diachronic developments that are still in progress.

The corpus findings indicated that the collective nouns that had a clear preference for singular agreement over a hundred years ago have continued to move further towards the singular when compared to present-day British English, and those collectives that clearly preferred plural agreement in the past have continued to move further towards the plural. The variable collectives have remained variable, yet some of them are slowly moving towards the singular or the plural (see section 6.1 above).

The rate at which the low-frequency items' preference has changed in the 'mostly singular' and 'mostly plural' groups seems to indicate that the developments observed in the corpus data would continue with an increasing speed. On the other hand, the analysis of the low-frequency items also seems to suggest that the rate of change for the 'variable' collectives would slowly decelerate (see section 6.3 above). However, it should be noted that since the results obtained by comparing low-frequency items must be by definition interpreted with caution, further research is needed to confirm the findings as regards the future prospects of the three groups.

In addition, a previously unknown and very interesting tendency was discovered in the corpus analysis: the tense of the verb following a collective noun influences the choice between the singular and the plural. Moreover, the tense was found to be connected to the direction of change in the overall agreement patterns of the collectives: as to the verbs

following the ‘mostly singular’ collective nouns, the past tense forms of *be* were relatively more often found in the plural than the present tense forms – and with respect to the verbs following the ‘mostly plural’ collective nouns, the past tense forms of *be* were relatively more often found in the singular than the present tense forms (see section 6.2 above). Therefore, the past tense forms of *be* seemed to be more resistant to changes in number preference than the present tense forms. This resistance was found to be valid until a certain level of constancy (app. 93-99% or 1-7% correspondingly, expressed in singular percentages) in the collective’s number preference is reached.

The present thesis has shed some light on the less acknowledged factors influencing subject-verb agreement with collective nouns, as well as on the recent developments in these agreement patterns in British English. Together with other recent studies, the present thesis has demonstrated that collective nouns have surprisingly individualised patterns of use, and therefore future research should seek to explore these still largely uncharted territories profoundly by examining the collective nouns’ agreement patterns one by one.

## Primary Sources

BNC = The British National Corpus (1994) [Internet] Oxford: Oxford University Computing Services. Available from <<http://www.sketchengine.co.uk>>. [Accessed 13 November 2007]

CLMETEV = The Extended Version of the Corpus of Late Modern English Texts [CD-ROM] Leuven: University of Leuven. [Accessed 16 December 2007]

FLOB = The Freiburg-Lancaster-Oslo/Bergen Corpus (1997) [CD-ROM] *ICAME Collection of English Language Corpora*. 2<sup>nd</sup> ed. Bergen: The HIT Centre [Accessed 8 January 2008]

LOB = The Lancaster-Oslo/Bergen Corpus (1978) [CD-ROM] *ICAME Collection of English Language Corpora*. 2<sup>nd</sup> ed. Bergen: The HIT Centre [Accessed 8 January 2008]

## Secondary Sources

Andersen, H. (2001) *Actualization: Linguistic Change in Progress*. Amsterdam: John Benjamins Publishing Company.

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

Bache, C. and Davidsen-Nielsen, N. (1997) *Mastering English: An Advanced Grammar for Non-native and Native Speakers*. Berlin: Mouton de Gruyter.

Battistella, E. L. (1990) *Markedness: The Evaluative Superstructure of Language*. Albany: State University of New York Press.

Battistella, E. L. (1996) *The Logic of Markedness*. Oxford: Oxford University Press.

Bauer, L. (1994) *Watching English Change*. London: Longman.

Bauer, L. (2004) "Inferring Variation and Change from Public Corpora." In *The Handbook of Language Variation and Change*. Chambers, J. K., Trudgill, P., and Schilling-Estes, N. 97-114. Malden: Blackwell Publishing Ltd.

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

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

- BNC User Manual*. (2005) [Internet] Oxford: The University of Oxford. Available from <<http://www.natcorp.ox.ac.uk/corpus/index.xml?style=.pdf>>. [Accessed 29 February 2008]
- BNCweb Manual*. (2002) [Internet] Zurich: The University of Zurich. Available from <<https://bncweb.uta.fi/html/manual/bncwebman-limitations.htm>>. [Accessed 29 February 2008]
- Carter, R. and McCarthy, M. (2006) *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S., Santiago-Diaz, I., Fotokopoulou, O., Pita, R., and Gil, D. (2004) "How Deep Are Effects of Language on Thought? Time Estimation in Speakers of English, Indonesian, Greek, and Spanish". In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. 575-580.
- Chalker, S. (1990) *English Grammar Word by Word*. Walton-on-Thames: Nelson.
- Collins Cobuild English Grammar*. (1990) London: Collins ELT.
- Corbett, G. C. (2000) *Number*. Cambridge: Cambridge University Press.
- Corbett, G. C. (2006) *Agreement*. Cambridge: Cambridge University Press.
- Cruse, D. A. (1986) *Lexical Semantics*. Cambridge: Cambridge University Press.
- De Smet, H. (2005) "A Corpus of Late Modern English Texts". In *ICAME Journal* 29. 69-82.
- Dekeyser, X. (1987) *Foundations of English Grammar*. Antwerpen: INKA.
- Depraetere, I. (2003) "On Verbal Concord with Collective Nouns in British English". In *English Language and Linguistics* 7:1. 85-127.
- Downing, A. and Locke, P. (1992) *A University Course in English Grammar*. New York: Prentice Hall.
- Fowler, H. W. (1926) *A Dictionary of Modern English Usage*. 1<sup>st</sup> edition. London: Oxford University Press.
- Fowler, H. W. (1965) *A Dictionary of Modern English Usage*. 2<sup>nd</sup> edition. London: Oxford University Press.
- Fowler, H. W. (1996) *The New Fowler's Modern English Usage*. 3<sup>rd</sup> edition. Oxford: Oxford University Press.
- Greenbaum, S. and Quirk, R. (1990) *A Student's Grammar of the English Language*. Harlow: Pearson Education Limited.

- Greenberg, J. H. (1966) *Language Universals: With Special Reference to Feature Hierarchies*. The Hague: Mouton de Gruyter.
- Hooper, J. B. (1976) "Word Frequency in Lexical Diffusion and the Source of Morphophonological Change". In Christie, W. M. (ed.) *Current Progress in Historical Linguistics*. Amsterdam: North Holland Publishing Company. 95-105.
- Hopper, P. and Traugott, E. C. (2003) *Grammaticalization*. 2<sup>nd</sup> edition. Cambridge: Cambridge University Press.
- Huddleston, R. (1988) *English Grammar: An Outline*. Cambridge: Cambridge University Press.
- Huddleston, R. and Pullum, G. K. (2002) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Jespersen, O. (1924) *The Philosophy of Grammar*. London: G. Allen & Unwin Ltd.
- Jespersen, O. (1961) *A Modern English Grammar. Part II: Syntax (First Volume)*. London: Bradford & Dickens.
- Johansson, S., Atwell, E., Garside, R., and Leech, G. (1986) *The Tagged LOB Corpus: User's Manual*. [Internet] Bergen: Norwegian Computing Centre for the Humanities. Available from <<http://icame.uib.no/lobman/lob-cont.html>>. [Accessed 29 February 2008]
- Kennedy, G. (1998) *An Introduction to Corpus Linguistics*. London: Longman.
- Kilgarriff, A., Rychly, P., Smrz, P., and Tugwell, D. (2004) "The Sketch Engine". In *Proceedings of the Eleventh EURALEX International Congress*. 105-116.
- Krug, M. (2003) "Frequency as a Determinant in Grammatical Variation and Change". In Rohdenburg, G. and Mondorf, B. (eds.) *Determinants of Grammatical Variation in English*. Berlin: Mouton de Gruyter.
- Langacker, R. W. (1987) *Foundations of Cognitive Grammar. Volume 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Leech, G. (1989) *An A-Z of English Grammar and Usage*. London: Edward Arnold.
- Leech, G. and Svartvik, J. (2002) *A Communicative Grammar of English*. 3<sup>rd</sup> edition. Harlow: Pearson Education Limited.
- Lehmann, C. (1989) "Markedness and Grammaticalization". In Tomić, O. M. (ed.) *Markedness in Synchrony and Diachrony*. Berlin: Mouton de Gruyter. 175-190
- Lehmann, C. (1995) *Thoughts on Grammaticalization – LINCOM Studies in Theoretical Linguistics*. Munich: LINCOM Europa.



- Leppälä, R. (2007) *Tilastollisten Menetelmien Perusteet I, TILTP2, Luentorunko, Syksy 2007*. [Internet] Tampere: University of Tampere. Available from <<http://mtl.uta.fi/tilasto/tiltp2/syksy2007/luennot.pdf>>. [Accessed 4 March 2008]
- Levin, M. (2006) "Collective Nouns and Language Change". In *English Language and Linguistics* 10:2. 321-343.
- Mair, C. (2002) "The Changing Patterns of Verb Complementation in Late Modern English: A Real-Time Study Based on Matching Text Corpora". In *Language and Linguistics* 6:1. 105-131.
- Moravcsik, E. A. (1988) "Agreement and Markedness". In Barlow, M. and Ferguson, C. A. (eds.) *Agreement in Natural Language: Approaches, Theories, Descriptions*. Chicago: The University of Chicago Press. 89-106.
- Morris, S. B. and Lobsenz, R. E. (1998) *Significance Tests and Confidence Intervals for the Adverse Impact Ratio*. [Internet] Illinois: Illinois Institute of Technology. Available from <[www.ipmaac.org/files/morris.pdf](http://www.ipmaac.org/files/morris.pdf)>. [Accessed 20 April 2008]
- OED = The Oxford English Dictionary*. (2008) [Internet] Oxford: The Oxford University Press. Available from <<http://helios.uta.fi:2243>>, accessed via <<http://www.nelliportaali.fi>>. [Accessed 16 December 2007]
- Persson, G. (1989) "On the Semantics of Collective Nouns in English". In Odenstedt, B. and Persson, G. (eds.) *Instead of Flowers: Papers in Honour of Mats Rydén on the Occasion of His Sixtieth Birthday, August 27, 1989*. Stockholm: Almqvist & Wiksell International. 179-195.
- Peters, P. (2004) *The Cambridge Guide to English Usage*. Cambridge: Cambridge University Press.
- Phillips, B. S. (1984) "Word Frequency and the Actuation of Sound Change". In *Language* 60:2. 320-342.
- Phillips, B. S. (2001) "Lexical Diffusion, Lexical Frequency, and Lexical Analysis". In Bybee, J. and Hopper, P. (eds.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins Publishing Company.
- Pinker, S. (1994) *The Language Instinct*. New York: William Morrow and Company.
- Poutsma, H. (1904) *A Grammar of Late Modern English. Part I: The Sentence*. Groningen: P. Noordhoff.
- Pullum, G. K. (1991) *The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on the Study of Language*. Chicago: University of Chicago Press.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.

- Reid, W. (1991) *Verb and Noun Number in English: A Functional Explanation*. London: Longman.
- Schibsbye, K. (1970) *A Modern English Grammar*. 2<sup>nd</sup> edition. London: Oxford University Press.
- Svensson, P. (1998) *Number and Countability in English Nouns: An Embodied Model*. Uppsala: Swedish Science Press.
- Swan, M. (1980) *Practical English Usage*. Oxford: Oxford University Press.
- Sweet, H. (1891) *A New English Grammar, Logical and Historical. Part I – Introduction, Phonology, and Accidence*. Oxford: Oxford University Press.
- Sweet, H. (1898) *A New English Grammar, Logical and Historical. Part II – Syntax*. Oxford: Clarendon Press.
- Thomson, A. J. and Martinet, A. V. (1960) *A Practical English Grammar*. 3<sup>rd</sup> edition. Oxford: Oxford University Press.
- Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins Publishing Company.
- Visser, F. T. (1963) *An Historical Syntax of the English Language. Part One: Syntactical Units with One Verb*. Leiden: E. J. Brill.
- Westergren-Axelsson, M. (1999) "The Use of a Corpus of Students' Written Production in University English Teaching". In Byrman, G., Lindquist, H. and Levin, M. (eds.) *Corpora in Research and Teaching – Papers from the ASLA Symposium Corpora in Research and Teaching, Växjö, 11-12 November 1999*. Uppsala: ASLA.
- Whorf, B. (2000) *Language, Thought and Reality*. Cambridge: MIT Press.
- Zandvoort, R. (1966) *A Handbook of English Grammar*. 4<sup>th</sup> edition. London: Longman.

## Appendix

The raw frequencies and singular percentages for the figures in section 5.

<b>Collective noun</b>	CLMETEV singular	CLMETEV plural	CLMETEV sg %	BNC singular	BNC plural	BNC sg %
<i>government</i>	154	21	88.0%	3,370	64	98.1%
<i>army</i>	122	13	90.4%	313	21	93.7%
<i>tribe</i>	6	6	50.0%	26	3	89.7%

Frequencies for Figure 1: The changes in the frequencies of the ‘mostly singular’ collectives followed by singular and plural verbs in the CLMETEV and in the BNC

<b>Collective noun</b>	<b>Level of statistical significance</b>
<i>government</i>	$p \leq 0.001$
<i>army</i>	$p \leq 0.104$
<i>tribe</i>	$p \leq 0.003$

Statistical significances for Figure 1: Levels of statistical significance for the changes in the singular percentage between the CLMETEV and the BNC

<b>Corpus</b>	<i>is</i>	<i>are</i>	<b>present sg %</b>	<i>was</i>	<i>were</i>	<b>past sg %</b>
CLMETEV	63	1	98.4%	29	12	70.7%
BNC	728	19	98.6%	761	19	97.6%

Frequencies for Figure 2: *Government* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

<b>Collective noun</b>	<b>Level of stat sign / present</b>	<b>Level of stat sign / past</b>
<i>government</i>	$p \leq 0.449$	$p \leq 0.001$

Statistical significances for Figure 2: Levels of statistical significance for the changes in the singular percentage of *government* followed by the forms of the verb *be*

<b>Corpus</b>	<i>is</i>	<i>are</i>	<b>present sg %</b>	<i>was</i>	<i>were</i>	<b>past sg %</b>
CLMETEV	31	0	100.0%	63	9	87.5%
BNC	70	5	93.3%	122	8	93.8%

Frequencies for Figure 3: *Army* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

<b>Collective noun</b>	<b>Level of stat sign / present</b>	<b>Level of stat sign / past</b>
<i>army</i>	$p \leq 0.071$	$p \leq 0.061$

Statistical significances for Figure 3: Levels of statistical significance for the changes in the singular percentage of *army* followed by the forms of the verb *be*

<b>Corpus</b>	<i>is</i>	<i>are</i>	<b>present sg %</b>	<i>was</i>	<i>were</i>	<b>past sg %</b>
CLMETEV	0	1	0.0%	2	4	33.3%
BNC	6	2	75.0%	9	1	90.0%

Frequencies for Figure 4: *Tribe* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

<b>Collective noun</b>	<b>Level of stat sign / present</b>	<b>Level of stat sign / past</b>
<i>tribe</i>	$p \leq 0.067$	$p \leq 0.010$

Statistical significances for Figure 4: Levels of statistical significance for the changes in the singular percentage of *tribe* followed by the forms of the verb *be*

<b>Collective noun</b>	CLMETEV singular	CLMETEV plural	CLMETEV sg %	BNC singular	BNC plural	BNC sg %
<i>family</i>	91	99	47.9%	885	583	60.3%
<i>crowd</i>	34	8	81.0%	131	51	72.0%
<i>gang</i>	3	3	50.0%	26	19	57.8%

Frequencies for Figure 5: The changes in the frequencies of the ‘variable’ collectives followed by singular and plural verbs in the CLMETEV and in the BNC

Collective noun	Level of statistical significance
<i>family</i>	$p \leq 0.001$
<i>crowd</i>	$p \leq 0.118$
<i>gang</i>	$p \leq 0.360$

Statistical significances for Figure 5: Levels of statistical significance for the changes in the singular percentage between the CLMETEV and the BNC

Corpus	<i>is</i>	<i>are</i>	present sg %	<i>was</i>	<i>were</i>	past sg %
CLMETEV	33	24	57.9%	33	53	38.4%
BNC	281	110	71.9%	276	240	53.5%

Frequencies for Figure 6: *Family* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

Collective noun	Level of stat sign / present	Level of stat sign / past
<i>family</i>	$p \leq 0.016$	$p \leq 0.005$

Statistical significances for Figure 6: Levels of statistical significance for the changes in the singular percentage of *family* followed by the forms of the verb *be*

Corpus	<i>is</i>	<i>are</i>	present sg %	<i>was</i>	<i>were</i>	past sg %
CLMETEV	4	0	100.0%	23	7	76.7%
BNC	24	13	64.9%	75	27	73.5%

Frequencies for Figure 7: *Crowd* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

Collective noun	Level of stat sign / present	Level of stat sign / past
<i>crowd</i>	$p \leq 0.077$	$p \leq 0.370$

Statistical significances for Figure 7: Levels of statistical significance for the changes in the singular percentage of *crowd* followed by the forms of the verb *be*

<b>Corpus</b>	<i>is</i>	<i>are</i>	<b>present sg %</b>	<i>was</i>	<i>were</i>	<b>past sg %</b>
CLMETEV	2	0	100.0%	1	1	50.0%
BNC	9	0	100.0%	7	9	43.8%

Frequencies for Figure 8: *Gang* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

<b>Collective noun</b>	<b>Level of stat sign / present</b>	<b>Level of stat sign / past</b>
<i>gang</i>	-	$p \leq 0.437$

Statistical significances for Figure 8: Levels of statistical significance for the changes in the singular percentage of *gang* followed by the forms of the verb *be*

<b>Collective noun</b>	CLMETEV singular	CLMETEV plural	CLMETEV sg %	BNC singular	BNC plural	BNC sg %
<i>staff</i>	0	2	0.0%	73	1,306	5.3%
<i>crew</i>	9	19	32.1%	78	201	28.0%
<i>clergy</i>	3	24	11.1%	1	82	1.2%

Frequencies for Figure 9: The changes in the frequencies of the 'mostly plural' collectives followed by singular and plural verbs in the CLMETEV and in the BNC

<b>Collective noun</b>	<b>Level of statistical significance</b>
<i>staff</i>	$p \leq 0.371$
<i>crew</i>	$p \leq 0.335$
<i>clergy</i>	$p \leq 0.009$

Statistical significances for Figure 9: Levels of statistical significance for the changes in the singular percentage between the CLMETEV and the BNC

<b>Corpus</b>	singular	plural	<b>sg %</b>
LOB	3	5	37.5%
FLOB	1	9	10.0%

Frequencies for Figure 10: The changes in the frequencies of *staff* followed by singular and plural verbs in the LOB and in the FLOB

<b>Collective noun</b>	<b>Level of statistical significance</b>
<i>staff</i>	$p \leq 0.452$

Statistical significances for Figure 10: Level of statistical significance for the change in the singular percentage between the LOB and the FLOB

<b>Corpus</b>	<i>is</i>	<i>are</i>	<b>present sg %</b>	<i>was</i>	<i>were</i>	<b>past sg %</b>
CLMETEV	0	0	-	0	0	-
BNC	24	378	6.0%	26	261	9.1%

Frequencies for Figure 11: *Staff* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

<b>Collective noun</b>	<b>Level of stat sign / present</b>	<b>Level of stat sign / past</b>
<i>staff</i>	-	-

Statistical significances for Figure 11: Levels of statistical significance for the changes in the singular percentage of *staff* followed by the forms of the verb *be*

<b>Corpus</b>	<i>is</i>	<i>are</i>	<b>present sg %</b>	<i>was</i>	<i>were</i>	<b>past sg %</b>
CLMETEV	2	2	50.0%	5	16	23.8%
BNC	16	54	22.9%	32	95	25.2%

Frequencies for Figure 12: *Crew* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

<b>Collective noun</b>	<b>Level of stat sign / present</b>	<b>Level of stat sign / past</b>
<i>crew</i>	$p \leq 0.110$	$p \leq 0.449$

Statistical significances for Figure 12: Levels of statistical significance for the changes in the singular percentage of *crew* followed by the forms of the verb *be*

<b>Corpus</b>	<i>is</i>	<i>are</i>	<b>present sg %</b>	<i>was</i>	<i>were</i>	<b>past sg %</b>
CLMETEV	0	5	0.0%	0	10	0.0%
BNC	0	10	0.0%	0	48	0.0%

Frequencies for Figure 13: *Clergy* followed by the present and past tense forms of the verb *be* in the CLMETEV and in the BNC

<b>Collective noun</b>	<b>Level of stat sign / present</b>	<b>Level of stat sign / past</b>
<i>clergy</i>	-	-

Statistical significances for Figure 13: Levels of statistical significance for the changes in the singular percentage of *clergy* followed by the forms of the verb *be*