

UNIVERSITY OF CAPE COAST

TOWARDS A PROBABILISTIC GRAMMAR OF COLLECTIVE NOUN  
AGREEMENT IN BRITISH ENGLISH



ENOCK APPIAH TIEKU

2020

UNIVERSITY OF CAPE COAST

TOWARDS A PROBABILISTIC GRAMMAR OF COLLECTIVE NOUN

AGREEMENT IN BRITISH ENGLISH

BY

ENOCK APPIAH TIEKU

Thesis submitted to the Department of English of the Faculty of Arts, College  
of Humanities and Legal Studies, University of Cape Coast, in partial  
fulfilment of the requirements for the award of Master of Philosophy Degree  
in English

NOVEMBER 2020

## DECLARATION

### Candidate's Declaration

I hereby declare that this thesis is the result of my own original work and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature: ..... Date: .....

Name: Enock Appiah Tiekou

### Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature: ..... Date: .....

Name: Dr Joseph Arko

Co-supervisor's Signature: ..... Date: .....

Name: Dr Richmond S. Ngula

## ABSTRACT

This thesis examines collective noun agreement, using the notion of probabilistic grammar. Most accounts on the phenomenon appears not to have been fully explained the factors that determine collective noun agreement in English suggesting the existence of multiple factors that inform collective noun agreement relations. The most established explanation (that plural agreement is the preferred agreement in cases that the individuals in the collection are the focus and singular when the collection is the referent) is taken as circular reasoning and as the gap. And the study departs from this line of reasoning by examining other nuance factors that impact collective noun agreement. The study used the corpus linguistic methodology to study 6,467 clauses of singular and plural collective noun agreement forms from the British National Corpus. The data has revealed accessibility, humanness, definiteness, distance, part-whole and number features of predicates as variously informing agreement choices. These factors translate into two-level predictors which inform singular and plural agreement at relative significance and confidence levels. The results show that there is a higher tendency for singular agreement choice when the CNs are conceived as ‘whole’. Plural agreement is significantly determined by human of humanness and part of part-whole constraints. The observation that the predictors were sensitive to specific CNs suggests that CNs differ somewhat from each other. The results also show that it is easier to predict plural agreement than singular agreement and plural agreement choices are often primed. Further studies are encouraged to examine the impact of the constraint herein identified as determining collective noun agreement in other varieties of English.

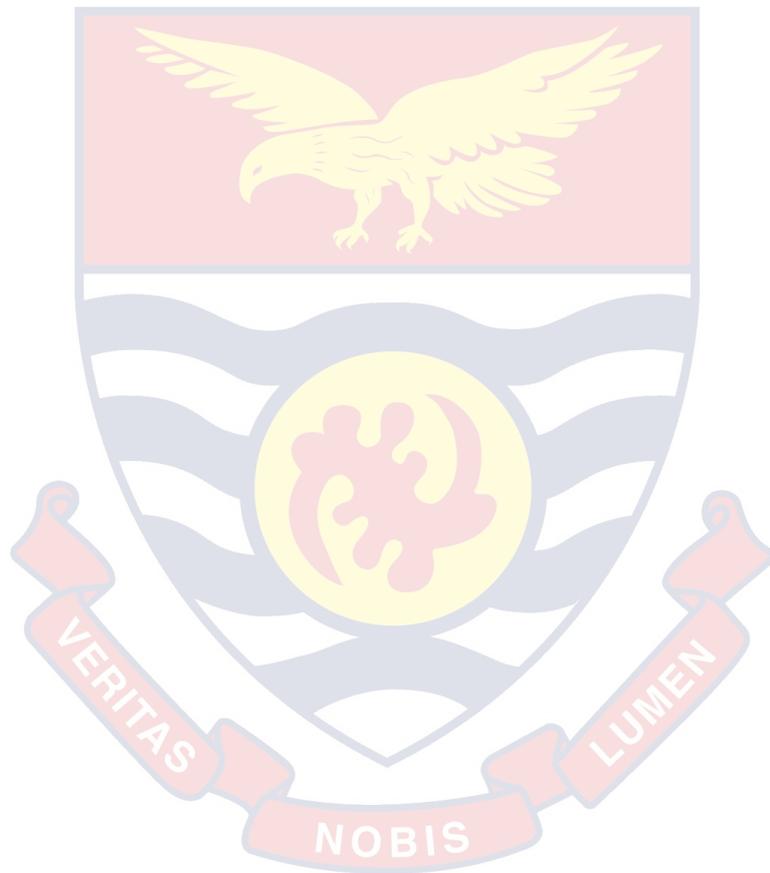
## ACKNOWLEDGEMENTS

I appreciate the supervisory role played by Dr Joseph Arko and Dr Richmond S. Ngula who have mentored and led me to certain truths and processes required for success in the academy. From Dr Arko, I have gained knowledge, skills and interest in theoretical linguistics. The training I received while I assisted you in teaching Semantics and Grammar has impacted my professional skills. I would not have heard of “Corpus Linguistics” nor accessed the British National Corpus but for Dr Ngula who used this period as my initiation ceremony into Corpus Linguistics. Dr Isaac Mwinlaaru has also served as the most needed third mentor. I owe a lot of the ideas in this text to his insightful suggestions and recommendations. Also, Isaac did not only introduce me to functional linguistics through analyzing texts during language classes I assisted him in teaching, I developed relevant research skills, especially text analysis skills. I very well appreciate you all for your contribution. If ever there is an academic family to boast of, I surely will boast of you.

I must appreciate the priceless services in the forms of proofreading and editing offered me by Louisa Osafo Adjei, Elijah Alimsewen, and Isaac Afful – all at the University of Ghana. I equally owe a lot to my mates and great friends. Alex Ohemeng and Kojo Oduro Ofori have been companions whose encouragements kept us going. Also, I thank my special coursemate, Emerald Tawiah and my friend, Mr Kwakye for their support. Finally, I appreciate the students whose thought-provoking questions set the desire to examine collective noun agreement. My siblings, Mark Oteng and Faustina Appiah, and my pastor, Rita Koranteng have been my emotional, and spiritual supports, always believing in me – we did it!

## DEDICATION

In memory of my father, Matthew K. Appiah

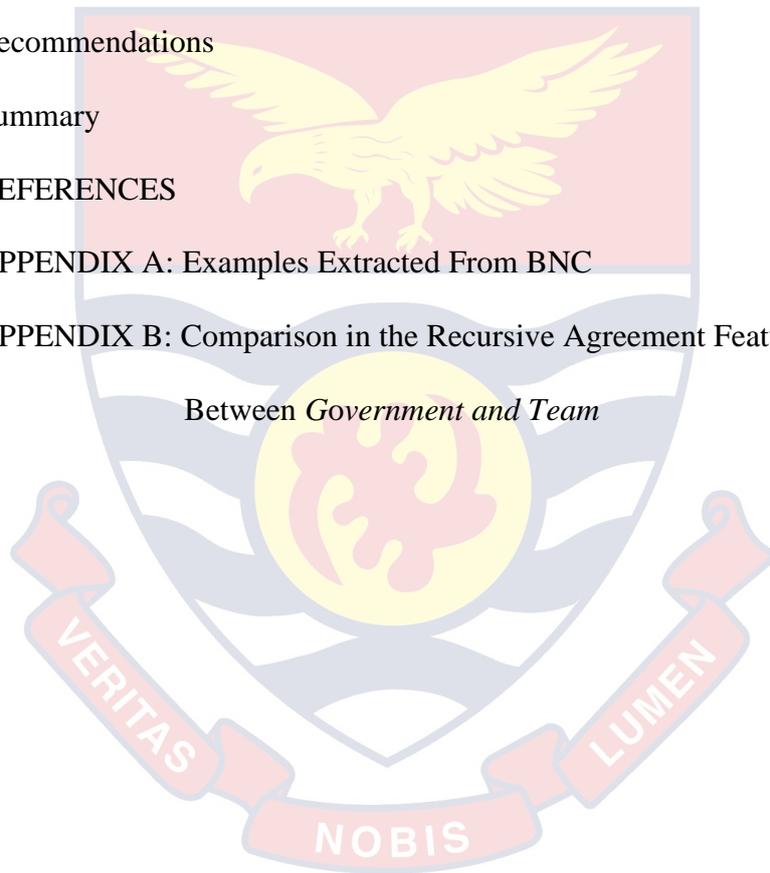


## TABLE OF CONTENTS

Content	Page
DECLARATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
DEDICATION	v
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	xi
CHAPTER ONE: INTRODUCTION	
Overview	1
Background of the Study	1
Statement of the Problem	8
Purpose of the Study	12
Research Questions	12
Motivation for the Study	12
Significance of the Study	13
Delimitation	15
Synopsis of the Study	16
CHAPTER TWO: LITERATURE REVIEW	17
Overview	17
Conceptual Review	17
Collective Nouns (CNs)	17
Agreement	21
Types of Agreement	22

Syntactic Agreement	22
Semantic Agreement	27
Pragmatic Agreement	29
The Pragmatics of Possessives	31
Pragmatics of Determiners	35
Government vs Agreement	37
Theoretical Review	41
Probabilistic Linguistics	42
Empirical Review	52
Summary	62
<b>CHAPTER THREE: METHODOLOGY</b>	
Overview	63
Corpus Linguistics (CL)	63
Data	68
British National Corpus	68
Collective Noun Sample	69
Data Analysis Procedure	72
Logistic Regression	73
Chapter Summary	75
<b>CHAPTER FOUR: ANALYSES AND DISCUSSION</b>	
Overview	76
Distribution of CNAgre in BNC	76
Discussion	81
Factors Determining Collective Noun <i>Agreement</i>	82
Number Features of Predicates	99

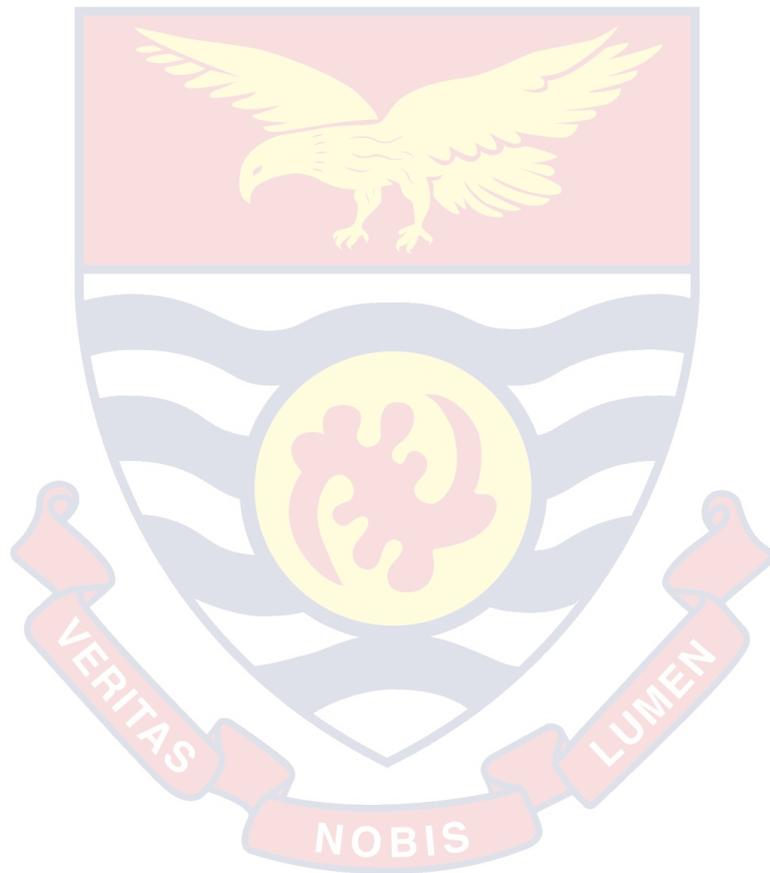
Humanness	99
CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATION	117
Overview	117
Summary of the Study	117
Conclusions	121
Implications of the Study	122
Recommendations	124
Summary	125
REFERENCES	126
APPENDIX A: Examples Extracted From BNC	147
APPENDIX B: Comparison in the Recursive Agreement Feature Choice <i>Between Government and Team</i>	150



## LIST OF TABLES

<b>Table</b>	<b>Page</b>
1 Functions of Proper Noun Modifiers	34
2 Differences between Agreement and Government	39
3 Frequently Studied CNs	70
4 Hits of CNs in BNC	72
5 Extract of CNAgre in BNC	73
6 Frequency & Percentage Distribution of CNAgre for the CNs in BNC	76
7 Distribution of Texts and Hits for Selected CNs in BNC	77
8 Distribution of plural agreement relations between ‘team’ and ‘have’	80
9 Distribution of plural agreement relations between ‘government’ and ‘have’	80
10 Variables that determine CNAgre	104
11 Coded Results of Predictors of CNAgre in the 6467 Clauses	107
12 Logistic Regression Analysis of Strength of Predictors of Agreement Relations for <i>Committee</i> : Full Model with the Predictor Variables, Beta Co-efficient, Standard Error, P-Value and the Odds Ratio	107
13 Logistic Regression Analysis of Strength of Predictors of <i>Agreement</i> Relations for <i>Family</i> : Full Model with the Predictor Variables, Beta Co-efficient, Standard Error, P-Value and the Odds Ratio	111
14 Logistic Regression Analysis of Strength of Predictors of Agreement Relations for <i>Government</i> : Full Model with the Predictor Variables, Beta Co-efficient, Standard Error, P-Value and the Odds Ratio	109

- 15 Logistic Regression Analysis of Strength of Predictors of Agreement  
Relations for *Team*: Full Model with the Predictor Variables,  
Beta Co-efficient, Standard Error, P-Value and the Odds Ratio 107
- 16 Logistic Regression Analysis of Strength of Predictors of *Agreement*  
Relations for *Party*: Full Model with the Predictor Variables,  
Beta Co-efficient, Standard Error, P-Value and the Odds Ratio 112



## LIST OF FIGURES

Figure		Page
1	Syntactic and Semantic Agreement	28



## CHAPTER ONE

### INTRODUCTION

#### Overview

The present study examines collective noun agreement.<sup>1</sup> It is conducted based on what is perceived as the need for a detailed and explicit explanation of agreement relations that involve collective noun headed NPs in English. This introductory chapter provides a context and direction for the study. The chapter discusses the background, motivation for the study, the purpose of the study and the problem that the study hopes to solve. Research questions are also provided. The chapter proceeds to state the significance of the study, delimitation, and a synopsis.

#### Background of the Study

Agreement relations with collective noun (CN) headed NPs in English has received much scholarly attention (e.g. de Vries, forthcoming; Fernández-Pena, 2016; Örlegård, 2014; Annala, 2008; Hundt, 2006; Levin, 1998; Algeo, 1988). In the scholarship, different factors have been identified to explain Agreement Relations with CN Headed NPs (CNAgre hereafter).<sup>2</sup> The explanations that have been given can be seen to come from two key sources. The first set of explanations comes from reference grammars (e.g. Downing,

---

<sup>1</sup> “Collective noun” as used in this thesis does not refer to the very technical sense of ‘collective’ as used in model semantic theory. Until otherwise stated, ‘collective noun phrases’ in this context is similar to “group NPs” as used by model semantic theorists (like de Vries, 2015) and “committee-type” nouns by Generative Grammarians (like Smith, 2020; 2017a). The source of the conflict will be resolved in chapter 2.

<sup>2</sup> Agreement, though mostly used synonymously with ‘concord’, is considered a different phenomenon in some linguistic contexts. Bloomfield (cited in Corbett, 2006) mentioned that agreement is a super-ordinate term of three other terms – *concord/congruence*, *government*, and *cross-inadequate explanation reference*. Greenberg (1978b) however considers concord as the broader term. In this work, the term agreement shall be used without any alignment to these schools of thought. It is used to refer to a relationship between the abstract number features of CNP subjects and finite VPs.

2015; Quirk, Greenbaum, Leech, & Svartvik, 1985; Eastwood, 1994) and the second set of explanations has been based on findings from empirical studies (e.g., Fernández-Pena, 2015a, b; Örlegård, 2014; Pearson, 2011; Levin, 2001).

In the reference grammars (e.g. Downing, 2015; Greenbaum, 1996; Eastwood, 1994; Quirk et al, 1985), two basic explanations are identified. The first is semantic and the second sociolinguistic. In terms of the semantic explanation, these grammars claim that plural agreement is the preferred CNAgre when the constituents of the CN are referred to and the singular when the single collective body is referred to. Since semantics is considered to account for the perception of CNs as a collection of individuals or a single unit, the grammars have generalized that semantics account for variable CNAgre.<sup>3</sup>

Scholars who closely studied number agreement relations of CNs found weaknesses in the semantic explanation. One way to identify the reference to individual members but not the group as a whole has been instances where body parts (like feet, eyes, arms) of the constituents of the collection are mentioned. Logically, the plural agreement is preferred in such cases since the individual members of the group are considered (Depraetere, 2003).<sup>4</sup> Plural “bodily parts” do not always require plural agreement (Levin, 2001). Sometimes, the ‘metaphorical’ reference to body parts may constrain the singular agreement choice which then contests the claim that mere reference to body parts determines plural CNAgre (Depraetere, 2003) and questions the established semantic explanation<sup>5</sup>. More so, distributivity explains instances where

---

<sup>3</sup>The semantic explanation is sometimes called the notional or individuatedness factor.

<sup>4</sup> Depraetere (2003, p. 118) exemplifies thus - *The crowd were on their feet again as substitute Murray Fraser picked up a quick throw-in and dived over after an 80-metre-run. (today/11)*

<sup>5</sup> Depraetere (2003) exemplifies thus: *But his audience was **all eyes and ears** as he rattles off wisecracks.*

predicates scope over individuals in the group though the singular verb is used - as in example (1) (Pearson, 2011, 163).

1. The Pearson family always has big feet.

Even as the singular agreement choice does not contest the interpretation that “for every generation [g] of the family, the members of [g] have big feet” (Pearson, 2011, 163), it becomes clearer that the lexical semantics of the CN does not fully explain CNAgre.

Also, the semantics of VPs confirmed the claim that semantics determine CNAgre. On the one hand, verbs that designate states or activities performed by referents of CNs were identified as determining both singular agreement choices. Those that designate activities, on the other hand, often foreground the individuals actors thereby informing plural agreement choices. Yet there are instances where activities performed by individuals is referenced but the singular verb is used (Annala, 2008).<sup>6</sup> This situation reveals the difficulty involved in expounding how the group or individual members of the group are referenced. The difficulty has led some of these grammars to caution, especially non-native speakers of English, that one could always resort to the grammatical agreement (i.e. singular agreement) “when in doubt” (Yankson, 1994, 24).

These grammars, probably due to the weakness in the semantic explanation, add sociolinguistic factors as another determinant of CNAgre. Selecting plural or singular number agreement for CNs has been said to depend on the (standard) regional variety of English involved. This observation is

---

<sup>6</sup> Annala’s (2008) example, *Their staff is busy compiling lists of works of art*, indicates that regardless of the semantics of the interpretation of the verb phrase ‘is compiling’, the singular agreement can be used.

attributed to the claim that whereas the preference for singular and plural agreement for CNs is distinctive of British English (BrE), the singular agreement is the typical agreement form in the American variety of English (Biber et al., 1999). Insofar as all varieties of English allow both plural and singular CNAgre – though at relatively different degrees – variable CNAgre becomes a feature of English and its prevalence in any variety does not fully explain its use in that variety nor English (Smith, 2017a; Corbett, 2006). The regional variation factor, thus, requires nuance factors as a basis. From these two explanations, agreement relations involving CNs seem not to have been fully explained by the reference grammars.

The second set of explanations for CNAgre is from researchers who have studied CNAgre. Generally, four factors are discussed: semantic, pragmatic, syntactic and sociolinguistic factors. In addition to the lexical semantic explanation, animacy has been identified as a determinant of CNAgre. CNs that consist of animate entities (mostly humans) typically select plural agreement than those whose constituents are inanimate (see de Vries, *fc*; and the literature therein; but also, Nixon, 1972). This explanation is supported by the observation that human and animate occupy the highest rank on the hierarchy of individuation and as such “are more easily accessible in the conceptualization of the whole than other collective wholes” (Gardelle, 2017, 8). When Gardelle (2019) and Persson (1989) concluded that it is difficult to select agreement forms for CNs whose constituents are animate, they put the problem with this semantic factor in perspective that the animacy factor only highlights the need to account for variable agreement relation of CNs if it motivates variable agreement (de Vries, *forthcoming*; Gardelle, 2019; Pearson,

2011).<sup>7</sup> Further research is still needed to shed light on how semantic factors determine variable number agreement relations between CNPs and finite verbs – both the individuatedness and animacy (which are semantic) factors appear questionable.

In terms of sociolinguistics, several explanations are put out by researchers. The explanations include

- young persons prefer singular agreement than adults (Bailey, 1987)
- context of use determines agreement (Fries, 1988; Hundt, 1998)<sup>8</sup>
- the medium of communication determines agreement (Levin, 2001)<sup>9</sup>
- regional variety of use and period of use constrains agreement (Lakaw, 2017; Pearson, 2011)
- the CN involved determines agreement relation (Annala, 2008; Levin, 2006).<sup>10</sup>

The effort to rely solely on variations in use as a means of explaining CNAgre looks aprioristic. Such explanations may neglect variety or language-internal semantic, syntactic and pragmatic factors (such as definiteness, accessibility, distance) which could provide basic explanations for why the variants are present in the language in the first place to serve as the basis for other sociolinguistic explanations that relate to gender, region, age, medium etc. In

---

<sup>7</sup> Persson (1989) shows that it is obvious that a majority of the nouns typically classified as CNs are animate and these are those whose agreement relations appear not to have been accounted for.

<sup>8</sup> For example, between officialese and journalesse use of language, Fries (1988) found singular as the preferred agreement form for journalists.

<sup>9</sup> Written medium aligns with singular and spoken with plural; formal contexts prefer singular and informal, variable (Levin, 2001).

<sup>10</sup> Annala (2008) was overwhelmed by her findings that CNs such as *staff*, *crew*, and *clergy* have had an increasing preference for singular agreement just as others (like *government*, *army* and *tribe*) had an increasing preference for plural agreement (Annala, 2008). CNs such as *Family* and *team* vary in agreement choices; *government* dominantly selects singular agreement while *police* is normally followed by plural verbs (Levin, 2006, p. 322).

other words, unless these studies share in the notional explanation, their lists of frequencies of use in terms of age, region, context, etc. might not be taken as an apt explanation of why variable agreement exists in the first place. Perhaps, they need to understand that "it is one thing to describe this difference [e.g. between BrE and AmE], but quite another to explain it" (Levin, 1998, 23).

Syntax has been identified as a determinant of CNAgre. While some premodifiers (e.g. *whole, a, one, every, this*) have been identified as fixing singular agreement relations, others (*several, both, every*) are said to constrain plural agreement (Depraetere, 2003; Levin, 1998; Poustma, 1904). Nevertheless, the same premodifiers identified as typically selecting singular number features for CNs and determining singular CNAgre also trigger plural agreement (Depraetere, 2003). For instance, in the earlier accounts of how determiners and premodifiers inform CNAgre, Dekeyser (1975) found *each* to determine singular CNAgre but Fowler (1992) contests that it selects both singular and plural agreement. Perhaps, the challenge is not to identify premodifiers that a particular number agreement feature is typically associated with but to explain how these premodifiers point to either plural or singular number features of CNs. After such an explanation, one would understand that though these premodifiers could be involved in variable agreement relations, they may prefer one agreement feature, making it needful to establish the relative strength they possess in determining CNAgre. This has been the approach Fernández-Pena (2017b; 2016; 2015a, b) seems to use in determining how *of-dependent* postmodifiers determine CNAgre.<sup>11</sup> As Fernández-Pena

---

<sup>11</sup> For example, in a group *of boys* and a set *of paints* etc., the italicised are *of dependent* PPs (Fernández-Pena, 2017b).

(2017b) identifies the number feature of the oblique object; the distance between the head of the CNP and finite verb; and the animacy of the oblique object as determining CNAgre, she succeeds not only in pointing to syntactic factors that determine CNAgre,<sup>12</sup> but explaining how such factors determine CNAgre.<sup>13</sup> Upon identifying syntax as a determinant of CNAgre, one needs to provide some interpretations of the various syntactic units and how they trigger number features. By far, it is only Fernández-Pena and Levin (2001) who have made this attempt.<sup>14</sup>

The efforts to provide pragmatic accounts of CNAgre seemed promising. Speech acts of *deciding, hoping or wanting* have, for example, been identified as licensing plural CNAgre (Swan, 2005, 519). Aside from a minimal consideration of the pragmatics of the VP as a determinant of CNAgre, the pragmatic account, especially the pragmatic features of premodifiers and postmodifiers that identify CNs (the number features of CNs and also fixing CNAgre), has not been examined.<sup>15</sup>

The inability of the two sources, reference grammars and empirical studies, to add nuance explanations to the general explanation of CNAgre makes it necessary that the phenomenon be examined in greater detail. As can be seen from the ongoing, though there appear to be a plethora of explanations on agreement relations with CN headed NPs, most of the explanations seem less

---

<sup>12</sup> Plural CNAgre is fixed when the *of-dependent* structure brings distance between the head of the CNP and the verb; when the oblique object is animate and when the oblique object is plural.

<sup>13</sup> This is not to say that Fernández-Pena's findings are overly explanatory of CNAgre. Her perception that postnominal possessives only construe relative distance which is measured by the number of words is a narrow view of how *of-dependent* postmodifiers could input CNAgre. As an example, she is silent on other semantic relations, such as constitutivity, epithetical relations etc. that are involved in the use of such *of-dependent* postmodifiers.

<sup>14</sup> For *of-dependent* structures, see Fernández-Pena (2016)

<sup>15</sup> The most notable studies of the pragmatic features of the NP have been comments on Nunberg's (1997) classic *The French fries is getting impatient* illustration. None has looked at the pragmatics of modifiers of the CN.

adequate in providing all the factors that determine collective noun agreement and predicting with a relatively higher degree of certainty, the variables that constrain (either singular or plural) collective noun agreement. The question – how are agreement relations between CNs and finite verbs determined? – does not seem to have a straight forward answer. This study seeks to provide some answers to such a question.

Hypothesizing that a multiple of factors determine CNAgre is essential and demands a view at the phenomenon from a perspective that perceives language as comprising a multiple of choices from which individuals pick. A study of huge samples of use could help identify some of the factors that inform the choices and predict which factor has a crucial effect on determining the choices. The probabilistic linguistic (PL) theory (Bod, Hay, Jannedy, 2003) perceives language this way and specifically establishes that a linguistic study is to provide a probabilistic grammar of the phenomenon under-study. Using PL as an underlying theory and the British National Corpus as data, the present study hopes to provide a probabilistic grammar of CNAgre.

### **Statement of the Problem**

The tendency to choose either singular and plural agreement in agreement relations that involve collective noun headed NPs as subjects has made CNAgre a linguistic phenomenon that is difficult to explain. Till date, it is difficult to explain what influences singular or plural agreement choices for CNs. Besides, studies on CNAgre have primarily focused on variation according to use, user, medium, style – that is, sociolinguistic factors to account for CNAgre (including Fernández-Pena, 2016; Hundt, 2006; Levin, 2001; Mcknight, 1925; Hundt, 1998; Trudgill & Hannah, 1994; Algeo, 1988;

Johansson, 1979; Dekeyser, 1975). Though the value of sociolinguistics in explaining phenomena cannot be disputed or understated, when sociolinguistic explanations precede accounts that explain language-internal features that explicate why the variants of a linguistic phenomenon exist in the language in the first place, much is unaccounted for (Geeraerts, Kristiansen, Peirsman, 2010). That is, to give a sociolinguistic explanation of variable agreement, one may have to first explain why the variants are allowed in English; what explanation informs the choice of any of the variants before expounding why gender, age, and geography distinguish choice of any agreement form. Therefore, the presence but minimal use of any number agreement features for CNs by a certain category of people (professionals, adults), for example, does not annul the fact that variable number features exist neither does it explain why it exists; it only shows what one group prefers, which is an “unsatisfactory” proposal because it does not explicitly explain what informs the choice of any of the number features (Pearson, 2011, 191). Following Smith (2017a), there is the need to explain how variable agreement is at all allowed in English.

Further, the need for comprehensive explanation for CNAgre has been expressed (Kairis, 2017; Fernández-Pena, 2015a,b).<sup>16</sup> The semantic explanations (like those provided by Smith, 2017a; Levin, 2001; Quirk et al., 1985; Nixon, 1972; Jespersen, 1961) do not fully explicate how speakers indicate individual or group referents in their use of CNs; neither do the animacy and semantic features of verb factors adequately explicate variable CNAgre. Alternatively, though some insightful pragmatic and syntactic factors have been

---

<sup>16</sup> Fernández-Pena (2015a, b) mentions the need for more investigation to carry out a more fine-grained analysis of the implications of overt vs. non-overt morphology as well as to **consider further variables**

provided, they do not fully explain CNAgreement: aside from identifying several premodifiers and postmodifiers as licensing CNAgreement (Depraetere, 2003; Poustma, 1994), Fernández-Pena (2017b) has partially provided some interpretations of how *of-dependent* postmodifiers determine CNAgreement; explanations on how premodifiers assign plural or singular interpretations to CNs and constrain CNAgreement could add to the solutions to the CNAgreement problem. Also, the only pragmatic explanation for CNAgreement is that offered on how the pragmatics of verbs determine CNAgreement (Swan, 2005). Consequently, it would not only “be of great interest to further investigate the use of collective nouns in complex noun phrases taking into consideration different factors such as syntax, semantics and pragmatics” (Karis, 2017, 30), but it will be highly useful that the pragmatics, semantics, and syntactic explanations of all grammatical items used in structures that overtly display CNAgreement are studied.

With the above in mind, one can guess that several factors determine CNAgreement. To fully account for CNAgreement, therefore, it would be appropriate to evaluate the determinants of CNAgreement for their relative strengths. In other words, since it is obvious that a multiple of factors influences CNAgreement, all the predictor variables need to be identified. This is essential because one needs to consider the gradient nature of language and understand that linguistics need not necessarily be a report on what is ideal but must first account for all possible options and thereafter determine the ideal categorial predictor or evaluate the strength of the predictors (Bresnan & Ford, 2010; Bresnan, 2007; Gries, 2003). Such a consideration will help establish the strength of the many factors that determine either singular or plural agreement.

It is difficult to entirely nullify the argument that the ideal form of the grammar of any language is the reflection of native speakers' knowledge, the systems of the language which is shaped by context. It is equally true that the use of language is constantly modified by the context of use. Halliday's (1991) observation is important here. He argued that native speakers can through intuition guess that *active* is preferred over *passive* structures; *positive* over *negative* forms, *declarative* over *interrogatives*; *this* over *that*; but will be uncertain about the relative frequencies of singular and plural forms. This shows that "some systems tend towards being equiprobable and hence have a higher information value while others are notably skewed and thus display greater degree of redundancy" (Halliday, 1991; 37). For a system like collective noun agreement whose choices are difficult to predict, he adds, studying huge samples of language use is highly useful. Studying dative alternation, Bresnan (2001) has underscored how useful data is since the phenomenon is highly gradient and intuitive studies overlook such gradient features. Also, the key proponent of lexical functional grammar, Beth Levin, as far as 1993, advised that data be used in describing very controversial features of a language. CNAgre is one such controversial features of English and intuitive studies (such as Smith, 2015; Sauerland & Elbourne, 2002; Den Dikken, 2001) have merely given an abstract view of the phenomenon but not adequately provided such native-speaker knowledge on CNAgre.<sup>17</sup> It is needful that usage data be studied to explain the phenomenon. On the whole, this study is an attempt to occupy the

---

<sup>17</sup> This is not to suggest that empirical studies have not been conducted but most of such studies have already been mentioned as suggested the need to identify nuanced factors that determine CNAgre.

niche created by the need for a comprehensive explanations of CNAgre to provide a probabilistic grammar of CNAgre.

### **Purpose of the Study**

Generally, this study contributes to the attempts that follow the need to explain CNAgre (Lakaw, 2017). The purpose of the study is captured in three specific statements of intent: (1) to show how the distribution of agreement forms for CNs provides explanations for choices in CNAgre; (2) to identify some nuance factors that determine CNAgre and (3) to define the relative strength of the factors that determine CNAgre.

### **Research Questions**

The following specific questions are posed:

1. How does the distribution of agreement forms between CNs and verbs in the BNC explain CNAgre?
2. What factors determine CNAgre?
3. What are the relative strengths of the factors that determine CNAgre?

### **Motivation for the Study**

The motivation for this research came in three folds. The first was an encounter with a student during a lecture on *Agreement*. This student was not convinced with the explanation that singular agreement choice is preferred when one considers the collective as a unit but plural when the individuals are considered. She questioned further as to why the noun – *family*, even though singular in form, mostly engage in plural agreement relations (specifically with its use in Ghana). Her inquiry drew my attention to the tendency for other factors to account for agreement. The second was a class discussion on *Subject Verb-*

*Agreement*. While I sat at the back of my class as Demonstrator, a student was asked to solve a subject-verb agreement problem: the subject of the sentence was a plural common noun and the verb was singular. Interestingly, the student changed the plural subject to singular. His answer was not accepted as the instructor and some other students agreed that the verb must agree with the subject but not the subject with the verb. The obvious question that came to mind was whether *Agreement* was a *government* type of relationship or not.

The final observation was a seminar presentation made by a member of staff in my department on parliamentary discourse. His study used corpora on British and Ghanaian parliamentary discourses. It was striking that whereas the British text had both plural and singular agreement relations with the noun *government*, the Ghanaian text almost always had singular agreement relations. These three observations, at varied times, called for readings of the literature on agreement and collective noun agreement but my inability to find a very convincing explanation specifically on *collective noun agreement relations* urged me to examine the issues in this thesis. I hope this work provides the answers my students would find useful.

### **Significance of the Study**

There are quite many areas where this study has relevance. Broadly, the study is relevant to both applied and theoretical linguistics. For applied linguists, especially those involved in teaching English, the findings of this study add to the explanation for CNAgre: they would be able to explain CNAgre instead of resting on inconclusive explanations, for example, that when in doubt one should resort to the grammatical agreement principle (Yankson, 1994).

The study also has some theoretical relevance. This study uses naturally occurring data to explain a linguistic phenomenon. By examining agreement through the use of such data, the study sheds light on discourse-oriented conceptualizations on agreement. This will add to Barlow's (1999) observation that discourse defines agreement relations. Moreover, the study shows that the context and co-texts are relevant resources to consider in determining the number features and agreement relations with collective nouns. This is relevant to the debate on semantics and pragmatics (Hawley, 2002).

Agreement is a controversial subject (Fernandez-Pena, 2017b). The presence of inconsistencies in agreement patterns makes it difficult to hold on to any factor that determines agreement. Noting that collective noun agreement has been one of the main areas of contention in *agreement*, the results of this study could address the controversies. Establishing the strength of the factors that determine collective noun agreement relations provides further evidence of the probabilistic nature of language. That is, the study shows that the choice of singular and plural agreement is informed by different factors and at different degrees.

Finally, researchers describing collective nouns often study them as a unified class (De Vries, fc; Fernández-Pena, 2017b; 2016; 2015a; 2015b; Levin, 2001). The few which sets out to distinguish them have done so through animacy – human, animate, inanimate (Gardelle, 2019) or their denotation – family: couple, family (Lakaw, 2017). Upon the following key observations from the results that:

- a. Different factors explain agreement for different CNs
- b. The texts in which the CNs are used differ

c. The construal of their grouping processes differs,

this study makes significant contribution to the conception of collective nouns. Since these observations were made by perceiving CNs from above (looking at the CN itself), below (exploring contextual features such as the context of use, text type, formality), and round-about (examining the co-texts such as premodifiers, determiners, postmodifiers that relate with the CN as ‘head’), the present study contributes a *cryptogrammar* perspective to the conception of CNs.<sup>18</sup>

### **Delimitation**

The analysis of agreement relations that involve CN headed constructions will consider only agreement relations between CNs, leaving agreement with other grammatical units like pronouns. The noun-pronoun agreement (and other agreement relations) is not the subject of this study because per the time allotted for this study and, more importantly, the in-depth analysis required, a focus on only one type of agreement will be helpful.

Verbal categories that overtly show agreement features are those that will be selected for analysis. Modal auxiliaries, non-finite verbs, imperative verbs, past tense forms of some lexical verbs, and subjunctive verbs do not mark agreement overtly (Biber et al. 1999:180) and will therefore not be selected for analysis. These types of verbs are not inflected for plurality. Also, number and person are marked overtly by present tense forms of verbs except the verb *be* whose past, *was/were*, show number agreement (Huddleston & Pullum 2002; Greenbaum, 1996).

---

<sup>18</sup> Below, above, and round-about are terms borrowed from Michael Halliday (Halliday, 1996). Cryptogrammar is used by Whorf to describe covert grammatical features (cf. Martin, 2016).

This study is limited to the analysis of clauses in the corpora where agreement is marked by the present tense form of regular verbs or finite forms of the verb *be*. Further, lexical verbs such as *set* and *cut* whose past and present forms are the same were taken out of the data. In brief, only constructs that overtly display agreement relations between subject CNPs and the finite verbs are selected for this study.

Also, certain syntactic structures that overtly display the number features of CNs were not considered. Quantifiers such as *all*, *both*, *few* were not selected because such quantifiers could occur as the head of the NP such that the CN becomes a postmodifying element.

### **Synopsis of the Study**

The study is organized into five chapters. This introductory chapter has established the need for and aims of the study. The second chapter is dedicated to the discussion of the main concepts underpinning the study: collective nouns, and agreement – the conceptual issues upon which this study is framed, and also to theoretical and empirical review. Chapter three describes the methodology. The corpus linguistic methods and the corpora used for the study will be discussed. Chapter four presents the analysis and discussion. Firstly, the distribution of CNAgre in the BNC is analysed and is followed by the identification and explanation of the predictor variables that determine CNAgre. The analysis and discussion finally establish the extent to which the identified factors determine agreement. The final chapter, Chapter Five, gives a summary of the entire study, a list of key findings and their implications, as well as recommendations for further studies.

## CHAPTER TWO

### LITERATURE REVIEW

#### Overview

Chapter Two is a review of related literature. The overriding aim of the chapter is to identify most of the scholarly inputs on key aspects of the topic, examine the inputs, and indicate how my study fits into existing arguments. To do this, this chapter is divided into three broad sections. The first is a conceptual review, discussing the concept “collective noun” and explaining various theory-shaped perspectives on *Agreement*. Section two discusses the Probabilistic Linguistic (PL) theory as the theoretical perspective of the study. The final section reviews empirical studies on CNAgre.

#### Conceptual Review

##### Collective Nouns (CNs)

The most defining feature of collective nouns in English has been their variable agreement relations emanating from notional plural and singular, and formal singular nature of CNs (de Vries, fc; Joosten, 2010; Levin, 2001). Defining collective nouns by their agreement relations will be a pre-theoretic attempt to explain CNAgre, considers only nouns in the class of CNs that engage in variable agreement (de Vries, fc; Gardelle, 2019), and disregard CNs in languages that do not mark agreement (Joosten, 2010; Joosten et al, 2007). The study departs from this perspective because its objective, which is to identify nuance factors that inform agreement relations, require a conception of CNs independent of their agreement relations. I follow the view that CNs are nouns with plural meanings at lexical level but singular in form (de Vries, fc;

Gardelle, 2019; Acquaviva, 2016) is informing.<sup>19</sup> From such a perspective, the following emerge as acid tests for class membership of CNs:

1. A CN requires plural noun complement when it takes an *of* phrase postmodifier (i.e., Det + N1 + of + NPs) (Barker, 1992). For instance, *the committee of women* but not *\*the committee of woman* is the typical postmodification structure for CNs.
2. The plurality feature exists at the lexical level of meaning (Gardelle, 2019). Plurality is part of the lexical content of CNs (e.g. *team*), differentiating CNs from singular count nouns which have no semantic plurality at the lexical level (e.g. *player*) and plural count nouns whose plurality is marked morphosyntactically at the lexical level (e.g. *players*).
3. Non-additivity: the referent of a CN is beyond the sum of the entities that form it (Gil, 1996; Barker, 1992). “A team is not just a collection of team members; it’s an entity in itself, with its own internal structure, its own complex way of functioning and its own independent goals” (de Vries, fc, 5). The entities that form CNs are constituted as parts (a *player* is a part of *team*) which is meronymic but not kinds or types (*chairs* and *tables* are kinds of things that form *furniture*) which result in taxonomic relations (Gardelle, 2019).
4. CNs require singular DP-internal agreement (*this team* but not *\*these team*) (Smith, 2017b; Birkenes & Sommer, 2014; Elbourne, 1999). This

---

<sup>19</sup> The term “collective” presents several terminological and theoretical debates. We do not pursue the debate here but refer the reader to Joosten et al. (2007), Gardelle (2019) and Acquaviva (2016).

test rejects nouns which denote lexical level plurality, such as cattle and people (these cattle, these people).

5. A CN disallows plural agreement in existential constructions (e.g. “**There is/\*are** a committee deciding the budget for next year” - Smith, 2017b; Elbourne, 1999).

Tests 2 and 4 suggest that the plural number feature of CNs cannot be explained in grammatical terms. This highlights the challenge of identifying how the number features of CNs are referenced, pointing to the need to examine semantic and pragmatic explanations. Joosten (2010) and Gardelle (2019), studying the qualities adjectives denote in their relation with CNs, have demonstrated that the number features of some CNs are more accessible than others. Explaining that CNs that display higher levels of opacity typically reference singular entities, Gardelle (2019) opines that CNs in human category favour both plural than those in animate and inanimate categories. Their success in using the meaning of adjectives to identify number features of CNs is the basis for my conclusion that a systematic semantic and pragmatic study of linguistic resources in the CN headed NP would reward one with the ways of accessing number features of CNs. Levin (2001) did not find determiners important in identifying number features of CNs because his interest was not in their meaning and how the meanings guide reference.

I appeal to the semantics and pragmatics of demonstrative determiners, possessives, and postmodifiers as anchoring references to CNs, and their ‘covert’ number features. To borrow a term from Micheal Haliday, this would mean that the noun is studied from “round about” – its premodifiers and postmodifiers.

Another way of identifying the number features will be to look at CNs from above. That is, to consider the linguistic and social context of the collective noun. Here, the observation that plural reference is a common feature of spoken texts while singular reference is common in written texts (Levin, 2008) needs to be highlighted. No known study, to the best of my knowledge, has examined the co-text sense of context. Here, linguistic resources that have co-referential functions with collective nouns could help determine their number features. The findings that the medium of communication and level of formality determine CNAgre (Levin, 2001) has been an attempt to look at CNAgre from above, the context of use. But context also includes co-texts – even including clauses before the CNAgre construct.

Lastly, CNs need to be studied by themselves. That is, there must be an attempt to describe the CNs by their lexical semantics features. The efforts in the literature have been to group them based on their animacy features (Gardelle, 2019 and the literature therein), their meanings Lakaw (2017). It will be useful to pay close attention to each noun first, before any classification. In this study, I analyze each CN in its own right, with the understanding that each has unique lexical features.

In this first subsection of the literature review, I have reviewed the widely subscribed definition of collective nouns, identified the weaknesses in the definition – showing a gap in the conception of CNs. It was further shown that Joosten (2010) and Gardelle (2019) have made efforts to fill the gap. I have added to their debate by showing that plural number is a covert feature that can be assessed by studying CNs from below, above and round-about. In the next section, I discuss agreement. The objective of the discussion is to show that

agreement is not just grammatical or semantic but also a discourse phenomenon. Following the understanding that agreement is pragmatic and that the number features are accessible from round-about the CN, I will discuss the pragmatics of determiners and possessives.

### *Agreement*

Several views have been shared on *agreement*. These views are typically shaped by the various grammar schools, conceptions and/or theories because agreement has become a key feature to test syntactic theories (Corbett, 2006). The section first provides a general definition of agreement then a discussion of the types of agreement before a working definition of agreement is provided.

Steele's (1978) classic definition summarizes views on agreement and also details what has sparked several arguments on what constitutes agreement. He defines agreement as a "systematic covariance between a semantic or formal property of one element and a formal property of another" (Steele, 1978, 610). 'Property' in his definition is today seen as the abstract number (singular, plural, dual etc.), gender (masculine, feminine, neuter) features which could be singular, plural, dual, etc; masculine, feminine, neuter (Corbett, 2006). Relating to his claim that agreement could be accounted for semantically or formally, whether agreement is triggered by semantic/pragmatic features or formal features is still a relevant question.

For generative grammars, agreement forms part of the valuation processes through which abstract features, such as number and case, are assigned to linguistic units (see Smith, Mursell, & Hartmann, 2020 and the

literature therein).<sup>20</sup> This viewpoint reveals the formal approach to agreement. Other scholars believe agreement is purely semantic (Dowty & Jacobson, 1988; Lapointe, 1980), and yet a great number of scholars indicate that agreement is semantic/pragmatic and formal (Corbett, 2006; Pollard & Sag, 1994, 1991; Kathol, 1991). *Agreement* has been variously classified. *Index* and *Concord* are identified types by Pollard and Sag (1994), *Canonical* and *Non-canonical agreement* by Corbett (2006). The traditional types of semantic, syntactic and pragmatic agreement types (Baker, 2008; Corbett, 1991, 1983, 1979; Quirk et al., 1985) will be used because the various classifications share similarities with the traditional groupings. For instance, Corbett's canonicity is a scalar representation of the syntactic, semantic, and pragmatic types by which he perceives canonical agreement to instances where syntactic and semantic features coincide while those which are less canonical would be those determined either by semantic features or syntactic features. Pragmatic agreement maps his non-canonical type of agreement.

### **Types of Agreement**

Since the types of agreement - syntactic, semantic, and pragmatic agreement - serve as a basis for all classifications, they will be discussed in this section.

### **Syntactic Agreement**

Syntactic agreement describes agreement relations where the forms of the items that relate in agreement consistently vary (Baker, 2008; Corbett 2006). That is, the form of the unit that controls agreement is token identical with the

form of the target.<sup>21</sup> For instance, the formal features of a noun phrase, such as the plural maker *\_s* (as in *boys*) relate to a verb in syntactic agreement if the verb has morphologically plural features (seen in the dropping of the singular marking morpheme – *s* for regular verbs). In example (2), the number features of the subject NP, *Kwakyé* is singular and it identifies with the singular verb *has*.

2. *Kwakyé* has finished the assignment.

It is not only the token identical syntactic features that mark syntactic agreement but Corbett (2019; 2006) observes that the semantic properties of the agreement units coincide in such relations. In example (2), *Kwakyé* is semantically a singular noun and the singular semantic interpretation of the verb *has* makes the agreement features coincide semantically.

In addition to subject-verb agreement, internal noun phrase agreement (i.e., agreement between premodifiers and the head nouns) is another way syntactic agreement is accounted for (Keizer, 2007). The example (1) where *committee* allows only the singular demonstrator (*This committee* vs *\*these committee*) illustrates how premodifying items must syntactically agree with the number features of heads of the NP. “Generally speaking, nominal heads are expected to agree in number with any determiners, quantifiers and numerals” (Keizer, 2007, 18): premodifiers like the quantifiers ‘many’, ‘some’ and ‘few’ are perceived as plural and typically display plural concord, while others are singular and display singular concord. Such plural premodifiers are thus not expected as modifiers of CNs because the concord in the CNP must be singular.

---

<sup>21</sup> Several terms are common to all the studies on agreement (Corbett, 2006, 4)  
Controller: “The element which determines the agreement (say the subject noun phrase)”  
Target: “The element whose form is determined by agreement”  
Domain: “The syntactic environment in which agreement occurs”.

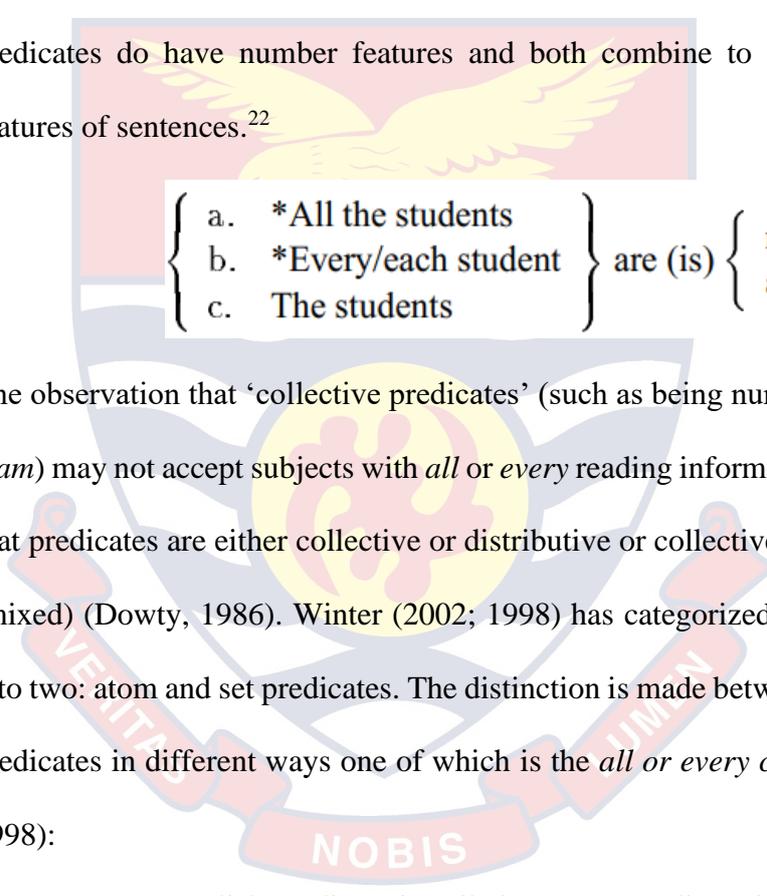
Earlier, collective nouns were defined as lacking overt plural markings unless they denote set of sets. Any given singular agreement may, therefore, be interpreted as a case where semantic and syntactic features coincide as a singular verb choice would mean that the CNP is interpreted as referring to a singular entity even as it lacks the plural marker (Corbett, 2019; Corbett, 2006).

Also, the explanation that agreement forms need to be token identical is very relevant. It allows an interpretation of agreement to be multidimensional, in that, we begin to understand how features of both the subject and the verb are involved in constraining agreement. As a result, finding the predictors of agreement automatically requires consideration of both the subject and the verb, if not the predicate. From this exposition, the claims of Swan (2005) and Levin (2001) that certain features of verbs or verb phrases determine agreement are valuable claims. For example, Swan's (2005) explanation that CNAgre is pragmatic and verbs denoting speech acts of *deciding, hoping and wanting* license plural agreement is a relevant indicator of plural features of predicates. Taking the number features of predicates as a predictor of CNAgre in the analysis section of the study corresponds with the conception that agreement is a matter of token identity between the subject and the VP.

The understanding that syntactic agreement involves token identical semantic and morphological features implies that predicates also participate in agreement. Model semanticists have shown that not only does the subject NP display number features but the semantic interpretation of predicates does the same because sentences have number features. This way, syntactic agreement in this study encompasses the semantic interpretation of the number features of both the subject and the verb or verb phrase. But to understand the number

features of predicates and how they semantically constrain agreement, there is the need to explain number features of predicates in the model semantics framework.

Model semantic theorists have advanced important arguments on number as they find number features to pertain not only to subjects but also to predicates. One most influential work on this is Dowty's (1986) which followed Vendler (1967) that regardless of the number features of certain subjects, some predicates do have number features and both combine to give the number features of sentences.<sup>22</sup>



$$\left\{ \begin{array}{l} \text{a. *All the students} \\ \text{b. *Every/each student} \\ \text{c. The students} \end{array} \right\} \text{ are (is) } \left\{ \begin{array}{l} \text{numerous} \\ \text{a good team} \end{array} \right\}$$

The observation that 'collective predicates' (such as being numerous or a good team) may not accept subjects with *all* or *every* reading informed the conclusion that predicates are either collective or distributive or collective and distributive (mixed) (Dowty, 1986). Winter (2002; 1998) has categorized these predicates into two: atom and set predicates. The distinction is made between set and atom predicates in different ways one of which is the *all or every criterion* (Winter, 1998):

An English predicate is called an atom predicate if and only if the sentence obtained by it to every noun phrase is equivalent to the sentence we get when it is combined with all noun phrase. Predicates that lead to non-equivalent sentences are referred to as set predicates. (Winter, 1998, 251)

---

<sup>22</sup> Also, Kratzer's (2012) *On the Plurality of Verbs*.

Winter (1998) provides a list of adjectives, nouns, verbs that evoke set and atom predicates in A and B.

A. Atom predicates

- i. Sleep, smile
- ii. Be a good team, be numerous, form a pyramid, elect Clinton, be alone
- iii. Vote to accept a proposal, weigh 1kg
- iv. (in some dialects) perform Hamlet, lift a piano, write a book
- v. Student(s), child(ren), shop(s), team(s), committee(s)

B. Set predicates

- i. meet, gather, disperse
- ii. be similar, be alike, be together
- iii. like each other, look at one another
- iv. perform Hamlet *together*, lift piano *together*
- v. Colleague(s), brother(s), friend(s), similar student(s), student(s) who met

The traditional view on number and agreement, per extension, would suggest that set predicates (B) accept singular subjects but they hardly accept plural subjects. It is yet the same extension that is indicative that the plural number of the subject NP is construed only for atom predicates (A). I use examples (3 & 4) from Barker (1992) to clarify the point:

3. **The men** *met on Tuesday.*
4. **The committee** *met on Tuesday.*

Subject NPs need to have plural number features before they can be in a syntactic configuration with the predicate *met on Tuesday*. This cannot be true of singular subjects as in example (5).

5. \*The boy *met on Tuesday*.

The tendency to have (4) thus shows first that the CN *committee* has plural interpretation and second that the plural meaning is involved in this context. Number features of predicates are therefore considered as part of the predictors of CNAgre.

### **Semantic Agreement**

When agreement is accounted for by the semantic features of the items involved in the agreement relation, agreement is said to be semantic. Semantic agreement also called “agreement ad sensum” or notional agreement, is, therefore, a type of agreement in which units of agreement are “consistent with [their] meaning” (Corbett, 2006). Such agreement occurs with units which have two agreement tracks, semantic specification and morphological specification, yet the semantic specification triggers agreement and not the morphological (Smith, 2015). An example of semantic agreement is said to be the agreement relations with collective nouns. Typically, such nouns are morphologically singular yet they can select plural agreement based on what has been argued in the literature as the semantic features of the noun. Whereas CNs are conceptualized as groups containing individual members, it appears to have different semantic features from other nouns and this feature allows agreement relations with CNs to be semantic but not always grammatical (Baker, 2008).

In French, plural pronouns are used even when a singular entity is referred to mark of politeness (see Corbett, 2000 and the literature therein). The

third person plural *vous* is used to refer to an individual when the referent has a higher social status than the speaker. Agreement relation with such use of "vous" is said to be semantic.

Also, the phenomenon of reference transfer (Nunberg, 1997; Pollard & Sag, 1994) shows semantic agreement relation. The possibility for speakers to make metaphorical references leads to agreement relations being accounted for based on the meaning assigned to the subject NP. Nunberg's (1997) famous example (8) explains this with instances where individuals who patronize food from a certain restaurant are referred to with the names of the dishes they order. The agreement between the verb, *is* and the subject NP *French Fries* then is said to be determined by the meaning of the subject NP.

6. That French fries is getting impatient.

I will conclude the discussion of syntactic and semantic agreement with Figure (1).

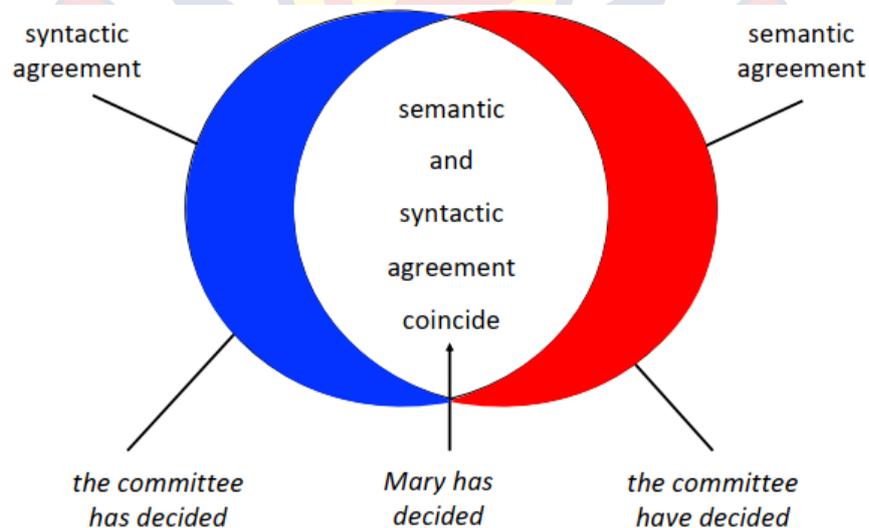


Figure 1: Syntactic and Semantic Agreement

Source: Corbett, 2019

Syntactic and semantic agreement may coincide in agreement (as seen in the middle of the figure) or vary to give syntactic (*the committee has decided*) and semantic agreement when the meaning of an agreement unit is said to determine agreement but not the form (*the committee have decided*).<sup>23</sup> In syntactic agreement, the plural reading of the subject does not input agreement whereas the syntactic feature of the subject, for example, does not inform the semantic agreement but the meaning of the agreement units. While the “systematic covariation” in Steele’s definition is typical of syntactic agreement and semantic agreement, each of these types of agreement does not display consistencies concerning the forms of the items engaged in agreement. In looking at the meaning of the agreement forms as purely semantic, context-specific meanings that accompany usage needs to be mentioned. What has largely been seen as semantic agreement also extends to pragmatic agreement.

### **Pragmatic Agreement**

Pragmatic agreement needs to be distinguished from semantic agreement. Distinguishing pragmatic agreement might involve the distinction between semantics and pragmatics (Sperber & Wilson, 2005; Hawley, 2002; Levinson, 2000; Grice, 1957). The Gricean project must have set this distinction on a proper footing as more than ever, post Gricean scholarship perceived meaning not to only emanate from the meanings of sentences or words exclusive of the context. For instance, Austin’s (1962) *How to do Things with Words* firmly grounds how different words, phrases, sentences could have different interpretations given different contexts. In any utterance, linguistic context (i.e.

---

<sup>23</sup>In considering this example as a description of semantic agreement, we need not forget that it is not explanatory that plural CNAgre is determined by semantic factors – instead, semantics here is used as a basis for features beyond morphosyntax.

co-text) and non-linguistic (participants; their roles, class, etc) context generate different meanings and conceptions. This is what is considered pragmatic.

Pragmatic agreement is a type of agreement determined by number features construed through the context of use: “speaker’s social relation and the attitude to the addressee, and sometimes to third persons too” (Corbett, 2006, 137). Corbett’s (2006, 231) example from Russian, where in showing respect, *vy* ‘you’ (PL) could point to either singular or plural agreement even when a singular addressee is referred to is useful here. In accounting for pragmatic agreement, however, the common assumption “that semantics inputs to pragmatics (and possibly cognitive disciplines)” must be highlighted (Zimmermann, 2012, 2).

The pragmatic nature of agreement also plays a crucial role in the use of corpora for this study. If the claim that “a corpus linguistics methodology offers access to the language output of a cross-section of speakers and can, therefore, reveal meaningful usage which would otherwise elude the attention of one conscious mind” were true, then it would be easier to find context-based predictors of CNAgre in corpora than in “introspection based approach which relies on the intuitions of a single scholar” (Hoffman’s, 2005, 2). Noting that variable agreement is more likely to be context-based, I use corpus methodology which is more useful a method in revealing pragmatic predictors of agreement than others like elicitations.<sup>24</sup>

There is the need to interrogate how pragmatic factors may help in the identification of the number features of CNPs. The earlier proposal that the semantic/pragmatic features of adjectives (Gardelle, 2019; Joosten, 2010),

---

<sup>24</sup> Johansson (1979) used elicitation methods to study CNAgre.

possessives, determiners, and articles (modifiers of head nouns) assign definite reference for nouns needs to be pursued.<sup>25</sup> The next sub-sections will show how the pragmatics of possessives and determiners would help the analysis to identify the pragmatic means to access the number features of CNPs. It is a fact that most of the semantic/pragmatic features to be discussed exist in the literature, their specific role in modifying collectives have not been outlined. As the discussion proceeds, I highlight essential aspects of the conceptions that are relevant to this study.

### **The Pragmatics of Possessives**

Possession is an asymmetrical relationship in which a less salient entity (the possessee) is given definite reference (through the possessor) (Rybarczyk, 2015; Langacker, 1995). Based on the syntax, two types of possession are identified: nominal and clausal (Heine, 1997).<sup>26</sup> I focus on the nominal type of possession because the clausal type which normally involves paraphrases of nominal possession through verbs like *belong*, *own*, *have* coincides somewhat with how the nature of verbs would render a predicate singular or plural which has been discussed in the model semantic conception of the number features of predicates. In considering only nominal possession, I also take a cue from Peters and Westerstahl (2013, 3) that “English is unusual in having two possessive constructions, one prenominal and one postnominal” but not clausal.

Prenominal possessives are typified by the genitive case or a possessive pronoun while postnominal possessives embody an *of-headed* prepositional phrase both as possessors of the head nouns which are normally non-salient or

---

<sup>25</sup> This proposal was made at the section where collective nouns were con

<sup>26</sup> The study uses these names for the two types instead of the attributive and predicative types of possessives because of the explicit syntactic reading of the terms nominal and clausal levels.

indefinite nouns (Kolkman, 2016; Peters & Westerståhl's, 2013; Rybarczyk, 2015; Willemse, 2010). Assigning definite reference through possessives emanates from the relation between the possessor and the possession. Traditionally, the ownership relation (my phone) the assigned interpretation for all possessives so that the possession is what the possessor owes (Heine, 1997 as cited in Kolkman, 2016). This traditional semantic interpretation is narrow as kinship (my sister), part-whole (the cliff's edge), location or spatiotemporality (Ghana's capital), etc. emerged as semantic relations in possession (Kolkman, 2016; Aikhenvald, 2012; Willemse, 2010 Langacker, 1995).

These semantic relations construed by possessives offer tremendous perspectives for understanding CNAgre hence they are useful in the identification of the predictors of CNAgre. Part-whole relations could reveal the number features of CNs. When the possessor is part of the collection (my *family*, my *government* or my *team*), it is more likely for the CN to be taken as a whole and singular but not plural. On the contrary, the possessor could allow a plural reading when they are not part of the collection. Also, ownership provides a pragmatic resource to test animacy. Unless one strictly aims at pointing to a slave type of relation, no individual can possess (in the sense of own) another individual. In this sense, the ownership relation between possessors and CNs would render CNs like *government*, *team*, *committee* as (non-human) institutions. Here also, the sense of ownership in constructs like *John Major's government*, *Richard Hannon's team*, *Mr Anguita's party* may be construed as leadership relations where power and control are perceived of the possessors. Kolkman (2016, 15) maintains that “the [possessive] NPs in isolation then

merely signal the availability of some unspecified relation” but the actual relation is ascertained in a given speech context. She, therefore, concludes that “the interpretation of English pre-nominal possessives cannot be explained in grammatical terms” but must “be transferred to the remit of pragmatic theory” (Kolkman, 2016, 15).

In accounting for the pragmatics of possession, relations of power, control, proximity, intimacy, obligatoriness, privileged access etc. are surface (Rybarczyk, 2015; Aikhenvald, 2012). In the views of Aikhenvald (2012), Jessen and Vikner (2003; 2002), Taylor (1996), possessive relations imply either alienable or an inalienable possession – kinship and part-whole relations being typically involved in inalienable possession while power and control relations normally deal with inalienability. Using the alienability relation, the analysis of the data will examine how possession relations that see the possessor as inseparably joined to the collection imply a singular referent while the denotation of part of the collection to show alienability and focus on the part of the collection will denote alienability and suggest plural reading of the CNP. Proper noun modifiers (e.g. *Obama administration*), however, are not considered as instances of possession marking.

There have been debates on whether or not proper noun modifiers (PNMs) are part of possessive constructs. For Kolkman (2016), *Obama government* and *the Mozart sonata* are examples of possessives yet there is much understanding in Breban (2018) and Breban, Kolkman and Payne’s (2019; 2015) claim that the ability to turn the proper name into genitive is not proof enough that such constructs are possessives. Breban (2018) indicates that it is impossible to form genitive variants of proper noun modifiers in many

instances (e.g. *a Mona Lisa smile* will have different interpretations when the genitive is used). I will not focus much on the reasons for not classifying proper nouns modification as possessives, Breban (2018) offers a very good discussion of the reasons. Instead, I focus on the functions of PNMs.

Relying on the Hallidayian functional linguistics and through a corpus study of BrE, Breban (2018) finds four functions of PNMs. PNMs may be identifying, describing, classifying or complementing the noun they modify.

**Table 1: Functions of Proper Noun Modifiers**

Function of PNM	Example	Meaning of function and example
1. Identifying	<i>Our Berlin Boots</i>	The PNMs serve as epithets which are renditions of long qualifying descriptions with a feature associated with the referent. Breban (2018) indicates that <i>Berlin boots</i> refer to the boots they bought at Berlin.
2. Descriptive	<i>A John Major answer</i>	A description is offered of the head noun to mean it is in a style of the proper noun modifier as in e.g. the answering style of <i>John Major</i> .
3. Classifying	<i>The Yorkshire terrier</i>	The proper noun modifier indicates a subtype of the noun which normally is a generic expression. <i>Yorkshire terrier</i> is a subtype of terrier.

---

4. Complements	A Kerry supporter	Denotations of part of Thing. The referent forms part of the supporters of <i>Kerry</i> .
----------------	-------------------	---

---

Source: Author's Construct following Breban (2018)

These functions in Table 1 emanate from the context of use. All the four functions give the NPs a degree of properhood as their interpretations make them appear proper. To relate Breban's (2018) functions of PNMs to the present study, CNs are defined when they take PNMs. Breban, Kolkmann and Payne (2019) in the workshop on *Different Perspectives on Proper Noun Modifiers* recorded several instances where PNMs were interpreted as proper names. If their findings are true of many native speakers of English, then it is obvious that there will be an interpretation of proper noun modified CNs as denoting a proper noun. PNMs are therefore to be considered as typical cases of definiteness and part of the definiteness predictor.

Pragmatic functions of possession have been shown to play crucial roles in identifying the number features of CNs. I now turn to show how the pragmatics of determiners offer resources that serve as useful resources for identifying the pragmatic features of CNs.

### **Pragmatics of Determiners**

There is a huge body of literature that shares the position that demonstratives are used for several purposes other than the basic deictic function of displaying spatial distance (some are Mwinlaaru & Yap, 2017; Rybarczyk, 2015; Lakoff, 1974 and the literature therein). For some time, two basic uses of the demonstrative pronouns (*this, that, these, those*) were commonly identified but Lakoff (1974, 345-346) and others (Rybarczyk, 2015

Diessel, 2006; Lyons, 1977b; Janssen 2002; Langacker, 1991) have identified rhetorical functions of determiners. To Lakoff (1974), these demonstratives, beyond being (i) “indicators of spatio-temporal deixis” by basically serving as “pointing words” or (ii) being “used as discourse deixis” to refer “back (sometimes forward) to prior (or future) discourse”,<sup>27</sup> may also be (iii) used to show emotional deixis.

The first function appears to be the most common and includes the use of demonstratives to identify objects that are proximal or distal spatially. In the second function, demonstratives refer to discourse either before or after the present discourse. The third function, Lakoff (1974) elucidates, is pragmatic as the demonstratives are used to show “speaker’s emotional involvement in the subject-matter of his utterance.” The emotional involvement implies the speaker’s closeness and the nature of his participation or relation with the subject and such uses are described as *vividness* and occur most often in colloquial contexts. Lakoff identifies three types of emotional deixis: (i) where *this* is used to refer to someone or something mentioned outside the discourse; (ii) used to replace indefinite referents in emotional deictic contexts and (iii) both the emotional and discourse deixis are used. The emotional deixis may show the speaker’s participation in the event (as in example (7) provided in Lakoff, 1974) or show vividness (what Lakoff considers insulting as exemplified in example (8)).

7. He kissed her with this unbelievable passion.

8. This Richard Nixon is gonna get his.

---

<sup>27</sup> These are the two basic functions of demonstratives

Whereas example (7) shows the speaker's position as an observer and how the addressee was fully involved in the event, there appears some sort of insult (in 8).

Aside from the functions identified by Lakoff, several other scholars have made insightful observations of the uses of the determiners. Arguments that “[i]n the extended meanings of demonstratives, the proximity measured in terms of physical space becomes replaced with metaphorical closeness conceptualized within the various mental domains” points to different metaphors that may be construed in the use of determiners (Diesel, 2013 as cited in Rybarczyk, 2015). The spatial construal of relative distance present in the use of demonstratives may be extended to temporal, epistemic, and social distance among others. The underlying function of showing the distance between the speaker and the entity spoken about, from the ongoing, is not abandoned entirely but extended to include social or emotional, temporal, epistemic distance. The distance between a speaker and the CN could show whether or not he/she can identify or is focused on the constituents of the collection or just the single body. This feature of demonstratives will be useful in the explanation of distance as a predictor of CNAgre.

This subsection has looked at agreement and the types of agreement. Agreement has been seen as syntactic, semantic and pragmatic. I will further distinguish between agreement and *government* for further clarification of agreement.

#### Government vs Agreement

In this subsection, I explain agreement as a feature sharing relation between the subject and the verb. The difference is important because it situates

number agreement as an abstraction which is not controlled by any of the resources that engage in the relation. In taking this position, the present study differs from Smith (2020), Barker (2002), Pollard and Sag (1994). The position however aligns with Barlow (1999) who have expanded views that what selects agreement is abstract. I review the government-like feature of agreement first and note points differences in the following.

The meta-language used in describing agreement relations could mislead one to perceive agreement as a government-like relation. To exemplify, expressions like “the subject agrees with the verb”, “probe”, “target”, “determine agreement” and “controller” do suggest that a linguistic unit/resource ‘constrains’ agreement for another which then makes the agreement relation not any different from relations like assigning ‘case’.<sup>28</sup> In other words, clarifications on whether a subject determines the number feature of the verb, for instance, needs to be made.

Kathol (1999, 229), concluding on Pollard and Sag’s (1994) explanation on agreement, stated that “[t]echnically, there is little difference in P&S’s theory between the relations where a verb selects a subject with a particular case marking (like NOM) and the one in which that verb selects a 3rd person singular subject because there is no structure sharing in either case, we can regard the distribution of the features in question essentially as in terms of government” (Kathol 1999, 7). Meanwhile, Zwicky (1986) identified that what separates government relations from agreement relations is the tendency for the latter to

---

<sup>28</sup> Probably this is what makes Chomsky consider agreement to involve the valuation process of assigning of case and number.

share (number or gender) features about a single phenomenon (quantity for example).

Corbett (2006) summarizes and merges the various ways in which agreement relations have been accounted for in the literature. To do this, he enumerates certain terminologies typical of discussions on agreement. The terminologies include *controller*, *target*, *domain*, *features* and *conditions*. Irrespective of the fact that the terms *controller* and *target* also point to the government-like-nature of agreement relations, Corbett's (2006) conclusion that agreement is asymmetrical since targets are constrained by their controllers even makes agreement the more government-like.

Should agreement be asymmetrical as Corbett (2006) posits, and involve a controller which specifies forms targets take, then the relationship is similar to a government relationship. Corbett is quick, however, to distinguish between agreement and government. Table 2 is his summary of the distinctions.

**Table 2: Differences between Agreement and Government**

	AGREEMENT	GOVERNMENT
1. feature specification of target/governee is determined by:	feature specification of controller	presence of governor
2. controller/governor:	has the relevant feature specification	does not have the relevant feature specification <sup>10</sup>
3. element which is normally nominal:	controller	governee
4. features involved are:	gender, number, person, i.e. 'direct' features (§4.2.4)	case, i.e. an 'indirect' feature
5. multiple targets/governees are:	same as each other	different from each other

Source: Corbett (2006, 8)

From Table 2, it appears that agreement and government will only share the quality of being asymmetrical. But this quality is what could lead one blindly to accept the view that in the case of subject-verb agreement, the subject which is the controller specifies what number features ought to be selected by verbs. I believe this is what accounted for Pollard & Sag's (1994) argument that agreement is purely abstract and not a feature of any of the constituents.

Kathol (1999, 11) explains that “selector categories [need] to contain all the information that covaries with that specified on the selected category”. This then becomes different from the other explanations of agreement and puts at an abstract level, the agreement phenomenon. From this conceptualization, therefore, Kathol (1999, 12) could be right to conclude that the agreement relationship between subject and verb in English involves relations where “subject and verb generally agree in number and person”. In this study, agreement relations between CNs and verbs involve the sharing of abstract number features between CNPs and VPs. As seen in the discussion of the number features of predicates, this study focuses not just on how the CNP constrains agreement since that would implicate a government-like relation but focuses on how different linguistic units and items in a clause contribute to constraining CNAgre. In taking this position, this study considers the types of agreement as means to identify how the abstract number (and person) features may be traced to several factors which relate to syntactic, semantic and pragmatic aspects of the language.

In this study, I take the position of Barlow (1999) that there are issues with the morphosyntactic view of agreement and that “the consideration of a range of attested agreement patterns leads naturally to an account in which

agreement relations are seen as links between discourse information structures”. Therefore, the analysis of the data will take “a discourse perspective” which will prevent “the descriptive problems associated with current syntactic approaches to agreement and leads to a revealing reconsideration of the nature of agreement relations” (Barlow; 1999).

### **Theoretical Review**

The argument made up to this stage is that two choices are involved in CNAgre, singular agreement choice or plural. To fully account for factors that constrain CNAgre, we to align with a viewpoint that considers language as embodying choices and offers resources to account for what constrains the choice. Michael Halliday’s functional grammar is a potentially useful framework for such a study:

Research describing functional grammars is often prefaced with strong assertions that the grammars (and therefore the systems, constraints, constituencies and dependencies) are probabilistic, with aspects of language variously described as a gradational, fuzzy and/or cline (Hasan, 1987; Halliday, 1994; Tucker, 1998; Fawcett, 2000; Halliday, 2002). While functional categories have long been described as meaningful tendencies in a continuous space, these shades of grammar have rarely been explored. (Munro, 2008; 2)

‘Probabilistic linguistics’ primarily aligns with deterministic theories that define confidences across (multiple) deterministic models rather than a single gradational model (Monro, 2008). In stating that collective noun agreement choices are graded, singular mostly preferred over plural, one provides a

gradational model which will be different from a focus on modelling multiple factors that constrain agreement. The objective to study multiple determining factors and establish deterministic models to predict their context-free use which is central to probabilistic linguistics is not shared by these functional grammars. The few functional grammar researchers who take up the matter of determinism add the deterministic perspective to their research. Monro's, (2008) "A *Probabilistic* Representation of Systemic Functional Grammar" and Bod & Kaplan's (1998) "A *Probabilistic* Corpus-driven Model for Lexical Functional Analysis" show the integration of probabilistic linguistics in the research of these functional grammars. Probabilistic linguistics but not systemic functional grammar is suitable for investigating constraints of plural and singular agreement choices to predict context-free uses – the objective of this study.

The next subsection discusses Probabilistic Linguistic (PL) theory highlighting its suitability for the present study. In the discussion, I will first define PL, provide a historical account, key assumptions and outline further the reasons that inform its choice for the present study.

### **Probabilistic Linguistics**

Probabilistic Linguistics embodies the conception of language as gradient and stochastic which differs from the perception that language is categorial (Claes, 2017; Bresnan et al., 2007). Contrary to claims that language systems are either grammatical or ungrammatical, probabilistic linguistics recognizes grammaticality as a continuum, a random variable (i.e., stochastic) and that linguistic choices are made from scalar magnitudes of several factors, displaying gradience.

The conception is based on the observation that language consists of systems of choice and that such choices depend on soft constraints such as socio-cultural factors (Halliday, 1961; Zipf, 1949; Sapir, 1921) but not categorial constraints.<sup>29</sup> As an example, the morphological aspect of language, Baayen (2003) has shown that the productivity of word-formation processes is gradient and stochastic. The fact that certain affixes generate many words (e.g. *-ness* as in *goodness*), some more words (e.g. *-ee* in *employee*) and others few words (e.g. *-th* in *warmth*) shows how frequency is interwoven in the fabric of morphology (Bod, 2010; Baayen, 2003) and of course language as a whole. PL generally uses probability theory to model gradient linguistic phenomenon so to explain, describe and predict the phenomenon. To be more precise, PL recognizes the gradient and stochastic features of language and applies an already existing theory – probability theory – to the field of linguistics to account for the grammar of language – a grammar PL refers to as probabilistic grammar.

There are difficulties in establishing exactly when and who started PL. However, conceptualizing and theorizing probability is known to have started in Plato and Aristotle's days and developed in Galileo Galilei's dice throwing challenge. Answers from Pierre de Fermat and Blaise Pascals' game of chance question also added to the development of probability theory and much later Pierre Simon Laplace and Thomas Bayes' classification of probability.

---

<sup>29</sup> Hard constraints are constraints that emanate from the view of language system to comprise of neatly drawn lines between grammatical and ungrammatical structures determined by a few linguistic factors that explain grammaticality. On the contrary, a recognition of grammaticality as a scale and that the choice of any structure on the scale depends on many factors, including those that are not linguistic but pragmatic point to constraints as soft.

Associating probability theory with linguistics (as done in PL), however, is traced to different sources.

Firth's (1957) *polysystematism* view of language which has been profoundly developed by Michael Halliday is robustly built around the notion of probabilities in language. Zipf's (1949) classical observation that the length of words determines their frequency of use as relatively short words are often used is taken by Griffiths (2010) as the key study that starts PL. Manning (2003) suggests Sapir (1921) as the key authority in the origin of PL. He cites Sapir's (1921) statement that "Everyone knows that language is variable" and further adds how language vary between and among individuals, a basis established by Sapir and developed by Labov (1966) as part of the earliest recognition of how language at all levels reflect continua. The emergence of the field of Information Theory (Shannon, 1948) and Harris' call for a marriage between information theory and linguistics is also considered foundational to the development of PL (Pereira, 2000). For Goldsmith (2002), it is in the 1950s that PL started in the works of Ray Solomonoff.

Regardless of the diversities in the historical accounts of PL, it appears clear enough that gradience has played a foundational role in ancient grammars like that of Panini and featured in earlier versions of the generative tradition (Bod, 2010). The emergence of the minimalist program (Chomsky, 1995) which sought to reduce all grammar rules to the minimal form possible and whose thesis that competence is categorial but not gradient brought to the limelight preference for categorial rules to the neglect of the study of gradience. For minimalists, considering language as gradient mix in world knowledge, does not model grammaticality and does not meet the goal of describing the mind-

internal I-language as opposed to the observed-in-the-world E-language (Manning, 2003).

It was the realisation that categorial rules do not fully account for linguistic phenomena that led to a renewed interest in PL. The observation that categorial rules do not fully account for the dative alternation phenomenon is one such example. For the dative alternation phenomenon, categorial grammars' claim that the VP NP<sub>i</sub> NP<sub>d</sub> structure is used when a change of state (possession) is implied (e.g. *Susan gave the children toys*) and the V NP<sub>d</sub> NP<sub>i</sub> structure is used when a change of place (movement to goal) is involved (e.g. *Susan gave toys to the children*). PL studies found these two factors to be only extreme among many others that influence dative alternation at relative degrees. Bresnan et al. (2007) identify discourse accessibility, definiteness, pronominality, topicality, discourse accessibility, semantic class and syntactic complexity as predictors of dative alternation. Roland et al. (2005) and Jaeger and Snider (2008) have also identified several factors in addition to the categorial ones as influencing genitive alternation and presence and absence of complementizer respectively. These observations highlighted unaccounted-for-gradient features of language and called for the need to study them.

Further, both human and animal cognition has been proven to be probabilistic (Lau, Clark & Lappin, 2017; Bod, Hay & Jannedy, 2003; Manning, 2003). Also, results of experiments in psycholinguistics revealed that well-formedness judgements of words and sentences are largely determined by the “combined probabilities of their subparts” (Bod, Hay & Jannedy, 2003). These imply that language systems in themselves are probabilistic. The realisation of the probabilistic nature of both the system and text instances invigorated the

interest of researchers from the middle of the twentieth century (as exemplified by important studies such as Shannon, 1948; Zipf, 1949; and Labov, 1966) and even more by the end of the twentieth century and the beginning of the twenty-first century (as studies such as Bod, Hay & Jannedy, 2003; Jurafsky, 2003; Goldsmith, 2002; Bybee and Hopper, 2001; Coleman & Pierrehumbert, 1997; Halliday, 1991). Since the underlying objective of this study is to use the deterministic theory to predict context-free constraints of collective noun agreement, PL, as conceived as such, will be discussed further.

Throughout the twentieth century, several attempts to extensively discuss the deterministic theory of probabilistic linguistics were made among which is Bod's (1998) *Beyond Grammar: An Experience-Based Theory of Language* which put the theory in perspective. After the linguistic symposium organized by the Linguistic Society of America on "Probability theory in linguistics", PL had gained much scholarly attention as the first book that fully addresses the theory was compiled after this meeting in Bod, Hay, and Jannedy (2003). By the first half of the twenty-first century, therefore, numerous studies on PL were carried out and the discussion of PL appeared in several contexts. Again, Bod's (2010) discussion of PL in Heine and Narroq's (2010) "Oxford Handbook of Linguistic Analysis" has provided an insightful discussion of PL. In the following discussion, Bod (1998; 2010) Bod, Hay, and Jannedy (2003) served as the primary reference.

Three central propositions identify PL which are that: (1) language is probabilistic; (2) the object of any linguistic study should be language use data but not intuitions; (3) linguistics thrives on predictions and inferences from such data. These central claims are discussed in turns.

I. Language is probabilistic

Evidence from several sources supports the fact that language is inherently probabilistic. Bod, Hay and Jannedy (2003) identify the frequency, variation, universals, gradience etc. as evidence for the probabilistic nature of language. I will discuss how variation, frequency and universals point to the probabilistic nature of language.

1. Variation: It is clear that no one person speaks in the same way given different speech contexts so it is that no two people speak a language in the same way. For these statements to be true, knowledge of variation must be part of the linguistic system to allow speakers manipulate and implement syntactic to “phonetic variants to portray linguistic and extralinguistic information” (Bod, Hay & Jannedy, 2003, 12).
2. Frequency: It is well established that language representation, processing and change are replete with frequency effects (Bod, Hay and Jannedy, 2003). The effects manifest in the easy identification, interpretation and use of frequently occurring words, meanings and combinations. PL allows modelling of the frequency effects with the help of probability theory.
3. Universals: the linguistic universals postulated by generative linguists are not verifiable by empirical studies of different languages (Bod, 2010). Through probabilistic viewpoint, it is easy to establish that though different languages share certain universal phenomenon, the distribution is relative (Bod, Hay & Jannedy, 2003). The task for the linguist then is to explain “universal representation for linguistic

experiences that should apply to all languages” but not to find the Universal Grammar (Bod, 2010).

The display of variance, how frequency affects language and the innate ability for humans to use experiences to ‘generate’ variants of linguistic forms are indications that language is probabilistic and must be studied as such. With the frequency effect manifesting everywhere, language appears to have all the features of *a probabilistic system* and therefore “knowledge of language should be understood not as a minimal set of categorical rules which may be characterised by a statistical distribution” (Bod, Hay & Jannedy, 2003, 10).

## II. Past linguistic experiences as the object of study

The second key tenet of PL is its ascription to performance as the object of linguistic study but not competence. Probabilistic Linguistics thrives on the observation that both the use and interpretation of language involve a representation of concrete experiences (Bod, 1998). Interpreting an utterance such as “it’s cold in here” to mean a request to close the window is an understanding of the utterance that is highly possible given a list of possible meanings and a consideration of the context (Manning, 2003). This shows that arriving at an interpretation of any utterance is probabilistic and that a listener or interpreter’s experience is highly likely to help resolve any ambiguity in comprehending such utterance. Again, there are a few options to consider in inquiring about someone’s age: by asking “How old are you?”, “What age do you have?”, “How many years do you have?” (Bod, 2010, 648). Yet, I agree with Bod that producing the first clause is more acceptable and is the “conventional way of asking someone’s age in English” though the other two are not wrong options (Bod, 2010; 1998). The fact that *How are you?* will be dominant in any

given corpus of English further proves that its acceptability and use is a representation of the experience of its use.

From the ongoing, it is obvious that a speaker/hearer's knowledge cannot be seen as grammar but as "a statistical ensemble of language experiences that change slightly every time a new utterance is processed" (Bod, 1998). Past language experience (i.e. discourse) serves as a more appropriate object of study for the linguist than the native speaker/hearer's knowledge since his/her knowledge is largely shaped by the experiences. This means that the advancement in technology which saw the collection of huge samples of language use and its subsequent conversion to machine-readable formats and the availability of software to allow easy analysis an essential factor in the development of PL (Baayen, 2003) and corpus linguistic provides a very reliable source of data for linguistic studies (Ngula, 2014; McEnery & Hardie, 2012; Leech, 1992).

III. Linguists could make predictions and inferences from such data

Finally, given a huge corpus of language use like the BNC, PL allows inferences to be made of a phenomenon. PL allows predictors to be evaluated through statistical models (such as logistic regression) after coding to ascertain how well they constrain the linguistic phenomenon. "A logistic regression model permits simultaneous evaluation of all the factors in a model and assesses the strength of each of the factors relative to others" (Bod, 2010). Bresnan et al. (2007) tested the reliability of the logistics regression model and stated that the model could predict with 94 per cent accuracy. From this, a categorical claim of belief in a predictor could be established after the strengths of the predictors have been found. In doing this, PL ascribes more to the Bayesian type of

probability. In this study, the relative strength of the predictors of CNAgre shall be determined to predict CNAgre to establish the Probability Grammar of BrE speakers' CNAgre.

So far, PL has been defined, a brief historical view of the theory has been given and its key assumptions have also been explicated. I proceed to highlight the strength of PL which is part of the factors that motivated its choice for this study. The first is its relation with other theories of language.

Unlike other linguistic theories that abandon previous observations and claims of 'conflicting' theories, PL "integrates" previous "knowledge with a probabilistic perspective", thereby, 'enriching' linguistic theory with statistics (Bod, 2010). With the assumptions that units of grammar coincide with that of language use and comprehension, context-free grammars such as Tree-Adjoining Grammar, Lexical Functional Grammar, Head-Driven Phrase-Structure Grammar (see Bod [2010, 653] for references) and Cognitive Grammars (e.g. Claes, 2017; Lau, Clark & Lappin, 2017) have benefited from the integration of probabilistic perspectives. Most PL studies identify predictors, adopt or adapt them and through probability models, determine the relative strength of the factors (Claes, 2017; Bod, 2010). The predictors could be those established in previous studies by Cognitive Grammars, Context-Free Grammars, Systemic Functional Grammars through to those by Traditional Grammars. In doing this, PL attains a rare eclectic feature which is very essential to the present study as it gives the avenue for various understandings from previous empirical studies on the factors that determine CNAgre, as well as different grammars like cognitive grammars, model semantic theory, and lexical semantics to be assessed, modified and variously used in the analysis.

To be more specific, this study adapted previous observations. First, that animacy determines CNAgre as a predictor of CNAgre is adapted as a factor in this study but the present study examines further resources that point to animacy. Also, the model semantic theory's conception of plural and singular predication with previous findings that certain verbs, verb phrases and speech acts constrain agreement form (Swan, 2005; Leech & Svartvik, 2002; Levin, 2001) are combined to make number features of predicate a predictor of CNAgre. Finally, the study adopts the conception of accessibility and definiteness which appear universal constraints in PL studies (Claes, 2017) as predictors of CNAgre.

The methodology and analysis aspects of this study are shaped by the underlying assumptions of PL. In the first place, the point has been made in passing that the second objective of identifying the predictors of CNAgre will benefit from PL by accounting for both existing predictors and exploring the data to find other predictors of CNAgre. Also, the use of logistic regression modelling to evaluate the predictor variables is one of the benefits this study derives from the use of PL. PL often studies the gradient features of binary outcomes such as *genitive alternation* (Roland et al., 2005), *dative alternation* (Bresnan et al., 2007), verb complementation – *ing vs. to complements* (Deshor & Gries, 2016) by using the *logistic regression* probability model to assess the predictor variables. The present study involves binary outcomes of singular and plural CNAgre and uses the logistic regression model to simultaneously evaluate and assess the predictor variables.

Despite the above strength, there are criticisms against PL. Paramount among the arguments has been the argument against statistical testing of sample

of language. Koplenig (2017) argues that the basis for statistical testing is that the corpora is a representative sample of a language but regardless of the size of a corpus, no corpus of any language is representative enough of the language because identifying the sample space of language which is an ever-growing object is utterly impossible. What Koplenig (2017) and even Gries (2005) lose sight of is that any linguistic study freezes language use in time to account for the features about the language. Nothing, therefore, prevents a researcher from statistically testing a frozen corpus to make predictions. Of course, the fact that language is an ever-developing system is the more reason why it keeps varying and the more reason why statistical tests are needed to predict what changes are more likely.

In summary, this section has discussed the PL theory as the underlying theory for the study. The descriptive, explanatory and predictive values of PL have been identified as factors that inform the choice of PL as the theoretical framework of this study. I have argued that a theory that accounts for the gradient feature and the inherently probabilistic feature of language offers a valuable perspective for the study of CNAgre.

### **Empirical Review**

This section reviews empirical studies on collective noun agreement. The review highlights the objectives of the studies, the methods and key findings and relates the empirical studies to the present study. In relating the studies to this research, the observed constraining factors of the various studies are evaluated, abandoned, adopted or adapted as predictors of CNAgre. Also, the methodologies the studies used will be assessed just as the impact of the

conclusions they arrive at. In all, the empirical studies affirm the relevance of using the corpus linguistics methodology in accounting for CNAgre..

Bailey (1987) studied variable agreement relations for *mass, collective and generic nouns* in English. He did not state the data or the methodology of his study but it appears he used much introspection and relied on personal experience to validate his position. He observed that young people who were in sync with 'current developments' preferred the singular agreement and the plural agreement relations are special uses. His observation, no matter how weak his conclusions are, has some relevance to the present study as it maintains the biases involved in a mere consideration of age as a factor that motivates CNAgre. In this study, age will not be considered because except Bailey (1987) no known study has shown the influence of age on CNAgre. Fries (1988) set out to test the validity of Bailey's claim that young people preferred singular agreement by observing extracts from British magazines published for and by young people. After observing that young people selected both plural and singular forms, Fries concluded that young people are not careful in their choice of agreement forms since they use both plural and singular agreement forms indiscriminately. These two studies shape the perspective of the present study that even though age, gender, social class play crucial roles in the variation in terms of language use, they do not appear to be predictors of CNAgre. This is because such sociolinguistic predictors are mostly at play when the alternations or variants do not result in much difference in meaning. In the case of CNAgre, there are contexts in which a plural verb cannot be replaced by a singular verb without a very significant change in meaning or rendering the statement ungrammatical or ambiguous. Owing to the recommendations by Fries that

regional variation could be the best avenue to explain CNAgre, many studies have focused on variational distinctions.

Levin (2001) is among the first studies that used the corpus linguistic methodology to extensively study collective noun agreement. His thesis was to describe the use of CNAgre in American, British and Australian English by examining frequency and variation across different genres and media. Both spoken and written data were used.<sup>30</sup> Levin *sampled* twenty-six nouns based on their frequency of occurrence in the corpora and their representativeness. From his analysis, he found BrE as having a high preference for plural CNAgre, AusE making moderate use of plural CNAgre and AmE having the lowest proportion of plural CNAgre use in both the written and the spoken genre. In testing how medium influences agreement, I found plural *agreement* to be a feature of spoken discourse while singular the written discourse. In the sub-genres of the newspaper data, Levin observed that the more formal the genre was the more likely it is for singular agreement to be selected while less formal genres favour the plural agreement choice.

In relative pronoun agreement, Levin (2001) noted that the relative pronoun *who* influences plural agreement relation with CNs, *which* does attract singular agreement relation and *that* in both singular and plural agreement. This finding offers relevant insights to this study. Since the relative pronoun *who* is related to animate, and *which* and *that* to inanimate objects, his finding serves an essential basis for a consideration of how animacy greatly informs CNAgre.<sup>31</sup>

---

<sup>30</sup> The spoken component of BNC, Longman Spoken American Corpus [LSAC]) and written data from *The New York Times*, *The Independent (Ind)*, and the *Sydney Morning Herald (SMH)* representing American, British and Australian English respectively were used.

<sup>31</sup> This study is not focused on relative pronoun agreement but on subject-verb agreement but considers relative clauses as contributors to the assignment of a by number of definite reference to CN heads of CNPs. They are therefore helpful in identifying the number features of CNs.

To clarify, the animacy predictor would consider not just the nature of the collective noun because the collective nouns studied are all collections of human but the changes for non-human relative clauses would point to non-human reference to the collections.

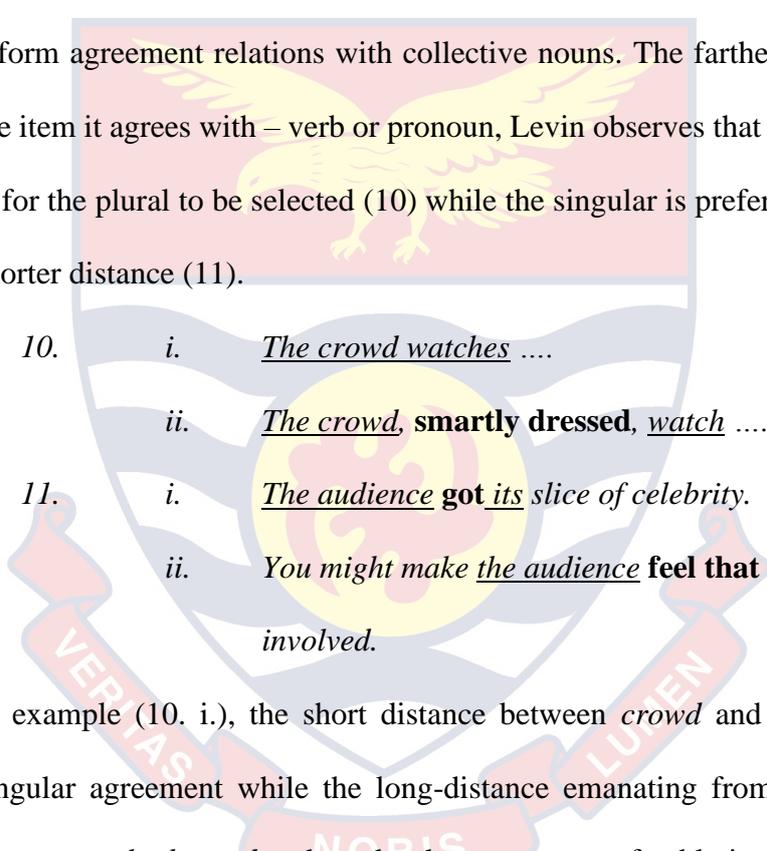
Levin (2001) further observed how semantically conceptualized animacy influenced agreement. Inanimacy as seen to relate to a perception of CNs as abstract referents that entail a large number of people prefer singular agreement. This will add to the strategies involved in identifying humanness as a predictor of CNAgre. Levin distinguished humanness as manifested in the use of relative pronouns which require human antecedent as marking both syntactic and semantic agreement for CNs. The semantics pertains to consideration of CNs as abstract entities which render them singular. Also, Levin (2001) observed that nouns are abstract when they refer to ‘generic entities’ such as his example (9):

9. I get clues as to how *a particular family* organises its life, whether room is a room for the whole family or a room that’s perhaps excludes children (...)(BNC)

He adds indefinite reference as a means of generic reference to the collection by identifying determiners like *each, every, this, that a/an* as indicators of indefinite reference to CNs and a means to render CNs abstract. In this study, definiteness of the CNP which would either be definite or indefinite is taken as a predictor of CNAgre in itself. Definiteness need not be considered a measure of humanness. Therefore, the indefinite determiners he identified still do not necessarily render CNs abstract or non-human. For instance, nothing about “each/this family” shows that it refers to human constituent or non-human.

Besides definiteness, some of these determiners will be seen to point to other predictors. For example, *whole* and *entire* as determiners in a CNP, would make listeners perceive the CNP as comprising a unit or a whole. Therefore, *whole* and *entire* would have to form part of the part-whole predictor; *a*, *each*, *every* will relate to indefiniteness; ‘*this*’, ‘*that*’ relate to distance but not animacy as Levin explains.

Further, Levin (2001) found syntactically conceptualized distance to inform agreement relations with collective nouns. The farther the CN is from the item it agrees with – verb or pronoun, Levin observes that the more likely it is for the plural to be selected (10) while the singular is preferred for relatively shorter distance (11).

- 
10. i. *The crowd watches* ....  
ii. *The crowd, smartly dressed, watch* ....
11. i. *The audience got its slice of celebrity.*  
ii. *You might make the audience feel that they are involved.*

In example (10. i.), the short distance between *crowd* and *watches* favours singular agreement while the long-distance emanating from the two lexical items *smartly dressed* makes plural agreement preferable in example (10. ii.).

In (11), the use of the plural pronoun *they* (11.ii) becomes of the distance as seen in the presence of the two lexical items,<sup>32</sup> *feel that* but ‘*its*’ (11.i) when the distance is shortened because of the presence of only one lexical item between ‘*audience*’ and the pronoun ‘*its*’.

---

<sup>32</sup> Conceptualizing distance as measured by the number of words is also a reflection of the concept of ‘agreement attraction’ which makes technically ungrammatical agreement relations acceptable because it is understood noun a noun closer to the verb, but not the subject itself could influence agreement. There is still debate on this phenomenon (Wagers et al., 2009)

Fernández-Pena (2017b; 2016; 2015a; 2015b) has taken the challenge to investigate how relative distance constrains collective noun agreement. Her main focus has been to investigate how relative distance and complexity in the use of *of-dependent* postmodifiers determine agreement in different dialects of English. Fernández-Pena (2017b) used the BNC, COCA and the Corpus of Global Web-based English (GloWbE) and found, contrary to Levin (2001), that more complex and distant *of-dependent* postmodifiers preferred singular agreement in 21 varieties of English she studied. Her use of GloWbE which embodies a less formal type of language use might have accounted for this as she observed lesser frequencies of plural agreement in the British and American English components of GloWbE than in the BNC and COCA respectively. Although her suggestion that genre type and formality is more plausible a constraint of plural agreement than distance and complexity, it neglects the fact that postmodifiers do not only construe distance but are also resources to ascertain the humanness nature of CNs, whole part relations if the postmodifier is a postnominal possessive and others. Though this study would ascribe distance as a predictor variable, distance is conceptualized as the relative distance between the speaker and the collection which is not measured by the number of words between the CN and the verb that it agrees with, but by spatial, temporal, epistemic, social distance<sup>33</sup>. Demonstrative determiners would hence help in the conception of distance (see the pragmatics of agreement for explanations; read Kolkman, 2016; Rybarczyk, 2015 and literature therein for the explanation of pragmatic construals of distance) but not necessarily

---

<sup>33</sup> Rybarczyk (2015) and Diesel (2013) explain the various extension of the conception of distance as they found demonstrative determiners to construe relative distance in many terms in addition to spatio-temporality.

postmodifiers. Although it has been repeated several times that the explanation of linguistic choices taking a cross variety perspective is highly valuable, this study maintains the position that we first need to find out why the variants exist in any variety in the first place and what factors influence choices of variants.

Another problem that might have made the findings of Levin (2001) and Fernández-Pena (2017b) inconclusive is the collective nouns they studied. Levin (2001) studied twenty-six collective nouns. Levin's (2001) twenty-six CNs comprised those that usually engaged in plural agreement relations, those that usually had the singular agreement choices and those that had variable agreement. Fernández-Pena (2017b) does not also show which CNs she studied but her examples indicate her use of CNs which are known for mostly selecting singular agreement (*majority, band, group*). Some of the nouns, they study (like *population* and *majority*) are not part of what Depraetere (2003) has listed as typical CNs. It is because of this that the present study focuses only five CNs which are typical CNs and part of the CNs that have variable agreement features.

The diachronic evidence of Annala (2008) who investigated variation in CNAgre within hundred years provides insights into collective noun agreement. Annala (2008) used the corpus methods to find out if CNs have internal stability in preference for singular and/or plural agreement by measuring such stability in hundred years. Analysing instances of CNAgre in Extended Version of the Corpus of Late Modern English Texts (CLMETEV) and BNC, he found that agreement preferences are relatively stable so that CNs which prefer singular agreement continue to prefer singular agreement while those which prefer plural agreement would continually prefer plural agreement just as those that prefer

variable agreement - though the last group seems to slightly move towards the singular. This finding validates the synchronic approach of the present study in that since CNAgre appears more stable in time, inferences can be made of a 'frozen' data of BrE.

Also, Annala (2008) observed in both corpora that tense influence the preference of agreement form: high preference for plural agreement with past tense but the lower preference of singular agreement for past tense. Even though he studied only the tense forms of *be*, this study aligns with the position that tense constrains agreement. Annala's (2008) position, the present study would consider past tense as relating to different predictors of CNAgre. First, since past tense is a feature of the verb, it adds to how predicates constrain CNAgre. Also, tense as a grammatical conception of time implies that tense could show relative temporal distance so that the tense forms part of the distance predictor of CNAgre. The latter position is in line with Kolkman (2016) and Rybarczyk's (2015) conception that distance helps in fixing referents for non-salient or common nouns.

Again, even though the nature of verbs influenced agreement with collective nouns, the type of verbs depended on the type of collective noun involved. For *family*, Annala (2008) exemplifies that verbs like *live*, *say*, *do*, *know*, *get* and *feel* denote concrete meanings when used with *family* whereas *need* and *own* are abstract and make descriptions of states but not actions. This observation validates the semantic concept that predicates could be plural or singular. This study will not consider the verb per se as a predictor but the predicate so that resources like pronouns and adverbs as well as verbs which constrain agreement in the predicate will all be accounted for together. Annala's

finding also collaborates Levin's (2001) which found VPs that indicate inclusion (*e.g. contain, comprise, include* and *be made*), passive verbs that refer to how collectives are constituted (*e.g. be established, be formed, be set up etc.*), and VPs that focus on the size of the collective (*e.g. be big/small, decrease, double*) typically select singular verb agreements. These will help in the discussion section of the study.

Lakaw (2017) is also a diachronic study of CNAgre in BrE and AmE from 1810 to 1909. He takes samples of use from *Corpus of Historical American English* (COHA), *Old Bailey Corpus* (OBC), and *Corpus of Late Modern English Texts*. His categorization of CNs into semantic groups (like employment – *e.g. crew, staff*; family – *e.g. couple, family*) is commendable as it supports sampling CNs from the category for a careful study to explain what characterise nouns in that category. Far more useful is his observation that the development towards the singular agreement preference in American English must have happened in the 20<sup>th</sup> century since both varieties had a stable development towards the singular before then. The relevance of this observation lies in his conclusion that language external features like those given in prescriptive grammars might have influenced the choice of singular agreement. This claim perhaps is seen in the caution grammar texts, like Quirk et al. (1985) advise that individuals should resort to singular verbs for CNs when they are not sure. Since these claims are made in the face of the lack of established constraining factors for CNAgre, this study finds its relevance in Lakaw's conclusion and call for the determination of the relative strength of the constraining factors of CNAgre.

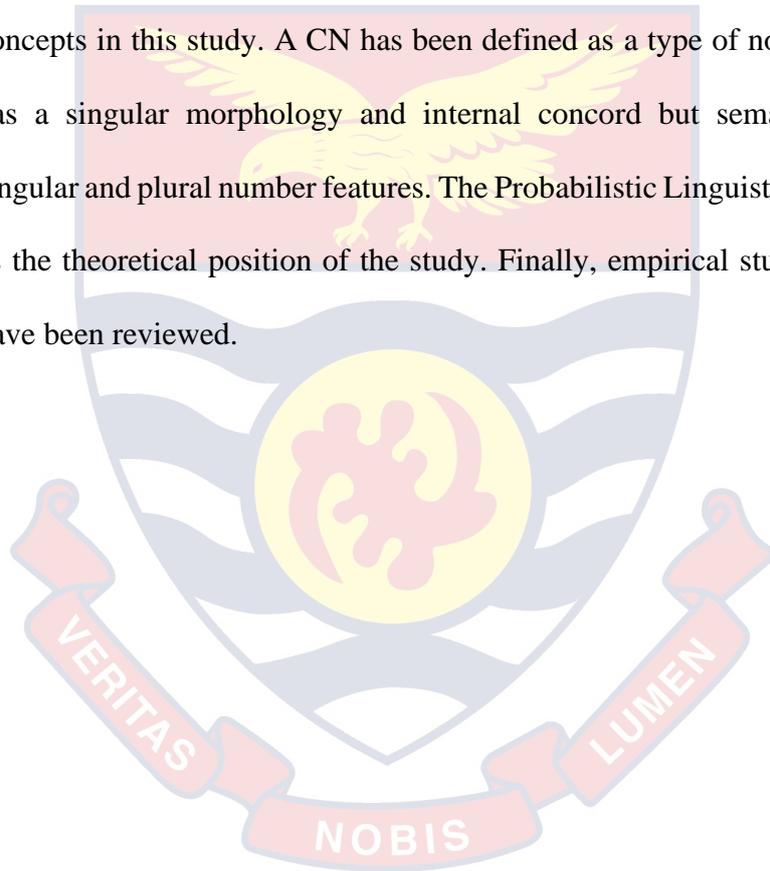
A careful observation of the empirical studies reviewed here would question the use of the corpus linguistic methodology and the BNC specifically in this study since they have been used by previous studies. Even though the review has not highlighted most of the studies that were not situated in Corpus Linguistics (CL), some studies have used different methodologies. Before the use of CL in studying CNAgre, introspection and elicitation studies (like Bailey, 1987 and Johansson, 1979) offered little explanation to how CNAgre is determined. For example, Johansson gave two sentences representing singular and plural CNAgre to his respondents to elicit their reaction. While he found both AmE speakers and BrE speakers to have positive reactions towards singular, he noted that AmE speakers, unlike BrE, had negative reactions towards the plural sentence. Even though this is experimental evidence for AmE's preference for singular CNAgre, it does not explain what influences any of the choices. It was the early corpus studies (like Levin, 2001) that found some factors that to some extent influence CNAgre. That is, these CL studies could even nullify Bailey's (1987) claim that plural agreement was the current phenomenon on the grounds of huge examples of singular agreement preference from corpora.

This study chose BNC although others who have studied the phenomenon have used it. From Levin (2001) through to Kairis (2017), the BNC seems to be used by so many researchers to explain CNAgre. The reason for this situation is that it offers many instances of variable agreement that helps to explain CNAgre. It is interesting, further, that each of the many studies that use the BNC observe something relatively different from the other studies that used the same data. I believe that the approach used, the objective, the

methodology and more especially the theoretical framework of a study shape what could be found of the same corpora. This is why the BNC is used with the probabilistic linguistic framework to provide the probabilistic grammar of CNAgre.

### **Chapter Summary**

This chapter has reviewed the literature relevant to the study. Collective nouns and Agreement have been defined, described and discussed as key concepts in this study. A CN has been defined as a type of noun that typically has a singular morphology and internal concord but semantically involve singular and plural number features. The Probabilistic Linguistics was discussed as the theoretical position of the study. Finally, empirical studies on CNAgre have been reviewed.



## CHAPTER THREE

### METHODOLOGY

#### Overview

The previous chapter was a review of literature. It specifically provided explanations on key concepts such as collective noun and agreement, discussed probabilistic linguistics as the theoretical position of the study, and reported on previous empirical studies on CNAgre. This chapter describes the methodological procedures used in conducting this study. I discuss approaches in probabilistic linguistics used by asserting its theoretical implications and the need for the Corpus Linguistic Approach.

#### Corpus Linguistics (CL)

The corpus linguistic methodology is used in this study. Huge samples of clauses that where CNPs engaged in verbal agreement were easy to identify and examine through this method. I will explain what motivates my choice of the methodology and how it helps this study in the following.

The previous chapter established the relevance of naturally occurring data in linguistic studies. The basic assumption that language is characterised by variations proves that comprehension and use of language thrive on various approximations made from previous language experiences rather than abstract grammar rules (Bod, 1998). By viewing the grammar of a language as “a statistical ensemble of language experiences that changes slightly every time a new utterance is processed” (Bod, 1998), the linguist understands that the knowledge of a competent speaker cannot be understood as grammar and intuition need not be the object of study. Far more useful an object of study in linguistics then are samples of actual language use and corpus linguistics (CL)

creates the avenue for huge data of such samples to be accessed and studied (Bod, 2010; Bresnan et al. 2007; Manning 2003).

Corpus is generally defined as a body of naturally occurring text (written and/or spoken) considered representative of a given language which is compiled for linguistic analysis (Bednarek & Caple, 2017; McEnery & Hardie, 2012; Tognini-Bonelli, 2001; Sinclair, 1991; Francis, 1982). Commenting on the several definitions of corpus, Nartey and Ngula (2009) note that the naturalness in the definitions of corpus embodies the efforts in making a corpus ‘authentic’ so that the texts added to any corpora are free from biases. Since the samples of language use are collected mostly without the producers’ prior knowledge that such texts are to be used for linguistic studies, the texts are considered reflective of actual linguistic practices of speakers and devoid of biases emanating from observers and respondents’ manipulations.

Concerning representativeness, the fact that it is difficult to attain data on all instances of language use even of a specific group of people and that language is ever-growing proves that no corpora will ever be representative of any language. This might answer Koplenig’s (2017) question of the representative sampleness of any corpora. Notwithstanding, gathering data from varied contexts of usage offers a better resource for language study as several examples are provided in the varied contexts and many instances of use (Bednarek & Caple, 2017; Nartey & Ngula, 2009). Koplenig (2017) advises that since language use of secluded genres is hard to attain and the population of whom samples of language use is gathered is mostly hard to define, null hypothesis testing should be abandoned in CL. His position - though captivating on the face of it and appearing to challenge null-hypothesis testing - is very hard

to accept based on his recommendation. He recommends, in agreement with Schmid (2010), that because CL offers descriptions of E-language, its results should compare with other empirical evidence from cognitive linguistics to account for “how our brain works” (Schmid, 2010 cf Koplenig, 2017, 22).<sup>34</sup>

From the probabilistic linguistics (PL) perspective, most of the descriptions offered after corpus studies are descriptions that align with already established cognitive linguistics descriptions. For example, cognitive linguistics generally consider markedness, preemption and priming as constraining factors of linguistic choices (Claes, 2017). Claes points to semantic roles, definiteness etc. to operationalize markedness in data while tense is operationalized for preemption in his study.<sup>35</sup> The probabilistic linguistics literature also identifies definiteness, givenness, distance, semantic roles as predictors. This means that at the underlying levels, cognitive linguistic constraints of linguistic choices manifest in PL’s constraints. One can, therefore, make descriptions, explanations and inferences of a corpus if the study aligns with PL. By far, it is priming that appears a difficult concept for PL and also CL but since less frequent items are never considered ungrammatical or ignored, the chances for less frequent items to be preferred based on priming cannot entirely dispute the relevance of CL and PL (Bod, 2010).<sup>36</sup> Even so, regression analysis does not

---

<sup>34</sup> *E-language* defines “a language as a collection of actions, or utterances, or linguistic form (words, sentences) paired with meaning, or as a System of linguistic forms or events” (Chomsky 1986: 19).

<sup>35</sup> Markedness explains the observation “that approximating archetypical conception [tends to be] coded linguistically by a category taking that conception as its prototype” (Langacker, 1991, 298). Preemption is an entrenched form of a structure. Goldberg (2011) explains: “consider for a moment, the more easily tractable case of morphology. How is it that we know we should use *went* instead of *\*goed*? Clearly it is because we consistently use *went* in contexts where *goed* would have been at least as appropriate: this is statistical preemption”.

<sup>36</sup> (Structural) Priming is a cognitive constraint that shows that a language user will more likely use a feature or form that has just been used or processed (Claes, 2017)

determine the strength of a predictor by its frequency in informing choice, it does so by expression of belief that the predictor can predict a choice.

The predictor variables that are identified in the present study align with cognitive explanations of how the human mind works as they align with established predictors in PL. That is, predictor variables identified in the data will be related to more general conceptions like definiteness, accessibility, distance whose explanations show how language systems work as shown in the cognitive linguistics literature. Therefore, Koplenig's argument was addressed when this study took CL as a methodology and PL as the underlying theory. In doing so, I reiterate Thompson and Hunston (2006, 8) that "corpus linguistics is a methodology that can be aligned to any theoretical approach to language". This position is not to merely ignore claims that CL is a theory (Stubbs 1996; Leech 1992) but to establish that even when a theory ably functions as a methodological tool, the theory does not lose much when it serves a methodological need, a theoretical need or both.

More so, CL has become a very useful methodological tool for research in linguistics (Bednarek & Caple, 2017; Ngula, 2009; Hunston, 2002; Meyer, 2002; Leech, 1992) and will inform the present study on "interesting lexical, phraseological, semantic and discourse insights" that may be hidden from the researchers who use introspection (Ngula, 2014, 522).

The above notwithstanding, CL methodology has some weaknesses. One is the temptation for studies that use CL methods to provide frequencies of occurrence as findings and draw conclusions on such frequency distributions. This *uncritical approach* (Hoffman, 2005) rests on the frequency of use to make claims to the neglect of detailed explanations. To benefit from the strengths of

CL methodology and still avoid its limitations, the present study – which is different from other studies – does not merely provide frequencies but focuses on explaining factors that determine agreement in addition to noting their frequencies. Thus, the guiding theory, which of course informs the objectives and research questions, requires first the identification and explanation of predictor variables before the reliability in the predictor variables are determined through statistical modelling.

The ‘uncritical-approach’ (as Hoffman, 2005 calls the approach of merely listing frequencies) weakness of CL also leads to yet another important part of the CL methodology. Bednarek and Caple (2017) and Hunston (2002) note that the use of corpora for (linguistic) studies gives room for both qualitative and quantitative analysis. These translate into the descriptive and explanatory aspects of PL and the potency of PL to make inferences.

In this study, the quantitative dimension follows the recent probabilistic approach to language studies in sharing the view that grammaticality is highly sensitive to and dependent on probabilities of occurrence (Geeraerts, et al. 2010; Bresnan & Ford, 2010; Bod, Hay & Jannedy, 2003). My interest then is in how probability accounts for the grammar of CNAgre. Explanations are provided first on factors that determine CNAgre before the evaluation of the strength of the predictors is given. Meyer (2002) summarizes the reason for this approach when she argues that unless the linguist is bent on just providing a list of frequencies of occurrences and computations, one may have to necessarily provide explanations of what is seen in the corpora through a linguistic framework.

## Data

The British English (BrE) was selected for this study because among all varieties of English, it is noted for its variable agreement relations. Therefore, enough samples of use of both singular and plural agreement choices will be found in the BrE than in any variety. To acquire samples of use of BrE, I selected the British National Corpus (BNC) as the data for the study. The BNC is a huge database and provides enough samples of past linguistic experiences of BrE than any existing corpus on British English.

### British National Corpus

The BNC comprises about 100 million words. It is designed to represent both spoken and written samples of British English between 1980 and 1993 and was released in 2007. Ninety per cent of the BNC is written while ten per cent is spoken. The written component comprises ‘extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays’ and several others (BNC, 2017). The 10 per cent spoken component comprises transcripts of ‘unscripted informal conversations (recorded by volunteers selected from different ages, regions and social classes in a demographically balanced way) and spoken language collected in different contexts, ranging from formal business or government meetings to radio shows and phone-ins’ (BNC, 2017).

As can be seen, different domains of language use are represented in the BNC. This makes the corpus representative (in a general sense of the word) of BrE. Both the written and spoken components of the BNC shall be studied to aid the generalizability of the findings. The study of both the spoken and written

component could allow stylistic variations to be accounted for should they determine CNAgre.

### **Collective Noun Sample**

Five collective nouns (*committee, family, government, party, team*) are sampled from about 200 collective nouns in English for this study (Gardelle, 2019; Fowler, 1992).<sup>37</sup> Three purposive sampling techniques were used in the sampling process. The first method selected CNs which have received relatively high scholarly attention in the collective noun agreement scholarship. This is determined through the selection of collective nouns studied by twelve research articles on CNAgre and used in existing grammars. The choice of articles (Kairis, 2017; Smith, 2017a; Wong, 2017; Örlegård, 2014; Outratova, 2013; Folkesson & Lantz, 2011; Annala, 2008; Bock et al. 2006; Hundt, 2006; Sauerland & Elbourne, 2002; Den Dikken, 2001; Levin, 2001) depended on recency (article should have been written between 2000 and 2017), language use (must have studied CNs in English) and number of CNs studied (more than two collective nouns). A total of 292 collective nouns were sampled from all the twelve articles. CNs that had been studied by at least six authors were considered to have received much scholarly attention. Seventeen (17) CNs were used in at least six and at most ten of the studies. The seventeen CNs and how frequently they have been studied are given in Table 3.

---

<sup>37</sup> The total number of CNs in the English Language is hard to find (airis, 2017; Annala 2008; Fries, 1988; Svensson 1998), yet it is estimated to be above 200 (Gardelle, 2019; Fowler, 1992).

**Table 3: Frequently Studied CNs**

COLLECTIVE NOUN	FREQ. OF OCCURRENCE IN ARTICLES
Audience	6
Army	7
Club	6
Committee	10
Council	8
Couple	6
Crew	7
Crowd	7
Family	10
Gang	6
Government	9
Group	8
Majority	6
Minority	6
Party	7
Staff	9
Team	8

Also, CNs discussed in grammar texts such as Quirk et al. (1985), Downing (2015), and Greenbaum (1996) were considered in the selection of CNs. The present study's objective to identify nuance factors constraining CNs and determine how they have emerged from complications in existing explanations. As such, studying the nouns which are frequently studied by these research works and grammars would be informing. CNs such as *family*, *committee*, *team* were also mentioned in the reference grammars. This further confirmed the attention given to the study of certain CNs. From the first sampling method, eight CNs (*committee*, *crew*, *crowd*, *family*, *staff*, *government*, *team*, *group*) which received more scholarly attention were selected.

Possessing key features of CNs was used as a basis to further select from these ten nouns. *Group* for example is taken from the sample because it is argued to be indeterminate, is an umbrella term for “any number of entities considered together” (Gardelle, 2019, 5). Among the three types of collective nouns, non-human CNs (eg., *archipelago*, *consortium flora*) prefer singular agreement mostly (Corbett, 2006); animates (e.g., *flock*, *gaggle*, *herd*, *horde*, *pack*) has a very low preference for plural agreement as compared to the human category (e.g. *government*, *team*) which shows relatively high plural agreement. This study hopes to contribute to collective noun agreement and as such, it was prudent to study CNs in the human category. Therefore, *group* which lies across all the three types was not taken.

Finally, the frequency of use in the BNC determined the choice of CNs to study. That is, the study also considered the chances of acquiring enough examples of the use of the CNs in the BNC to select them. This is in agreement with Levin’s (2001, 49) claim that:

*While some collectives are common enough to necessitate a limitation of the number of included examples, others are too infrequent to yield material that is large enough for comparisons.... An alternative approach [to the frequency approach] would have been to study how the rarer noun types are treated since these are more likely to have variable usage patterns than the more frequent ones. Yet the problem remains that very few tokens do not allow firm conclusions.”*

The five CNs selected after these sampling processes are *committee*, *family*, *government*, *party*, *team*.

### Data Analysis Procedure

In analysing the data, a word search was carried out in the BNC for all the collective nouns under study. The search displayed all instances of the use of each of the CNs in the BNC which has been summarized in Table 4. Even after a cursory observation of hits that display agreement, there was a huge concordance which would make interpretation difficult (Hunston, 2002). This study normalized the hits for all CNs. Considering that the least number of hits was 18, 757 (i.e. team) the study focused on the first 19, 000 hits of all the CNs. Therefore, the first 19, 000 hits were carefully observed to select instances where overt CNAgre relations were marked.

**Table 4: Hits of CNs in BNC**

Collective noun	Hits in BNC
Committee	18, 852
Family	33, 693
Government	61, 798
Party	39, 710
Team	18, 757

Source: Author's construct

In scanning the text for samples, some of the instances were rejected because they did not display subject-verb agreement features. Hits with the following features were not considered:

- CNs used as pre-modifiers (e.g. *government* officials)
- CNs that form part of postmodifying units of head nouns
- CNPs in relations where the verbs do not overtly display agreement features
- CNs used in conjunction with other NPs such that the subject NP is plural syntactically.

Table 5 is a summary of the hits selected for analysis.

**Table 5: Extract of CNAgre in BNC**

CN	Singular Agreement	Plural Agreement	Total
<b>Committee</b>	1261	129	1390
<b>Family</b>	600	286	886
<b>Government</b>	1806	45	1851
<b>Party</b>	787	24	811
<b>Team</b>	1074	455	1529
	<b>5528</b>	<b>939</b>	<b>6467</b>

In all, 6,467 clauses which mark CNAgre were extracted as data for analysis. This comprised 939 instances of plural verb use (representing 14.5 per cent) and 5528 singular agreement relations (forming 85.5 per cent). These instances were analysed to answer the questions.

In analysing the data for the predictors of CNAgre, the nature of the subject CNP, the predicates and contexts of use were considered. Using examples from the extracted data, these predictors are explained as part of the analysis and discussion.

### **Logistic Regression**

In order to meet the third objective of the study, I used the logistics regression to model the constraints of agreement for the CNs. Logistic

regression is a statistical approach to describe and estimate the relationship between one variable – in the case of the present study, the choice of singular or plural CNAgre – and another variable, factors that constrain CNAgre. Its inferential tools are used in the prediction of the choice of singular or plural agreement relations in terms of the factors that determine CNAgre. It predicts the chances of the factors that determine CNAgre on the probability scale of 0 and 1 and also assigns to each singular and plural CNAgre a coefficient for measuring the independent contribution of the variation given the constraining factors. The intercept is the value of the logit model when the values of all the singular and plural agreement choices are zero (no predictor variables). The regression coefficients describe the amount of contribution of the factors that determine CNAgre on the logit model. A positive value of the regression coefficient means the constraining factor increases the probability of the model and a negative value means the constraining factor decreases the probability of the model (Agresti, 1996).

### **Inferential Statistics in Logistic Regression Modelling**

**Wald Test:** This is used in determining the statistical significance of the coefficients in the model.

**Odds Ratio:** this measures the relation between categorical outcomes like binary responses – singular and plural agreement. The odd of singular occurring measures the probability of its occurrence against the occurrence of plural agreement.

Logistic regression enables researchers to assess models to “predict categorical outcomes with two or more categories” (Pallant, 2005, 160). The use of the model provides the advantage of evaluating independent

contributions of multiple factors that constrain agreement and in the same process allows one to evaluate the joint contribution of specific constraining factors (Marneffe, Grimm, Arnon, Kirby & Bresnan, 2012). In the literature, the model has been used by linguists to predict the choices certain linguistic forms given potential variables. Scholars attest to how useful logistic regression modelling is to language studies (e.g. Baayen, 2008b; Forster & Masson, 2008; Jaeger, 2008). Also, studies like Bresnan et al. (2007) on the dative alternation, Weiner and Labov (1983) on the active/passive voice alternation and Jaeger (2010) the presence and absence of complementizer have well employed logistic regression in explaining binary outcomes. Since the study investigates binary outcomes – singular or plural CNAgre – logistic regression modelling will be particularly helpful.

### **Summary**

This chapter has described the methodological processes involved in the study. The chapter discussed the need for corpus linguistic methodology to be used since the probabilistic linguistic theory, the underlying theory of this study relies on huge discourse data to determine the probabilistic grammar of language users. The corpus for the study has been described, the methods for extracting clauses has been established, the data size determined and procedure for data analysis has been spelt out. In explicating the procedure used in data analysis, it has been observed that detailed descriptions of the factors that determine CNAgre will first be provided and through logistic regression modelling, the relative strength of these determinants will be integrated.

## CHAPTER FOUR

### ANALYSES AND DISCUSSION

#### Overview

The present chapter is the analyses and discussion of the data. A descriptive analysis is given first of the data extracted from the BNC. I first give an overview of the extracted clauses that overtly mark agreement in relations where *government*, *committee*, *family*, *team* and *party* are the heads of the subject CNP. After a careful study of the extracts, the analysis identifies six factors as constraining CNAgre: animacy, accessibility, definiteness, distance, part-whole and predicate. Descriptive statistics of the distribution of these factors in constraining CNAgre is given. The strengths of these factors in determining CNAgre are also determined through inferential statistics, the logistic regression model.

#### Distribution of CNAgre in BNC

A total of 6,467 clauses were extracted from the BNC as clauses involving CNAgre with the CNs *government*, *committee*, *family*, *team* and *party*. The distribution in Table 6 shows that the majority of the agreement relations (5528 of the extracts, that is 85%) were singular agreement choices while the lowest (939 extracts forming 15%) were plural agreement relations.

Despite the general preference for singular agreement for all the CNs, the CNs differ in their distribution of plural and singular agreement.

**Table 6: Frequency & Percentage Distribution of CNAgre for the CNs in BNC**

CN	Singular Agreement	Plural Agreement	Total
----	--------------------	------------------	-------

<b>Committee</b>	1261 (91%)	129 (9%)	1390 (24%)
<b>Family</b>	600 (68%)	286 (32%)	886 (13%)
<b>Government</b>	1806 (98%)	45 (2%)	1851 (28%)
<b>Party</b>	787 (97%)	24 (3%)	811 (12%)
<b>Team</b>	1074 (70%)	455 (30%)	1529 (23%)

Source: Author's Results (2018)

Two of the CNs, *family* and *team* have relatively high preferences for plural agreement (32% and 30% plural agreement respectively) and another two, *government* (2% of the extracts being in plural agreement) and *party* (3% of instances in plural agreement relations) have less frequent plural agreement forms. With 9 per cent of hits being plural, *committee* has moderately lower plural preference than *family* and *team* but fairly than *party* and *government*. This seems to confirm the observation that different CNs have different degrees of preferences for singular and plural agreement choices (Fernández-Pena, 2015a, b; Levin, 2001). However, a very important concern of this study is to determine why the CNs have such preferences. Two reasons account for the variation in the disparity in this distribution and these reasons provide an underlying explanation for the CNAgre choices.

1. Difference in their Distribution in a Single Text in the BNC

CNs which have fairly more plural agreement choices (*family*, *team*, and also *committee*) have fewer hits in a single text than those which have less plural agreement choices. This is calculated by finding the average of the hits by the text type in which they occur and also by selecting the results on frequency per million words in the BNC and this distribution is shown in Table 7.

**Table 7: Distribution of Texts and Hits for Selected CNs in BNC**

CN	Number of Texts	Hits	Mean	Frequency	Per
					Million Words
Committee	1754	18852	10.7	191.75	
Family	2836	33693	11.8	342.71	
Government	2299	61798	26.8	628.58	
Party	2439	39710	16.3	403.91	
Team	1958	18757	9.5	190.79	

Source: Author's Results

*Family*, *team*, and *committee* have more plural agreement relations (32%, 30% and 9% plural agreement respectively) and fewer hits in the BNC (33693, 18757 and 18852 respectively) as compared to *government* and *party* which have lesser plural agreement choices and respectively had many hits of 61798 and 39710 (see Table 7). Moreover, the mean scores (in terms of the total number of hits per number of different texts) for *family* (11.8), *team* (9.5), and *committee* (10.7) are smaller than those for *government* (26.8) and *party* (16.3). This means that given any one single text, *family*, *team*, and *committee* are used less frequently than *government* and *party*. Moreover, per million words in the BNC, *family* is 342.71 times frequent, *committee* has a frequency of 191.75 and *team* 190.79, representing lower frequencies when compared with *government* and *party*, which have 628.58 and 403.91 frequency of occurrence per million words respectively. The distributions reveal that CNs which have larger frequencies of occurrence in different texts in a corpus have very high singular agreement relations and low plural agreement relation whereas those that have lesser frequencies have a relatively lower number of singular agreement relations and slightly higher frequency of plural agreement. The results suggests

that frequency of use of the CN itself informs the choices of singular and plural agreement.

## 2. Priming

The distribution of the data further points to priming as an underlying explanation for collective noun agreement choices. Priming refers to the act of constraining the choice of a feature because it has just been used or processed (Claes, 2017). Agreement choices formed clusters of plural agreement choices in single texts. That is, given a text in which the plural agreement choice is selected first, there was a higher tendency for plural agreement to be selected for many instances of the use of the CN in that text. Appendix B is a summary of how the distribution points to priming. The data also shows several instances where the choice of singular agreement in a text did not constrain subsequent singular agreement choices though the choice of plural agreement first often triggered subsequent plural agreement relations. In addition to the fact that singular agreement is the preferred choice for all the CNs, it appears that singular agreement is the entrenched agreement feature. In other words, the choice of singular agreement in a text is predicated on its prototypicality which makes its choice a case of markedness (Claes, 2017). Because a plural agreement choice requires previously used forms of agreement feature, it can be concluded that plural agreement is primed.

Further, the nouns that have relatively high plural agreement features (*team, family*) differ somewhat from those that have a fewer plural agreement (*government, party*). The distribution of plural agreement for *team* (a query of the search terms *team have* and *government have*) shows that plural agreement choice is made at most four times in a single text. For *government*, as many as

171 plural agreement records are seen in a single text. These results are presented in Tables 8 and 9.

**Table 8: Distribution of plural agreement relations between ‘team’ and ‘have’**

Text	JXV	K25	K4T	KIR	J1M	HPC	HBK	CH3	CEP	AK6	A2E
Hits	2	2	3	3	2	2	2	3	4	4	2

**Table 9: Distribution of plural agreement relations between ‘government’ and ‘have’**

Text	ABU	K5D	G3H	G5G	HDT	HHV	HHW	HHX
Hits	2	3	9	2	2	162	118	171

The results demonstrate how the very frequent nature of a CN in a single text informs the kind of agreement relations they engage in. The HH texts of *government* (Table 9) are parliamentary section records and indicate the tendency for extensive interactions about *government* to be held so that several instances of agreement relations could be identified for analysis. It is obvious that texts in which people hold an extensive conversation about *family*, *team*, *committee* are rare but quite common for nouns like *government* and *party* and this would shape the frequencies of agreement forms. The conclusion by previous studies that certain CNs prefer plural agreement choices (Levin, 2001) appears to have been confirmed. This results here however contests the conclusion on the grounds that it is very likely that given a similar text for all CNs, the difference in preference for either singular or plural will be very similar. The main point here is that CNs have preferential occurrences that have a tenuous connection to their frequency of occurrence in a single text.

## Discussion

In response to the question, “how does the distribution of CNAgre relations provide an underlying explanation for CNAgre?”, the study has found the most plausible explanation for plural CNAgre choice to be priming. It found that the choice of agreement feature is usually preferred after a specific agreement feature has been selected. It is likely then that specific texts would display more choices of specific agreement form which would be a confirmation of the claim that text type informs agreement relations. What this means, however, is that at the underlying levels, there is a trigger in these text types for a kind of agreement feature. The study has also shown that singular agreement choice is entrenched when compared to plural, which is mostly primed.

From the analysis, this section has also found that even though some CNs have higher frequencies of occurrence in the plural agreement relation (*team, family*), it must not be taken as explanatory enough a factor that informs CNAgre as Annala (2008) and Levin (2006) concluded. That is, the observation that plural agreement choice is a feature of certain CNs need to be revisited because given a huge text of balanced data, there appears a high tendency for all the CNs to display quite similar agreement relations. The above also suggests a weakness of the Corpus Linguistic methodology (Koplenig, 2017) as the difficulty in the balance of corpora may mislead us to make certain conclusions. It also highlights the danger in relying solely on frequencies to explain linguistic phenomena (Koplenig, 2017). However, the probabilistic linguistic theory which allows multiple theoretical views to be adopted in the conduct of research has served an important role here. The cognitive linguistic constraints of priming and markedness have therefore assisted in explaining how the

distribution characterises CNAgre choices. After proposing “that probabilistic grammars may benefit from insights from cognitive (socio)linguistics”, Claes (2017) found markedness, priming and preemption to constrain agreement relations with existential *there* constructs. In identifying markedness and priming as explaining *agreement*, the present study corroborates Claes’ conclusion that “with the support of Cognitive Sociolinguistics, Probabilistic Grammar may move beyond description towards explanation”.

### **Factors Determining Collective Noun Agreement**

Generally, the analysis and discussion here aim to combine theoretical ideas advanced in cognitive linguistics, probabilistic grammars, and other grammars to offer “a social-interactional perspective” on collective noun agreement – making the analysis discourse-oriented. The present section is written from an analysis of modifiers, appositives, adverbials, and clauses which do reveal the number features of collective nouns by noting how they, in turn, constrain agreement. A close analysis of articles, adjectives, adverbs, determiners, prepositional phrases, relative clauses, adverbials and linguistic resources in the complement of the 6467 clauses has brought out themes such as definiteness, accessibility, part-whole conception of NPs to form the factors identified as determinants of CNAgre. These factors will therefore be discussed as identified constraints of CNAgre.

#### **I. Accessibility**

From the data, the number feature of the CN is sometimes explicitly expressed in the linguistic context, that is the co-text. The indicators include nominals that share coreferential features with the CN. Such number features of

the nominals are seen to be a reflection of the number features of the CN. When the number feature of the CN is known, the choice of agreement relation is easily determined because the known number feature constrains agreement. These nominals may be preceding the CN in clauses or adverbials - an example of this is (13):

13. Scrambling ashore, and scuffing *their* way inland through ankle-deep ash, *the party were able to see* that the centre of all the excitement was the small cone of Perboewetan, at the northern end of Krakatoa. (ASR 401)

The plural pronoun *their* in the adverbial makes a cataphoric reference to *the party* and highlights the plural nature of the CNP (e.g. 13). As the plural number is given, the plural agreement relation between *the party* and the finite form *were* is determined by the given number feature.

In addition to adverbials, other clauses preceding CNAgre constructs may also have nominals which spell out the number features of the CNP to inform CNAgre. The Example (14) illustrates a committee deliberating issues and the plural number of the participants informs their use of the plural pronoun “We” in addressing themselves in utterances before the CNAgre construct – ... *do we know ... we're talking.... We've budgeted.*

14. But it seems <-|-> to be a lot less than it might have been last year, I mean I, do we know how many that's, we're talking about? We've budgeted this year, er, for a maximum of six, but as Committee were informed at the last meeting, that number has exceeded during the year, erm, and that did create some financial difficulties .... (J3P 402)

It would be understood in this context that *Committee* refers to individuals in the meeting who have been referred to with the plural pro form. In effect, the preceding context indicates whether the members (plural) or the institution (singular) is the given number feature of the CN. Here also, it becomes obvious that the known number feature determines CNAgre choices.

Nominal appositives<sup>38</sup> were also seen to be highlighting the number features of CNs and determines CNAgre. In the data, a pattern was recorded of appositives that indicate whether their antecedents, CNs in this case, are plural or singular; the following (15 & 16) are examples.

15. Mr Fitch a former barman had Korsakoff's disease, a kind of brain failure occurring as a result of alcoholism. His behaviour was becoming unmanageable at home. The background was that Stanley's first wife had died six years previously, and Joan had been her best friend. Joan herself was a widow, and as an ex-nurse she befriended Stanley and began to care for him. She offered to be his housekeeper, but he wanted her to move in with him. ***The family, her children, were Roman Catholic***, and advised against this without marriage, so she married him. (CGD 610)

16. At first, it seemed that ***the family - mother, father, and two sisters – were more or less alone in their grief.*** (CAH 1253)

The appositives in examples (15 & 16) are plural in number and they render *family* plural. *The family* (in 15) can be replaced with *her children*, in that Joan has lost her husband and what remains as her family are her children: more so, it was these children who *were Roman Catholic* and who *advised* her. The plural

---

<sup>38</sup> This is one of the five types of syntactic apposition Meyer (2002) identified.

nature of the appositive (her children) informs one that the referent of is a plural number of individuals and therefore the plural agreement is determined by such given plural number. For example (16), *mother, father, and two sisters* are the referents of *family* since they *were more or less alone in their grief*. In these examples (15 & 16), the co-referentiality between the two noun groups (*the family*, and *mother, father, and two sisters*) allows the overt plural nature of the appositive to point to a plural interpretation of *the family*; an interpretation which supports plural agreement relation between the CN and the finite verb. This shows that the number features which determine agreement for CNs can be accessed in the linguistic context of the CN.

Further analysis of the data revealed the appositive use of *of NP postmodifiers* as another type of appositive that gives the number feature of CNs and determines CNAgre. Instead of analysing *of NPs* as possessives (as done by Langacker, 1995; etc), one can analyse them as appositives performing two key appositive functions: (1) referential or (2) qualifying function (a position taken by Taylor, 1999). Each of the two functions in the Taylor analysis reveals the actual referent, including the number features, of the CN making it easy to determine agreement relations. I will analyse examples (17, 18, 19 & 20) to show how number features are given in such appositives.

17. *The party of skylarks were taking\_a breather* from their incessant high-rise singing to indulge in an early-morning splashing.

18. A *team of top American managers*, who have taken over running the 97-year-old firm, are now its only hope. (CRB 2188)

The *party of skylarks* (e.g.17) does not actually refer to *a party* but to *skylarks* which is the complement of the preposition *of*; the preposition semantically

helps the location of the actual referent of *party*. Further evidence to support this analysis of the CNP is seen in the nature of the predicate – *taking a breather from their incessant high-rise singing ...* is easily associated with *skylarks* than with *party* though the *skylarks* exist as a *party*. The number feature of *skylarks*, the actual head of the referential expression, therefore gives a plural reading of the CNP and constrains plural agreement. For example (18), *team* is an epithet and *top American managers* serves as the head (actual referent) of the NP. The analysis here follows Taylor's (1999) that in a construct like *an angel of a child*, *an angel* is not the referent but serves as a qualifier of *a child*, the referent of the NP, making an *angel* an epithet. Also, evidence from Example (18) supports this analysis. Firstly, the act of *running the firm* is not a reserve for *a team* but definitive of the professional practices of 'managers' - *managers* participate in the inner meaning of *running the 97-year-old firm* than *team*. Secondly, the *only hope of the 97-year-old firm* is not in 'a team' but in *top American managers*. In this case, the actual referent, *top American managers* gives the plural number feature of the CNP and determines agreement relations with the finite verb.

I have by far demonstrated that plural number is given in the *of NP* appositive, I turn to show how, similarly, the singular number is accessed in such contexts. For CNPs such as *team of ministers* and *a committee of MPs* (e.g. 19 & 20), the CN serves as the actual referent and head of the phrase and the noun phrase in the appositive specifies the referent of the head as a body of individuals. *Team*, *committee*, *party* are the referents so that the NPs of the postmodifier serve as qualifiers that distinguish them from other *teams* and *committees*.

19. A *team of ministers* is assigned to the Northern Ireland Office. (AK9 198)

20. A *committee of MPs* meets to decide the fate of a controversial bill under which the main line and underground stations would be transformed and a new terminal platform built for Channel Tunnel trains. (A9N 64)

In example (19), *team* can be analysed as the referent because it is the body assigned to the Northern Ireland Office but to clarify what team it is, the appositive – *ministers* is used. In this reading of the CNP, the actual referent becomes a body unless other resources existed to point to the number features of the CN. Taylor (1999) illustrated this use of NP constructs with *a distance of 10 miles* and analysed *distance* as the referent which is further specified by *10 miles*. It is possible to interpret example (19) that it is *a team* that has been assigned to the Northern Ireland office and though it could have been any team, this particular team entails ministers. Also, the meeting in (20) is a committee meeting but such meeting will be attended by *MPs*; *MPs*, therefore, specify that the committee does not comprise any category of people but *MPs*. In such usages, the singular agreement form is preferred because the plural postmodifier merely specifies the constituents of the body but it is not the referent or head, the actual referent which is the singular institution determines agreement.

From the analysis above, the number features of a CN may be given as either singular or plural in the linguistic context. If the number features are accessible, the number agreement choice for the CNP is seen as depending on that given number feature. Therefore, accessibility is said to determine collective noun agreement. This finding differs from Fernández-Pena (2016)

who analysed *of NP* postmodifiers as construing syntactic distance between the CN and agreement unit and found that relative syntactic distance modestly influenced CNAgre as the present study takes *of NPs* to form appositives that point to the number features of CNs. The observation, however, confirms Joosten et al.'s (2007) findings of Dutch collective nouns that the tendency to access the number features of CNs indicates whether the plural or singular number features of CNs are referred to in the particular speech context though the observation here slightly modifies their position.<sup>39</sup> In this study, accessibility is seen to define instances where number features are given in context. Context here appears as the co-text (Bach, 2005) that is found in “the immediate linguistic environment” within which a unit of discourse, a collective noun or noun phrase in this case, “occurs and is interpreted” (Janney, 2002).

## II. Distance

The data also revealed a pattern where the speakers and listeners or interlocutors construed distance between themselves and the collective noun and such distance implicates one's (in)ability to perceive the constituents of the collection, their number feature and consequently determine agreement. Demonstrative determiners were found in the data as construing such distance. The demonstrator *that* typically mark distal relations and such distance normally leads individuals to take the collection as one single body. Examples (21 and 22) substantiate how distance informs CNAgre.

---

<sup>39</sup> They consider accessibility to be a feature of the CN as certain CNs allow easy access to the members while others do not (Joosten et al., 2007).

21. Then, too, Jackie made a misjudgement in 1966 and 1967: he stayed in BRM when *that team was already on the slippery slope towards oblivion.* (CD9 67)

The use of *that in* the example (21) is to designate emotional distance: the speaker emotionally distances himself as she perceives the British Racing Motors, the *team*, to have been *on a slippery slope* – he was not on the slippery road. By distancing herself emotionally from the collection, the speaker allows the interpretation of the collection as a distant ‘other’ they are not part of. A distant view of the collection allows conceptualization of the collection as one singular body, a conception that eventually licenses singular agreement.

The demonstrative determiner that traditionally marks proximity (i.e. *this*) has been argued to encode both proximal and distal relations (see Rybarczyk, 2015 & Lakoff, 1974 and the literature therein). When it shows proximal relations in its use with CNs, it allows a reading of the constituents of the CNs. Once the constituents are perceived, the plural number feature is revealed and it triggers plural CNAgre. An example is given in (22):

22. Mr. Alan Williams: To ask the Secretary of State for Trade and Industry what plans he has to combat short-termism among investors in British industry.

The Secretary of State for Trade and Industry and President of the Board of Trade (Mr Peter Lilley): The most effective way of avoiding short-termism is to defeat inflation, a task which *this Government are well on the way to achieving.* Mr. Williams: Does the Minister recall that, after collapsing by 30 per cent at the start of the last decade,

manufacturing investment took the rest of the decade to return to its starting point? Now it has collapsed by 20 per cent.

*Government*, in example (22), is viewed from the viewpoint of the Secretary of State for Trade and Industry who relates with *government* as a member – he is a minister of the ruling government. Relatively, he is closer to the *government* than non-members like Mr. Williams who prefer highlighting the shortfalls of the *government* and his closeness further manifests in his positive remarks about the government. Due to his emotional attachment to *this government*, Peter is able to construe *government* as embodying a plural number of people, including himself, who are all doing well. This analysis adds to the evidence that a speaker's solidarity (to the government in the case of this example) is also a way of marking social proximity (Acton, 2014). Conceptualizing proximity allows him to construe a plural number for *government* and such plural number triggers plural CNAgre. That is, Peter's expression, *this government* is to a plural referent – himself and the others who form the government – and such plurality informs plural verbal agreement choice.

In the data, distance was marked by the use of *this*. There were several instances where *this* pointed to a body the speaker is removed from in terms of time, space, emotions, etc. Such use matches Lakoff's (1974) observation on derogatory use of the demonstrative determiner (such as *this Henry Kissinger*) to signal attitudinal distance.<sup>40</sup> In example (23), *government* is perceived as a distal body and therefore the singular number feature associated with a singular body is triggered.

---

<sup>40</sup> Other recent studies like Mwinlaaru (2018), Rybarczyk (2015) and Kolkman (2016) have confirmed this.

23. THE Government's coalfield massacre plan was in total disarray yesterday after the High Court ruled it was illegal.

Miners at threatened pits greeted the ruling with the demand: Now let us get back to work.

As Government and British Coal lawyers last night pored over the fine print of Lord Justice Glidewell's surprise decision, Board of Trade President Michael Heseltine faced an avalanche of calls to quit the Cabinet.

It will take a political Houdini escape trick to keep him from being sucked further into the mire....

Calls for Mr Heseltine's resignation were growing louder last night.

Labour's Trade and Industry spokesman, Robin Cook, said: 'This judgment is a humiliation for Michael Heseltine. ***This Government has been pursuing an energy strategy*** that was wrong in conception and incompetent in execution.'

*Government* refers to the conservative *government* which Mr Michael Heseltine is a member of but the utterance with "... this government has..." was made by Robin Cook who is a member of the opposition Labour Party. For Robin Cook, 'This Government' is a body he is not part of and his social distance manifests in his conception of *government* as a group whose membership he does not know. A further demonstration of his social distance is seen in his view that *government* is "incompetent". *This* in this context will then have an emphatic sense, a pragmatic strategy to isolate the government for criticism. The analysis here also follows common knowledge which is further intimated in the *Face* conception in *Politeness* that for the purposes of politeness, speakers may prefer

being indirect. The derogatory comments that usually distance speakers from the collective nouns require speakers to be indirect and such indirectness is what makes them refer to the singular collective institution but not the individuals that form it. Also, in line with the analysis of *this* as a means of showing politeness would imply that individuals who prefer being ‘bald’ would rather refer to the constituents even in passing derogatory remarks to show distal relations and the plural number and plural agreement would be constrained.

One’s relative distance to the collective nouns may allow him or her to perceive the CN as a plural denoting noun or as a singular noun. Also, the analysis has shown that emotional distance marking is usually associated with insults and such insults require indirect referents for politeness purposes which is why the reference is normally to the singular body but not to known constituents. This is the first study to have analysed demonstratives as informing CNAgre in English. Even though distance has been said to inform CNAgre (Fernández-Pena, 2017b; 2016; 2015a, b), such distance was defined by the number of lexical items between the head of the CNP and the unit it agrees with which differs from distance as construed socially. However, Bauer’s (1994) observation that in British English, the singular agreement is the preferred relation when government refers to a non-British government and the plural for British government has been confirmed by the results of the present study. The analysis here has also shown that demonstratives “do not just specify where something is” but serves as a resource for specifying “which one [we] are talking about” (Enfield, 2003).

Distance does not categorically determine agreement. In some instances, plural and singular agreement forms were preferred when respectively the distal

and proximal relations were conceptualized. Even though these were few, they are evidence to support the probabilistic grammar orientation of this study. That is, it supports the position that linguistic choices are not determined by hard categorial rules but soft gradient constraints (Bod, 2010).

### III. Part-whole conception of CNP

The analysis and discussion in this section briefly report findings of previous studies to set the pace for the analysis and discussion of part-whole conception as a determinant of CNAgre. Levin (2001) in studying how premodifiers and determiners inform CNAgre found the determiners *whole* and *entire* to be the only modifiers among others like *this*, *that*, *a/an*, *the* that establish a clear pattern of frequently determining singular CNAgre. His explanation that the adjectives *whole* and *entire* evoke “a holistic conceptualization” of CNs and such whole reading is what makes them prefer singular agreement is an important point. It offers the foundation for the explanation of the various ways this study observed as a means of conceiving CNs as wholes, a conception which shapes agreement choice. In furtherance of Levin’s analysis, I noted prenominal possessives as determiners used for such holistic conceptualization of CNs in the data.

Possessives that construe membership relations, ownership relations (*his team*), leadership relations (*Thatcher’s party*; *de Klerk’s government*), locative relations (e.g., *Macao’s government*) show that the speaker refers to the whole (*team*, *party*, *government*) but not the constituents of the collection. When the CN is located through the member, the whole collective body is referenced. Also, leadership relations which are usually institution sanctioned render the CNs institutions, instead of individuals. A similar case will be made

for locatives, the possessor which is the location serves as the base of the institution but not the people in it so that the institution, *Liverpool's team* is based in Liverpool is a plausible reading even if all the members of the team live outside Liverpool. Following Jessen and Vikner (2003; 2002) who classified possession into two main groups (alienable and inalienable) for the four semantic classes of control (alienable) and inherent, part-whole, producer (inalienable) relations, the results of the present study show that membership, leadership, and locative relations are types of inalienable relations with CNs, a relation that renders CNs as institutions.<sup>41</sup> Paraphrasing such CNPs in constructions like *the party Thatcher leads*, *the government Klerk presides over*, *the team he is a member of* allows a reading of the CN as an institution or body which the possessors are inalienably linked to. And just like the use of *whole*, this holistic conception of CNs in the use of possessors renders the CNP singular which therefore explains the singular CNAgre relations in such contexts (see I in appendix A for examples).

Again, Levin (2001) noticed *all* and *half* to be very frequent triggers of plural CNAgre.<sup>42</sup> Even though this observation appears far-fetched because both *all* and *Half* form quantifier phrases where they may be heads though they appear as premodifiers. Levin's explanation that their use informs plural agreement because they designate plural portions of the collective nouns, however, is a relevant basis for the analysis here. The data revealed other linguistic resources, mostly possessives but also numerals, as denoting

---

<sup>41</sup> Inalienable possessive relations imply an inseparable relation between the possessor and the possession (Heine, 1997).

<sup>42</sup> *All* and *half* need to be taken out of these determiners because there are ambiguities in their use: they, in most cases, are head nouns with the *of dependent* postmodifiers deleted. This is what makes it sometimes difficult to distinguish between *all students* and *all of the students*.

reference to part of the CNs, suggesting plural number and determining plural agreement. Possessives that designate manager relations and leadership relations where the leader is perceived as responsible for the management of the individuals who form the collection conceptualizes the collection as a body of individuals managed or led by the possessor. The manager usually is part of the collection but his responsibility as a manager is to manager members of the collective. Since the individuals the manager discharges his administrative responsibility for are plural in number, there appears to be a pattern where the plural agreement feature is foregrounded. In example (34),<sup>43</sup> *his* refers to the manager - *Manager Steve Perryman*; therefore, the team referred to is the part of the team, the players Perryman manages and they are understood as the subject. Also, in example (35), *Ron Atkinson* is the manager and *team* refers the players who form an aspect of the team – the people he manages.

In the data, it appears that numerals reveal the part-whole reading of the CNP. Numerals either enumerate the total number (holistic conception) or a part of constituents in the collection (part conception). In example (37), the referent of the CNP, *The 80-strong party* is not every individual member but an a fraction of them.

37. With 125 people on board – students, teachers, education officials – the vessel left Norway for Aberdeen and called at Lerwick before returning to Bergen during the ten-day itinerary .... *The 80-strong party* were given a briefing on Sullom Voe’s environmental care programme at the Fraser Peterson Centre before being taken on a tour of the terminal site.

---

<sup>43</sup> See examples in section IV of appendix A.

The 80 people taken out of the 125 members of the party point to a part of the whole constituents. When a speaker focuses their attention on the part of a CN their attention is directed to number, leading them to interpret the CNP as plural which would trigger plural agreement.

The part-whole constraint needs to be distinguished from the notional factor that determines CNAgre. As seen in Quirk et al. (1985), the consideration of the collection as a whole or reference to the members of the collection forms the basis for the notional distinction. The finding here differs from this understanding. On the one hand, whole as construed by possessives is used to define instances where certain possessive relation as an institution. On the other hand, the result of the present study shows part of the CNs may be referenced and the plural number feature of the parts constrains agreement. This is different from the notional view that reference to the individuals determines agreement. That is, the reference in part conception is not just to the individuals but to a part of the individuals that form the collective.

#### IV. **Definiteness**

The data showed instances where indefinite CNPs were seen as increasing in preference for singular agreement. The first pattern was found of the use of the indefinite article *a*. Given contexts where there were other modifiers such as premodifiers (“an *energetic* team” and “a *LOCAL* family” [see 38 & 39 in IV of Appendix A]) or postmodifiers (“a team *of six from Weatherall’s London office*, e.g. 37), the indefinite article did not explicitly impact agreement choices. Yet, when the collective nouns combined with only the indefinite articles (e.g. *a team & a government* – see examples of this as 40 and 41 in Appendix A, IV), the singular agreement choice was most preferred.

In explaining this observation, indefinite CNPs are seen as denoting institutions that are not known by the interactants: predicates of constructs with indefinite CNPs as subjects would be about the unknown body. This brings to the fore the genericity of indefinite referents as they could refer to any collection at all (e.g. *a committee* would thus refer to any committee [institution] at all). The singular number choice then lies in the total shift of focus from the constituents of the collection to the institution or body and this is what constrains singular agreement.

In the presence of other modifiers in addition to the indefinite article, CNs are further defined and the definition sometimes specifies that the focus of the CNP is on the constituents which is why usages of such nature do not distinguish singular or plural agreement choices. For example (38 in appendix), *an energetic team* allows an interpretation that the *team* comprises of an indefinite plural number of individuals and such would prefer a plural agreement relation and the choice would depend on *energetic* because it contributes humanness quality to *team*.<sup>44</sup> Levin (2001) noticed that certain determiners conceptualized abstract reference which is why they prefer singular agreement. Like earlier studies that found articles to lack a clear pattern in constraining collective noun agreement choices, Levin (2001) also concluded that articles do not determine CNAgre. The observation in this study differs because it has taken indefiniteness as a constraint of CNAgre but not determiners and shown that indefiniteness in the CNP typically constrains singular agreement choice.

---

<sup>44</sup> Gardelle (2019) has shown foregrounding the human constituents of CNs leads to a plural reading of the noun.

Unlike indefinite CNPs, definite CNPs did not have any pattern of preference for either singular or plural agreement but it still reveals interesting features of collective nouns and their agreement relations. The definite article is not considered a marker of definiteness in this study but proper noun modifiers share their properhood feature with CNs they modifier to define CNs.<sup>45</sup> Expressions like *the Turkish government*, *the Royal family*, *the Labour Party*, *the England team*, *the Plowden committee* (all taken from BNC) were preferred designation of properhood. The singular agreement choice is possible because the definition of the CNP such as *The Barlow family* in examples 43 (see section V of appendix A), defines an institution as the institution still stays in Cambridge. When the plural agreement is preferred in the contexts where the CNP is definite (such as example 47, see appendix), there must be further evidence in the predicate to suggest that the reference is to a plural number of people but not to the single institution. It will be unlikely to conceive a team as residing at a hotel but very likely for the members of a team to conceived as residing at a hotel, this explains the plural agreement choice in “the Scottish Football team were residents in the hotel”.

#### V. Existing Factors

The data also point to existing factors that have been identified as determining CNAgre. Two of such factors were found: number features of predicate and humanness. These factors as seen in the data are presented below and they are compared with existing explanations.

---

<sup>45</sup> See Breban et al. (in press) define PNM as use of proper nouns to modify other nouns. literature review on proper noun modifiers

## Number Features of Predicates

Predicates are said to have number features (Model semantics). Plural predication is distinguished from singular based on the semantic features of verbs, verb phrases, pronouns (like reflexive pronouns *themselves*, *myself*), and other complements. For example (48), the plural verb ‘are’ is selected based on the plural nature of the complement “country people” as the expression (*\*The Royal family is country people*) that assigns singular verb choice for the predicate will be ungrammatical.

48. The Royal family *are country people*.

49. Honestly, Les, this team is the biggest load of rubbish we have had there since I’ve been a supporter .... (FRS 250)

Similarly, the predicate of example (49) would typically not allow plural agreement; the complement of the verb “the biggest load of rubbish” cannot first be used to describe a plural number of people. Identifying the number features of predicates as a predictor of CNAgre is motivated by the study’s alignment with model semantics theory that number is not merely a feature of subjects but of predicates. It also provides a basis to explain Annala (2008), Swan (2005), Leech and Svartvik (2002), Levin’s (2001) attempt to show how verb phrases determine CNAgre.

## Humanness

Animacy has variously been seen as determining CNAgre (Fernández-Pena, 2017b; Depraetere, 2003; Levin, 2001; Nixon, 1972). In this study, I studied CNs in the human category (Gardelle, 2019) therefore animacy is narrowed to humanness which refers to the feature of a CNP in designative either human constituents or an institution. It is the linguistic context which

would show that the human constituents are accessed and spoken of (Joosten et al., 2007). In perceiving the CN as an institution, the singular would be preferred because the institution is taken as a singular entity. In addition to the parameter identified in this study as indicating humanness feature of CNPs, the data confirmed parameters such as adjectives (Joosten et al., 2007) and the relative pronouns (Depraetere, 2003) reveal whether or not their antecedents are human or an institution.<sup>46</sup> In the data for the analysis, I have identified possessive relations to determine humanness feature of CNPs. The following examples will substantiate the claims I have made.

In example (50), the adjective *terrified* shows that human constituents are referenced since being terrified is a quality of humans. In example (51), the adjective *nuclear* assigns an institution reading to *the nuclear family* as it only renders reference to the family as a non-human institution.

50. *The terrified family* were handcuffed, had pillows pulled over their heads and were held at gunpoint overnight. (CEN 6480)

51. *The nuclear family* is not only smaller, as it is based essentially on the marriage bond and children, but it encourages privatisation. (CCE 1634)

It is possible to have the humanness feature of the CNP revealed in the context even without the presence of such modifiers. For example (51), the predicate “is the matrix of identity” suggests that *family* refers to an institution, a non-human institution.

---

<sup>46</sup> Joosten et al. (2007) have shown how adjectives like *old* and *big* enable animate members who form a collection to be accessed or not to reveal the number feature of the CNP. For example, *big family* and *big team* encode reference to non-human institutions but not animate constituents so that a singular number is more likely to be associated because the institutions are singular.

52. ‘The family’, says Minuchin, ‘is the matrix of identity.’ (CEE 47)

53. If the new government — whoever heads it — wants to avoid an embarrassing, inflationary and recession-reinforcing rise in mortgage rates, it must allow societies to turn increasingly to the international money markets. (AJH 312 )

*New* in example (53) renders the government non-human. *The new government* refers to the *new* institution designated for the administration of society. Also, the adjective *most common* and *quality* make *team* (in 55 and 56) institutions.

54. *The most **common** team* is the manager and his direct reports or subordinates. (EAA 2284)

55. And *the **quality** team* was an M S two and an M S one, and it's now only an ....(FUK 1643)

As further shown in example (55), the pronoun ‘it’ confirms that “the quality team” refers to an institution. This also is indicative of the possibility for the conclusion that the non-human level of humanness constrains singular agreement whereas human constrain plural agreement but this claim needs to be confirmed after the inferential analysis in the next section.

The analysis and discussion In this section, four predictors of CNAgre have been identified. Two existing factors given in the literature also have potentials in determining CNAgre. The four predictors are accessibility, part-whole, distance and definiteness. It has been seen that the linguistic context in which CNs occur sometimes specify the number feature of the CN thereby constraining the choice of agreement feature in CNAgre. Because both singular and plural number features are accessible, accessibility is taken as a binary constraint of given singular and given plural.

In addition to accessibility, the data has shown that distance informs CNAgre. Proximal and distality measured in spatial, temporal and even social terms motivates the choice of singular or plural agreement. The assumption underpinning this constraint is that one's relative proximity to the collection shows whether he sees the unit perceives a plural number of people in the group or a singular number. Furthermore, the part-whole constraint is seen to involve instances where modifiers point to some members of the CN or the whole unit of the CN. Possessives and numerals are modifiers that help construe such part-whole relations. The analysis also points to definiteness as a predicting CNAgre. Indefinite CNPs refer to unknown institutions and trigger singular agreement just as definite CNPs would prefer singular agreement because a singular institution is usually defined in the construal of definite CNPs. However, features of the predicator could affect the singular agreement choice for definite CNPs.

The study has confirmed that predicates may have number features that constrain number agreement relations for collective nouns (de Vries, forthcoming; Katzer, 2012). It has also confirmed that humanness feature which indicates whether the human constituents of the CN are the referent or the non-human institution informs agreement choices (Gardelle, 2019; Joosten, 2010; Levin, 2001). What is left in this analysis and discussion is to determine the strength of these factors in their constraint of agreement. That is, the analysis and discussion further ascertain whether the identified factors are significant determinants of CNAgre and can predict CNAgre relations.

### Relative Strength of the Factors that Constrain CNAgre

The previous section of the analysis and discussion identified factors that determine CNAgre in response to the second research question: “what are the factors that determine CNAgre?”. This section analyzes the data to respond to the third research question which is “what is the relative strength of the factors that determine CNAgre?”. This section, therefore, seeks to establish how effective the determinants are in determining CNAgre. I first briefly describe the constraints, highlighting the coding processes before the actual analysis of the strengths.

#### *Coding*

Six explanatory variables were considered as determining collective noun agreement and the 6467 clauses of CNAgre involving *committee*, *family*, *government*, *team* and *party* were manually annotated for these variables. Each clause has undergone a series of coding process so that when multiple factors inform the choice of agreement feature in a particular clause, the clause was assigned to multiple factors. The explanatory variables include the adaptation of the humanness (previously animacy) factor which has been identified in the literature and the adoption of the number features of predicates factor. These two are added to the constraints which this study has identified (*accessibility*, *definiteness*, *distance* and *part-whole*). All these predictors were seen as operating at binary levels; Table 10 summarizes the factors.

**Table 10: Variables that determine CNAgre**

Constraining Variable	Level of Constraint
Definiteness	Definite
	Indefinite
Distance	Distal
	Proximal
Humanness	Human
	Non-human
Predicate	Singular
	Plural
Part-whole	Part
	Whole
Accessibility	Singular given
	Plural given

I now briefly describe the parameters used in coding the data.

**Accessibility:** Accessibility, as seen earlier, refers to the tendency to have the number features of the collective noun given in the linguistic context. Tokens of this constrain will be appositives that construe the number features of CNs including the appositive use of *NPs* and nominal appositives. Also, the number features of nouns that cataphorically refer to the CN are taken as providing access to the number features of CNs.

**Definiteness:** A CN is definite when its referent is conceived as familiar and/or identifiable to both the speaker and addressee (Ionin, 2006; Heusinger, 2002; Lyons, 1999). Proper noun modifiers define CNPs. CNPs with the structure *indefinite article + head* are taken as indefinite.

**Humanness:** Humanness denotes reference to the human constituents of the CN or the institution, which is non-human. Parameters which indicate humanness

include adjectives that denote human qualities (Joosten et al., 2007),<sup>47</sup> the relative pronouns (Depraetere, 2003) and possessive pronouns. In coding therefore, tokens that have adjectives, possessives, and relative clauses that reveal the humanness of CNs were coded as instances of humanness feature informing CNAgre.

**Distance:** The relative distance between the speaker or the interlocutors and the collective noun may allow either a plural reading or a singular reading of the CNP. Distance is construed as epistemic, spatial, temporal, social etc. (Kolkman, 2016; Rybarczyk, 2015) but not as the syntactic relations measured by the number of words between the head of the CNP and the unit it agrees with (as seen by Fernández-Pena, 2017b; Levin, 2001; Corbett, 1979). To code distance, demonstratives determiners, and adverbs that encode distance were considered.

**Number features of predicates:** [d]e Vries (forthcoming) shares in the various positions (including Joosten et al., 2007; Pollard and Sag, 1994; Baker's, 1992) that the predicate has number features that may restrict the number agreement relation for CNs in a similar manner as the number features of CNPs constrain CNAgre. The coding here thus follows the understanding that when properties of the group itself or the members of the group are expressed by the predicate,<sup>48</sup> singular and plural number would be constrained respectively (de Vries, forthcoming; Pollard & Sag, 1994). Predicates that show quantity of their

---

<sup>47</sup> Joosten et al. (2007) have shown how adjectives like old and big<sup>47</sup> are helpful in allowing the animate members who form a collection to be accessed or not to reveal the number feature of the CNP. For example, big family and big team encode reference to non-human institutions but not animate constituents so that a singular number is more likely to be associated because the institutions are singular.

<sup>48</sup> Pollard & Sag (1994) believe that the predicate relates with groups but not individuals and therefore the plural agreement for the structure, *\*A new committee have been constituted* renders the structure ungrammatical.

subject will, therefore, be taken as informing agreement so that those that show singular number will be taken as the singular predicate level and those that display plural number would be taken for plural predicates. Ways in which these predicates construe number are by having plural complements.

To conduct the logistic regression analysis, I coded the independent variables in binary mode. That is, 0 and 1 were respectively assigned for singular and plural agreement relations. The predictor variables were also coded using 1 and 2. Humanness was coded as 1 = *human* and 2 = *non-human*; definiteness was coded as “1=definite” and “2=indefinite”; distance as “1=part” and “2=whole”; accessibility: “1=given singular” and “2=given plural” and finally number feature of predicates (predicate for short) was coded as “1=singular predicate” and “2=plural predicate”. All these were coded in excel before they were imported to SPSS for analysis. The binary logistic regression was then run because the expected outcome of the independent variables was two (Bresnan, et al, 2007). This coding was modelled with binary logistic regression for each of the CNs (see table 11 for results). The results of the regression analysis are presented and discussed in turns below. Table 12 is the result upon testing the significance of the predictor variables in determining singular and plural agreement for *government*.

**Table 11: Coded Results of Predictors of CNAgre in the 6467 Clauses**

CN	Animacy		Definiteness		Distance		Part-whole		Accessibility		Predication		Total
	Animate	Inanimate	Definite	Indefinite	Proximus	Distal	Part	Whole	Given Singular	Given Plural	Singular Predicate	Plural Predicate	
Gov't Singular	13	1002	540	106	3	31	0	104	1200	5	400	0	3404
Gov't plural	19	4	40	0	10	0	3	1	3	42	0	35	157
Team singular	5	593	343	712	0	24	12	750	906	0	633	0	3978
Team Plural	147	0	205	0	20	0	216	20	2	354	0	234	1198
Family Singular	10	512	115	198	0	23	19	51	260	0	252	0	1140
Family Plural	172	14	155	0	2	0	69	6	0	142	0	130	690
Com't Singular	0	412	389	216	7	25	2	322	705	0	859	0	2792
Com't Plural	72	4	62	0	3	0	105	6	0	62	0	72	540
Party Singular	45	412	489	316	0	12	2	322	605	0	68	0	1662
Party Plural	17	2	19	0	4	0	10	1	0	11	0	17	690
Total	500	2955	2357	1548	49	155	438	1583	3681	616	2212	488	16254

**Table 12: Logistic Regression Analysis of Strength of Predictors of Agreement Relations for Committee: Full Model with the Predictor Variables, Beta Co-efficient, Standard Error, P-Value and the Odds Ratio**

Variable	B	S.E	Wald	Df	Sig.	Exp(B)
Accessibility						
Given singular	-	-	-	-	-	-
Given plural	-2.430	.132	336.373	1	.000	.088
Definiteness						
Definite	19.364	2734.791	.000	1	.994	256819062
Indefinite	-21.203	2734.791	.000	1	.994	.000
Distance						
Proximus	20.356	8038.594	.000	1	.998	692346409
Distal	-	8038.594	.000	1	.998	.000
	21.203					
Humanness						
Human	25.840	4770.028	.000	1	.996	166797779896
Non-human	-4.637	.502	85.188	1	.000	.010
Part-whole						
Part	7.947	.824	92.965	1	.000	2826.250
Whole	-3.986	.412	93.586	1	.000	.019
Predication						
Singular predicate	-	-	-	-	-	-
Plural predicate	-2.478	.123	407.86	1	.000	.084

6

**Committee**

The alpha levels (sig) (of 0.00, see Table 12) shows that construing part and whole relations for *committee* (i.e., part-whole constraint), providing plural number features in the context in which committee is used, designating

*committee* as non-human, and having a plural number feature of predicates are significant in determining agreement choices for *committee*. There are higher odds ( $\text{Exp}[B]$ ) of part constraining plural agreement than singular agreement and of whole constraining singular agreement than plural agreement. Also, a plural predicate is very likely to predict plural agreement choices for *committee* than singular agreement choices and singular predicate does not predict agreement for *committee*. Also, when plural number features of committee are given in context (i.e., given-plural), there is a high tendency for committee to engage in plural agreement. Definiteness at both definite and indefinite and distance at both the proximal and distal levels are not significant in determining agreement relations involving *committee* as they have p-values greater than .05, the alpha level.

### ***Family***

In Table 13, the logistic regression results reveal that referring to the human constituents or the non-human institution when the referent term *family* is used influences its agreement choices. Human is more likely than non-human to predict plural agreement and non-human has a higher probability of predicting singular agreement. With p-values of .00, social-attitudinal proximity influences agreement choices of *family*, having a higher probability of predicting plural agreement than singular agreement. The context of use as seen in the predicate and co-referential resources used before the agreement relations usually indicate plural agreement choices of *family*. Singular number is however not often given in context as given-singular is not significant in determining agreement for *family*. Also, the proximal level of distance and singular predicate level of predicate have no impact on agreement choices of *family*.

**Table 13: Logistic Regression Analysis of Strength of Predictors of Agreement Relations for Family: Full Model with the Predictor Variables, Beta Co-efficient, Standard Error, P-Value and the Odds Ratio**

Variable	B	S.E	Wald	Df	Sig	Exp(B)
Accessibility						
Given singular	-	-	-	-	-	-
Given plural	-.578	.105	30.340	1	.000	.561
Definiteness						
Definite	21.493	2856.387	.000	1	.994	2158601839
Indefinite	-21.203	2856.387	.000	1	.994	.000
Distance						
Proximus	-	-	-	-	-	-
Distal	-2.442	.737	10.976	1	.001	.087
Humanness						
Human	6.520	.430	230.193	1	.000	678.738
Non-human	-3.675	.281	171.264	1	.000	.025
Part-whole						
Part	3.503	.506	47.950	1	.000	33.222
Whole	-2.159	.431	25.086	1	.000	.115
Predicate						
Singular predicate	-	-	-	-	-	-
Plural predicate	-.666	.108	38.074	1	.000	.514

**Government**

The regression model of factors that constrain agreement highlights humanness at both levels of human, non-human, whole of part-whole, given singular of accessibility, and plural predicate of number features of predicate as significant in determining agreement relations for *government* as they have p-values (Sig) less than the established alpha level of 95% (0.05). The p-values of both definite and indefinite levels of definiteness, part of part-whole, given plural of accessibility and distance at both proximal and distal levels exceed 0.05, the alpha level and are therefore not significant predictors of agreement relations for *government*. From the regression model, there is also a positive impact on agreement relation with *government* for a unit increase in human of the humanness constraint and the odds of human predicting singular agreement for *government* is greater than predicting plural agreement for *government*.

Non-human negatively affects agreement relations of *government* and its odds of predicting plural agreement for government is lower than singular agreement. Under part-whole, whole has negative impact on agreement choices for *government* and its odds of predicting singular agreement for *government* is lower than predicting plural agreement for *government*. At the level of given-singular of accessibility, there is a negative impact on constraining singular agreement for government and the odds of predicting singular agreement for *government* is lower than predicting plural agreement for *government*. Plural predicate impacts agreement choices of *government* negatively and the odds of plural predicate predicting plural agreement for *government* is lower than predicting singular agreement.

**Table 14: Logistic Regression Analysis of Strength of Predictors of Agreement Relations for Government: Full Model with the Predictor Variables, Beta Co-efficient, Standard Error, P-Value and the Odds Ratio**

Variable	B	S.E	Wald	df	Sig.	ExP(B)
Accessibility						
Given singular	-6.733	1.239	29.555	1	.000	.001
Given plural	-0.511	.730	.489	1	.484	.600
Definiteness						
Definite	18.669	3885.609	.000	1	.996	128164970.945
Indefinite	-21.203	3885.609	.000	1	.996	.000
Distance						
Proximal	22.502	7105.180	.000	1	.997	5923407975.632
Distal	-21.203	7105.180	.000	1	.998	.000
Humanness						
Human	5.879	.612	92.178	1	.000	357.607
Non-human	-5.574	.501	123.810	1	.000	.004
Part-whole						
Part	25.866	23205.4	.000	1	.999	171240335777.216
Whole	-4.663	1.005	21.544	1	.000	.009
Predication						
Sin. predicate	-	-	-	-	-	-
Plu. predicate	-2.168	.156	193.976	1	.000	.114

**Team**

The logistic regression modelling of the strength of the factors that predict agreement choices for *team* are presented in Table 15. Amongst the

twelve constraining factors, the regression model shows part-whole as significant in determining agreement relation for *team*. At both levels of part and whole, part-whole informs agreement choices. Humanness (at both human and non-human levels), definiteness (at indefinite and indefinite levels), distance (proximal and distal), accessibility (given singular and plural) are not significant predictors of agreement relations involving *team* as the subject NP. The variable ‘part’ has a positive impact on *team* and the odds of predicting singular agreement for *team* is higher than predicting plural agreement. There is a negative impact of whole on predicting agreement choices of *team* and the odds of predicting singular agreement relation for *team* is lower than predicting *team* plural. Also, plural predicate has a negative impact on agreement choices for *team*.

**Table 15: Logistic Regression Analysis of Strength of Predictors of Agreement Relations for Team: Full Model with the Predictor Variables, Beta Co-efficient, Standard Error, P-Value and the Odds Ratio**

Variable	B	S.E	Wald	df	Sig.	Exp(B)
Accessibility						
Given singular	-27.319	2136.232	.000	1	.990	.000
Given plural	21.203	2136.232	.000	1	.992	1615474721
Definiteness						
Definite	20.685	1506.298	.000	1	.989	962708253
Indefinite	-21.203	1506.298	.000	1	.989	.000
Distance						
Proximus	-	-	-	-	-	-
Distal	-0.182	0.303	0.363	1	0.547	0.833
Humanness						
Human	24.584	1649.138	.000	1	.988	47494974485
Non-human	-21.203	1649.138	.000	1	.990	.000
Part-whole						
Part	6.516	.373	304.813	1	.000	675.900
Whole	-3.626	.227	256.090	1	.000	.027
Predication						
Sing. predicate	-	-	-	-	-	-
Plu predicate	-.994	.077	168.582	1	.000	.370

**Party**

From results in Table 16 only distal, definite and indefinite are insignificant predictors of CNAgre involving *party*. The remaining predictors, namely: part and whole, given plural, plural predicate, human, and non-human significantly impacts the choice of agreement features for *party*. The results further show that perceiving part of the constituents of a party positively impacts *party*'s agreement relation.

**Table 16: Logistic Regression Analysis of Strength of Predictors of Agreement Relations for *Party*: Full Model with the Predictor Variables, Beta Co-efficient, Standard Error, P-Value and the Odds Ratio**

Variable	B	S.E	Wald	df	Sig.	Exp(B)
<b>Accessibility</b>						
Given singular	-	-	-	-	-	-
Given plural	-4.006	.304	173.343	1	.000	.018
<b>Definiteness</b>						
Definite	17.846	2257.463	.000	1	.994	56276810.0
Indefinite	-21.203	2257.463	.000	1	.993	.000
<b>Distance</b>						
Proximus	-	-	-	-	-	-
Distal	-	.577	3.621	1	.057	.333
<b>Part-whole</b>						
Part	1.099	7.387	1.266	34.040	1	.000
Whole	-	1.002	33.278	1	.000	.003
<b>Predication</b>						
Sing predicate	-	-	-	-	-	-
Plu predicate	-	.272	25.505	1	.000	.254
<b>Humanness</b>						
Human	4.377	.764	32.805	1	.000	79.591
Non-human	-	.709	56.498	1	.000	.005
	5.328					

The odds that reference to part of the constituents of *party* will impact singular agreement choice is higher than its odds of impacting plural agreement. Whole is a negative indicator of agreement relations with *party*. For a unit increase in whole, the odds of whole impacting singular agreement relation for *party* is lower than impacting plural agreement. Distal impacts agreement

relations with *party* negatively and the odds of impacting *party* singular is higher than the odds of impacting plural agreement. Also, given plural negatively impacts agreement relation of *party* and the odds of it impacting singular agreement for a unit change is lower than impacting on plural agreement. Plural predicate negatively impacts on agreement relation involving *party* and the odds ratio of plural predicate impacting singular agreement choice is lower than impacting plural agreement. Human and non-human have positive and negative impacts on agreement choices of *party* but the odds of human informing singular agreement choice for *party* is higher than impacting plural agreement more than non-human. Finally, the impact of definite in determining agreement choices of *party* is positive while indefinite has a negative impact. The odds of definite impacting singular agreement choice for *party* is higher than impacting plural agreement. Definite has no predictability for agreement relations with *party* since the odds ratio is zero.

### **Discussion**

The logistic regression analysis has examined the strength of determinants of agreement with collective noun headed subjects. The determinants, accessibility, animacy, definiteness, distance, number features of predicates, and part-whole are assessed to show their strength in constraining singular and plural agreement choices for *government*, *team*, *committee*, *family* and *party*. From the results, part-whole influences agreement relations with all the CNs. The logistics regression analysis shows that it is more likely for part of part-whole to predict plural agreement for *committee*, *family*, *team* and *party* but less likely for part to predict singular agreement for any CN or predict plural agreement for *government*. Whole of part-whole is, however, more likely to

determine singular agreement choice for all the collective nouns. The finding that reference to part of the constituents designate plurality has not been reported by existing studies. The current study, therefore, adds to existing literature that the presence of linguistic resources such as possessives and numerals that indicate reference to part of the entities forming *committee*, *family*, *team*, and *party* would usually constrain plural agreement relations with CNs.

Humanness significantly influences agreement relations with *committee*, *family*, *government* and *party*. From the results of the odds of prediction, human is more likely to determine plural agreement choices for *family*, *government* and *party* and non-human usually determines the choice of singular agreement for *committee*, *family*, *government* and *party*. The present study has confirmed that when relative clauses and adjectives signal the humanness of the collective nouns, agreement choices are defined (Gardelle, 2019; Levin, 2001) and also adds possessives and co-texts as important determinants of humanness. The negative impact of humanness on agreement relations with *team* highlights the differences in the accessibility of humanness feature of certain CNs, confirming the findings of Joosten (2010) and Gardelle (2019) that accessibility of human entities of CNs depends on the CN involved.

The binary regression analysis also shows a higher likelihood for plural number features (that is, given-plural of accessibility predictor) to be given in linguistic contexts where *family*, *committee* and *party* are used. However, the singular number feature is less likely to be stated in the context since only the singular agreement relation with *government* is determined by given-singular of accessibility. This means that interactants through the use of co-referential

linguistic resources such as appositives and cataphoric references signal plural number features for some CNs (*family, committee, and party*) even before they are used as subject CNPs and such number features usually determine plural agreement relations for the CNs. The current study is the first to show that the plural number feature of CNs exist in the linguistic context.

The number features of predicates factor also has similar results as that of accessibility. While plural predicates are significant determinants of agreement choices for *party, family* and *government*, singular predicates do not influence agreement choices of any of the CNs studied. These two findings (of the strength of the accessibility and number features of predicate predictors) suggest the high tendency for the plural number feature to be determined by the context of use than the singular number. After finding that existential *there* subjects constrain plural agreement choices for anticipatory CNPs, Smith (2015), Sauerland and Elbourne (2002), and Den Dikken (2001) have concluded that plural agreement choices are constrained in specific contexts than singular agreement. The results of the present study supports the conclusion by showing that plural number feature is usually constrained by the context of use.

Distance significantly influences singular agreement for *party* and *committee* at the distal level. The proximal level is not significant in determining agreement choices for any CN. This shows that when a speaker construes a distal relation (in terms of social-attitude) for *party* and *committee*, it is more likely for singular agreement choices to be made for *committee* and *party*.

The logistic regression analysis has shown that different factors have different likelihoods of informing agreement choices. Although whole of part-whole is more likely to determine singular agreement for all the CNs, non-

human of humanness influences the chances of singular agreement choice for *government*, *family*, *committee* and *team* but not for *party* and given-singular of the accessibility predictor is also a significant predictor of singular agreement for *government* and *family*. This observation contests established claims that the singular agreement is predicted when one refers to a CN as a whole (of course the whole conception in this study is not the same) (Depraetere, 2003; Levin, 2001; Quirk et al., 1985). The findings, however, shows that depending on the collective noun involved, other factors will determine singular agreement choices. The observation is also evidence to support the probabilistic view of language (Bod, 2010). That is, instead of the category ‘whole’ constraining agreement for CNs to signal that CNAgre is determined by the categories part and whole, other factors also inform singular agreement choices, highlighting the gradient nature of language. Gradience is not only seen in the factors that inform CNAgre but the collective noun system itself. This is highlighted in plural agreement choices of the various CNs.

For plural agreement, none of the predicting factors significantly determine agreement for all the CNs. This reveals variability in the CNs. Part of the part-whole constraint significantly influences plural agreement choices for *team*, *family*, *committee*, and *party* but does not determine plural agreement for *government*. A plural predicate is more likely to determine the choice of plural agreement for *government*, *committee* and *party* but not for *team* and *family*. Given plural number features are significant determinants of plural agreement for *family*, *committee*, and *party* but not for *government* and *team*. In addition to showing differences among the CNs, this variability presents a justification for the sub-categorization of CNs as done by Lakaw (2017).

## CHAPTER FIVE

### SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

#### Overview

This final chapter of the study summarizes the thesis by drawing conclusions and making suggestions for further studies. I will restate the need for identifying nuances that explain agreement relations with collective nouns as the niche the present study set out to occupy, reflection on how the gap in knowledge as discussed in Chapter One has been filled. Recommendation for further studies will also be provided in this chapter to guide future research.

#### Summary of the Study

The principal aim of this study has been to establish the probabilistic grammar of collective noun agreement. Based on the theoretical position and the problem statement, I stated three objectives. These are to (1) show how the distribution of agreement forms for CNs provides explanations for choices in CNAgre; (2) identify the factors that determine CNAgre and (3) define the relative strength of the predicting factors in constraining CNAgre. The purpose of this study is worthwhile as it finds its significance in identifying subtle features that motivate singular and plural agreement choices for collective nouns. Agreement and collective nouns were relevant concepts reviewed in chapter two. In the same chapter, I discussed the theory rich perspective the study draws from probabilistic linguistics (PL) theory. PL's position that language is gradient and grammar must provide the gradient choices involved in the description of any linguistic phenomenon justifies the presence of many factors that determine CNAgre and the tendency for some to be significant determinants of CNAgre. In Chapter Three, five collective nouns were selected

as samples from the over 200 CNs (Fowler, 1992) to be studied. Also, I describe the data I used for the analysis. In the analysis and discussion in chapter four, I identified the factors that inform CNAgre and determined their relative strengths. These findings are summarized in the next section.

### **Key Findings**

For the first research question, the study found priming to explain the choices of plural agreement relations. The results show that plural agreement choice once selected in a text with several agreement relations increases the chances of subsequent plural agreement in that text. Therefore, the plural agreement relation is primed. On the other hand, the choice of singular agreement in a single text did not constrain subsequent choices in that text. This reveals the entrenched nature of the singular agreement choice as compared to the plural agreement choice. Entrenchment (like markedness) explains linguistically coded archetype whose choice requires less cognitive efforts (Claes, 2017; Langacker, 1991).

In responding to question two, the findings of the study revealed six nuance features that determine collective noun agreement choices. These are accessibility, definiteness, distance, humanness, part-whole, predication. The factors included an addition to the humanness and number features of predicates as identified in the literature (see de Vries, forthcoming). This study identified possessive marking as a parameter in determining the humanness feature of CNs which is an addition to the use of relative pronouns and adjectives (Depraetere, 2003; Levin, 2001). In addition to the two, the data revealed other factors like accessibility, definiteness, distance and part-whole as influencing CNAgre. The present study is the first to have identified these as informing CNAgre.

The study found accessibility which refers to the cases in which the number features of the CN is given in the linguistic context to inform the choice of agreement features. The finding that accessibility determines CNAgre has further expanded Joosten et al.'s (2007) understanding that the members of CNs are accessible for some CNs than others. It also adds to the cognitive linguistics understanding that givenness informs linguistic choices (Claes, 2017). Numerals and possessives show whether part of the constituents of the collection or the whole are referred to. Singular and plural agreement choices therefore preceded knowledge of reference to part or whole which indicates the language user's ability to distinguish number features of the CN. Furthermore, distance, defined in terms of emotion, social relation, spatial, temporal relations, informs CNAgre. What motivated the identification of this factor was the observation that the more one was proximal, the more likely the language user gets at construing the individuals who form the collection. Lastly, when definiteness is marked, the collective nouns appeared to encode a singular institution and informed agreement choices. The findings here prove that language use embodies choices which are motivated by several factors (Bod, 2010) and that these factors stand out when one closely observes discourse data (Ngula, 2014; Bod, 2010).

In answering the last research question, I used the logistic regression model to analyse the 6467 clauses extracted from the BNC. I coded the clauses by the factors that determine their agreement relations. The coded data was analysed in the SPSS statistical tool and the results outline the gradient grammar of CNAgre.

The regression analysis revealed that part-whole, humanness, accessibility and number features of predicates are significant determinants of collective noun agreement choices whereas distance and definiteness were not significant predictors. The observation shows that choice is determined by several factors which have relative strengths in predicting agreement choices (Bod, 2010). Their significance varies across different CNs and the agreement feature involved as well as at different levels of prediction. For singular agreement, whole is more likely to define singular agreement relations for all the CNs under study. Also, the study has found that non-human is a significant predictor of singular agreement for *government*, *committee*, *family* and *team*. Lastly, given singular is a significant predictor of singular agreement for *family* and *government*. In terms of singular agreement therefore, whole, non-human, and given-singular (when the singular number of the CN is given in context) emerged as a significant determinant of agreement. This means that multiple of factors influence the choice of singular agreement.

For plural agreement, the results of the regression model show that part of the part-whole constraint of agreement determines agreement for *team*, *family*, *committee* and *party* but it is an insignificant predictor of agreement relation for *government*. Moreover, plural predicate was a significant predictor of singular agreement of plural agreement for *government*, *committee*, and *party*. Lastly, “given-plural” (when plural number feature of CNs is provided in the context) informs agreement choice for *party*, *committee* and *government*. Part is more likely to determine plural agreement choice for *team*, *family* and *committee* and human has a higher likelihood in terms of predicting agreement relations for *government* and *party*. This shows that in any data set, there is a

high probability for *government* and *party* to be conceptualized as human entities and such will predict their plural agreement choices.

The agreement relations for collective nouns, from the results, are choices made following preemption and markedness. Markness explains how singular agreement is entrenched and is selected regardless of the absence of any factor positively impacting its choice. This study has thus contributed to a probabilistic grammar for CNAgre.

### Conclusions

The following conclusions are drawn from the key findings.

1. For plural agreement choices are often primed while singular agreement choices are entrenched. At the underlying level, collective noun agreement is informed by cognitive constraints.
2. At least six factors are involved in constraining collective noun agreement: accessibility, definiteness, distance, humanness, part-whole, predication. These factors are mutually inclusive in determining agreement.
3. Part of the part-whole and human of humanness determinants of agreement have the most effect on predicting plural agreement for many of the CNs studied. Whole influences singular agreement for all CNs studied.
4. CNs system displays gradience in their characteristics as different factors impacted their agreement relations differently.
5. Gradience also manifests in the strength of the factors that inform agreement choices. Different factors in different contexts impact agreement.

6. Plural agreement choices are far often determined by context than singular agreement choices. This supports Elbourne and Sauerland's (2002) conclusions that plural choices are restricted uses.

### **Implications of the Study**

The research findings and conclusions have implications for Probabilistic Linguistic (PL) theory, agreement and collective noun agreement. In the first place, the study confirms PL's position that language is probabilistic as several factors from which learners make systematic choices in determining collective noun agreement has been established (Bod, 1998). The study has shown that the choice of singular and plural agreement is informed by different factors and at different degrees. For singular agreement, it has shown that non-humanness, whole of part-whole, given singular of accessibility influences agreement. Moreover, human of humanness, part of part-whole, given plural and plural predicate inform plural agreement choices. PL's assumption that several factors exist to inform the choices of linguistic phenomena is also confirmed making the theory more suited for the study of complex linguistic phenomenon such as CNAgre. This study has also used PL's assumption that inferences could be made after testing the strength of the predictor variables. This assumption is confirmed as, for example, the study infers that part and non-human positively impacts plural agreement.

There are implications of this study for the conception of agreement. Agreement is usually defined in the view of Steele (1978) as the "systematic covariance between a semantic or formal property of one element and a formal property of another". From this definition, linguists have identified the problem about the linguistic resource that constrains agreement, resulting in debates on

semantic or syntactic perspectives to agreement. The findings of this study have shown that features of the subject, verb phrase (predicate) and linguistic contexts contribute to agreement. These features are normally defined by the linguistic context. For example, part-whole as marked by possessives is context defined,<sup>49</sup> accessibility which is construed as given number features in context is determined by context; so are the number features of the predicate and animacy. Thus, all the predictors of CNAgre are context determined. This situation makes it necessary for agreement to be conceptualized as a discourse phenomenon (Barlow, 1999). Such a conception will align with one of the first grammar of agreement,<sup>50</sup> Apollonius Dyscolus'. Dyscolus's observation that *person* (first person, second person, third person) determines agreement crucially points to the role of context in determining agreement. The observation in the present study that agreement for collective nouns is highly context-based makes it essential that the discourse perspective to agreement be taken in explaining noun agreement in future studies.

With regards to collective noun agreement, unlike previous studies and grammar texts which have mainly provided a general explanation to CNAgre, this study has offered several factors as impacting CNAgre. Collective noun agreement is therefore seen as being informed by several factors at relative degrees and significance levels. It contributes to explanations of agreement as it takes away 'doubts' associated with the need to resort to the grammatical agreement (Yankson (1994).

---

<sup>49</sup> See pragmatic agreement in chapter two

<sup>50</sup>Apollonius Dyscolus is the first discussion of agreement in English and he sees agreement as relating strictly to person: first person has agreement forms, so do second and third persons (Bod, 2013).

Another implication of the study is the relevance it has in the use of linguistic data. The subtle contextual features that determine agreement were identified because of the discourse grammar perspective that required the analysis corpus of naturally occurring data. Therefore, the relevance of discourse data to linguistics (Bod, 1998; Levin, 1993; Halliday, 1991) is further confirmed by this study.

Further, the findings have implications for sociolinguistic studies. Normally, social variables such as age, gender, region are used to explain a linguistic phenomenon. This study has not used these factors but used variables that are generally seen as informing choices in the probabilistic linguistic framework. These factors (like animacy, part-whole, distance) relate very well with the internal syntactic features of the language and offers explanations on why the choices are allowed in the language in the first place. For example, the plural agreement, from this study, could result from the humanness feature of the CNP, the context-determined plural number feature, the plural nature of the predicate, and the reference to part of the whole constituency of the CNP. It is from these internal factors that one may examine collective noun agreement from the perspective of social variables such as age, region, gender class, text type, etc.

### **Recommendations**

Using a few CNs and data from BNC, CNAgre has been studied in this work. The study was limited to five collective nouns which means further studies could be conducted on different nouns to test the findings of this study. Since Fowler (1992) mentioned that CNs are over two hundred, it would be particularly useful that a study is conducted on other CNs to either confirm or

disconfirm the findings of this study. The findings from such studies could help ground the factors identified here as determining CNAgre. Also, I encourage future studies to examine the factors identified in non-native contexts. This could help in expanding generalizability of the results to other varieties of English.

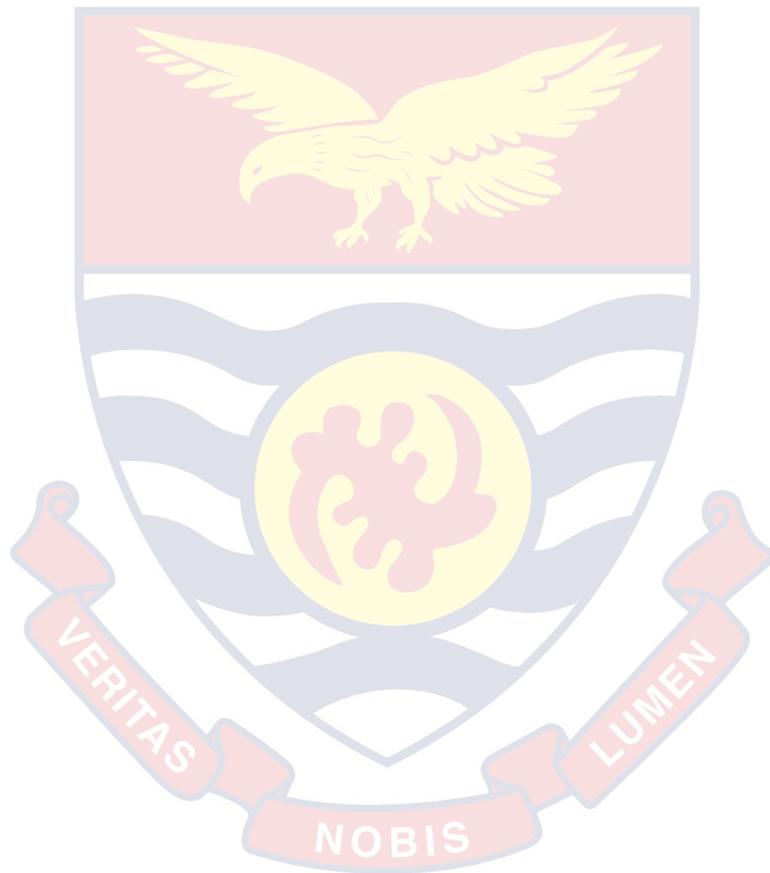
The six constraining factors relate to cognitive linguistics constraints on linguistic choices. Yet the relationship is not explicitly stated. It will be insightful for future studies to examine how cognitive linguistic constraints relate to these factors to determine agreement. Specifically, cognitive studies that investigate the brain's behaviour during CNAgre choices will be informing. It could focus on processing time, difficulty in processing agreement to explain CNAgre.

Finally, the analysis drew from research findings that possessives have several pragmatic functions. These functions help in defining referents. Yet these pragmatic functions have not been so well established. By far, only Kolkman (2016) and Rybarczyk (2015) have examined the pragmatics of possessive nouns and not related that to specific subgroups of nouns. Further studies on the pragmatics of possessives for subclasses of nouns (collective nouns preferably) could reveal how native speakers of English fix the profiles for referents through possession.

### Summary

This concluding chapter has reviewed the motivation of the study, the study objectives relative to the findings. I have shown that the results contribute a discourse grammar perspective to the collective noun agreement problem. The chapter concluded by highlighting the findings and the factors that constrain

CNAgre. I have recommended further studies to test the findings in different varieties of English.



## REFERENCES

- Acton, E. K. (2014). *Pragmatics and the social meaning of determiners*. Unpublished doctoral dissertation. Stanford, California: Stanford University.
- Acquaviva, P. (2016). Linguistic plurality and the conceptualization of part Structure. In M. Carrara, A. Arapinis & F. Moltmann (eds.), *Unity and Plurality: Logic, Philosophy, and Linguistics*, (pp. 194-218). Oxford: Oxford University Press.
- Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.
- Aikhenvald, A. (2012). Possession and ownership: A cross linguistic perspective. In Alexandra Y. Aikhenvald & Robert M. Dixon (eds.), *Possession and Ownership: A Cross Linguistic Typology* (pp. 1-63). Oxford: Oxford University Press.
- Algeo, J. (1988). British and American grammatical differences. *International Journal of Lexicography*, 1, 1-31.
- Annala, H. (2008). *Changes in subject-verb agreement with collective nouns in British English from the 18th century to the present day*. Unpublished doctoral dissertation, University of Tampere.
- Austin, J. L. (1962). *How to do things with words*. Oxford: Oxford University Press
- Baayen, R. H. (2008b). *Analyzing linguistic data: A practical introduction to Statistics*. Cambridge: Cambridge University Press.
- Baayen, H. (2003). Probabilistic approaches to morphology. In R. Bod, J. Hay, & S. Jannedy (eds.), *Probabilistic Linguistics*, (pp. 229-288).

Cambridge & Massachusetts: The Massachusetts Institute of Technology Press.

- Bach, K. (2005). Context ex Machina. In Z. Szabó (ed.), *Semantics versus pragmatics* (pp. 1-31). Oxford: Oxford University Press.
- Bailey, C. J. N. (1987). 'Marginalia' on singular and plurals in English. *Arbeiten aus Anglistik und Amerikanistik* 12, 3-11.
- Baker, M. C. (2008). *The syntax of agreement and concord*. Cambridge: Cambridge University Press.
- Barker, C. (1992). Group terms in English: Representing groups as atoms. *Journal of Semantics*, 9, 69-93.
- Barlow, M. (1999). Agreement as a discourse phenomenon. *Folia Linguistica*, 33(2), 187-201.
- Bauer, L. (1994). *Watching English change*. London: Longman.
- Bednarek, M, & Caple, H. (2017). *The discourse of news values: How news organizations create newsworthiness*. Oxford: Oxford University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited.
- Birkenes, M. B., & Sommer, F. (2014). The agreement of collective nouns in the history of ancient Greek and German. In C. Gianollo, A. Jaeger, & D. Penka, (eds.), *Language change at the syntax-semantics interface* (pp. 183-222). Berlin: De Gruyter Mouton.
- Bock, K., Cutler, A., Eberhard, K. M., Buttefield, S., Cutting, J. C., & Humphreys, K. R. (2006). Number Agreement in British and American English: Disagreeing to Agree Collectively. *Language*, 82(1), 64-113.

- Bod, R. (2010). Probabilistic grammar. In Bernd Heine, & Heiko Narrog (eds.), *the Oxford Handbook of Linguistic Analysis* (pp. 11-38). Oxford: Oxford University Press.
- Bod, R. (1998). *Beyond grammar: an experience-based theory of language*. Stanford: Centre for the Study of Language and Information.
- Bod, R., Hay, J., & Jannedy, S. (2003). Introduction. In R. Bod, J. Hay, & S. Jannedy, *Probabilistic linguistics* (pp 1-11). Cambridge & Massachusetts: The Massachusetts Institute of Technology Press.
- Bod, R., & Kaplan, R. (1998). A probabilistic corpus-driven model for lexical functional analysis. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and COLING 17 (ACL 36/COLING 17)*, Montreal, Canada.
- Breban, T. (2018). Proper names used as modifiers: A comprehensive functional analysis. *English Language and Linguistics*, 22(3), 381-401.
- Breban, T. (2013, May 22-25). A new analysis of proper noun modifiers in PDE NPs. Paper presented at *International Computer Archive of Modern and Medieval English (ICAME34)*, Santiago de Compostela.
- Breban, T., Kolkman, J. & Payne, J. (2015, July 26-31). Is the Ghana problem Ghana's problem? Differing interpretations of two English NP constructions. *International Pragmatics Association Conference (IPrA14)*, Antwerp.
- Breban, T., Kolkman, J., & Payne, J. (2019). The impact of semantic relations on grammatical alternation: An experimental study of proper name modifiers and determine genitives. *English Language and Linguistics*, 23(3), 797-826.

- Bresnan, J. (2007). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In S. Featherston & W. Sternefeld (eds.), *Linguistics in search of its evidential base* (pp. 75-96). Berlin: Mouton de Gruyter.
- Bresnan, J. (2001). *Lexical-functional syntax*. Oxford: Blackwell Publishers
- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. *Cognitive Foundations of Interpretation*, 3, 69-94.
- Bresnan, J., & Ford, M. (2010). Predicting syntax: processing dative constructions in American and Australian varieties of English. *Language*, 86(1), 168-213
- Bybee, J. L., & Hopper, P. J. (2001). *Frequency and the emergence of linguistic structure* (Vol. 45). Amsterdam: John Benjamins Publishing.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, Mass.: Massachusetts Institute of Technology Press.
- Claes, J. (2017). Probabilistic grammar: The view from cognitive sociolinguistics. *GLOSSA: A Journal of General Linguistics*, 2(1), 62.1-3.
- Coleman, J. S., & Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. John Coleman (ed.), *The Third Meeting of the ACL Special Interest Group in Computational Phonology: Proceedings of the Workshop*. East Stroudsburg, PA: Association for Computational Linguists.
- Corbett, G. G. (2019, October). *Agreement hierarchy revisited*. Paper presented at the Conference on “the Many Facets of Agreement”, Zurich: University of Zurich.

- Corbett, G. G. (2006). *Agreement*. Cambridge: Cambridge University Press.
- Corbett, G., G. (2000). *Number*. Cambridge: Cambridge University Press.
- Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett, G. G. (1983). *Hierarchies, targets and controllers: Agreement patterns in Slavic*. London: Croom Helm.
- Corbett, G. G. (1979). The agreement hierarchy. *Journal of Linguistics*, 15, 203-224
- De Vries, H. (fc). Collective Nouns. In P. C. Hofherr & J. Doetjes (eds.), *Oxford Handbook of Grammatical Number*, (pp. n.a.). Oxford: Oxford University Press.
- De Vries, H. (2015). *Shifting sets, hidden atoms: The semantics of distributivity, plurality and animacy*. Unpublished doctoral dissertation, Utrecht University.
- Dekeyser, X. (1975). *Number and case relations in 19th century British English: A comparative study of grammar and usage*. Antwerp: Bibliotheca Linguistica.
- Den Dikken, M. (2001). "Pluringulars", pronouns and quirky agreement. *The Linguistic Review*, 18(1), 19-41.
- Depraetere, I. (2003). On verbal concord with collective nouns in British English. *English Language & Linguistics*, 7(1), 85-127.
- Deshor, S., & Gries, S. (2016). Profiling verb complementation constructions across New Englishes: A two-step random forests analysis of -ing vs. to- complements. *International Journal of Corpus Linguistics*, 21(2), 192-218.

- Diessel, H. (2013). Distance contrasts in demonstratives. In Martin Haspelmath, Matthew Dryer, David Gil, and Bernard Comrie (eds.), *World atlas of language structures*, (pp. 170-173). Oxford: Oxford University Press.
- Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, 17(4), 463-489.
- Downing, A. (2015). *English grammar: A university course*. Abingdon & New York: Routledge.
- Dowty, D. (1986). The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics? *Linguistics and Philosophy*, 9(1), 37-61.
- Dowty, D. & Jacobson P. (1988). Agreement as a semantic phenomenon. In J. Powers & K. de Jong (eds.), *Proceedings of the Fifth Eastern States Conference on Linguistics* (pp. 95-108). Ohio State University.
- Eastwood, J. (1994). *Oxford guide to English grammar*. Oxford: Oxford University Press.
- Elbourne, P. (1999). Some correlations between semantic plurality and quantifier scope. *The North East Linguistic Society (NELS) 29*, (pp. 81-92). Amherst, MA: GLSA.
- Enfield, N. J. (2003). Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language*, 79(1), 82-117
- Fawcett, R. (2000). *A Theory of syntax for systemic functional linguistics*. Philadelphia: John Benjamins Publishing Company.

Fernández-Pena, Y. (2017b). Patterns of verbal agreement with collective nouns taking plural of-dependents: A corpus-based analysis of syntactic distance. *Corpora*, 12(2), 207-241.

Fernández-Pena, Y. (2016). Revisiting the role of plural of-dependents on verbal agreement with collective nouns: A syntactic pilot study on New Englishes. In A. Ibarrola-Armendariz & J. O. D. U. Arruabarrena (eds.), *On the move: Glancing backwards to build a future in English studies*, (pp. 237-245). Bilbao: University of Deusto.

Fernández-Pena, Y. (2015a). *A Corpus-based study on verbal agreement with collective nouns in the recent history of English*. Unpublished thesis. University of Vigo.

Fernández-Pena, Y. (2015b). Verbal agreement with collective noun-based constructions: syntactic and lexical implications of *of*-dependents. In A. L. Lafuente & M. D. P. Requejo (eds.), *English and American studies in Spain: New development and trends*, (pp. 206-217). Alcalá de Henares: University of Alcalá de Henares.

Firth, J. R. (1957a). Applications of general linguistics. In F. R. Palmer, (ed.), *Selected Papers of J.R. Firth*, (pp. 126-136). Indiana University Press.

Folkesson, E. & Lantz, F. (2011). Collective nouns. *Uppsalla University, English HS2: Linguistics*

Forster, K., & Masson, M. (2008). Introduction: Emerging data analysis. *Journal of Memory and Language*, 59(4), 387-388.

Fowler, I.H. (1992). *The new Fowler's modern English usage*. Oxford: Oxford University Press.

- Francis, W. N. (1982). *Problems of assembling and computerizing large corpora* (pp. 7-24). n.a.
- Fries, U. (1988). The crew have abandoned the ship: Concord with collective nouns revisited. *Arbeiten aus Anglistik und Amerikanistik*, 13, 99-104.
- Gardelle, L. (2019). Semantic plurality: English collective nouns and other ways of denoting pluralities of entities. In J. C. Salmons, & E. F. K. Koerner (eds.), *Current Issues in Linguistic Theory: Amsterdam Studies in the Theory and History of Linguistic Science*, Amsterdam: John Benjamins Publishing.
- Gardelle, L. (2017). Are superordinates such as furniture and belongings collective nouns? *Linguisticae Investigationes*, 40(2), 150-172.
- Geeraerts, D., Kristiansen, G., & Peirsman, Y. (2010). Introduction. In D. Geeraerts, G. Kristiansen & Y. Peirsman (eds.), *Advances in cognitive sociolinguistics* (pp. 1-19). Berlin & New York: Mouton De Gruyter.
- Gil, D. (1996). Maltese collective nouns: A typological perspective. *Rivista di Linguistica*, 8(1), 53-87.
- Goldberg A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, 22, 131-153.
- Goldsmith, J. (2002). Probabilistic models of grammar: Phonology as information minimization. *Phonological Studies*, 5, 21-46.
- Greenbaum, S. (1996). *The oxford English grammar*. Oxford: Oxford University Press.
- Greenberg, J., H. (1978b). Generalizations about numeral systems. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (eds.), *Word Structure*, (pp 249-296). Stanford: Stanford University Press.

- Grice, H. P. (1957) Meaning. *Philosophical Review*, 66, 377-388.
- Gries, S., T. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement*. London & New York: Continuum Press.
- Gries, S., T. (2005). Null-hypothesis significance testing of word frequencies: A follow-up on Kilgarriff. *Corpus Linguistics and Linguistic Theory*, 1(2), 77-294.
- Griffiths, T. L. (2010). Bayesian models as tools for exploring inductive biases. In M. Banich & D. Caccamise (eds.), *Generalization of knowledge: Multidisciplinary perspectives*. New York: Psychology Press.
- Halliday, M. A. K. (2002). *On grammar*. London: Continuum.
- Halliday, M. A. K. (1996). On grammar and grammatics. In R. Hasan & C. Butt (eds.). *Functional descriptions: Theory into practice*, (pp. 1-38). Amsterdam: John Benjamins.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2<sup>nd</sup> ed.). London: Edward Arnold.
- Halliday, M. A. K. (1991). Corpus studies and probabilistic grammar. In K. Aijmer, & B. Altenberg (pp. 30-43), *Advances in corpus linguistics*. Amsterdam: Rodopi.
- Halliday, M. A. K. (1961). Categories of the theory of grammar. *Word*, 17, 241-92.
- Hasan, R. (1987). The grammarian's dream: Lexis as most delicate grammar. In M. A. K. Halliday & R. Fawcett (eds.), *New developments in systemic linguistics*. London: Pinter.
- Hawley, P. (2002). What is said. *Journal of Pragmatics*, 32, 969-991.

- Heine, B. (1997). *Cognitive foundations of grammar*. Oxford: Oxford University Press.
- Heine, B. & Narrog, H. (2010). Introduction. In B. Heine & H. Narrog (eds.), *Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press.
- Hoffmann, S. (2005). *Grammaticalization and English complex prepositions: A corpus-based study*. New York: Routledge.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of English language*. Cambridge: Cambridge University Press.
- Hundt, M. (2006). The committee has/have decided...: On concord patterns with collective noun in inner-and-outer-circle varieties of English. *Journal of English Linguistics*, 34(3), 206-232.
- Hundt, M. (1998). *New Zealand English grammar—fact or fiction? A corpus-based study in morphosyntactic variation*. Amsterdam: John Benjamins Publishing.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Ionin, T. (2006). This is definitely specific: Specificity and definiteness in article systems. *Natural Language Semantics*, 14(2), 175-234.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23-62.
- Jaeger, T. F. (2008) Categorical data analysis: Away from ANOVAs (Transformation or Not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59, 434-446.

- Jaeger, T. F. & Snider, N. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In L. Wolter & J. Thorson (eds.), *University of Rochester Working Papers in the Language Sciences*, 3(1), 26-44.
- Janney, R. W. (2002). Cotext as context: Vague answers in court. *Language & Communication*, 22(4), 457-475.
- Janssen, A. S. (2002). Negative affect and sensitization to pain. *Scandinavian Journal of Psychology*, 43, 131-137.
- Jespersen, O. (1961). *A modern English Grammar*. London: Bradford & Dickens.
- Johansson, S. (1979). American and British English grammar: An elicitation experiment. *English Studies*, 60(2), 195-215.
- Joosten, J. (2010). Collective nouns, aggregate nouns, and superordinates: where 'part of' and 'kind of' meet. *Linguisticae Investigationes*, 33, 25-49.
- Joosten, F., de Sutter, G., Drieghe, D., Grondelaers, S., Hartsuiker, R. & Speelman, D. (2007). Dutch collective nouns and conceptual profiling. *Linguistics*, 45, 85-132.
- Jurafsky, D. (2003). Probabilistic modelling in psycholinguistics: linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (eds.), *Probabilistic Linguistics*, Cambridge & Massachusetts: Massachusetts Institute of Technology Press.
- Kairis, P. (2017). *The rest of the family is or are? A quantitative analysis of collective nouns that are pre-modified by quantifying noun expressions in British and American English*. Unpublished master's thesis, Stockholms Universitet.

- Kathol, A. (1999). Agreement and the syntax-morphology interface in HPSG. In Robert D. Levine and Georgia Green, (eds.), *Studies in Contemporary Phrase Structure Grammar* (pp. 209-260). Cambridge & New York: Cambridge University Press.
- Kathol, A., & Kasper, B. (1991). *Agreement in HPSG revisited*. Unpublished master's thesis, Ohio State University.
- Keizer, E. (2007). *The English noun phrase: the nature of linguistic categorization*. Cambridge: Cambridge University Press.
- Kolkman, J. (2016). *The pragmatics of possession: Issues in the interpretation of pre-nominal possessives in English*. Unpublished doctoral dissertation, University of Manchester.
- Koplenig, A. (2017). Against statistical significance testing in corpus linguistics. *Corpus Linguistics and Linguistic Theory*, 15(2), 321-346
- Kratzer, A. (2012). On the plurality of verbs. In J. Dolling, T. Heyde-Zybatow, & M. Schafer (eds.), *Event Structures in Linguistic Form and Interpretation*, (pp. 269-300). Berlin: De Gruyter.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington DC: Center for Applied Linguistics.
- Lakaw, A. (2017). Diachronic shifts in agreement patterns of collective nouns in 19<sup>th</sup> century American and British English". In S. Hoffmann, A. Sand & S. Arndt-Lappe (eds.), *Studies in variation, contacts and change in English 18, Exploring Recent Diachrony: Corpus Studies of Lexicogrammar and Language Practices in Late Modern English* (pp. 1-33). Helsinki: VARIENG, University of Helsinki.

- Lakoff, R. (1974). Remarks on *this* and *that*. *Proceedings of the Chicago Linguistics Society*, 10, 345-356.
- Langacker, R. W. (1995). Raising and transparency. *Language*, 71(1), 1-17.
- Langacker, R. W. (1991). *Foundations of Cognitive Grammar: Descriptive applications*. Stanford: Stanford University Press.
- Lapointe, S. G. (1980). *A theory of grammatical agreement*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.
- Lau, J. H., Clark, A., & Lappin, S. (2017). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science*, 41(5), 1202-1245.
- Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (ed.), *Directions in Corpus Linguistics. Proceedings of the Nobel Symposium 82*, (pp. 105-122). Mouton de Gruyter.
- Leech, G. & Svartvik, J. (2002). *A communicative grammar of English* (3<sup>rd</sup> ed). London: Routledge.
- Levin, M. (2006). Collective nouns and language change. *English Language and Linguistics*, 10(2), 321-343.
- Levin, M. (2001). Agreement with collective nouns. *Department of English*, Lund: Lund University.
- Levin, M. (1998). Concord with collective nouns in British and American English. *Acta Wexionensia Humaniora, Humanities*, 1, 193-204.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
- Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge & Massachusetts: Massachusetts Institute of Technology Press.

- Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- Manning, D. C. (2003). *Probabilistic syntax*. In R. Bod, J. Hay & S. Jannedy (eds.), *Probabilistic Linguistics*, (pp. 289-342). Cambridge, Massachusetts: Massachusetts Institute of Technology Press.
- Marneffe, M.C., Grimm, S., Arnon, I., Kirby, S., & Bresnan, J. (2012). A statistical model of the grammatical choices in child production of dative sentences. *Language and Cognitive Processes*, 27(1), 25-61.
- Martin, J. R. (2016). Meaning matters: A short history of systemic functional linguistics. *Word* 62(1), 35-58.
- McEnery, T., & Hardie, A. (2012). Corpus linguistics: Method, *theory and practice*. *International Journal of Corpus Linguistics*, 18(2), 290-294.
- McKnight, G. H. (1925). Conservatism in American speech. *American Speech*, 1(1), 1-17.
- Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge University Press.
- Munro, R. A. (2008). *Probabilistic representation of systemic functional grammar*. London: Endangered Languages Archive Department of Linguistics School of Oriental and African Studies University of London.
- Mwinlaaru, I. N., & Yap, F. H. (2017). A tale of two distal demonstratives in Dagaare: Reflections on directionality principles in grammaticalization. *Language Sciences*, 64, 130-151.
- Nartey, M. & Ngula, R. S. (2009). Language and corpora: The case for Ghanaian English. *The Southeast Asian Journal of English Studies*, 20(3), 79-92.

- Ngula, R. S. (2014). Corpus linguistics and language development in Ghana. *Asian Journal of Humanities and Social Studies*, 2(4), 2321-2799.
- Ngula, R. S. (2009). *Corpus-based investigation into the semantics of modal verbs in Ghanaian English*. Unpublished master's thesis, University of Cape Coast.
- Nixon, G. (1972). Corporate-concord phenomena in English. *Studia Neophilologica*, 44, 120-126.
- Nunberg, G. (1997). *The pragmatics of reference*. Unpublished doctoral dissertation, CUNY Graduate Centre.
- Örlegård, P. (2014). *Collective nouns in English used in Sweden: A corpus-based study on number concord with collective nouns*. Unpublished thesis, Linnaeus University, Faculty of Arts and Humanities, Department of Languages.
- Outratová, T. (2013). *Subject-verb concord and pronominal reference after collective nouns in British and American English*. Unpublished master's thesis, Institute of English Language and Didactics.
- Pallant, J. (2005). *SPSS Survival Manual: A step by step guide to data analysis using SPSS for windows (version 12)*. Australia: Allen & Unwin.
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal Memory and Language*, 41, 427-456.
- Pearson, H. (2011). A new semantics for group nouns. In M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, & B. Tomaszewicz, (eds.), *Proceedings of WCCFL 28*: 160-168, Somerville, MA: Cascadilla Press.

- Pereira, F. (2000). Formal grammar and information theory: Together again. *Philosophical Transactions of the Royal Society*, 358, 1239-1253.
- Persson, G. (1989). On the semantics of collective nouns in English. In B. Odenstedt & G. Persson (eds.), *Instead of flowers: Papers in honour of Mats Rydén* (pp. 179-188). Stockholm: Almqvist & Wiksell.
- Peters, S. & Westerstahl, D. (2013). The semantics of possessives. *Language*, 89(4), 713-759.
- Pollard, C., & Sag, I. (1994). *Head-driven phrase structure grammar*. Chicago & London: University of Chicago Press.
- Poutsma, H. (1904). *A grammar of late modern English*. Groningen: Noordhoff.
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik. (eds.). (1985). *A comprehensive grammar of the English language*. London & New York: Longman.
- Roland, D., Elman, J.L., Ferreira, V.S. (2005). Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition*, 20, 1-28.
- Rybarczyk, M. (2015). *Demonstratives and possessives with attitude: An intersubjectively-oriented empirical study*. Amsterdam: John Benjamins Publishing Company.
- Sapir, E. (1921). *Language: An introduction to the study of speech*. Brace: Harcourt.
- Sauerland, U., & Elbourne, P. (2002). Total reconstruction, PF-movement, and the derivational order. *Linguistic Inquiry*, 33, 283-319.
- Schmid, H. J. (2010). Does frequency in text instantiate entrenchment in the cognitive system? In D. Glynn & K. Fischer (eds.), *Qualitative methods*

- in cognitive semantics: Corpus-driven approaches*, (pp. 101-133).  
Berlin & New York: De Gruyter.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(4), 623-656.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Smith, Peter W. (2017a). *Possible and impossible agreement mismatches*. Unpublished doctoral dissertation, University Frankfurt.
- Smith, P. W. (2017b). The syntax of semantic agreement in English. *Journal of Linguistics*, 53(4), 823-863.
- Smith, P. W. (2015). *Feature mismatches: Consequences for syntax, morphology and semantics*. Unpublished doctoral dissertation, University of Connecticut.
- Smith, P. W., Mursell, J., & Hartmann, K. (2020). Some remarks on agreement within the Minimalist Programme. In P. W. Smith, J. Mursell, & K. Hartmann (eds.), *Agree to agree in the minimalist programme* (pp. 1-30). Language Science Press.
- Snider, N. (2008). *An exemplar model of syntactic priming*. Unpublished doctoral dissertation, Stanford University.
- Solomonoff, R. (1977 [1995]). The discovery of algorithmic probability: A guide for the programming of true creativity. *Journal of Computer and System Sciences*, 56(2), 153-173.
- Sperber, D. & Wilson, D. (2005). Pragmatics. In F. Jackson & M. Smith (eds.), *Oxford Handbook of Contemporary Philosophy* (pp. 468-501). Oxford: Oxford University Press.

- Steele, S. (1978). Word order variation: A typological survey. In J. Greenberg (ed.), *Universals of human language*, (pp. 585-623). Stanford: Stanford University Press.
- Stubbs, M. (1996). *Text and corpus analysis: Computer assisted studies of language and culture*. Oxford: John Wiley and Sons Ltd.
- Swan, M. (2005). *Practical English usage* (3<sup>rd</sup> ed.). Oxford: Oxford University Press.
- Taylor, J. R. (1999). Possession. In K. Brown & J. Miller (eds.), *Concise encyclopaedia of grammatical categories* (pp. 300-303). Amsterdam: Elsevier.
- Taylor, J. R. (1996). *Possessives in English: An exploration in cognitive grammar*. Oxford: Oxford University Press.
- Taylor, J. R. (1995). On construing the world. In J. R. Taylor & R. E. Maclauray (eds.), *Language and the cognitive construals of the world* (pp. 1-22). Berlin: Walter de Gruyter & Co.
- Thompson, G. & Hunston, S. (eds.) (2006). *System and corpus: Exploring connections*. London: Equinox.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam & Philadelphia: John Benjamins.
- Trudgill, P. & Jean H. (1994). *International English* (3<sup>rd</sup> ed.). London: Arnold.
- Tucker, G. H. (1998). *The lexicogrammar of adjectives: A systemic functional approach to lexis*. London: Cassell.
- Vendler, Z. (1967). *Linguistics in Philosophy*. New York: Cornell University Press.

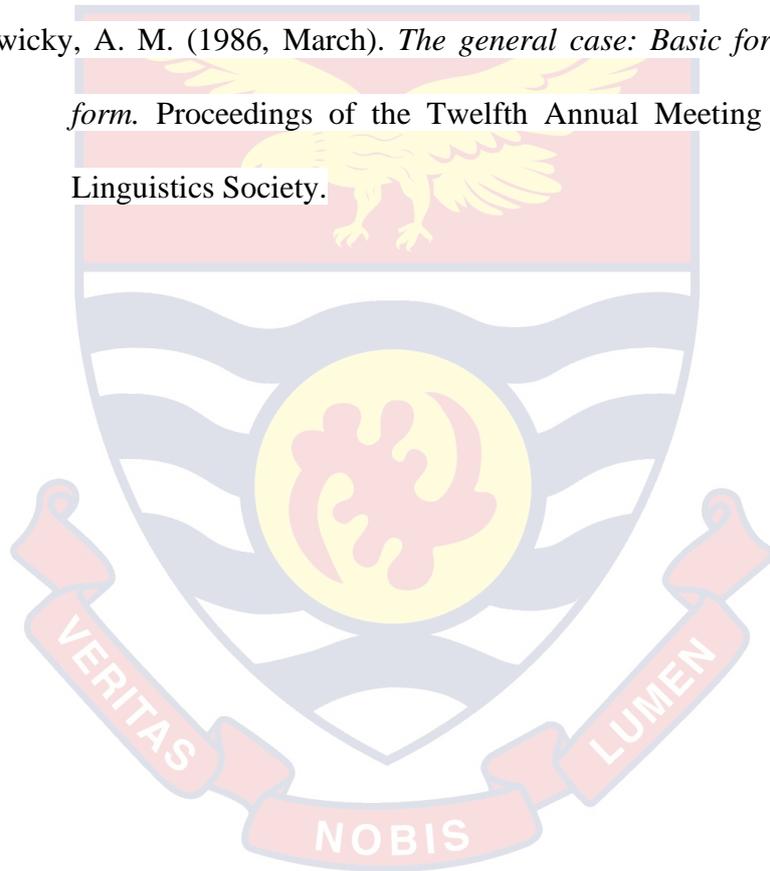
- Vikner, C. & Jensen, P. A. (2002). A semantic analysis of the English genitive. Interaction of lexical and formal Semantics. *Studia Linguistica*, 56(2), 191-226.
- Vikner, C. & Jensen, P. A. (2003). Producer interpretations of the English prenominal genitive. In M. Weisgerber (Ed.), *Proceedings of SUB7-Sinn und Bedeutung* (pp. 173-183). University of Konstanz, Germany.
- Von Heusinger, K. (2002). Specificity and definiteness in sentence and discourse structure. *Journal of Semantics*, 19(3), 245-274.
- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206-237.
- Weiner, E. J., & Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics*, 19(1), 29-58.
- Winter, Y. (2002). Atoms and sets: A characterization of semantic number. *Linguistic Inquiry* 33(2), 493-505.
- Winter, Y. (1998, May). Atom predicates and set predicates: towards a general theory of plural quantification. *Proceedings of Semantics and Linguistic Theory, SALT8*. <https://doi.org/10.3765/salt.v8i0.2801>.
- Willemse, P. (2010). A discourse perspective to nominal reference-point constructions. In Elzbieta Tabarkowska, Michal Choinski, & Lukasz Wiraszka (eds.), *Cognitive Linguistics in action: From theory to application and back* (pp. 209-204). Berlin/New York: Mouton de Gruyter.
- Wong, M. L. (2017). Concord patterns with collective nouns in Hong Kong English. *Linguistik Online*, 37, 60-76.

Yankson K. T. (1994). *Better English through concord for West African students*. Accra: Commercial Associates Limited.

Zimmermann, T. E. (2012). Compositionality problems and how to solve them. In M. Werning, W. Hinzen, & E. Machery, *Oxford handbook of compositionality*, (pp. 81-106). Oxford: Oxford University Press.

Zipf, G. K. (1949). *Human behaviour and the principle of least effort*. Cambridge: Addison Wesley Press.

Zwicky, A. M. (1986, March). *The general case: Basic form versus default form*. Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society.



APPENDIX A  
EXAMPLES EXTRACTED FROM BNC

I. Extracts of Whole Conception of CNs through the Use of Possessives

25. Gentler salute President FW de Klerk's government is very keen to present a modern progressive image, to dispel any nagging doubts the rest of the world may harbour about what a deeply repulsive ideology apartheid really is. (A1V 698)
26. Macao's government does not deny the existence of a confidence crisis, but places responsibility for it squarely on Peking. (A1V 17)
27. These delays cost small businesses well over £1bn a year but Mrs Thatcher's Government has not followed our EC partners in providing for automatic charging of interest on overdue bills, a step recommended 10 years ago by the Law Commission. (A4G 188)
28. If India's government is persuaded that warming produce better and more reliable monsoons, then it might decide that the interests of its burgeoning population would be better served by global warming than by attempts to hold it in check. (AB 692)
29. If his team is not doing too well he doesn't perform. (J1C 233)
30. His team has proved a successful nursery for a succession of young drivers who have made the Grand Prix circuit. (CBG 7166)
31. Describing herself as 'Absolutely Disgusted of Tunbridge Wells' she represented a new breed of recession victims living deep in the Tory heartlands who feel their party has betrayed them. (CEN 5127)
32. What he is saying is that he always wanted to drive, his family was against it and he defied them and did it. (CD9 395)

33. Their family fears that they may have been tortured to death or extrajudicially executed, or that they are still being held in incommunicado detention.

## II. Examples of Possessives Showing Part Reading of CNs

34. Manager Steve Perryman maintains his team have the talent — they simply lack consistency. (AK6 752)

35. In fact Ron Atkinson's team are playing well (AAW 206)

## III. Numerals as Constraints of CNAgre

36. With 125 people on board – students, teachers, education officials – the vessel left Norway for Aberdeen and called at Lerwick before returning to Bergen during the ten-day itinerary .... *The 80-strong party* were given a briefing on Sullom Voe's environmental care programme at the Fraser Peterson Centre before being taken on a tour of the terminal site.

## IV. Extracts in which Indefiniteness Constrains CNAgre

37. A team of six from Weatherall's London office have visited the sites, and provided detailed reports on location, facilities, likely capital expenditure requirements and site attractiveness.

38. Land for the playing field was acquired on lease some years ago and an energetic committee have worked ever since to maintain and develop it.

(C93 1049)

39. A LOCAL family have given a £6,000 boost to the Rainbow Trust funds.

(C88 1587)

40. A team is coming together based in the church with different ministry giftings and callings — Ephesians 4:11,! (CC5 206)

41. Finally, a government controls people by providing remedies for breaches of laws and for the violation of people's rights. (ANH 1127)
42. As we approached, a party of surly children were being taught how to traverse this wire contraption by Outward Bound teacher from Midlands. (AS3 1119)

**V. Instances of Definite CNPs**

43. The Barlow family, too, *still lives* in Cambridge. (A7D 587)
44. And in Italy opinion polls suggest the Socialist Party has lost a larger share of support since the election a year ago than any other party. (CR9 1959)
45. Erm we [unclear] which is actually putting forward a budget which is responding to what the Policy Committee asks us to do. (JJ9 174)
46. Even if, as I believe, the British Paralympic team are not affiliated to the BOC, the real reason is, as ever, money. (AKE 915 )
47. The reception had told her the Scottish Football team were residents in the hotel which increased her hopes of picking up the odd tip in order to ease her way through University. (B1L 812)

APPENDIX B

COMPARISON OF THE RECURSIVE AGREEMENT CHOICE  
BETWEEN *GOVERNMENT* AND *TEAM*

This table shows how frequently plural agreement choice occurs in one text, especially for *government*. It shows how *team* how plural agreement is often primed.

Government have		Government were		Team have	
Text	Hits	Text	Hits	Text	Hits
A07	1	A55	1	AA2	1
ABA	1	A6G	1	A2E	1
ABU	1	AAC	1	A33	1
ANR	1	AAX	1	A40	1
B1D	3	ABA	1	AS2	1
C8U	1	ABU	2	A5C	1
CRD	2	ADM	1	A61	2
EEC	1	AM0	1	AA7	2
FNX	2	B01	1	ABR	3
FRA	1	B0H	1	ADE	1
FX5	1	B0S	1	AHK	1
G3H	1	B1D	1	AJY	3
G3L	1	B2A	1	AKE	2
HHV	11	CE8	1	AJ4	4
HHW	16	CN5	1	B1L	5
HHX	17	E9P	1	B2H	1
HNK	1	EA4	1	BN6	1
HRJ	1	EBW	1	BN9	1
HUW	1	EEC	1	BNT	1
HWA	1	EF9	1	C88	1
J16	1	EFN	1	C8L	2
JJD	1	EUU	1	C98	1
JSH	1	EVP	1	C9R	1
		FBP	1	CAD	1
		FPX	1	CC8	1
		FXT	1	CCJ	1
		GOC	1	CCS	1
		G3H	9	CEP	1
		G4R	1	CER	1
		G5G	1	CH3	2
		HAU	1	CH7	2
		HB3	1	CHV	1
		HBK	1	CJ6	1
		HDT	2	CLD	1
		HE5	1	CML	1
		HH3	2	CN2	1

HHV	162	EB3	1
HHW	118	EBU	1
HHX	171	ED9	1
HJ7	1	FT9	1
HJM	1	G27	1
HSL	1	GXA	1
HYX	1	GXJ	1
J1M	1	H7B	2
J3P	1	HAT	1
J9B	1	HB2	1
J9D	1	HBH	2
J9K	1	HBJ	1
J9L	2	HD7	1
JAE	2	HGM	2
JJE	1	HJ3	1
JNJ	2	HJG	1
JSG	3	HMN	1
JSH	1	HNW	1
K1J	1	HP8	1
K1K	1	HPY	1
K5D	2	HRR	1
K5H	1	HS1	1
K76	1	HU8	1
KB1	1	HU9	1
KDJ	1	J0W	1
KGX	1	JIF	1
KLY	1	JIG	2
KM0	1	JBS	1
KM8	1	JYB	1
KRT	1	K1D	1
		KJH	2
		K1L	1
		K1M	1
		K1P	1
		K1T	1
		KIW	1
		K4T	3
		K52	2
		K5A	4
		K5J	2
		K5M	1
		K97	1
		K9M	1
		KAG	1
		KBC	1
		KBK	1
		KLV	1