

**Specification in the English nominal group
with reference to student writing**

Y.V. Botha

Thesis submitted for the degree Doctor of Philosophy in Linguistics and Literary Theory
at the Potchefstroom Campus of the North-West University

Promoter: Prof. A.J. van Rooy

2012

Acknowledgements

I thank the Lord for granting me the opportunity and means to undertake this study and for giving me the strength to complete it. To Him be the glory and honour.

I wish to thank the following people:

- My promoter, Prof Bertus van Rooy, for his patient guidance, helpful comments and encouragement;
- My husband, Willem, for his support and encouragement and for the sacrifices he made to allow me to focus on my research;
- Our family and friends (in particular our parents) for their moral support and prayers;
- Colleagues at the Schools of Languages at the Vaal Triangle Campus and the Potchefstroom Campus of the NWU, CText and the Research Unit, in particular, Attie de Lange, Martin Puttkammer, Nico Oosthuizen, Righardt Pretorius and Ansu Berg;
- Nelleke Oostdijk at Radboud University, Nijmegen, for her input in the early days of this study when I had the opportunity (by means of a Huygens scholarship) to spend some time in the Netherlands to work on my thesis.

This work is based upon research supported by the National Research Foundation. Any opinion, findings and conclusions or recommendations expressed in this material are those of the author and therefore the NRF do not accept any liability in regard thereto.

This thesis is dedicated to my son Max.

Table of contents

CHAPTER 1: INTRODUCTION

1.1 Introduction	1
1.2 Contextualization and problem statement	1
1.2.1 Black South African English.....	1
1.2.2 The nominal group in descriptions of BSaE.....	8
1.2.3 Focus of this study.....	11
1.3 Research questions and aims.....	11
1.4 Prospectus.....	12

CHAPTER 2: THEORETICAL FRAMEWORK

2.1 Introduction	13
2.2 A usage-based understanding of language	13
2.2.1 Functions and structural categories in SFG	18
2.2.2 Functions and structural categories in Givon	23
2.3 Framework for analysis	31
2.3.1 Construal of experience at the rank of the clause	32
2.3.1.1 Construal in the clause according to Hallidayan SFG	33
2.3.1.2 Semantic roles and situation types in clauses	41
2.3.1.3 Types and instances of things and processes	44
2.3.2 Specifying functions in the nominal group.....	53
2.4 Framework for interpretation	64
2.4.1 Constructions and schemas	64
2.4.2 Patterns and the idiom principle.....	66
2.4.3 Conventionalization in new varieties of English	67
2.5 Conclusion	68

CHAPTER 3: METHODOLOGY

3.1 Introduction	70
3.2 The corpora	70
3.3 Approaches to corpus data	71
3.4 Data processing	76
3.4.1 Wordlists, frequencies and keywords.....	76
3.4.2 Concordances and random sampling	79
3.4.3 Part-of-speech tagging	80

3.4.4 Computer-aided error analysis in the <i>ICLE</i> project.....	80
3.5 Conclusion	81

CHAPTER 4: GRAMMATICAL PATTERNS OF THE GENERAL NOUN *PEOPLE*

4.1 Introduction	82
4.2 Analysis categories	84
4.3 Unspecified <i>people</i>	86
4.4 Definite <i>people</i>	93
4.4.1 Instances of <i>people</i> marked as definite without additional quality specification in the nominal group	97
4.4.2 Definite <i>people</i> with quality specification.....	104
4.4.2.1 Definite <i>people</i> -groups where quality is specified prenominally only ...	105
4.4.2.2 Definite <i>people</i> -groups where quality is specified postnominally only .	112
4.4.3 Definite <i>people</i> -groups in quantifying <i>of</i> -partitions	117
4.4.4 Summary	120
4.5 Quality-specified <i>people</i>	122
4.5.1 Prenominal qualities	127
4.5.1.1 Relative qualities	129
4.5.1.2 Thing-oriented qualities	134
4.5.2 Postnominal word groups	141
4.5.3 Postnominal prepositional phrases	148
4.5.4 Postnominal qualities in the form of clause	159
4.5.5 Summary	169
4.6 Quantified <i>people</i>	170
4.7 Left-dislocated <i>people</i> -groups	179
4.8 Conclusion	192

CHAPTER 5: DETERMINATION

5.1 Introduction	195
5.2 Theoretical concepts	196
5.2.1 The determiner function in the nominal group	196
5.2.2 Grounding and the distinction between type and instance	199
5.2.3 Referential intent: particular vs. non-particular instances	200
5.2.4 Definite/indefinite	212
5.2.4.1 Sources of identifying information	217
5.2.4.2 Demonstrative and possessive determiners	221
5.2.5 Summary of conceptual distinctions.....	221

5.3	Determiners in non-native varieties of English.....	222
5.4	The articles in the TLEC and LOCNESS.....	239
5.4.1	The definite article.....	242
5.4.1.1	Identifying uses of <i>the</i> in the TLEC and LOCNESS.....	244
5.4.1.2	Non-particular uses of <i>the</i> in the TLEC and LOCNESS.....	253
5.4.1.3	The use of <i>the</i> in fixed expressions.....	260
5.4.1.4	Summary of the discourse functions of the definite article.....	262
5.4.2	The indefinite article.....	263
5.5	Demonstrative and possessive determiners.....	271
5.5.1	Demonstrative determiners.....	274
5.5.2	Possessive determiners.....	278
5.6	Conclusion.....	281
CHAPTER 6: QUANTIFICATION		
6.1	Introduction.....	283
6.2	Quantifiers in non-native varieties of English.....	286
6.3	Quantitative overview of quantifiers in the TLEC and LOCNESS.....	288
6.3.1	Single-word quantifiers.....	290
6.3.2	Multi-word quantifiers.....	297
6.4	Concordance analyses of key quantifiers: <i>some</i> and <i>most</i>	301
6.4.1	The quantifier <i>some</i>	302
6.4.2	The quantifier <i>most</i>	308
6.5	Number marking.....	319
6.6	Conclusion.....	323
CHAPTER 7: CONCLUSION		
7.1	Introduction.....	324
7.2	Summary of findings.....	324
7.2.1	The syntagmatic patterns of the noun <i>people</i>	324
7.2.2	Determiner choice.....	326
7.2.3	Quantification.....	327
7.3	Conclusion.....	328
	Bibliography.....	330

List of Tables

Table 2.1	Representation and coding	28
Table 2.2	Process types, their meanings and characteristic participant	34
Table 2.3	The categories of relational clause	37
Table 2.4	Contrast between noun and verb archetypes.....	46
Table 3.1	Excerpt from frequency-ordered WordList (TLEC)	77
Table 3.2	Excerpt from alphabetically-ordered WordList (TLEC)	77
Table 3.3	Statistics tab in WordList (TLEC).....	78
Table 3.4	Top ten key words in the TLEC when compared to LOCNESS	79
Table 4.1	The noun <i>people</i> : frequency and keyness	83
Table 4.2	Analysis categories of <i>people</i> concordances	85
Table 4.3	<i>If</i> -conditional clauses, <i>wh</i> -circumstance clauses and modal auxiliaries with unspecified <i>people</i>	90
Table 4.4	Frequencies of patterns of definite <i>people</i> -groups in the TLEC and LOCNESS.....	94
Table 4.5	Realizations of definite determiners in <i>people</i> -groups in the TLEC and LOCNESS	95
Table 4.6	Quality specifications of <i>people</i> in the TLEC and LOCNESS.....	125
Table 4.7	Modifiers of <i>people</i> in the TLEC and LOCNESS	126
Table 4.8	Prenominal thing-oriented qualities of <i>people</i> in the TLEC and LOCNESS....	135
Table 4.9	Postnominal groups as modifiers of <i>people</i> in the TLEC and LOCNESS.....	142
Table 4.10	Prepositional phrases that modify <i>people</i> in the TLEC and LOCNESS.....	150
Table 4.11	Postnominal clauses as modifiers of <i>people</i> in the TLEC and LOCNESS.....	160
Table 4.12	Quantity specifications of <i>people</i> in the TLEC and LOCNESS	171
Table 4.13	Left-dislocated Subject nominal groups with <i>people</i> in the TLEC	183
Table 5.1	The definite/indefinite system	225
Table 5.2	The specific/non-specific system	227
Table 5.3	Sand's (2004) frequencies for articles in student writing compared to article frequencies in the TLEC and LOCNESS (per 1000 words in the corpus)	240
Table 5.4	Usage of articles in the error-tagged subcorpus of the TLEC	242
Table 5.5.	Discourse functions of the definite article in the TLEC and LOCNESS	244
Table 5.6	Nominal group internal identification in the concordances of <i>the</i> from the TLEC and LOCNESS: Position of identifying information.....	251
Table 5.7	Nominal group internal identification in the concordances of <i>the</i> from the TLEC and LOCNESS: Quality type and form.....	252
Table 5.8	Discourse functions of the indefinite article in the TLEC and LOCNESS	264
Table 5.9	Frequencies of the main types of definite determiners in the TLEC and LOCNESS	273
Table 5.10	Usage of demonstrative determiners in the error-tagged subcorpus of the TLEC	275
Table 5.11	A comparison of demonstrative determiners in the TLEC and LOCNESS	277
Table 5.12	A comparison of possessive determiners in the TLEC and LOCNESS.....	279
Table 6.1	Single-word determinative and pronominal quantifiers in the TLEC and LOCNESS	291
Table 6.2	Types of determinative and pronominal quantifiers in the TLEC and LOCNESS	292
Table 6.3	Multi-word quantifiers in the TLEC.....	297
Table 6.4	Multi-word quantifiers in LOCNESS.....	298
Table 6.5	The frequencies of <i>some</i> and <i>most</i> in the TLEC and LOCNESS	302
Table 6.6	A comparison of <i>some</i> in the TLEC and LOCNESS.....	302
Table 6.7	A comparison of the <i>most</i> in the TLEC and LOCNESS	308
Table 6.8	Patterns of <i>most of</i> in the TLEC.....	311

List of figures

Figure 2.1	The stratification of language	18
Figure 2.2:	The three types of structure in the clause according to Hallidayan SFG.....	20
Figure 2.3	Preposition phrase and preposition group structure	22
Figure 2.4	The functional realms of language distinguished by Givon.....	25
Figure 2.5	Material Processes and their Participants.....	35
Figure 2.6	Mental Processes	36
Figure 2.7	Behavioural Processes	38
Figure 2.8	Verbal Processes.....	38
Figure 2.9	Existential Processes	39
Figure 2.10	Experiential structure of the nominal group.....	40
Figure 2.11	The experiential structure of the verbal group.....	40
Figure 2.12	Semantic and syntactic roles.....	41
Figure 2.13	The lexico-grammatical continuum	50
Figure 2.14	Elements in the experiential structure of the nominal group	55
Figure 2.15	Specification in the nominal group.....	56
Figure 2.16	Head and Thing in <i>of</i> -partitions	61
Figure 3.1	Extract from concordance	81
Figure 4.1	Semantic types of definite <i>people</i>	104
Figure 4.2	The extension of <i>other</i>	134
Figure 4.3	A categorization network of the meanings of prepositions.....	149
Figure 4.4	Subsets of <i>people</i>	174
Figure 5.1:	The hierarchical position of the determiner	198
Figure 5.2	Referential intent system (part 1).....	211
Figure 5.3	Referential intent system (part 2).....	212
Figure 5.4	Referential intent system (part 3).....	214
Figure 5.5	Referential intent system (part 4).....	217
Figure 5.6	The definite/indefinite distinction of “the more established Englishes”	226
Figure 5.7	The specific/non-specific distinction of the New Englishes	227
Figure 5.8	Identification system.....	230
Figure 5.9	Relative frequencies of definite determiners in the TLEC and LOCNESS.....	273
Figure 5.10	Relative frequencies of demonstratives in the TLEC and LOCNESS.....	278
Figure 5.11	Relative frequencies of possessive determiners in the TLEC and LOCNESS	280

Summary and key terms

In this thesis the structure of the nominal group in Black South African English (BSAfE) is investigated by means of a comparison of data from the *Tswana Learner English Corpus* (TLEC) and the *Louvain Corpus of Native English Essays* (LOCNESS). Both corpora consist of student essays and are sub-corpora of the *International Corpus of Learner English* (ICLE). The TLEC represents a non-native variety of English, namely BSAfE, while LOCNESS represents native English from the United States and the United Kingdom.

In the existing literature there are observations about and examples of (non-standard) characterizing features of BSAfE pertaining to nouns, determiners and quantifiers (e.g. Gough 1996), but until now, no in-depth study of the grammar of the nominal group in BSAfE has been undertaken. This study is an attempt to fill that gap. I present a description of the grammatical features of BSAfE observed in the corpus data in terms of linguistic functions and without assuming that they are errors or evidence of deficiencies. Though the approach is comparative (in the sense that a control corpus is used), it is primarily descriptive and non-normative, and as such, function-based. This study is conducted within the theoretical framework of functional linguistics, drawing on systemic functional linguistics as well as other functional and cognitive approaches to language.

The specifying functions that the nominal group and its elements may fulfill form the basis of the descriptive framework. These functions are (1) the type-specifying function of the noun, (2) the (referent-)specifying function of the determiner, (3) quantification and (4) quality specification by modifiers and complements. The type-specifying function of the noun is particularly relevant in Chapter 4, which deals with the grammatical patterns of the noun *people*, although it also informs analyses in Chapters 5 and 6 which deal with determiners and quantification respectively. Quality specification is especially relevant in the description of the grammatical patterns of the noun *people*.

The noun *people* is the most frequent noun in each of the corpora, but is at the same time a positive keyword in the TLEC, which means that it occurs much more frequently in the TLEC than in LOCNESS. Analysis of the full corpus concordances of this noun provides much evidence of anti-deletion in BSAfE (as first postulated by Mesthrie,

2006) and also sheds some light on left dislocation patterns involving the noun *people* (cf. Mesthrie, 1997). This analysis also reveals unique uses of the definite article and certain quantifying constructions in the TLEC data, which are investigated in the next two analysis chapters.

A comparison of concordance samples of the articles indicates that the definite/indefinite distinction is made in both corpora and that there is not enough corpus evidence to postulate that there is a different system underlying the choice of article in BSAfE, such as a system based principally on the specific/non-specific distinction, as postulated by Platt, Weber and Ho (1984) for New Englishes in general. Analysis of the concordances of demonstrative and possessive determiners indicate that these determiners are used proportionally more frequently in the TLEC than in LOCNESS. Concordances of the words that are normally classified as quantifiers indicate that there are many more partitive-*of* quantifying constructions in the TLEC than in LOCNESS. The words *some* and *most* are positive keywords in the TLEC. After analyses of their concordances, it is concluded that their relative frequency can be attributed to the fact that *some* is often used merely as an indefinite marker and that *most* is often used as a synonym for *many*.

The study shows that BSAfE largely shares its general grammar of the nominal group with other (including native) varieties of English, but at a finer level of analysis, some characteristic constructions and uses are detected. The corpus data indicate that the unique constructions in the TLEC data are mostly functionally motivated. These constructions represent conventionalized innovations in the sense used by Van Rooy (2010), rather than mere language learning errors.

Key terms: Black South African English, nominal groups, noun phrases, specification, nouns, determiners, quantifiers, student writing, systemic functional grammar, cognitive grammar, pattern grammar, corpus linguistics.

Opsomming en sleutelterme

In hierdie proefskrif word die struktuur van die naamwoordgroep in Swart Suid-Afrikaanse Engels ondersoek aan die hand van 'n vergelyking van data uit die *Tswana Learner English Corpus* (TLEC) en die *Louvain Corpus of Native English Essays* (LOCNESS). Beide korpora bestaan uit studente-opstelle en is subkorpora van die *International Corpus of Learner English* (ICLE). Die TLEC verteenwoordig 'n nie-moedertaalvariëteit van Engels, naamlik Swart Suid-Afrikaanse Engels, terwyl LOCNESS verteenwoordigend is van moedertaal-Engels van die Verenigde State en die Verenigde Koninkryk.

In die bestaande literatuur is daar opmerkings oor en voorbeelde van die (nie-standaard) kenmerke van Swart Suid-Afrikaanse Engels wat betref naamwoorde, bepalersⁱ en kwantifiseerders (bv. Gough, 1996), maar tot op hede is geen in-diepte studie van die grammatika van die naamwoordgroep in Swart Suid-Afrikaanse Engels onderneem nie. In dié studie word daar gepoog om hierdie leemte te vul. Die studie bied 'n beskrywing van die grammatikale kenmerke van Swart Suid-Afrikaanse Engels soos waargeneem in die korpusdata in terme van taalfunksies sonder om aan te neem dat sodanige kenmerke taalfoute of -gebreke is. Hoewel die benadering vergelykend is (in die sin dat daar van 'n kontrolekorpus gebruik gemaak word), is dit primêr beskrywend en nie-normatief en derhalwe funksie-gebaseerd. Die ondersoek word onderneem binne die teoretiese raamwerk van die funksionele linguïstiek en ontgin sistemies-funksionele linguïstiek sowel as ander funksionele en kognitiewe benaderings tot taal.

Die spesifiserende funksies wat die naamwoordgroep en sy elemente kan vervul, vorm die grondslag van die beskrywende raamwerk. Hierdie funksies is (1) die benoemingsfunksie van die naamwoord, (2) die (referent-)spesifiserende funksie van die bepaler, (3) kwantifisering en (4) kwalifisering. Die benoemingsfunksie van die naamwoord is veral van toepassing in Hoofstuk 4 wat handel oor die grammatikale patrone van die naamwoord *people*, maar dit het ook betrekking op die ontledings in Hoofstukke 5 en 6 wat handel oor bepalers en kwantifisering onderskeidelik. Kwalifisering is veral relevant in die beskrywing van die naamwoord *people*.

ⁱ Bepaler is hier die vertaling van “determiner” en nie, soos gebruiklik, van “modifier” nie.

Die naamwoord *people* is die mees frekwente naamwoord in albei korpora, maar is tegelykertyd 'n positiewe sleutelwoord in die TLEC, wat beteken dat dit heelwat meer gereeld in the TLEC as in LOCNESS voorkom. 'n Ontleding van die volle korpuskonkordansie van dié naamwoord bevestig voorkoms van die anti-skrapping-verskynsel in Swart Suid-Afrikaanse Engels (soos aanvanklik aan die hand gedoen deur Mesthrie, 2006) en belig ook sommige linkerverplasingspatrone met die naamwoord *people* (sien Mesthrie, 1997). Hierdie ontleding onthul ook unieke gebruike van die bepaalde lidwoord en sekere kwantifiserende konstruksies in die TLEC-data wat in die opvolgende twee hoofstukke ondersoek word.

'n Vergelyking van ewekansig-geselekteerde konkordansiereëls van die lidwoorde dui daarop dat die bepaalde/onbepaalde onderskeid in beide korpora gemaak word en dat daar nie genoeg korpusbewyse is om te postuleer dat die keuse van lidwoord in Swart Suid-Afrikaanse Engels onderlê word deur 'n ander sisteem, soos byvoorbeeld 'n sisteem wat gegrond is op die spesifieke/nie-spesifieke onderskeid soos wat Platt, Weber and Ho (1984) vir Nuwe variëteite van Engels in die algemeen postuleer, nie. 'n Ontleding van die konkordansies van aanwysende en besitlike bepalers dui daarop dat hierdie tipe bepalers proporsioneel meer frekwent in die TLEC as in LOCNESS voorkom. Konkordansies van woorde wat normaalweg as kwantifiseerders geklassifiseer kan word, dui daarop dat daar meer gelede kwantifiserende konstruksies met *of* in die TLEC is vergeleke met LOCNESS. Die woorde *some* en *most* is positiewe sleutelwoorde in die TLEC. Na 'n ontleding van hul konkordansies word die gevolgtrekking gemaak dat hul relatiewe frekwenheid toe te skryf is aan die feit dat *some* dikwels bloot as 'n onbepaaldheidsmerker aangewend word en dat *most* dikwels as sinoniem van *many* gebruik word.

Die studie toon aan dat Swart Suid-Afrikaanse Engels grootliks ooreenstem met ander (insluitend nie-moedertaal-) variëteite van Engels wat betref die grammatika van die naamwoordgroep in die algemeen, maar op 'n fyner vlak van ontleding blyk daar sommige kenmerkende konstruksies en gebruike te wees. Die korpusdata toon aan dat die unieke konstruksies in die TLEC-data meestal funksioneel-gemotiveerd is. Hierdie konstruksies verteenwoordig gekonvensionaliseerde innoverings in die sin waarin Van Rooy (2010) die term gebruik, eerder as blote taalaanleerdersfoute.

Sleuteltermes: Swart Suid-Afrikaanse Engels, naamwoordgroepe, naamwoordstukke, spesifisering, naamwoorde, bepalers, kwantifiseerders, studenteskryfwerk, sistemies-

funksionele grammatika, kognitiewe grammatika, patroongrammatika,
korpuslinguistiek.

CHAPTER 1

Introduction

1.1 Introduction

As the title indicates, this study investigates aspects of the grammar of Black South African English (henceforth BSAfE). Section 1.2 below contextualizes this study in terms of existing literature on BSAfE, starting with a brief overview of the historical origins of the variety and proceeding to a survey of scholarly work on BSAfE in which the changing attitudes to the variety and its study can be traced (Section 1.2.1). In Section 1.2.2 the focus is narrowed to observations on the grammar of BSAfE relating to the nominal group (noun phrases). It will be shown that this is an area of BSAfE grammar that is still very much underexplored in linguistic enquiry. In Section 1.3 the research questions and aim of this study are presented. Section 1.4 provides an overview of the chapters to follow.

1.2 Contextualization and problem statement

1.2.1 Black South African English

BSAfE is the variety of English spoken by the majority of the black people in South Africa. Its characteristic features with regard to pronunciation and grammar is often related to its origins in education (De Klerk, 2003a:463ff). From the early 1800s when South Africa first came under British rule until the mid-1900s when the nationalist government came into power, a small minority of black people received education in English by means of missionary schools (De Klerk, 2006:11). The incentive for learning English was one of survival. The ruling white cultural group that provided employment (for example on the mines) were English-speaking. It is commonly purported that the relatively small number of black people who attended missionary schools had access to mother-tongue speakers of English as teachers. Van Rooy (in preparation) points out

that the situation in the missionary schools were not as good as they are made out to be in some accounts (for example Wright, 1996:150). They have become idealized in the light of the education system introduced by the nationalist government through the Bantu Education Act of 1953, which deprived black pupils of access to competent teachers as well as appropriate learning materials and infrastructure (Wright, 1996:150-151). Apart from the fact that only a small minority of the black population attended missionary schools, those who did rarely completed more than the first few years of school and many of the teachers were not in fact first language speakers of English (though they were European). Of those who were first language speakers of English, many spoke regional British dialects.

In the (black) Department of Education and Training (DET) schools (of the apartheid years), the medium of instruction in the first four years of school was the mother tongue (for instance, Setswana) (Buthelezi, 1995:242; Gough, 1996:54). Thereafter the medium of instruction was English. The practice of introducing English as language of instruction after the fourth year of school and the absence of native English teachers in township schools persist. Gough (1996:54) remarks on the detrimental effects of this sudden transition from the mother tongue to English and point out that “[t]eachers are overwhelmingly non-native speakers and products of Bantu Education themselves”.

It is an undeniable fact that a highly deprived education system has had some effect on BSAfE. However, Van Rooy (in preparation) cautions against viewing poor education as the sole factor in the formation of a New English. Viewing BSAfE simply as a product of a poor education system implies that it is an interlanguage or an insufficiently acquired second language, and also implies that the perceived deficits of this second language variety will improve if the education system improves. Instead, Van Rooy (in preparation) argues in favour of viewing BSAfE as a New English with “unique linguistic features which are different from native varieties of English, but not deficient” (cf. also Van Rooy 2008b:337-338).

The view of BSAfE as an interlanguage riddled with learner errors is based on the much less than perfect conditions in which it is learnt (which includes little to no exposure to native speakers of English). The socio-economic consequences of not being proficient in English and the importance of being internationally intelligible are used as rationale to

improve the English of BSAfE speakers in the direction of native Standard English. Eradication of language errors and proper mastery of Standard English are seen as the only way to ensure upward mobility. This provides a rhetoric in which it is possible to “empower” a majority group (speakers of BSAfE) by forcing upon them the standards of a minority (white English speakers in South Africa). Depictions of BSAfE by Finn (1986) and Scheffler (1978) are typical of the deficit perspective. Wright suggests that “BSAfE could be defined as an arrested stage in a learner-language continuum” (1996:153) and argues strongly against “the development of BSAfE as a separate but equal variety of English” which he considers “misguided” (1996:151-152). This view is not limited to the status of English in South Africa. In the context of English as a world language, scholars like Quirk (1985) also argue against acceptance of non-standard varieties of English (New Englishes) and in favour of upholding the native standard – ostensibly for allowing learners of English the best chance of getting ahead in a global economy, but with the thinly-veiled, ulterior motive of conserving the native standard and warding off the collapse of Standard English. These approaches are characterized by the wildly exaggerated threat that varietal features supposedly pose to mutual intelligibility in contexts where speakers from different language backgrounds have to converse with each other. Wright issues the following warning:

The very real danger exists that an educational language dispensation too hospitable to linguistic variation in English could prove a socio-political disaster, firstly by not reflecting the aspirations of the society and secondly by encouraging politically expedient initiatives which inhibit the development of effective intra-national and international communication.

The truth may be unwelcome to some, but it seems unavoidable that in today’s world of interdependent national states, the only English standard which educational authorities can reasonably implement is the international standard (1996:155).

The (unwarranted) concern with intelligibility can be ascribed to what Van Rooy (2008b:342) terms a view of grammar “as some kind of precondition for communication”. Citing Hopper’s proposal of emergent grammar, Van Rooy proposes an analysis of BSAfE grammatical features from the perspective that grammar is a by-product of communication, rather than being aprioristic. This approach will be discussed in more detail later in this section.

At the same time that scholars like Wright (1996:152) were still clinging to a denial of the status of BSAfE as a variety of South African English, scholars such as Buthelezi (1995) and Wade (1995), as well as Gough (1996) in more carefully phrased terms, were starting to argue for the recognition of BSAfE as a distinct variety of English with its own grammatical features. Buthelezi uses the term “the emerging dialect of English used by blacks (Africans) in South Africa” and maintains that this variety has “distinctive lexical and syntactic features” (1995:242). Gough (1996:59, 61) points out some similarities between BSAfE and varieties of African English with regard to pronunciation and between BSAfE and new Englishes with regard to grammatical features. Wade (1995) goes as far as exploring the possibility of BSAfE influencing the standard native variety of English in South Africa as the contexts in which BSAfE is used expand.

Since the late 1990s BSAfE has been regarded as a distinct variety of English and studied as such (De Klerk, 1999, De Klerk & Gough, 2002, Van der Walt and Van Rooy, 2002). Four waves of scholarly work on BSAfE since the late 1990s to the present can be distinguished. The first wave can be seen as pre-corpus work and comprises studies that attest the existence of unique features of BSAfE. Gough (1996) lists a number of phonological, grammatical and stylistic features of BSAfE based on previous observations of the writing of matric pupils and university students. Since the data was not collected with the aim of quantifying features, no attempt is made to quantify the data. Van der Walt and Van Rooy (2002) go beyond merely listing features that seem to occur regularly in BSAfE by investigating the extent to which these features have gained acceptability in the language of BSAfE speaking students and teachers. Their data allow for quantification in that acceptability rates of questionnaire items are expressed as percentages.

The second wave of BSAfE scholarship makes use of authentic data to illustrate the features under discussion (Wade, 1995; De Klerk, 2003; Makalela, 2004). Wade (1995) finds much evidence of copy pronouns in his small corpora of spoken and written BSAfE. He judges copy pronouns to occur frequently in BSAfE, without explicit comparison to control data. Makalela (2004) discusses four often mentioned groups of features of BSAfE, namely the extension of the progressive aspect to stative verb, tense sequencing, topic promotion devices and modality markers. He makes use of spoken data to exemplify these features, but no attempt is made to indicate frequencies. De

Klerk (2003a&b) lists a number of examples from her corpus of spoken Xhosa English to exemplify features of BSAfE discussed in previous literature (chiefly Van der Walt and Van Rooy, 2002, but also Gough, 1996). Despite having access to corpus data designed to allow for rigorous quantitative accounts of features, De Klerk (2003:a,b as well as 2006:139-156), provides only raw frequencies of the number of examples of a specific grammatical feature found in her data with no attempt to relativize these frequencies so that an accurate idea of the scarcity or ubiquity of a feature cannot be ascertained. De Klerk (1999, 2006) places her studies using the Xhosa English corpus in the context of the World Englishes paradigm.

The third wave of studies of the features of BSAfE adds a comparative and quantitative dimension to the use of authentic data. Mesthrie (1997) makes use of 44 interviews to investigate topicalization phenomena in BSAfE. He finds topicalization phenomena in 5.6% of the sentences in these interviews. In order to determine his judgement of this figure as high, Mesthrie (1997:127) compares it to interview data from 150 South African Indian English speakers (8.1%) and 10 white native speakers from Cape Town (1.8%). Also illustrative of this comparative approach are the studies by Minow (2010) and Siebers (2012). Minow (2010) makes use of 27 interviews (involving 45 speakers) representing spoken BSAfE to investigate four selected features of BSAfE, namely past tense marking, the progressive usage, article omission and left dislocation. Siebers (2012) makes use of recordings of 16 informants to investigate a wide variety of features of BSAfE, which include phonological features and grammatical features pertaining to the noun phrase, the verb phrase (tense and aspect) and the clause in general. The work of Mesthrie, Minow and Siebers share a socio-linguistic perspective in which variance in the data along the lectal scale (basilect-mesolect-acrolect) is taken into consideration. De Klerk (2006:87) classifies her chapter on formulaic utterances in spoken Xhosa English as a sociolinguistic study. In this study she compares the data of her L2 corpus with L1 New Zealand English. By her own admission this comparison is superficial (De Klerk, 2006:87). A comparison is made to New Zealand English rather than another L1 variety because of “the obvious link between the two ‘colonial’ varieties of English” (De Klerk, 2006:101 note) and because a corpus of spoken South African English was not readily available. Only raw frequencies, not normalized frequencies are reported since De Klerk (2006:96) deems the “fairly similar size” of the corpora sufficient for comparative purposes. Unsurprisingly De Klerk finds many examples of formulaic

utterances in Xhosa English and also finds that many of them do not occur in New Zealand English. Her conclusion would suggest that Xhosa English speakers rely on these utterances to help them sound like native speakers of English (a view which is not reconcilable to a New Englishes perspective of BSAfE):

As was stated at the outset, **to sound like a native** one has to know 'how things are said', one needs to know the formulae and, 'conventional expressions' used by a linguistic subculture, and the **standard** ways of talking about familiar ideas (Langacker, 1983). It is clear the speakers of Xhosa English in this corpus have mastered many such lexical strings, and use them to sound both natural and **normal** (and thereby claim **membership of the group**) while at the same time playing the language game with apparent ease (De Klerk, 2006:100-101, my emphasis – YVB).

One assumes that De Klerk (2006) meant that Xhosa English speakers attempt to sound fluent in Xhosa English and to claim membership of that linguistic subculture or group. De Klerk also compares the discourse markers *actually* and *well* in Xhosa English and New Zealand English (2006:155-186, based on De Klerk, 2004 and De Klerk, 2005a) reporting percentages of occurrences. In another chapter, which is based on an earlier article (De Klerk, 2005b), De Klerk (2006:187-201) compares Xhosa English to New Zealand English with regard to intensifying adverbs. This time frequencies per 1000 words are reported.

What all of the studies of BSAfE cited above have in common is that the features they examine are based on differences between BSAfE and Standard English. Many of the distinctive features under investigation were first described as "errors". De Klerk actually predicts:

Many of the other characteristic features, such as those listed by Adey (1977), Scheffler (1978), Gough (1994, 1996a, 1996b), Buthelezi (1995), de Klerk and Gough (2002) and others have yet to be explored in the corpus, and will no doubt be the topic of future research (2006:211).

It would seem that we have not seen the last of studies on BSAfE where the focus is on irregularity (in comparison to Standard English) and not on regularities or patterns that can be seen in the data itself. With the exception of Mesthrie's work on left dislocation and anti-deletion (1997, 2006), the BSAfE studies mentioned so far are not conducted

from the perspective of a clearly defined linguistic theory. In the present study, corpus data will be interpreted in terms of a usage-based theory of language and grammar.

The fourth wave of enquiry into the characteristics of BSAfE goes beyond a usage-based comparison of BSAfE and Standard English, or native English, and attempts to uncover the grammatical system of BSAfE in its own right. In this sense it is data-driven. Van Rooy's work since 2006 typifies this approach (Van Rooy, 2006, 2008a, 2008b, 2010, 2011; Van Rooy & Terblanche, 2006, 2009, 2010). Here the focus is not on how a feature of BSAfE differs from some external norm (typically Standard English), but on the extent to which a new norm emerges from the data, i.e. the extent to which endonormative stabilization has taken place. Van Rooy (2008b) demonstrates that an investigation of syntagmatic structure in the context of usage can lead to the identification of previously unidentified patterns with regard to tense and aspect in BSAfE. The studies cited here have made use of corpus data, in particular the *Tswana Learner English Corpus* (TLEC) and fall into two categories with regard to study object (though the theoretical premises are the same), namely multi-dimensional accounts of grammatical features that cluster together to serve certain discourse functions (Van Rooy & Terblanche, 2006, 2009; Van Rooy, 2008a) and in-depth analysis of one area of the grammar. The latter kind of study has focused on the verbal group (Van Rooy, 2006; Van Rooy 2008b, Van Rooy, 2009) or on morphological processes (Terblanche, 2009; Van Rooy & Terblanche, 2010).

The current study aligns itself with the view that grammar features are at least partly attributable to the discourse functions that they serve in the communicative contexts in which they are used and that BSAfE is an independent variety of English about which much can still be discovered with regard to its grammar. In-depth grammatical analyses in the spirit of the fourth wave of BSAfE scholarship have focused on the verbal group (Van Rooy, 2006; Van Rooy 2008b). Lexico-semantic work on BSAfE using the TLEC has also focused on verbs (Partridge, 2011). To date no in-depth corpus linguistic investigation into the grammar of nominal groups in BSAfE have been undertaken. In fact, not much in the line of detailed discussion of noun-related features of BSAfE can be found in the literature as will be seen in the section below.

1.2.2 The nominal group in descriptions of BSAfE

One aspect of the grammar of BSAfE relating to the nominal group that has received a lot of attention in the literature is pronoun copying. It is mentioned, and sometimes incorrectly conflated with the “resumptive pronoun”, in most studies that cover a number of characteristic features of BSAfE (Gough, 1996:61; De Klerk & Gough, 2002:362; Van der Walt & Van Rooy, 2002:124; De Klerk 2003a:467; De Klerk, 2003b:225; Mesthrie, 2004:972; Makalela, 2004:362; Mesthrie, 2006:124-126). It is discussed in depth by Wade (1995), Mesthrie (1997), Minow (2010) and Siebers (2012). Wade (1995:195) provides the following example of pronoun copying from his BSAfE data:

- (1) and er the lady who is up here at the crèche she still needs some help for that ... (Wade, 1995:195).

Wade (1995) regards the use of copy pronouns as fossilized in BSAfE. Wade (1995:194) acknowledges that copy pronouns also occur in native English, but points out that copy pronouns in native English is a marked strategy reserved for contexts where there is a strong motivation to explicitly signal the topic, typically where an old (“given”) topic is reintroduced. In contrast, the copy pronoun in BSAfE is a less marked construction and serves a more general purpose, namely “simply to signal a change in topic referent” (Wade, 1995:194). Mesthrie (1997:127) establishes through quantitative comparison with white South African English that topicalization phenomena are used much more frequently in BSAfE. Mesthrie’s data indicate that the most prevalent topicalization strategy in BSAfE is left dislocation with subjects. The examples below show that left dislocation is basically what is meant by the use of the copy pronoun.

- (2) Oh, Haroun, he was the coordinator. Farouk, that’s my economics teacher (Mesthrie, 1997:127, 131).
- (3) Tswana, I learnt it in Pretoria (Mesthrie, 1997:127, 131).

Like Wade (1995), Mesthrie (1997:131ff) explores the pragmatic motivations for left dislocation. Mesthrie (1997:130-133) shows that much of his data cannot be accounted for by the pragmatic functions typically ascribed to left dislocation in descriptions of native English, which are the reintroduction of given information, contrast and listing. He

also points out that the left dislocations in his data that cannot be attributed to the aforementioned pragmatic functions also cannot all be explained by the signalling of topic changes, particularly those that involve the noun *people* (Mesthrie, 1997:132). Of the 783 uses of the appositive (copy) pronoun that De Klerk (2003a:468) counts in her data, she reports 221 instances of *people they* without further comment. Minow's database of spoken BSAfE yielded 221 instances of left dislocation of which 31 involved the noun *people* (2010:188). Minow (2010:189) hypothesizes that for at least one speaker in her database, the use of *people* in subject position triggers left dislocation, but does not explore the matter further. Siebers (2012:210) observes the referent tracking function of copy pronouns where the noun *people* is followed by a postmodifier and also notes the use of the pronoun *they* after the partitive *of*-constructions as in *some of the people they [...]*.

Wade (1995:193) acknowledges the possible influence of the substrate languages in the use of the copy pronoun (in the form of the subject agreement marker affixed to the verb or the absolute pronoun following the subject in Bantu languages), but also points out that pronoun copying is widespread in New Englishes and creoles. With regard to attributing left dislocation in BSAfE to the influence of the substrate languages, Mesthrie remarks:

However, the idea that it is the **concord** pattern of Xhosa (or other language) *per se* that induces left dislocation is patently simplistic, given that subject prefixes to verbs are purely grammatical in African languages, whereas left dislocation serves a range of pragmatic purposes (1997:139, emphasis in the original).

In spite of the widespread occurrence of copy pronouns across New Englishes and the pragmatic functions of left dislocation, Makalela insists that topic promotion devices in BSAfE are due to “an underlying Bantu substrate system” (2004:361-362) arguing that they serve both morpho-syntactic logic and pragmatic function. Neither morpho-syntactic logic nor pragmatic functions adequately explain left dislocation in the syntactic environment of the noun *people*. This suggests that the noun *people* itself deserves further investigation.

The non-standard use of articles in BSAfE is noted in most overviews of the features of BSAfE (Gough, 1996:61; Van der Walt and Van Rooy, 2002:120; De Klerk, 2003a:472-

472; De Klerk, 2003b:234; Mesthrie, 2004:970). Both Minow (2010) and Siebers (2012) select article usage as features to focus on in their book-length accounts of selected sets of BSAfE features. Both refer to the specific/non-specific distinction made in New Englishes versus the definite/indefinite distinction made in native English (Platt, Weber & Ho, 1984) but they do not probe their data to establish whether one or the other system is operative in BSAfE. In the current literature, article usage is described in terms of omission, insertion and substitution, i.e. in terms of deviations from the standard. There has been no explicit attempt to check whether article choice in BSAfE can be ascribed to a different conceptual distinction. The view taken in this study is that articles need to be investigated closely in their context of usage to determine the system underlying the use of articles in BSAfE and that a specific/non-specific distinction rather than a definite/indefinite distinction should not be assumed.

Other features pertaining to nouns or the nominal group noted in overviews of features of BSAfE are the use of mass (or non-count) nouns as count nouns (Gough, 1996:61; De Klerk, 2003a:472; Siebers, 2012:134), the omission of the plural suffix on plural count nouns (Gough, 1996:61; Siebers, 2012:136) and the use of quantifiers normally associated with mass nouns (e.g. *much*) with count nouns (Siebers, 2012:136-137). The use of *this* and *that* as determiners of plural nouns are also mentioned (Siebers, 2012). Under the heading of quantifiers the use of the *other(s)...other(s)* construction and combined quantifiers such as *some few* are also mentioned (Gough, 1996:63). These features all pertain to quantification. Until now observations about quantification in BSAfE literature have been restricted to listing of non-standard constructions with no attempt at discovering what motivates the use of these constructions. De Klerk (2003a: 472) concludes from a number of non-standard collocations of (quantifying) determiners and nouns that there is a “loss of distinction between mass and count nouns” in BSAfE, but does not substantiate this claim with any analyses of the available corpus data.

Non-standard use of relative pronouns (Gough, 1996:62), *X's first time for the first time that X* (Gough, 1996:63), and gender conflation in third person singular pronouns (Gough, 1996:59) are also mentioned in BSAfE literature.

1.2.3 Focus of this study

As was indicated in Section 1.2.1 above, there are no in-depth studies of the nominal group in BSAfE. The present study will complement studies that focused on the verbal group (Van Rooy, 2006, 2008b, 2009) by giving a linguistic account of the grammar of nominal groups. This study does not pretend to be a comprehensive account of the grammar of nominal groups in BSAfE. Rather, it is an attempt to gain insight into grammatical patterns pertaining to phenomena that have previously been misunderstood or underexplored. This study will provide a usage-based account of such phenomena, making use of a corpus of written BSAfE (the *Tswana Learner English Corpus*) and a native control corpus (*The Louvain Corpus of Native English Essays*).

Since the noun *people* is specifically mentioned in BSAfE literature (with reference to left dislocation) and since it is a high-frequency noun due to its general meaning and potential to function as a cohesive device, concordances of *people* will be taken as a point of departure in exploring the corpus data. Non-standard uses of determiners and quantifiers are also frequently mentioned in BSAfE literature, but to date there has not been a thorough linguistic investigation of determiners and quantifiers.

1.3 Research questions and aims

The overarching aim of this study is to contribute to a more thorough understanding of the grammar of the nominal group in BSAfE by answering the following three questions:

1. What are the syntagmatic patterns associated with the noun *people* and what do they reveal about the structure of nominal groups in BSAfE?
2. What are the systems that underlie determiner (particularly article) choice in BSAfE?
3. What are the typical structures that realize quantification in BSAfE and how do they differ from Standard English.

1.4 Prospectus

The first question is addressed in Chapter 4, in which concordances of the noun *people* are retrieved from both corpora and all concordance lines are analyzed with regard to the elements of the nominal group. The second question is addressed in Chapter 5 which centres on analyses of concordances of the articles and demonstrative and possessive determiners. The third question is addressed in Chapter 6 by means of analyses of concordances of the words typically functioning as quantifier. Chapter 2 provides a theoretical framework for analysis and interpretation of the data, while Chapter 3 provides a methodological framework.

CHAPTER 2

Theoretical framework

2.1 Introduction

The investigative aims and methods of this study situate it firmly within a usage-based, as opposed to a formalist, approach to language. The aim of this chapter is to establish a framework for analyzing and interpreting corpus data. Section 2.2 provides an overview of the approaches to language and linguistic phenomena that are usage-based in nature and introduces a number of theoretical concepts relevant to this study. Section 2.3 presents the concepts and categories relevant to the analyses of the corpus data with specific reference to the functional elements of the nominal group. Section 2.4 provides an exposition of additional concepts that are useful in interpreting the corpus data, especially high-frequency phenomena. These include constructions, patterns and conventionalization.

2.2 A usage-based understanding of language

A corpus-linguistic investigation of linguistic phenomena implies a usage-based approach, but usage-based studies of language need not necessarily make use of corpora. Bybee traces usage-based approaches to the “tradition of explicitly studying the functions of grammatical constructions” (2010:11) starting in the 1970s and exemplified by the work of Givon, and the surface-oriented approaches to grammar that gave rise to the notion of construction as form-meaning pair piloted by Langacker in the 1980s and further developed by, for instance, Goldberg (1992) and Croft (2001). Cognition and conceptualization as well as the role of meaning in grammar are taken into account in both functionalist and constructionist approaches to language, as will be demonstrated later this chapter. It is not always easy to separate functionalist and constructionist approaches from each other. For instance, Butler (2003:34) regards

Cognitive Grammar (henceforth CG) by Langacker, Systemic Functional Grammar (henceforth SFG) (primarily associated with Halliday) and West Coast Functionalism (under which Givón resorts) as examples of structural-functional grammars. The commonalities he observes among these include a view of language as a means of communication; a rejection of syntax as self-contained in favour of a view of grammar where semantics is central and where syntax is motivated by meaning; a rejection of the self-containedness of grammar in favour of a functional view of language that takes cognitive and socio-cultural factors into account; as well as a recognition of the cognitive dimension of language usage (Butler, 2003:33). Each of these four aspects will be considered in the remainder of this section.

Functionalist approaches to linguistics have in common the view that language should be studied from the perspective of the functions that it fulfills. Because it asks, “What is language used for?” / “What is the purpose/function of language?” it is a teleological approach to language (Givón, 1995:3). In functionalist approaches to language, meaning (semantics) and context (discourse/pragmatics) are important aspects of language function, entailing that morphology and syntax are not described without reference to meaning and discourse. Though both Halliday and Givón view language as a means of communication, they differ in the way in which they define the functions of language. In Hallidayan SFG, the focus is on the social and cultural functions of language (Butler, 2003: 44). Givón recognizes the social aspects of language usage, but gives primacy to “the information-processing function of language” (1993:21). Givón (2001:7) regards the **representation and communication of knowledge (experience)** as the primary functions of language. Cognitive linguists subscribe to the idea that “[l]anguage is shaped and constrained by the functions it serves” (Langacker, 2008: 7) and recognize the communicative function of language, but cognitive linguistics is distinguishable from other functionalist approaches in the priority which is given to **conceptualization**: “It fully acknowledges the grounding of language in social interaction, but insists that even its interactive function is critically dependent on conceptualization” (Langacker, 2008:8). Cognitive linguistics emphasizes the semiological function of language, namely the symbolization of conceptualizations by means of sounds and gestures (Langacker, 2008:7). Since a demarcation of the functions of language is central to a function-based analysis of grammatical features and other linguistic phenomena, the functions of language as it is defined by Halliday and

Givon will be discussed in more detail in separate subsections below (see 2.2.1 and 2.2.2).

Functionalist approaches can be distinguished from a formalist view of language with respect to the status of syntax within the grammar. Croft defines grammar as “an individual’s knowledge of their language” which includes in addition to knowledge of syntax, “knowledge of the conventional semantic, pragmatic and discourse functions of the syntactic forms” and concludes that “grammar is a semiotic system including both syntax and semantics” (1995:492). While the notion of grammar as a semiotic system is not denied in formalist approaches to language, a clear distinction is made between the syntactic component and the semantic component, which privileges syntax to the extent that it is sometimes used as a synonym for grammar (see Croft’s (1995:495) interpretations of Chomsky (1977)). In contrast, functionalist linguistics “argue for a certain degree of interpenetration of semiotic function and syntactic form” (Croft, 1995:493). In formalist linguistics the lexicon is also seen as distinct from the grammar, or (morpho)syntax. In some strands of functionalism, lexis and grammar are seen as inseparable and often referred to as the lexicogrammatical continuum (Langacker, 2008:5, Halliday and Matthiessen, 2004:25, 30).

Since functionalists explain grammatical phenomena in terms of their functions, the delineation of grammatical categories in functional linguistics differ somewhat from that of formal linguistics. In formalist approaches the category of noun, for example, is defined initially in terms of its morphological properties but principally in terms of its syntactic properties (Radford, 2004:28). In terms of such an approach, a noun is a word which can inflect for plural. Radford (2004:29) acknowledges that mass nouns and nouns such as *sheep* which do not inflect for plural are problematic if only morphological properties are taken into account, and concludes that syntactic properties should therefore also be considered. Within a purely formalist categorization no reference is made to the meaning of nouns, let alone the discourse functions of nouns. The following remark typifies the formalist approach to word categorization:

Given that different categories have different morphological and syntactic properties, it follows that we can use the morphological and syntactic properties of a word to determine its categorization (i.e. what category it belongs to) (Radford, 2004:33).

The opposite of the above approach is typified by Fawcett (2000:204), a systemic functionalist, who advocates that “we should use the internal structure and semantics of a unit to determine its class” and further that “an element of structure should be defined in terms of the **function** it serves” (Fawcett, 2000:220). The systemic view of grammar provides for a description of grammatical categories “by reference to what they mean” (Halliday & Matthiessen, 2004:10). Halliday and Matthiessen describe their grammar of English as “a ‘semanticky’ kind of grammar” (2004:31).

Somewhere between these formalist and functionalist extremes are accounts of English grammar which take both form and function into account when characterizing grammatical units such as nouns or nominal groups. However, due to the fact that they do take function into account, such approaches are regarded as functional. The major reference grammars of English (Quirk et al., 1985; Biber et al., 1999; Huddleston & Pullum, 2002) are examples of such approaches which consider the internal structure of units (in the case of nouns, its morphology; and in the case of noun phrases, its constituent elements), their syntactic role in the higher units in which they occur (in the case of nouns, in phrases; and in the case of noun phrases, in clauses for instance) as well as the meanings they encode. Givon (1993:51ff) also characterizes the lexical word classes of English in terms of semantic, morphological and syntactic criteria.

With cognitive grammar’s emphasis on cognition in the delineation of grammatical categories, it is more extreme than the approaches which consider internal structure, syntactic role and semantics in the classification of grammatical units, such as the reference grammars mentioned above. Langacker (2008:96) points out that even where cognitive grammar does retain traditional standard grammatical terminology, the classes defined by cognitive grammar are not precisely co-extensive with traditional ones. The differences arise from the centrality of cognition in defining basic categories such as nouns or verbs.

It is only by recognizing the crucial role of cognition – how situations are apprehended and conceptualized – that semantic characterizations become feasible. Especially relevant are two aspects of construal: profiling and level of specificity (Langacker, 2008:98).

The notion of grammar as a semiotic system is illustrated in Halliday's brand of SFG in which **semiosis**, "the making and understanding of meaning" (Halliday & Matthiessen, 2004:5) is central. Grammar is regarded as a meaning-making resource "and meaning resides in systemic patterns of choice" (Halliday & Matthiessen, 2004:23). Here the word system is used in a slightly different sense than it has been used so far in this chapter. Croft's use of the term system as a synonym for grammar, "the psycholinguistic system of the individual – called grammar" (1995:493), exemplifies the way in which "system" is generally employed in usage-based linguistics. In SFG, system refers to sets of choices that a speaker makes with regard to the meanings which he wishes to realize through the lexicogrammar. These two senses of the term "system" are not incompatible, but related.

Butler (2003:48, 51, 55) demonstrates that the cognitive dimension is recognized in SFG and by Givon, but that it is "paramount" for Langacker. However, where Givon invokes cognition to explain linguistic phenomena, Halliday and Matthiessen (1999:x) attempt to "explain cognition by reference to linguistic processes". The term "usage-based" is normally reserved for an approach that aims to explain linguistic phenomena in terms of cognition.

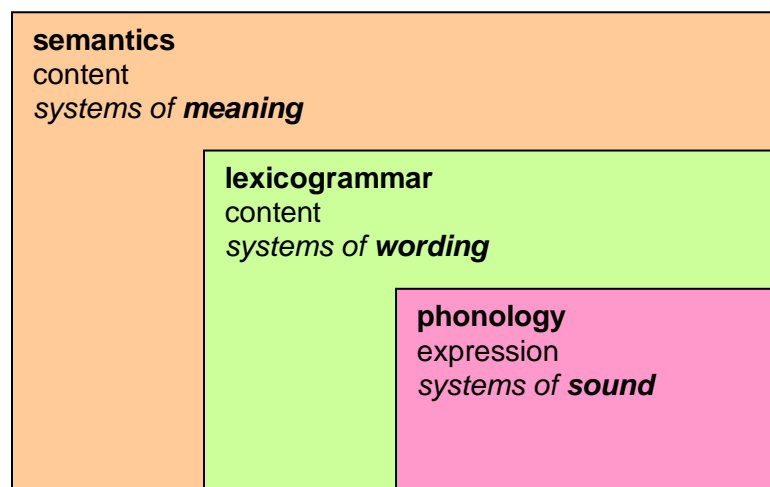
Bybee maintains that "the hypothesis that instances of use impact the cognitive representation of language" (2010:14) is central to the usage-based position and argues for an exemplar representation of language. This basically entails that generalizations are made from exemplars. Kemmer and Barlow remark that "[a] usage-based model is one in which the speaker's linguistic system is fundamentally grounded in 'usage events': instances of a speaker's producing and understanding language" (2000:viii). They emphasize three aspects of the link between instances of use ('usage-events') and the individual speaker's linguistic system. Firstly, the system is abstracted from the instances of use, or as Langacker (2000:4) would have it, schemas are abstracted from instantiations. Secondly, the link between the schema and its instantiations remains strong so that general representations are activated in concert with specific instances of a given pattern. Thirdly, language productions, i.e. the product of an individual speaker's linguistic system, are in themselves usage-events that serve as input for other speakers' or the speaker's own system.

With regard to the first aspect, Langacker maintains that “‘rules’ can only arise as schematizations of overtly occurring expressions. However far this abstraction may proceed, the schemas that emerge spring from the soil of actual usage” (2000:3). This claim demonstrates the bottom-up approach of cognitive grammar and usage-based approaches in general (also see Bybee 2010:15) as opposed to the top-down approach of generative grammar that emphasizes general rules and universal principles (Langacker, 2000:2).

2.2.1 Functions and structural categories in SFG

Halliday and Matthiessen (2004:4, 24ff) distinguish a content and an expression plane of language. In current Hallidayan functional linguistics, language is viewed as a **tri-statal** system in which a semantic, a lexicogrammatical and a phonological level can be distinguished (Halliday & Matthiessen, 1999:5) as can be seen from the diagram below (based on those provided by Halliday and Matthiessen, 2004:25, 30; 1999:5):

Figure 2.1 The stratification of language



The strata of semantics and the lexicogrammar occur on the content plane, whereas phonology (in the case of speech) and graphology (in the case of written texts) occur on the expression plane (Halliday & Matthiessen, 1999:4; Halliday & Matthiessen, 2004:24-25). The diagram of the strata above (and also in the various sources on which it is

based) attempts to “show the stratal environment of each level” and therefore the concentric shapes¹ should not be read as a relationship of “consist of” or “is a subset of”, but should be interpreted as environments: “thus lexicogrammar appears in the environment of semantics and provides the environment for phonology” (Halliday & Matthiessen, 1999:4). The relationship between these strata is that of **realization**: meaning is realized in wording (i.e. the lexicogrammar) which, in turn, is realized through speech sounds or orthography.

The stratum of the **lexicogrammar** consists of the lexis and the grammar of a language. These are not separate, but two ends of a lexicogrammatical continuum (Halliday & Matthiessen, 2004:43) – as Halliday and Matthiessen (1999:5) put it: “The lexical region, or lexis, is not a separate component, but simply the most ‘delicate’ end of the (unified) lexicogrammar.” On the **semantic** stratum, three **modes of meaning** are distinguished. These modes of meaning are called the **‘metafunctions’**² of language. The **ideational** metafunction pertains to language as a construal of human experience, i.e. “language as reflection”, whereas the **interpersonal** metafunction refers to language as a means of enacting personal and social relationships, i.e. “language as action” (Halliday & Matthiessen, 2004:29-30). The **textual** metafunction refers to the grammar’s ability to organize information and create cohesion and continuity (Halliday & Matthiessen, 2004:30). The textual metafunction of language relates propositions (clauses as “messages”) in a text to each other and to the context (Thompson, 2004:28; Halliday & Matthiessen, 2004:309).

Meaning is realized by the lexicogrammar. Halliday and Matthiessen (2004:21) refer to grammar as “the powerhouse where meanings are created”. It is in the structure of the clause that all three modes of meaning – (a) the ideational, (b) the interpersonal, and (c) the textual – are realized (Halliday & Matthiessen, 2004:10; Thompson, 2004:33). Three different structures of the clause can be distinguished in accordance with the three kinds of meaning on the semantic level (Halliday & Matthiessen, 2004:309). The **transitivity**

¹ Concentric circles in Halliday and Matthiessen (2004:25 and 1999:5); rectangles in Halliday and Matthiessen (2004:30).

² The term ‘metafunction’ rather than just ‘function’ is used in SFG, because ‘function’ has very wide usage and may simply mean “purpose” or “way of using language”, whereas the term ‘metafunction’ emphasizes the fact that function is an integral component of the theory (Halliday & Matthiessen, 2004:30-31). Literally ‘metafunctions’ imply a kind of umbrella usage, so that each of the three metafunctions may subsume more specific functions (Thompson, 2004:28).

structures of the clause are said to realize **ideational/experiential/representational** meaning. When the clause is analyzed as **representation**, one or more Participants, a Process and, optionally, one or more Circumstances can be distinguished³. Interpersonal meaning is realized in **modality structures**. When the clause is regarded as **interpersonal exchange** a Mood ‘region’ and a Residue ‘region’ can be distinguished. The Mood consists of the Subject and the Finite. In the realization of **textual** meaning, the clause is regarded as “**message**” with an **information structure** in which a Theme and a Rheme can be distinguished. The functional constituents in a clause structure are realized by groups or phrases of various classes. The sequencing of these groups and phrases is referred to as a **syntagm** (Halliday & Matthiessen, 1999:9). It is this syntagm that can be analyzed into three different clause structures based on the three metafunctional lines of meaning which are realized by the clause. This is illustrated in Figure 2.2

Figure 2.2 The three types of structure in the clause according to Hallidayan SFG

Example ⁴	<i>many</i>	<i>children</i>	<i>are</i>	<i>born</i>	<i>everyday</i>
textual structure	Theme		Rheme		
interpersonal structure	Mood			Residue	
	Subject		Finite	Predicator	Adjunct
experiential structure	Participant		Process ⁵		Circumstance
syntagm	nominal group		verbal group		adverbial group

The block-diagram above shows how the structural constituents are mapped onto the syntagm, i.e. the sequence of grammatical units (groups, in this case). The fact that the structure of the clause can be analyzed in three different ways, depending on the metafunctional meaning, entails that the units in the syntagm (groups or phrases) are multifunctional. For instance, the **nominal group** in the table above serves three functions. It realizes the function of Participant in the experiential structure of the clause, while it realizes the function of Subject in the interpersonal structure of the clause, and the function of Theme in the textual structure of the clause. The **verbal group** in the

³ In SFG, the functional labels of constituents have an initial capital letter by way of convention.

⁴ Example from corpus essay <ICLE-TS-KIMC-0294.1>.

⁵ In Hallidayan SFG, the Process is regarded as being realized by the entire verbal group, i.e. the main verb, which designates the type of process, as well as the auxiliaries which express tense, aspect, modality and polarity (Halliday & Matthiessen, 2004:177).

example above realizes the function of Process in the experiential structure of the clause. With regard to the interpersonal structure of the clause, the auxiliary verb realizes the Finite in the Mood, while the lexical verb realizes the Predicator in the Residue. In the textual structure of the clause above, the verbal group and the adverbial group are part of the Rheme. The **adverbial group** realizes the function of Circumstance in the experiential structure and the function of Adjunct in the interpersonal structure of the clause.

The classes of **group** distinguished in SFG, namely **nominal** groups, **verbal** groups and **adverbial** groups are named after the three major word classes which are recognized in SFG. The classification of words in SFG differs from the word classes (parts of speech) of traditional grammars. This is because word classification in SFG is based more on function than on (morphological) form. In SFG, the name of a word class, such as “nominal”, “verbal” and “adverbial”, indicates “its potential range of grammatical functions” (Halliday & Matthiessen, 2004:52). The word class label “**nominal**” includes the traditional classes of *noun*, *adjective*, *numeral* and *determiner* (also see Halliday & Matthiessen, 2004:320). However, in the Cardiff strand of SFG, adjectives are not treated as a kind of nominal and an additional class of group, the quality group is recognized (Fawcett, 2008; Tucker, 1998). *Pronoun* is regarded as a type of *noun*. Payne and Huddleston (2002:327) also treat the pronoun as a kind of noun, because pronouns can head phrases which typically have the same syntactic role as phrases headed by nouns.

Traditional grammars favour the term “phrase” (as in noun phrase, verb phrase, etc.) and do not use the term “group”. In Hallidayan functional grammar a distinction is made between groups and phrases although they both rank below the clause. “Group” means “group of words” (Halliday & Matthiessen, 2004:310). In a certain sense a group is a **word complex**, which means that the nominal group might be seen as a complex of nominals, or nominal words. The term “word complex” means that the words are not randomly grouped together, but are combined on the basis of a **logical** relation. For example, in the clause *many children are born every day*, there is a logical relation between *many* and *children*: *children* is the **head** of the group *many children*. It is not possible to recognize any logical relation between *children* and *are*; i.e. it is not possible to identify a head and a modifier for *children are*. It is due to the fact that groups are

word complexes that they are called word groups as opposed to phrases. Halliday and Matthiessen (2004:311) explain the difference between groups and phrases as follows: “A phrase is different from a group in that, whereas a group is an expansion of a word, a phrase is a contraction of a clause.” SFG recognizes both prepositional groups and prepositional phrases in English (Halliday & Matthiessen, 2004:359) and these are useful to illustrate the distinction between group and phrase. In the examples from the Tswana Learner English Corpus below, the preposition groups are underlined and the preposition phrases are in boldface. In the groups *all through* and *all across*, *all* serves as **modifier** of the preposition, and is non-obligatory. The head-modifier relationship is a logical relationship associated with word groups. In the examples below the prepositions are followed by an obligatory **Complement**. Complement is a syntactic role, i.e. an element of the clause (cf. Quirk et al., 1985:49, 54). Halliday & Matthiessen (2004: 359-360) regard the preposition as a kind of “minor verb” which encodes the Predicator function in the “clause-like” interpersonal structure of the prepositional phrase, while the nominal group following the preposition encodes the Complement. The preposition phrase (italics in the examples below) can also be analyzed like a clause experientially. The preposition (underlined in the examples below) then functions as a “minor Process” and the nominal group which follows it as “Range” (Halliday & Matthiessen, 2004:361) (Also see Figure 2.3 below).

- (4) ...it runs *all through* *the year* <ICLE-TS-NOUN-0098.1>
- (5) ...not only *in* *Africa* but *all across* *the continent*. <ICLE-TS-KIMC-0365.1>
- (6) ...they are (sic) *just like* *a pumpkin* <ICLE-TS-NOUN-0034.1>
- (7) ...put their hands *deep in* *their pocket* <ICLE-TS-NOUN-0098.1>

Figure 2.3 Preposition phrase and preposition group structure

Preposition phrase: (TLE:TSNO1098)	<i>all</i>	<i>through</i>	<i>the</i>	<i>year</i>
Phrase structure: Interpersonal	Minor Predicate		Complement	
Phrase structure: Ideational	Minor Process		Range	
Phrase syntagm	prepositional phrase			
	preposition group		nominal group	
Logical structure of the word group	Modifier	Head	Modifier	Head
Group syntagm	adverb	preposition	determiner	noun

The position in Hallidayan functional grammar, then, is that on the rank between clause and word, two kinds of grammatical units can be found; phrases and groups, the former more clause-like and the latter more word-like. Langacker also (2008:310) argues against the use of the term “noun phrase” pointing out that “nominals are not always phrases, nor do they always contain a noun (as traditionally understood)”. For Langacker (1972:194), the defining characteristic of a nominal is its syntactic function “rather than internal structure”, entailing that “the ability to function as a subject or object” is what distinguishes a grammatical unit as a nominal. In terms of this definition elements headed by an adjective (e.g. *the rich*) or a pronoun (e.g. *he, those, everyone*), and even subordinate nominal *that-* and *what-*clauses (e.g. *what distinguishes a grammatical unit as a nominal*), can be regarded as nominals. All of these are capable in SFG terms to realize the clausal function of Participant, and it is therefore perfectly sensible to use the term “nominal”. However, groups and (subordinate) clauses differ significantly in terms of their internal structure and therefore some kind of terminological distinction is necessary. The terms “nominal group” and “nominal clause” indicate the functional similarity as well as the structural difference between these two units.

2.2.2 Functions and structural categories in Givón

Givón (1993:2), who is representative of American functionalism, regards the function of language as an integral part of its description. He summarizes his perspective as follows: “human language is a purposeful instrument designed to code and communicate information, [...] and its structure is not divorced from its function”. He views grammar as “a unique, human-specific device for coding and communicating information” (Givón, 1993:4). This view of the role of grammar is very similar to the view of grammar as a meaning-making instrument held in systemic functional linguistics. Givón’s (1993:21) description of the functions of language, however, differs subtly from the three metafunctions distinguished in Hallidayan SFG. He identifies two major tasks of language, namely (a) “mental representation of experience”, and (b) “its communication to others” (Givón, 1993:21). In a more recent description of his brand of functional linguistics, he refers to these primary functions of language as (a) the representation of knowledge and (b) communication of knowledge (Givón, 2001:9). Givón (1993:21) also mentions three (although he points out that there are more) meta-communicative

functions of language, namely (i)⁶ socio-cultural cohesion functions, (ii) interpersonal affective functions, (iii) aesthetic functions. These are clearly relatable to the interpersonal metafunction of language in SFG, especially as it was formulated by Halliday in 1970:

Language serves to establish and maintain social relations: for the expression of social roles, which include the communication roles created by language itself – for example the roles of questioner or respondent, which we take on by asking or answering a questions; and also for getting things done, by means of the interaction between one person and another. Through this function, which we may refer to as interpersonal, social groups are delimited, and the individual is identified and reinforced, since by enabling him to interact with others language also serves in the expression and development of his own personality(Halliday, 2002:143).⁷

Grammar is involved in the realization of all of the aforementioned meta-communicative functions of language, but grammar is primarily employed in the coding and communication of information; as Givon (1993:21) puts it: “The bulk of our grammatical apparatus finds its primary use in the **information-processing** function of language, that is in the **mental coding** and **verbal communication** of information” (my emphasis – YVB).

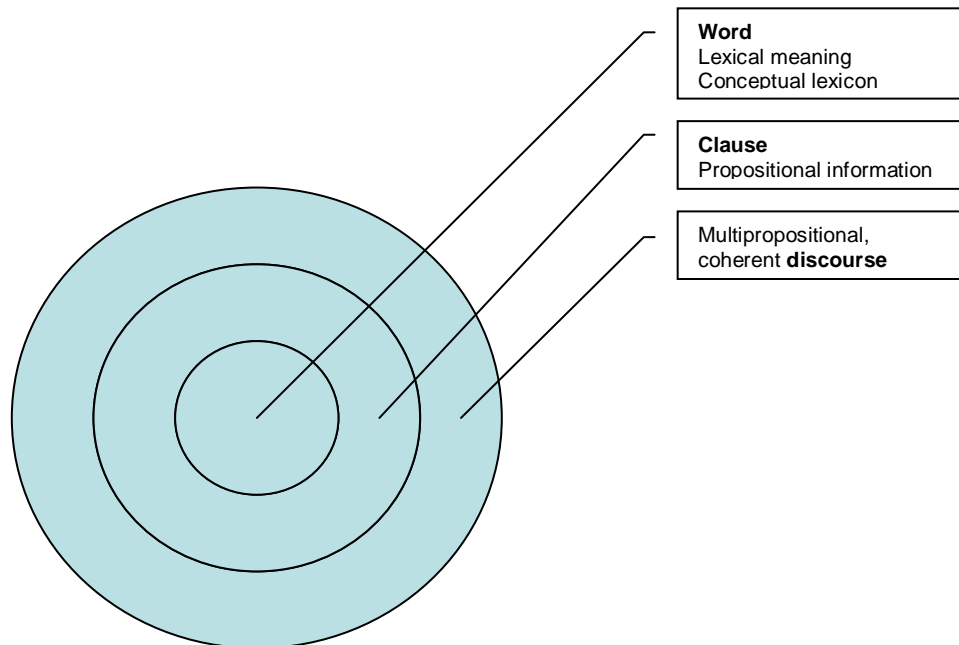
Givon (2001:7) derives two sub-systems of the human communicative system from the two major functions of language, namely: (a) the **cognitive representation** system (for knowledge representation) and (b) the **communicative coding** system (for the communication of knowledge). There are two coding instruments in the communicative coding system, namely, the **peripheral sensory-motor codes** and the **grammatical code** (Givon, 2001:7). Givon (1993:21) distinguishes three functional realms of language, which are arrayed concentrically. These are (1) Words (meaning), (2) Clauses (information) and (3) Discourse (coherence). Words code our **concepts of entities**, which in turn are combined into propositions in clauses which provide information about **relations, qualities, states** or **events** in which the entities partake. Propositions are combined into **coherent text** in discourse. In earlier work, Givon (1984:31) refers to the functional realms of language in terms of semantics and

⁶ The in-text numbering formats are varied here to separate the metacommunicative functions of language, the functional realms, and the coding systems from each other.

⁷ This is from a paper originally published in 1970. The page reference is to the 2002 publication in which the article is reprinted.

pragmatics, distinguishing (1) lexical semantics, (2) propositional semantics and (3) discourse semantics. In more recent work, Givon (2001:7) refers to “three concentrically linked levels” of the human **cognitive representation system**, and identifies these as (1) the conceptual lexicon, (2) propositional information and (3) multi-propositional discourse. The diagram below shows the functional realms of language associated with the cognitive representation system and the linguistic units which realize them:

Figure 2.4 The functional realms of language distinguished by Givon



The **conceptual lexicon** is defined as “a repository of relatively time-stable, socially-shared, relatively well-coded concepts, which, taken together, constitute a cognitive map of our experiential universe” (Givon, 2001:7). This experiential universe refers to the external-physical universe, the social-cultural universe and the internal-mental universe (Givon, 2001:8). The meanings of words may change over a fairly long period of time, but they are not “in rapid flux” and are therefore regarded as time-stable. Due to the assumption of speakers that everybody attaches roughly the same meaning to a word,

the conceptual lexicon is seen as socially shared. Its perception as well-coded results from the fact that there is a strong tendency for unique relationships between a code (word) and its meaning, with homophony being the exception (Givon, 2001:8 and 1995:395-396). The conceptual lexicon is regarded as a network of nodes and connections (Givon, 2001:8). At each node an individual concept with its own distinct meaning and code-label occurs (Givon, 1995:396 and 2001:8). Here 'concept' refers to "types of conventionalized experience, rather than individual token" (Givon, 1995:396). Concepts, or types, arise from conventionalization, i.e. sufficient repetition to allow for grouping according to similarities, while tokens are unique experiences which one is able to recognize as belonging to an established type. It is important to note that concepts/types are generic, whereas tokens are specific (Givon, 2001:8). Givon (1995:396) links concept representation to word classification:

A lexical concept may represent a relatively time-stable entity - physical object, landmark, location, plant, animal, person, cultural institution, or abstract concept – thus typically a *noun*. Or it may represent an action, event, process or relations, thus typically a *verb*. It may represent a quality, property or temporary state, thus typically an *adjective*. Or it may convey some adverbial meaning, thus typically an *adverb*.⁸

In cognitive psychology, the conceptual lexicon is referred to as "permanent semantic memory" (Givon, 2001:8). The word classes referred to in the quote above (noun, verb, adjective and adverb) are lexical word classes. Givon (1993:46) remarks, "Individually and as a network, this lexical vocabulary represents our shared physical and cultural universe". Givon (1993:46) distinguishes between lexical ('content') words and non-lexical ('function') words. The latter are not regarded as "words" in the ordinary sense of the word, but as **morphemes**. Grammatical morphemes contribute to the grammatical structure of clauses, meaning that they contribute to the coding of propositional meaning and discourse pragmatics (Givon, 1993:47). Grammatical morphemes are therefore part of the grammatical code, and not the conceptual lexicon. The grammatical morphemes of English occur as affixes, but in some cases, are written as separate words, "as if they were independent word-stems, i.e. **free morphemes**" (Givon, 1993:51). Grammatical

⁸ Givon's (1995) description of the meaning of adverb as 'adverbial' is semantically empty. Adverbs typically denote 'circumstance' ('time', 'place', 'manner', etc.), 'degree' and 'stance' (Halliday & Matthiessen, 2004:268-269). Biber et al. (1999:552-560) provide a detailed description of the meaning of adverbs.

morphemes which are written as separate words in English include determiners, prepositions, auxiliaries and conjunctions. Givon (1993:51 and 2001:48) attributes this writing convention to the fact that grammatical morphemes arise from the grammaticalization of lexical words over time and that writing systems are slow to register such grammaticalizations. In earlier work he remarks “today’s morphology is yesterday’s syntax” (Givon, 1971:12).

Propositional information is encoded in clauses through the combining of concepts (words) (Givon, 1993:25). Clauses convey “propositional information about relations, qualities, states or events in which entities partake” (Givon, 1995:396). In cognitive psychology the ability to process and store propositional information is referred to as “episodic-declarative memory” (Givon, 2001:1). Clauses are combined to form multi-propositional, coherent **discourse** (Givon, 2001:8). Grammatical units which are smaller than clauses, but larger than words, i.e. the groups and phrases, are absent from Givon’s three-tiered model of the cognitive representation system and the realms of language. This does not mean, however, that these units are disregarded. Strictly-speaking words refer only to lexical content words. When these lexical content words combine with grammatical morphemes and/or other lexical words by way of the grammatical code, they become part of propositional information encoded by the clause. Givon (2001:45) points out that “grammatical morphemes partake in the grammatical structure of clauses”.

As mentioned earlier, the communicative coding system consists of the sensory-motor coding system and the grammatical coding system. The **peripheral sensory-motor codes** pertain to the physical and neurological aspects of language production and perception, including the production and perception of speech sounds (Givon, 2001:11). The conceptual lexicon is coded by the sensory-motor code, or what Givon (1984:31) also calls the sound code. Apart from signaling lexical meaning, the sensory motor code, by way of intonation and stress, can also signal grammatical morphology (Givon, 1995:404). **Grammar** is the coding system through which the communication of information (in clauses as components of coherent discourse) is achieved. Grammar is the coding system for both propositional semantics and discourse coherence (Givon, 1995:401 and 2001:13). The grammatical coding system thus operates on the level of

the clause as well as on the level of discourse, although it is located in the clause. Givon (2001:13) remarks:

This is indeed one of the most baffling facts about grammar-as-code: Although it is located wholly in the clause, its functional scope is not primarily about the propositional information couched in the clause in which it resides. Rather, grammar is predominantly about the coherence relations between the propositional (clause) and its wider discourse context.

According to Givon (1995:405), the part of the grammatical code that encodes propositional semantics, namely the semantic participant roles and semantic transitivity (states, events and actions), is relatively small in relation to the part that encodes discourse pragmatics: “The bulk of the grammatical code is deployed in discourse pragmatics, signaling the coherence of information within its wider – cross-clausal, situational, cultural – context”. To illustrate this point, Givon (2001:13) lists ten grammatical sub-systems that are primarily discourse oriented. These include, with regard to nominal groups, grammatical roles (subject, direct object), definiteness and reference, anaphora and pronouns, topicalization and relativization, as well as, with regard to verbal groups, tense, aspect, modality and negation. The table below shows the relation between the cognitive representation system and the communicative coding system.

Table 2.1 Representation and coding⁹

cognitive representation			coding system
component of the cognitive representation system	functional realm	structural unit	
conceptual lexicon	lexical semantics	word	sensory-motor code
propositional information	propositional semantics	clause	grammar
multi-propositional discourse	discourse pragmatics	discourse (coherence)	

For Givon the conceptual lexicon and the grammar-as-coding-device are separate; i.e. he distinguishes between grammar and lexis. Although both Givonian and Hallidayan

⁹ This table is based on explanations and representations in the work of Givon (1984:30; 1993:21-25; 1995:395, 405).

strands of functionalism regard grammar as a coding device, such a distinction between the grammar and the lexicon is not made in Hallidayan grammar where the lexicogrammar is seen as a cline. Langacker (2008:18) also argues “that a clear demarcation between lexicon and grammar is far from evident” referring specifically to idioms which are syntactically analyzable, but have a different meaning than the one that would be predicted by their parts. For Langacker (2008:43) conceptualization and construal are interrelated and inseparable:

... a meaning consists of both conceptual content and a particular way of **construing** that content. The term construal refers to our manifest ability to conceive and portray the same situation in alternate ways. ...

The distinction between content and construal is not at all a sharp one. For instance, level of specificity – an aspect of construal – has direct bearing on the content evoked: precisely because they differ in specificity, *the glass with water in it* has more content than *the container with liquid in it (...)*.¹⁰

For Halliday and Matthiessen (1999:9 ff) **construal** is what the ideational metafunction is about. One could argue that what Givon (2001:9) terms the “representation and communication of knowledge” correlates with the ideational metafunction of SFG. In their study on the linguistic construal of experience, Halliday and Matthiessen (1999) focus on the representational function of the lexical words. For Givon (1984:31), lexical knowledge (the conceptual lexicon) is a “cognitive map of the phenomenological universe”; and for Halliday (2002:174-175), “[l]anguage serves for the expression of ‘content’: that is, of the speaker’s experience of the real world, including the inner world of his own consciousness”. Givon’s description (above) of the propositional information of the clause resonates with Halliday and Matthiessen’s account (see above and the next section) of the clause’s realization of the ideational metafunction. According to Givon (1984:31) the syntactic coding of propositional information mainly involves two aspects of the proposition, namely:

- (i) Its characterization as *state*, *event*, or *action*; and
- (ii) The characterization of the *participants* (‘arguments’, ‘case-roles’) in the proposition as to their **semantic roles** vis-à-vis the predicate.

¹⁰ The citation above not only explains the interrelation between conceptualization and construal, but also provides a definition of construal and illustrates what is meant by specification. It will therefore be revisited later.

Givon (1984:47) further distinguishes between **specific** and **generic** propositional information. Specific information pertains to specific events/actions/states occurring at a particular time and place and involving particular participants, whereas generic information pertains to reality in general; what is valid for all times, places or persons/things (Givon, 1984:47).

Givon's functional realm of multi-propositional discourse (pragmatics) correlates with the textual metafunction of language in SFG. Compare the following descriptions of the text/discourse-constructing function of language:

... language has to provide for making links with itself and with features of the situation in which it is used. We may call this the textual function, since this is what enables the speaker or writer to construct 'texts', or connected passages of discourse that is situationally relevant; and enables the listener or reader to distinguish a text from a random set of sentences (Halliday, 1970:143).

The functional realm of discourse pragmatics involves the sequencing or placing of atomic propositions within a wider communicative context, i.e. in discourse (Givon, 1984:31)

The textual metafunction (or discourse coherence) pertains not only to constructing text/discourse from propositions, but also to linking the discourse (speech acts) to the communicative context (of situation and of shared cultural knowledge) and the identification of speech roles. Similar to the distinction made between the context of culture and the context of situation in systemic functional linguistics (Egins, 2004) is Givon's (1984:31-32) distinction between the "cultural common denominator" of the lexicon as a "generically-shared body of stable knowledge" and the "*specific* context" which includes the speaker's communicative and pragmatic goals; the interaction between the speaker and the hearer based on the relationship between them and what they assume about the other's knowledge, goals and predisposition; as well as the information which had already been given in the discourse and the status of information as important or not. Givon's (2001: 13) observation that the function of grammar is not only to code propositional information, but is "predominantly about the coherence relations between the propositional (clause) and its wider discourse context", underlines the fact that the textual and the ideational are distinguishable, but not separable. Givon (2001:11) regards the relation between propositional information and discourse coherence as a one-way conditional:

One can understand the meaning of clauses independent of the discourse in which they are embedded; but one cannot understand the discourse without understanding the propositions that make it up.

The ideational and textual metafunctions of language can be subsumed under what Givon terms the representational and communicative function(s) of language. The ideational metafunction maps onto representation, through the lexicon and the combination of words into clauses to convey propositional information about states, events and actions and the participants involved in these. The textual metafunction maps onto communication in the form of coherent multipropositional discourse. A proposition may be worded in more than one way (Halliday & Matthiessen, 2004: 23, 33, 175; Langacker, 2008:43) and propositions may be arranged into discourse in more than one way. Although construal and information packaging are aspects of the ideational/representational and textual functions of language respectively, construal and information packaging are influenced by the communicative situation and particularly by who the speaker(s) and hearer(s) are, and therefore also ties in with the interpersonal metafunction. Givon (1993:21) refers to the interpersonal aspects of language as meta-communicative functions. In this study, the ideational/representational, textual/discourse-pragmatic and interpersonal/interactive/social functions of language are recognized, but the focus will fall on the first two.

2.3 Framework for analysis

In order to establish an initial framework for the analysis of corpus data, structural-functional descriptions of the nominal group in the standard reference grammars of English, as well as in SFG, CG and by Givon will be surveyed. Sub-section 2.2.1 focuses on the role of the nominal group within the clause, whereas Sub-section 2.2.2 is an exposition of the internal specifying elements of the nominal group.

2.3.1 Construal of experience at the rank of the clause

Both CG and Hallidayan SFG occupy themselves with the important role of language in **the construal of a situation** (cf. Halliday & Matthiessen, 1999; Halliday & Matthiessen, 2004:29). As Langacker (2008: 4) puts it, “an expression imposes a particular construal, reflecting just one of the countless ways of conceiving and portraying the situation in question”. Specification is an important part of construal (Langacker, 2008:55; Halliday & Matthiessen, 1999: 82-84).

According to Langacker (2008:43), the term **construal** refers to the “ability to conceive and **portray** the same situation in alternate ways” (my emphasis – YVB). Construal is part of the ideational/representational function of language and also involves the conceptual lexicon. Language expresses ‘content’, “gives structure to experience, and helps to determine our way of looking at things” (Halliday, 1970, 143) Halliday and Matthiessen (1999:9) say of the **ideational** metafunction: “it is language as a theory of reality, as a resource for reflecting on the world”. In SFG, the ideational metafunction is regarded as having an **experiential** as well as a **logical** component (Halliday & Matthiessen, 2004:29). The focus here is on the construal of *experience*. The realization of logical meaning will be considered later. In realizing the ideational/representational function of language, the clause is “a mode of reflection, of imposing order on the endless variation and flow of events” (Halliday & Matthiessen, 2004:170). The experiential structure of the clause is summarized as follows:

The experiential structure of the clause is a figure consisting of “[1] a process unfolding through time [2] the participants involved in the process [3] circumstances associated with the process” (Halliday & Matthiessen, 2004:175).

The process, which is realized by a verbal group, is at the core of the clause as a configuration of change. Participants “are directly involved in the process, bringing about its occurrence or being affected by it in some way” (Halliday & Matthiessen, 2004:176). Participants are realized by nominal groups. The process and participants are **central** elements in the clause (Halliday & Matthiessen, 2004:176). Circumstances, which are typically realized by the adverbial group or prepositional phrase, are mostly optional and do not occur in all clauses, and are therefore regarded as peripheral. The core role that the functional-semantic categories of Process and Participant play in the

construal of experience implies that the structures which realize them, namely nominal group and verbal group, are the **key** structures in the linguistic construal of experience. In CG nouns and verbs – and by implication their expansive grammatical units, e.g. nominal (group) – are also afforded central importance in construal (cf. Langacker, 2008:103ff and also Langacker, 1987:53ff). Langacker remarks, “Among the conceptual archetypes helping to structure our mental universe, few if any are more fundamental than physical objects and events, the respective prototypes for the noun and verb categories” (2008:296). Deutscher (2005: 213) maintains that the distinction between things and actions must have been made even before the evolution of language, and that words for things and words for actions simply reflect a pre-existing conceptual distinction.

2.3.1.1 Construal in the clause according to Hallidayan SFG

The reflection on or representation of ‘goings-on’ in the clause is achieved through the grammatical system of transitivity (Halliday & Matthiessen, 2004:170). The various types of process derive from the system of transitivity:

The transitivity system construes the world of experience into a manageable set of PROCESS TYPES. Each process type provides its own model or schema for construing a particular domain of experience as a figure of a particular kind ... (Halliday & Matthiessen, 2004:170).

In SFG, a distinction is made between processes that represent ‘outer’ experience and processes that represent ‘inner experience’ (Halliday & Matthiessen, 2004:170). According to Halliday and Matthiessen (2004:170), the prototype of outer experience is that of actions and events: “things happen, and people or other actors do things or make them happen”. Outer experience is represented by **material process** clauses. Inner experience is partly a reflection on outer experience and partly an “awareness of our states of being” and is represented by **mental process** clauses (Halliday & Matthiessen, 2004:170). Humans also have the ability to relate one fragment of experience to another through identification (‘X is the same as Y’) and classification (‘X is a kind of Y’) (Halliday & Matthiessen, 2004:170). These relations are encoded in **relational process** clauses. Material, mental and relational processes are the primary types of process. If they are arranged on a wheel, three more types of process can be distinguished on the boundaries between them. According to Halliday and Matthiessen, 2004:171, they are:

behavioural processes (between material and mental processes), which “represent the outer manifestations of inner workings, the acting out of processes of consciousness and physiological states”; **verbal processes** (between mental and relational processes), i.e. processes of saying and meaning which construct symbolic relationships; and **existential processes** (between relational and material processes).

The constituents of the experiential structure of the clause in SFG were introduced earlier as “[1] a process unfolding through time [2] the participants involved in the process [3] circumstances associated with the process” (Halliday & Matthiessen, 2004:175). Semantically a process will involve at least one participant, the inherent participant, although the participant need not be expressed linguistically. For example, *Stop!* encodes the Process of stopping, but it is implied that someone must stop. Some types of processes require more than one participant, or a Circumstance. For example: *You can put it there.* – The Process *put* requires a second Participant (*it*) as well as a Circumstance (*there*). Relational processes have two inherent participants. Some meteorological processes are not analyzed further and do not have an inherent participant, e.g. *it’s raining* (Halliday & Matthiessen, 2004:175). The table below provides an overview of the different types of process and the participants involved in these processes.

Table 2.2 Process types, their meanings and characteristic participants

PROCESS TYPE	category meaning	participants, directly involved	participants, obliquely involved
material: action event	'doing' 'doing' 'happening'	Actor, Goal	Recipient, Client; Scope; Initiator; Attribute
behavioural	'behaving'	Behaver	Behaviour
mental: perception cognition desideration emotion	'sensing' 'seeing' 'thinking' 'wanting' 'feeling'	Senser, Phenomenon	
verbal	'saying'	Sayer, Target	Receiver; Verbiage
relational: attribution identification	'being' 'attributing' 'identifying'	Carrier, Attribute Identified, Identifier; Token, Value	Attributor, Beneficiary Assigner
existential	'existing'	Existent	

(reproduced from Halliday & Matthiessen, 2004:260)

The examples in block-diagrams in the discussion of participant types below are provided by Halliday and Matthiessen (2004:180ff) to elucidate the various types of process and the participants involved in them.

In **material** processes, the **Actor** “brings about the unfolding of the process through time, leading to an outcome that is different from the initial phase of the unfolding” (Halliday & Matthiessen, 2004:180). In **intransitive** clauses the outcome of the process is confined to the Actor, and not extended to another entity. An intransitive clause represents a **happening**. In **transitive** clauses the process is extended to another participant who is directly affected by the process, namely the **Goal** – in the sense of ‘goal of impact’. Transitive clauses may be in the active (operative) or passive (receptive) voice. In the operative voice, Actor maps onto the Subject (of the interpersonal structure) and Goal maps onto the interpersonal Complement. In the receptive voice, the Actor, if it is expressed, maps onto the Adjunct of the interpersonal structure, but it may also be left unexpressed. Both **Recipient** and **Client** are participants that benefit from the process, but “[t]he Recipient is one that goods are given to; the Client is one that services are done for” (Halliday & Matthiessen, 2004:191) (my underlining – YVB). **Scope** is treated as a participant grammatically, although it does not actually “participate” in the process: “Rather, it either construes the domain over which the process takes place (...) or (ii) construes the process itself, either in general or specific terms” (Halliday & Matthiessen, 2004:192). The function of Scope is found in processes which are transformative and where the transformation is an elaboration of the Actor. Similarly, without actually “participating”, **Attribute** and **Role** can elaborate the Goal. The following examples provided by Halliday and Matthiessen (2004: 180-195) illustrate the participants associated with material processes:

Figure 2.5 Material Processes and their Participants

<i>the lion</i>	<i>sprang</i>
Actor	Material Process (intransitive)

<i>the lion</i>	<i>caught</i>	<i>the tourist</i>
Actor	Material Process (transitive: active)	Goal
Subject	(Finite) Predicator	Complement

<i>the tourist</i>	<i>was</i>	<i>caught</i>	<i>by the lion</i>
Goal	Material Process: (transitive: passive)		Actor
Subject	Finite	Predicator	Adjunct

<i>the tourist</i>	<i>was caught</i>
Goal	Material Process (passive)

<i>They</i>	<i>built</i>	<i>me</i>	<i>a house</i>
Actor	Material Process	Client	Goal

<i>They</i>	<i>gave</i>	<i>him</i>	<i>a house</i>
Actor	Material Process	Recipient	Goal

<i>They</i>	<i>played</i>	<i>the piano</i>
Actor	Material Process	Scope (entity)

<i>They</i>	<i>played</i>	<i>a game of tennis</i>
Actor	Material Process	Scope (process)

<i>She</i>	<i>crossed</i>	<i>the room</i>
Actor	Material Process	Scope (entity)

<i>They</i>	<i>washed</i>	<i>the plates</i>	<i>clean</i>
Actor	Material Process	Goal	Attribute (resultative)

<i>They</i>	<i>cut</i>	<i>it</i>	<i>into cubes</i>
Actor	Material Process	Goal	Role (product)

In the **mental** clause, the inherent participant is the **Senser**, “the one that ‘senses’ – feels thinks, wants or perceives” (Halliday & Matthiessen, 2004:201). The Senser is therefore a ‘conscious being’ or an entity that is endowed with **consciousness** in the context of the discourse. The **Phenomenon** may be a **thing**, an **act** or a **fact**. If the Phenomenon is an act, it is realised by a **macrophenomenal** clause, and if it is a fact it is realised by a **metaphenomenal** clause, as the following examples provided by Halliday and Matthiessen (2004:204) show. Note that it is possible to omit overt expression of the Senser in the passive voice.

Figure 2.6 Mental Processes

<i>Mary</i>	<i>liked</i>	<i>the gift</i>
Senser	Mental Process	Phenomenon (thing)

<i>The gift</i>	<i>pleased</i>	<i>Mary</i>
Phenomenon (thing)	Mental Process	Senser

<i>He</i>	<i>saw</i>	[[the sand dredger heading for the cruiser]] ¹¹
Senser	Mental Process	Phenomenon (act)

[[That this was not the ideal solution]]	was recognised
Phenomenon (fact)	Mental Process

The participants of **relational** processes differ from the participants in other processes, in that there are two inherent participants (not just one), because the clause sets up a

¹¹ The double square brackets indicate an embedded clause.

relationship between the two participants. The relational clause is a configuration of **'being' or 'having'**. Halliday and Matthiessen (2004:214) point out the importance of the Participants in relational clauses, stating that “the experiential ‘weight’ is construed in the two participants, and the process is merely a highly generalized link between these two participants”. Due to the non-prominent role of the process, the (copula) verb in relational clauses is typically non-salient – in English it might be contracted, as in *she’s happy*, and in some languages it is completely absent so that the configuration is simply ‘Be-er1’ + ‘Be-er2’ (Halliday & Matthiessen, 2004:214). There are two modes of relational, namely **attributive** (which construes class membership) and **identifying** (which construes identity), and there are three types of relation, namely **intensive** (‘being’), **possessive** (‘having’) and **circumstantial** (‘being’ in terms of location in time or space). From these two modes and three types of relation, six categories of relational clause can be distinguished (Halliday & Matthiessen, 2004:216):

Table 2.3 The categories of relational clause

	(i) attributive 'a is an attribute of x'	(ii) identifying 'a is the identity of x'
(1) intensive 'x is a'	<i>Sarah is wise</i>	<i>Sarah is the leader; the leader is Sarah</i>
(2) possessive 'x has a'	<i>Peter has a piano</i>	<i>the piano is Peter's; Peter's is the piano</i>
(3) circumstantial 'x is at a'	<i>the fair is on a Tuesday</i>	<i>tomorrow is the 10th; the 10th is tomorrow</i>

The distinction between attributive and identifying relations is an important one with regard to specification. In intensive attributive processes, the participants are **Carrier** and **Attribute**. In intensive identifying processes the participants are **Identified** and **Identifier**. The Attribute is realized by a nominal group which is non-specific or indefinite and which is headed by an adjective or a common noun which construes class membership. The Attribute is not typically realized by a proper noun or a pronoun, because these do not typically construe class membership. The identifier is realized by nominal groups which are definite. In intensive attributive relations, the participants are not interchangeable, whereas they are reversible in identifying intensive relations.

If one considers the following statement by Halliday and Matthiessen (2004:228), one can gather that identity is actually the most extreme case of class membership, where the class is so specific that its membership (always within a certain discourse context) is narrowed down to one:

One way of looking at the ‘identifying’ clause would be to say that here we are narrowing down the class in question to **a class of one**. If we say *Alice is the clever one*, or *Alice is the cleverest*, this does serve to identify Alice, because we have specified that there is only one member in the class, **a single instance**. It does not say, of course, that there are no other clever people in the world; only that there are no others within a previously specified population, e.g. *There are three children in the family* (my emphasis – YVB).

The **Behaver** is the participant inherent in **behavioural** processes, but need not be expressed, for instance in imperative clauses. Where the process is worded in very general terms, the **Behaviour** expresses more specifically what the Behaver did. The participants of behavioural clauses are shown in the block-diagram below, which is an analysis of examples provided by Halliday and Matthiessen (2004:250-251).

Figure 2.7 Behavioural Processes

<i>I</i>	<i>was dreaming</i>		
<i>He</i>	<i>gave</i>	<i>a great yawn</i>	
<i>he</i>	<i>could have dreamt</i>		<i>about it</i>
	<i>breathe</i>		<i>deeply</i>
Behaver	Process	Behaviour	Circumstance of Matter or Manner

Verbal processes construe quoting or reporting (Halliday & Matthiessen, 2004:253). The **Sayer** is the inherent participant of verbal processes. The **Verbiage** is typically realized by a representation of direct or reported speech. In verbal processes there may also be a **Receiver**, the one to whom the saying is directed, which is realized by a nominal group denoting a conscious being. The following examples from Halliday and Matthiessen (2004:253, 255) illustrate the participants of verbal processes:

Figure 2.8 Verbal Processes

<i>John</i>	<i>said</i>	<i>'I'm hungry'</i>
Sayer	Process (Quoting)	Verbiage (Quoted)
<i>John</i>	<i>said</i>	<i>he was hungry</i>
Sayer	Process (Reporting)	Verbiage (Reported)
<i>Describe</i>	<i>to the court</i>	<i>the scene of the accident</i>
Process	Receiver	Verbiage

The **Existent**, the entity or event which exists, is the participant inherent in **existential** clauses which convey that something exists or happens (Halliday & Matthiessen, 2004:

256). The following block-diagrams are provided by Halliday and Matthiessen (2004:259) to show the participants of existential processes:

Figure 2.9 Existential Processes

<i>there</i>	<i>was</i>	<i>a storm</i>
	Process	Existent: event

there	is	a man	at the door
	Process	Existent: entity	circumstance

Regardless of the type of process, the constituents of the experiential structure of the clause construe the situation in terms of **time** and **space**. Participants exist in three-dimensional space; processes take place over time. Circumstance elements contribute to locating the process or one or more of the participants in space or time. The facet of transience is associated with the verbal group as realization of the process, whereas the facet of permanence is associated with the nominal group as realization of participant, as Halliday and Matthiessen (2004:177, 178) explain:

Transience is the experience of **unfolding** through time; it is construed by a verbal group serving as the process. Permanence is the experience **lasting** through time and being located in (concrete or abstract) space; it is construed by nominal groups serving as participants. Thus participants are relatively stable through time; and an **instance** of a participant can take part in many processes. ... Change is thus construed as involving both transience and permanence, and the phenomena of experience are construed either as **transient processes** or as **permanent participants** [my emphasis – YVB].

The ideational metafunction has two components: the experiential and the logical (Halliday & Matthiessen, 2004:29). In the clause the ideational function of the nominal group is to realize Participant, and that of the verbal group is to realize Process. The nominal group and the verbal group are structured internally to fulfill this function. Figures 2.10 and 2.11 below show the functional elements in the experiential structure of the nominal group and verbal group respectively. The logical component refers to the ways in which groups and clauses can be expanded through elaboration, extension and enhancement (Halliday & Matthiessen, 2004:373ff). With regard to the nominal group, it refers to elaboration through modification and submodification (Halliday & Matthiessen, 2004:489).

Figure 2.10 Experiential structure of the nominal group

Example	<i>those</i>	<i>two</i>	<i>splendid</i>	<i>old</i>	<i>electric</i>	<i>trains</i>
Experiential structure	Deictic	Numerative	Epithet		Classifier	Thing
Word class syntagm	determiner	numeral	adjective	adjective	adjective	noun

Halliday and Matthiessen (2004:329)

The experiential structure of the nominal group arises from its specifying function. The nominal group has the function of **specifying a class of things** (e.g. job) and then **specifying some category of membership** within that class (e.g. *domestic jobs, any job*) (Halliday & Matthiessen, 2004:312). The noun realizing the element of Thing expresses the class. All of the elements preceding the Thing in a nominal group, i.e. the Deictic, the Numerative, the Epithet and the Classifier, serve to establish categorization within the class.

According to Halliday and Matthiessen (2004:335) “[t]he experiential structure of the finite verbal group is **Finite** (standing for ‘Finite operator’) plus **Event**, with optional elements **Auxiliary** (one or more) and **Polarity**”. They (Halliday & Matthiessen, 2004:336) provide the following representation of the experiential structure of the verbal group:

Figure 2.11 The experiential structure of the verbal group

<i>has</i>	<i>not</i>	<i>been</i>	<i>working</i>
Finite	Polarity	Auxiliary	Event

There is a correlation between the Finite element of the verbal group and the Deictic element of the nominal group. A similar correspondence is to be found between the Thing of the nominal group and the Event of the verbal group.

Halliday and Matthiessen’s (2004) account of the experiential structure of the clause differs from traditional accounts of clause structure found in standard reference grammars of English which are more syntactic and less semantic. Traditional syntactic clause roles are Subject, Verb, Complement and Adjunct (or Adverbial). The Complement of a transitive verb is an Object, direct and/or indirect. The Object itself may also take a Complement. The Complement of intransitive copula verbs is a Subject

Complement (Subject Predicative). Nominal groups can fulfill the role of Subject, Object or Complement. Adjective phrases (i.e. nominal groups headed by adjectives) can also fulfill most of the roles which noun phrases can, as can nominal clauses. In Hallidayan SFG, the analysis of the clause as “message” most closely corresponds to analysis of the clause as found in the reference grammars of English. A view of the clause as representation of experience, i.e. one in which a Process, Participants and Circumstances are distinguished, helps us understand the systemic choices behind the wordings of verbal and nominal groups in the analysis of actual examples of language usage since it allows for a semantic sub-classification of verbs based on Process types and of nominal groups based on Participant roles.

2.3.1.2 Semantic roles and situation types in clauses

Halliday and Matthiessen (2004:176) distinguish various types of Participant based on the type of Process they are involved in and their role in it, asserting that “[t]he nature of participants will thus vary according to the type of process”. Givon (2001:105), on the other hand, characterizes verbs semantically in terms of the participants in the event or state coded by the verb, pointing out that each verb (type) has a characteristic cluster of obligatory participants. For Givon (2001:105), participants fulfill a semantic role. The simple clause can semantically encode a state, event or action – each of which can be semantically sub-classified. Participant roles are semantic roles. In the simple (main, declarative, affirmative, active) clause, there is virtually a one-to-one relationship (i.e. an isomorphism) between grammatical roles (e.g. Subject, Direct Object, Indirect Object, Predicate/Verb, Adverb[ial]) and semantic roles (e.g. Agent, Patient) (Givon, 2001:105). In order to illustrate this isomorphism between the syntax and semantics of simple clauses, the examples provided by Givon (2001:107) are given in block diagrams below:

Figure 2.12 Semantic and syntactic roles

	<i>Mary</i>	<i>kicked</i>	<i>John</i>
SEMANTIC ANALYSIS	Agent		Patient
SYNTACTIC ANALYSIS	Subject		Direct Object

	<i>Mary</i>	<i>is</i>	<i>tall</i>
SEMANTIC ANALYSIS	Patient		
SYNTACTIC ANALYSIS	Subject		

	<i>John</i>	<i>scared</i>	<i>Mary</i>
SEMANTIC ANALYSIS	Agent		Dative
SYNTACTIC ANALYSIS	Subject		Direct Object

	<i>John</i>	<i>talked</i>	<i>to Mary</i>
SEMANTIC ANALYSIS	Agent		Dative
SYNTACTIC ANALYSIS	Subject		Indirect Object

	<i>She</i>	<i>chopped</i>	<i>firewood</i>	<i>with an axe</i>
SEMANTIC ANALYSIS	Agent		Patient	Instrument
SYNTACTIC ANALYSIS	Subject		Direct Object	Adverbial/Adjunct

	<i>He</i>	<i>fixed</i>	<i>the roof</i>	<i>for his mother</i>
SEMANTIC ANALYSIS	Agent		Patient	Benefactive
SYNTACTIC ANALYSIS	Subject		Direct Object	Adverbial/Adjunct

	<i>She</i>	<i>lives</i>	<i>in Philadelphia</i>
SEMANTIC ANALYSIS	Patient		Locative
SYNTACTIC ANALYSIS	Subject		Adverbial/Adjunct

	<i>He</i>	<i>went</i>	<i>to the store</i>
SEMANTIC ANALYSIS			Locative
SYNTACTIC ANALYSIS	Subject		Adverbial/Adjunct

	<i>She</i>	<i>worked</i>	<i>with her father</i>
SEMANTIC ANALYSIS			Associative
SYNTACTIC ANALYSIS	Subject		Adverbial/Adjunct

	<i>He</i>	<i>left</i>	<i>in a hurry</i>
SEMANTIC ANALYSIS			Manner
SYNTACTIC ANALYSIS	Subject		Adverbial/Adjunct

The semantic terms state, event and action distinguished by Givon (2001:106) are reminiscent of Vendler's (1957) classification of verbs based on time schemata. Vendler (1957:144) first distinguishes between verbs that readily admit "continuous tenses" and those that do not, making use of the verbs *run* and *know* with question-prompts to illustrate the distinction:

What are you doing?
 I am running.
 *I am knowing.

Do you know ...?
 Yes I do.

Do you run?
 Yes I do. [Note though that this is not equivalent in meaning to "Yes, I am running".]

According to Vendler (1957:144-145), verbs that admit the “continuous”, e.g. *running*, are “processes going on in time, i.e. ... they consist of successive phases following one another in time”, whereas verbs like knowing are “not processes going on in time”. Within the category of processes going on in time, a distinction is made between activity terms and accomplishment terms (Vendler, 1957:145-146). Accomplishments have a terminal point and must be finished, whereas activities “go on in time in a homogeneous way; any part of the process is of the same nature as the whole” (Vendler, 1957:146). The difference is illustrated as follows:

- (8) He is running a mile. (Accomplishment)
- (9) He is running. (Activity)

The verbs that do not denote processes going on in time may still “be predicated of a subject for a given time” (Vendler, 1957:146). Here a distinction is made between verbs predicated for a moment and verbs predicated for a period in time, the first being achievement terms and the latter being state terms (Vendler, 1957:147). This distinction is illustrated as follows:

- (10) He reaches the hilltop. (Achievement)
- (11) He knows something. (State)

Huddleston (2002:118ff) also refers to states, achievements, activities and accomplishments, but the taxonomy into which these fit is slightly different than the one originally conceived by Vendler. With regard to the situation type expressed in a clause, the first distinction Huddleston makes is between states (static) and occurrences (dynamic). Under dynamic situations, processes (durative) and achievements (punctual) are distinguished. Under processes, activities (atelic) and accomplishments (telic) are distinguished. The situation type classification by Quirk et al. (1985:177, 200ff) is more refined, but also starts with the stative/dynamic distinction. Under stative situation types a clear distinction is made between qualities and states, illustrated respectively by the following two examples:

- (12) Mary is Canadian, and Mary has blue eyes. [Stative:quality]
- (13) Mary is tired and Mary has a bad cold. [Stative:state]

Under dynamic situation types durative and punctual situations are distinguished. Both may be either conclusive or non-conclusive. A further distinction which applies to all the aforementioned dynamic situation types is that of agentive/non-agentive. These three binary oppositions give rise to eight dynamic situation types, namely goings-on, activities, processes, accomplishments, momentary events, momentary acts, transitional events and transitional acts.

Langacker (2008:147) makes a **conceptual** distinction between the two major subclasses of verbs, **perfective** and **imperfective**, based on **temporal bounding**. Perfective verbs are conceived as being bounded in time, whereas imperfective verbs are not bounded. Like Givon (2001), Langacker (2008:356) also distinguishes semantic roles and grammatical roles for clause participants. The clause types and associated roles distinguished by Langacker (2008) need to be understood against the background of the definition of “noun” and “verb” in Cognitive Linguistics, and will therefore be discussed after an account of the cognitive definition of nouns and verbs in the next subsection.

In this study the focus will fall on nominal groups. Where the clausal role of the nominal group potentially influences its internal structure, reference will be made to the type of process in which the nominal group is a “participant”. This will be relevant in Chapter 4 in respect of nominal groups in which the noun *people* realizes the Thing element and in Chapter 5 in respect of the choice of determiner as specifying element.

2.3.1.3 Types and instances of things and processes

Like systemic functional linguistics, cognitive linguistics also distinguishes ‘process’ and ‘participant’ as conceptual categories. However, although there is much conceptual overlap, these are not the exact terminological equivalents of ‘Process’ and ‘Participant’ in SFG. The similarities in the terminology arise from the fact that in both SFG and CG the structural components of the clause are derived from ‘**experience**’. The differences in the senses in which the terms ‘participant’ and ‘process’ are used in SFG and CG stem from the fact that their use in CG is based explicitly on conceptual archetypes.

Compare the following remarks by Halliday and Matthiessen (2004) and Langacker (2008):

The concepts of process, participant and circumstance are semantic categories which explain in the most general way how phenomena of our experience of the world are construed as linguistic structures (Halliday and Matthiessen, 2004:178)

Despite its diversity and complexity, clause structure is readily seen as being grounded in basic human experience. It is best described and understood with reference to certain archetypal conceptions representing fundamental aspects of such experience. Conceptual archetypes function as the prototypes for clausal elements and are a major factor in determining their structural arrangement. One archetype is the organization of a scene into a global **setting** and any number of smaller, more mobile **participants** (Langacker, 2008:355).

The idea that it should be possible to define nouns and verbs semantically (Langacker, 2008:103, 1987:54) arises from CG's position that "grammar is meaningful" (Langacker, 2008:3). According to Langacker (2008:5), "CG's most fundamental claim is that grammar is **symbolic** in nature". Here 'symbol' is defined as "the pairing between a semantic structure and a phonological structure, such that one is able to evoke the other" (Langacker, 2008:5). In terms of this definition of symbol, it is clear why lexical items are regarded as symbolic structures, but in CG, it is argued that grammar too can be described in terms of symbolic structure, because it is through grammar that symbolic structures (i.e. words) combine. As Langacker (2008:5) puts it:

Grammar, of course, is concerned with how such elements [lexical items as symbols – YVB] combine to form complex expressions. The basic tenet of CG is that nothing beyond symbolic structures need be invoked for the proper characterization of complex expressions and the patterns they instantiate. More specifically: lexicon and grammar form a gradation consisting solely in assemblies of symbolic structures. An immediate consequence of this position is that all constructs validly posited for grammatical description (e.g. notions like 'noun', 'subject', or 'past participle') must in some way be meaningful. Langacker (1987:53, 2008:103) claims that it is possible to provide notional (i.e. semantic) definitions of the categories of noun and verb, which he regards as the most fundamental grammatical categories. Notional definitions of prototypical nouns (i.e. nouns referring to concrete objects) and prototypical verbs (verbs referring to dynamic actions) are quite plausible, whereas a notional definition which is valid for all nouns or all verbs, i.e. a schematic definition of noun or verb, is more difficult to conceive, and yet, this is what Langacker (2008:103) proposes, starting from the category prototype. The

noun and verb archetypes functioning as category prototypes are represented and contrasted in the table below:

Table 2.4 Contrast between noun and verb archetypes

Noun archetype			Verb archetype	
1	A physical object is composed of material substance.	materiality	1	An energetic interaction is not itself material, consisting instead of <u>change</u> and the transfer of energy.
2	We think of an object as residing primarily in <u>space</u> , where it is bounded and has its own location.	primary location	2	Thus an event resides primarily in <u>time</u> ; it is temporally bounded and has its own temporal location.
3	In <u>time</u> , on the other hand, an <u>object</u> may persist indefinitely, and it is not thought of as having any particular location in this domain.	non-primary dimension	3	By contrast, an <u>event</u> 's location in <u>space</u> is more diffuse and also derivative, as it depends on the locations of its participants.
4	An object is conceptually autonomous , in the sense that we can conceptualize it independently of its <u>participation</u> in any event.	conceptual autonomy	4	This is so because an event is conceptually dependent ; it cannot be conceptualized without conceptualizing the <u>participants</u> who interact to constitute it.

(Langacker, 2008:104) (my underlining – YVB)

The table above not only characterizes noun and verb archetypes, but also serves to contrast nouns and verbs in terms of materiality, location in time and space and conceptual autonomy. The fact that prototypical nouns and verbs can be contrasted directly in terms of these four aspects indicates that noun and verb are conceptual counterparts. The contrasts in the table above are also found in SFG accounts of the noun-verb parallelism. In SFG the clause is regarded as a “figure of change” (compare line 1 in the table above under verb archetype). Halliday and Matthiessen (2004:177) point out that “[t]he units that realize the process, participant, and circumstance elements of the clause make distinct contributions to the modelling of the quantum of change”. According to them, the nominal group and the verbal group, as realizations of the central elements of the clause, “construe complementary facets of the change”, namely **transience** and **permanence** (Halliday & Matthiessen, 2004:177). This ties in with the third property in the table above. In terms of the fourth property in the table above, conceptual autonomy, there is a difference between CG and SFG. Halliday and Matthiessen (2004:176) afford primacy to the process in the determination of the clause type and the nature of participants, whereas Langacker (2008:104) asserts that event conceptualization depends on participant conceptualization.

Schematic notional definitions of noun and verb are generalized and abstracted from the noun and verb archetypes represented in the table above. Schematization refers to “the process of extracting the commonality inherent in multiple experiences to arrive at a conception representing a higher level of abstraction” (Langacker, 2008:17). The cognitive abilities of grouping and reification are important in arriving at a schematic definition of nouns, whereas the apprehension of relationships and their tracking through time are necessary for defining verbs schematically.

Entities can be grouped together based on similarity, contiguity or some other perception of interconnectedness. Once grouped together, the group can function as a unit in a higher-level group. This is reification, the “capacity to manipulate a group as a unitary entity for higher-order cognitive purposes” (Langacker, 2008:105). In CG a **thing** is defined as “any product of grouping and reification” (Langacker, 2008:105). Langacker (2008:105) argues that because grouping and reification are general cognitive phenomena which, importantly, are not limited to space or perception, “things can emerge from constitutive entities in any domain or at any level or conceptual organization”. He makes use of the nouns *recipe*, *committee* and *moment* to illustrate how grouping and reification produce a thing: The steps in producing a dish are grouped together on the basis of being conceived as occurring in a sequence and are then regarded as a single unified procedure - i.e. reified – as ‘recipe’ with a goal of creating a dish. *Committee* refers to a group of members, but is reified by the conception of the members as working together as one with a common goal. *Moment* is a thing which is different in nature from the previous two examples; as Langacker (2008:106) explains,

As a continuous span of time, its constitutive entities (points in time) are grouped on the basis of temporal contiguity. The group is reified through its conception as being a single unit of temporal experience, quite brief in overall duration.

The schematic characterization of a noun is that it profiles a thing (Langacker, 2008:106). Keeping in mind the definition of thing as “any product of grouping and reification”, the proposal that “**a noun profiles a thing**” is applicable to “the many nouns whose referents clearly consist of multiple, individually recognizable entities”, such as *group*, *constellation*, *list*, *forest*, *family*, *colony*, *lexicon*, etc. (Langacker, 2008:106).

It would seem problematic to apply the notion of thing as "a set of **interconnected entities** which function as a **single entity** at a higher level of conceptual organization" (Langacker, 2008:107, my emphasis – YVB) to prototypical nouns, such as *rock* or *board*, because it is difficult to think of these as consisting of interconnected entities. Langacker (2008:107) points out, however, that an **entity** need not be discrete, cognitively salient, or individually recognized. The "entities" that make up a *rock* or a *board* are not processed as discrete or individuated, and herein lies their prototypicality as nouns:

Indeed, the absence of individuation is precisely what makes physical objects prototypical. They represent the special circumstance where grouping and reification are so **automatic** that constitutive entities are never consciously accessible. It is only when these operations are extended to other cases, where they are nonautomatic, if not atypical, that we can be cognizant of their effect.

With physical objects it is thus the **product** of grouping and reification, the conception of a unitary entity, that predominates at the conscious level.

Langacker (2008:107) posits that the abstract notion of thing also accommodates mass nouns and abstract nouns. In order to fully grasp the schematic definition of noun, namely "a noun profiles a thing", it is necessary to keep in mind the definition of "thing" (as expounded above). The **profile** is "what the expression is conceived as **designating** or referring to ... the entity that an expression designates, i.e. its **referent**" (Langacker, 2008:66, 67 my emphasis – YVB). A noun profiles a thing, because a noun **designates** a thing.

In CG, it is accepted that a verb profiles a process (Langacker, 2008:112), i.e. a verb designates a process. The term **process** refers to "a complex relationship that develops through conceived time and is scanned sequentially along this axis" (Langacker, 2008:112). In CG, for something to be a process, it needs to be cognitively processed through **sequential** scanning. Sequential scanning refers to "mentally tracking an event as it unfolds through time" (Langacker, 2008:111). If the event is not construed as unfolding through time, but is viewed as "a single gestalt" the scanning is summative, not sequential (Langacker, 2008:111). The same conceptual content can be construed either as a process or non-processually depending on the type of scanning involved. For a verb form to be regarded as a process, and thus as a true verb in CG terms, it needs to be finite, i.e. involve sequential scanning (through time). Infinitives and

participles which represent the same semantic content in a summary fashion are regarded as a separate grammatical category that profiles a non-processual relationship (Langacker, 2008:112, 19-120).

In CG, then, a noun is an **expression** that profiles a thing, and a verb is an **expression** that profiles a process. It is important to note that in CG “noun” and “verb” classify expressions, not words only. The definition of a noun as “any expression that profiles a thing” includes “both lexical nouns and nominal expressions of any size” (Langacker, 2008:310). Likewise the term “verb”, defined as “any expression that profiles a process” includes “both lexical verbs and complex verbal constructions” such as clauses (Langacker, 2008:354). Langacker (2008:264) likens lexical nouns to nominals (what SFG would call nominal groups), and lexical verbs to clauses based on the similarities of their respective profiles:

Nominals and finite clauses resemble nouns and verbs in being universal and grammatically fundamental. Like a noun, a nominal profiles a thing; like a verb, a clause profiles a process. Because they profile things and processes, nominals and finite clauses are themselves nouns and verbs as broadly defined in CG.

The difference between lexical nouns and nominals and between lexical verbs and finite clauses lies in the fact that lexical nouns and lexical verbs are **designators of type** whereas nominals and finite clauses are **grounded instances of the type** (Langacker, 2008:134, 263-264).

They [nominals and finite clauses] are distinguished from nouns and verbs in the usual narrow sense - ‘lexical’ nouns and verbs - by further semantic properties pertaining to their cognitive and discourse function. The primary function of lexemes is **classificatory**. As fixed expressions, they provide an established scheme for apprehending the world in terms of culturally sanctioned categories of proven relevance and utility. By contrast, the primary function of a nominal or finite clause is **referential**. It directs attention to a particular thing or process accorded a certain epistemic status in relation to the ground. Through grounding, its characterization of the profiled entity serves to distinguish it from other members of its category and identify it for immediate discourse purposes. (Langacker, 2008:264)

The **classificatory** function of lexical nouns and lexical verbs and the deictic function of the Deictic (determiner) and the Finite form the basis of the noun-verb parallelism recognized in SFG. The “identification of the item in terms of the **here-&-now**” (Halliday

& Matthiessen, 2004:322) by the **Deictic** element is what is referred to as **grounding** in CG (Langacker, 2008:259), and what Langacker (2008:264) means by the **referential** function of nominals in the quote above.

In Halliday and Matthiessen’s (2004:329) example, *those two splendid old electric trains*, the element with the **least specifying potential** is *train*, which specifies **a class of thing**, and the element with the **most specifying potential** is the determiner *those* which **anchors** (or grounds) the nominal group in the context of situation. Note, though, that *train*, the Thing, is quite rich in lexical content, whereas, *those*, the Deictic element, has very little lexical content and rather serves a grammatical function. It is thus possible to map the experiential structure of the nominal group, excluding postmodifiers (Qualifiers), onto the lexicogrammatical cline. Items with the greatest specifying potential lie on the grammatical side of the cline, whereas items with the least specifying potential lie on the lexical side of the cline. Langacker (2008:263) also places the grounding elements of nominals and clauses “toward the grammatical pole of the lexicon/grammar continuum” and lexical nouns and verbs on the lexicon pole.

Figure 2.13 The lexico-grammatical continuum

	<i>those</i>	<i>trains</i>
Experiential structure	Deictic	Thing
Word class (syntagm)	determiner	noun
Lexicogrammatical cline	GRAMMAR ----- LEXIS	
Cognitive functions	grounding	designation

Deicticity or grounding ties in with permanence and transience. The internal organization of the nominal group and the verbal group has evolved to express permanence and transience, the two facets of change, respectively and to **specify in terms of location in space and time** respectively: “while nominal groups have evolved the system of DETERMINATION¹² for locating **referents in a referential space**, verbal groups have evolved the system of TENSE for locating **a unique occurrence of a process in time**” (Halliday & Matthiessen, 2004:178) (my emphasis – YVB). Specifying potential, then, as it is meant in Hallidayan SFG, is not so much about lexical content,

¹² In SFG, the names of systems are conventionally written in upper case.

but rather about location in space and time relevant to the speech event. A word denoting a thing, such as *trains*, has less potential to locate a referent in space than a word such as *those*.

The system of DETERMINATION is associated with the Deictic element in the experiential structure of the nominal group which orientates the (nominal) group in terms of the speaker by reference to person (by possessive determiners) or proximity (by demonstrative determiners) (Halliday & Matthiessen, 2004: 313, 336). The system of TENSE is associated with the Finite element in the verbal group which relates the process to the speaker-now by tense or modality (Halliday & Matthiessen, 2004: 336). Halliday and Matthiessen (2004:336) point out that both the verbal group and the nominal group “begin with the element that ‘fixes’ the group in relation to the speech exchange” – the **Deictic** of the nominal group and the **Finite** of the verbal group. This “fixing” of the expression in relation to the speech exchange is referred to in CG as grounding. The ground is “the speech event, its participants (speaker and hearer), their interaction, and the immediate circumstances (notably, the time and place of speaking)” (Langacker, 2004:259)

In SFG, Event is regarded as the verbal equivalent of Thing as “both represent the core of the **lexical** meaning” (Halliday and Matthiessen, 2004:336) (my emphasis – YVB). In SFG terms, lexical nouns specify a class of Thing and lexical verbs specify a class of Event. The group-final position of Thing (in the absence of Qualifier) and Event is significant. Halliday and Matthiessen (2004:336) point out that “[b]oth verbal and nominal group ... end with the element that specifies the representation content” and that “[f]inal position is informative; and the newsworthy component of a process or participant is some aspect of its lexical content”.

To summarize: Nouns and verbs as lexemes specify the type of thing and the type of process respectively. The addition of a determiner to a noun creates a nominal group which has the potential to refer to a grounded instance of the type. Similarly, the finite clause refers to an actual (i.e. grounded) instance of the process type. Lexical nouns and verbs specify **types**, determiners and finite markers specify **instances**. Borrowing CG terminology, these two kinds of **specification** will be called **designation** and **grounding** respectively.

As the core elements of the clause as representation of experience, both participants and processes lend themselves (or more accurately their structural realizations as nominal groups and verbal groups) to a corpus-based investigation of specification as a facet of construal. Ideally a comprehensive study of specification should include both, but time and space constraints necessitate a choice between nominals and processes for this study. As mentioned in Chapter 1, there are studies on the construal of processes in BSAfE (Van Rooy: 2006, 2008b, 2009 with regard to temporal grounding; Partridge, 2011 with regard to designation). Another reason for opting for nominal groups is the conceptual autonomy of the noun archetype:

An object is **conceptually autonomous**, in the sense that we can conceptualize it independently of its participation in any event.

...an event is **conceptually dependent**; it cannot be conceptualized without conceptualizing the participants who interact to constitute it (Langacker, 2008:104).

The implication of this is that it should be possible, theoretically, to investigate nominal groups in a corpus without reference to the processes in which they participate. To what extent this will be possible will be discovered in the course of analyzing the corpus data. The main reason, however, for deciding on nominal groups as object of investigation, lies in the potential of nouns to be elaborated into very complex nominal groups. The nominal group has a much greater potential for **complexity**, whereas the verbal group's complexity is restricted. The following remark by Halliday and Matthiessen (2004:336) captures this: "since **things are more highly organized than events**, there are additional lexical elements in the nominal but none in the verbal group". Apart from the additional lexical elements preceding the Thing (e.g. words which might realize the specifying functions of Numerative, Epithet or Classifier), nominal groups may also contain one or more Qualifiers (postmodifiers) typically in the form of prepositional phrases or relative clauses. **Designation** and **grounding** as the core kinds of specification can be found in both the nominal group and the verbal group, but the nominal group also allows for examining the specification of quantities as well as attributive and identifying qualities. These additional kinds of specification will be referred to as **quantification** and **qualification**. Thus, insight into more areas of the grammar of BSAfE is possible through an investigation of nominal groups.

2.3.2 Specifying functions in the nominal group

Langacker (2008:55) provides a broad definition of specificity as “the level of precision and detail at which a situation is characterized” and relates specificity to instantiation and elaboration:

Alternate terms are **granularity** and **resolution**. A highly specific expression describes a situation in fine-grained detail, with high resolution. ... The converse of specificity is **schematicity**. Thus *relative* is schematic with respect to *aunt*, and *rodent* with respect to *large brown rat*. A schematic characterization is **instantiated** by any number of more specific ones, each serving to **elaborate** its coarse-grained specifications. Elaborating *rodent*, ..., are *rat*, *large brown rat*, *vole*, *curious mouse*, *ground squirrel*, *ferocious porcupine with sharp quills*, and so on indefinitely.” (Langacker, 2008:56)

Two of the examples in the quote above suffice to illustrate the paradigmatic and syntagmatic dimensions of specification. The word *rat* is more specific than the word *rodent*, and in turn, the expression *large brown rat* is more specific than *rat*. The relationship between *rodent* and *rat* is one of **paradigmatic specificity**. From a “list” of possible lexical nouns to head the nominal expression a selection is made which can be either more or less specific. The relationship between *rat* and *large brown rat* is one of **syntagmatic specificity**. The lexical noun *rat* is elaborated “horizontally” by additional elements in the nominal group. These two kinds of specification corresponds to the two kinds of specifying function in the nominal group described by Halliday and Matthiessen (2004:312), namely “the function of specifying (i) a class of things, ... and (ii) some category of membership within this class” (Halliday & Matthiessen, 2004:312). Consider the following corpus example:

(14) Africa as a developing continent faces a lot of challenges including HIV/AIDS and, poverty, <illegible> and floods and war. <ICLE-TS-NOUN-0250.1>

In the nominal group *a developing continent*, the noun *continent* specifies a class of thing, while *a developing* serves to further delimit a category of membership within this class.

Lexical specificity is part of the classificatory function of lexemes, and therefore pertains to all lexical word classes. For instance *amber* (an adjective) is a more specific kind of *brown*. Lexical specificity is a result of language's construal of experience into **taxonomies** (Halliday & Matthiessen, 2004:29).

It is clear that language does – as we put it – **construe** human experience. It names things, thus construing them into categories; and then, typically, goes further and construes the categories into taxonomies, often using more names for doing so. So we have *houses* and *cottages* and *garages* and *sheds*, which are all kinds of *building*; *strolling* and *stepping* and *marching* and *pacing*, which are all kinds of *walking*; *in*, *on*, *under*, *around* as relative locations and so on ... (Halliday & Matthiessen, 2004:29)

The semantic relation between the sets of words quoted above is that of **hyponymy** – 'x is a kind of y': for instance, *surgeon* is a kind of *doctor*. Another semantic relation associated with specification is **meronymy**: *She hurt her arm* is more specific than *She hurt herself*. Meronymic relations are formulated as 'x is a part of y': *arm* is a part of the whole body designated by *herself*. In the nominal group the word that **names** or **designates** the Thing, specifies the class of Thing lexically (Halliday & Matthiessen, 2004:40). Langacker (2008:265) remarks that lexemes serve their classificatory function by making **type specifications**:

A noun **designates** a **type of thing**, and a **verb** a **type of process**. Through the nouns and verbs of a language, speakers have ready access to immense inventories of thing and process types that are generally recognized and easily expressed. These allow an initial classification of conceived entities for linguistic purposes. The usual starting point for a nominal or clausal expression is thus a lexical noun or verb, which specifies what type of thing or process is being referred to (Langacker, 2008:265) (my emphasis – YVB).

In the nominal group, the word naming the 'Thing' can be specified further by the elements that precede or follow it. For example, in the nominal group "*the HIV/Aids epidemic in Africa*", "*epidemic*" is further specified by the words "*the*" (determination) and "*HIV/Aids*" (sub-classification) and the phrase "*in Africa*" (qualification). Here the specification is effected by the internal structure of the nominal group as grammatical unit. Specification through nominal **modification** is syntagmatic specification.

The 'Thing' (i.e. "class of things") may also be further specified by units external to the nominal group in which it occurs. At this point it must be remembered that the nominal

group realises the functional element of 'Participant' within the ideational structure of the clause. A single participant role need not necessarily be realised by a single nominal group, but can also be realised by a combination of a nominal group and another grammatical unit which further specifies it. A typical case is that of apposition, where two appositive nominal groups form a nominal group complex which fills a single participant role. Specification by way of a nominal group complex can also be regarded as syntagmatic specification.

As mentioned in Section 2.3.1.1, the experiential structure of the nominal group as conceived in Hallidayan SFG (Halliday & Matthiessen, 2004:312ff; Halliday & Hasan, 1976:42) is explicitly based on the specifying function that nominal groups fulfil and the specifying functions of the elements within the nominal group. A block-diagram of the elements of the experiential structure of the nominal group (given in Figure 2.10) is repeated here for ease of reference, this time using an example from Halliday and Hasan (1976:40).

Figure 2.14 Elements in the experiential structure of the nominal group

	the	two	high	stone	walls	along the roadside
Experiential structure	Deictic	Numerative	Epithet	Classifier	Thing	Qualifier
Syntagm	determiner	numeral	adjective	noun	noun	[prepositional group]

Halliday and Hasan (1976:42) summarize the ways in which each of the functional elements in the nominal group specifies:

... the general meaning of the functions Deictic, Numerative, Epithet, Classifier and Qualifier is that of SPECIFICATION. The Deictic specifies by **identity**, ... including identity based on **reference** ...; the Numerative by **quantity** or ordination ...; the Epithet by reference to a **property**...; the Classifier by reference to a **subclass** ...; and the Qualifier by some **characterizing** relation or process (my emphasis - YVB).

Halliday and Matthiessen's (2004:322) description of the progression "from the kind of element that has the greatest specifying potential to that which has the least" in the nominal group echoes these different bases of specification:

Proceeding from left to right, we begin with the immediate context, the **identification** of the item in terms of the here-&-now... From there we go on to **quantitative** features: place in order, and number. Next come the **qualitative** features. ... Lastly comes **class membership**; (my emphasis - YVB)

The kinds of specification can be mapped onto the elements of the experiential structure of the nominal group as follows:

Figure 2.15 Specification in the nominal group

Specification in terms of:	Identity (reference)	Quantity	Quality (feature/property)	Sub-class	Class	Quality (characterization by way of relationship)
Experiential structure	Deictic	Numerative	Epithet	Classifier	Thing	Qualifier
Syntagm	determiner	numeral	adjective	noun	noun	[prepositional group]

The SFG-based diagrams in this section represent the nominal group as having certain functional slots which may or may not be filled, for instance, in the nominal group *the HIV/Aids epidemic*, the Deictic (*the*), Classifier (*HIV/Aids*) and Thing (*epidemic*) slots are filled, whereas the Numerative, Epithet and Qualifier slots remain empty. Such a slot-and-filler representation runs the risk of going too far with the iconicity of grammar, which is widely recognized in functional approaches (Givon, 2001:34). The relation between function and structural element is not necessarily one-to-one; for instance, both the Thing and the Classifier have a classificatory or denotative function and both the Epithet and the Qualifier provide a characterization (qualification) of the Thing. Based on this, and the concepts of type specification (designation) and grounding from CG introduced in Section 2.3.1.3, the following sub-functions of specification are proposed: (1) Designation (denotation or type specification), (2) Determination (or grounding), (3) Quantification and (4) Qualification.

Designation is common to all lexical word classes, i.e. nouns, verbs, adverbs and adjectives. The kind of specificity that obtains here is **paradigmatic (lexical) specificity**.¹³ For example, the noun *doctor* is more specific than *healthworker* but less specific than *surgeon*; the verb *walk* is more specific than *move*, but less specific than *saunter*, the adjective *bad* is less specific than *inferior* or *wretched*; the adverb *slowly* is less specific than *languidly* or *unhurriedly*. Nominal designation deals with how classes

¹³ See Partridge (2011) for an investigation into the specificity of lexical verbs in the BSAfE.

of things are organised taxonomically, and again the distinction made by the grammar ('particular'/'general') represent the ends of a cline. (Halliday & Matthiessen, 2004:327).

Also relevant here is the use of 'general nouns', for example, *thing, stuff, place, idea, nonsense* as cohesive devices, anaphora (Halliday & Matthiessen, 2004:327), or conceptual shells (Schmid, 2000:3). With designating nouns, the count/mass distinction is an important aspect of construal. The count/mass distinction is based on four conceptual factors, namely bounding, homogeneity, contractibility and replicability (Langacker, 2008: 139). Other distinctions pertaining to nominal designation are the animate/inanimate distinction (Halliday & Matthiessen, 2004:327) and the abstract/concrete distinction (Quirk et al., 1985:247). The designating function of nouns is more clearly observable in common nouns than in proper nouns. The reason for this difference is actually a conceptual one: common nouns are "type-coding", whereas proper nouns are "token-coding" (Givon, 2001:58).

Nouns ('common nouns') do not *refer* to individual entities ('tokens'), but only *connote* classes ('types') of entities. Names ('proper nouns'), on the other hand, refer to individual entities (or specific groups) (Givon, 2001:58).

The function of designation, as defined in this study, corresponds to Givon's (2001:58) notion of "*connote* classes ('types') of entities". The SFG definition of Thing as the element that specifies the "class of things" (Halliday & Matthiessen, 2004: 312) clearly obtains when the Thing is realized by a common noun (i.e. in nominal groups where Thing = Head = common noun), but becomes a little problematic when Thing is realized by proper nouns and personal pronouns which refer to a specific entity or a group of entities.

The definition of designation for purposes of this study corresponds to what Langacker (2008:264) terms the **classificatory** function of lexemes. In Thing=Head nominal groups (e.g. *the government of South Africa*), where the head is realized by a common noun, it is obvious that the common noun, as a lexeme, fulfills a classificatory function, indicating a type of thing (*government*), whereas it is the whole nominal group (*the government of South Africa*) that singles out, or refers, to a grounded instance of the type. Langacker (2008:316) cautions against the classic view that proper names do not have a type description (due to the fact that they refer to instances), pointing out that

“many proper names are conventionally employed for particular types of entities”. For example, typical boys’ names such as *John* are used for entities of the type ‘human male’ and typical girls’ names such as *Sarah* for ‘human female’. Inherent to names is the supposition that within a given context there is just one entity referred to in this manner. Langacker (2008:317) explains the difference between common nouns and proper nouns as follows:

A common noun is one whose type has multiple instances ... On the other hand, a proper noun labels just a single entity, so there is no basis for abstracting a separate type. We can either say that the type/instance distinction is neutralized or, equivalently, that the type has just one instance.

For purposes of comparison with the other types of noun, the latter formulation is preferred in this study, i.e. that proper nouns are types with only one instance in a given context. Whereas the referent of a proper noun is defined experientially, the referent of a pronoun is defined interpersonally (Halliday & Matthiessen, 2004:325). According to Langacker (2008:314) “[o]ne facet of grounding, that of locating the referent vis-à-vis the ground, is thus intrinsic to personal pronouns”. Personal pronouns, like demonstratives, have the potential to single out a unique referent – a particular entity or a particular group of entities, but their meanings are very schematic. Halliday and Matthiessen (2004:327) observe “the most general type of noun is in fact a pronoun, which is the limiting case of anaphoric generalization”.

The attributes of prototypical nouns (time-stable, multi-featured, concrete, spatially compact and countable) (cf. also Givon, 1993:54) are central to distinguishing nouns from other word classes, but the notions of prototype and markedness are also useful in subcategorizing common nouns on semantic grounds. Temporality and concreteness are the prototype attributes that are most relevant for establishing a noun classificatory hierarchy, or as Givon (2001:56) calls it, a “markedness hierarchy”. The features¹⁴ of this hierarchy are arrayed as follows:

ENTITY \supset TEMPORAL \supset CONCRETE \supset ANIMATE \supset HUMAN (Givon, 1984:56 and 2001:56).

¹⁴ Givon (2001:56) uses the term “feature” here. As will be seen later in the discussion, Givon does not present these “features” as gradable and therefore arguably not as “attributes” in the true functional typological sense.

Givon (1984:56) defines the terms in the array above as follows:

Entity means 'that which exists', **temporal** 'exists at a particular time', **concrete** 'exists in both time and place'; **animate** adds ... the feature 'living organism', and **human** also the feature 'be human'.

All nouns denote entities. Being an entity entails some kind of existence. The existence of **abstract** entities (e.g. *poverty*) cannot be defined in terms of time or space. Entities that exist in time only (e.g. *day, week, month* and *year*), are denoted by **temporal** nouns, which may be regarded as "*semi-abstract*" (Givon, 1984:56) because they have temporal dimension, but lack spatial dimension. Concrete entities have both spatial and temporal dimensions. Standard grammars of English make similar distinctions between concrete and abstract (non-concrete) nouns (Quirk et al., 1985:347; Huddleston & Pullum, 2002:337, 375; Biber et al., 1999:241).

The semantic attributes in the array above can be expressed in systemic terms as a series of binary choices (*either...or...*), and as Givon (2001:56) terms it, a "chain of one-way conditional implications" (*if...then...*).

An entity is either abstract or temporal. If temporal, then it is either non-spatial or concrete (spatial). If concrete, then it is either inanimate or animate. If animate, then it is either non-human or human.

Langacker's (2008:103) definition of the noun prototype as a physical object confirms the unmarked status of the feature 'concrete', but because 'physical object' can be understood as denoting both animate and inanimate objects, the prototype cannot be employed to confirm markedness status beyond the node of 'concrete' in the hierarchy. From this point in the system, the linguist has to turn to corpus data to establish which feature is the unmarked one.

Both concrete and non-concrete (abstract) nouns can be sub-classified in terms of countability. Here the distinction is between **count** nouns, i.e. individuated entities, or **mass** nouns (Givon, 2001:57). Standard grammars of English likewise distinguish between countable and uncountable, or count and mass, nouns (Quirk et al., 1985:245; Huddleston & Pullum, 2002:328, 334-340; and Biber et al., 1999:241-244). Construal as

countable or uncountable corresponds to whether something is conceived as **bounded** or **unbounded** (Langacker, 2004:81).

Halliday and Matthiessen (2004:326) identify three **semantic clines** on which nouns can be located, namely **countability** (count/mass), **animacy** (conscious/non-conscious) and **generality** (general/particular). Their distinction in terms of animacy is not between factually animate and inanimate (i.e. living and non-living) but between “(a) conscious things, which are referred to as *he/she*, [and] (b) non-conscious things, those referred to as *it*” (Halliday & Matthiessen, 2004:326). From the use of personal pronouns in their definition, one can conclude that “conscious” implies a thing which is construed as human or as having human-like consciousness, i.e. a personified thing (Halliday & Matthiessen, 2004:198). Classifying nouns or adjectives preceding the head noun are also regarded as being involved in designation, rather than in qualification, for example, *electric* trains (compare *steam* trains).

Grounding elements occur with nouns and verbs. In a nominal group where a common noun realizes the function of thing, the determiner is the grounding element. Whereas designation is lexical, grounding is grammatical. **Determination** ties in with the referential function of nominal groups in discourse. With regard to reference, a number of terminological issues arise that will be clarified in Chapter 4. Important distinctions are specific/non-specific (Halliday & Matthiessen, 2004:312-313) and definite/indefinite.

The **quantifying** function in nominals is typically realized by numerals which indicate a definite, or exact, quantity (e.g. *ten*) or by indefinite quantifying determiners (*many*, *several*, *much*) which indicate an inexact quantity. In grammar textbooks, a distinction is normally made between quantifying numerals, i.e. Numeratives (‘quantitatives’) in SFG, which may specify an either an exact or inexact number, for example, *three*, *many*, and ordering numerals, i.e. Numeratives (‘ordinatives’) in SFG, which, likewise, may specify either an exact or inexact place in order (Halliday & Matthiessen, 2004:318), for example, *third*, *subsequent*. However, ordinatives are more quality-like in nature and although they are not strictly grounding elements in CG-terms, they evoke the ground. The number features of the lexical noun also need to be taken into account when the quantifying function is considered. With regard to quantification, it is important to note

that partitive *of*-constructions sometimes give rise to nominal groups where the logical head is not the experiential Thing, as illustrated below:

Figure 2.16 Head and Thing in *of*-partitions

	<i>most</i>	<i>of</i>	<i>the</i>	<i>people</i>
Logical structure	Head	Postmodifier		
Experiential structure	Numeral		Deictic	Thing
Terms used in this study	Quantifier		Determiner	Denotation

Qualification is realized in various ways in the nominal. A Thing-designation may be qualified in terms of its order in place or relation to something else (e.g. *subsequent*, *same*), a subjective, evaluative property (e.g. *awful*) an objective property (e.g. *large*, *yellow*) or its involvement in a relationship, e.g. *the guy on the roof*, *the guy sitting on the roof*.

In English, the choice of grammatical unit through which a quality is expressed in the nominal group (word, group, phrase or clause) largely corresponds to the position of the quality expression (prenominal or postnominal). Prenominal qualities are typically expressed by words or word groups, whereas postnominal qualities are typically expressed by clauses or prepositional phrases (Halliday & Matthiessen, 2004:323; Quirk et al., 1985:1239). Notable exceptions are postposed adjectives, e.g. *something different* (Quirk et al., 1985:1293-94; Biber et al., 1999:519; Payne & Huddleston, 2002:445) and “nonce-formations” as in *my do-it-yourself skills* (Payne & Huddleston, 2002:444). Since clauses and preposition phrases represent major and minor processes, they characterize the thing in terms of some process in which the thing is a direct or indirect participant, when they function within the structure of the nominal group (Halliday & Matthiessen, 2004:4).

The fact that qualities can be expressed within the nominal group allows the nominal group to become very complex. Biber et al. (1999:576) remark that “[i]n principle, there is no limit to the complexity of noun phrases” and point out that modifiers can “present a wealth of additional propositional information beyond that presented by the matrix clause

itself” (also see Quirk et al., 1985:1238 in this regard). This provides a basis for a working definition of the term qualification for purposes of this study, namely additional propositional information about the Thing within the nominal group, but excluding, for practical purposes only, the information implied by the very name of the Thing. Pre- and postnominal modifiers are typically the vehicles of this additional propositional information. The view taken here – in accordance with Payne and Huddleston (2002:438), Quirk et al. (1985:1238-9) and Biber et al. (1999:574) and contrary to Givon (2001:1) and Halliday and Matthiessen (2004:329) – is that determiners are not a kind of premodifier, but distinct from modifiers, because they do not convey additional propositional information in the narrow sense of the word *propositional*. Bear in mind, though, that although Halliday and his followers treat determiners as part of premodification in the logical structure of the nominal group, they do distinguish different functions for determiners and adjectives in the experiential structure of the nominal group: determiners typically occurring in the Deictic slot and adjectives typically occurring in the Epithet and Classifier slots (Halliday & Matthiessen, 2004:312). The information conveyed by modifiers, as well as other kinds of quality expressions, is **characterizing** in nature. This characterization serves primarily to further specify the **type** of Thing **denoted** by the head noun.

Apart from their function as premodifiers of nouns, adjectives can also occur as the heads of adjective phrases that fulfil the syntactic roles of subject complement or object complement (i.e. predicative functions). Since adjectives can themselves be premodified or followed by a complement, it is useful to employ the term “adjective phrase” in the way that Pullum and Huddleston (2002a:542-550) do to refer to the adjective and its modifiers or complement. They supply the following example:

(15) Kim was *very keen to take part*.¹⁵

The notion of adjective phrase is not reconcilable with the terminology of Hallidayan SFG, but it can be argued that it is not conceptually contrary to a systemic-functional grammatical analysis, as Tucker’s (1998) comprehensive study of the lexicogrammar of adjectives demonstrates.

¹⁵ The Adjective phrase is italicized. The head is underlined. The degree adverb *very* is a modifier of the adjective *keen* and *to take part* is the complement.

In Hallidayan SFG, the adjective is regarded as a kind of nominal since it can head groups realizing the function of Participant in the clause (Halliday & Matthiessen, 2004:52, 219, 320-321), thus the conventional term “predicative adjective phrase” translates to “adjective-headed nominal group”, i.e. a nominal group where the Thing is absent from the experiential structure of the nominal group.

In the Cardiff school of SFG (Fawcett, 2008), the term “quality group” roughly covers the conventional notions of “adjective phrase” and “adverb phrase”, but “adjectival expressions” are typically associated with qualities of things and “adverbial expressions” with qualities of situations (Tucker, 1998:64). The preference for the term “quality group” in the Cardiff Grammar rather than, for example, “adjectival group” is due to the fact that word class labels “have no part to play in grammatical description within the Cardiff Grammar” (Tucker, 1998:53). The notion of “class” is relevant only for the rank of group, but in this regard, Tucker remarks,

and even here classes of group [...] are defined not with reference to the structure of the unit next above, as Halliday originally defined them, but according to their internal structure, i.e. their componence (1998:43).

Since the current study is not subject to the theoretical and ideological constraints on terminology imposed by the Cardiff grammar, the term **adjectival group** will be used here. This is in keeping with the general use of “group” in SFG, but emphasizes the fact that the pivotal element in this group is an adjective, and thus allows for comparison with the concept of adjective phrase found in the reference grammars of English. In describing types of qualities (as realized by adjectives), reference to the positions of adjectives, or adjectival groups, is inevitable because of the strong correlation between the meaning of an adjective and its grammatical position. Tucker (1998:126) remarks, “Foremost among the lexicogrammatical correlates of semantic distinctions is the order in which Qualities are presented in modifier sequences”. There are many cases in English where the meaning of a particular adjective imposes a restriction on the positions in which it can occur in the nominal group. Some adjectives are restricted to a position directly following the central determiner (the post-determiners or Post-deictics such as *same*); some to a position directly before the noun (the classifiers as in *legal matters*); some to a post-positive position (e.g. *heir apparent*) and some to a predicative

position (e.g. *afraid*). It goes without saying that this semantic restriction on the position of adjectives is not restricted to English. It is, in fact, even better illustrated in languages such as French where the same adjective may occur to the left or right of the noun depending on the sense in which it is used (Bouchard, 2002:73ff). It is also worth noting that in Setswana, the first language of the subjects who contributed to the TLEC, there are no prenominal adjectives, only qualificative word groups which follow the noun (Krüger, 2006:301; Cole, 1955:139).

2.4 Framework for interpretation

A corpus concordance of a word or combination of words provides us with a great many instances of use, or usage-events. In order to make sense of these surface patterns and the meanings they represent, the notion of a construction is very useful. The notion of construction is central to usage-based (cognitive) approaches to language. The notion of pattern that emerged in data-driven strands of corpus linguistics has some aspects in common with constructions and is particularly useful in a corpus study. Both will be employed in the interpretation of corpus data in Chapters 4, 5 and 6.

2.4.1 Constructions and schemas

Construction grammar originates from attempts to account for idiomatic expressions that form part of a speaker's linguistic knowledge (Croft & Cruse, 2004:225). A construction is a direct pairing of form and meaning (Bybee, 2010:76; Goldberg, 1992:1). Goldberg explains that "a construction is posited in the grammar if it can be shown that its meaning and/or its form is not compositionally derived from other constructions existing in the language" (1992:4). Postulation of a construction can account for cases where the "whole is not built up from the lexical items in a straightforward way", for example, *Pauline smiled her thanks* (Goldberg, 1992:12). Langacker's (2008:161) definition of constructions as symbolic assemblies does not impose the strict condition above. However, he points out that the ordering of the components in a composition is itself part of the meaning of the composition, as can be seen from the patterns *jar lid factory*, *lid factory jar* and *jar factory lid* (2008:168). In each of these expressions the meanings of

the individual components (words) is analyzable, but they mean different things because the components are ordered differently. For Langacker the meanings of the component structures are insufficient to predict the meaning of the composition: “the basis for prediction also has to include a conventionally established compositional pattern” (2008:169). Since the components are themselves form-meaning pairings one, can say that they are constructions which make up the longer expression, which in turn is also a construction. This is what is meant by the compositional nature of constructions.

According to Bybee a construction “has sequential structure and may include positions that are fixed as well as positions that are open” (2010:9). She goes on to explain the usefulness of constructions in an exemplar model of language usage:

Representing grammatical patterns in constructions is particularly appropriate in an exemplar model, since constructions are direct pairings of form with meaning with no intermediate level of representations, such as phrase structure rules would provide. This is appropriate since what language users experience is specific instances or tokens of constructions. They map similar tokens onto one another to establish exemplars and these exemplars group together to form categories that represent both the fixed and the schematic slots in constructions (Bybee, 2010:26).

Bybee uses eight instances (from the BNC¹⁶) of a construction where the verb *drive* is used with an adjective (e.g. *mad*) or prepositional phrase (e.g. *up the wall*) to illustrate how this mapping and grouping works to arrive at a construction with fixed and open slots. From the eight instances, she arrives at the following representation (Bybee, 2010:27):

SUBJECT [DRIVE] {me/you/him/her/the producer} {crazy/mad/up the wall}

The verb *drive* is fixed. In the representation above exemplars of what might fill the open positions after the verb occur in parenthesis { }. The exemplars in the first position after the verb are mostly pronouns, but crucially they refer to animate, human beings. The exemplars in the second position after the verb all mean ‘crazy’. The SUBJECT in the representation is a truly open position, which can be filled by virtually any nominal group or nominal clause. This is the most schematic element of the construction, whereas the verb *drive* is most specific element.

¹⁶ *British National Corpus*

2.4.2 Patterns and the idiom principle

The Pattern Grammar of Hunston and Francis (1999) is intended as a lexically-based, linear representation of grammatical structure. They focus on “the patterns that are associated with particular lexical items and that are observable from investigation of an electronically-stored corpus of written and spoken texts” (Hunston & Francis, 1999:1). The fact that patterns are derived from usage instances (in corpus concordances) makes it an exemplar approach. The patterns associated with certain words have much in common with constructions where one or more lexical items occur in fixed slots, but as we shall see patterns, as opposed to constructions, are more limited in the extent to which they allow for schematization. Nonetheless, the concept of pattern is very useful in the interpretation of corpus concordances of individual words, which is the kind of data that is used in this study. The concept of pattern is explained as follows:

The patterns of a word can be defined as all the words and structures which are **regularly** associated with the word and which **contribute to its meaning**. A pattern can be identified if a combination of words occurs **relatively frequently**, if it is dependent on a particular word choice, and if there is a clear meaning associated with it (Hunston & Francis, 1999:37, my emphasis – YVB).

Meaning and frequency are important in the identification of patterns in concordances of a word. Hunston and Francis (1999:71) caution that “frequent co-occurrences of words do not necessarily indicate the presence of a pattern, and that **interpretation** of concordance lines is necessary to the identification of patterns” (1999:71, my emphasis – YVB). Hunston and Francis (1999:56-57) identify a number of patterns that are generalized according to word class, for instance **a N**, **the N** and **N of n**, where “N” denotes the primary noun and “n” the secondary noun. Examples of specific nouns with these patterns are *a cinch*, *the blues* and *series of acts* respectively. When one studies a specific word’s concordance, one might find that the general patterns identified for its word class do not all occur and that some patterns occur much more frequently than others. Hunston and Francis (1999:181) postulate that new subclasses of a word class can be established on the basis of the patterns that certain words frequently share. An example of this would be shell nouns, which are nouns that require a lexicalization in their immediate context. This lexicalization is frequently given in a complementing *that*-clause, as in *There was a suggestion that the whole thing was a joke* (Hunston & Francis, 1999:57, 185). In other words these nouns frequently occur in the pattern **N**

that. In the case of shell nouns it is important to note that the *that* in the pattern introduces a complement clause which is a lexicalization of the shell noun and not a relative clause which modifies a noun with its own lexical content. This is illustrative of the interpretation that is needed for the identification of patterns instead of merely looking at the word forms to the left and to the right of the search word in a concordance.

Pattern grammar is suited to a data-driven approach to corpus data (see Chapter 3) and as such builds on earlier work in corpus studies of collocations most notably that of Sinclair (1991) (Hunston & Francis, 1999:14ff). With regard to establishing a framework for the interpretation of corpus concordance data, Sinclair's distinction between the open-choice principle and the idiom principle is relevant. Sinclair (1991:109) defines the open-choice principle as follows:

This is a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word, a phrase or a clause), a large range of choice opens up and the only restraint is grammaticalness.

The description above fits SFG like a glove. Sinclair (1991:109) goes on to categorize this way of describing language as a slot-and-filler model. However, "the open-choice principle does not provide for substantial enough restraints on consecutive choices" (Sinclair, 1991:110). As a solution, Sinclair (1991:110) suggests the idiom principle, which is defined as follows:

The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.

According to Sinclair (1991:112), the idiom principle should be at least as important as grammar in explaining how meaning is created in a text.

2.4.3 Conventionalization in new varieties of English

Conventionalization underlies both the concepts of construction and pattern. If the meaning of a construction is not predictable from the meanings of its components, some

kind of conventionalization in the use of the construction is implied (Croft & Cruse, 2004:230). Likewise, the relative frequency that a pattern with a certain meaning needs to occur is indicative of conventionalization. Van Rooy (2010, 2011) draws on Croft's definition of conventionalization to arrive at a principled distinction between error and conventionalized innovation in the analysis and description of the so-called non-standard features of African Englishes. Innovation entails the creation of new forms by an individual speaker in an attempt to communicate (Van Rooy, 2011:192). These new forms then become "part of the feature pool of potential linguistic forms in a speech community" (Van Rooy, 2011:192). Van Rooy (2011:192-193) points out that in a language contact situation and especially in the contexts in which Outer Circle varieties of English are used, the potential for innovating new forms increases. Some of these innovations may be learner errors. If they become accepted through social processes, they become conventionalized innovations, and thus become entrenched in the grammatical system of individuals and conventionalized in the speech community. If not, they remain errors and do not become conventionalized. Van Rooy (2011) goes on to provide three examples of linguistic features that originated as errors, but became acceptable in their speech communities and thus became conventionalized innovations. He concludes that "[g]enuine new linguistic conventions emerge from forms that may have started out as errors" (Van Rooy, 2011:189).

The implication of this for the present study is that new, even erroneous forms, encountered in the corpus data are potentially conventionalized innovations. The frequency of the innovation compared to the frequency of the "standard" equivalents in the data will provide an indication of the extent to which innovative forms have become accepted and thus established as constructions, i.e. conventionalized form-meaning pairs, in BSAfE.

2.5 Conclusion

In the first section of this chapter, the current study was situated in the broader framework of functional or usage-based approaches to the study of language. In the second section a framework of analysis based on a function-based description of clauses, nouns and nominal groups were set out. Clauses construe experience in terms

of Processes, Participants and Circumstances. The nominal group is the realization of a Participant in a Process. Based on the experiential structure of the nominal group in SFG and the referential and classificatory functions of nominals in CG, it is postulated that the elements of a nominal group serve four potential specifying functions, namely Designation, Grounding, Quantification and Qualification. The third section provided key concepts to interpret corpus data. In constructions and patterns observed in the data will be interpreted either as conventionalized innovations or as learner errors.

In Chapter 3, a general methodological framework for the corpus analyses in Chapters 4, 5 and 6 is presented. The analysis of the concordances of the noun *people* in Chapter 4 starts with sorting and grouping the concordance lines according to the elements present in the nominal group. In this chapter, all four specifying functions of the elements of nominal groups where the noun *people* realizes the Thing element will be taken into account, i.e. the denotation meaning of the noun *people*, the choice of determiners (grounding), as well as quantity and quality-specification of *people*. The analysis is based on the experiential structure of the nominal group set out in Section 2.3.2. Here a distinction is made between determiners that serve only a grounding (Deictic) function and those that serve a primarily quantifying function. With regard to the specifiers of quality (Epithets, Classifiers and postnominal Qualifiers), the concordance lines are further classified with regard to the position of the modifiers (i.e. the logical structure of the nominal group). Prenominal adjectives of *people* are sub-classified in terms of the Quality type realized by the adjectives (based on Tucker, 1998). Postnominal modifiers are sub-classified according to their forms (groups, preposition phrases or clauses). Where applicable the type of postmodifying clause is further sub-classified in terms of its formal properties and the Process type it encodes, for example, “a finite *who*-clause encoding a relational process”.

The Deictic slot of the experiential structure of the nominal group (Section 2.3.2) is important for the concordance analyses in Chapter 5 which focuses on the articles and definite determiners. Here the syntactic role (both experiential and logical) (Section 2.3.1) in which the relevant nominal group functions is taken into account to the extent that it potentially influences the choice of determiner. The Deictic and Numerative slots of the experiential structure of the nominal group (Section 2.3.2) are relevant for the analysis of quantifiers in Chapter 6.

CHAPTER 3

Methodology

3.1 Introduction

In Chapter 1 it was pointed out that early corpus studies of BSAfE (e.g. De Klerk, 2003a,b) did not go beyond the use of corpus data to exemplify previous claims (e.g. by Gough, 1996 and Van der Walt & Van Rooy, 2002). Minow (2010:1) criticizes the incomplete account of BSAfE in which such studies result:

Unfortunately, most accounts of BSAE often merely list the features and hence do not provide much insight into how common these features actually are. Even if corpora or databases of BSAE are searched for occurrences of a particular BSAE feature, the reports often simply state that a feature occurs a given number of times without providing how often the speakers adhere to the rules of standard English (cf. de Klerk 2003a, 2003b).

In the present study, corpus data will be examined with a view to uncover innovative patterns in the grammar of BSAfE. The focus will not be on non-standard approaches. Frequencies will be normalized and compared to a control corpus. The corpora used in this study are described in Section 3.2 below. In Section 3.3 this study is characterized as a hybrid of corpus-driven and corpus-based methods. In Section 3.4 the data processing tools and steps used in this study are set out.

3.2 The corpora

Both the corpora used in this study form part of the *International Corpus of Learner English* (Granger et al., 2009). The main source of data is the *Tswana Learner English Corpus* (TLEC). *The Louvain Corpus of Native English Essays* (LOCNESS) serves as control corpus (Granger et al., 2009:42). Both corpora consist of essays written by students ranging from age 18 to 25, but with the majority being between 19 and 21 years old (Van Rooy, 2008a:275) and both corpora consist of approximately 200 000 words of running text.

The TLEC consists of 519 essays provided by four South African institutions, namely Kimberley College (64 essays), University of the North-West (428 essays), Potchefstroom University (20 essays) and the University of the Witwatersrand (7).¹⁷ The essays in the TLEC were contributed by humanities or social science undergraduate students who speak Setswana as a first language and who were enrolled for a course in English as part of their degree programme (Van Rooy, 2008a:273). The medium of instruction for the last eight years of their school careers had been English and they were all enrolled at tertiary institutions where the medium of instruction was English. Van Rooy (2008a:273) points out that most of the contributors were educated at township schools (as opposed to racially integrated schools) in towns or small cities and were also studying at tertiary institutions where contact with native speakers of English would be limited (the exception being the seven essay writers from the University of the Witwatersrand). He goes on to argue that the contributors to the TLEC should not be regarded as “learners who are still a long way off the eventual target of some acquisition process. They are all relatively close to the terminal point” (Van Rooy, 2008a:274). The TLEC is regarded as representative of BSAfE (Van Rooy, 2008a:269).

LOCNESS is comparable with the other ICLE corpora in terms of text-type (argumentative essays) and level of writing skill and style (novice writers as opposed to professional writers or edited published writing) (Granger et al., 2009:42). Roughly half of the LOCNESS texts consist of exam essays and argumentative essays by humanities students from Surrey (United Kingdom). The other half comprises argumentative essays by students from Marquette or South Carolina Universities (USA) (Van Rooy, 2008a:275).

3.3 Approaches to corpus data

As we have seen in Chapter 1, the identification of the characteristic features of BSAfE have until quite recently been based on the aspects in which it differs from Standard

¹⁷ In the references of each example quoted in this study essays from these four institutions are indicated as TSKC (Kimberley College), TSNO (University of the North-West), TSPO (Potchefstroom University) and TSWI (University of the Witwatersrand, also known as Wits).

English that were most salient to native speakers of English or speakers with native-like competence. The first corpus investigations into the features of BSAfE were still based on characterizations arising from intuition. The best example of this kind of application of a corpus is De Klerk's (2003a, 2003b) work which provides corpus evidence for features that were initially noticed in the speech and writing of BSAfE speakers based on their obvious difference to Standard English (De Klerk & Gough, 2002). Biber et al. (1998:3) argue that corpora should be used to "uncover typical patterns rather than making judgments of grammaticality". They also maintain:

Because we are looking for typical patterns, analyses cannot rely on intuitions or anecdotal evidence. In many cases, humans tend to notice unusual occurrences more than typical occurrences, and therefore conclusions based on intuition can be unreliable (Biber et al., 1998:27).

In this study the focus will fall on identifying typical patterns in the corpus data. The use of a native English control corpus does not necessarily imply a focus on the differences or on "unusual occurrences" (although these will obviously be reported if they are observed). A good example of a corpus linguistic investigation of BSAfE focusing on the typical rather than the unusual is that of Van Rooy and Terblanche (2009). They focus on "'normal' grammatical features, rather than errors" (Van Rooy & Terblanche, 2009:178) in the TLEC and LOCNESS and gain insight from both the differences and the commonalities between the corpora.

Lindquist cautions that the corpus linguist should not only report frequencies, but should "go back to individual examples from the corpus to look at the reality behind the figures" (2009:26). This is the purpose of electronic concordance tools which will be discussed further in section 3.4 below. However, it is not enough to simply list examples without further interpretation. As Mahlberg says, "[a]utomatically retrieved data has to be interpreted by the human observer and a critical attitude is necessary to avoid a skewed picture" (2005:15).

With regard to the retrieval of data by means of concordance searches, for instance, and the interpretation of the retrieved data, two broad approaches are distinguished, namely corpus-based studies and corpus-driven studies (Tognini-Bonelli, 2001:11; Mahlberg, 2005:17). Biber et al. (1998) are self-proclaimed adherents of the corpus-based

approach. They define the goal of corpus-based investigation as the identification and analysis of “association patterns” by which they mean “the systematic ways in which linguistic features are used in association with other linguistic and non-linguistic features” (Biber et al., 1998:5). There are two perspectives from which association patterns can be investigated, namely (1) the use of a certain linguistic feature, or (2) the characteristics of texts or varieties. Biber et al. (1998:5) emphasize that a corpus-based approach should go beyond reporting frequencies of linguistic features and should include “qualitative, functional interpretations of quantitative patterns”.

Tognini-Bonelli (2001:65) points out that the term “corpus-based” can be vague and general and refers to a variety of work that makes use of a corpus. For purposes of distinguishing between a corpus-based and a corpus-driven approach, she applies a stricter definition to “corpus-based” than what can be gleaned from Biber et al.’s (1998:4-5) description. She defines a corpus-based approach as

a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study (Tognini-Bonelli, 2001:65).

One might argue that Biber et al.’s (1998) approach, with its aim of identifying association patterns between linguistic features, goes beyond the expounding, testing and exemplification of existing theories. However, Tognini-Bonelli (2001:66) regards any approach where corpus data is analyzed in terms of pre-existing linguistic categories, which is the case with the work of Biber and his collaborators (e.g. Biber et al., 1998; Biber et al., 1999), as corpus-based.

We could say, therefore, that corpus-based linguists adopt a ‘confident’ stand with respect to the relationship between theory and data in that they bring with them models of language and descriptions which they believe to be fundamentally adequate, they perceive and analyse the corpus through these categories and sieve the data accordingly (Tognini-Bonelli, 2001:66).

It would seem that what bothers Tognini-Bonelli (2001:81, 84) about corpus-based linguistics is the idea that the corpus data is used to validate or supplement existing theoretical models. The corpus-driven approach, on the other hand, is one where theories and hypotheses are derived from the corpus data (Tognini-Bonelli, 2001:84, 87; Mahlberg, 2005:18). Mahlberg (2005:18-19) problematizes the application of a purely

data-driven approach with regard to her own study of general nouns. For instance, she asks, “can we take it for granted that we know what a noun is, or would we have to start even earlier with a description of all words before we can form classes such as nouns?” (Mahlberg, 2005:18). Mahlberg goes on to point out that “[e]ven if the aim is to follow a corpus-driven path, what is observed in the corpus needs to be named and described”, although she concedes that the terminology used to describe the data might impact on the results and should therefore be treated with caution (2005:19).

Rayson (2008:521) overcomes similar concerns by calling his approach in the investigation of key words and semantic domains “data-driven” rather than “corpus-driven”. Although his approach has “the corpus as main informant” in common with corpus-driven linguistics, he makes use of part-of-speech tags based on existing categories. This means that his approach uses elements from both corpus-driven and corpus-based linguistics.

The current study also draws on both corpus-based and corpus-driven approaches. It is corpus-based in the sense that it draws on grammatical categories established in broadly functional (including cognitive) approaches to language in order to determine search words and strings for concordance retrieval, as well as to analyze the concordance data. However, it is not taken for granted here that the linguistic structures of BSAfE and the functions that they serve are the same as those in Standard English. As was pointed out earlier, non-native speakers (of Standard English) are the main primary providers of input to the feature pool of BSAfE and BSAfE is not regarded as an interlanguage on the way to some target (Van Rooy, 2008a:274). In this context it is only natural to question to what extent the descriptive categories for Standard English are applicable to this variety. The analysis framework set out in the previous chapter (Section 2.3) is based on descriptions of Standard English and was taken as a starting point. It should be kept in mind that in establishing the analysis framework itself, a single, uniform theory was not adopted, because no single theory, such as Halliday’s SFG (Halliday & Matthiessen, 2004), or Quirk et al.’s reference grammar (1985), was deemed sufficient to illuminate the functional aspects of the nominal group. This implies a critical stance to existing categories, especially as they apply to non-native varieties. However, one has to start somewhere, and these categories provided a good starting point for the initial sorting of the concordance lines.

In Chapter 4, it is accepted that the word *people* belongs to the class of noun (except in cases where it means ‘to populate’, but no such usage appears in the corpus data). The concordance lines of the word *people* are initially sorted according to the elements of the nominal groups in which it occurs. The first step is to code instances where *people* occurs in a nominal group that functions as the complement of the preposition *of* in a partitive quantifying construction, e.g. *most of the people*. The second step is to distinguish instances of *people* as only element in the nominal group (i.e. without determiner or modifiers) from instances where other elements such as determiners or modifiers are present. Eventually separate set codes are given to (1) occurrences with a determiner, but without modifiers, (2) occurrences with both determiners and modifiers, and (3) occurrences with no determiner but with modifiers. The semantic kinds of qualities conveyed by the modifiers are mostly categorized in terms of Tucker’s (1998) systemic network of quality type.

However, the decision to start the corpus linguistic investigation of nominal group internal functions in BSAfE with a concordance of the noun *people* is corpus-driven in that *people* is the most frequent noun in both corpora and yet has a high positive keyness in the TLEC. The aim of the concordance analysis is to find out the reasons for the high keyness of this noun and also to determine aspects of nominal structure that warrant further investigation. The findings of the concordance analysis of the noun *people* indicate that determiners and quantifying constructions warrant further investigation. These observations happen to partly correspond to some of the qualitative observations about the characteristics of BSAfE mentioned in the literature (see Chapter 1). Determination and quantification are dealt with in Chapters 5 and 6 respectively. Concordance searches of the word forms that function as articles, possessive determiners and demonstrative determiners were used as main source of data in Chapter 5. These obviously constitute well-established categories. Likewise, the analysis in Chapter 6 draws mainly on concordance searches of the words that are typically classified as quantifiers in reference grammars of English. However, the concordances of certain quantifiers, like *most*, showed that there are quantifying constructions in BSAfE which are not accounted for, neither in descriptions of Standard English, nor in BSAfE literature – except for being mentioned by way of example. In

these cases, interpretation of the patterns observed in the corpus data can be used to enrich the theory.

In summary then, the approach taken in this study is a hybrid of corpus-based and corpus-driven approaches. The initial research questions about the functional elements of the nominal group are rooted firmly in functional and usage-based theories of language. Decisions about which aspects of the nominal group to investigate are motivated by frequency data and the patterns observed in the concordance of *people*. The initial sorting (set coding) of the concordance data are based on existing grammatical categories, but in the interpretation of the concordance data due attention is given to potential areas where existing descriptions can be expanded.

3.4 Data processing

Frequency lists and concordances form the core of the corpus data processing (Rayson, 2008:520; Evison, 2010:122). Frequency counts, by means of wordlists, and concordance searches can be carried out on raw, unannotated corpora or on corpora that have some kind of annotation, for example, part-of-speech tagging. Most of the analyses in this study have been carried out on the unannotated versions of the TLEC and LOCNESS. Wordlists, keyword lists and concordances were generated with *WordSmith Tools* (Versions 4.0 and 5.0) (Scott, 2010).¹⁸ The notion of keywords and the use of frequencies, in general as well as how they are deployed in this study, are described in Section 3.4.1. The use of concordances is described in Section 3.4.2. Sections 3.4.3 and 3.4.4 describe the limited extent to which annotated data was utilized in this study.

3.4.1 Wordlists, frequencies and keywords

Frequency lists provide a count of all the words (tokens) in a corpus. The results are displayed according to frequency or alphabetically (Evison, 2010:124). The table below represents the ten most frequent words in the TLEC according to a list generated by the

¹⁸ When the first part of the research was carried out WordSmith 4.0 was the current version, but WordSmith 5.0 subsequently became available and was used in the latter part of the study.

WordList function in Wordsmith Tools 4.0. The word *the* is ranked first in the frequency list and occurs 9086 times in the corpus. This means that 4.51 percent of the corpus consists of the word *the*. It occurs in 517 (99.61%) of the 519 texts that make up the corpus. The wordlist below also indicates the raw frequency, corpus percentage and frequency ranking of the article *a* (ranked 10th most frequent token).

Table 3.1 Excerpt from frequency-ordered WordList (TLEC)

N	Word	Freq.	%	Texts	%
1	THE	9,086	4.51	517	99.61
2	TO	7,015	3.49	518	99.81
3	OF	5,444	2.70	518	99.81
4	AND	5,159	2.56	516	99.42
5	THEY	4,324	2.15	491	94.61
6	IS	4,063	2.02	516	99.42
7	IN	3,844	1.91	514	99.04
8	THAT	3,384	1.68	502	96.72
9	ARE	3,354	1.67	507	97.69
10	A	3,213	1.60	507	97.69

Should one be interested in the number of times that the indefinite article form *an* occurs, one can view the wordlist alphabetically. The table below shows the word *an* in its alphabetical ranking. From this table it is also clear that spelling errors affect frequencies, not so much with regard to types, but with regard to type-token ratios.

Table 3.2 Excerpt from alphabetically-ordered WordList (TLEC)

N	Word	Freq.	%	Texts	%
371	AMUNITION	1		1	0.19
372	AN	378	0.19	226	43.55
373	ANABIDING	1		1	0.19

The statistics tab in the wordlist generated by Wordsmith allows one to distinguish between the number of running words (tokens) and the number of distinct words (types) and indicates the ratio between them, as can be seen from Table 3.3 below:

Table 3.3 Statistics tab in WordList (TLEC)

N	Overall
tokens (running words) in text	201,283
tokens used for word list ¹⁹	200,897
types (distinct words)	9,706
type/token ratio (TTR)	4.83

Spelling errors like *anabiding* add to the number of types and therefore affect type-token ratio. However, type-token ratio is not relevant to the current investigation since it is mainly concerned with grammatical features and not with lexicological issues.

The KeyWords function Wordsmith Tools compares two wordlists to identify key words. Scott (2010:152) defines key words as “those whose frequency is unusually high according to some norm.” In the case of the current study, LOCNESS is taken as the norm, or the control corpus for determining which words in the TLEC are key words. The list of key words is based on a comparison of the wordlists of the two corpora created with the WordList tool. Keyness is automatically calculated in the KeyWords function by means of either the classic chi-square test of significance with Yates correction for a 2 X 2 table or Ted Dunning's Log Likelihood test (Scott, 2010: 166).

Table 3.4 below shows the first 10 items of the TLEC-LOCNESS keywords list generated by the KeyWords tool in WordSmith 5.0. using log likelihood to calculate the keyness. This list contains the nouns *Aids*, *HIV*, *Poverty*, *Africa*, *money* and *people*. Though *people* is the most frequent noun in each corpus according to their word lists, it is not the noun with the highest keyness. The other nouns mentioned all have higher keyness in the TLEC. However, the high frequency of these nouns in the TLEC compared to LOCNESS can be attributed to the essay topics assigned to students contributing to the corpus. The same cannot be said of *people*. Due to its high frequency in both corpora and its high positive keyness in the TLEC, which is not due to an essay topic, the noun *people* was chosen as a starting point for exploring the nominal group. In Chapter 4 the structure of nominal groups containing *people* as Thing element

¹⁹ The difference in number of running words and number of words included in the wordlist is mainly due to the exclusion of items that are not orthographically written as words, e.g. digits, from the word list.

are analyzed in an attempt to uncover ways in which nominal group structure in the TLEC differs from that in LOCNESS, while at the same time attempting to explain the comparatively high frequency of the noun *people* in the TLEC.

Table 3.4 Top ten key words in the TLEC when compared to LOCNESS

N	Key word	Freq.	%	RC. Freq.	RC. %	Keyness
1	THEY	4,331	2.16	1,257	0.62	1,842.40
2	AIDS	1,411	0.70	53	0.03	1,591.36
3	HIV	1,155	0.57	9		1,522.48
4	POVERTY	1,037	0.52	5		1,393.72
5	AFRICA	987	0.49	2		1,354.11
6	ARE	3,355	1.67	1,454	0.72	799.95
7	YOU	1,505	0.75	354	0.17	781.46
8	MONEY	953	0.47	118	0.06	751.46
9	SOUTH	697	0.35	42	0.02	709.38
10	PEOPLE	2,068	1.03	747	0.37	662.08

In analyzing the structure of the nominal groups with *people* as Thing, it came to light that the TLEC contributors used proportionally more demonstrative and possessive determiners rather than the definite article and that there was a very high usage of quantifying partitive *of*-constructions, many of which involved the word *most*. Returning to the keywords list (based on log likelihood), it is interesting to note that the articles are very obvious negative keywords in the TLEC: *the* (-825.88), *a* (-131.11), *an* (-157.1). There are three possessive determiners in the TLEC with a positive keyness of more than 100, namely, *our* (276.12), *their* (175.57), and *your* (107.45). The quantifiers *some* and *most* also have a high positive keyness in the TLEC: *some* (253.04), *most* (249.10).

3.4.2 Concordances and random sampling

Concordances allow one to see all the instances of a given search word in the corpus in their context of usage. The Concord tool in WordSmith also allows one to classify these

instances by assigning an alphanumeric code in the Set column. All the instances in both the concordances of *people* are subclassified in this manner in Chapter 4. In Chapter 5, the large size of the concordances of *the*, *a*, *their* and *that* necessitates random sampling to arrive at 400 concordance lines per search item per corpus. In their random sampling of concordances from the TLEC and LOCNESS, Van Rooy and Terblanche (2006:167) also made use of random samples of 400 based on the fact that a sample size of at least 384 was necessary in analyzing proportions of these corpora to achieve a confidence level of 95% according to Sison and Glaz's (1995) method of simultaneous determination of sample size and confidence level. In Chapter 6, the corpus concordances of all of the words that potentially serve a quantifying function are subclassified in order to distinguish their syntactic role. In each chapter, an exposition is given of the analysis categories represented by the set codes in the concordance files.

3.4.3 Part-of-speech tagging

Both corpora were part-of-speech-tagged using the UCREL CLAWS7 Tagset. In spite of it proving to be the most accurate tagger for the TLEC (Van Rooy & Schäfer, 2003), a number of tagging errors still arose due to spelling errors in the TLEC and "hard returns" in LOCNESS. The part-of-speech-tagged versions of the corpora were not used in the frequency or concordance analyses in the present study.

3.4.4 Computer-aided error analysis in the *ICLE* project

As part of the ICLE project, a portion of the TLEC essays (amounting to roughly 40 000 words) were manually corrected by a native English editor and subsequently tagged with the Louvain error tagset (Dagneux et al., 1996). This kind of error analysis allows for the flagging and classification of errors as well as the insertion of corrections. The research aims of the present study are not aligned with the pedagogical aims of error analyses in general since the variety of English represented by the TLEC is not seen as an interlanguage on the way to the target of Standard English (Van Rooy, 2008a:274). However, the error-tagged data is very useful to quantitatively check some of the anecdotal claims that have been made with regard to certain features of BSAfE (e.g. De

Klerk and Gough, 2002). In the current study, the error-tagged data are employed in Chapter 5 to paint a preliminary picture of the differences between the TLEC and Standard English with regard to the use of articles and demonstrative determiners. To this end concordance searches of the words in question (e.g. *a*, *the*, *this*, etc.) marked as correction were conducted. The error-tags themselves were ignored. For example, the search string **\$the\$** would yield all the instances where *the* is the correction or Standard English variant. The concordance lines were then manually coded according to the original forms which were deemed erroneous. The concordance extract below (Figure 3.1) illustrates that cases where *a/an* was used where *the* is expected in Standard English were coded as “A” and cases where a demonstrative determiner was used where *the* is deemed more appropriate in Standard English were coded as “D”.

Figure 3.1 Extract from concordance

N Concordance	Set
27 \$0\$. This young woman has (GA) %% a \$the\$ dream of (FS) %% furnishing \$finishing\$	A
28 are \$he or she is\$ educated and %% these \$the\$ people who are not working %% are	D

In the cases where an article, demonstrative determiner or possessive determiner was inserted where no determiner is expected in Standard English, the search string includes the “erroneous” determiner and the correction, e.g. **the \$0\$**. Such searches imply that the concordance lines need not be sifted further.

3.5 Conclusion

This study is both corpus-driven and corpus-based and entails both quantitative and qualitative methods. The frequency lists (word lists and keyword lists) described in Section 3.4.1 represent the more quantitative and corpus-driven aspects of the study, while the concordance analyses described Sections 3.4.2 and 3.4.4 represent the more qualitative and corpus-based aspects of the study. Additional detail on analysis categories is provided in the analysis chapters.

CHAPTER 4

Grammatical patterns of the general noun *people*

4.1 Introduction

In this chapter the full concordances of the noun *people* will be examined. The aim of this analysis is to compare the specification of referent identity, quality and quantity in the two corpora in terms of the same noun. It is contended that much insight can be gained about nominal specification from the data if the choice of Thing is constant. At the same time this chapter aims to explore the reasons for selecting a general noun such as *people* over a more specific noun.

The noun *people* is of special interest in this study. It is the most frequent noun in both corpora, but at the same time it is one of the top ten positive keywords in the TLEC compared to LOCNESS. Processing of the untagged corpora in Wordsmith yields a count of 2068 instances in the TLEC and 747 instances in LOCNESS.²⁰ If these frequencies are normalized to the number of occurrences in every hundred thousand words in the corpus, the TLEC has a frequency of 1029 and LOCNESS a frequency of 368. The table below indicates the positive keyness of *people* in the TLEC compared to LOCNESS (the reference corpus).

²⁰ The semi-edited wordlists generated from the POS-tagged corpora yield an even greater number of instances of *people* in both corpora: 2102 in the TLEC and 765 in LOCNESS. This discrepancy in frequency figures is largely due to the fact that the apostrophe is allowed as a character in a word by Wordsmith with the effect that the genitive case form *people's* is counted as a separate word. The CLAWS POS-tagger splits the genitive morpheme from its base (*people*) and tags it separately as "Germanic genitive marker" while the base is counted as an instance of *people*. In the case of the TLEC some misspellings further contribute to the greater number of instances of *people* in the wordlist based on the tagged version of the corpus. The slightly higher frequencies for *people* in the wordlists based on the annotated corpora do not contradict the trends that can be observed in the processing of the unannotated corpus data. However, for purposes of further investigation of the usage of the noun *people*, only the correctly-spelled, whole-word occurrences of *people* in the unannotated corpora are considered.

Table 4.1 The noun *people*: frequency and keyness

Raw frequency in TLE	2068
Normalized frequency per 100 000 in TLEC	1029
Raw frequency in LOCNESS	747
Normalized frequency per 100 000 in LOCNESS	368
Rank in keyword list based on loglikelihood	10 th
Keyness based on loglikelihood	662.0828857
P-value for keyness based on loglikelihood	3.22878E-18
Rank in keyword list based on chi-square	9 th
Keyness based on chi-square	636.1116333
P-value for keyness based on chi-square	3.66487E-18

The noun *people* belongs to the special subcategory of noun termed ‘(very) general noun’ in SFG (Halliday & Hasan, 1976:274, Halliday & Matthiessen, 2004:327, Mahlberg, 2005:1). Along with nouns like *thing* and *place*, *people* is situated on the general side of the lexical specificity cline (Halliday & Matthiessen, 2004:327). Due to the very general (schematic in CL terms) meaning of such nouns they can function as discourse anaphora, much like pronouns. For this reason Halliday and Hasan (1976:274) remark that a general noun is on the borderline between lexis and grammar and that the cohesive function of a general noun lies on the borderline between grammatical and lexical cohesion. Mahlberg’s (2005:161, 173) corpus study of general nouns (including *people*) assumes that “[g]eneral nouns are characterized by local textual functions” and that “their world meaning is less central than their textual or evaluative meanings”. Rephrasing this in pure SFG terms, one could postulate that the textual and interpersonal meanings of *people* are more prominent than its experiential meaning. Indeed, the experiential meaning of *people* is rather vague: the word *people* denotes a plural number of human entities (which implies animate and endowed with consciousness).

All instances of the Wordsmith-generated concordances of *people* in the untagged versions of both the corpora were analyzed to determine which functions (experiential, textual or interpersonal) are predominantly realized by this general noun. To explore the functional meanings of the noun *people*, the structures of the nominal groups in which *people* is the Thing element will be examined. With the choice of Thing as a constant

(*people*), much insight is to be gained about the choice of the other elements. The influence of the functions of *people*, as well as the structures of the nominal groups in which it occurs, will also be considered.

4.2 Analysis categories

In order to efficiently classify a total of 2815 occurrences of *people* (in both corpora) in terms of functional meaning, recourse to syntactic form was necessary to establish initial classification categories (“sets” in Wordsmith terminology) (See Table 4.2 for a breakdown of these categories). In the initial form-based classification of the concordances of *people*, the nominal groups in which *people* occurred were sorted according to the presence and/or absence of structural elements that serve to identify the referent and/or to quantify and/or to qualify *people*.

The cases where the noun *people* occurred in a nominal group which served as a complement of the preposition *of* in a partitive construction (e.g. *a number of people*) were classified separately, since these partitions imply a specification of either quantity or quality (Quirk et al., 1985: 249-251). In addition to the quantification specified by the quantifying noun preceding the preposition *of*, or the qualification specified by the species noun (Biber et al., 1999:255), the noun *people* may be modified internally within the nominal group complementing *of* as can be seen in the example below.

(16) Most of the people affected by HIV/AIDS are not educated. <ICLE-TS-NOUN-0140.1>

Another major distinction made in the classification of the concordance lines was between instances that contain no specifiers and instances where the noun *people* is specified further by determiners and/or modifiers. The specified instances were then further classified in terms of the types of specification realized in the nominal group: namely, the identity of the referent and/or quantity and/or quality. This classification was based on the forms that typically realize these specifying functions.

Table 4.2 Analysis categories of *people* concordances

Set description according to only the specifying elements present in the nominal group headed by <i>people</i>	TLEC		LOCNESS	
	Raw freq.	Per 1000	Raw freq.	Per 1000
(no specifiers)	627	303	235	315
Definite determiner only	211	102	120	161
Indefinite determiner with implied quantification only	85	41	37	50
Explicit quantifier only	161	78	92	123
Prenominal quality specification only	198	96	52	70
Postnominal phrase only	96	46	23	31
Postnominal clause only	158	76	30	40
<i>these two people</i>	0	0	1	1
Definite determiner & prenominal quality specification only	48	23	13	17
Definite determiner & postnominal phrase only	30	15	45	60
Definite determiner & postnominal clause only	78	38	15	20
Definite determiner, premodifiers and postmodifiers (no quantifiers)	21	10	6	8
Indefinite determiner/quantifier & prenominal quality only	17	8	11	15
Indefinite determiner/quantifier & postnominal phrase only	39	19	9	12
Indefinite determiner/quantifier & postnominal clause only	41	20	9	12
Indefinite determiner/quantifier, premodifiers and postmodifiers (no quantifiers)	7	3	2	3
<i>such people</i>	6	3	0	0
Premodifiers and postmodifiers (no determiners nor quantifiers)	38	18	8	11
(<i>People</i> -group is complement in qualifying <i>of</i> -construction)	12	6	2	3
∅ quant <i>of people</i> -group <i>People</i> is specified for quality, i.e. has modifiers	37	18	3	4
∅ quant <i>of people</i> -group <i>People</i> has no specifiers	35	17	21	28
Definite quant <i>of people</i> -group <i>People</i> has no specifiers	1	<1	3	4
Definite quantifier <i>of people</i> -group <i>People</i> is specified for quality, i.e. has modifiers	9	4	3	4
∅ quant <i>of people</i> -group <i>People</i> has definite determiner and quality-specifying modifiers	52	25	4	5
∅ quant <i>of people</i> -group <i>People</i> has definite determiner, but no modifiers	42	20	2	3
None of the above	19	9	1	0
Total lines in concordance of <i>people</i>	2068		747	

Demonstrative (*these, those*) and possessive (e.g. *his, our*) determiners, as well as the definite article (*the*), were taken as definite markers. Definite determiners typically signal that the referent is identifiable from the context of situation or from a prior or anticipated occurrence in the text (Halliday & Matthiessen, 2004:314; Quirk et al., 1985:265). The indefinite determiners *any, all* and *some* imply quantification. Determiners like *any, some, enough* and *each* are called 'quantitative determiners' by Quirk et al. (1985:258). Cardinal numbers and the closed-class quantifiers (*many, a few, few, several*) (Quirk et al., 1985:261-263) explicitly specify a quantity.

Qualitative specification is realized through prenominal and/or postnominal modifiers. Prenominal classifying adjectives and classifying nouns serve to refine the denotative meaning of people, and thus fulfill an experiential function. Post-determiners and some postnominal modifiers assist in specifying the referent and thus serve a textual function. Modifiers that are descriptive in nature can be evaluative and thus fulfill an interpersonal function.

The sections below will describe the identifying, qualifying and quantifying constructions associated with the noun people, and in so doing attempt to explain the high keyness of people in the TLEC.

4.3 Unspecified people

The first set of occurrences to consider is the unspecified use of *people*. In purely formal terms, any *people*-group with no determiner or modifier should count as an unspecified use of *people*, even when it is the complement of the preposition *of* in partitive constructions such as *a lot of people*. However, these partitive constructions convey quantity, or imply a certain quality, and will be discussed separately. This section deals with nominal groups in which *people* is both the logical Head and the experiential Thing, without any other specifying elements.

In both corpora just under one third of all the occurrences of *people* are unspecified. In the TLEC 303/1000 (627 instances in the full concordance of 2068) occurrences of people are unspecified. In LOCNESS the number is 314/1000 (235 instances in the full

concordance of 747). Nominal groups with unspecified *people* as head are used **generically** in the assertion of general truths or **non-specifically** in hypothetical and other non-assertive contexts.

(17) People have a lot of needs that they want to be met. <ICLE-TS-NOUN-0321.1>

(18) If our country could be able to create jobs I think this problem could be minimized and people would start to think about better things to do <ICLE-TS-NOUN-0081.1>

Basic sorting of the data in the concordance file reveals the typical syntactic patterns and semantic contexts in which the noun *people* occurs without further specification. In experiential terms, the most frequent patterns that emerge in one-word-left (L1) and one-word-right (R1) sorting of the concordance lines relate to processes of causation, the construal of non-assertive (hypothetical) contexts and processes of informing.

In both corpora, unspecified *people* occurs frequently as the subject of clauses that present a reason for a certain state of affairs²¹ and as complements of verbs denoting causative processes.

(19) Prostitution is the main cause of Aids/Hiv **because** people do not have jobs or money to take care of their families. <ICLE-TS-NOUN-0073.1>

(20) this can **lead** people to commit crime <ICLE-TS-NOUN-0081.1>

(21) Poverty has **made** people want to do things that are risky. <ICLE-TS-WITS-0006.1>

There are a total of 93 constructions of causation involving unspecified *people* in the TLEC compared to the 41 in LOCNESS.²² In terms of registerial field, this suggests that

²¹ In TLEC (total 55x): *because people* 31x; *because of people* x2, *so that people* 11x. In LOCNESS: *because people* 5x.

talking about causes of phenomena often entails talking about people in the most general sense. The prominence of the notion of ‘cause’ in the TLEC essays therefore accounts for some of the difference in the raw frequency of *people* in the two corpora. The token *cause* ranks 39th in the keywords list with a keyness of 157.²³ In their study of causative constructions in BSAfE (based on an earlier version of the TLEC), Van Rooy and Van Huyssteen (2003) note that the verb *cause* is the second most frequent causative *verb* in their corpus after *make*. They point out that the frequent choice of *cause* as a verb is noteworthy in the light of the fact that the verb *cause* is not amongst the seven most frequent verbs in the *Longman Spoken and Written English Corpus* (Biber et al., 1999). With regard to the current study, it should be noted that there is very little difference between the two corpora in the frequency of unspecified *people* as complement of a causative verb: 38x in the TLEC versus 36x in LOCNESS. The difference arises from the high use of *because* and *so that* in the TLEC. The token *because* ranks 11th in the keywords list with a keyness of 560 (based on chi-square).

In the longer text excerpt from the TLEC below, a variety of expressions of causation are illustrated – not only with regard to unspecified *people*, but also other nominal groups headed by *people* and their anaphora.

(22) Our continent is facing a problem of unemployment. We have well educated people but they are not able to get jobs **as** they are not available. So this **led people end up** changing their behaviour **in order to** earn a living.

Most people end up resorting to prostitution as a means of survival. Unintentionally **because of** Poverty and the chances of her being infected will be very high.

Poverty also **make some people** to abandon their characters and disciplines for other people's interest. This is very common in rural

²² In TLEC (total 38x): *result* in people* 4x; *cause* people* 3x; *force people* 2x; *influence people* 1x; *lead* people* 7x; *lead* to people* 5x; *let people* 1x; *lobby people* x2; *make* people* 13x (excluding *make people aware*) [Asterix indicates that number represents verb lemma]

In LOCNESS (total 36x): *allow* people* 2x; *cause* people* 3x; *encourage* people* 2x; *force* people* 3x; *get* people* 3x; *help* people* 5x; *lead* to people* 2x; *let* people* 2x; *make* people* 12x; *stop people* 2x.

²³ Generated in Wordsmith from the untagged versions of the corpora and calculated with chi-square. This token count includes both nouns and verbs. Only the base (i.e. uninflected) forms are included in this count, since the wordlists are not lemmatized.

areas where people tend to undermine themselves and they will **let** to be used by other people who think they are living a luxurious life and **as a result** they **end up** being infected. <ICLE-TS-NOUN-0056.1>

Of note is the use of *end up*²⁴ to express a resulting state (also see Van Rooy, 2008b:349). Double, or blended, expressions of causation can be seen in the third sentence and the last sentence of the excerpt: *So this led **people** end up changing their behaviour [...]* and *as a result **they** end up being affected*. As can be seen from the example above, the use of *end up* is not restricted to unspecified *people*. In addition to the 93 constructions of causation mentioned above, there are 18 instances of unspecified *people* that are the syntactic subjects of the verb *end up* in the TLEC, as well as a further 7 of *people* with specification of identity, quality or quantity. The examples below are further illustrations of *end up* with unspecified *people* as subject. In the last two examples, double marking of causation (before and after the nominal group) can be seen again.

(23) because of lack of money, material things, people **end up** giving their bodies unto men in order to get the money. <ICLE-TS-NOUN-0328.1>

(24) Most of the people are unemployed and these cause poverty **as the results** people **end up** only not committing crime but also sell their bodies as way of getting income even though is unprotected sex <ICLE-TS-KIMC-0382.1>

(25) Unemployment is also a major contributory factor. the number of jobs that are available do not equal the number of graduants that schools produce; **that being the result**, people **end up** rooming the streets with nothing to do. **In the end**, they resort to doing bad things that further worsen the HIV/AIDS virus. <ICLE-TS-NOUN-0286.1>

²⁴ The token *end* ranks 81st on the keywords list and has a keyness of 88. A concordance search yields 395 instances of the various grammatical forms of both the verb lemma and the noun lemma. Of these 113 are nouns. Of the 282 instances that are verbs, 252 are prepositional verbs followed by *up*.

The high keyness of *people* in the TLEC is partially explained by the fact that the noun *people* is frequently used in the context of causation and that ‘cause’ is a prominent field feature in the register of the TLEC. Unspecified *people* also occurs frequently in non-assertive contexts where *people* is interpreted as non-specific (see Table 4.3). The noun *people* occurs as subject of circumstantial or conditional subordinating clauses much more frequently in the TLEC than in LOCNESS (47 times vs 19 times).

(26) If **people** were employed in industries there will be less starvation resulting from poverty. <ICLE-TS-NOUN-0070.1>

(27) If only **people** can be given the skills on how to provide for themselves and stop demanding jobs from the Government. <ICLE-TS-POT-0199.1>

In a multidimensional comparison of the TLEC and LOCNESS, Van Rooy (2008a:279) also reports a higher usage of causative and conditional adverbial subordinators in the TLEC, although LOCNESS makes more frequent use of subordinating structures overall (Van Rooy, 2008a:281).

Table 4.3 *If*-conditional clauses, *wh*-circumstance clauses and modal auxiliaries with unspecified *people*²⁵

Pattern in concordance lines	Raw frequency in TLE	Raw frequency in LOCNESS
<i>if people</i>	16	9
<i>when people</i>	17	8
<i>where people</i>	14	2
<i>people can</i>	37	31
<i>people could</i>	5	2
<i>people may</i>	1	1
<i>people might</i>	1	0
<i>people must</i>	16	2
<i>people should</i>	25	5
<i>people will</i>	31	10
<i>people would</i>	8	5
<i>people have to</i>	4	2
<i>people need to</i>	0	1
<i>people ought to</i>	0	1

²⁵ Negative forms of modals (e.g. *cannot*, *can't*, *won't*) are included in these counts.

Clauses with marked modality are also non-assertive. Unspecified *people* frequently occurs as subject of clauses with marked modality: 128x in the TLEC versus 60x in LOCNESS. The modal auxiliaries construe a hypothetical scenario in which *people* is interpreted as generic or non-specific.

(28) what I am saying is **people** should stop blaming poverty for causing HIV/Aids in Africa. <ICLE-TS-WITS-0003.1>

(29) **People** must be taught how to create jobs for themselves - so that crime would be decreased. <ICLE-TS-NOUN-0102.1>

In the TLEC, but not in LOCNESS, *people* in its unspecified, general meaning, frequently occurs as generalized recipient(s) of information – i.e. as complements of verbs of communication.²⁶ There are 35 such occurrences in the TLEC concordance, 23 of which involve teaching. Unspecified *people* complements the verb lemma *educate* 12 times and the verb lemma *teach* 11 times.

(30) organisations and medical centres which **inform** people about this epidemic. <ICLE-TS-NOUN-0255.1>

(31) they are trying everything to **make** people **aware** of this killer disease. <ICLE-TS-NOUN-0313.1>

These processes can be represented as nominalizations without the overt mention of the very general Recipient(s), viz. *provide information, raise awareness*. Wording these processes as clauses instead of nominalizations necessitates overt clause Participants and has the effect of making the human Recipients more salient. In a study of nominalization in both spoken and written BSAfE (compared to native English), Terblanche (2009:58) reports less abstract nominalizations in written BSAfE (represented by the TLEC) compared to written native English (represented by LOCNESS).

²⁶ In TLEC (total 35): *advise* people* 2x, *alert people* 2x; *educate people* 12x; *inform people* 1x; *make people aware* x3; *teach* people* 11x; *tell* people* 4x.
In LOCNESS: *showing people* 1x.

There are 5 instances in the TLEC where unspecified *people* is used in appositive as-phrases. Three of these occur in consecutive sentences in one essay. In these examples, the general noun *people* is employed to emphasize the humanity of the clausal Participant.

(32) These are the things that **we as people** came across every single day. They show you ways that how **we as people** can contract this disease and how we can prevent it. **We as people** choose to be ignorant. <ICLE-TS-NOUN-0313.1>

(33) AIDS/HIV is not someone else's problem, it is my problem it is your problem. **We as people** have to face this danger because millions of our youth will not rich adult stage. <ICLE-TS-NOUN-0105.1>

(34) Prison people are also our problem because they affect our progresses. **As people we** have to think about prisoners as being criminals or decrease crime rather than labelling them by name of criminals. <ICLE-TS-NOUN-0403.1>

As-phrases with an indefinite nominal group (often with no determiner) as prepositional complement serving to express role or capacity are not unusual and are illustrated by these examples:

(35) If there is no trust amongst our troops, then how can we, as civilians, trust in them? <ICLE-US-MRQ-0013.1>

(36) When teachers, as role models, do not take just stands then students tend to be less reluctant to cheat. <ICLE-US-MRQ-0044.1>

(37) As a BEd student, I have been given a chance to go to the field and actually taste how it feels like to be a teacher. <ICLE-TS-NOUN-0467.1>

The high denotational specificity of the nouns in these nominal groups is noteworthy: *civilians*, *role models*, *BEd student*. However, when role is expressed by the most

general way in which reference to humans can be made (i.e. *as people*) a question about the sense in which *people* is used arises. The use of *we* in this context suggests inclusivity and an activation of the ‘humanity’ meaning of *people*. This interpretation is quite plausible for the examples of *as people* quoted above. A concordance search of *as* in the TLEC yields at least 5 occurrences of *as (a) human (being)s* serving to expressed role or capacity (see example 24). This confirms the ‘humanity’ interpretation of *people*.

(38) As we live in the modern society there a many thing that we do as human beings. <ICLE-TS-KIMC-0361.1> [first sentence of essay].

In this section the comparatively high frequency of the instances of unspecified *people* in the TLEC has been explained in terms of their use in causative constructions, in non-assertive contexts, in clausal constructions of processes with overt Recipients and in appositive *as*-phrases. In many of these constructions, the human Participant is made salient.

Without further specification, the noun *people* realizes experiential meaning – albeit at a very general level – and does not really fulfill a text-cohesive function. The most likely occurrences of *people*-headed nominal groups serving a textual function would be occurrences containing a definite determiner. These will be examined in the next section.

4.4 Definite *people*

The possessive and demonstrative determiners specify the identity of the referent, whereas the definite article *the* indicates that the identity of the referent is recoverable from elsewhere in the discourse itself or from the larger communicative context. Since the corpora consist of written essays it is expected that where referents (of the noun *people*) are regarded as identifiable, the identification is likely to be based on textual rather than contextual information. It should be kept in mind that while the definite article is typically associated with specific reference, it may also occur in nominal groups with generic or non-specific interpretations (see Chapters 2 and 5). Where definite

determiners are accompanied by further specifications of quantity and/or quality, such additional specification might contribute to the identification of the referent. For this reason nominal groups with definite determiners but no modifiers are considered separately from definite nominal groups with additional specifications of quantity and/or quality. For purposes of analysis, the presence or absence of modifiers was used as an indication of quality specification. The few instances where the predeterminer *all* preceded the definite determiner were not counted here as separate specifications of quantity. In order to get a full picture of the distribution of *people* with a definite determiner, the frequencies for definite nominal groups that function as heads of partitive *of*-constructions are also reported.

Tables 4.4 and 4.5 provide an overview of the distribution of definite nominal groups in the *people* concordances. In this table the various patterns of definite-marked *people* as well as the choice of type of definite determiner can be seen.

Table 4.4 Frequencies of patterns of definite *people*-groups in the TLEC and LOCNESS

	TLEC		LOCNESS	
	Raw freq.	per 1000	Raw freq.	per 1000
<i>def people</i>	211	102	120	161
<i>these two people</i>	0	0	1	1
SUBTOTAL:	211	102	121	162
<i>def qual people</i>	48	23	13	17
<i>def people qual:group/phrase</i>	31	15	45	60
<i>def people qual:clause</i>	78	38	15	20
<i>def qual people qual</i>	21	10	6	8
SUBTOTAL:	178	86	79	106
<i>quant of def people-group people-group has modifiers</i>	52	25	4	5
<i>quant of def people</i>	42	20	2	3
SUBTOTAL	94	45	6	8
<i>Other definite people +quality</i>	1	--	0	--
<i>Other definite people -quality</i>	9	--	0	--
SUBTOTAL	10	5	0	0
TOTAL definite people per corpus	493	241	206	276

Table 4.5 Realizations of definite determiners in *people*-groups in the TLEC and LOCNESS

Type of determiner	Definite article		Demonstrative determiners				Possessive determiners									
	T	L	T		L		TLEC							LOCNESS		
Corpus	T	L	T	L	T	L	my	our	your	her	its	their	such	its	his	our
Definite determiners →	the	the	these	those	these	those										
Pattern in people concordance (concordance set) ↓																
<i>def people</i>	69	79	79	20	25	2	1	21	2	2	9	8	0	2	12	0
<i>these two people</i>	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
SUBTOTAL:	69	79	79	20	26	2	1	21	2	2	9	8	0	2	12	0
<i>def qual people</i>	32	10	6	0	0	1	0	6	0	0	2	2	0	0	0	2
<i>def people qual:group/phrase</i>	23	45	3	3	0	0	0	2	0	0	0	0	0	0	0	0
<i>def people qual:clause</i>	49	14	2	26	0	1	0	1	0	0	0	0	0	0	0	0
<i>def qual people qual</i>	20	5	0	1	1	0	0	0	0	0	0	0	0	0	0	0
SUBTOTAL:	124	74	11	30	1	2	0	9	0	0	2	2	0	0	0	2
<i>quant of def people-group people-group has modifiers</i>	42	2	4	1	1	1	0	5	0	0	0	0	0	0	0	0
<i>quant of def people</i>	29	1	6	1	1	0	0	4	0	0	2	0	0	0	0	0
SUBTOTAL	71	3	10	2	2	1	0	9	0	0	2	0	0	0	0	0
SUBTOTAL of other patterns	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL occurrences per definite determiner per corpus	274	156	100	52	29	5	1	39	2	2	13	10	0	2	12	2
TOTAL occurrences per type of definite determiner per corpus	274	156	152		34		67							16		
Frequency of definite determiner as percentage of all definite determiners of <i>people</i> per corpus	56	76	31		17		14							8		

In terms of frequency, the definite *people*-groups are considered from three perspectives: firstly, the raw frequencies; secondly, as a number of occurrences per 1000 concordance lines; and thirdly, as a percentage of a certain pattern. The first contributes to explaining the high keyness of the token *people* in the TLEC compared to LOCNESS. The second provides an indication of the portion of instances of *people* in each corpus specified for identity. The third indicates the relative frequency of the patterns used in each corpus, particularly with regard to the choice of definite determiner, in the concordances of *people*.

In terms of raw frequency, the number of instances of definite *people*-groups in the TLEC are more than double the number in LOCNESS: 493 vs 206. Discovering the reason for the greater use of definite determiners with *people* in the TLEC will further add to the explanation of the high keyness of *people* in the TLEC. The most typical discursive function of a nominal group with a definite determiner and headed by a general noun is that of establishing an anaphoric chain.

If one normalizes the frequency of definite nominal groups to the number of occurrences per 1000 concordance lines in each concordance of *people*, the difference between the two corpora is less pronounced: 238/1000 in the TLEC vs 276/1000 in LOCNESS. In fact, this normalized frequency indicates that once the noun *people* is selected, it is marked as definite slightly more frequently in LOCNESS than in the TLEC.

The two corpora differ widely in respect of the choice of definite determiner with *people*. The definite article is selected 55.5% (274x) of the time in the TLEC and 75.7% (156x) of the time in LOCNESS. The obvious preference for the definite article in LOCNESS is balanced by a higher usage of demonstrative and possessive determiners in the TLEC: 31% (152x) vs 16.5% (34x) and 13.5% (67x) vs 7.8% (16x), respectively. The high raw frequencies of demonstrative and possessive determiners in the TLEC will be probed below for a possible explanation of the keyness of *people* in the TLEC. The question that arises is whether the nominal groups with demonstrative and possessive determiners fulfill a purely text-cohesive function.

The remarkably high frequency of definite nominal groups that function as complements of quantifying partitive *of*-constructions in the TLEC (94x vs the 6 instances in LOCNESS) will be discussed in detail in Sections 4.4.3 and 4.5 which deal with quantification. Definite nominal groups without quantity or quality specification will be discussed in Section 4.4.1, and definite nominal groups with quality specification will be considered in Section 4.4.2.

4.4.1 Instances of *people* marked as definite without additional quality specification in the nominal group

In this section, definite *people*-headed nominal groups without group-internal specification of quality will be considered, i.e. the pattern **def *people***, e.g. *the people*, *those people*, *our people*, etc. Instances of definite *people* with neither quality, nor any quantity specification, number 209 (101/1000) in the TLEC and 120 (161/1000) in LOCNESS concordances of *people*. However, the two instances in the TLEC where the predeterminer *all* precedes the definite article to specify quantity (*all the people*)²⁷ and the single instance in LOCNESS where the demonstrative determiner is followed by a cardinal numeral (*these two people*)²⁸ have been added to this category in Table 4.4 (211x in the TLEC and 162x in LOCNESS).

In the TLEC only 33% (69/211) of the instances in this category have *the* as determiner, whereas in LOCNESS 65% (79/121) of the instances in this category have *the* as determiner. These percentages (33% vs 65%) indicate a clear preference for the definite article in LOCNESS within this category. In terms of raw instances, there are actually 10 more instances with *the* as determiner in LOCNESS, which do not help to explain the positive keyness of *people* in the TLEC. For that, the demonstrative and possessive determiners need to be considered. The choice of definite determiner in this category is much more evenly spread in the TLEC than in LOCNESS. In this category, demonstrative determiners are selected in 47% (99x) of the instances in the TLEC and in 23% (28x) of the instances in LOCNESS, while possessive determiners are selected 20% (43x) of the time in the TLEC and 12% (14x) in LOCNESS.

²⁷ <ICLE-TS-NOUN-0172.1>, <ICLE-TS-NOUN-0059.1>

²⁸ <ICLE-US-MRQ-0046>

The choice of definite determiner needs to be interpreted in terms of referent identification. Demonstrative and possessive determiners tell the reader ‘which people’ (*these people, their people*) and thus identify the referent deictically or anaphorically, whereas the definite article merely indicates that the reader is expected to be able to identify the referent based on prior knowledge or information that is recoverable from the communicative context or from elsewhere in the text. The meaning of the head-noun *people* also plays a role in the choice of determiner as will be shown below.

Where *people* occurs with a demonstrative determiner, the nominal group serves a text-cohesive function. The meaning of *people* is restricted to a subclass of people defined earlier in the text, for instance HIV-positive people, as illustrated in the example below.²⁹

- (39) HIV can be eliminated by educating people all over the country. There must be groups of people who are victims of HIV to talk to the community members and tell them how hard it is to accept that you are HIV positive may people can understand the situation in which those people are in <ICLE-TS-NOUN-0081.1>.

Nearly half of the nominal groups with the pattern **def people** in the TLEC are determined by demonstratives. In cases where *people* functions as an anaphor for a specific subclass of people delineated earlier in the text, the TLEC writers preferred to explicitly point the reader to the antecedent by using a demonstrative determiner. It is interesting to note that the sub-classes of people represented by the antecedent is often related to the essay topics of the TLEC: people who provide sex to meet their financial needs (13x), people who give money or gifts in return for sex (4x), people with HIV/Aids (3x), soccer players (5x), criminals and prisoners (38x), the poor or unemployed (10x), people who are illiterate, uninformed or uneducated (6x), people from rural areas (5x), leaders in society (2x). In the TLEC concordance *these/those people* are defined in terms of antecedent information about their health, economic or social status, their societal roles or geographic provenance.

²⁹ In examples in this section (4.4.1), solid underlining is used for *people* groups with a demonstrative determiner and broken underlining is used for other *people*-groups. Nominal groups that are antecedents of *people*-groups are indicated with italics.

The TLEC excerpt below illustrates the text-cohesive function of the demonstrative determiners. Two subclasses of people are referenced anaphorically by *these people*.

(40) The government is also responsible for *prisoners who got this disease in prison* because over-crowding and separation from their families. Yet these people are doing this, not because of poverty and hunger but because of loneliness and desire for sex.

Another group is of *rich or people who earn a better living*; having big houses, luxury cars, but they still have aids why? These people were firstly greedy, selfish and playing big in a small town. <ICLE-TS-POT-0190.1>

The use of the full nominal group *these people* in the essay above, instead of the pronoun *they* as anaphor emphasizes the fact that the antecedent defines a subclass of people. In the excerpt below, *these people* is used three times as anaphor of the nominal group *the poor or unprivileged people*, but is alternated with anaphoric *they* (boldface). One might argue that in each case, *they* is actually the anaphor of *these people*, which in turn is the anaphor of *the poor or unprivileged people*, but each instance of *they* and *these people* are co-referential with each other and with the class-defining nominal group *the poor or unprivileged people*.

(41) [...] It infects both young and old, rich or poor, but people who seem to be more affected by this epidemic are the poor or unprivileged people. The reasons why these people end up being victims is because *they* are deprived of a lot of resources one most important resource being education. These people don't know much or in most cases anything about HIV or AIDS and **they** are more likely to get infected because you can't exactly take precautions for something you know nothing about. Once these people are infected and do find out, **they** are still faced with the problem of acquiring money for the necessary medication. [...]<ICLE-TS-NOUN-0094.1>

As anaphor, *these people* is more specific and less general, or schematic, than *they*. It seems as though the writer deems it necessary, or helpful to the reader, to emphasize

that he/she is referring to a predefined subclass of people intermittently in this long anaphoric chain, even though there is no chance of an ambiguous interpretation of *they*, since there are no other nominal groups which might be referents or antecedents of *they*. The use of a demonstrative determiner and the general noun *people* to serve as anaphor, instead of a personal pronoun, goes some way in explaining the keyness of *people* in the TLE. It accounts for a potential 71 instances of the noun *people*.

Of the 79 proximal demonstrative determiners (*these*) in the *people*-concordance of the TLEC, 16 take the form *this*. This is not necessarily an indication of construal as 'singular' or even 'mass'. In (mesolectal) BSAfE the [i] and [ɪ] vowels tend to converge to [i] with variable length (Van Rooy & Van Huyssteen, 2000:20-22; Van Rooy, 2004:945; Mesthrie, 2005:136). The example below, in which both forms (*this* and *these*) are used before a plural countable noun, illustrates the difficulty a data-analyst has in inferring a specific writer's construal of a thing with regard to countability. Nominal groups that have a direct or indirect anaphoric or cataphoric link to *these people* are italicized.

(42) Personally I partly agree with the above statement when it says "the *prison* system is outdated. The way the police treat *prisoners* is uncalledfull. The initial establishment of *this institutions*, I believe was to correct *these people*, that is why it is called correctional services. Hence they seem to be ignoring the fact that *these people* (*prisoners*) are supposed to get out of prison, and go back to the society. <ICLE-TS-NOUN-0009.1>³⁰

The single instance of the singular form of the distal demonstrative *that* before people cannot be explained in terms of phonology.

(43) They <there> are differend way that can be used to punish rather than to put criminals in prison. They must use social workers, pastors and their friends to counsil them when you tell people for what they had done directly, it is easy for *that people* to repent. <ICLE-TS-NOUN-0168.1>

³⁰ Note that in the last instance of *these people* in this example, the identifying information is restated in an appositive nominal group.

The confusion of singular and plural forms of the demonstratives cannot be directly attributed to transfer, since the demonstratives in the substrate language, Setswana, inflect to concord with the number features of the noun (Krüger, 2006:132-134; Cole, 1955:130).

Although the usage of demonstrative determiners in the LOCNESS concordance category is lower, it is functionally similar to that of the TLEC in that the demonstratives serve a text-cohesive function by pointing to a specific class of *people* defined earlier in the text, but there is a slight difference in the type of retrievable quality in terms of which the subclass of *people* is to be identified. The antecedent information in terms of which the 27 instances of *this/these people* in LOCNESS are identified include ideological affiliation (11x), geographic provenance (4x), health status (4x), social status or role (4x), behaviour towards others (2x, namely *accident stagers*, *mercy killings*) and events (2x, namely *crime*, *earth-quake*). Note the many instances in LOCNESS where *these people* refer to a group of people who hold a certain belief.

Where the noun *people* is collocated with a possessive determiner, it has the meaning of 'nation' or 'populace'. The possessive determiner serves as anaphor of an antecedent person or place in terms of which this 'population' is defined. The possessive determiner is used in 20% of the cases of the pattern **def people** in the TLEC. In the LOCNESS concordance of *people*, possessive determiners are used in only 12% of the instances of **def people**.

A closer look reveals that in all 14 occurrences of **poss people**³¹ in LOCNESS singular third-person possessives are used (*his*=12³² and *its*=2). In the TLEC only a quarter of the occurrences of **poss people** (11/43) have a singular third-person possessive determiner (*its*=9 and *her*=2). Here *her* refers to Africa and *its* to either Africa or a country. In these concordance lines the subclass of *people* is identified anaphorically in terms of geographic provenance. There are also 8 occurrences of the plural third-person possessive determiner *their*. About half of the occurrences of **poss people** in the TLEC (22/43) contain first-person possessive determiners, the majority of which is plural (*our*=21, *my*=1). There are also two occurrences of *your people* in the TLEC. The

³¹ **poss people** stands for the pattern "possessive determiner *people*".

³² In all 12 of the occurrences with *people* in the LOCNESS concordance category, *his* refers to Caligula, a character in a work by Camus.

high frequency of first person pronouns suggests a greater emphasis on interpersonal meanings in the TLEC. In a multi-feature comparison of the TLEC and LOCNESS with regard to involvement, Van Rooy and Terblanche (2006:178) find that the TLEC writers use more personal pronouns and generally assume more shared background between writer and reader than the L1 writers. The preference for first and second person possessive determiners in the TLEC *people*-concordance is in line with the high usage of first and second person pronouns in the corpus. There are more than twice as many first person pronouns in the TLEC than in LOCNESS and more than four times as many second person pronouns (Van Rooy & Terblanche, 2006:169). The use of first and second person pronouns to express the writer's solidarity with the people he/she is writing about is prevalent in the TLEC data, but not in the LOCNESS data.

Where the noun *people* is collocated with the definite article, the 'populace' meaning of people is activated in all the instances of the pattern **def people** in LOCNESS and in the majority of instances in the TLEC – all the exceptions will be discussed below. In particular, the extended meaning of 'populace' as 'members of a certain society' is activated. The identity of the society is typically recoverable, directly or indirectly, from the text. The examples below illustrate the societal meaning of *people* when it is determined by the definite article, as well as the information unit from which the identity of *the people* is recoverable.

(44) *This is a democratic county and the people have rights.* <ICLE-TS-KIMC-0374.1>

(45) *The Government had to make jobs available for the people so that it must lessen their stress.* <ICLE-TS-POT-0198.1>

The next example is a notable exception to the use of *the people* to express the societal meaning of *people*, but it clearly illustrates how the identity of *the people* is recovered later in the text, i.e. cataphorically. Here *the people* refers to potential speakers who might address the youth.

(46) *The way the information comes across is also to blame. You have to choose your words as well as the people when speaking to teenagers*

with raging hormones. The moment you bring in some old woman or man instructing us on what to do with our sexlives we are obviously going to rebel and do exactly the opposite of what we are told to do. <ICLE-TS-NOUN-0093.1>

In the example below *the people* refers to 'people who have been or could be sent to prison'. One might argue that the 'populace' meaning is activated if *the people* is interpreted as the population of a prison.

(47) If the government changes the prison from being hostel to a hotel the people will be relaxed and not find out the reason to be outside while you can get anything inside without working for it. <ICLE-TS-KIMC-0374.1>

In the next pair of examples *people* has a societal meaning, but it would appear that only a subset of the society is meant, namely 'those who turn to prostitution' and 'those who sleep around', respectively. The unmodified societal use of *the people* in these examples might be interpreted as indicative of the perceived largeness (i.e. generality) of the subset in question.

(48) Lack of food made **the people** to be prostitute. <ICLE-TS-POT-0202.1>

(49) The cause that make **the people** to sleep with everyone is that there's nothing that people can entertain themselves with rather than in the urban areas. <ICLE-TS-NOUN-0301.1>

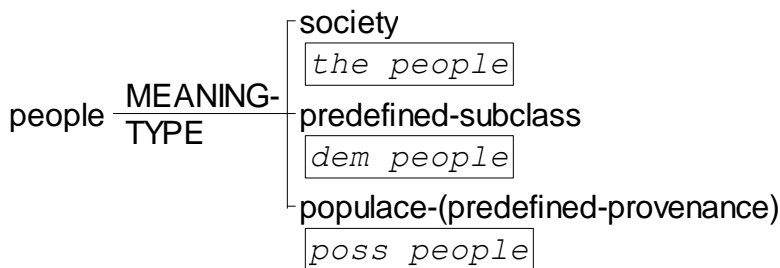
In the next example *the people* occurs in an appositional *as*-phrase. Here we have double apposition, first by means of the *as*-phrase and then by the copy pronoun *we*.

(50) This spreading of HIV/AIDS we as **the people** we take it for granted <ICLE-TS-NOUN-0301.1>

The differences between the corpora pertaining to the choice of definite determiner, which can be observed in the frequencies reported in Table 4.4, can be explained by the

intended meaning of *people*. In this analysis category (i.e. the pattern **def *people***), the demonstrative determiners are not in systemic competition with the definite article if, as I contend, the denotative meaning of *people* is preselected. Where *people* carries the meaning of a clearly defined sub-class of *people*, a demonstrative determiner is selected. The definite article is typically selected when the societal meaning of *people* is intended. When *people* has the meaning of ‘populace’ and where the population in question is identified by prior mention of a geographical place or a person (typically a leader) (i.e. in terms of provenance), the appropriate possessive determiner is selected. These selections are observed in both corpora.

Figure 4.1 Semantic types of definite *people*



The real difference between the corpora in this analysis category is not so much in the choice of definite determiner, but in the meanings intended for the noun *people*. In other words, the difference does not arise from the realization of the text-cohesive function, but in the realization of the experiential meaning of the noun. One could go as far as saying (in the case of this concordance category) that the field affects the choice of definite determiner.

4.4.2 Definite *people* with quality specification

In this section nominal groups that are specified both in terms of definiteness and quality are considered. This supercategory comprises the patterns where quality is specified only prenominally (**def qual *people***), where quality is specified only postnominally by a clause (**def *people* qual:clause**) or a phrase (**def *people* qual:phrase**) as well as

cases where quality is specified both pre- and postnominally (**def qual people qual**). This supercategory excludes definite, quality-specified *people* in *of*-partitions. These will be discussed in Section 4.4.3 below.

In both corpora the definite article is the most frequent determiner in this supercategory (124 in the TLEC and 74 in LOCNESS), but it is used proportionally less in the TLEC (69%) than in LOCNESS (94%). Again, a proportionally higher usage of demonstrative determiners can be observed in the TLEC in this supercategory (23% vs 4%). There are only 13 instances in the TLEC and 2 in LOCNESS where a possessive determiner occurs in a quality-specified *people*-group.

The definite article is by far the most frequent determiner in definite *people*-groups with both pre- and postnominal quality specification in both of corpora: 20/21 in the TLEC and 5/6 in LOCNESS. In both corpora the use of the definite article is generally attributable to the fact that identifying information is given in the postmodifier. In the rest of this section, the choice of definite determiner in nominal groups with prenominal quality specification only (see 4.4.2.1) and postnominal quality specification only (see 4.4.2.2) will be considered. The semantic types of quality realized by the modifiers will be discussed in Section 4.5.

4.4.2.1 *Definite people-groups where quality is specified prenominally only*

The first concordance category under consideration is definite nominal groups with prenominal quality-specification only, i.e. the pattern **def qual people**. In the *people*-concordance there are 13 instances (17/1000) of this pattern in LOCNESS and 48 (23/1000) in the TLEC.

Ten of the 13 instances in the LOCNESS data (77%) have the definite article (*the*) as determiner. In nine of these instances, the prenominal adjective is a classifier denoting provenance, e.g. *the American people* (6x), *the Russian people* (1x), *the French people* (2x). Here the definite article indicates that the referent is identifiable and the identifying information is given by the classifying adjective. In the remaining instance, the premodifier indicates economic status, *the lower-income people* (1x).

In the TLEC, 67% of the instances of the pattern **def qual people** are determined by the definite article. Two of these 32 instances have the relative adjective *other* as prenominal modifier. A relative adjective does not denote a quality, but evokes another entity in relation to which the Thing is specified. In the examples below reference is made to ‘people other than the Batswana’ and to ‘people other than the wealthy’ respectively.

(51) They are also interested in our culture, *we the Batswana* we are very proud of our culture and heritage because we are still practising it in many villages and we have built Museums to show those coming behind how we used to practise our culture and we expect them to follow our foot steps. Mmabana Cultural Centres are also there to teach the other people the tswana dances and practice the values and norms <ICLE-TS-NOUN-0035.1>

(52) Most of the people who are affected by this epidemic are *the wealthy man*. then they spread to the other people. <ICLE-TS-NOUN-0128.1>

In these examples the use of the definite article could suggest reference to an identifiable subclass of people. However, in standard English, the definite article would not be used in this context. The class of people in question is very general and not clearly defined. Its interpretation can only be very generic, meaning ‘others’. As will be seen from the rest of the discussion, the use of the definite article in *people*-groups with very generic interpretations in the TLEC is not restricted to these two examples above.

Of the 32 instances of the pattern **the qual people**, 12 have an adjective that classifies *people* with regard to geographic provenance or ethnic identity: *the rural people* (2x), *the local people* (1x) *the African people* (3x), *the South African people* (1x), *the Batswana people* (2x), *the Baralong (Boo-Ratshidi) people*, *the Black people* (1x). In these examples, the adjectives evoke the ‘nation’ sense of *people* with its connotations of ‘solidarity’. Construal as a single unified group explains the choice of definite article in these cases. The adjectives *rural* and *local* also evoke the notion of a unique class of

people; viz. *rural vs urban, local vs foreign*. There is one instance of classification in terms of social class, *the married people*.

- (53) The sex workers they don't care about what is going to happen to them at last, they just want some money even the married people they can do so because of poverty. <ICLE-TS-NOUN-0322.1>

Though it is true that people can be either 'married' or 'not married', this social distinction is not as clear-cut as the provenance distinction between 'local' and 'foreign'. For one thing, the 'local/foreign' distinction can pertain to all people, whereas the 'married/not married' distinction is really only relevant to people old enough to marry. In standard varieties of English, the nominal group would not be marked as definite in co-texts and contexts similar to that of the example above, probably on strength of the generic or non-specific interpretation of such nominal groups (indicated with boldface in the LOCNESS examples below).

- (54) Before getting married she and her fiance more than likely discussed what they planned to achieve as **a married couple** <ICLE-US-SCU-0014.2>

- (55) If my cousin did have the child, it would have caused problems because she was pregnant for **a twenty-two year old married man**. <ICLE-US-SCU-0008.4>

There is one instance where the noun *man* is used to classify *people* in terms of gender (example below). The hyphen is an indication of the morphological process of compounding, or even of the use of *man-* as a gender prefix. The definite nominal group *the man-people* can be substituted for indefinite nominal groups, e.g. *males* or *men*, in a Standard English reformulation.

- (56) Mental prisons created by the man-people against us women is all physical. <ICLE-TS-NOUN-0425.1>

In the majority of the instances, *the* is followed by a descriptive adjective. The choice of the definite article in the nominal groups in the set of examples below can be attributed to the fact that the discourse (the essay topic and all that is associated with it) and the grammar (by means of identifying clauses and the superlative) provide a referential context within which the referent of the nominal group can be identified (indirectly) with the additional help of the descriptive information conveyed by the adjective.

(57) I have this perception that the illiterate people are more dangerous in our community more especially in the spread of misconduct education. <ICLE-TS-NOUN-0128.1>

(58) But also the older people will be the best. <ICLE-TS-NOUN-0102.1>

(59) the very people to target are the poverty striven people because they need money. <ICLE-TS-NOUN-0160.1>

(60) but people who seem to be more affected by this epidemic are the poor or unprevilledged people. <ICLE-TS-NOUN-0094.1>

(61) In some parts of the world the most talented people are kept behind with the aim of rehabilitating them. <ICLE-TS-NOUN-0274.1>

For the remainder of the instances of ***the qual:descriptor people*** in the TLEC, the use of the definite article cannot be explained in terms of an anaphoric chain, an identifying process, or contextual uniqueness. These need to be considered more carefully in terms of their Standard English agnates, especially since similar constructions do not occur in the LOCNESS data.

(62) Let the concerned people not deny student their chances of being capable and knowledgeable of performing their duties satisfactorily. <ICLE-TS-NOUN-0489.1>

The example above illustrates prenominal placement of the kind of qualifying information that would occur after the noun in more standard varieties, cf. *the people concerned*.

The same might be said of the two examples below if it were clear that *affected* is used in its usual sense. However, it is possible that *affected* and *infected* are used interchangeably. This possibility will be considered in Section 4.5 below.

(63) poverty can also cause HIV/AIDS epidemic in Africa. Due to the fact that the affected people are poor it is clear that their condition of living is also poor. <ICLE-TS-NOUN-0149.1>

(64) The Aids virus brings shame, nightmares, hate, loneliness was it attacks someone. Everyone would never want to glance back again to the infected person even though they could be their relatives, mother, brother, sister or father. The affected people are it not treated like any other humanbeings even if they deserve the right to do so. <ICLE-TS-NOUN-0265.1> [from the beginning of the essay]

In the examples below, the zero article would be preferred in standard English. In these examples reference is made to subclasses of *people* – subclasses which are defined in terms of an objective quality, such as age (*young*), or economic status (*poor, needy, wealthy*).

(65) I think if all the people can work and earn enough money at the end of the month. there will be no depression causing the young people becoming prostitutes <ICLE-TS-NOUN-0172.1>

(66) The young people are gone. <ICLE-TS-NOUN-0120.1>

(67) The goverment have tried so many times to warn the people about this disease, but this goes on because the needy people have no money to buy food. <ICLE-TS-NOUN-0186.1>

(68) Without resouces been taken and be given to the poor people, we are going to die because of AIDS. <ICLE-TS-NOUN-0133.1>

(69) Due to poverty and starvation most inhabitants of Africa are engage in sexual intercourse with the wealthy people so as to enable them to survive accordingly<ICLE-TS-NOUN-0128.1> [from the beginning of the essay]

(70) Some end up jobless and the poor people remain poorer while *the rich* become richer. <ICLE-TS-NOUN-0221.1>

The descriptors in these examples (above) have opposites which allow for *people* as a general class to be divided into **one** of two classes (albeit very big classes) of humanity in general, e.g. 'young/old' and 'poor/rich'. The construal of a unique class might explain the use of *the* in the examples above. However, in spite of the identifiability of the subclass of people, the indefinite agnates would be more acceptable in standard English because of the strength of the class meaning (i.e. a generic interpretation) of these nominal groups and the perceived largeness of the class. Where a non-specific or generic meaning is attached to a plural noun, the zero article is selected in standard English.³³ The following examples are further evidence of the use of the definite article with quality-specified *people*-groups in the TLEC with generic meanings.

(71) It means that the government will have to budget for that disabled person as he will be qualifying for all the benefits that are received by the disabled people. <ICLE-TS-NOUN-0508.1>

(72) We know that nowadays it is very difficult for a person to get a job so when these people are trained they will create employment for the helpless people and thereby earning a living. <ICLE-TS-NOUN-0397.1>

More examples of plural definite nominal groups with generic interpretations need to be considered before generalizations about determiner choice and article usage in BSAfE can be made, but an investigation of the most frequent noun in the corpus suggests that there is a systematic difference between BSAfE and Standard English. One possible explanation, based on the examples above, is that with plural nouns, the generic interpretation carries more weight than the identification of a subclass in the choice of

³³ Compare: *Tigers are scary* and *The tigers are scary*.

determiner in Standard English than in BSAfE. From the perspective of BSAfE this would entail that the identifiability of a sub-class is enough to warrant the selection of the definite article, even if the subclass in question is sufficiently large to be interpreted as generic. In other words, the criterium of 'identifiability' for the selection of the definite article might be overgeneralized in BSAfE. The use of the pattern **the qual people** (with generic meaning) should also be considered in the wider context of determiner usage in BSAfE, specifically with regard to the generic and ascriptive uses of the definite article (see Chapter 5) and the insertion of indefinite determiners (such as *some*) before plural nouns (see Chapter 6).

It is possible that the pattern **the ADJ people** is an extension of the pattern **the NADJ**.³⁴ Where a subclass of people is sub-classified in terms of a certain quality, it is also possible, in some restricted cases, to omit the general noun and nominalize the adjective. Such class-denoting, adjective-headed nominal groups require a definite article, e.g. *the poor*, *the needy*, *the disabled*. The use of **the ADJ people** instead of **the NADJ** in the examples below can be attributed to undeletion.

(73) Without resouces been taken and be given to the poor people, we are going to die because of AIDS. <ICLE-TS-NOUN-0133.1>

(74) ... because the needy people have no money to buy food. <ICLE-TS-NOUN-0186.1>

(75) ... for all the benefits that are received by the disabled people. <ICLE-TS-NOUN-0508.1>

Undeletion is a kind of anti-deletion in which the following principle is at work: "If a grammatical feature can be deleted in [Standard English], it can be undeleted in [Black South African English] mesolect" (Mesthrie, 2006:129). Mesthrie (2006:129) points out that such undeletions are not mandatory and adds the following corollary to the principle of undeletion: "If a grammatical feature can be deleted in StE, it can also be (variably) deleted in BISAfE mesolect, at a lower rate of frequency."

³⁴ ADJ = adjective; NADJ = nominalized adjective

One might postulate that the TLEC writers are choosing this fused-head (nominalized adjective) construction without deleting the noun. Mesthrie (2006) does not list the undeletion of the head noun in fused-head constructions, because it probably does not occur in his data. However, he does list “adjective as substantive” as evidence of deletion (the opposite of undeletion) in BSAfE illustrating this from his data with *the rurals* for ‘rural people’ (Mesthrie, 2006:137). In the TLEC there are no instances of *the rurals*, but two instances of *the rural and the poor* and three instances of *the rural ones*. There are also many more examples of the pattern **the adjective** where the adjective is a nominalization denoting a group of people in the TLEC.³⁵ This is evidence that adjective-headed references to classes of people do occur in BSAfE as in Standard English. Mesthrie’s examples (2006:137) (and *the rural* in the TLE) suggest that this process, which is a kind of deletion, is applied more widely in BSAfE than in Standard English. However, examples such as *the rural ones* and *the needy one*, along with examples such as *the poor people*, from the TLEC are also suggestive of undeletion of the original head element. This illustrates Mesthrie’s point (2006:129, 142), namely that deletion is variable, not mandatory. In other words, evidence of deletion of a certain syntactic element does not preclude a postulation that such an element is also undeleted in other cases. Given Mesthrie’s (2006) evidence that anti-deletion (which includes undeletion) characterizes the grammar of BSAfE, it is postulated here that some of the instances of the pattern **the qual people** is indeed attributable to the principle of undeletion. The use of *one* as dummy head³⁶ in the TLEC data also supports the undeletion-thesis.

4.4.2.2 Definite people-groups where quality is specified postnominally only

Concordance lines of *people* marked for definiteness with postnominal quality specification only (**def people qual**) were further classified in terms of the structure realizing the postnominal quality. A distinction was made between phrasal and clausal postmodifiers (**def people qual:phrase** vs **def people qual:clause**). There are fewer instances of phrasal postmodifiers in the TLEC than in LOCNESS (15/1000 vs 60/1000).

³⁵Further examples from a concordance of *the* in the TLE: *the poor* (13x), *the whites* (9x), *the needy* (2x), *the unemployed in all races*, *the greats*, *the havenots*, *the haves*, *the hungry and poor*. Also see Mesthrie, 2004:970.

³⁶*the needy one*, *the unemployed ones* (TLE)

On the other hand, definite nominal groups qualified only by a postnominal clause are relatively more frequent in the TLEC than in LOCNESS (38/1000 vs 20/1000). The raw frequencies show the discrepancy between the corpora more clearly: 78 vs 15.

In the concordance data of both corpora, the postnominal phrases in definite nominal groups specify geographic provenance in the majority of cases, e.g. *the people of Argos*, *the people in Africa*. This is the case in 84% (38/45) of the instances of the pattern **def people qual:phrase** in the LOCNESS data and in 77% (24/31) of the instances in the TLEC data. In these cases reference is made to a specific population of people identifiable in terms of some location.

It is also interesting to note the three instances of appositive nominal groups in the TLEC data. Appositive nominal groups are regarded as non-clausal qualifiers, i.e. as word groups, and on this basis, as well as the presence of a definite determiner, these examples are classified as a subcategory of the pattern **def people qual:phrase**. There are no such examples in the LOCNESS data. In the first two examples below, *these people* serves as an anaphor of a previously defined sub-class of *people*. However, the writers also add a more explicit description of the sub-class in question in the appositive nominal group.

it is obvious that these people especially the one on the Rural areas they dont know that a person must stick to one partner <ICLE-TS-NOUN-0127.1>

This can lead to H.I.V/AIDS because these people (sugar daddies and mummies) also have their partners and most of them are married. <ICLE-TS-NOUN-0160.1>

Because of no job, low education, lack of facility the people more over the youth are comited in useless things <ICLE-TS-NOUN-0156.1>

The types of qualities conveyed by postnominal phrases will be returned to in Section 4.5 below. With regard to the choice of determiner, it is interesting to note that only the definite article is employed in the LOCNESS data for the **def people qual:phrase** pattern. The definite article signals that the referent is identifiable. This suggests that the identifying information is given within the nominal group (i.e. esphorically) by the postmodifier. This is also the case in three quarters of the instances in the TLEC, but

these (2x), *this* (1x), *those* (3x) and *our* (2x) are also used as determiners in the pattern **def *people* qual:phrase**. In the example below, the underlined nominal group is the first mention of the class of people in question. This usage of the demonstrative determiner corresponds to the colloquial usage of *this* to introduce referents in L1 English (Biber et al., 1999:274). Here *this* is not definite (Stirling & Huddleston, 2002:1510).

- (76) And this illness is caused because of poverty. In the other countries like Kenya people are so poor that they think sex is the only entertainment they can afford. So they sleep with each other most of the time And this sleeping together with each other causes this drastic disease.
And also this people with the idea that it will never happen to me. Because I'm too smart and too intelligent that something so low to happen to me. So he can pick and choose anyone he wants because his got a smart car, big house and a deseant job. And they are the people who are promoting this deadly disease. <ICLE-TS-KIMC-0290.1>

It is postulated that the underlined *people*-groups in the first two examples below are evidence of undeletion. The italicized nominal groups in the last three examples illustrate the deletion that could have taken place. These (italicized) nominal groups are called fused-head constructions by Stirling and Huddleston (2002:1504) because the determiner and the head are fused into one surface element (*those*). In SFG terms, the Deictic element in the experiential structure corresponds to the Head of the logical structure.

- (77) Those people against me should therefore reconsider their view of the topic and analyse it in terms of what will the future bring for us if we continue to deny modernization. <ICLE-TS-NOUN-0477.1>

- (78) The counsellors should be those people with a heart, *those who are willing to help criminals*. <ICLE-TS-NOUN-0397.1>

- (79) People really do not understand, especially *those in the rural areas*. <ICLE-TS-NOUN-0453.1>

(80) You are not likely to be hired as compared to *those from Accountancy colleges*. <ICLE-TS-NOUN-0468.1>

The use of *our* in the examples below is motivated by geographic provenance. This is consistent with the use of a possessive determiner when the 'populace' ('population of a certain place') meaning of *people* is activated, as discussed in Section 4.4.1. The first-person plural possessive determiner conveys a strong sense of solidarity with people from South Africa and the countries neighbouring South Africa, respectively.

(81) Poverty play a major role in our continent, because our people in (South Africa) want to be spoonfed, <ICLE-TS-NOUN-0306.1>³⁷

(82) So keep on inviting our people from our nearest countries to see this owsome province with owsome people and very friendly animals. <ICLE-TS-NOUN-0057.1>

As was mentioned earlier the difference between the two corpora with regard to the pattern **def *people* qual:clause**, is more appreciable when the raw frequency of concordance lines in this category is considered: 78 instances in the TLEC vs 15 in LOCNESS. This massive difference indicates that the instances with this pattern contribute to the high relative frequency of *people* (i.e. to the positive keyness of *people*) in the TLEC. Fourteen of the fifteen instances in LOCNESS (93%) have the definite article as determiner. Only 63% of the instances in the TLEC take the definite article.

What is interesting is the relatively high frequency of the distal demonstrative *those* in the TLEC data (26x). Of the 26 instances of the pattern ***those people* qual:clause** in the TLEC, 21 have a *who*-clause. This pattern is also illustrated in the single instance where a demonstrative determiner occurs in this analysis category in LOCNESS. The example below is typical of the use of *those* to establish esphoric reference. The identifying information is provided group-internally by the postmodifying clause.

³⁷ The writer probably intended to place the brackets around (*in South Africa*).

(83) Robert Rygor's words are fitting for those people who are against the idea. <ICLE-US-MRQ-0021.1>

The word *people* can be omitted (deleted) in the example above. Since the relative pronoun *who* is restricted to humans (or entities construed as human), the general noun *people* is not necessary. The pattern **those who-clause** “is understood as denoting a set of people” (Stirling & Huddleston, 2002:1504). This pattern occurs 107 times in the TLEC and 69 times in LOCNESS.³⁸ The pattern **those N who-clause**³⁹ occurs 33 times in the TLEC (and 11 times in LOCNESS). In two thirds of the instances of this pattern in the TLEC, the noun is *people* (21 times in this analysis category and once in an *of*-partition). In each of these cases, deletion of *people* is possible. This is illustrated by the examples below.

(84) Even though we believe that Aids has no class it affects those people who are poor. <ICLE-TS-NOUN-0165.1>

(85) According to me our government must try to help those people who came from different universities with jobs, because they spend more money for that degrees <ICLE-TS-NOUN-0479.1>

In the majority of cases where the distal determiner *those* is combined with a *who*-clause to denote a subclass of people, the pattern used in the TLEC corresponds to the preferred pattern in LOCNESS, namely that of noun-deletion. The noun is deleted in 107 instances and retained in 33 instances in the TLEC. Again, this shows that undeletion is variable. Still, the 22 cases of undeletion of the noun *people* contribute to the keyness of *people* in the TLEC. The type of qualifying information contained in the postnominal clauses of this category will be discussed in Section 4.5

³⁸ In concordances of *those*.

³⁹ N = noun. A concordance search was done for “those * who”.

4.4.3 Definite *people*-groups in quantifying *of*-partitions

In this section instances where the noun *people* is marked as definite and quantified by an *of*-partition will be considered. Again, two analysis categories can be distinguished, namely one without quality specification and one with quality specification. The first category takes the pattern **quant of def *people***⁴⁰ and is represented by the following example:

(86) Most of the people use their bodies to support their families more especially women. <ICLE-TS-NOUN-0062.1>

This pattern is much more frequent in the TLEC than in LOCNESS, both in terms of raw frequency (42x vs 2x) and in terms of relative frequency (20/1000 vs 3/1000). The large discrepancy between the two corpora with regard to raw frequency indicates that the pattern **quant of def *people*** contributes to the keyness of the noun *people* in the TLEC. The 2 instances in LOCNESS have *the* and *these* respectively as determiner. In the TLEC 69% of the instances of this pattern have the definite article as determiner, 17% have demonstrative determiners and 14% have possessive determiners. As was the case for instances of **def *people*** without quantification (Section 4.4.1), the demonstrative determiners point to a previously defined subclass of people.

(87) I believe that the cause of HIV/Aids in *adults or married adults* is caused by extramarital affairs. Some of these people do not believe that HIV/Aids exist. <ICLE-TS-WITS-0007.1>

In 4 of the 6 instances with a possessive determiner in the TLEC (**quant of poss *people***), the determiner is the first-person plural determiner, i.e. *our*.

(88) In Africa most of our people are not working. <ICLE-TS-KIMC-0363.1>
[opening sentence].

The frequent use of the possessive determiner and the preference for the first person form *our* is consistent with the higher frequency of possessives (and first person forms)

⁴⁰ Quant = Quantifier

noted earlier. In other words, we find the same phenomenon, but only in a different syntactic environment, namely in a quantifying *of*-partition.

With regard to the use of the definite article in this pattern, it is postulated that the pattern **quant of the people** is used in the TLEC in contexts where the pattern **quant people** (e.g. *most people, some people*) would be preferred in Standard English. The absence of the definite article in corresponding Standard English constructions might be indicative of a non-identifying use of the definite article in the TLEC. A cursory look at each of the TLEC essays in which the 23 instances of *most of the people* occurred, revealed that the referent of *the people* in each case was not a direct anaphor, but only indirectly identifiable in terms of the essay topic and other things mentioned in the context of the essay. The three examples below illustrate the association of *people* to an antecedent mention of a place.

(89) Just like at Congo in the year 1999. There was big war there at *Congo* that left most of the people in a trouble. <ICLE-TS-NOUN-0086.1>

(90) The people of Africa especially the children are being traumatised by high levels of crime, domestic violence and the impact of death and suffering in their families due to HIV/AIDS. Although there are lot of factors that cause the HIV/AIDS epidemic in Africa "poverty" plays a major role. Most of the people are unemployed and these cause poverty as the results people end up only not committing crime but also sell their bodies as way of getting income <ICLE-TS-KIMC-0382.1> [quoted from beginning of essay]

(91) *Most of our countries in Africa* are poor and most of the people do not have jobs <ICLE-TS-NOUN-0023.1> [opening sentence of essay]

In the examples above, the meaning of *people* is still generic and the choice of the definite article cannot directly be attributed to referent identifiability. Mesthrie (2006:139-140) postulates that quantifying *of*-partitions such as those in the examples above are due to the insertion of morphemes, i.e. a kind of anti-deletion. This possibility will be explored further in Section 4.6.

Nominal groups with definite determiners, as well as modifiers, that function as prepositional complements in quantifying partitive *of*-constructions are also much more prevalent in the TLEC than in LOCNESS, both in terms of raw frequency in the corpus (52x vs 4x) and as a normalized portion of the *people*-concordance (25/1000 vs 5/1000). The examples below represent the patterns in this analysis category:

quant of def *people* qual

(92) Most of the people who get this sickness or virus are those who are not educated because of poverty. <ICLE-TS-NOUN-0063.1>

quant of def qual *people* qual

(93) Most of our working people in Africa, earn little income [...] <ICLE-TS-NOUN-0317.1>

quant of def qual *people*

(94) Some of our African people have a belief that [...]

In the four instances in LOCNESS the choice of determiner is as follows: *the* (2x), *these* and *those*. 81% of the instances in this category in the TLEC are determined by *the*. Five instances have demonstrative determiners (*these* 4x, *those* 1x) and the other five the first-person plural possessive *our*. The use of this possessive pronoun emphasizes the writers' solidarity with African people, or in one case, with South Africans. In all five instances the provenance is also made explicit by a modifier, as can be seen in the last two examples above and the examples below:

(95) I think the problem come from us for example most of our south african people I think they [...] <ICLE-TS-NOUN-0304.1>

(96) I've notice that most of our people in Africa a poor. <ICLE-TS-KIMC-0290.1>

(97) About 60% of our people in Africa are not employed <ICLE-TS-NOUN-0217.1>

The examples with *our* quoted above can be reformulated as non-partitive quantifying construction, e.g. *most South African people*, *most people in Africa*. Such constructions would be more likely in Standard English. Note, however, that these non-partitive, indefinite agnates do not express solidarity or inclusivity. A Standard English agnate of the last example above would retain the *of*-partition, but would probably substitute *the* for *our*, as in *about 60% of the people in Africa*.

The use of the definite article may be ascribed to the identifying information provided by the postmodifier. In 9 of the 11 times that the pattern **quant of the (qual) people qual:phrase** occurred in the TLEC, the identification is in terms of geographic provenance. With regard to the pattern **quant of the people qual:clause** (23x), the reduced pattern (i.e. non-partitive with zero determiner, **quant people qual:clause**) would be preferred in Standard English in many of the cases. The nominal groups in the examples below, for instance, can be reformulated as *most people who stay in urban areas* and *most people who are not working*, respectively. Note the non-identifying, generic use of *the* with *urban area* in the first example and the appositive pronoun *they* in second example.⁴¹

(98) Most of the people who stayed in the urban area are employed because they have skills of working. <ICLE-TS-NOUN-0075.1>

(99) Most of the people who are not working **they** drink a lot. <ICLE-TS-NOUN-0217.1>

4.4.4 Summary

In this section all the instances where the noun *people* is marked as definite were considered (493 instances in the TLEC and 206 instances in LOCNESS). With regard to

⁴¹ The past tense marking on the verb *stayed* and the absence of plural marking on the noun *area* are notable, but fall outside the scope of the current discussion.

choice of definite determiner, there are proportionally more demonstrative and possessive determiners in the TLEC than in LOCNESS. The sense in which the noun *people* is used also plays a role in the selection of the definite determiner. Where a previously defined sub-class of *people* is referenced, the TLEC writers prefer to use a demonstrative determiner to make the anaphoric link more explicit. Possessive determiners are used when the 'populace' meaning of *people* is activated. In this regard, the TLEC shows a preference for the first person plural possessive (*our*) whereas the third person singular (*his*) is preferred in LOCNESS. Possessive determiners are used to express 'solidarity' in the TLEC, but not in LOCNESS. Where the 'society' meaning of *people* was activated, the definite article is typically chosen as determiner. A closer look at the data also showed that the use of the definite article is often not motivated by referent identification. In the TLEC the definite article is used with the noun *people* in contexts where the zero article would be used in Standard English. This usage of the definite article in generic contexts is attributed to what Mesthrie (2006) calls anti-deletion, i.e. the insertion of a morpheme.

In the supercategory of definite *people*-groups, there are two general patterns that show a discrepancy in frequency between the two corpora, namely the pattern **def people qual:clause** and the pattern **quant of def people** (both with and without modifiers specifying quality). The high frequency of these patterns in the TLEC compared to LOCNESS contributes to the high keyness of *people* in the TLEC and can be partially explained by two kinds of anti-deletion, namely undeletion and insertion.

21 of the 78 instances of the pattern **def people qual:clause** comprise the subpattern **those people who-clause**. The use of a pattern containing the noun *people* rather than the use of a nounless agnate, **those who-clause**, is regarded as a manifestation of undeletion.

Nearly all of the quantifying *of*-constructions in the pattern **quant of def people** in the TLEC (94), except those indicating specific amounts, can be reformulated as non-partitive quantifying constructions, i.e. **quant people**. The use of quantifying *of*-partitions rather than non-partitive quantifying constructions is attributed to the insertion of the morphemes *of the*.

4.5 Quality-specified *people*

For purposes of this study, qualification refers to any specifying information which is not purely referential (in the narrow sense of this term) or quantifying. Such a broad definition of quality specification covers both pre- and postnominal qualities and includes the so-called postdeterminers, such as *different*, as a type of quality. In Chapter 2 quality specification within the nominal group was defined in terms of modification which in turn is defined as additional propositional information about the Thing. In the discussion below, *modification* and *quality specification* are used interchangeably, the term *modification* being preferred where the rank and/or class of the grammatical units which realize qualities are discussed, and the terms *quality specification* or *qualification* being preferred where the semantic types of quality are discussed. In other words, *modification* is used here in a sense which is slightly more restricted than the conventional sense of the term. Reference to *modifiers* and *modification* in this Section excludes quantifying modifiers.

The addition of qualifying information to a noun (i.e. modification) has the effect of delimiting the **denotation** of the Thing to a subclass of Thing. This subclassification, which is an **experiential function**, is most obvious where the specifying information takes the form of a classifying adjective or noun, e.g. *African people*. However, a descriptive adjective can also be regarded as information that delimits the denotation, or type specification, of the noun. The set *poor people* is a subset of the larger set *people* and excludes some of its members, namely, 'all those who are not poor'. When used with a definite determiner, the qualifying information can also help to identify the **referent** of the nominal group, as discussed in Section 4.4 above, i.e. contribute to the realization of the **textual function** of the nominal group as a whole. Both the experiential and the potential textual functions of quality specification will be taken into account in the discussion below, but the focus will fall on the experiential function and therefore on the meanings realized by the various grammatical units that can express quality.

Whereas word groups typically realize quality prenominaly, phrases or clauses typically realize quality postnominaly (Halliday & Matthiessen, 2004:323). In other words, qualities that occur prenominaly in English are typically denoted by adjectives or

adjectival groups, whereas postnominal qualification is typically specification in terms of a Process in which the Thing is involved.

With regard to quality specification, the *people*-concordances of the two corpora are analyzed to determine the types of qualities of *people* that are expressed, as well as the grammatical units employed for doing so, in order to establish whether there are systematic differences between the two corpora with regard to the quality meanings expressed or their lexicogrammatical realizations. It is expected that a difference in either or both will help to explain the keyness of *people* in the TLEC, as well as show up systematic features of the grammar of BSAfE. Table 4.6 provides an overview of the various form-based categories according to which the concordance lines of *people* were analyzed. As with the analysis of definite determiners (in Section 4.4), the concordance frequencies with regard to quality specification can be viewed from three perspectives, namely, (1) raw frequency, (2) relative frequency per 1000 concordance lines, and (3) percentage of a form-based analysis category within the concordance. The first two are given in Table 4.6.

There is a very large difference between the two corpora in terms of raw frequency of instances where the noun *people* is specified for quality by modifiers: 883 in the TLEC vs 234 in LOCNESS. The modified instances of *people* therefore contribute to the keyness of this noun in the TLEC. The question is how? Quality specifiers can serve to sub-classify the Thing denoted by the noun. The noun *people* has a very general meaning. The possibility that sub-classes of human beings are denoted extensively in the TLEC by the noun *people* accompanied by one or more modifiers instead of using a more specific noun, such as *students*, will be explored below.

Relative frequency per 1000 can be used to compare the distribution of the various forms of modifiers in the two corpora (see Table 4.6), but it is also useful in contrasting the corpora with regard to the proportion of *people*-groups with quality specification in the whole concordance versus those without quality specification. In the TLEC 427/1000 concordance lines are specified in terms of quality, compared to 313/1000 in LOCNESS. All the instances of *people* with quality specification form a kind of supercategory, which will be called “qualified *people*” hereafter for ease of reference. The smaller form-based categories can also be grouped together as intermediate categories of varying levels

based on the types of modifiers they have, for example “*people*-groups with premodifier” and “*people*-groups with postmodifiers”.

There is a remarkable similarity between the two corpus concordances with regard to the use of premodifiers and postmodifiers. In each corpus 41%⁴² of the instances of “qualified *people*” have premodifiers. Similarly, 68% of “qualified people” in the TLEC contain postmodifiers, compared to 66% in LOCNESS.⁴³ In systemic terms, this means that the two concordances both traverse the selection path *QUALITY ROLE: ‘quality of thing’: ‘quality as modifier’* with similar frequency **after** the simultaneous selection of the noun *people* in the THING TYPE system network (Halliday & Matthiessen, 1999) and ‘quality presented’ in the Quality system network (Tucker, 1998). In other words, once the writer has chosen to use the noun *people* with a specification of quality, the positional type of modifier (prenominal or postnominal) is selected at approximately the same rate.

⁴² (284+76)/883 in the TLEC and (79+17)/234 in LOCNESS

⁴³ (76+186+334+3)/883 in the TLEC and (17+78+59+1)/234 in LOCNESS

Table 4.6 Quality specifications of *people* in the TLEC and LOCNESS

Pattern	TLEC		LOCNESS	
	Raw freq.	Per 1000	Raw freq.	Per 1000
\emptyset qual <i>people</i>	198	96	52	70
definite qual <i>people</i>	48	23	13	17
quant qual <i>people</i>	17	8	11	15
quant of qual <i>people</i>	8	4	2	3
def quant of qual <i>people</i>	1	<1	0	0
quant of definite qual <i>people</i>	10	5	1	1
... qual <i>people</i>	2	1	0	0
Prenominal quality specification only	284	137	79	106
def qual <i>people</i> qual:phrase	6	3	1	1
def qual <i>people</i> qual:clause	14	7	5	7
def qual <i>people</i> qual:phrase+clause	1	<1	0	0
quant qual <i>people</i> qual:phrase	4	2	2	3
quant qual <i>people</i> qual:clause	3	1	0	0
\emptyset qual <i>people</i> qual:phrase	17	8	3	4
\emptyset qual <i>people</i> qual:clause	21	10	4	5
quant of qual <i>people</i> qual:phrase	1	<1	0	0
quant of qual <i>people</i> qual:phrase+clause	1	<1	0	0
quant of qual <i>people</i> qual:clause	2	1	1	1
def quant of qual <i>people</i> qual:phrase	0	0	1	1
quant of def qual <i>people</i> qual:phrase	4	2	0	0
quant of def qual <i>people</i> qual:clause	2	1	0	0
Pre- and postnominal quality specification	76	37	17	23
zero <i>people</i> qual:phrase	96	46	23	31
definite <i>people</i> qual:phrase	30	15	45	60
quant <i>people</i> qual:phrase	39	19	9	12
quant of <i>people</i> qual:phrase	6	3	0	0
def quant <i>people</i> qual:phrase	4	2	0	0
quant of def <i>people</i> qual:phrase	10	5	1	1
<i>the most people</i> qual:phrase	1	<1	0	0
Quality specified by postnominal phrase only	186	90	78	104
\emptyset <i>people</i> qual:clause	158	76	30	40
def <i>people</i> qual:clause	78	38	15	20
quant <i>people</i> qual:clause	41	20	9	12
quant of <i>people</i> qual:clause	19	9	0	0
def quant of <i>people</i> qual:clause	4	2	2	3
quant of def <i>people</i> qual:clause	26	13	2	3
[other patterns]	7	3	0	0
<i>the types of people</i> qual:clause	1	<1	1	1
Quality specified by postnominal clause only	334	162	59	79
Quality specified postnominally only, by phrase + clause	3	1	1	1
TOTAL <i>people</i>-groups with quality specification	883	427	234	313

Differences between the two corpora are therefore to be sought in the selection of the noun *people* itself and/or in the selection of ‘quality presented’ for *people* (i.e. the decision to specify *people* further in terms of quality). As indicated above, differences between the two corpora can also be explored with regard to the lexicogrammatical forms that realize ‘quality as modifier’ as well as QUALITY TYPE. The corpora differ significantly with regard to the type of postmodifier preferred.

As can be seen in Table 4.6, there are many instances where *people* is specified for quality by more than one modifier. If one considers the forms of modifiers as a percentage of all modifiers of *people* (rather than a percentage of all instances of quality-specified *people*), the differences between the two corpora in respect of choice of modifier position and form is clear (see Table 4.7).

Table 4.7 Modifiers of *people* in the TLEC and LOCNESS

	TLEC		LOCNESS	
	Raw freq	%	Raw freq	%
Prenominal modification	360	37	96	38
Postnominal word group	40	4	11	4
Postnominal prepositional phrase	183	19	75	30
Postnominal clause	381	40	70	28
Total modifiers of <i>people</i>	964	100	252	100

There are more postnominal clauses than phrases in the TLEC data, whereas there are more postnominal phrases than clauses in the LOCNESS concordance of *people*. Van Rooy (2008a:281, 297) also reports that the TLEC writers make less frequent use of prepositional phrases to present additional information and links this to a lower informational density in the TLEC.

In the remainder of this section, quality types will be discussed in subsections based on the structures that realize them, ranging in complexity from prenominal words, postnominal word groups, postnominal phrases and, lastly, postnominal clauses.

4.5.1 Prenominal qualities

In the TLEC there are 360 (174/1000) concordance lines where the noun *people* is specified for quality prenominally. Of these, 76 also contain postnominal specifications of quality. In 32% (284/883) of the instances in the supercategory “qualified *people*” in the TLEC, *people* is specified for quality prenominally only. In terms of raw numbers there are fewer instances of *people* with prenominal modifiers in LOCNESS (96) than in the TLEC (360); however, as was pointed out above, the same percentage (41%)⁴⁴ of instances of *people* specified for quality have prenominal modifiers in each of the two corpora. In LOCNESS, 34% (79x) of the instances of qualified *people* have prenominal modifiers only.

This section deals with the intermediate category “people with prenominal specification(s) of quality”, i.e. all patterns containing the sequence **qual *people***. The patterns and form-based concordance categories involved in this intermediate category can be found in Table 8.5. In the remainder of this section the focus will fall on the meanings realized by these prenominal modifiers, which are typically adjectives.

In the TLEC there are 365 words that modify *people* prenominally, since there are 5 instances where the noun is modified by more than one modifier. In LOCNESS all prenominal modifiers represented a single quality per instance of *people*. This confirms Tucker’s observation (1998:211) that examples of modifier sequences are quite scarce. Although no generalizations can be made about the complexity or length of prenominal specifiers based on the concordance of only one noun, it should be noted that for the noun *people*, which is the most frequent noun in both corpora, there are rarely more than one premodifier, and that such premodifiers are rarely submodified, or tempered in the terminology of the Cardiff Grammar (Tucker, 1998). This is true of both the TLEC and LOCNESS.

A manually extracted wordlist of the prenominal adjectives and classifying nouns in the TLEC concordance of *people* yields 98 separate (unlemmatized) premodifiers (i.e 98 types and 360 tokens). There are 37 types of premodifier and 95 tokens in the LOCNESS concordance of *people*. The most frequent lexical item realizing prenominal

⁴⁴ 360/883 in the TLEC and 96/234 in LOCNESS

quality in both corpora is the relative adjective *other*. However, there are more than twice the number of instances of *other people* in the TLEC compared to LOCNESS (56 vs 23). In the TLEC, *other* (56x) is followed by *poor* (54x), *young* (31x), *African* (19x) and *innocent* (12x) with regard to frequency ranking. In LOCNESS, the most frequent premodifiers after *other* (23x) are *American* (9x), *young* (9x), *old* (7x) and *innocent* (5x). *African* and *American* denote the geographic provenance of the student writers who contributed to the TLEC and LOCNESS respectively. The correspondence in usage frequency of *young* and *innocent* is probably an effect of the essay topics for the two corpora. What is striking is the high frequency of the adjective *poor* (54x) in the TLEC, which will be returned to later.

The lists of pronominal modifiers of *people* were semantically classified in order to determine whether there were differences in the semantic types of qualities occurring pronominally.

The quality types referred to in Chapter 2, which are based on Tucker's system network for Quality (1998) as well as descriptive grammars of English (e.g. Biber et al., 1999:507-509), are taken as a starting point for classifying the 365 and 95 pronominal qualities in the TLEC and LOCNESS respectively. It should be noted, though, that the smallness of the data set does not warrant a semantic classification that goes into as much detail as a system network does, since too fine a classification will place undue emphasis on individual words and not allow a comparison in terms of broader semantic types. The purpose of a system network is to provide all potential meanings which can be selected from. Each selection gives rise to a new set of semantic features to be selected from, until a realization can be generated. Any given selection path across the Quality system network will culminate in an individual adjective or other qualificative word. The analysis of the concordance data traverses the system in the other direction, starting with the individual quality-denoting words. For instance, in the TLEC data, the adjectives *African* (19x), *American* (1x) and *South African* (5x) denote 'geographic provenance'.⁴⁵ There are also other specifications of 'provenance', e.g. *black* (7x) which is a specification of 'ethnic provenance'. 'Provenance' qualities, along with 'domain' qualities, e.g. *professional*, are 'classifying' qualities. The various semantic classification

⁴⁵ Meanings are placed between single quotation marks.

paths will not be discussed in any further detail. Only the broader categories of Quality type, and their subcategories, are reported below.

Based on Tucker's (1998) system network for Quality, three broad categories of prenominal quality can be distinguished: 'relative' qualities, 'classifying' qualities and 'epithetic' qualities. The latter two are Thing-oriented qualities. An additional category of Thing-oriented qualities dealing with 'social status' qualities was deemed necessary for purposes of analysis of the corpus data. Tucker (1998) treats adjectives like *employed* and *married* as 'classifiers', since they indicate systemic distinctions pertaining to 'social status': 'working/not working' and 'married/not married'. Adjectives like *poor* and *rich*, which are not explicitly addressed by Tucker, indicate 'economic status', while adjectives like *uneducated* indicate 'educational status'. It is contended here that adjectives like *poor/rich* and *educated/uneducated* should be grouped with adjectives like *employed/unemployed* and *married/unmarried*, since they are indicators of social status. Yet *poor* is not treated as a classifier in descriptive grammars of English (e.g. Biber et al., 1999:508-509). Adjectives like *poor*, *employed* and *married* are not as central to the category of classifying adjective as are the affiliative (e.g. *Christian*), domain (e.g. *professional*) and provenance adjectives, and for this reason they are given their own category here. In conceptual space, the social status adjectives are located at the point where description shades into classification. Treating the social status qualities as a separate analysis category turned out to highlight the biggest difference between the two corpora with regard to Thing-oriented qualities (i.e. epithets and classifiers). These will be discussed in more detail after the discussion of relative qualities.

4.5.1.1 *Relative qualities*

In the TLEC, 8 different words (with a total of 78 tokens) are classified as relative qualities. In LOCNESS there are 7 different words (with a total of 33 tokens) denoting relative qualities. In the TLEC, 21 % (78/365) of the prenominal qualities of *people* are relative qualities, while relative qualities make up 35% (33/95) of the prenominal qualities of *people* in LOCNESS. *Own* (3x in the TLEC, 1x in LOCNESS) only occurs after possessive pronouns and, in the *people*-concordances of both corpora, without further specification of quality. In both corpora *only* (2x in the TLEC, 1x in LOCNESS) occurs

after definite determiners and with postnominal qualifiers. *Same* also only occurs with definite determiners. The single instance of *same* in LOCNESS has no further quality specification, whereas both instances in the TLEC also have postnominal qualifiers. All instances of *different people* were without overt determiners. Two of the six instances of *different people* in the TLEC have postnominal qualifiers. Likewise, in LOCNESS two of the five instances of *different people* have postnominal qualifiers. The 7 instances of *very* as a relative quality in the TLEC are restricted to definite nominal *people*-groups and also occur with postmodifiers, for example:

(100) And they are the very people who spread HIV Aids<ICLE-TS-NOUN-0258.1>

It was pointed out above that *other* is the most frequent quality of *people* in both corpora (56x in the TLEC, 23x in LOCNESS). Since a peculiar usage of *other* has been observed in descriptions of BSAfE (Van der Walt & Van Rooy, 2002:124; De Klerk and Gough, 2002:363; De Klerk, 2003a:470; Mesthrie, 2004:969), the 56 instances of *other* as a relative quality of *people* in the TLEC will be considered more closely. The example below represents prenominal *other* used in its standard sense as a contrastive relative quality.

(101) If somebody is isolated from other people he/she feels that they don't belong <ICLE-TS-NOUN-0064.1>

Use of the pattern ***other...other*** is widely reported in literature on BSAfE. Van der Walt and Van Rooy (2002:124) report that 82% of the learners in their study accepted “quantifier” constructions such as *other...other* as standard. De Klerk (2003a:470) reports 67 instances of *other...other* in her corpus of Xhosa English, which she regards as a subcategory of Black South African English (2003a:463). These are described as “[u]niquely characteristic usage of quantifiers” (De Klerk 2003a:470).

It should be noted that, in the current study, *other* was initially analyzed and counted as a relative quality, following Tucker (1998) and not as a quantifier. The quantifier label given to *other...other* in BSAfE literature arises from the correspondence of this pattern to *some...other* in Standard English. *Some* is a quantifier. The question is whether a

'quantifying' meaning of *other* is intended by BSAfE speakers. Here explicit quantification through words like *most* and *many* should be distinguished from the quantification inherent to indefinite determiners such as *some* and *any*. It is contended here that the primary function of the latter is determinative rather than quantifying, and that the quantifying function of an indefinite determiner, such as *some*, is a by-product of its referential function. In the *people*-concordance of the TLEC, there are only two instances of the *other...other* pattern, but at least 19 instances where *other* is used in a non-standard manner as an indefinite determiner rather than relative quality. In order to illustrate this determinative function of *other*, it is necessary to consider longer stretches of text. The essay below (quoted in full) contains four instances of *other* used as a determiner rather than as a relative adjective.

(102) Africa as a centre of Attraction. to tourists there is high rate of Aids transmission. Most tourist are coming to Africa for different purposes. It is believed that Aids spread like veld fires among the poorest. Poor especially women they are advertising themselves to different men which at the end increasing the rate of transmission. These poor people ware doing this to have a better living. Other people believed that to be married to a foreigner you are to be rich and recognised. Instead this are the people who travelled different countries and met different people and at the end spread Aids. Also aids can be spread to this poorer countries when there is poor sanitation, contamination of water and food. *Other countries* can stay some days without water and food. This resulted in making their bodies weak and not able to fight the diseases that attack. Other people in this poor countries tend to take care of sick patient in their homes. They are doing this without any instructions being followed. They handle the patient the way they want and at the end there is transmission of the disease among the families. Other people are ignorant in the use of methods that prevent aids like using of a condom. <ICLE-TS-NOUN-0211.1> [essay quoted in full]

In the nominal groups highlighted in the essay above *other* is used as a referent-introducing determiner corresponding to the stressed *some* of Standard English. In the

latter two instances of the pattern *other people* in the essay above, it is also possible, in Standard English, to omit the determiner completely and to attach a generic (i.e. not referring to a specific group) interpretation to *people*. The nominal group *other countries* in the essay above shows that this determinative usage of *other* also applies to nouns other than *people*. The examples below are further illustrations of the referent-introducing determinative function of *other* in the *people*-concordance.

(103) As we live in the modern society there a many thing that we do as human beings. Because we want to employ our selves but jobs are nowhere to be fund thus. there are so many deaths and crime that causes other people to be negative towards life. <ICLE-TS-KIMC-0361.1> [quoted from beginning of essay]

(104) And as we know that the is different type of crime that man can do. Sometimes other people hijack the car's and they become the victim of that whitout realising that there is danger behind. <ICLE-TS-KIMC-0341.1>

In an agnate of the first example, *other* can be replaced with *some*: *that causes some people to be negative*. It is possible to leave out the “determiner” in the second example: *Sometimes people hijack cars*. These examples illustrate two phenomena in BSAfE, namely (1) the use of *other* instead of *some* and (2) the insertion of an indefinite determiner (such as *other*) before a generic plural noun in contexts where no determiner would be preferred in Standard English. With regard to the first, it should be noted that the words *some* and *other* translate to the same Setswana root: *-ngwe* so that *batho ba bangwe* can mean either ‘other people’ or ‘some people’.⁴⁶ The use of *other* as synonym of *some* in BSAfE is therefore attributable to substrate language influence. The use of *some* as indefinite marker will be discussed in detail in Chapter 6.

The next example illustrates the referent-introductory use of *some* with the contrastive *other* in the TLEC, i.e. the *some...other* construction also found in Standard English.

⁴⁶ Personal communication with Prof Rigardt Pretorius and Ms Ansu Berg.

(105) Another thing, because of poverty people migrate from their places of origin in a seek of a better life. Like people who are living in rural areas and informal settlement like farm's they move to the urban areas like cities, town and suburbs in a seek of a job. when they arrived there they tried to start a better life, **some** are the parents they left their children at home's and **some** are married husband's they left their wives at their places, the first thing they will meet the other people they dont know and they will enjoyed being far from home because they meet **other new faces**. **Some** are going to fall in love with **others** and what will be consequenses. AIDS, pregnancy and death. <ICLE-TS-NOUN-0311.1>

The corpus data show that the *other...other* pattern has not fully replaced the *some...other* pattern, but rather that they co-exist in BSAfE. In the example above, instances of *some* (as pronoun) refer to sub-classes of migrant people from rural areas whereas the instances of *other(s)* refer to the people they encounter in their new urban environment. The *some...other* construction is employed here to contrast two groups of people and to clearly track them as separate referents. The use of *other* in the last two instances in the excerpt above can be omitted, as in: *they will meet people they don't know* and *they meet new faces*.

Below are the only two examples of the *other...other* pattern in the concordance of *people*.

(106) I hear other people laughing about they are doing to other people, forgetting that what goes around comes around. <ICLE-TS-KIMC-0348.1> [beginning of new paragraph, no anaphoric link to prior info]

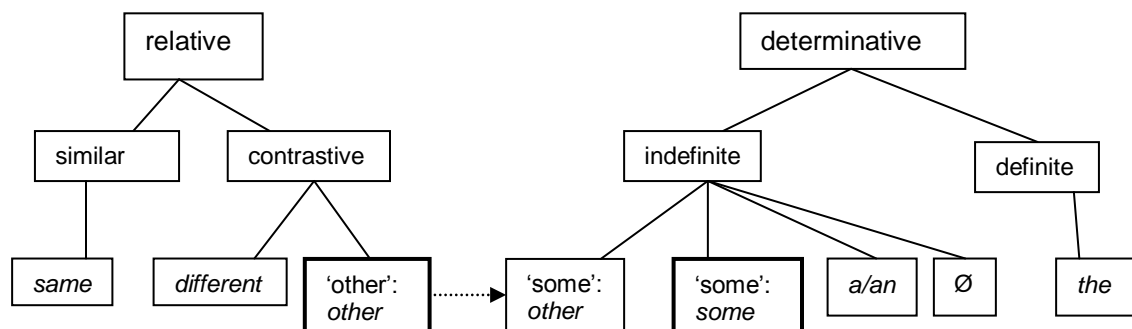
(107) Sometimes other people rape because they know that they had this virus and they want to spread it amongst other innocent people, even young kids. <ICLE-TS-NOUN-0305.1>

In the analysis of relative qualities the focus fell on the most frequent premodifier of *people* in both corpora, namely *other*. Careful analysis of the concordance lines

containing *other people* in the TLEC indicates that use of *other* in BSAfE is often extended from a contrastive usage to a purely determinative usage, i.e. as synonym of *some*. At least 19 of the 56 instances of *other people* in the TLEC manifest this extended usage.

Traversal of the quality system network (Tucker, 1998) from meaning to lexical realization only reveals the contrastive (relative) meaning of *other*. However, the TLEC data have shown us that *other* is also a realization of the determiner system network (Halliday & Matthiessen, 2004) in BSAfE. When starting from the surface patterns (the realizations themselves) rather than the meanings, it is easier to discover meaning extensions of words such as *other*. That is, a usage-based approach starting with exemplars (Bybee, 2006:714) is more effective in revealing meaning shifts and extensions than a systemic network approach. Such meaning extensions can be represented by a categorization network (see below) or as constructions.

Figure 4.2 The extension of *other*⁴⁷



4.5.1.2 Thing-oriented qualities

After subtraction of the relative qualities, we are left with 287 ‘thing-oriented’ qualities of *people* in the TLEC and 62 in LOCNESS. The quality types discussed below are reported as percentages of ‘thing-oriented’ pronominal qualities in the respective corpus

⁴⁷ Elaborative (taxonomic) relations are indicated by solid lines, while extensions are indicated by dotted horizontal lines (see Van Rooy, 2006). This categorization network is not maximal in the sense that not all potential elaborations and extensions of each node have been given – only enough to illustrate the extension of the meaning of *other* in the direction of *some*.

concordances. Three broad categories are distinguished, namely, ‘classifiers’, attributes pertaining to ‘social status’ and ‘epithets’. In terms of raw frequency there are more instances of each of the three prenominal, ‘thing-oriented’ qualities in the TLEC. Percentage-wise, the LOCNESS writers use twice as many ‘classifiers’ (as qualities of *people*) as the TLEC (42% vs 21%), but qualities that pertain to ‘social status’ (and which are potentially a means of social classification) are used much more in the TLEC than in LOCNESS: 38% vs 5% (see Table 4.8).

Table 4.8 Prenominal thing-oriented qualities of *people* in the TLEC and LOCNESS⁴⁸

Quality type and example(s)	TLEC	LOCNESS
Classifier: (misc) <i>specialist, Tswana-speaking</i>	4	3
Classifier: affiliation <i>Christian</i> (LOCNESS)	0	3
Classifier: domain <i>professional</i>	6	2
Classifier: provenance: ethnic <i>black</i>	15	1
Classifier: provenance: geographic <i>African</i>	36	17
SUBTOTAL: Classifier	61 (21%)	26 (42%)
Epithet: objective: (misc) <i>hungry, drunk, homeless</i>	7	1
Epithet: age <i>old, young</i>	45	18
Epithet: evaluation: (misc) <i>innocent</i>	16	9
Epithet: evaluation: behaviour to others <i>friendly</i>	7	0
Epithet: evaluation: emotive <i>awesome</i>	1	0
Epithet: evaluation: epistemic <i>real</i>	7	3
Epithet: evaluation: worth <i>good, bad, important</i>	8	0
Epithet: physical health <i>sick, infected, HIV-positive</i>	13	2
Epithet: extent <i>biggest</i>	1	0
Epithet process-related <i>affected</i>	12	0
SUBTOTAL: Epithet	117 (41%)	33 (53%)
Social status: economic <i>poor</i>	81	1
Social status: education <i>educated</i>	17	2
Social status: employment <i>working</i>	7	0
Social status: marital <i>unmarried</i>	4	0
SUBTOTAL: Social status	109 (38%)	3 (5%)
TOTAL: Thing-oriented qualities	287	62

In 81 of the 109 instances where a ‘social status’ quality is ascribed prenominal to *people* in the TLEC, that quality relates to ‘economic status’, the most frequently

⁴⁸ Raw frequencies are given in this table. Examples are from the TLEC data unless otherwise indicated. The percentages are percentages of thing-oriented prenominal qualities of *people*.

employed adjective being *poor* (54x) (see example below).⁴⁹ The high usage of *poor*, as well as the other 'economic status' qualities, ties in with the fact that *poverty* is used in the wording of the most frequently chosen essay topic in the TLEC.

(108) I would suggest to the government that they must let poor people attend schools for free even intitution they should pay for them so that they can make better lives for themselves. <ICLE-TS-NOUN-0302.1>

In both corpora, classifying adjectives most frequently realize 'geographic provenance' (36 in the TLEC vs 17 in LOCNESS), the most frequent of which is *African* (19x) in the TLEC and *American* (9x) in LOCNESS. In terms relative frequency as a percentage of 'thing-oriented' prenominal qualities, 'geographic provenance specifiers' are more frequent in LOCNESS than in the TLEC (27% vs 12%). However, in the TLEC, there are also 'ethnic provenance' specifiers, the most frequent of which is *black* (7x), and they comprise 5% (15/287) of prenominal 'thing-oriented' qualities.

Nouns or embedded nominal groups are also used prenominally to classify *people*, but much more rarely than adjectives. There are 6 instances where *people* is modified by a classifying noun or nominal group in LOCNESS, e.g. *business people* and "*new age*" *type people*.⁵⁰ Apart from the two standard uses of nouns as classifiers of *people* in the TLEC (*business people*, *specialist people*⁵¹), there are also seven non-standard utilizations of nouns as classifiers of *people* (exemplified below). The non-standardness of the example below has to do with the structure of the nominal group as a whole.

(109) At the end we will have many HIV people who are positive, and that will lead to AIDS. <ICLE-TS-KIMC-0363.1>

In the example above, the information unit *HIV-positive*, which is expected before the noun, is fragmented so that some of the information (*HIV*) occurs before the noun and some of it (*who are positive*) after the noun. There are two instances of *HIV-positive*

⁴⁹ In the TLEC: *rich* (7x), *poverty-stricken* (7x), *poorer* (4x), *unprivileged* (2x), *needy* (2x), *disadvantaged* (1x), *wealthy*(1x), *well-off* (1x), as well as *poverty-related* (1x) and *poverty* (1x) as prenominal modifiers. The reported counts include misspellings of the adjectives listed here.

⁵⁰ <ICLE-BR-SUR-0018.3> and <ICLE-US-MRQ-0005.1>

⁵¹ <ICLE-TS-WITS-0003.1> and <ICLE-TS-NOUN-0019.1>

people in the TLEC. The next example can also be reformulated as *HIV-positive people*, but it should be noted that *HIV/Aids* can function as a classifier in a nominal group if the noun is more specific in its denotation. Compare *HIV/Aids patients*.

(110) that's why Africa experience a high rate of HIV/AIDS people. <ICLE-TS-NOUN-0311.1>⁵²

In the next set of examples it is the generality of the noun *people* which is incongruent with the use of classifying nouns.

(111) It kills people even if it is not through intercourse, because when we look at the case of car accident people trying to help each other not knowing whether one is a victim we touch the blood of that person trying to help if you have a scar it mean you are a victim because it go through the blood stream. <ICLE-TS-NOUN-0120.1>

(112) Why do they think for only nurses and sciences people, are this equipments of other programmes expensive? <ICLE-TS-NOUN-0461.1>

(113) Prison people are also our problem because they affect our progresses. <ICLE-TS-NOUN-0403.1>

One would expect more specific nouns in Standard English reformulations, e.g. *car accident survivors*, *science students* (or *scientists*) and *prison inmates* (or *prisoners*). It is as if *people* serves a function similar to that of a derivational affix, like *-ist* in *scientist* and *-er* in *prisoner*. The plural form *sciences* in the second example above might be due to reference to a number of related disciplines in the context of a university, e.g. *sciences* as opposed to *humanities*. However, the use of plural suffixes with uncountable nouns in BSAfE has been noted and is discussed in Chapter 6.

In the next example, it is the classifying noun that is used in a function similar to that of a derivational suffix. Since *man-* indicates gender and since it is prefixed to *people* with a

⁵² The quantifying *of*-construction will be discussed in Section 8.4 below.

hyphen, it is difficult to say whether it is intended as a bound morpheme or as a classifying premodifier. It is hard to say why the writer did not choose *men* or *males*.

(114) Mental prisons created by the man-people against us women is all physical. <ICLE-TS-NOUN-0425.1>

The use of *poverty* in the example below may be primed by the wording of one of the essay topics, but *poverty-stricken* occurs only 7 times compared to the 54 times that *poor* is used.

(115) Poverty people they wish to satisfy their need <ICLE-TS-NOUN-0233.1>

In both corpora the type of epithetic quality most frequently realized is ‘age’, followed by ‘evaluative qualities’. In the TLEC, 16% (45/287) of the prenominal, ‘thing-oriented’ qualities of *people* denote ‘age’. Percentage-wise, ‘age’ qualities are chosen even more frequently in LOCNESS: 29% (18/62). In both corpora the most frequent ‘age’ adjective is *young*, followed by *old*.⁵³ The second most prevalent kind of epithetic quality in both corpora are ‘evaluative’ epithets, comprising 14% (39/287) of prenominal ‘thing-oriented’ qualities in the TLEC and 19% (12/62) in LOCNESS. As was mentioned above the adjective *innocent* is the most frequent ‘evaluative epithet’ in both corpora.⁵⁴ Apart from two denotations of ‘health status’, namely *handicapped* and *infected*, there are no other ‘physical state epithets’ in the LOCNESS concordance data. The adjective *big* (2x) in LOCNESS is used not solely to denote physical size, but in its extended sense to refer to ‘adults’ as opposed to ‘children’ and is therefore classified here as ‘age’ in the sense of ‘life stage’, as can be seen from the example below.

(116) Children draw their own conclusions from what they see **big people or adults** do. <ICLE-US-SCU-0010.2>

⁵³ In TLE: *young* (31x); *old* (6x). In LOCNESS: *young* (9x), *old* (7x).

⁵⁴ The adjective *innocent*: 12x in TLEC, 5x in LOCNESS.

In addition to the 13 instances where the health state of *people* is indicated in the TLEC,⁵⁵ there are also ten descriptors (18 instances) which denote more objective 'epithetic' qualities, i.e. not purely 'subjective evaluations'. These are: *sexual active*,⁵⁶ *helpless*, *homeless* (2x), *outcoming*,⁵⁷ *hungry*, *affected* (8x), *raped*, *reformed*, *rehabilitated*, *suffering*. There are a number of participial forms amongst these adjectives. Participial forms are also found amongst the 'social status' and 'health state' qualities, e.g. *married*, *employed*, *working*, *educated*, *infected*, etc.

Some of the participial forms in the TLEC are used in ways that differ from Standard English usage. These will be considered in more detail, starting with the most frequent participial form, *affected*. In the discussion of *the affected people* in Subsection 4.4.2.1, two potential explanations were mentioned, namely, that *affected* is used prenominal where it is expected postnominally in Standard English, or that no clear lexical content distinction is made between *affected* and *infected*. A Standard English reformulation would place *affected* after the noun, as in *the people affected*. In the absence of clear evidence to the contrary, one has to assume that the writer is not confusing *affected* and *infected*. In the example below both *infected* and *affected* are used as adjectives suggesting that they are distinguishable words.

(117) The Aids virus brings shame, nightmares, hate, loneliness was it attacks someone. Everyone would never want to glance back again to the **infected** person even though they could be their relatives, mother, brother, sister or father. The affected people are it not treated like any other humanbeings even if they deserve the right to do so. <ICLE-TS-NOUN-0265.1> [from the beginning of the essay]

In the essay below (quoted in full), the writer is using *affected* (in boldface) in the sense of 'infected'. *People-groups* are underlined in the essay below. It is possible that *affected* is used euphemistically.

⁵⁵ In TLE: *infected* (4x), *HIV-positive* (4x), *sick* (2x), *crippled* (1x), *disabled* (1x), as well as *HIV-Aids* as classifying noun.

⁵⁶ Standard English: *sexually-active*

⁵⁷ In the sense of 'outgoing'.

- (118) Poverty is when someone is on need of so many things for survival. HIV/AIDS is a disease which is caused by staying with so many partners. Poverty leads to abuse of bodies morespecially ladies, poor education and no shelter at their homes.
- Poor people suffer mostly because they do not have money, so this causes ladies to use their bodies to earn money for better lifes. They give themselves to everyman they come across so that they could have money. Man no longer stick to their wives because of those people who sell their bodies. when they sell their body, they do it with affected people and they got the disease.
- Affected** men sometimes use young ladies by paying school fees for them. If the poor lady come across this men, she stays with him for good support. In some African countries government does not support children, so this come difficult for poor people to get better education.
- Poor people do not have enough shelter for the family. This results in some children goes to town for better living. They are streets children who end up eating dirty things with bacteria. They can even use the things used by affected people which will end up being **affected**. After being **affected** they go back to their home village. Their parents did not know what their children are suffering from. Without any knowledge from the school. They treat their children others so this leads to them be **affected** because they do not know.
- Poor people need employment. Rich people can abuse them. if they have **affected** person in their home, they give poor people the job, to wash and makeup things for **affected** person. Morespecialy people living in rural areas know not much about the disease. They treat **affected** person like everyone, believing that is not HIV/AIDS. In African countries is not god to say someone is affected by Aids, this results in high rate of the disease. Affected people do not like to die without spreading. They use their money to get partners because you can not tell if he/she is positive or negative. Others use cars as a prestige to collect ladies with. People from different African countries can ran from their country to the other to affect.
- The better of improving poor people is to give them support, so that they can live a good life as others. Give them a good education.
<ICLE-TS-NOUN-0214.1>

There are a small number of instances in the TLEC *people*-concordance where the comparative form of the adjective is used in non-comparative contexts where the uninflected form is expected. The adjectives involved are *older* (4x), *younger* (1x) and *poorer* (4x), for example:

- (119) Even the older people sometimes get involved with the youth, [...]
- Older people will handle their kids unprotected and thats when aids affect them. <ICLE-TS-NOUN-0165.1>

(120) Younger people are most severely affected by the diseases with around 60% of all adults who acquire HIV becoming infected before they turn 25. <ICLE-TS-KIMC-0357.1>

(121) Firstly most poorer people will migrate to other countries to search for jobs but they will end up being unemployed [...]

Poorer people may end up becoming street kids that means they will eat contaminated foods, drinking alcohol also smoke glue which is unhealthy because they will be easily raped and affected by this serious disease.

(122) Most poorer people are illiterate so they do not use the contraceptive like condom correctly ... <ICLE-TS-NOUN-0223.1>

In this subsection prenominal modifiers denoting social class, epithetic qualities and classifying qualities were examined. The analysis showed that there are many more social status specifiers in the TLEC than in LOCNESS (109 vs 3) and that the majority of social status adjectives specified economic status (81x), with *poor* being the most frequently used social status adjective (54x). With regard to classifiers, the seven non-standard uses of classifying nouns are evidence that **qual people** is sometimes used where a more specific noun would be used in Standard English, e.g. *prison people* vs *prisoners*. With regard to lexical specificity, Partridge (2011:138) notes less specificity in the lexical verbs used in the TLEC compared to those used in LOCNESS. In the discussion of epithetic qualities above, a few instances of non-standard uses of participial adjectives, specifically *affected*, were observed, as well as a few instances of the use of the comparative form of adjectives in non-comparative contexts.

4.5.2 Postnominal word groups

Of the 223 non-clausal postnominal modifiers of *people* in the TLEC, 40 are word groups and the rest are prepositional phrases. In the LOCNESS concordance of *people*, 11 of the 86 non-clausal postnominal modifiers are word groups rather than prepositional phrases. The following word groups have the potential to postmodify nouns: appositive nominal groups, postposed adjectival groups and adverbial groups. Appositive nominal

groups show the largest discrepancy between the two corpora with regard to postnominal word groups as modifiers of *people*, as can be seen from Table 4.9 below.

Table 4.9 Postnominal groups as modifiers of *people* in the TLEC and LOCNESS⁵⁸

	TLEC	LOCNESS
Appositive nominal groups	28	1
Adjectival groups	3	8
Adverbial groups	9	2
TOTAL	40	11

Apart from a single instance of the reflexive pronoun *themselves*⁵⁹ and the expression *big people or adults* (example below), there are no further nominal groups that postmodify the noun *people* in LOCNESS.

(123) Children draw their own conclusions from what they see big people or adults do. <ICLE-US-SCU-0010.2>

In contrast, there are 28 nominal groups in the TLEC which are in an appositive relation to *people*. The use of *people* with appositives might contribute to its high frequency in the TLEC compared to LOCNESS. In the LOCNESS example above, the noun *adults* is denotationally more specific than *people*, but it is synonymous with the nominal group *big people*. In other words, the appositive nominal group *or adults* is a restatement of *big people*, i.e. an elaboration in SFG terms (Halliday & Matthiessen, 2004:489). In all 28 cases in the TLEC concordance of *people*, the pivotal noun of the appositive nominal group is more specific than the noun *people* too. However, unlike in the LOCNESS example, there are no instances where the appositive nominal group is a synonymous restatement of the content of the *people*-group. In every instance the appositive nominal group is a hyponymous reclassification, i.e. a subclassification. In fact, in the majority of cases, there are no prenominal quality specifiers of *people*. The devices used to establish these hyponymic appositive reclassifications, as well as the subclasses denoted by the appositive nominal groups are of interest. The most frequent way of establishing a hyponymic apposition of *people* is by making use of wordings which

⁵⁸ Only raw frequencies are given here.

⁵⁹ <ICLE-BR-SUR-0013.1>

include *especially* (12x). In 8 of the ten cases where appositive noun phrases subclassify *people* in terms of gender, the word *especially* is used in the appositive nominal group.

(124) They should continue to teach people especially the youth about Aids and most importantly in rural areas. <ICLE-TS-NOUN-0073.1>

(125) Yes poverty causes HIV/AIDS in Africa because; people or especially black males who do not have jobs turn to use sex as the only thing they can do during the day. <ICLE-TS-NOUN-0312.1>

(126) The most startling thing in Prostitution is that people especially men do not want to do it with protection <ICLE-TS-NOUN-0250.1>

(127) People especially women try to support minor kids by being prostitutes. <ICLE-TS-NOUN-0105.1>

The one instance of *moreover* as initial word of an appositive nominal group is also used in the sense of 'especially' rather than in the conventional sense of the word. The word *even* is used once to initiate an appositive nominal group. In seven of the cases the writing system itself is employed to offset the appositive from the *people*-group, for instance by using brackets and commas.

(128) Because of no job, low education, lack of facility the people more over the youth are comited in useless things <ICLE-TS-NOUN-0156.1>

(129) they want to spread it amongst other innocent people, even young kids. <ICLE-TS-NOUN-0305.1>

(130) they start falling in love with many people (men). <ICLE-TS-NOUN-0233.1>

(131) Young people go-out with elderly people (rich people) in order to make a living for them. <ICLE-TS-NOUN-0315.1>

There is one instance in the concordance data where the forward slash is used to symbolize *or*. There are a further 6 instances where *or* is used to initiate the apposition.

(132) it must creat jobs so that people/victims of HIV/AIDS could be on the position to buy those Aids drugs. <ICLE-TS-NOUN-0132.1>

(133) In this period they are locked in cells where they mix with different people or other prisoners who comitted different crimes. <ICLE-TS-NOUN-0384.1>

(134) When people or colleagues find that a graduate is posted to their institution, they fill very happy and proud thinking that they are going to gain a lot from him <ICLE-TS-NOUN-0457.1>

(135) Poverty is a situation where people or the family are poor. <ICLE-TS-NOUN-0003.1>

(136) Because things that are done by people or our parents make our hearts broken. <ICLE-TS-NOUN-0212.1>

(137) People or men in most cases are the ones who decide whether to have protected sex or unprotected one <ICLE-TS-NOUN-0070.1>

(138) One of the people or a friend who was close to me died because of Aids. <ICLE-TS-KIMC-0153.1>

Although apposition is not one of the linguistic features considered separately in the multidimensional comparisons of the TLEC and LOCNESS (Van Rooy, 2008a; Van Rooy & Terblanche, 2009), insights gained from these studies with regard to the packaging of information in the TLEC shed light on the use of appositions in this corpus. The type of apposition exemplified above represents an add-on strategy typical of spoken discourse with the pressure of real-time production (Biber et al., 1999:1068). However, the essays comprising the TLEC were hand-written in a class situation without

“severe time constraints” (Van Rooy & Terblanche, 2006:165). Use of the add-on strategy rather than subordination implies that the TLEC “resembles spoken language more than LOCNESS does in respect of the way information is presented” (Van Rooy, 2008a:281). However, Van Rooy (2008a:288) points out that, although the TLEC resembles spoken discourse more than LOCNESS in some respects, it is not “like spoken language”, i.e. speech-like. The appositive nominal groups above fit nicely into Van Rooy and Terblanche’s Dimension 2, an indication of transparency, which they describe as follows:

Our Dimension 2 features tend to overlap more with features that show evidence of real-time constraints, resulting in **generalised lexical choice** and **sequentially structured, non-integrated information**, combined with very explicit marking of particular cohesive relations (Van Rooy & Terblanche, 2009:246, my emphasis – YVB).

In the cases with appositive nominal groups after *people*, the general lexical item *people* is edited (without being deleted) by an apposition with a more specific description, which is not syntactically integrated into the nominal group.

The raw frequency of postposed adjectival groups is higher in LOCNESS than in the TLEC (8x vs 3x). In order to consider the potential corresponding constructions that may be preferred over the postposed adjectival group in the TLEC, all eight instances of postnominal adjective in LOCNESS and all three instances in the TLEC are given below. Adjectives with a past participle form, such as *involved* (see examples 137 and 138), are amongst the examples of postposed adjectives given by Biber et al. (1999:519). There is one example of postposed *involved* in the TLEC concordance and one in the LOCNESS concordance.

(139) It used many first hand accounts from people involved, <ICLE-US-MRQ-0005.1>

(140) Poverty is the cause of HIV/AIDS because people involved are mostly illiterate <ICLE-TS-NOUN-0069.1>

In the cases with *involved*, as well as the examples below, the postposed adjectival group can be rephrased as a finite postnominal clause by inserting *who are/were* directly

after *people*, as in *people who are guilty of crimes*. However, this is a less economical, more explicit way of conveying the informational content of the adjective or participle, as these four examples of from LOCNESS illustrate:

(141) It meant that he was in a milieu of guilt, since he was defending people guilty of crimes, or trying to prove people guilty of crimes. <ICLE-BR-SUR-0008.1>

(142) At my high school the people who were caught smoking underage received the same punishment as the people old enough to smoke. <ICLE-US-SCU-0005.2>

(143) since the Lisbon earthquake was an actual occurrence which shocked and stupefied many people, unable to comprehend its nature or the force behind it. <ICLE-BR-SUR-0009.2>

(144) This scenario sounds very realistic to people still angry at the L.A.P.D. about that incident. <ICLE-US-MRQ-0045.1>

In all of the examples above the adjective is descriptive (*guilty*, *old*, *unable* and *angry*). In the example below, the adjective (*good*) is also descriptive, but the adjectival group as a whole specifies economical status. Here a familiar, high-frequency adjective with a more general meaning is chosen with *economically* instead of an adjective with a more restricted meaning such as *sound* or *secure*, as in *economically sound* or *economically secure*. In a Standard English reformulation, the 'social status' information presented postnominally in the example below would be presented prenominally.

(145) I strongly believe that is high time our governments should stand and their people, so that they are not affected and abused by people economically good. <ICLE-TS-NOUN-0077.1>

Reformulation with a finite *be*-clause is not possible in the two examples below. In the first example the adjective *possible* is not a direct modifier of the noun *people*, but

involved in the quantification of *people*. In the latter example the coordinated adjectives, *young and old*, serve as a supplement.

(146) Like any other group of ideas or beliefs, the only way to reach the most people possible is through teaching others and then those people teaching other people and so on. <ICLE-US-MRQ-0005.1>

(147) HIV/AIDS is an incurable disease that can even kill people, young and old. <ICLE-TS-KIMC-0419.1>

Both postnominal adverbial groups that modify *people* in LOCNESS take the form *out there*. There are also two examples of *people out there* in the TLEC. The phrase *out there* emphasizes general existence. It denotes a general conceptual space in which people exist rather than specifying a specific place.

(148) However, I'm sure that many other people out there found it impossible to pass up such an offer. <ICLE-US-MRQ-0030.1>

(149) There are still people out there that believe women are not feminine because of their athletic ability. <ICLE-US-MRQ-0020.1>

(150) There are lot of educated people out there who will do any job they can find. <ICLE-TS-NOUN-0485.1>

(151) It is also important for those young people out there who are talented. <ICLE-TS-NOUN-0158.1>

All of the postnominal adverbial groups in the TLEC *people*-concordance are place adverbials, the most frequent being *there* (5x). When the adverb *there* is used on its own, it functions referentially as an anaphor of a previously described or named place.

(152) Most of the people there are unemployed <ICLE-TS-NOUN-0313.1>
[there = in informal settlements]

(153) Unlike, in Europe, people there, are not struggling that is why the European countries experience low rate of HIV aids. <ICLE-TS-NOUN-0061.1>

The LOCNESS writers use proportionally more postposed adjectival groups than the TLEC writers. Eight of the 11 postnominal word groups modifying *people* in the LOCNESS concordance are adjectival groups. The key difference between the corpora in terms of the use of postnominal word groups, lies in the high frequency with which appositive nominal groups are used in the TLEC (28x) to designate a more specific subclass of *people*.

4.5.3 Postnominal prepositional phrases

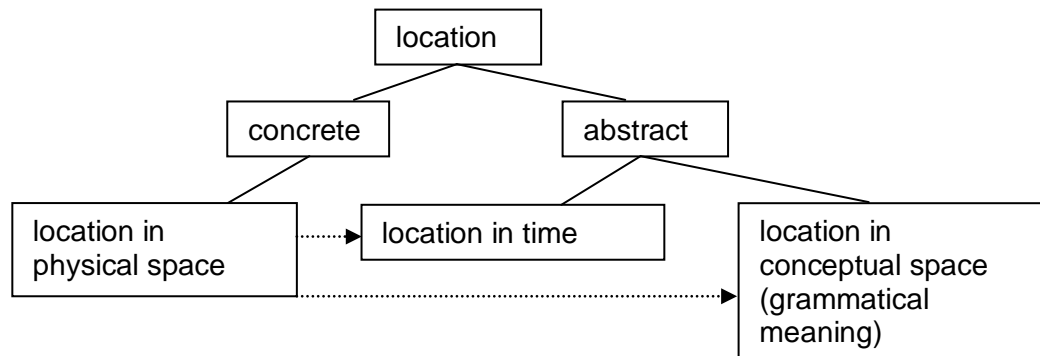
In both corpora, prepositional phrases comprise the majority of the non-clausal postnominal modifiers of *people*. There are 183 postnominal prepositional phrases in the TLEC concordance of *people* and 75 in LOCNESS. In the TLEC, 21% of the instances in the supercategory “qualified people” have prepositional phrases as qualifiers, compared to 32% in LOCNESS. As mentioned earlier, while phrasal modifiers are used proportionally less frequently, clausal postnominal modifiers are chosen proportionally more frequently in the TLEC than in LOCNESS (44% vs 30% of instances of “qualified people”). This is clearly a matter of more frequent selection of the clausal pattern and not of non-acquisition of the phrasal pattern. The absolute frequency of the phrasal pattern in the TLEC (183) exceeds that of LOCNESS (75), so there can be no implication that the mainly clausal options available in the substrate language constrain the L2 output, though it may play a role in the preference of the clausal pattern.

In this section the forms of the prepositions that introduce the postnominal modifiers of *people*, as well as their meanings will be discussed. In the rest of the discussion, counts of forms and prepositional meanings will be reported as a percentage of the total amount of postmodifying prepositional groups in each concordance of *people*. For example, 13% (i.e. 10/75) of the postmodifying prepositional phrases of *people* in LOCNESS consist of *with*-phrases. When the meaning relation established by the preposition is

discussed, the nominal core of the primary nominal group will be referred to as N1 and the nominal group which is the prepositional complement will be referred to as N2. For instance, in the nominal group *the Tswana-speaking people of South Africa*,⁶⁰ N1 refers to *Tswana-speaking people* and N2 to *South Africa*.

The concordance lines in question were further subclassified in terms of the prepositions used as well as the meanings conveyed. Each preposition encountered in the concordance data was classified into one of three broad semantic categories, namely basic (concrete) spatial meaning denoting place, extended spatial meaning (e.g. location in time) and grammatical (rather than lexical) meaning (Pullum & Huddleston, 2002b:647-653). Figure 4.3 below is an attempt to represent these semantic distinctions. The smallness of the data set did not warrant further subclassification of the various grammatical functions that a preposition might fulfill.

Figure 4.3 A categorization network of the meanings of prepositions



The prepositional phrases that modify *people* in the TLEC and LOCNESS and their functional meanings are set out in Table 4.10.

⁶⁰ <ICLE-TS-NOUN-0031.1>

Table 4.10 Prepositional phrases that modify *people* in the TLEC and LOCNESS

Preposition	TLEC (183 instances)			LOCNESS (75 instances)		
	Place	Abstract space (circumstance)	Grammatical function	Place	Abstract space (circumstance)	Grammatical function
<i>of</i>	18	2	4	35	0	2
<i>in</i>	86	7	2	3	8	3
<i>from</i>	14	4	0	4	0	0
<i>with</i>	0	0	17	0	0	10
other	16	0	13	3	0	7
TOTAL	134 (73%)	13 (7%)	36 (20%)	45 (60%)	8 (11%)	22 (29%)

In both corpora, postnominal prepositional phrases are employed most frequently to qualify *people* in terms of physical location, but relatively-speaking, even more frequently in the TLEC than in LOCNESS. 73% (134x) of the prepositional phrases in the TLEC concordance specify a physical place, compared to 60% (45x) in the LOCNESS concordance data. The 134 concordance lines in the TLEC where people are specified in terms of their location in a physical place together with the other 13 instances where people are specified in terms of some other circumstance (typically non-physical location) make up 80% of the instances where prepositional phrases function as postnominal modifiers of the noun *people*. In LOCNESS, 71% of the prepositional phrases that modify *people* realize circumstantial rather than grammatical meanings. The discussion below will focus on these circumstantial meanings and the prepositional phrases most frequently employed to express them, namely, *of*-phrases, *in*-phrases and *from*-phrases.

Apart from this quantitative difference, the two corpora also differ in the choice of preposition to convey these spatial meanings. In the TLEC, *in*-phrases (86x) are most frequently used to indicate the physical location of *people*, followed by *of*-phrases (18x) and then *from*-phrases (14x). In LOCNESS, the preposition most frequently involved in qualifying *people* in terms of 'place' is *of* (35x). In the LOCNESS data, *of*-phrases are followed by *from*-phrases (4 instances) and *in*-phrases (3 instances). In both corpora the prepositions *around* (9x in TLE; 1x in LOCNESS) and *on* (1x in TLE; 2x in

LOCNESS) are also used to introduce 'place'-specifying modifiers of *people*. In the TLEC, *across* (1x), *all over* (2x) and *at* (3x) also introduce phrases denoting 'place'.

In further exploring the differences between the two corpora regarding preposition choice, the focus will fall on the three most frequent prepositions, namely *in*, *of* and *from*. Of these three, *of* is the most neutral expression of geographic provenance, while *in* and *from* have more concrete positional meanings. 'Provenance'-specifying *of*-phrases greatly overlap in meaning with prenominal denotations of 'provenance'; compare *the people of Africa* and *African people*. However, not all *of*-phrases denoting geographical provenance can be used prenominally, as the following example illustrates:

(154) You look at the people of this continent you sow poverty in their eyes.
<ICLE-TS-NOUN-0107.1>

The preposition *in* evokes the notion of 'containment' of *people* within a bounded physical space, whereas the preposition *from* conveys 'source of movement' and is a more literal way of conveying 'origin' (Pullum & Huddleston, 2002b:648-650), often implying that the *people* in question are away from the source at the time of the discourse. The two examples below illustrate these basic meanings:

(155) There are no counsellings provided to people in hospitals, <ICLE-TS-NOUN-0070.1>

(156) People from neighbouring countries flock to Africa [...]<ICLE-TS-NOUN-0235.1>

In the LOCNESS data, 34 of the 35 'place'-specifying *of*-phrases are modifiers in definite nominal groups. Of these 34 occurrences of the pattern ***the people of N2***, the place name *Argos* realizes N2 in 24 instances. Since this place name is not native to English, it is not readily inflected so that it can be used prenominally,⁶¹ in the way that *British people* (2x) and *American people* (9x) might be used interchangeably with *the people of Britain* (1x) and *the people of America* (4x), respectively. The student writers' choice of an essay topic necessitating reference to the population of a place with a Greek name

⁶¹ the Argine population <ICLE-BR-SUR-0011.1>

therefore probably artificially heightened the use of postmodifying *of*-phrases, and thus of postmodifying phrases in general, in 24 instances in the *people*-concordance of LOCNESS. Note that this would not have had an effect on the frequency of the noun *people*, merely on the distribution of prenominal 'provenance' qualities versus postnominal *of*-phrases as 'geographical provenance' specifiers.

In the TLEC data, 13 of the 18 occurrences of the pattern **people of N2** are definite nominal groups, and of these 13, N2 is realized twice by *Africa* and once by *South Africa*. Of the 5 indefinite occurrences of **people of N2**, N2 is realized by *Africa* three times and by *South Africa* once. So, in only 7 of the 18 occurrences of the pattern **people of N2**, is N2 realized by a proper name. In the majority of instances of this pattern in the TLEC data (11x), N2 is realized by an anaphoric nominal group headed by a common noun. This contrasts with the LOCNESS data, where N2 is a proper name in the majority of cases, there being only 3 occurrences of the pattern where N2 is a nominal group headed by a common noun.

(157) People of this continent does not have money to pay for medication,
<ICLE-TS-NOUN-0018.1>

(158) the loving and hospitable people of this province <ICLE-TS-NOUN-0057.1>

In the TLEC concordance of *people*, there are also two *of*-phrases that convey 'circumstance', but where N2 does not denote a physical place. In the first example below *people* are specified in terms of their 'location in time' and in the latter in terms of their 'physical composition'.

(159) They are also places known as the battlesites, of which people of the past fought for the province against the boers and English who wanted to take it and use it for their own interest. <ICLE-TS-NOUN-0031.1>

(160) They do not want to behave like people of blood and flesh who have feeling instead choose to take animals positions. <ICLE-TS-NOUN-0332.1>

There are only 3 instances in LOCNESS where postnominal *in*-phrases qualify *people* with regard to their location in physical space. Of these, one occurs in a definite nominal group and two in indefinite nominal groups. In two of these, the prepositional complement (i.e. N2) is a proper name.

(161) The people in Argos wallow in their guilt. <ICLE-BR-SUR-0006.1>

(162) Though Madonna's novel was not considered to be a baimed text at this time, several people in Arizona did not want to even give it a chance. <ICLE-US-SCU-0015.4>

(163) Many of the people in jail are there for illegal drug or drug-related offenses <ICLE-US-SCU-0011.2>

The choice of *in* rather than *of* emphasizes that the people in question as well as the process in which they are involved ('wallowing in guilt', 'rejecting a book') occur within the boundaries of the named place. In the prepositional phrase *in jail*, N2 is a common noun denoting a physical place, but here the choice of preposition is idiomatically fixed and substitution with another preposition is not possible in this context. Though *jail* has a 'non-specific' interpretation, evoking any and/or all instances of the type *jail*, it still denotes a physical location.

The majority of the 86 *people*-groups with postnominal, 'place'-specifying *in*-phrases in the TLEC are indefinite (67x). In more than half of the 86 'place'-specifying *in*-phrases (44x), the noun, or nominal, that follows *in* (i.e. the prepositional complement, or N2) is a proper noun. The pattern **people in propN** occurs 44 times. *Africa* occurs 39 times and *South Africa* 5 times as N2 after *in*.

(164) People in Africa are not educated because of poverty. <ICLE-TS-NOUN-0245.1>

In the TLEC data, 32 of the instances of the pattern **people in propN** also contain some form of quantification. The following subpatterns are distinguished:

quant *people in* propN⁶² (18x)

(165) Poverty contribute to the escalating HIV/AIDS epidemic, considering the fact that most people in Africa are illiterate. <ICLE-TS-NOUN-0326.1>

quant of *people in Africa* (4x)⁶³

(166) Most of people in Africa are infected by Aids especially the young ones. <ICLE-TS-NOUN-0316.1>

most qual *people in Africa*⁶⁴ (2x)

(167) Most poor people in Africa don't like talking about sexual intercourse in the public, they think that, you will be encouraging immorality. <ICLE-TS-NOUN-0325.1>

quant of definite *people in Africa* (5x)

(168) According to my own view poverty causes aids in the sense that most of the people in Africa are poverty stricken [...]<ICLE-TS-NOUN-0248.1>

quant of definite qual *people in Africa*.⁶⁵ (3x)

(169) Many of the poor people in Africa are illiterate. <ICLE-TS-NOUN-0004.1>

In these 32 nominal groups where the noun *people* is quantified, the *in*-phrase cannot readily be substituted with an *of*-phrase since the quantification relies on a bounded space within which the intended instances of *people* can be “counted”. The high incidence of *in*-phrases compared to *of*-phrases in the TLEC concordance of *people* can therefore largely be explained in terms of quantification. Of the 41 ‘place’-specifying *in*-

⁶² quant=quantifier, propN=Proper Noun.

⁶³ These include two instances of the non-standard quantifying construction *most of people in Africa*

⁶⁴ qual=quality specifier (e.g. an adjective).

⁶⁵ definite=definite determiner, qual=quality specifier (e.g. an adjective).

phrases that have common noun-headed nominal groups as complements, there are 9 which occur in quantified *people*-groups, e.g.

(170) Most people in rural areas do not get enough information about HIV/AIDS <ICLE-TS-NOUN-0062.1>

In 'place'-specifying postnominal *in*-phrases where N2 is a common noun-headed nominal group (41x in the TLEC), the common nouns used are sometimes co-referential with *Africa* (continent 4x) and *South Africa* (country 5x) and thus establish lexical cohesion. There are 12 instances of the sequence *people in rural areas*, one instance of *people in those areas*⁶⁶ where *those areas* refer to rural areas, as well as the two examples below which specify *people* in terms of their location in rural as opposed to urban areas.

(171) What about the people in the rural villages or areas? <ICLE-TS-NOUN-0492.1>

(172) I personally think or plead to the Government to help those people in the rural areas because they suffer the most. <ICLE-TS-KIMC-0294.1>

The 'place'-specifying *in*-phrases discussed so far help to narrow down the sub-class of *people* being referred to, but there are also a few instances where locative *in*-phrases are used to emphasize the general meaning of *people*.

(173) All people in the universe leave their jobs because of low salaries. Everybody need money for himself to improve his lifestyle. <ICLE-TS-NOUN-0506.1>

(174) You can also ask yourself why especially you of all the people in the world and you could wish if you were sleep and never wake up some other time because of the challenge you have met. <ICLE-TS-NOUN-0082.1>

⁶⁶ <ICLE-TS-NOUN-0256.1>

In the examples above, the physical space in which people are contained is very large (*the universe, the world*). The largeness of the 'conceptual container' has the effect of generalizing rather than specifying the meaning of *people*. The next two examples represent other locative prepositional phrases that are used to generalize rather than particularize the denotation of *people* by relating them to a very large space, namely the world.

(175) they persevered and are a delight to many people across the world.
<ICLE-TS-NOUN-0066.1>

(176) We as the people of this world wherever the silly method is applied also have to look into things like our moral values [...]<ICLE-TS-NOUN-0011.1>

In the LOCNESS concordance of *people*, non-locational prepositional phrases are employed to emphasize generality, as the following examples illustrate.

(177) [...] development in astronomy affected the people as a whole. <ICLE-BR-SUR-0007.2>

(178) But he realises later that he is guilty of judging the people he deals with and people in general as he feels he dominates them. <ICLE-BR-SUR-0007.1>

There are also 8 instances in LOCNESS and the 7 instances in the TLEC where *in*-phrases convey 'containment' and therefore have locational meanings, but where the 'container' (i.e. N2) is not a physically delimited place such as a city or a country, but a more abstract location such as *society*.⁶⁷ The example below illustrates this use of *in*-phrases to locate *people* in a more abstract domain. The two corpora are fairly similar with regard to this construction.

⁶⁷ It should be noted that the semantic classifications did not rely on form only. Not all prepositional phrases following *people* are postmodifiers. Many are Adjuncts (adverbials), e.g. *there are more old people in society than young* <ICLE-BR-SUR-0017.1>

- (179) The government and people in the advertising industry, they are trying everything to make people aware of this killer disease. <ICLE-TS-NOUN-0313.1>

Three of the four 'place'-specifying *from*-phrases in LOCNESS occur in one essay⁶⁸ in which 'people from warm climates' are compared to 'people from cold climates'. The other instance is given below. Here the 'origin' meaning conveyed by the preposition *from* is clearly illustrated.

- (180) Many people not from the South do not understand the southern way of life, simply because they have never lived it before. <ICLE-US-SCU-0001.2>

In the TLEC, there are 14 instances where *from*-phrases specify 'physical place of origin' and 4 instances where they specify an 'origin' which is not a physical place. Only two of the 14 'place'-specifying *from*-phrases have a proper noun as complement (*Cape Town*, *Ghana*). The rest have nominal groups headed by common nouns as complement, for example:

- (181) People from neighbouring countries flock to Africa [...] <ICLE-TS-NOUN-0235.1>

- (182) People from rural areas are easily affected by the disease because of poor health and lack of facilities of use to prevent the disease. <ICLE-TS-NOUN-0275.1>

- (183) The people from rural places have clinics to consult to and nurses do go around such places making people aware of such a disease <ICLE-TS-NOUN-0005.1>

The examples below are representative of the use of *from*-phrases to specify provenance which is not geographic (4 instances in the TLEC).

⁶⁸ <ICLE-BR-SUR-0014.2>

(184) Most people from disadvantage family rely on them. <ICLE-TS-NOUN-0007.1>

(185) and people from health departments should visit poor villages several times to deliver this useful beneficial information for everybody <ICLE-TS-NOUN-0256.1>

As a common noun denoting concrete, countable entities, which are human, the noun *people* is a prototypical noun. It is therefore not surprising that the basic (original/prototypical) meanings of the prepositions in the prepositional phrases that are employed to modify the noun *people* are activated in the majority of the cases where *people* is modified by a prepositional phrase. While a comparison of prepositional phrases modifying *people* does not allow for a generalization of the meanings of prepositional phrases across the two corpora, it does highlight some interesting differences between the two corpora with regard to the way in which prepositional phrases are employed to modify *people*. The first of these is the preference for the use of *in*-phrases to denote geographic provenance of *people* in the TLEC, while *of*-phrases are preferred for this in LOCNESS. With regard to *of*-phrases as specifiers in terms of physical space, it is also interesting to note that the majority of the prepositional complements in the LOCNESS concordance are headed by proper nouns, while the majority of the nominal groups that complement *of* are headed by common nouns in the TLEC data. The biggest quantitative difference between the two corpora lies in the use of *in*-phrases to specify people in terms of spatial location: 86 instances in the TLEC compared to only 3 instances in LOCNESS. In 41 instances, the choice of an *in*-phrase rather than an *of*-phrase in the TLEC data can be attributed to additional quantity specification. The preference for *in*-phrases in the TLEC and the preference for *of*-phrases in LOCNESS can also be related to the basic meanings of the prepositions *of* and *in*. The preposition *of* indicates origin without profiling the boundaries of the place of origin, while the preposition *in* implies a bounded space due to its invocation of a 'container'. One can conclude that the writers of the TLEC prefer to specify *people* in terms of bounded spaces.

4.5.4 Postnominal qualities in the form of clauses

This section deals with patterns containing the sequence **people qual:clause**. While postnominal clauses are the most frequent form of modifier in the TLEC concordance of *people*, it is the least frequent in the LOCNESS concordance of *people*. The aim of this section is to account for the high frequency of postnominal clauses in the TLEC data.

Occurrences of *people* with qualifying clauses number 381 (184/1000 concordance lines; 43 % of “qualified *people*”) in the TLEC and 70 (94/1000 concordance lines, 30% of “qualified *people*”) in LOCNESS. This difference between the two corpora relates to the use of groups and prepositional phrases as postmodifiers. Disregarding the instances where there are both postnominal groups or phrases and clauses,⁶⁹ there are nearly two and a half times more clausal (376x) than non-clausal (154x) postmodifiers of *people* in the TLEC, whereas there are more non-clausal (85x) than clausal (69x) postmodifiers of *people* in LOCNESS. It should be noted that postnominal clauses are the most frequent type of modifier in the TLEC concordance of *people*, occurring in 43% of the instances of “qualified people” (40% of all modifiers of *people*), followed by prenominal word groups (41%) and then by postnominal phrases (25%).⁷⁰ Information is presented in a less dense and less integrated manner in the TLEC compared to LOCNESS. In LOCNESS, prenominal word groups are the most frequent form of quality specifier of *people* (41%), followed by postnominal groups or phrases (37%) and then by postnominal clauses (30% of instances of quality-specified *people*; 28% of all modifiers of *people*). As shown in Table 4.6, the same trend can be observed if percentages of all modifiers of *people* are taken into account. So, whereas postnominal clauses are the most preferred qualifier of *people* in the TLEC, it is the least preferred form in LOCNESS. Percentage-wise, the biggest difference between the two corpora lies in nominal groups where postnominal clauses are the only quality specification of *people*, i.e. in patterns containing **people qual:clause**.⁷¹ The two corpora are very similar with regard to the form frequency rankings of the clausal postmodifiers of the noun *people*. The three most frequent patterns in both corpora are:

⁶⁹ 5x in TLEC and 1x in LOCNESS

⁷⁰ These percentages of “qualified *people*” add up to a number higher than 100 (viz. 108), because of the nominal groups that have both pre- and postmodifiers.

⁷¹ 334 in TLEC vs 59 in LOCNESS

∅ people qual:clause (158 in the TLEC and 30 in LOCNESS)

(186) People who are poor are likely to die very quickly of the epidemic than people who are well off financially <ICLE-TS-NOUN-0516.1>

def people qual:clause (78 in the TLEC and 15 in LOCNESS)

(187) Even though we believe that Aids has no class it affects those people who are poor. <ICLE-TS-NOUN-0165.1>

quant people qual:clause (41 in the TLEC and 9 in LOCNESS)

(188) There are many people who are poor in Africa <ICLE-TS-NOUN-0302.1>

Table 4.11 provides a form-based overview of the postnominal clauses in the *people*-concordances. The majority of the clauses that modify *people*, in both corpora, are finite: 336/381 in the TLEC and 53/70 in LOCNESS.

Table 4.11 Postnominal clauses as modifiers of *people* in the TLEC and LOCNESS

	TLEC		LOCNESS	
	Raw frequency	Percentage of postnominal clauses	Raw frequency	Percentage of postnominal clauses
<i>ing</i> -participle clause	33	9	12	17
<i>ed</i> -participle clause	10	3	5	7
<i>to</i> -infinitive clause	2	<1	0	0
unmarked finite clause	10	3	6	9
finite <i>who</i> -clause	307	81	37	52
finite <i>that</i> -clause	11	3	8	11
finite <i>which</i> -clause	4	1	0	0
finite <i>whose</i> -clause	0	0	1	1
finite <i>than</i> -clause	0	0	1	1
finite <i>whom</i> -clause	4	1	0	0
Total postnominal clauses	381	-----	70	-----

In both corpora, the most frequent clause form used as postnominal specifier of *people* is a finite relative clause introduced by *who*, but the corpora differ with regard to the extent to which *who*-clauses are preferred as postmodifiers. In the TLEC concordance

of *people* 81% (307x) of the postmodifying clauses are *who*-clauses. In LOCNESS 53% (37x) of clauses postmodifying *people* are *who*-clauses. In both corpora, the pattern \emptyset ***people qual:who-clause*** (i.e. a subpattern of \emptyset ***people qual:clause*** and of ***people who-clause***) is the most frequent pattern, occurring 131 times (34% of postnominal clauses) in the TLEC and 18 times (25% of postnominal clauses) in LOCNESS.

Nearly a quarter of the postmodifying *who*-clauses in the TLEC *people*-concordance are copula clauses (75 of 307, i.e. 24%), i.e. contain the pattern ***people who copV***.⁷² Just more than a quarter of the *who*-clauses in the *people*-concordance of LOCNESS (10 of 37, i.e. 27%) have a copula verb (the verb *are* in each case). In the TLEC data there are also one postmodifying *which*-clause with a copula verb, one postmodifying copula clause introduced by *whom* and two copula *that*-clauses. Here the choice of *whom* might be a case of hypercorrection. In Standard English *who*, not *which*, is typically used as relative pronoun after nouns denoting human beings or entities construed as 'human'. In other words, the choice of *wh*-relativizer is restricted by the meaning of the preceding noun. The four instances of *people which* in the TLEC do not constitute enough evidence to conclude that the same restriction does not apply in BSAfE, especially since two of the instances occurred in the same essay. However, a quick concordance and count of the words *which* and *who*, with no regard to their syntactic function, indicates that the *which:who* ratio in the TLEC (1.8)⁷³ is higher than in LOCNESS (1.3)⁷⁴ and that there are altogether more *wh*-pronouns in the TLEC than in LOCNESS.⁷⁵

The prevalence of postmodifying clauses in the TLEC concordance of *people* requires further investigation into the extent to which clausal modifiers are used in contexts where the quality could have been realized by a word group or phrase. This pertains especially to clauses representing ascriptive relational processes, since prenominal adjectival groups can also fulfill the function of ascription. In all 10 instances of copula relative clauses in the LOCNESS data, the predicative (i.e. Complement) position of the quality was grammatically required or at least justified. To illustrate, in the first example below,

⁷² copV = copula verb

⁷³ Concordance on untagged version of the corpus of only correctly spelt instances of the words yields 1767 *which* and 970 *who*.

⁷⁴ *which* 724, *who* 542.

⁷⁵ In the manually analyzed concordances of *that*, there are 682 occurrences of *that* as relative pronoun in the TLEC and 667 in LOCNESS.

the adjective has a complement (*in clothes*), while in the last two examples in the set below, the Complement of the copula clause is realized by a prepositional phrase.

(189) There tends to be a link between people who are very interested in clothes and the fact that this type of person is highly socially oriented <ICLE-US-SCU-0009.3>

(190) This means that people who are in need of real civil relief have to wait <ICLE-US-SCU-0016.4>

(191) Robert Rygor's words are fitting for those people who are against the idea. <ICLE-US-MRQ-0021.1>

Similar justifications for the use of a relative clause where the quality is situated in the Complement of the copula verb obtain for the majority of the 79 copula relative clauses modifying *people* in the TLEC, as illustrated in the set of examples below. In the first example, the adjective (*able*) has a complement. Here the principle of end-weight is at work. In the second example, the adjective (*aware*) is one which is restricted to predicative position only and in the last the Complement of the copula verb is a noun.

(192) The new introduced Internet banking only accommodate people who are able to afford computers. <ICLE-TS-NOUN-0492.1>

(193) People who were otherwise "aware" could have been misled into letting down their guard in matters of protection while the issue was being debated. <ICLE-TS-NOUN-0004.1>

(194) There are people who are now prostitutes but who did not choose to be so but they were forced by poverty to do that. <ICLE-TS-NOUN-0248.1>

As noted in Section 4.5.1 above, the adjective *poor* is the most frequent prenominal adjective in the TLEC concordance of *people*, but it also occurs as Complement in

postnominal clauses. Reformulations of the examples below substituting the sequence *people who are poor* with *poor people* are possible.

(195) In most cases people who are poor are not educated because they cannot afford to pay for education <ICLE-TS-NOUN-0239.1>

(196) People who are poor, if they have HIV/AIDS they die quickly unlike the rich. <ICLE-TS-NOUN-0249.1>

(197) There are many people who are poor in Africa <ICLE-TS-NOUN-0302.1>

Apart from the examples with *poor* above, there are at least 15 more examples where adjectives occur as complements in *who*-clauses where they might have premodified *people*, for example:

(198) They didn't know that there are people who are white. <ICLE-TS-NOUN-0503.1>

The example above can be perceived as non-standard due to the fact that the adjective has a classifying (viz. 'provenance: ethnic') and not an epithetic function – classifiers typically occur preminally. However, the postnominal copula *who*-clauses in the TLEC data are generally not ungrammatical, nor non-standard. The data simply indicates selection of the fullest modifier form where a denser form would be preferred in Standard English. This is also evidenced in the 4 cases in the TLEC where the complement of the copula *who*-clause is a prepositional phrase. In the examples below *who are* can be omitted.

(199) People who are at the top level having jobs must look at those who are having nothing. <ICLE-TS-POT-0198.1> [people at the top level who have jobs]

(200) For me, that shows that we have humans in there, we have people who are like us, who have hearts and want to repent.<ICLE-TS-NOUN-0157.1> [people like us]

(201) People who are from outside countries may not be allow to stay in our country <ICLE-TS-NOUN-0081.1> [people from outside]

(202) People who are in prison are given all sorts of punishments <ICLE-TS-NOUN-0510.1> [people in prison]

However, there are also 2 instances like these in the LOCNESS *people*-concordance:

(203) This means that people who are in need of real civil relief have to wait <ICLE-US-SCU-0016.4> [people in need of real civil relief]

(204) Robert Rygor's words are fitting for those people who are against the idea. <ICLE-US-MRQ-0021.1> [those against the idea]

As can be seen in Table 4.8, the majority of postmodifying clauses in the *people* concordances of both corpora are finite (336/381 in the TLEC and 53/70 in LOCNESS). Since the choice of a finite clause as modifier instead of a non-clausal modifier is often motivated by the fact that finiteness is required for negation and/or modality marking, some attention to the grammatical structure of the verbal group is necessary to explain the preference for finite relative clauses in the TLEC data. In the LOCNESS *people*-concordance, there are four *who*-clauses that are marked for modality. Twenty nine of the finite clauses functioning as postnominal modifiers of *people* in the TLEC are motivated by modality marking.

The finite also facilitates polarity marking. In addition to the negation accompanying the one instance of *should* in the LOCNESS data, there is one instance of a postmodifying clause of *people* which is marked for negative polarity, and one instance where the auxiliary *do* is used to emphasize positive polarity.

(205) You would be very surprised at the amount of people, who **don't** use their brakes [...] <ICLE-US-SCU-0002.4>

(206) In fact many people who **do** support the death penalty, <*>. Even those on the side of capital punishment acknowledge its flaws. <ICLE-US-MRQ-0016.1>

Except the three modality-marked postnominal clauses that are negated in the TLEC concordance of *people*, there are 20 more finite postmodifying clauses that are marked for negative polarity and one where the auxiliary *do* emphasizes positive polarity, as exemplified below.

(207) There are many South African people who **do not** work. <ICLE-TS-NOUN-0067.1>

(208) What about the people who **do** have the necessary information, but they are still doing what the information we are given tells us not to do <ICLE-TS-NOUN-0093.1>

Of the 20 negated postmodifying finite clauses in the TLEC concordance of *people*, there are 7 instances where the postnominal clause is also marked for progressive aspect. Remarkably, *working* is the lexical verb in each of these seven instances, which come from seven different essays. In the six examples below the verb *working* denotes a 'state' (viz. social status), rather than an 'activity' (i.e. atelic, non-conclusive dynamic process) (Vendler, 1957; Van Rooy, 2006:41), which can also be expressed by a prenominal participial adjective (e.g. *unemployed/employed*) but not by a non-finite postnominal clause, cf. **people not working do not have food to eat*.

(209) There are lot of people who are not **working**, claiming that there is no work, saying that the government should create job's. <ICLE-TS-NOUN-0306.1>

(210) People who are not **working**, do not have food to eat. <ICLE-TS-POT-0198.1>

(211) People who are not **working** they are suffering and they want to live a better life like other's; <ICLE-TS-NOUN-0311.1>

(212) these people who are not **working** are the ones who are involved in some activities like being prostitutes in order to make a living. <ICLE-TS-NOUN-0045.1>

(213) Most of the people who are not **working** they drink a lot. <ICLE-TS-NOUN-0217.1>

(214) Poverty is the cause of Aids because there are many people who are not **working** in South <ICLE-TS-KIMC-0421.1>

In the next example *working* denotes an 'activity' and here it is possible to omit the *who are*, i.e. substitute the finite relative clause for a non-finite *-ing*-clause.

(215) Poverty is the cause of HIV/AIDS EPIDEMIC IN Africa because the most affected are the people who are not **working** or living in the disadvantage areas. <ICLE-TS-POT-0191.1>

There are also two instances in the TLEC where *working* is the main verb of a finite *who*-clause (as modifier of *people*) not marked for polarity (i.e. positive). In both these instances, the lexical verb denotes an 'activity'.

(216) It is because that they do not have money to buy their own things, so they are stealing from people who were **working** so hard to put food on their tables. <ICLE-TS-NOUN-0438.1>

(217) Again, there is Manyane Game Lodge where you can get different kinds of animals and birds. The people who are **working** there can explain thoroughly to the students and people who tours. <ICLE-TS-POT-0194.1>

In Standard English reformulations of the two examples above, the aspectually unmarked form of the verb (*work*) would be used. The present tense would also be preferred over the past tense in a Standard English reformulation of the first example. Van Rooy demonstrates that the use of the progressive is extended to dynamic verbs with a general 'continuous' interpretation, i.e. those that extend over a long time and express a durational attribute (2006:57), as the examples of *working* illustrate.

There are 38 *people*-groups in the TLEC where the finite *who*-clauses are marked for progressive aspect⁷⁶ as opposed to only two finite postnominal clauses with an *-ing* verb in the LOCNESS concordance of *people* (exemplified below).

(218) Simply because that there are few people who are **thinking** of taking their own lives and are also of sane mind. <ICLE-US-MRQ-0033.1>

(219) they also take in to account the types of people who will be **reading** the essay <ICLE-US-MRQ-0041.1>

The comparatively high number of finite progressive clauses functioning as modifiers of *people* in the TLEC (38) is not surprising in the light of Van Rooy's observation that "the progressive aspect is used almost twice as often in the TLEC compared to the reference corpus LOCNESS" (2006:47). The most frequent lexical verbs occurring in these finite progressive clauses are *living* (10x) and *working* (9x), which account for half of the 38 instances. The senses of *working* have been discussed above. The verb *living* is also used in more than one sense, though both senses can be interpreted as 'static'. In six of the ten instances where *living* is the verb in the finite progressive clause, *living* denotes 'existence' and is followed by an Adjunct (Circumstance) representing the state in which the people in question find themselves. This sense is represented by the three examples below.

(220) You find that people who are affected by AIDS are people who are **living** under poverty datum line. <ICLE-TS-NOUN-0326.1>

⁷⁶ This number excludes instances of the semi-modal pattern **be going to**.

(221) Nowadays we cant just fold our arms or live on food which we get from the field crops as our people who were **living** before us <ICLE-TS-NOUN-0046.1>

(222) If we take an example of developed countries like Europe and the United states of America there is a very low percentage of people who are **living** with HIV/aids <ICLE-TS-NOUN-0282.1>

In the remaining four cases, the existential meaning of *living* is extended to denote 'residence' and is followed by an Adjunct (Circumstance) of 'place', as can be seen in the example below:

(223) I think people who are living in rural areas must be educated about sex and sexually transmitted deseases. <ICLE-TS-KIMC-0294.1>

There are also ten instances where the main verb of the postnominal clause is *live*, giving a total of 28 postmodifying clauses of *people* in the TLEC where the main verb is realized by a form of the lemma *live*. After forms of the lemma *be* (80x), *live* is the most frequent lemma functioning as main verb in the clauses that modify *people* in the TLEC. It is followed by the lemmas *have* (27x), *work* (20x), *suffer* (10x), *die* (9x) and *do* (9x).

In this subsection, the formal and semantic types of postnominal clauses that modify the noun *people* were considered. In both corpora, the most frequent type of clause as postmodifier of *people* is a finite *who*-clause. However, finite relative clauses are much more frequent in the TLEC than in LOCNESS. While the choice of some of the finite relative clauses in the TLEC data can be attributed to the expression of modality and/or polarity, the information contained in many of these can also be expressed in denser constructions with premodifiers or postnominal phrases. The use of the progressive aspect in these finite postnominal clauses confirm the extension of the progressive aspect in BSafE noted by Van Rooy (2006). The process types expressed by the postnominal clauses modifying *people* are chiefly relational processes (lemmas of *be* and *have*) and existential processes involving *living* and *working* in the sense of being employed.

4.5.5 Summary

Per 1000 concordance lines, there are more instances of *people* specified for quality in the TLEC than in LOCNESS (472 vs 313). In this section, the modifiers that specify *people* in terms of quality were compared with reference to their forms and meanings. With regard to the forms and positions of modifiers of *people*, the TLEC and LOCNESS are quantitatively very similar with regard to both prenominal and postnominal word groups, but quite different with regard to postnominal prepositional phrases and clauses. LOCNESS makes slightly greater use of postnominal prepositional phrases than clauses (75 vs 70) but compared to the TLEC the load is spread fairly evenly between prepositional phrases and clauses as modifiers of *people* in LOCNESS (30% and 28% of modifiers). With regard to postnominal modification of *people*, there is a clear preference in the TLEC for clauses rather than prepositional phrases (381 vs 183).

Following Tucker (1998) a distinction was made between relative qualities and thing-oriented qualities (Classifiers and Epithets) realized by prenominal word groups in Section 4.4.1. With regard to relative qualities, the most striking finding was that the word *other* is not used as contrastive relative quality of *people* in 19 of the 56 instances where it precedes *people* in the TLEC. In these 19 cases, *other* is used as an indefinite determiner synonymous to *some*. This is attributed to substrate language influence, since *other* and *some* are rendered by the same Setswana word, which in turn sheds some light on the *other...other* construction in BSAfE (Van der Walt & Van Rooy, 2002:124; De Klerk, 2003a:470). The discovery of this indefinite, rather than contrastive, usage of *other* in BSAfE is a result of starting with the usage data and not the semantic categories. This demonstrates the efficiency of an exemplar-based approach over a systemic approach in uncovering meaning extensions.

Initial classification of the concordance lines with regard to thing-oriented prenominal qualities, indicated that an additional category between descriptive (epithetic) and classifying qualities is necessary to properly account for the qualities of *people*. This category comprises adjectives that denote 'social status'. With regard to thing-oriented qualities the biggest difference between the two corpora lies in the specification of

people in terms of social status. There are 109 such quality specification in the TLEC (mostly indicating economic status) compared to only 3 in LOCNESS.

With regard to postnominal groups, the main difference between the corpora lies in the 28 uses of appositive nominal groups in TLEC compared to only one in LOCNESS. While the one instance of the appositive nominal group in LOCNESS is synonymous with the *people*-group (*big people or adults*), the 28 appositions of *people* in the TLEC are hyponyms of *people*. This demonstrates a linear, rather than syntactically integrated presentation of classifying information.

In both corpora, preposition phrases most frequently specify *people* in terms of a physical location. A similar trend is observed for prenominal classifiers, which typically specify *people* in terms of geographic provenance in both corpora. The corpora differ with regard to the preposition used to relate *people* to a physical space. There are 86 *in*-phrases in the TLEC concordance data that specify the physical location of *people*, compared to only 3 in LOCNESS. The choice of *in* rather than *of* shows a preference in the TLEC for specifying *people* within a bounded space, which in turn allows for quantification. In 41 of the instances in the TLEC where the noun *people* is qualified by an *in*-phrase, it is also quantified.

4.6 Quantified *people*

In the analyses of the determiners and quality specifications of *people* in Sections 4.4 and 4.5, a number of instances of non-standard quantification in the TLEC were observed. These will be discussed in more detail in this section. As was the case with definite and quality-specified instances of *people*, there are many more instances of quantity-specified *people* in the TLEC than in LOCNESS in terms of raw frequency: 574 versus 199. However, when these raw numbers are relativized to a frequency per thousand concordance lines, the distribution of quantity-specified instances of *people* versus instances not specified for quantity is close to the same for the two corpora: 264/1000 in the TLEC and 266/1000 in LOCNESS.

Quantity is normally expressed prenominally. A distinction can be made between indefinite determiners which imply quantity (e.g. *some people*) and explicit indications of quantity (e.g. *many people*). The first kind corresponds to the category of central determiners (Quirk et al., 1985:253-255), or non-specific Deictics in SFG (Halliday & Matthiessen, 2004:315), and cannot co-occur with other central determiners (cf **some the people*, **the some people*). The universal quantifier *all* can be used as an indefinite central determiner (*all people*) or as a predeterminer (*all the people*). The latter kind of quantifier belongs to the category of postdeterminer (Quirk et al., 1985:261), or the Numerative slot in SFG (Halliday & Matthiessen, 2004:318), and may be preceded by the definite article (*the two/many people*), but not by quantifying central determiners. Quantifiers that occur in determiner or postdeterminer position (Deictic or Numerative slot) will be referred to as determinative quantifiers to distinguish them from quantifying partitive *of*-constructions, as in *some of the people* (see Table 4.12).

Table 4.12 Quantity specifications of *people* in the TLEC and LOCNESS

		TLEC		LOCNESS	
		Raw frequency	% of quantity-specified instances	Raw frequency	% of quantity-specified instances
Determinative	<i>most</i>	71	13	16	8
	<i>many</i>	131	24	64	32
	<i>some</i>	93	17	34	17
	cardinal numbers and fractions	20	4	17	9
	other quantifiers	56	10	32	16
	SUBTOTAL: Prenominal quantifiers	371	68	163	82
Partitive	<i>most</i>	89	16	0	0
	<i>many</i>	3	<1	2	1
	<i>some</i>	15	3	0	0
	cardinal numbers and fractions	17	3	9	5
	other quantifiers	52	10	25	13
	SUBTOTAL: Quantifying of-partitions	176	32	36	18
<i>TOTAL quantifiers</i>		<i>547</i>	<i>----</i>	<i>199</i>	<i>---</i>

In both corpus concordances of *people* there are more determinative than partitive quantifiers. However, there are proportionally fewer non-partitive quantifiers of *people* in the TLEC compared to LOCNESS: 68% vs 82%. In other words, the TLEC writers make much more frequent use of partitive quantification than the LOCNESS writers. The most frequent quantifier of *people* in the TLEC is *most*, occurring 160 times. It is followed by

many (134x) and *some* (108). In LOCNESS, the word most frequently used to quantify people is *many* (66x), followed by *some* (34x). There are 16 instances in LOCNESS where *most* is used to quantify *people* (see Table 4.9).

In the TLEC concordance of *people*, *most* is used more often in an *of*-partition than in a (post)determiner position (89 vs. 71), while *many* and *some* are used as determinative quantifiers of *people* more often than they are used in partitive constructions. In LOCNESS, the word *most* never occurs in a partitive construction as a quantifier of *people*. However, a concordance retrieved for the search string *most of the* yielded twenty instances of this quantifying construction in LOCNESS. The most frequent sequence found in this concordance was *most of the time* (7x). Six of the concordance lines are quantifications of nouns denoting human beings: *most of the patients in the hospitals*,⁷⁷ *most of the population* (2x),⁷⁸ *most of the professional football players*,⁷⁹ *Most of the pupils leaving the Lycée technique*,⁸⁰ *Most of the opponents*.⁸¹ None of these nouns are as general in meaning as the noun *people*.

Of the 89 uses of *most* in *of*-partitions in the TLEC data, 70 instances of *people* are marked as definite. In the TLEC concordance data, just more than half of the partitive quantifying constructions contain a definite determiner before *people* (94x). The majority of *people*-groups that complement partitive *of* in quantifying partitions are indefinite in LOCNESS (30/36), the most frequent sequence being *a lot of people* (5x). This sequence also occurs 10 times in the TLEC data, along with one occurrence of *lots of people*.

In Standard English, the second nominal group (N2) in certain quantifying partitive *of*-constructions, e.g. those with *most*, *some* and *many*, has to be definite. In other words, if the partition has a quantifier like *most*, the pattern is **quant of def N2**.⁸² In this regard, two observations about the TLEC data can be made. Firstly, the standard pattern for

⁷⁷ <ICLE-US-MRQ-0019.1>

⁷⁸ <ICLE-US-MRQ-0006.1>; <ICLE-BR-SUR-0016.1>

⁷⁹ <ICLE-US-SCU-0008.2>

⁸⁰ <ICLE-BR-SUR-0016.1>

⁸¹ <ICLE-US-MRQ-0038.1>

⁸² quant = quantifier ; def = definite; N = nominal (i.e. noun plus internal modifiers or complements if there are any). The quantifier is the itself the first nominal (i.e. implied N1), the nominal in the prepositional complement is N2.

quantifying *of*-partitions is observed in the majority of the cases (70/89), but there is also a pattern emerging where *people* (i.e. N2) has no overt determiner and is therefore not marked as definite (19/89). Secondly, the LOCNESS concordance data show that the pattern **quant of def N2** is not the prototypical way of quantifying *people* in native English. The noun *people* is typically quantified by determiners or postdeterminers and not marked as definite in LOCNESS, as in *most people*. These two observations give rise to the following question: What motivates the use of the non-typical pattern **quant of def N2** rather than the pattern **quant N** in the TLEC? Two possibilities will be explored with reference to the use of *most* in the quantification of *people*. It is possible that a definite determiner is required in the unfolding discourse for referent-tracking purposes. On the other hand, it is also possible that *most* triggers the insertion of *of*, i.e. that expressions such as *most of the people* have become conventionalized in BSAfE in contexts where *most people* would be used in Standard English.

Of the 70 instances where *people* is marked as definite and quantified by means of the quantifier *most* in an *of*-partition, 30 instances are without other modifiers and 40 are specified for quality (i.e. contain modifiers). Three subpatterns of **most of def people** without quality specification are distinguished in the TLEC concordance data. The definite article is chosen in the majority of the instances in this subcategory, as can be seen from the first pattern.

***most of the people* (23x)**

(224) People happen to know that most of the people does'nt use their banks to bank their money <ICLE-TS-NOUN-0229.1>

In this example the most general, 'society' meaning of *people* is activated. With these general interpretations of *people*, a non-partitive quantifying construction (*most people*) would be used in Standard English. Here, the choice of the definite article is not motivated by the need to mark the referent of the quantified subset of *people* as identifiable. Unlike the definite article, the demonstrative and possessive determiners have clear identifying uses in the TLEC.

most of poss people (5x)⁸³

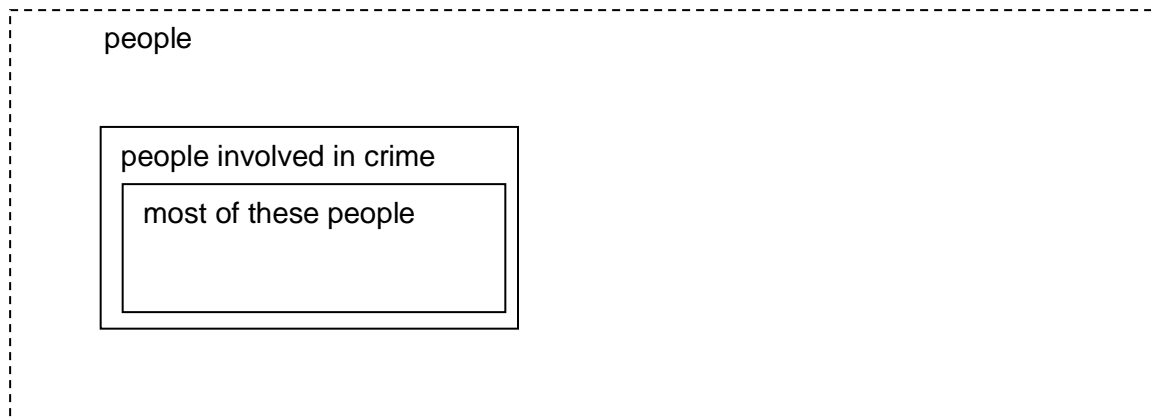
(225) In Africa most of our people are not working. <ICLE-TS-KIMC-0363.1>

The possessive determiner is used to identify *people* in relation to other person(s). In the example above *our* relates *people* to *Africa* and to the essay writer who expresses his/her own solidarity with this population. The expression of this personal identification is not possible in a non-partitive quantifying construction. In the example below, *these* points to a subclass of *people* defined earlier in the discourse (*people involved in crime*), and delineates a slightly smaller subset of the earlier subset.

most of these people (2x)

(226) In many situations where you find people involved in crime, most of these people dont have an interesting future. <ICLE-TS-KIMC-0334.1>

Figure 4.4 Subsets of *people*



The following patterns occur amongst the 40 instances where *most of* quantifies definite, quality-specified *people*. The definite article is chosen as definite determiner in 35 of the instances. The most frequent pattern is one with a definite article and a postnominal clause.

⁸³ poss=possessive. poss:*our* 4x; poss:*its* 1x.

most of the people qual:clause (20x)

(227) Most of the people who are affected by AIDS are the youth who must be taught on how dangerous this disease is, <ICLE-TS-NOUN-0244.1>

Other patterns with the definite article are:

most of the qual people (6x)

(228) Most of the South African people are poor, especially the Black people. <ICLE-TS-NOUN-0075.1>

most of the people qual:phrase (5x)

(229) Most of the people in rural areas had no access in schooling. <ICLE-TS-NOUN-0150.1>

most of the people qual:group (2x)

(230) Most of the people there are unemployed, <ICLE-TS-NOUN-0313.1>

most of the qual people qual:group (1x)

(231) Most of the unemployed people, more especially women practices promiscuous which is the sleeping around with too many men. <ICLE-TS-NOUN-0333.1>

most of the qual people qual:phrase (1x)

(232) Most of the poor people in Africa are the people who can't afford to put a loaf of bread on the table. <ICLE-TS-NOUN-0266.1>

In the 35 instances of patterns with the definite article the possibility that *people* is marked as definite due to identifying information provided in the modifiers was considered, but as the examples above illustrate, the definite article is not motivated by referent identification (also see subsection 4.4.3). Standard English reformulations would still allow a non-partitive, indefinite quantifying construction, e.g. *most poor people in Africa*. However, the four instances where a possessive determiner (*our*) and the two instances where a demonstrative determiner (*these*) are used require partitive quantification.

(233) I've notice that most of our people in Africa a poor. <ICLE-TS-KIMC-0290.1>

(234) Most of these people who take advantage of such puerty victims are HIV/AIDS positive. <ICLE-TS-NOUN-0034.1>

At this point, one can conclude that the majority of the cases of partitive quantifications of definite-marked *people* with the quantifier *most* cannot be attributed to the discourse function of the definite article as marker of referent identifiability. In Section 4.4.3 the use of quantifying partitions such as *most of the people* rather than *most people* is ascribed to morpheme insertion, a type of anti-deletion identified by Mesthrie (2006). A closer look at the 58 instances of the sequence *most of the* in TLEC *people* concordance confirms this explanation. Although *most* is the quantifier most frequently used in *of*-partitions in the *people* concordance (89x), there are also 87 instances of partitions in which other quantifiers occur (see Table 4.11), indicating a much wider use of quantifying partitions in the TLEC than in LOCNESS. The use of quantifying partitions is illustrated in the following essay below, which is quoted in full. Nominal groups where Thing element is realized by *people* are underlined. Other nominal groups with quantifying *of*-partitions are in boldface.

(235) Most of the South African people are poor, especially the Black people. **Most of them** who stay in the rural areas. In South Africa Poverty is caused by unemployment and many of the Blacks are not educated. In rural areas, **many of the blacks** are not educated because they did not have school and enough money to sent their children to school. In urban area they are educated but not all. Most of the people who stayed in the urban area are employed because they have skills of working. *Many people in rural areas* they are not working because in their places they did not have factories and other facillities. In Urban areas they do have job opportunities because they stayed in town. Most of the people who have HIV/AIDS are in rural areas. Most of the people in rural areas they did not have the stable relationship. Aids is essentially a disease caused by the relationship between people, which in turn is influenced by many social and economic factors. Maintaining stable, lasting and trusting relationship is at the heart of the control of the epidemic. The past, present and the future force and influence the community and family life in South Africa have and will continued and dertemine impact of the epidemic. The highest rate of HIV/AIDS infection is amongst women in South Africa, *most* in rural areas. Teenage girls are highly infected because

they fall in love with older man. Also in Urban areas **most of women and 14-15 years** are infected.

The reasons for women to be infected is that they make sure that their children eat everyday. They end up being prostitutes, because of the Unemployment. The government must do something about the HIV/AIDS, or else creates jobs because most of the people who have HIV/AIDS are blacks.

Aids is cause by poverty so the goverment should find the cure for Aids. They have the virus, because they did not have anything to support their families.

To say Aids is related to poverty is a big mistake. Aids is not caused by poverty but by men who failed to control their sexual behaviour. Men should lear to have control and respect for women. <ICLE-TS-NOUN-0075.1>

The 19 instances where *most* is used in a quantifying partition where *people* is not marked as definite now come under consideration. Not all partitive quantifying constructions require a definite determiner for N2; for example, *a lot of people*, *millions of people*, *lots of people*. However, as mentioned above, in Standard English, when *most*, *many* and *some* are used in partitive constructions, it is assumed that a quantity of a definite (identifiable) set of entities is meant, which precludes expressions like *most of people*. The following such patterns emerge from the TLEC data:

most of people (9x)

(236) In our country most of people are infected because they sell their body, because of lack of money and food. <ICLE-TS-POT-0422.1>

most of qual people (4x)

(237) Most of poor people go around looking for money <ICLE-TS-NOUN-0302.1>

most of people qual:phrase (3x)

(238) Most of people in Africa are infected by Aids especialy the young ones. <ICLE-TS-NOUN-0316.1>

most of people qual:clause (x3)

(239) Morespecially in Botswana, most of people who are doctors are from different countries <ICLE-TS-NOUN-0275.1>

With regard to these examples with quality specification, one might consider the possibility that the postnominal quality is providing identifying information and that the definite article is deleted, but still effecting partitive quantification. However, the qualities in most of these examples are quite general and still denote very large subsets of *people*. More importantly, this does not explain the 9 occurrences of the sequence *most of people* without any quality specification, as well as one instance of *more of people* in the TLEC.

The use of *most* in partitive *of*-constructions should be considered against the background of quantifying *of*-partitions in general. Of the 36 partitive quantifying constructions in the LOCNESS *people* concordance, only 6 quantify definite *people*. In another 6 of these partitive constructions in the LOCNESS data, the quantity is marked as definite, but the noun *people* has no determiner.

(240) **many** of **these** people <ICLE-US-SCU-0011.2>

(241) **the number** of people executed since 1976 <ICLE-US-MRQ-0011.1>

In the TLEC there are also 5 occurrences of *lot* as quantifier in *people*-groups that are further specified for quality. Two of these occur without the article *a*.

(242) There are **lot of** people who are not working <ICLE-TS-NOUN-0306.1>

(243) There are **lot of** educated people out there who will do any job they can find. <ICLE-TS-NOUN-0485.1>

In a concordance of *lot of*, there are altogether 26 instances where *a* is omitted before *lot* compared to 112 instances where *a* is retained. The omission of *a* before *lot* (e.g. *lot of people*) and the omission of *the* before *of* (e.g. *most of people*) may be an effect of the frequency of partitive quantifying constructions in BSAfE. High frequency phrases, such as *most of the* and *a lot of*, are more susceptible to phonological reduction than lower frequency phrases (Bybee, 2006:714). The phonologically and semantically less salient elements (*a* and *the*) of these high frequency sequences are the ones that are reduced.

In the patterns below, the less salient elements, i.e. those that are susceptible to reduction in BSAfE are underlined:

a lot of N
quant of the N

4.7 Left-dislocated *people*-groups

Left dislocation has received much attention in BSAfE literature (Wade, 1995; Gough, 1996:61; Mesthrie, 1997; De Klerk & Gough, 2002:362; Van der Walt & Van Rooy, 2002; De Klerk 2003a:467; De Klerk, 2003b:225; Mesthrie, 2004:972; Makalela, 2004:362; Mesthrie, 2006: 124-126; Minow, 2010; Siebers, 2012). The association of left dislocation and the noun *people* is well-documented (Mesthrie, 1997:132; De Klerk, 2003a:468; Minow, 2010:188-189; Siebers, 2012:210). The high frequency of the noun *people* in left-dislocated nominal groups is not entirely unexpected given the relatively high frequency of the noun *people* in the TLEC.

The high frequency of left dislocation in BSAfE compared to Standard English has been attributed to various pragmatic functions such as the signaling of topic change (Wade, 1995:194), contrast and list functions (Mesthrie, 1997:131), and referent tracking (Siebers, 2012:210), as well as to substrate language influence (Makalela, 2004:361-362). Mesthrie remarks that in 9.8% of the cases of left dislocation in his data (i.e. the category of “neutral predicates” which includes clauses that involve the noun *people*) “no apparent pragmatic function is served” (1997:131-132). In the analysis of left-dislocated *people*-groups in the TLEC, the extent to which left dislocation is attributable to either substrate language influence or pragmatic function will be considered.

There are 37 instances in the TLEC concordance of *people* that fit the description of left dislocation which, according to Mesthrie, “involves a fronted NP with a copy pronoun in the main or (less commonly) subordinate clause” (1997:124). Structurally, left dislocation is illustrated in the following example from the TLEC:⁸⁴

⁸⁴ In corpus examples in this section, *people*-groups are underlined and copy pronouns are in bold.

I think poverty is the cause of HIV/AIDS because people **they** do not have something to eat. <ICLE-TS-NOUN-0236.1>

Makalela (2004) argues that the copy pronoun is attributable to the noun class prefix system in Bantu languages, and provides the following example from Sepedi, a sister-language of Setswana in the Sotho group, to illustrate how this works:

Batho ba ja bogobe.

People + subject agreement marker + eat + porridge

People they are eating porridge (Makalela, 2004:361, emphasis in the original)

In Setswana, the subject agreement morpheme is prefixed to the verb and similar in form to the class prefix of the subject noun (Krüger, 2006:171) as illustrated below:

Dikgômo di-a-fula ('The cattle are grazing') (Krüger, 2006:171, my emphasis – YVB)

It could be argued that the copy pronoun fulfills the function of the substrate language subject agreement morpheme. However, as mentioned in Chapter 1, Mesthrie (1997:139) argues that attributing left dislocation to the concord pattern of the substrate language is “patently simplistic” given that Bantu subject agreement morphemes have a purely grammatical function, whereas left dislocation also has pragmatic functions. Mesthrie (2006:126) treats left dislocation as an instance of undeletion in BSAfE, but is careful to point out that left dislocation fulfills pragmatic functions in BSAfE that are similar to those in Standard English, except for some instances where “left dislocation does not seem to serve any pragmatic function”. It is my contention that such cases involving the noun *people*, as in the example below, are reflective of the substrate language pattern.

(244) I think poverty is the cause of HIV/AIDS because people **they** do not have something to eat. <ICLE-TS-NOUN-0236.1>

In the example above, the copy pronoun is in Subject position. The copy pronoun fills the Subject position in 35 of the 37 instances of left dislocations involving nominal groups with *people* as Thing. There are two cases where the copy pronoun fills the Object position. These will be considered first.

(245) Most of the people whom are victims of Aids or Hiv **their** parents and relatives leave **them** as they are. Only the Government will take care of them till the end of their lives. <ICLE-TS-NOUN-0087.1>

(246) Most of the people poverty leads **them** to do dirty jobs like becoming prostitutes because they find it is the only way to overcome this problem of poverty. <ICLE-TS-NOUN-0139.1>

The two examples above do not fulfill the pragmatic functions associated with left-dislocation in Standard English, namely to reintroduce information that has not been mentioned for some time or to contrast two or more items (Mesthrie, 2006:125, Minow, 2010:178). In both examples the left-dislocated nominal group contains a quantifying *of*-partition. In the first example, the noun *people* is also modified by a postnominal clause. One could postulate that the left dislocation in these examples is an information processing strategy where the complex nominal group is unplugged from the structure of the main clause. This has the effect of topicalizing the dislocated nominal group, but whether the speaker intends to focus the reader's attention on the new topic is not clear. The agnate wordings need to be considered:

(247) Most of the people whom are victims of Aids or Hiv **their** parents and relatives leave **them** as they are. <ICLE-TS-NOUN-0087.1>

Agnate: The parents of most of the people who are victims of Aids or HIV leave them as they are.⁸⁵

⁸⁵ No attempt is made here to provide a Standard English reformulation. As little as possible structural editing to the nominal groups in question has been done.

In the first example (directly above), left dislocation avoids adding an additional level of complexity to an already very complex noun phrase. The brackets below illustrate the depth of the embedding in the left-dislocated *people*-group, compared to the agnate.

Most [of [the people [whom are victims of Aids or Hiv]]]

The parents [of [most [of the people [who are victims of Aids or HIV]]]]

Here one can argue that the left dislocation is attributable to information processing rather than topic focus, although topic focus is always a by-product of left dislocation. In the second example (directly below), left dislocation is more likely to be attributable to topic focus rather than ease of production.

(248) Most of the people poverty leads **them** to do dirty jobs like becoming prostitutes <ICLE-TS-NOUN-0139.1>

Agnate: Poverty leads most of the people to do dirty jobs like becoming prostitutes.

The association between left dislocation and the complexity of the left-dislocated noun phrases have been noted in BSAfE literature (Mesthrie, 2006:126; Minow, 2010:191). In the remainder of this section, the internal structure of left-dislocated *people*-groups in Subject position will be examined in order to establish whether there is a link between the complexity of the dislocated nominal groups and the discourse function served by the left dislocation. An overview of all 35 left dislocated *people*-groups are given in Table 4.13

Table 4.13 Left-dislocated Subject nominal groups with *people* in the TLEC

	Quantified			Qualified			+intervening syntactic element
	+definite determiner	+quantifying determiner	+of-partition	+prenominal modifier	+postnominal phrase	+postnominal clause	
1. people they							
2. This people they	X						
3. those people they	X						
4. These people they	X						
5. Those people they	X						
6. Those people they	X						
7. This people Sometimes they	X						X
8. These people maybe they	X						X
9. these people when they get there in prison, they	X						X
10. some people they		X					
11. other people they		X					
12. Many people they		X					
13. Many people in rural areas they		X			X		
14. many people in rural areas they		X			X		
15. Many people in Africa they		X			X		
16. Black people they				X			
17. Poverty people they				X			
18. the married people they	X			X			
19. The rural people it is like they	X			X			X
20. People who are poor if they						X	X
21. People who use drugs they						X	
22. People who are not working they						X	
23. People who live in poverty, they						X	
24. People who live in informal settlement they						X	
25. these people who abuse they	X					X	
26. all those people who have or had Hiv/Aids they	X	X				X	
27. some of the poor people they	X		X	X			
28. most of our South African people I think they	X		X	X			
29. most of these poor people they	X		X	X			
30. Most of the people in rural areas they	X		X		X		
31. most of the people they	X		X				
32. most of our people they	X		X				
33. Most of the people who are not working they	X		X			X	
34. Most of the people who are staying at home doing nothing, they	X		X			X	
35. Most of people they			X				

The analysis starts with the least complex *people*-groups. There are nine instances where the noun *people* is not further specified in terms of quality or quantity. Of these, there is only one instance without a definite determiner.

(249) I think poverty is the cause of HIV/AIDS because people **they** do not have something to eat. <ICLE-TS-NOUN-0236.1>

As indicated earlier, the copy pronoun in the example above serves no obvious pragmatic function and is likely due to the application of the substrate language pattern. The question arises whether the same can be said of the eight cases with a definite determiner without quality or quantity specification. In all eight cases the definite determiner is a demonstrative determiner serving an anaphoric function. Although the *people*-headed nominal groups themselves are not complex, seven of them are anaphora of complex antecedents. These examples are quoted below (underlined with a solid line) together with their antecedents (italicized). In the first two examples the antecedent directly precedes the left-dislocated *people*-group:

(250) I give thanks to *South African government especially the department of Health minister Mrs Tshabalala-Mosimang*. This people **they** fight against HIV/Aids epidemic <ICLE-TS-NOUN-0172.1>

(251) ... *the ones who are selling their bodies just because they can't afford to meet their needs*, These people **they** need to be advised because they are putting their lives at risk. <ICLE-TS-NOUN-0139.1>

In the next example, all of the anteceding information is not contained within a single nominal group but spread over various causation structures. The anteceding information is 'people who are in prison because they ended up stealing and killing to survive because they were unemployed'.

(252) As a results most of our people **they** did not work they endup stealing and killing for survive because *they* spend their life in prison and those people **they** are afraid to be dumped. <ICLE-TS-NOUN-0218.1>

The antecedents in the next two examples are complex because they contain finite relative clauses. The copy pronoun in the relative clauses is also noteworthy.

(253) When the women sell themselves in the streets they increase HIV because they can't use condoms to the people that they want to buy. Those people they got house, wife and family. <ICLE-TS-NOUN-0179.1>

(254) that is why we have sex workers *those who die because of Aids*. Those people they want some money so they sell themselves so that they can have money. <ICLE-TS-NOUN-0322.1>

In the first example the antecedent refers to the clients of sex workers, i.e. 'people who want to buy [sex]'. The pronoun *they* in *the people that they want to buy* is a copy pronoun and evidence of structural undeletion rather than any pragmatic concern. In the second example above, the copy pronoun (*those*) is inserted between the head noun (*sex workers*) and the relative pronoun (*who*).

From the five examples of the pattern **dem people they** quoted above, the operation of an information-packaging strategy where the anaphoric **dem people** is placed as close as possible to its antecedent information (typically a complex nominal group) can be seen. The pattern **dem people** is then re-referenced by the copy pronoun *they*. In the next three examples, there is an intervening Adverbial element between the left-dislocated nominal group (**dem people**) and the copy pronoun. In all three these cases the antecedent of the left-dislocated *people*-group is also a *people*-group.

(255) There are many people who are unemployed who need that kind of work. This people Sometimes **they** go and earn money in an immoral way and they can get HIV/AIDS because they have lost hope and feel like there is nothing they can do about it. <ICLE-TS-NOUN-0302.1>

(256) ... people in prison. These people maybe **they** have brilliant ideas about helping S.A. <ICLE-TS-NOUN-0319.1>

(257) The society should not take people (criminals) to prison because these people when **they** get there in prison, they can be psychologically disturbed as a result of the punishments they are given. <ICLE-TS-NOUN-0391.1>

Based on the three examples above, it is postulated that **dem people** is used as an information-tracking device to make cohesive chains as explicit as possible, by being placed directly after a complex antecedent instead of some distance away from it. Agnate constructions of the examples above show this more clearly.

Sometimes these people go and earn money...

Maybe these people have brilliant ideas...

...because when these people get there...

If *these people* in the agnate constructions were to be replaced by *they*, the text would still be cohesive. The pattern **dem people** is a preferred cohesive device in BSAfE and used frequently in contexts where a pronoun or a nominal group with the definite article would be used in Standard English. In summary, it is argued here that in cases where Subject **dem people** is left-dislocated, it is because **dem people** has a cohesive function.

There are three instances where the left-dislocated *people*-group contains a quantifying determiner without any further specification of quantity, in other words has the pattern **quant people**. In these examples, the left dislocation serves no discernible pragmatic function. They can be attributed either to an application of the substrate template or to the extension of the pattern with a demonstrative determiner.

(258) Many people **they** suffer because of to be poverty. <ICLE-TS-NOUN-0233.1>

(259) If you commit crime you must know that you cannot always succeed but some people **they** succeed for a long time <ICLE-TS-KIMC-0339.1>

(260) In Africa we have many countries which some of those countries are regarded as the lowest standard countries. Meaning that they cannot do anything for their survival. So those countries or places are regarded as the poor of the poorest. So other people **they** don't

accept the situation they are in. <ICLE-TS-NOUN-0131.1> [Essay quoted from the start].

The last example in the set above has been quoted from the start of the essay to illustrate that *other* is not serving a contrastive function, but is being used as a synonym for *some*. As mentioned earlier the words *some* and *other* are translated into Setswana by the same root, *-ngwe*. The nominal group *batho ba bangwe* can mean 'some people' or 'other people' depending on the context.⁸⁶

It is also difficult to observe a clear function for the left dislocation in three of the four cases where *people* is further specified only by a premodifier. The possibility that the focus is placed on a subclass of people might be considered as a motivation for the left dislocation, but it is impossible to conclusively associate left dislocation with sub-classification based only on the examples below.

(261) Black people **they** are now married whites. <ICLE-TS-NOUN-0478.1>

(262) Poverty people **they** wish to satisfy their need <ICLE-TS-NOUN-0233.1>

(263) The rural people it is like **they** enjoy being like their great grandparents. <ICLE-TS-NOUN-0453.1>

(264) The sex workers **they** don't care about what is going to happen to them at last, they just want some money even the married people **they** can do so because of poverty. Those who use drugs **they** can be infected too. <ICLE-TS-NOUN-0322.1>

The last example shows left dislocation in three consecutive sentences. Here it is possible to discern the contrast and list function. The writer introduces three groups of people (like items from a list), viz, *The sex workers*, *the married people* and *Those who use drugs*, and says something about each of them.

⁸⁶ Personal communication with Prof Rigardt Pretorius and Ms Ansu Berg.

I shall now consider left-dislocated nominal groups that are more complex due to quantifying *of*-partitions (9 cases) and/or postmodification of the noun *people* (13 cases), starting with the pattern ***many people in-phrase***.

(265) Many people in rural areas **they** are not working because in their places they did not have factories and other facilities. <ICLE-TS-NOUN-0075.1>

(266) First of all many people in rural areas **they** are not well Educated <ICLE-TS-NOUN-0228.1>

(267) Many people in Africa **they** die because of AIDS/HIV <ICLE-TS-POT-0203.1>

With regard to the these three examples, one can argue that what on the surface appears to be left dislocation is in fact simply the insertion of the copy pronoun or undeletion of some agreement marker under the influence of the substrate language system. This argument can apply to most of the examples given so far in which no pragmatic function for the left dislocation could be seen.

There are 9 instances where *people* is modified by a relative clause. Six of these are not specified for quantity. They are given below. In the first example, the left dislocation is most obvious, since the subordinator *if* separates the *people*-group from its copy pronoun.

(268) People who are poor, if **they** have HIV/AIDS **they** die quickly unlike the rich. <ICLE-TS-NOUN-0249.1>

(269) People who use drugs **they** can get HIV/AIDS because five of them **they** can use one needle so they need essential services. <ICLE-TS-NOUN-0322.1>

(270) People who are not working **they** are suffering and they want to live a better life like other's; <ICLE-TS-NOUN-0311.1>

(271) People who live in poverty, **they** dont have enough money to send their children to schools, <ICLE-TS-NOUN-0328.1>

(272) People who live in informal settlement **they** are exposed to so many diseases, <ICLE-TS-NOUN-0313.1>

(273) ... these people who abuse **they** could be locked up and rotten in jail <ICLE-TS-NOUN-0060.1>

In three of the nine instances where the left dislocated *people*-group contains a postmodifier, the quantity of people is specified. In the first example below quantity is specified by the predeterminer *all*. In the remaining two examples quantity is specified by an *of*-partition.

(274) when you can take a further look you can see that all those people who have or had Hiv/Aids **they** firstly experienced something like a pimple on their private parts then after some thing worse than that <ICLE-TS-NOUN-0127.1>

(275) Most of the people who are not working **they** drink a lot. <ICLE-TS-NOUN-0217.1>

(276) Most of the people who are staying at home doing nothing, **they** turn to think of anything that can make them happy. <ICLE-TS-NOUN-0217.1>

Of the remaining instances where quantity is specified by an *of*-partition, there are another four instances where quality is specified by a premodifier (3x) or an *in*-phrase. These are quoted below. Again, the topicalizing function of the left dislocation in the first example is more obvious because of the intervening clausal material. The addition of the stance expression *I think* has the structural effect of subordinating the clause in which the copy pronoun functions as Subject.

(277) I think the problem come from us for example most of our south african people I think **they** did not take soccer as profision or as a job <ICLE-TS-NOUN-0304.1>

(278) Some of the poor people **they** are not employed even though they have all the qualifications the government needs. <ICLE-TS-NOUN-0302.1>

(279) Where they stay is not healthy and also you find that most of these poor people **they** do not bother themselves to use a condom. <ICLE-TS-NOUN-0001.1>

(280) Most of the people in rural areas **they** did not have the stable relationship. <ICLE-TS-NOUN-0075.1>

The remaining three instances of left-dislocated *people*-groups with partitive quantification have no further specification of quality.

(281) These problem of poverty must be solved immediately because most of the people **they** act before they and without is like shooting without aiming <ICLE-TS-NOUN-0139.1>

(282) Most of the people **they** don't want to protect themselves they just took it easy as everything will pass they even take it as a joke <ICLE-TS-NOUN-0046.1>

(283) Most of people **they** have not self control. <ICLE-TS-POT-0203.1>

Mesthrie (2006:126) reports that 17 out of 27 (62.9%) of relative clauses and 32 out of 62 partitive genitives (i.e. partitive *of*-constructions) (52.5%) in his data set (12 BSAfE speakers) involved left dislocation. Amongst the 37 instances of left dislocation in the TLEC concordance of *people*, there are 10 instances (9 Subjects and 1 Object) that contain relative clauses and 11 instances (9 Subjects and 2 Objects) that involve quantifying *of*-partitions. Since the current study does not consider all relative clauses or

of-partitions in the corpus, but is restricted to the concordance of the noun *people*, the results cannot be directly compared to that of Mesthrie (2006), but the general tendency for left-dislocated nominal groups to contain relative clauses and *of*-partitions is confirmed. However, earlier in this Chapter it was demonstrated that BSAfE has a preference for postnominal clauses to express quality (Section 4.5), and for partitive constructions to express quantity (Section 4.6). This needs to be taken into consideration in explaining left dislocation.

In spite of the potential frequency effects of the noun *people*, postnominal clauses and *of*-partitions in the TLEC, a closer look at the instances where *people*-groups are left-dislocated has demonstrated a general tendency to make anaphoric links to complex antecedents explicit by means of left dislocation. This kind of cohesive chain can take one of two forms:

Complex nominal group --- [left-dislocated] dem *people* ---- *they*

(284) ... *the ones who are selling their bodies just because they can't afford to meet their needs*, These people **they** need to be advised because they are putting their lives at risk. <ICLE-TS-NOUN-0139.1>

[Left-dislocated] complex nominal group --- *they*

(285) Most of the people who are not working **they** drink a lot. <ICLE-TS-NOUN-0217.1>

The referent-tracking pragmatic function of these left dislocations is most obvious where intervening clausal material (Adverbials or Subordinators) occurs before the copy pronoun. Amongst the examples discussed in this section there was one that illustrated the contrastive (list) function of left dislocation. This, together with the many examples of cohesive uses, shows that left dislocation in BSAfE does serve pragmatic functions. However, the instances in the data where an obvious pragmatic function cannot be discerned suggest that some left dislocations are attributable to substrate language influence.

4.8 Conclusion

The concordances of *people* were analyzed with the aim of explaining the positive keyness of this noun in the TLEC compared to LOCNESS and to explore the structure of nominal groups with specifying elements while keeping the noun (Thing element) constant. In order to discover the usage patterns of *people* in the two corpora, the concordances were compared in terms of four broad categories, namely, instances where the noun *people* had no specifiers (Section 4.3), instances where it was marked as definite (Section 4.4), instances where it was specified for quality (Section 4.5) and instances where it was specified for quantity (Section 4.6). In the last section (4.7), left-dislocated *people*-groups in the TLEC were investigated.

With regard to patterns that characterize the language of the TLEC, a number of trends can be observed, namely: the use of *people* to make the human participant “visible” in the discourse; very explicit referent-tracking devices; a linear, non-integrated presentation of information; substrate language influence; and evidence of anti-deletion. These trends are all relatable to the three possible causes that Schneider (2007:88-89) postulates for the fact that Post-colonial Englishes are characterized by the emergence of new constructions. These factors pertain to the relation between language and culture, the second-language acquisition context itself, and the possibility that a variety develops a character of its own by consistently selecting patterns that conform to an overarching type. Schneider (2007:90) cites Mesthrie’s (2006) investigation of anti-deletion in BSAfE as the only study to explicitly show that features of a variety (BSAfE) conform to an overarching tendency. This chapter has added extensive further evidence for anti-deletion in BSAfE as postulated by Mesthrie (2006).

One could argue that the use of the noun *people* to make human participants more salient and to highlight the humanness of referents is reflective of the African philosophy of *ubuntu*, as captured in the dictum “I am because we are” (Wolf & Polzenhagen, 2009:81). In other words, linguistic devices to make the concepts of community, solidarity and humanity more salient can be seen as reflective of a culture in which the individual is defined in terms of community. In the TLEC *people*-concordance, the following linguistic phenomena are attributed to the ‘visibility’ of human participants: the 35 instances of unspecified *people* as Recipient in a communicative process which could

have been expressed by a 'humanless' nominalization; the 5 instances of *as people* to elaborate the pronoun *we*; the high frequency of the plural possessive determiner *our*; the 21 instances where ***those people who-clause*** is used where ***those who-clause*** would also be idiomatically correct; as well as the high usage of *in*-phrases rather than *of*-phrases to relate *people* to a physical location.

Attempts to be as explicit as possible, a linear presentation of information, and evidence of substrate language influence are associated with the process of language acquisition. Explicit referent tracking is evident in the choice of the demonstrative determiner in anaphoric *people*-groups as well as the linear stacking of co-referential nominal groups brought about by left dislocation. Evidence of a linear, rather than syntactically integrated, information structure can be found in the 28 instances where the *people*-group is followed by an appositive nominal group with a more specific meaning. This demonstrates an add-on strategy normally associated with the pressure of speech production in real time. The cases where left dislocation involved the unplugging of a very complex nominal group from the syntactic structure of the clause also represent a linear rather than integrated ordering of information.

Schneider (2007:89) points out that a learner's first language competence also contributes to the pool of features that can potentially influence the emerging new variety. In the analysis of the data in the *people*-concordance of the TLEC, the possibility that some patterns can be attributed to transfer from the substrate language system was considered, particularly with reference to the high frequency of postnominal clauses as specifiers of quality, the use of *other* as indefinite determiner (synonymous to *some*) and the instances of left dislocation that serve no discernible pragmatic function.

Some of patterns of the noun *people* in the TLEC can be related to morpheme insertion and undeletion. The definite article is often inserted in *people*-groups with generic meaning. After quantifiers *of* or *of the* is inserted, resulting in partitive quantifying constructions such as *many of the people*. Undeletion is most evident in the 21 instances of the pattern ***those people who-clause*** where the pattern ***those who-clause*** would also be idiomatically and pragmatically appropriate.

The patterns that potentially contribute to the positive keyness of the noun *people* in the TLEC include causative constructions in which unspecified *people* is a participant, non-assertive clauses in which unspecified *people* is the Subject and occurrences of unspecified *people* as Recipient in communicative processes. The undeletion of *people* in the pattern ***those people who-clause*** and the 28 instances where the *people*-group has a more specific apposition also contribute to the high frequency of the noun in the TLEC.

The data-driven, bottom-up, usage-based approach taken in this chapter allowed for the identification of a number of phenomena that potentially represent emerging constructions or innovations in BSAfE. Many of these relate to the use of articles, determiners and quantifiers. The use of articles and definite determiners will be investigated further in Chapter 5. In Chapter 6, the use of indefinite determiners and quantifying constructions will be examined.

CHAPTER 5

Determination

5.1 Introduction

This chapter will investigate the use of the determiner element of the nominal group, which is referred to in the grammars of English variously as “Deictic”⁸⁷ (cf. Halliday & Matthiessen, 2004: 312), “determiner” (Payne & Huddleston, 2002:355), “determinative element” (Quirk et al., 1985:253) and “grounding element” (Langacker, 2008:263). This chapter will focus on the articles (*a/an, the*), as the primary lexical indicators of the definite/indefinite distinction, and demonstrative and possessive determiners as determiners with inherent definiteness. The aim of this chapter is to determine the extent to which usage of these central determiners in BSAfE can be attributed to a unique underlying system or to processes of conventionalization. Or, in Sinclair’s (1991) terms: Is the lexical realization of the determiner in BSAfE attributable primarily to the operation of the Open Choice Principle, or to the operation of the Idiom Principle? In the case of the first, semantico-grammatical (systemic) conceptualizations underlie the choice of the article. In the case of the latter, the determiner is learned as part of a pattern (Hunston & Francis, 1999) or a construction (Goldberg, 1992).

In order to explore the article system of BSAfE, it is necessary to first determine the conceptual distinctions of which such a system may consist. Section 5.2 deals with these conceptual distinctions. The definite/indefinite distinction is central to the use of articles. The referential status of the nominal group also plays a role. Section 5.2 presents an extensive exposition of how the definite/indefinite distinction and the specific/non-specific distinction are treated in SFG and the reference grammars of English. There is a complex interaction between definite/indefinite and specific/non-specific which is not adequately explained in the reference grammars. Section 5.2 is an attempt to make the relation between these concepts more explicit. This attempt is

⁸⁷ In SFG functional elements start with a capital letter. This convention is followed here where SFG terminology is used.

continued in Section 5.3 which comprises a literature survey of article usage in non-native varieties of English, including New Englishes and BSAfE. Non-standard article usage is noted in BSAfE literature, but the semantic and pragmatic motivation behind article choice in BSAfE has not been explained. In the remainder of the chapter, the findings of the analyses of the articles (Section 5.4), demonstrative determiners and possessive determiners in the TLEC and LOCNESS are presented.

5.2 Theoretical concepts

The various grammars of English not only use different terms to refer to the determiner slot of the nominal group, but also employ differing systems of categorization of the referential status of the nominal group, resulting in subtle differences in the way in which this functional element is conceptualized by each of them. Therefore a brief overview of some of the descriptions of the grounding element of the nominal group will be presented first in order to arrive at a working terminology for this chapter.

5.2.1 The determiner function in the nominal group

In SFG terms, “Deictic” denotes a slot in the experiential structure of the nominal group. The function of the Deictic, according to Thompson (2004:181), is to signal “**how the nominal group fits into the context of the text around it** and/or of the wider context in which the language event is situated” (my emphasis – YVB). The Deictic identifies the Thing in terms of the here-and-now, i.e. in terms of the speech event (Halliday & Matthiessen, 2004:322). This view of the Deictic function corresponds to the notion of grounding in CG. According to Langacker (2008:259), a grounding element specifies the status vis-à-vis the ground of the thing profiled by a nominal. The ground refers to the speech event, the speaker(s) and hearer(s) and their interaction as well as the immediate circumstances (time and place of speaking). Langacker (2008:260, 262) points out that not all expressions that evoke the ground are true grounding elements and that grounding elements actually do not refer to the ground explicitly. For example, in both the nominal groups *this person* and *the person near me*, the ground is evoked, by *this* and by *near me* respectively, but whereas *this* (as well as *the* in the second

nominal group) evokes the ground only implicitly, the expression *near me* makes an aspect of the ground, namely the speaker, explicit through the word *me*. The word *near* also has lexical meaning. Although the implicit meaning of *this* can be stated as 'proximal', the meaning is more grammatical than lexical.

The criteria for a grounding element presented by Langacker (2008:263) can be summarized as follows:

- (i) A grounding element evokes the ground **implicitly**, but does not profile a facet of the ground; rather, it contributes to the profiling of the thing referred to by the nominal or the process in the finite clause. (In the examples above *person* is profiled.)
- (ii) The lexical content of a grounding element is very schematic. Rather than being conceptually rich, grounding elements serve an **epistemic** function, indicating in a fundamental way what the speaker and hearer know about the status of events and the identifiability of participants (referents).
- (iii) Grounding elements have **grammatical**, rather than lexical meaning, “their schematized meanings residing more in **construal** than in any specific conceptual content” (my emphasis – YVB).

The addition of a grounding element to the thing is what constitutes a nominal group. As Langacker (2008:272) puts it: grounding is “a **semantic function**, an aspect of conceptual organization by which an expression qualifies as a nominal”. The grounding function may be **overtly** realized (e.g. *this train*) or **covert** (e.g. *beer, tigers*) making use of the so-called “zero determiner” (Langacker, 2008:272). With regard to the absence of overt determiners, Halliday and Matthiessen (2004:316) point out that for a nominal group without a Deictic element in its structure, a value is still selected in the Deictic system and realized by “a form having no Deictic in the expression”. Nominal groups that typically require a determiner (albeit the zero determiner) are those that are headed by common nouns.

Nominal groups headed by proper nouns and pronouns do not require determiners. The referents of proper nouns and personal pronouns are identifiable in the context of situation or discourse without the need for a separate grounding element. According to

Langacker, their grounding is intrinsic. In nominal groups with a possessive determiner, (e.g. *their train*, *John's book*) the grounding takes place indirectly via the intrinsic grounding of the possessive noun or possessive determiner (Langacker, 2008:272).

In spite of differences in terminology, the basic notion that a grounding element is a requirement for an expression to be regarded as a nominal group is also supported in more traditional grammars of English. *The Cambridge grammar of the English language* (Huddleston & Pullum, 2002:22)⁸⁸ uses the term “nominal” to refer to “[A] phrase consisting of a noun and the constituents that go with it most closely”, while the term “noun phrase” is used to refer to “a nominal plus a determinative”. According to this reference grammar a noun phrase (NP) minimally consists of a head, which is realized by a noun, and may be accompanied by dependents (Payne & Huddleston, 2002:329), which can be sub-categorized as determiners, modifiers and complements. Huddleston and Payne (2002:55, 329) postulate a three-level hierarchy of NP-nominal-noun which can be stated as: an NP (noun phrase) consists of a determiner plus a nominal, and a nominal consists of a noun with/without modifiers and/or complements. This entails that it is the whole nominal which is actually the head of the NP, and that the noun is the head of the nominal, and therefore the ultimate head of the NP. This hierarchy can be represented as follows:

Figure 5.1 The hierarchical position of the determiner

<i>the</i>	<i>same</i>	<i>people</i>
determiner	modifier	head
determiner	nominal	
noun phrase		

(based on Huddleston & Payne, 2002: 55, 329)

Apart from the grounding function of determiners, which centres on the identifiability of referents and ties in with the status of events, most words that have the potential to realize the determiner function in the nominal group also convey (implicit or explicit) quantifying information; the most notable exception being the definite article *the*.

⁸⁸ *The Cambridge grammar of the English language* (2002) is edited by Huddleston and Pullum, but bibliographical references are to the individual chapter authors.

5.2.2 Grounding and the distinction between type and instance

The distinction between instance and type is important in understanding the function of grounding in the nominal group (Langacker, 2008:264). Type specification (i.e. designation or denotation) is achieved by the lexical noun, whereas instance specification is only possible in a nominal group with direct (whether overt or covert) or intrinsic grounding. According to Langacker (2008:267), *cat* specifies a type, whereas *a cat* specifies an instance of the type *cat*, whether real or imagined. Langacker (2008:267) cautions that instantiation is a separate mental operation and not simply a matter of making a type conception more specific. He asserts that “[w]e often refer linguistically to an instance without having any particular instance in mind” and provides the following set of examples to illustrate this:

(286) Samantha wishes she had a cat. (Langacker, 2008:267)

(287) Since Jennifer had a cat, Julie also got a cat. (Langacker, 2008:267)

In both examples *a cat* refers to an instance. In the first example this instance is not a particular, actual cat, but an evoked instance which, according to Langacker (2008:267) “need not be characterized in any more detail than the type specification”. In the second example reference is made to instances of two particular cats. Langacker (2008:269) reminds us that “we refer to entities at the conceptual level, at the discourse level, and in the ‘world’”. In these terms, the common noun *cat* “refers” to a type at the conceptual level. Adding a grounding element (whether overt or covert) establishes reference to a grounded instance of the type *cat* at the level of the discourse, irrespective of whether a particular instance is intended or not. The nominal group *a cat* may refer to an actual, real-world cat or any hypothetical instance of a cat which is merely evoked to represent the type.

Langacker (2008:271) makes use of anaphoric relations to argue against the use of the term “non-referential” (by, for instance, Payne & Huddleston, 2002: 400, and Givon, 2001:440) to classify non-actual or non-factual instances, or generic nominal groups.

(288) If she had a Porche she would drive it to church. (Langacker, 2008:271)

(289) A hub lies at the center of the wheel it is part of. (Langacker, 2008:271)

(290) I don't have any pets, so I don't have to feed them. (Langacker, 2008:271)

Langacker (2008:271) maintains that the nominals in these examples establish referents at the discourse level, even though they do not have real-world referents, because they can serve as antecedents for anaphoric pronouns, and goes on to state:

From a linguistic standpoint these nominals are indeed referential, as they single out a grounded instance of a type as their referent. Their special property is that they profile a **virtual** instance rather than an **actual** one (Langacker, 2008:271).

For Langacker the criterion for reference is the singling out of a grounded instance of the type. Reference grammars of English, such as the *Cambridge Grammar*, require that the instance be specific for the nominal group to be interpreted as referential. This criterion is discussed in more detail in the next section.

5.2.3 Referential intent: particular vs. non-particular instances

Payne and Huddleston's (2002:399) criterion for reference differs from that of Langacker (2008). They maintain that

... a linguistic expression has reference if, by using it on a given occasion, a speaker intends it to pick out some **independently distinguishable** entity, or set of entities, in the real world (or in some fictional world). By 'independently distinguishable', we mean distinguishable by properties other than those inherent in the meaning of the expression itself. We will say that an expression used in this way is **referential**, that it is used to **refer to** the entity in question, and we call this entity the **referent** of the expression (Payne & Huddleston, 2002:399) (emphasis in the original).

From the excerpt above one can surmise that the criterion for referentiality is the speaker's intention to talk about a particular instance ("entity, or set of entities"). For a nominal group to be referential, it must "pick out" or point to something that exists in the real world or in a fictional world. Moreover, the *Cambridge Grammar* (Huddleston & Pullum, 2002), avoids using the term "reference" in describing the relation between anaphora and antecedents.

Traditionally an anaphor is said to refer to its antecedent, but that is a very different sense of 'refer' from that in which we say that an NP refers to a person or other entity in the outside world. We will not use 'reference' for the relation between anaphor and antecedent. 'Anaphora' itself is the term that names this relation (Stirling & Huddleston, 2002:1457-1458).

This view of anaphora entails that nominal groups that are in an anaphoric relation to each other are not necessarily referential in the sense of picking out a distinguishable (i.e. particular) entity. Stirling and Huddleston (2002:1458) maintain that two nominal groups, even if they are related anaphorically, can only be co-referential if both of them are referring expressions. Although they would acknowledge the anaphoric relationships between the nominal groups underlined in Langacker's (2008:271) examples (repeated below), they would argue that they are not co-referential, because the nominal groups *a Porche* and *A hub* have non-referential interpretations; non-specific and generic respectively.

(291) If she had a Porche she would drive it to church. (Langacker, 2008:271)

(292) A hub lies at the center of the wheel it is part of. (Langacker, 2008:271)

The grammaticality of the use of the anaphor does not necessarily prove co-referentiality in the strict sense in which "co-referentiality" is used in the *Cambridge Grammar* (Huddleston & Pullum, 2002). The following set of examples demonstrates how [truly] co-referential pronouns can be used to test for referentiality:

(293) Did Mary telephone while I was out? She promised to call today. (Payne & Huddleston, 2002:400)

(294) Did anyone call while I was out? *She/He/They promised to call today. (Payne & Huddleston, 2002:400)⁸⁹

Where Langacker (2008:271) regards all nominal groups as referential expressions, Payne and Huddleston (2002:399) maintain that they are used either referentially or non-referentially, depending on the speaker's intent. Givon (2001:439) shares the latter view:

⁸⁹ * indicates ungrammaticality.

Every Universe of Discourse is opened up – established – by a particular *speaker* who then intends entities in it to either refer or not refer. And it is this *referential intent* of the speaker that seems to be more relevant to the grammar of reference in human language.

The nominal groups in the following example are referential in the sense intended by Payne and Huddleston (2002:399) and Givon (2001:439):

(295) Felicity bought a TV and a video recorder, but she returned the video recorder because it was defective. (Quirk et al., 1985:267)

According to Quirk et al. (1985:267) the nominal groups in the example above have **specific reference**. The speaker intends to refer to a particular TV and a particular video recorder. In the remainder of this section we will consider the types of nominal groups that Payne and Huddleston regard as non-referring due to the fact that a speaker, in uttering them, will not intend to "pick out some **independently distinguishable** entity, or set of entities, in the real world (or in some fictional world)" (Payne & Huddleston, 2002:399) (emphasis in the original).

While Payne and Huddleston maintain that nominal groups with **generic** interpretations are non-referential, Quirk et al. (1985:281) use the term "generic reference". When predicative nominal groups specify the type and not a particular individual instance, they are regarded as non-referring by Givon (2001:440). Yet he remarks that, although generic subjects are normally regarded as non-referring (due to their universal nature), "[t]he facts of natural language, ..., tend to suggest that generic subject expressions [...] are in fact bona fide referring expressions, but referring to a *group* or a *type* rather than to an individual" (Givon, 2001:440). The reference is generic if the class denoted by the noun is meant, without any reference to particular specimens of the class. Specific reference entails that reference is made to particular members of the subclass. Quirk et al. (1985:265) provide the following examples to distinguish between specific and generic reference:

(296) A lion and two tigers are sleeping in the cage. (Quirk et al., 1985:265)
[specific]

(297) Tigers are dangerous animals. (Quirk et al., 1985:265) [generic]

Like Quirk et al. (1985:265), Biber et al. (1999:265) distinguish between generic and specific *reference*, stating that "[r]eference is **generic** when a noun phrase refers to a whole class rather than to an individual person or thing." According to Quirk et al. (1985:281), "the generic use of *a/an* picks out **any representative member of the class**." (emphasis in the original)

(298) The best way to learn a language is to live among its speakers. (Quirk et al., 1985:281)

(299) A bull terrier makes an excellent watch dog. (Quirk et al., 1985:281)

(300) A hub lies at the center of the wheel it is part of. (Langacker, 2008:271)

Due to the association of *a/an* with **singular** representatives of a type, nominal groups determined by *a/an* in subject position cannot be used with predications pertaining to the species as a whole (Quirk et al., 1985:281-282; Payne & Huddleston, 2002:407).

(301) Lions / *A lion / the lion will soon be extinct in this part of Africa. (Payne & Huddleston, 2002:407).

The definite article can be used generically with singular countable nouns to indicate "the class as represented by its typical specimen" (Quirk et al., 1985:282).

(302) The bull terrier makes an excellent watch dog. (Quirk et al., 1985:281)

(303) A great deal of illness originates in the mind. (Quirk et al., 1985:282)

(304) No one knows precisely when the wheel was invented. (Quirk et al., 1985:282)

(305) This chapter describes the English noun phrase. (Payne & Huddleston, 2002:407)

The generic definite article is also used with nationality names, to refer to a political or ethnic group of people (e.g. *the Chinese*, *the Romans*) or with nominal adjectives to refer to a group of people who share a certain attribute (e.g. *the unemployed*, *the rich*) (Quirk et al., 1985:283). According Quirk et al. (1985:282), "[t]he generic use of zero article

with both plural nouns and noncount nouns identifies the class considered as an **undifferentiated whole**.”

- (306) Bull terriers make excellent watch dogs (Quirk et al., 1985:281).
- (307) Cigarettes are bad for your health (Quirk et al., 1985:282).
- (308) Hydrogen is lighter than oxygen (Quirk et al., 1985:282).
- (309) Hunger and violence will continue to mark the future of mankind/humanity (Quirk et al., 1985:282).

The first example represents a special case of “the **categorial** meaning of zero” (Quirk et al., 1985:275), namely generic categorial reference implying ‘all members of the category’. This is distinguished from **specific categorial reference** where the use of the zero article with a plural noun evokes the category (class) without involving all members, e.g.

- (310) Joe’s been chasing women ever since he was young. (Quirk et al., 1985:275)

It is clear that *Joe* could not have chased all of the members of the class ‘women’, but the implication is that he would have chased some (probably many) members of the class. Yet the insertion of a quantifier like *some* or *many* would distort the intended categorial meaning of the utterance. This example is compatible with a special case of non-referential interpretation identified by Payne and Huddleston (2002:406), called **multiple-situation-bound** interpretations. There have been multiple situations in which Joe chased a woman. The interpretations of the nominal groups underlined in the examples below are also multiple-situation-bound.

- (311) The president has been assassinated three times. (Payne & Huddleston, 2002:406)
- (312) The police are getting younger. (Payne & Huddleston, 2002:406)
- (313) My sister goes to the theatre every month. (Quirk et al., 1985:269)
- (314) He always eats a sandwich (for lunch). [Habitual] (Givon, 2001:443)

In the first example above, there are three situations in which three different presidents were assassinated. In the second example, more than one generation of police is implied. In the third example, it is not relevant which theatre is attended every month. According to Quirk et al. (1985:269) "*the theatre* refers, rather, to the theatre as an institution", i.e. to the notion of the theatre. The last example is provided by Givon (2001:443) to indicate the correlation between "tense-aspect", in this case the "habitual", and non-referring interpretations of nominal groups.

Payne and Huddleston (2002:405) distinguish non-referring interpretations of object nominal groups which arise from the scope of subject nominal groups with quantifiers. They refer to these as NP-bound interpretations. In the current study the term **quantifier-bound** interpretations will be employed to denote this usage.

(315) Some students have a boyfriend. (Payne & Huddleston, 2002:405)

(316) Most people got the salary they deserved. (Payne & Huddleston, 2002:405)

In the context of copular clauses, Payne and Huddleston (2002:402-404) also identify ascriptive, descriptive and definiendum and definiens non-referring interpretations which would render the nominal group non-referential.

Ascriptive nominal groups are predicates that do not specify or identify the subject, but simply attribute a certain property to the subject. The difference between ascriptive and identifying subject predicates is illustrated by the following set of examples:

(317) John is a teacher (You asked me what he did for a living). (Givon, 2001:440)

(318) He's a teacher I met in college (You asked me who he was). (Givon, 2001:440)

In the second example the predicative nominal group identifies an individual in the 'Universe of Discourse', whereas it is used in the attributive sense in the first example, "connoting the type but denoting no particular individual" (Givon, 2001:440). This attributive usage is termed "ascriptive" by Payne and Huddleston (2002:402) and

“descriptive” by Quirk et al. (1985: 273). Further examples of ascriptive nominal groups provided in the reference grammars include:

(319) Paganini was a great violinist. (Quirk et al., 1985:273)

(320) Mary is a Manchester United supporter. (Payne & Huddleston, 2002:402)

(321) Kim became Pat’s lover / the heir to a large fortune. (Payne & Huddleston, 2002:402)

The ascriptive use of the indefinite article is also found in syntactic roles other than the subject predicative. It is also found in nominal groups functioning as the object predicative or as the complement of the preposition *as*, for example:

(322) My daughter is training as a radiologist. (Quirk et al., 1985:273)

(323) her duties as (a) hostess (Quirk et al., 1985:273)

When one nominal group in a copula clause provides information that helps to identify the referent of the other nominal group, the nominal group providing the identifying information is regarded by Payne and Huddleston (2002:402) as non-referring due to its “**descriptive interpretation**”.

(324) The Vice-Chancellor is that guy over there by the piano. (Payne & Huddleston (2002:402)

(325) I’m Kim Lane. (Payne & Huddleston (2002:402)

(326) The only member of the group who didn’t enjoy it was Jill. (Huddleston, 1984:250)

(327) He’s a teacher I met in college. (Givon, 2001:440)

Both nominal groups in clauses that are definitions are non-referring. The subject nominal group is the **definiendum**, “the entity whose definition is to be specified”, and the predicative nominal group is the **definiens**, “the entity providing the definition” (Payne & Huddleston, 2002:403).

(328) A pentagon is a regular figure with five sides. (Payne & Huddleston, 2002:403)

Givon (2001:301ff) presents a framework for deciding the referentiality of the nominal group which is based on the **propositional modality** of the clause in which it occurs. Givon (2001:301) redefines the **epistemic** modalities of the logical tradition which can be traced back to Aristotle so that the communicative context of the proposition is taken into account. Where the logical tradition distinguishes the following four epistemic modalities: (a) necessary truth, (b) factual truth, (c) possible truth and (d) non-truth, Givon (2001:301) distinguishes (a) presupposition, (b) realis assertion, (c) irrealis assertion and (d) negative assertion respectively. It is apposite to consider the definitions for each of the communicatively defined epistemic modalities provided by Givon (2001:301):

- a. *Presupposition*
The proposition is *taken for granted* to be true, either by definition, prior agreement, generic culturally-shared convention, by being obvious to all present at the speech situation, or by having been uttered by the speaker and left unchallenged by the hearer.
- b. *Realis assertion*
The proposition is *strongly asserted* to be true. [...]
- c. *Irrealis assertion*
The proposition is *weakly asserted* to be either possible, likely or uncertain (epistemic sub-modes), or necessary, desired or undesired (valuative-deontic sub-modes). [...]
- d. *Negative assertion*
The proposition is *strongly asserted* to be *false*, most commonly in contradiction to the hearer's explicit or assumed beliefs. [...]

For purposes of deciding the referential status of nominals, Givon (2001:441-442) groups the epistemic modalities under the mega-modalities of fact and non-fact. Presupposition and realis assertion fall under **fact**; irrealis and negative assertion under **non-fact**. He goes on to make the following predictions of nominal reference in terms of these mega-modalities:

- a. Under the scope of *non-fact*, nominals can be interpreted as *either* referring or non-referring.
- b. Under the scope of *fact*, nominals can *only* be interpreted as referring.

These predictions are then reformulated as: "Nominals may be interpreted non-referentially only if they are under the scope of some non-fact modality. Otherwise they must be interpreted referentially" (Givon, 2001:442). Huddleston's (1984:424) distinction between affirmative and non-affirmative contexts is very similar to the distinction between fact and non-fact modalities. A non-affirmative context "creates an environment where we can have a non-specific NP" (Huddleston, 1984:425).

The most obvious case of a non-affirmative context or negative assertion is a clause with negation. Payne and Huddleston (2002:405) maintain that predicate nominal group in these clauses have **negative-bound** interpretations.

(329) My sister doesn't have a car. (Payne & Huddleston, 2002:405)

(330) I don't have the slightest idea. (Payne & Huddleston, 2002:405)

Payne and Huddleston (2002:401) maintain that nominal groups determined by *no*, *neither*, *either*, *each*, *any*, and *every* are always non-referential due to the fact that they cannot have anaphora with which they are truly co-referential. The determiners *any* and *either* are regarded by Huddleston (1984:255) as "inherently non-specific". Givon (2001:443) indicates that "tense-aspects" also correlate with modality in a highly predictive way. The future, the habitual and the repetitive allow a non-referring interpretation of indefinite objects, for example:

(331) He will eat a sandwich. [Future] (Givon, 2001:443)

(332) He always eats a sandwich. (for lunch) [Habitual] (Givon, 2001:443)

The effect of tense is illustrated in the pair of examples below which Huddleston (1984:254) employs to distinguish between "specific" and "non-specific" noun phrases:

(333) Ed bought a house in Honour Avenue.

(334) Ed would like to buy a house in Honour Avenue.

In the first sentence, where the verb is in the past tense, it is clear that a specific house was bought by Ed. The second example, where a modal auxiliary occurs, allows for

either a specific or a non-specific interpretation: either Ed has set his heart on a specific house, or Ed has no particular house in mind, but does have some preference as to the street in which it should be. According to Huddleston (1984:254) the interpretation as specific (i.e. referential) or non-specific (i.e. non-referential) depends on the context. For instance, in the sentences above, the verb in (333), *bought*, denotes an "actual" event, whereas the verb in (334), *would like to buy*, denotes "potential" (Huddleston, 1984:255, 425).

The inherent modality of certain lexical verbs also plays a role in deciding whether a nominal group should be interpreted as referring or not. Givon (2001:442-443) points out that most lexical verbs have an inherent realis modality (e.g. *has*) and that some verbs carry inherent presupposition scope (*know, forget, regret, be [happy]*). He goes on to list typical examples of the few lexical verbs that are inherently irrealis (*want, like, look for, dream of, think of, believe in*) or negative (*lack, refuse, decline, miss, fail*). The effect of the inherent modality of lexical verbs is illustrated in the following example where the object of the verb *intend* may be interpreted as either referring or non-referring:

(335) I intend to date a Norwegian. (Payne & Huddleston, 2002: 404)

The referential interpretation of *a Norwegian* would obtain if the speaker has a specific person (who happens to be Norwegian) in mind, whereas the **non-specific** interpretation would obtain if the speaker did not have any particular person in mind, but meant *anyone* as long as they have the property of being Norwegian. According to Payne and Huddleston (2002:404) verbs like *intend, desire* and *seek* create hypothetical contexts. The following two examples are additional illustrations of verbs which cast an irrealis scope (*looking for* and *need*) and therefore allow for a non-referring interpretation of their objects:

(336) I'm looking for a millionaire, she says, but I don't see many around.
(Biber et al., 1999:260)

(337) I feel terrible. I need a friend. (Biber et al., 1999:260)

Epistemic adverbs (e.g. *maybe, probably*) cast an irrealis scope over the clause (Givon, 2001:444). Embedded adverbial clauses of condition, purpose or future are irrealis and therefore allow a non-referring interpretation of the indefinite object, for example: *When you get a loan, ...; If you get a loan, ...; In order to get a loan, ...* (Givon, 2001:446).

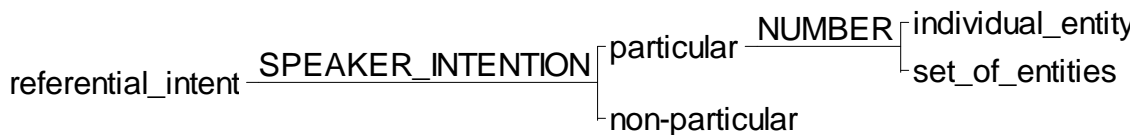
All the examples provided by the grammarians surveyed in this section have articles in the determiner slot of the nominal groups in question. While the articles can be used either with specific nominal groups or with nominal groups intended as generic or non-specific, some determiners are inherently specific (e.g. *this, that*) or non-specific (e.g. *any*) or generic (e.g. *all*).

As we have seen from the discussion so far, Payne and Huddleston (2002) and Givon (2001) explicitly distinguish between referential (referring) and non-referential (non-referring) interpretations of nominal groups, but Givon (2001:449) reminds us that although “referring/non-referring” is treated as a discrete binary choice, “[t]he facts of natural languages suggest the rather disturbing possibility of a continuum between a clearly referring and a clearly non-referring interpretation”. The operationalization of the referring/non-referring distinction in this study, is based not on the cognitivist definition of reference (set out in the previous section), but on Payne and Huddleston’s (2002:399) criterion for referentiality, viz. that referential noun phrases “pick out some independently distinguishable entity, or set of entities, in the real world (or in some fictional world)”.

According to this criteria the words *referring/non-referring* could be substituted with *particular/non-particular*, or for that matter with *specific/non-specific*. There are two reasons for avoiding the terms *specific/non-specific* as a major distinction, i.e. as synonymous with *referring/non-referring*. The first is that this distinction based on referential intent should not be confused with the similarly termed distinction made in SFG with regard to deixis and the Deictic slot (i.e. *specific* and *non-specific* Deictics) (Halliday & Matthiessen, 2004:312). The second is that the term “non-specific interpretation” will be reserved for a special kind of non-referring interpretation (in hypothetical contexts) in accordance with Payne and Huddleston’s (2002:404) operationalization of the term.

The interpretational subclasses of the nominal groups are also important. From the literature surveyed in this section, the following system of referential intent is derived: The speaker can intend either to talk about⁹⁰ a particular thing or a non-particular thing. Should the speaker choose to talk about a particular thing, he may refer either to a specific individual or a specific group of entities.

Figure 5.2 Referential intent system (part 1)⁹¹

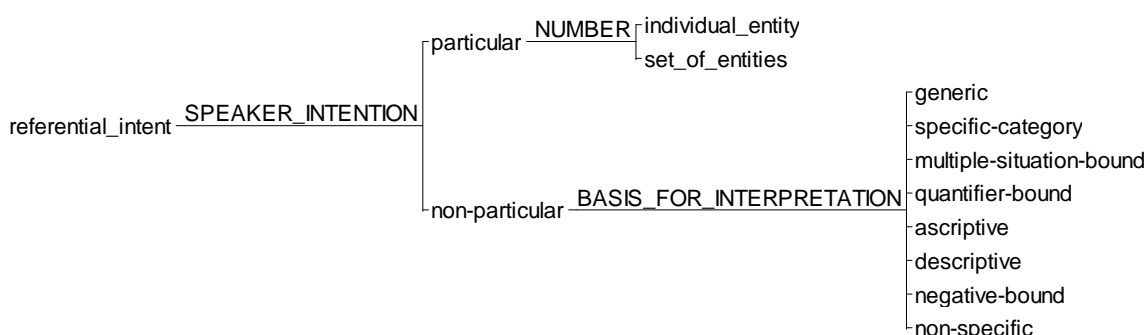


The speaker may talk about class as a whole, by mentioning the entire class, a single representative member or a typical member. This is regarded by Quirk et al. (1985:265) as generic reference, but, in accordance with Payne and Huddleston (2002:4-7) it will be treated here as a special kind of non-particular determination, i.e. as an evocation of the type. Should a speaker have no particular thing in mind it is possible to distinguish on what grounds the hearer/reader interprets the speaker's intention as non-referring (or non-particular). These include ascriptive and descriptive interpretations as well as quantifier-bound, negative bound and non-specific interpretations. Figure 5.3 includes these bases for non-particular interpretations as a subsystem.

⁹⁰ The phrase "talk about" is used here deliberately in order to avoid the verb "refer".

⁹¹ Systems have been generated by the *UAM-Corpus Tool* (O'Donnell, 2007).

Figure 5.3 Referential intent system (part 2)



5.2.4 Definite/indefinite

Quirk et al. (1985:265) point out that "[t]he distinctions between definite and indefinite, and between singular and plural, are important for specific reference", and "less crucial for generic reference". Definiteness is an **epistemological** matter. The epistemic function of the grounding elements of nominal groups have been mentioned earlier: they indicate what the speaker and hearer **know** about the **identifiability** of participants (referents) (Langacker, 2008:263). Identifiability is at the core of the definite/indefinite distinction. In standard grammars of English, the articles *the* and *a(n)* are usually employed to illustrate the definite/indefinite distinction, because they constitute "the special subcategory of determinatives that provide the most basic expression of definiteness and indefiniteness" (Payne & Huddleston, 2002:368), although the distinction can be extended to other determiners (Quirk et al.,1985:253). Demonstrative (*this/that/these/those*) and possessive determiners (*my, his, our, their, etc.*) are regarded as definite while determiners like *any* and *some* are regarded as indefinite.

According to Biber et al. (1999:263) the definite article "specifies that the referent of the noun phrase is assumed to be **known** to the speaker and the addressee" (my emphasis – YVB). Quirk et al. (1985:265) also relate the notion of definiteness to identifiability:

The definite article *the* is used to mark the phrase it introduces as **definite**, *i.e.* as **referring** to something which can be **identified uniquely** in the contextual or general knowledge shared by speaker and hearer (my emphasis – YVB).

Payne and Huddleston (2002:368) point out that "the concept of identifiability expressed by the definite article is best understood in terms of pre-empting a question with *which?*". The use of a definite determiner indicates that the speaker assumes the referent to be familiar to the addressee, or else identifying information is provided linguistically.

Thompson, working in SFG, paraphrases the identifiability signalled by words such as *the* as "the specific member of the type of Thing I am talking about is one that you, the hearer, are able to identify" (Thompson, 2004:181). The information needed to identify the subset of Thing marked as identifiable by *the* is typically to be found within the communicative situation, or has been or will shortly be given in the discourse (Halliday & Matthiessen, 2004:314; Thompson, 2004:149).

Indefiniteness is explained in descriptive grammars in terms of its contrast to definiteness: "With indefinite NPs the addressee is not being expected to be able to identify anything" (Payne & Huddleston, 2002:371). In the same vein Quirk et al. (1985:272) maintain that *a/an* will be used in contexts where the referent is not uniquely identifiable or still unfamiliar to the hearer. Whereas definiteness entails that the referent is "known", "familiar" or "identifiable", the use of the indefinite article entails that the referent is not yet known. The indefinite article is used to present a new entity into the discourse (Biber et al., 1999:260). According to Halliday and Matthiessen (2004:312) the discourse referent of the Thing is introduced by non-specific Deictic elements while, specific Deictic elements track the referent in the discourse. Bear in mind that Halliday and Matthiessen (2004:315) do not use the terms "specific" and "non-specific" in the sense used in the previous section (i.e. in the sense of 'referential intent'), but in a way which overlaps with the definite/indefinite distinction (i.e. in the sense of 'identification'). In this regard, they remark:

Thus, the so-called 'articles' of English, 'definite article' *the* and 'indefinite article' *a(n)*, are terms in, respectively, the specific and non-specific systems of nominal deixis (Halliday & Matthiessen, 2004:315).

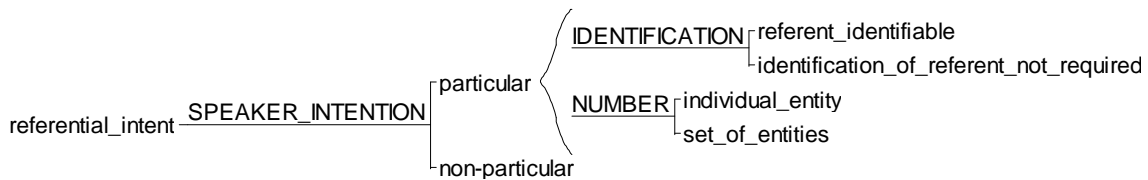
The distinction between specific and non-specific Deictics should not be confused with the distinction between specific and non-specific interpretations of nominal groups based on the intent of the speaker (discussed in the section above). The specific Deictics "have the function of identifying a particular subset of the 'thing' that is being referred to"

whereas non-specific Deictics "convey the sense of all, or none, or some unspecified subset" (Halliday & Matthiessen, 2004:314, 315). These observations for specific and non-specific Deictics are true for definite and indefinite determiners respectively. In SFG a distinction between presenting and presuming reference is made:

Whenever a participant is mentioned in a text, the writer/speaker must signal to the reader/hearer whether the identity of the participant is already known or not. That is, participants in a text may be either presented to use (introduced as 'new' to the text) or presumed (encoded in such a way that we need to retrieve their identity from somewhere) (Eggins, 2004:33).

The definite/indefinite distinction discussed so far, pertains to contexts where the speaker intends to refer to particular instances or sets of instances. This is represented in Figure 5.4.

Figure 5.4 Referential intent system (part 3)



The frequent correlation between definiteness and referentiality is noted by Givon (2001:441) who observes that “definite NPs are most commonly obligatorily referring” and goes on to suggest that in human language “‘definite’ may be aptly viewed as further sub-specification of ‘referring’”. However, as the examples of definite articles in nominal groups with ascriptive, descriptive, negative-bound, quantifier-bound and multiple-situation-bound interpretations illustrate, the definite article also has non-referring uses:

(338) Kim became Pat's lover / the heir to a large fortune. (Payne & Huddleston, 2002:402)

(339) The Vice-Chancellor is that guy over there by the piano. (Payne & Huddleston (2002:402)

(340) I don't have the slightest idea. (Payne & Huddleston, 2002:405)

(341) Most people got the salary they deserved. (Payne & Huddleston, 2002:405)

(342) The president has been assassinated three times. (Payne & Huddleston, 2002:406)

It is important that the definite/indefinite and referential/non-referential distinction not be conflated. The view taken here broadly corresponds to that of Quirk et al. (1985:265) and indirectly to that Givon (2001:441) presented above, namely that the definite/indefinite distinction pertains mainly to referring nominal groups. Nominal groups determined by determiners which are inherently non-referring (in the sense of not picking out a particular entity or set of entities) need not be analyzed as indefinite in the sense of “the hearer/reader is not expected to identify which entity” since no particular entity is intended in the first place. However, the terms “definite article” for the word *the* and “indefinite article” for the word *a(n)* have been conventionalized as a way to distinguish between the articles even where they are used in non-referring nominal groups where the notion of identification is of no relevance. This conventionalized usage will be retained here. In other words, when the words *definite* or *indefinite* precede the noun *article* they are only intended to distinguish between the words *the* and *a(n)*, but when these terms are used on their own their usage is based on the notion of referent identification in referring nominal groups.

To summarize: a referring nominal group is definite when the referent is deemed by the speaker/writer to be identifiable by the hearer/reader. This is an idealized notion of definiteness. It may happen in natural language usage that the hearer/reader is not able to identify what the speaker/writer has intended (and marked) as an identifiable referent. A referring nominal group is indefinite when the hearer/reader is not expected to identify the referent: because the referent is unidentifiable, not yet introduced, or because the identity of the referent is irrelevant.

A definition of definiteness based on the identifiability of the **referent** clearly precludes non-referring nominal groups from the application of the definite/indefinite distinction. However, it is important to note that sometimes an anaphoric relationship provides identifying information in cases where neither the anaphor nor the antecedent is referring to a specific, actual entity.

Standard grammars of English, despite their identifiability-of-referent definitions of definiteness, persist with a form-function isomorphism: *a* = 'indefinite' and *the* = 'definite'. This isomorphism gives rise to phrases such as "Nonreferring uses of the indefinite article" (Quirk et al., 1985:273) and "non-referential uses of NPs determined by the definite and indefinite articles" (Payne & Huddleston, 2002:407).

On the one hand, labelling a determiner as "definite" or "indefinite" in a nominal group that does not have a referent of which the identifiability is at issue seems pointless. On the other hand, the grammar presents a systemic choice between *a(an)*, *the* and no (zero) determiner, not only in nominal groups with referential interpretations, but also in those that are interpreted as non-referential (i.e. non-particular). In many non-particular contexts they cannot be used interchangeably without affecting the meaning, for example:

(343) He plays the piano better than the flute.

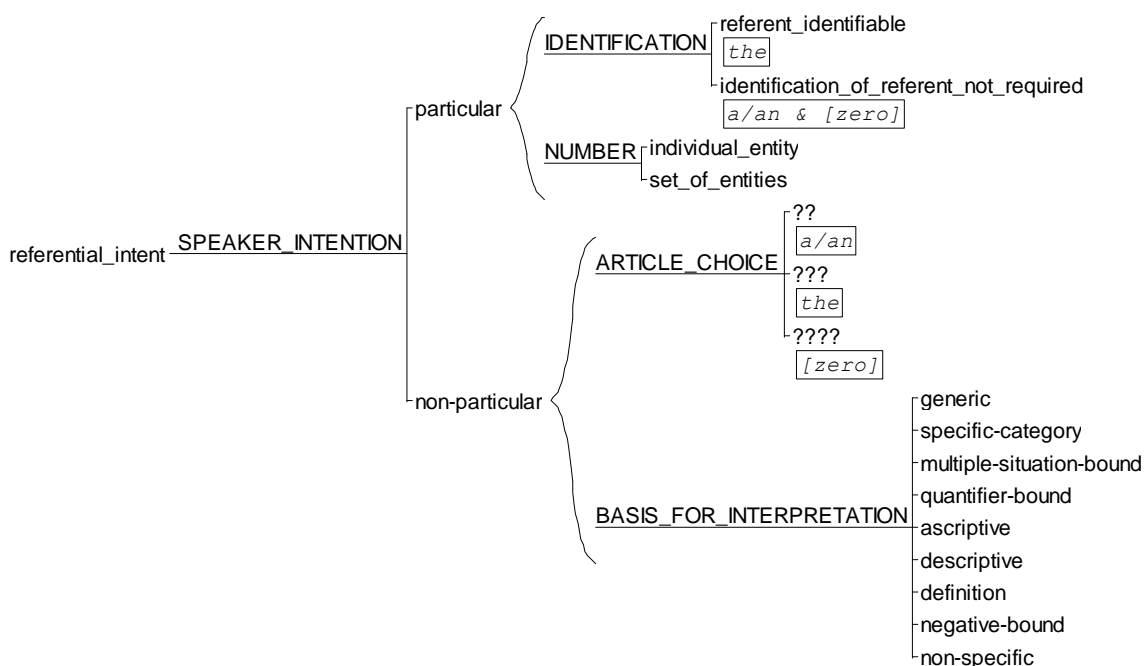
(344) He is playing on a piano, not a harpsichord.

(345) He plays piano every day.

What motivates the definite/indefinite distinction in the article system in non-referential nominal groups? With referring nominal expressions, as well as non-referring nominal groups in anaphoric relationships, the definite/indefinite distinction is based on identifiability. While recognizing a systemic distinction in referring nominal groups based on the identifiability of the referent, it is also necessary to acknowledge the operation of a so-called article system in non-referring nominal groups where identification is not at issue. The semantic terms of such a system are not clear at this point, as can be seen in Figure 5.5.

The *piano*-examples above suggest that conventionalization may be an important factor in the choice of articles. In order to discover what motivates the choice between *the*, *a(n)* and \emptyset (zero determiner) in native, as well as non-native varieties, close examination of referential intent and the non-particular (non-referring) nominal groups in the corpus concordances of the articles are necessary.

Figure 5.5 Referential intent system (part 4)



5.2.4.1 Sources of identifying information

With regard to definite nominal groups which mark the referent as identifiable the various sources of identifying information will be considered in this subsection and will be employed in the classification of the corpus concordances of the word *the* in Section 5.4.1.

The identification of a referent is based on extra-linguistic information or on information provided in the (written or spoken) linguistic utterance. In SFG terms, the Deictic element in the nominal group identifies the Thing “in terms of the here-&-now”, or “in terms of the surrounding text”, but in both cases, “the point of reference is still the speech event” (Halliday & Matthiessen, 2004:323). In other words, the Deictic element situates the nominal group within the discourse or in the wider context in which the discourse takes place (Thompson, 2004:181). Extralinguistic information derived from the communicative situation relates to the notion of deixis, whereas linguistic identifying information relates to anaphoric relations.

Deixis is a Greek word meaning 'pointing/showing' (Payne & Huddleston, 2002:1452). The concept of deixis in verbal communication is explained as follows:

The reference of certain kinds of expression is determined in relation to features of the utterance-act: the time, the place, and the participants, i.e. those with the role of speaker or addressee. This phenomenon is known as deixis and the expressions concerned are called deictic (Stirling & Huddleston, 2002:1451).

Deictic expressions refer to the extralinguistic situation in which an utterance is made, and are therefore dependent on this context for interpretation. The reference of an anaphor, on the other hand, is determined by its antecedent, i.e. the interpretation of an anaphoric expression is based on the previous or subsequent mention of the co-referent in the discourse (Payne & Huddleston, 2002:1455).

Extralinguistic information refers to information derived from the hearer/reader's knowledge of the world or from information presented by the immediate situation in which the discourse takes place (Quirk et al., 1985:266), in SFG terms, the context of culture and the context of situation (Egins, 2004:34). Identification based on extralinguistic information typically entails that the referent is unique within the context, and therefore identifiable, e.g. *the moon*, *the Prime Minister*, *the world* (Quirk et al., 1985:267). Egins (2004:34) uses the term homophoric reference for identity retrieval from the shared context of culture. The identity of the referent may also be retrieved from the immediate situation in which the discourse takes place. This is termed situational reference (Quirk et al., 1985:266) or exophoric reference (Egins, 2004:34).

(346) Go and put it in the fridge. (Huddleston, 1984:249)

(347) The boss has just returned from lunch. (Huddleston, 1984:249)

Huddleston (1984:249) points out that the referent in the examples above is not absolutely unique, but unique in the context of the discourse; i.e. there is only one fridge in the place where the first utterance is made and only one person who is the boss of the speaker and/or hearer involved in the second example. Further examples of identity retrieval from the immediate situation include:

(348) The roses are very beautiful. [said in a garden] (Quirk et al., 1985:266)

(349) These are the pistons. [explaining the engine of a car] (Quirk et al., 1985:266)

Identity retrieval based on the extralinguistic context of the utterance (both the wider context of shared cultural knowledge and the narrower context of the immediate situation) rests on the uniqueness of the thing being identified. The use of the definite article with body parts is partly due to uniqueness within the immediate situation and partly due to prior reference to the individual to whom the body part belongs.

(350) Mary banged herself on the forehead. (Quirk et al., 1985:270)

(351) They pulled her by the hair. (Quirk et al., 1985:270)

(352) Don't keep digging me in the ribs. (Quirk et al., 1985:270)

(353) She grabbed me by the arm. (Payne & Huddleston, 2002:370)

In the examples above the prior mention of the individual to whom the body part(s) belong sets up a context in which the body part is uniquely identifiable. With regard to the last example, one of two arms is implied, but it is not relevant whether it was the left or the right arm.

Intralinguistic bases of identification consist of phoric relations between two nominal groups (which is also a cohesive device) and specifying information provided within the nominal group itself (termed 'qualification' in this study and applied in the broadest sense of the word). An anaphoric relationship between two nominal groups is one where the second nominal group is identified by way of the antecedent nominal group (Quirk et al., 1985:267; Eggins, 2004:35).

(354) A man and a woman got on the bus; the man was carrying a heavy suitcase. (Huddleston, 1984:249)

(355) Felicity bought a TV and a video recorder, but she returned the video recorder because it was defective. (Quirk et al., 1985:267)

In the examples above the second nominal group refers to the same entity as the first nominal group. However, it is also possible that the antecedent nominal group is not co-

referential with the second nominal group, but merely helps to identify its referent through association. Quirk et al. (1985:267) call this indirect anaphoric reference.

(356) I took a taxi to work this morning; the driver turned out to be a part-time university student. (Huddleston, 1984:249)

(357) John bought a new bicycle but found that one of the wheels was defective. (Quirk et al., 1985:267)

The previous mention of *a taxi* pre-empts the question *which driver?* and similarly we know *which wheels* due to the prior mention of *a new bicycle*. Here common general knowledge also comes into play. We know that taxis have drivers and that bicycles have wheels.

The addition of premodifiers and/or postmodifiers can make the referent unique. Where the identifying information is provided in a postmodifier, Quirk et al. (1985:268) call it “cataphoric reference” based on the fact that the identifying information comes after the noun, whereas Eggins (2004:35) calls it “esphoric reference” since the identifying information is given within the nominal group of which the referent is being identified.

(358) The man who succeeded Gough Whitlam as leader of the Australian Labor Party was a very different sort of person. (Huddleston, 1984:249)

(359) The bicycle John bought has been stolen. (Quirk et al., 1985:268)

(360) The President of Mexico is to visit China. (Quirk et al., 1985:268)

Quirk et al. (1985:270) mention the use of 'the' in nominal groups where the meanings of certain adjectives and postdeterminers logically imply uniqueness. These include words like *first*, *last*, *next*, *only*, *same* and the superlative forms of adjectives.

(361) We must catch the next bus. (Quirk et al., 1985:270)

(362) It turned out that John had been to the same school as Max. (Huddleston, 1984:249)

5.2.4.2 Demonstrative and possessive determiners

Like the articles, demonstrative and possessive determiners are central determiners, meaning that they are in a choice relationship with other central determiners (Quirk et al., 1985:254). In Standard English, a nominal group cannot contain more than one central determiner, as in **this the ball*. Like the definite article, demonstrative and possessive determiners are regarded as specific Deictic elements in SFG (Halliday & Matthiessen, 2004: 314).

As central determiners with definite meaning (Quirk et al., 1985:254, 372), demonstrative and possessive determiners are in a choice relationship with the definite article, with the use of the definite article implying 'you know which thing', that of the demonstrative implying 'I'm pointing out which thing' (Halliday & Matthiessen, 2004:314) and that of the possessive indicating 'which thing' in terms of a person.

In addition to their specific referential meaning, the demonstrative determiners also have either proximal ('near') or distal ('far'/'distant') meaning (Quirk et al., 1985:372; Stirling & Huddleston, 2002:1504; Halliday & Matthiessen, 2004:313). The demonstratives *this* and *these* realize 'near', while *that* and *those* realize 'far'. In Standard English the inflectional form of the demonstrative is restricted by the number features of the noun it determines: singular countable nouns and uncountable nouns take the proximal form *this* and the distal form *that*, while plural countable nouns take the proximal form *these* and the distal form *those* (Quirk et al., 1985:372, Stirling & Huddleston, 2002:1504).

5.2.5 Summary of conceptual distinctions

In analyzing the semantic functions of determiners, the first distinction to keep in mind is that between type and instance. The noun itself specifies the type, while the whole of the nominal group represents an instance. The speaker may either refer to a particular instance or have no particular instance in mind. This is what is meant by the particular/non-particular distinction with regard to referential intent. In the latter case (non-particular) the instance merely represents the type. The definite/indefinite distinction is based on identifiability. With regard to intentional reference to particular

instances, the nominal group is marked as definite if its referent is identifiable and as indefinite if its referent is not yet identifiable or if the hearer/reader is not expected to identify the referent. Where the speaker/writer does not intend to refer to a particular instance of the Thing, we still find a grammatical distinction between definite and indefinite, but the semantic basis for this distinction is not clear.

Reference grammars of English (Quirk et al., 1985; Biber et al., 1999; Huddleston & Pullum, 2002) make the definite/indefinite and the particular/non-particular distinction, but do not explicitly relate these distinctions to each other. The system networks in this section (Figures 5.2, 5.3, 5.4 and 5.5) represent an attempt to make sense of these distinctions and to explicitly relate them to each other.

5.3 Determiners in non-native varieties of English

The articles receive far more attention than other determiners in the literature on the features of non-native English, New Englishes and Black South African English in particular. Minow (2010:160-161) points out that irregular article usage (omission or insertion of articles and substitution of the definite article for the indefinite article and vice versa) is attested in 33 out of the 46 varieties of English described in the *Handbook of the Varieties of English* (Kortmann et al., 2004). With regard to BSAfE, Siebers (2012:113) expresses the belief “that the irregular use of articles belongs to the most prototypical features of BSAfE.”

A recent, rigorous, corpus-linguistic account of article usage in varieties of English is Sand’s (2004) study. Sand compares the normalized frequency counts of the articles in six subcorpora of the International Corpus of English (of which two represent native English and four represent contact varieties of English) and one learner corpus (Sand, 2004:282). Her results show that the presence or absence of articles in the substrate language of any particular contact variety does not affect the frequency of articles, which means that “the substrate hypothesis with regard to article use cannot be substantiated through Quantitative analysis”. This finding is relevant for BSAfE since its substrate languages are article-less.

Another important conclusion that Sand (2004:294) draws from the qualitative analysis is that there are bigger differences in article frequencies between text-type grouping within a given corpus, than between the corpora themselves, which points to the important role of text-type in article usage, with written text-types consistently having higher article frequencies than spoken text-types. From a qualitative perspective, Sand (2004:295) further observes that there is a tendency in all contact varieties to use the definite article in contexts where it would not be used in Standard English. These consist of the use of the definite article with (a) collective nouns with generic meaning, (b) nouns denoting a type of institution rather than a specific institution, (c) proper names of places or institutions where the article is not part of the name, (d) temporal expressions which are not further specified through postmodification; (e) deverbal nouns ending in *-ing* where verbal expressions would be expected in Standard English. (Sand, 2004:290-291). She provides the following examples of these usages:

(363) But that is one problem the people don't believe in homeopathy still (ICE-IND S1A-037). (Sand, 2004:290)

(364) Also Jamaicans would not have to enter the university in the prime years of their adult lives... (ICE-JAM W1A-015). (Sand, 2004:291)

(365) ... the Panadol Hockey Tournament which started at the Jamhuri Highschool continued today at the same venue (ICE-EA(K) news broadcast). (Sand, 2004:291)

(366) The most universalized celebration is the Christmas (ICE-EA(K) exam essay). (Sand, 2004:291)

(367) We present a new concept in the broadcasting in India (ICE-IND S2A-007). (Sand, 2004:291)

Sand (2004:295) maintains that these uses of the definite article cannot be due to Second Language Acquisition (SLA) processes since no such uses were found in the learner corpus (ICLE-GER)⁹² and goes on to postulate that the use of the definite article with proper names and temporal expression (365 and 366 above) can be attributed to the principle of unique identifiability, while the other cases (363, 364, and 367 above) are linked to humans and their activities. Sand (2004:295) points out that “[t]here is a strong link in linguistic typology between the concept of definiteness and animacy and

⁹² International Corpus of Learner English: German component

humanness". Use of the definite article with common nouns denoting human institutions (see 364) is also attested for BSAfE by Siebers (2012:129).

(368) I didn't have the mentality that I am in the hospital. (Siebers, 2012:129)

Siebers (2012:123) provides evidence of the insertion of the indefinite article where reference is made to specific companies or institutions and points out that in other varieties of English only the insertion of the definite article in this context is attested.

(369) And then there's a lasertec security. (Siebers, 2012:123)

Sand (2004:295) acknowledges that further studies which include the zero article and reference to the influence of psychological salience are needed to complete the picture. In the analysis of articles, the normalized frequencies of articles in the TLEC and LOCNESS will be compared to Sand's (2004) frequency data. Sand's qualitative observations about the unique uses of the definite article in contact varieties will be considered in the analysis of the concordances of *the* extracted from the TLEC and LOCNESS.

Under features of African English pertaining to the noun phrase, Schmied (1991:71) lists the tendency to omit articles and other determiners before nouns. He considers the possibility that the article system of African English differs from that of standard English in that the former distinguishes specific/non-specific and the latter distinguishes definite/indefinite (Schmied, 1991:71):

In contrast to the system in Standard English some linguists (...) even see a completely different system of articles in 'New Englishes': whereas Standard English uses the definite/indefinite system (**known/not known**) as the basic distinction, the 'New Englishes' prefer to use the specific/non-specific (**particular/not particular**) system, as in the Standard English determiner pair *a certain – any*. In this system non-specific reference is expressed by no article (...) and specific by *the* (...), *one* or even *this/these* and *that/those* (...) (Schmied, 1991:71) (my emphasis – YVB).

Here Schmied (1991:71) refers to the work by Platt, Weber and Ho who maintain that "[m]any of the New Englishes ... appear to make the specific/non-specific distinction rather than follow the definite/indefinite division of the more established Englishes"

(1984:54). Mesthrie and Bhatt (2008:47-48) also start their description of the articles in World Englishes with a reference to Platt, Weber and Ho's (1984) classic distinction between definite/indefinite and specific/non-specific as separate "cognitive subsystems". Siebers (2012:111-112) likewise refers to this distinction in her study of BSAfE.

In order to clarify what is meant by the definite/indefinite distinction on the one hand and the specific/non-specific distinction on the other hand, Platt, Weber and Ho (1984:53-54) provide the following tables with examples:

Table 5.1 The definite/indefinite system

Definite	Indefinite
<p>The persons, things, etc. are thought by the speaker or writer to be <i>known</i> to the listener because</p> <p>1 He has come across them before <i>The girl who rang you yesterday was my secretary</i></p> <p>2 There is (or is thought to be) only one of them in the universe <i>The sun is rising</i></p> <p>3 There is (or is thought to be) only one of them in the particular context <i>Let's go to the park</i></p> <p>4 They belong to a known group or species <i>The penguin is a flightless bird</i></p>	<p>The persons, things, etc. are thought by the speaker or writer to be <i>unknown</i> to the listener or reader because</p> <p>1 He or she has not come across them before <i>I'll tell you about a nice restaurant we went to yesterday</i></p> <p>2 No particular person, thing, etc. is referred to, e.g. <i>Fred wants a job (any job)</i></p>

(Platt, Weber & Ho 1984:53)

Table 5.2 The specific/non-specific system

Specific	Non-specific
<p>The persons, things, etc. are thought by the speaker or writer either</p> <p>1 to have been previously unknown to the listener or reader e.g. <i>He get wan black buk</i> He has a black book (from Haiwaiian Creole)</p> <p>2 to have been previously known to the listener or reader, e.g. <i>Jan bai di buk</i> John bouth the book (meaning a particular book which has been mentioned before) (from Guyana Creole)</p>	<p>1 The persons, things, etc. are unknown to the speaker or writer or the identity of the item is thought by him to be irrelevant to the issue he is discussing or is thought to be obvious, e.g. <i>Jan bai buk</i> John bought a book (it is the book-buying that matters and it is not relevant here whether one or more or which book was bought) (from Guyana Creole)</p> <p>2 The persons, things, etc. are not particular ones but belong to a group, type or species <i>Dag smat</i> The dog (that is the species dog) is smart (from Hawaiian Creole)</p>

(Platt, Weber & Ho 1984:54)

From these examples, Platt, Weber and Ho (1984:54) conclude:

... the main distinction between *definite* and *indefinite* is *known: not known* whereas the main distinction between *specific* and *non-specific* is *particular: not particular*. Within the specific concept there is often a further division into *known: not known*.

Their diagrammatical representations of the definite/indefinite and specific/non-specific systems are reproduced below.

Figure 5.6 The definite/indefinite distinction of “the more established Englishes”
 (Platt, Weber & Ho, 1984:55)

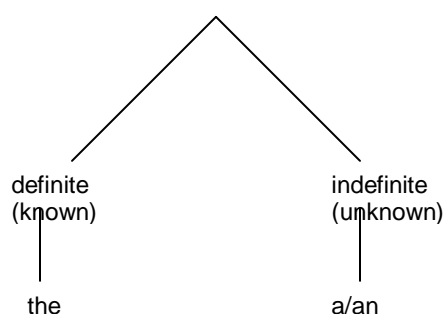
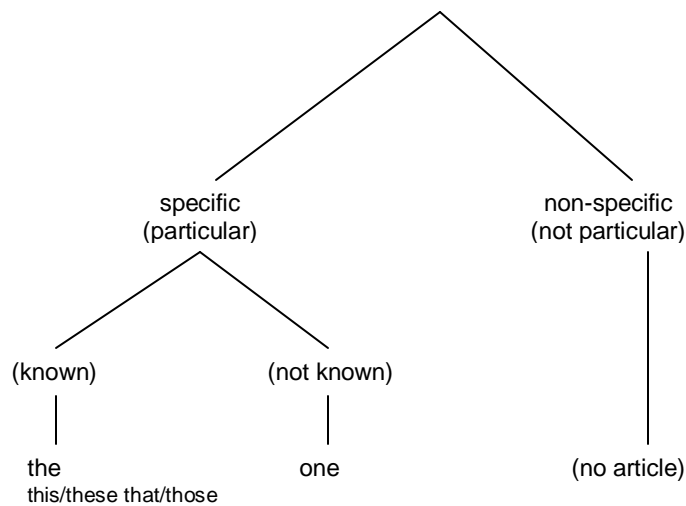


Figure 5.7 The specific/non-specific distinction of the New Englishes
(Platt, Weber & Ho, 1984:55)



It is indeed questionable whether standard or native varieties of English follow the definite/indefinite distinction while New Englishes follow the specific/non-specific distinction. Grammars of (native) English appearing in 1984 and later actually distinguish explicitly between the specific and non-specific uses of the articles (cf. e.g. Huddleston, 1984:254; Quirk et al., 1985:273, Huddleston & Pullum, 2002) and also between the specific and generic uses (cf. e.g. Huddleston, 1984:255); Quirk et al., 1985:265ff) of the articles, as the discussion in Section 5.2 showed.

In Sections 5.2.3 and 5.2.4 above it was demonstrated that a careful exegesis of the descriptions of articles in reference grammars of English (Quirk et al., 1985; Biber et al., 1999; Huddleston & Pullum, 2002) allows for the postulation of a determiner system (see Figure 5.5) for Standard English which makes a similar distinction than the specific/non-specific distinction postulated by Platt, Ho and Weber (1984:53-55) for New Englishes (see figure 5.7), the major difference being the typical lexical realizations of specific indefinite and non-particular. Referential (i.e. particular/specific) nominal groups which require no referent identification from the hearer/reader would require a word like *one* in a New English whereas *a/an* would be used in standard varieties of English. In non-particular contexts no overt article would be used in a New English, while *a/an* would be used in Standard English. It would seem as though Platt, Ho and Weber (1984:55) do

not make provision for the use of the *a/an* form of the article – only for either \emptyset (non-specific) or *one* (specific, unknown). Yet the *a/an* form does occur in some New Englishes.

A comparison of the examples of specific and non-specific contexts from Platt, Weber and Ho (1984:55-58) to examples of referring and non-referring (non-particular) contexts from the descriptive grammars of Standard English (above) illustrates the similarity in terms of the interpretation of referential intent as well as the differences with regard to the lexical choice of article.

The following examples can be compared to generic uses of articles described in the reference grammars of English:

(370) It looks like cat (It belongs to the species cat) (from Indian English)
Platt, Weber & Ho, 1984:55)

(371) ... like mudbrick hut with thatch roof (not a particular hut but referring to a type of hut and a type of roof) (from West-African English) (Platt, Weber & Ho, 1984:55)

The examples below correspond to non-specific interpretations of nominal groups mentioned in the reference grammars:

(372) I thought when I get job I will not be doing any more studies (the speaker didn't know what job he would be getting) (from Papua New Guinean English) (Platt, Weber & Ho, 1984:55)

(373) I really want to spend some time in village, definitely if I get chance. (Here no particular village is meant – just any village. The same goes for chance.) (From Indian English) (Platt, Weber & Ho, 1984:55).

In Payne and Huddleston's (2002:405) terms, the nominal group in the first example below will have a negative-bound (non-referring) interpretation, whereas the last example will have an NP-bound (i.e. quantifier-bound, non-referring) interpretation.

(374) I'm not on scholarship (not on any scholarship) (from East African English) (Platt, Weber & Ho, 1984:55)

(375) Everyone has car (No particular car or cars are being discussed) (from Indian English) (Platt, Weber & Ho, 1984:55)

The New Englishes examples of non-referring (non-specific) nominal groups above have no overt article, whereas in similar contexts, nominal groups in Standard English will have the indefinite article. Platt, Weber and Ho (1984:56) also provide an example of a referring nominal group without an article.

(376) I got very kind mother (from Singapore English) (Platt, Weber & Ho, 1984:56)

With regard to the example above, Platt, Weber and Ho (1984:56) claim, "Here mother is non-specific because it is obvious to the speaker that she has only one mother". This explanation is questionable. In fact, **reference** is made to a **specific** mother, but the uniqueness or identity of the mother is so obvious that the specificity is not marked by a determiner. It is unfortunate that Platt, Weber and Ho (1984:56), in this case, derive the function from the form instead of prioritizing the functional distinctions which give rise to the various forms of determiner. They provide the following examples of specific nominal groups with overt determiners:

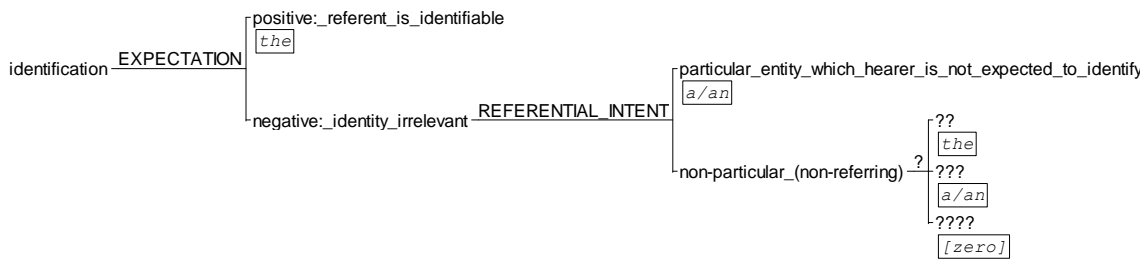
(377) Here got one stall selling soup noodles (from Singapore English) (Platt, Weber & Ho, 1984:57)

(378) I didn't buy the dress lah (referring to a dress that has been discussed before) (from Singapore English) (Platt, Weber & Ho, 1984:57)

(379) When he gave me the bible, he attached the memo which was written about his love about me. [Paraphrase: When he [the narrator's husband] gave me the bible, he attached a memo on which was written some text about his love for me] (from Korean English) (Ionin, Ko & Wexler, 2004:4)

The view taken in this study is that both the speaker/writer's intent to talk about a particular entity or set of entities (or not) (i.e. particular/non-particular) and the epistemic status of the referent (identifiable/not identifiable or known/not known) in the case of referring nominal groups are important in the production and comprehension of nominal groups in both native and non-native varieties of English, if not in all languages. This begs the question: how can seemingly similar pragmatic and semantic choices give rise to such varying lexical realizations of the articles? A possible explanation is that the referential intent and the referent identity subsystems may occur in varying points along the article system network, the more salient distinction occurring earlier. The system network in Fig 5.5 above represents varieties of English in which the particular/non-particular distinction is more salient. A system in which the identity of the referent is more salient might look like this:

Figure 5.8 Identification system



Support for the notion that either one of the two distinctions might be salient in a particular language or variety comes from Ionin, Ko and Wexler's (2004) study on article semantics. Although their study is conducted from the perspective of second language acquisition (as opposed to varietal studies) of English and within a formal (as opposed to functional) theoretical paradigm, their findings are relevant to the current investigation.

Ionin, Ko and Wexler (2004:5) define specificity, i.e. the semantic feature [+specific],⁹³ generally as “speaker intent to refer”. This view of specificity is similar to those of Platt, Weber and Ho (1984:54) and Quirk et al. (1985:273), as well as to Payne and

⁹³ In the current study semantic categories or meanings are generally enclosed in single quotation marks. However, in this case, the conventions of formal semantics used in the cited work have been retained, partly in order to highlight the binary (as opposed to scalar) nature of these features: [+specific] vs [-specific], [+definite] vs [-definite].

Huddleston's (2002) definition of referentiality. Though, it should be noted that Ionin, Ko and Wexler further restrict their definition of specificity by requiring that the speaker intends to refer to "an individual who exists in the actual world" (2004:9).

Uniqueness is central to their definition of [+definite] regardless of the speaker's intent to refer to a particular individual (Ionin, Ko & Wexler, 2004:6-7). They provide the following two (presumably artificial) examples to illustrate this:

(380) I saw a cat. I gave the cat some milk.

(381) The winner of this tournament will receive a prize.

In the first example *the cat* is unique on the grounds of previous discourse. In the second example the speaker and hearer need not be talking about "some salient winner" (i.e. have a particular person in mind), but according to their shared world knowledge they can presuppose that there will be a unique *winner of this tournament* (or any tournament for that matter) (Ionin, Ko and Wexler, 2004:7). Both examples also illustrate that in Standard English the feature [+definite] is marked by the article *the* and the feature [-definite] is marked by the article *a/an*. In the example above *a cat* is specific, while *a prize* is non-specific. Ionin, Ko and Wexler point out that "Standard English has no marker for the [+specific] feature in its article system" (2004:7), but that the word *this* is a marker of specificity in colloquial (non-standard) English. The examples below show that substitution of *a/an* with non-demonstrative *this*⁹⁴ is possible in specific contexts in colloquial (L1) English, but not in non-specific contexts.

(382) Peter intends to marry a/this merchant banker – even though he doesn't get on at all with her.

(383) Peter intends to marry a/??this merchant banker – though he hasn't met one yet. (Ionin, Ko and Wexler, 2004:7)⁹⁵

(384) John has a/this weird purple telephone.

(385) John has a/#this telephone, so you can reach me here. (Ionin, Ko and Wexler, 2004:7)⁹⁶

⁹⁴ Termed "the demonstrative *this* in its indefinite referential use" by Ionin, Ko and Wexler, 1984:7)

⁹⁵ These examples are from Lyons (1999) and cited in Ionin, Ko and Wexler (2004:7). "??" indicates that the grammaticality of *this* is questionable.

Based on an investigation of the articles of English and Samoan, Ionin, Ko and Wexler (2004:12) propose that a language with two articles, such as English (if non-demonstrative *this* and the zero article are excluded) and Samoan, base the choice of articles on either definiteness, as in the case of English, or specificity, as in the case of Samoan. They refer to these as “The Definiteness Setting” and “The Specificity Setting,” respectively, in “The Article Choice Parameter (for two-article languages)” (Ionin, Ko & Wexler, 2004:12). They then proceed to investigate the acquisition of L2 English articles by L1 speakers of languages that have no articles, namely Korean and Russian. The predictions and findings are directly relevant to the current study because the L1 of the contributors to the TLEC, Tswana, is also an article-less language. Ionin, Ko and Wexler’s (2004:17) hypothesis is that L2 learners have access to both settings of the Article Choice Parameter and fluctuate between these “until the input leads them to set this parameter to the appropriate value”. They (Ionin, Ko and Wexler, 2004:19) predict that Korean and Russian learners of L2 English will use *the* correctly in [+definite], [+specific] contexts and will use *a/an* correctly in [-definite, -specific] settings, but will overuse *a/an* in [+definite, -specific] contexts, i.e. where *the* is expected in Standard English, and will overuse *the* in [-definite, +specific] contexts, i.e. where *a/an* is expected in Standard English. Their predictions are supported by the results of an elicitation test (Ionin, Ko and Wexler, 2004:40) from which they conclude that

[...] errors of article misuse in L2 English are systematic, following from an (optional) association of *the* with the feature [+specific] rather than the feature [+definite] (Ionin, Ko and Wexler, 2004:41).

Ionin, Ko and Wexler (2004:42ff) also tested their predictions in a production task. Again, the results mostly supported the predictions, except that there were virtually no instances of [+definite,-specific] produced by the learners, which means that the prediction that *a* would be overused in these contexts was not attested (Ionin, Ko & Wexler, 2004:49). Their ultimate conclusion is that “errors in L2 English article choice are not random but actually reflect L2 learners’ access to universal semantic distinctions of definiteness and specificity” (Ionin, Ko and Wexler, 2004:50).

⁹⁶ These examples are from Maclaran (1982) and cited in Ionin, Ko and Wexler (2004:7). “#” indicates that *this* is ungrammatical or unidiomatic.

When they differentiated between the respondents on the basis of proficiency level, they found that the advanced L2 learners were more accurate with both definites and indefinites (Ionin, Ko and Wexler, 2004:31). This supports the fluctuation hypothesis mentioned above, i.e. that learners fluctuate between an article distinction based on definiteness and one based on specificity “until the input leads them to set the Article Choice Parameter to the appropriate value (the Definiteness value in the case of English)” (Ionin, Ko and Wexler, 2004:50). The nature of the input mentioned here deserves further consideration. The research subjects were late adolescent or adult Koreans or Russians who went to live in the United States for some time. In other words this is strictly speaking an EFL (English as a Foreign Language) immersion learning context.

The “input” for BSAfE is obviously very different. Although English is the medium of instruction in secondary schools, historically teachers (including teachers of English) in rural and township schools are not native speakers of English, but speakers of BSAfE (De Klerk & Gough, 2002:357). Coming from an SLA and formal linguistic paradigm, Ionin, Ko and Wexler (2004) assume native English as input and target and do not take into account the input provided by the learner community. In usage-based approaches to language, the speakers of a variety are regarded as the source of input. Drawing on Croft’s construction-based approach to linguistic patterns, Van Rooy (2010:7) remarks in this regard:

The feature pool is equivalent to the population of linguistic patterns, embedded in actual utterances, in the context where a language is spoken [...]
If this view of language is adopted when considering varieties of English, it must be clear that the linguistic patterns in the actual utterances contributed by speakers form the basis for understanding what happens in the development of new varieties.

Since the substrate languages of BSAfE (including Tswana) are without articles (De Klerk, 2003b:231, 234), the question arises whether Ionin, Ko and Wexler’s (2004:19, 43) predictions about the use of articles also obtain for BSAfE, in spite of the “input” being different. The fact that articles are mentioned (and sometimes discussed in some detail) in nearly all surveys of the grammatical features of BSAfE (Van der Walt and Van Rooy, 2002:120; De Klerk & Gough, 2002:362; De Klerk, 2003a:472-473; De Klerk, 2003b:234; Mesthrie, 2004:970; Minow, 2010:155ff; Siebers 2012:111ff) suggests that

the community of BSAfE speakers themselves rather than an external native norm is responsible for the development of the usage patterns of the articles.

The following example from the TLEC suggests that speakers of BSAfE make a *particular/not particular* distinction conceptually, but that their article choice is not as predicted by Ionin, Ko and Wexler (2004:19, 43).

- (386) It will be easy for them to get a job not just a job but the appropriate one, the one which suit their courses <ICLE-TS-NOUN-0463.1>
- (387) It will be easy for them to get a job - not just any job but an appropriate job which ... (my reformulation – YVB)

In the example above *the* is used in a non-referring context where *a/an* is expected in Standard English. This might be in analogy to first-mention referring nominal groups taking the indefinite article with subsequent-mention referring nominal groups taking the definite article. De Klerk (2003a:473 and 2003b:234) also provides some examples from a corpus of spoken Xhosa English where *the* is used in non-particular contexts, similar to the example above, but these cannot be explained in terms of an analogy to first and subsequent mention.

- (388) I'm the South African I have the right to build here. (De Klerk, 2003a:473 & 2003b:234)
- (389) I'm the vegetarian because I'm afraid of high blood pressure. (De Klerk, 2003a:473 & 2003b:234)

Siebers (2012:115) explains these examples of De Klerk in the light of the fact that they are predicates of a first person subject: “reference to oneself is naturally definite, as no other person can be more definite than oneself”. This ties in with the link between animacy (humanness) and definiteness pointed out by Sand (2004:295).

The following example (from spoken Xhosa English) illustrates the identifying use of the definite article, as opposed to the ascriptive use in the examples above, as well as the use of *a* where *the* would be expected in Standard English.

(390) I'm the man; you're a woman. (Siebers, 2012:121) [Context: "the man speaks about his expectations of who has which role in the house" (Siebers, 2012:122)]

The definite article is expected in both the nominal groups underlined in the example above since they denote unique roles. Siebers (2012:121-122) provides more examples of the use of the indefinite article in contexts where the definite article would be expected in Standard English.

(391) You can't be in a same situation as she is. (Siebers, 2012:121)
(392) And I told him that "you're a despicable man I ever come across". (Siebers, 2012:122)

With regard to the first example, Siebers (2012:122) points out that *same* is incompatible with the indefinite article and "used as if the context was 'in a *similar* situation.'" With regard to the second example, Siebers (2012:122) notes that the superlative is intended and that *a* replaces *the most*. In both these examples the grammatical co-text establishes logical uniqueness (Quirk et al., 1985:270). In the analysis of the article concordances drawn from the TLEC and LOCNESS, the collocation of the articles and postdeterminers like *same*, *similar* and *different* will be considered.

Siebers (2012:126) also finds evidence of the use of *one* or *some* where the Standard English would use the indefinite article:

(393) There's one mother, and there's father. (Siebers, 2012:126)
(394) You will see some child trying to pronounce like eh his name as Akona. (Siebers, 2012:127)⁹⁷

In their acceptability test, Van der Walt and Van Rooy (2002:120) report nearly perfect acceptance of the standard uses of the articles (including the zero article), but only 31% rejection of non-standard Art+X+N constructions and only 21% rejection of non-standard No-Art+X+N constructions. They take the acceptance of both the insertion and omission of articles (see their examples below) as an indication that there is still a lot of instability

⁹⁷ The use of *some* will be examined in Chapter 6.

with regard to establishing a norm for articles in BSAfE (Van der Walt & Van Rooy, 2002:120).

(395) smoking is health risk (Van der Walt & Van Rooy, 2002:120)

(396) my matric results gave a courage to apply for a bursary (Van der Walt & Van Rooy, 2002:120)

De Klerk and Gough (2002:362) list the omission of articles under the grammatical features of BSAfE. De Klerk (2003a:472-473 and 2003b:230) mentions a number of instances in a corpus of spoken Xhosa English where *a/an* and *the* are used with non-count nouns, as well as the use of *a* with quantifiers such as *one* (*a one*) and *some* (*a some*), and the use of *a/an* with plural countable nouns.

(397) so you've got already an information (De Klerk, 2003a:472)

(398) do you believe that the education is the key to success (De Klerk, 2003a:472).

(399) to train them as a bricklayers (De Klerk, 2003a:473)

De Klerk (2003a and 2003b) unfortunately reports only raw frequencies and not normalized frequencies which make it impossible to gauge the prevalence of the phenomena exemplified above (see Minow, 2010:1 for criticism of this kind of utilization of corpus data). Siebers (2012:119-120) also notes the use of the indefinite article with plural nouns and with uncountable nouns and points out that the examples of these usages from her data all involve non-specific entities and mostly occur in object position. Siebers (2012:121) attributes the insertion of indefinite articles before plural and uncountable nouns mainly to “the lack of differentiation between countable and uncountable nouns as well as between singular and plural.” De Klerk (2003a:472) likewise postulates conflation of countable and uncountable nouns in BSAfE. In Chapter 6, I will argue that the use of (quantifying) determiners in BSAfE with nouns with which they would not collocate in Standard English does not imply that BSAfE speakers do not make the countable/non-countable distinction.

Siebers (2012:128) observes instances of the use of the definite article in BSAfE in contexts where no article is expected in Standard English, e.g. in generic nominal

groups and before nouns denoting institutions and before **most of N** where *most* is used in its non-superlative sense. The collocation of *the* and *most* will be further investigated in the concordances of *most* (rather than in the very large concordances of *the*) drawn from the TLEC and LOCNESS (see 6.4.2 later).

Mesthrie (2004:970) mentions the variability between \emptyset , *a/an* and *the* and points out that *a/an* is often inserted in contexts where \emptyset is expected in Standard English. This, according to Siebers (2012:115), allows for the hypothesis that there are higher frequencies of indefinite articles in BSAfE. However, the normalized frequencies of the articles in a corpus of Xhosa English (overall and per speaker) reported by Siebers (2012:116) show “both overuse and underuse” of both the definite and the indefinite articles. The frequencies reported by Siebers (2012) will be compared to the frequencies of the articles in the TLEC and LOCNESS in the next section.

Siebers (2012:125, 130) finds some evidence of article omission in her data, but notes that omission is much less frequent than insertion. According to Mesthrie (2006:136) the omission of articles is not characteristic of mesolectal BSAfE. Based on the high article suppliance rates of her study, Minow (2010:199) concludes that “article omission is a fairly marginal feature overall and not very pervasive even among the lower-mesolectal/basilectal speakers” and asserts that article insertion is more common than omission.

Evidence of article substitution, insertion and omission in BSAfE presents only part of the picture. It does not take into account the instances where the articles are used in ways that correspond to their usage in Standard English. Studies reporting the suppliance rates of articles in spoken corpora of BSAfE (Siebers, 2012; Minow, 2010) provide a wider perspective. In Siebers’s (2012:130) investigation of the output of 16 speakers (with isiXhosa as L1), the indefinite article was supplied in 1751 out of 1864 contexts (overall suppliance rate: 94%), while the definite article was supplied in 3105 out of 3234 contexts (overall suppliance rate: 97%). The overall suppliance rates in Siebers’s data confirm her contention that “the indefinite articles are indeed more problematic than the definite article” (2012:131). The same trend of a slightly lower suppliance rate for the indefinite article than for the definite article is also found in Minow’s (2010) investigation of the output of 45 speakers of BSAfE. Minow (2010:166)

reports an overall suppliance rate of 87.67% for the indefinite article and 92.21% for the definite article. Siebers (2012:130) points out that there are notable differences between the suppliance rates of individual speakers, ranging from 78% to 100% for the indefinite article and from 88% to 100% for the definite article.

Minow (2010:167) divided her respondents into five groups ranging from acrolectal to basilectal and calculated the mean suppliance rates for each article overall and per group. The overall mean suppliance rate for the definite article is 93.1% (standard deviation is 7.03), while the overall mean suppliance rate for the indefinite article is 88.51 (standard deviation is 9.81) (Minow, 2010:167, 169). Unsurprisingly, the mean suppliance rates of the acrolectal group are highest and that of the lower mesolectal/basilectal group the lowest.

Compared to articles, very little is written about demonstrative and possessive determiners in the literature on BSAfE. Van der Walt and Van Rooy (2002:121-122) report a rejection (by teachers) of only 2% (one out of 47 instances) of demonstratives which are non-standard in terms of number features, together with a 100% acceptance of demonstratives with standard English number features. This suggests the establishment of “a new norm, which disregards the number distinction for demonstrative pronouns”. De Klerk (2003a:469) reports 40 cases in her corpus of spoken Xhosa English where *that* is used with plural nouns, 68 cases where *this* is used with plural nouns, 90 cases where *those* is used with singular nouns and 29 instances where *these* is used with singular nouns. Unfortunately, she does not indicate which percentage of the demonstrative determiners is used in this way out of the overall usage of demonstrative determiners. Van Rooy (2008a:282) compares the general trend in the usage of demonstratives (determiners as well as pronouns) statistically and finds that demonstratives are more frequent in LOCNESS than in the TLEC.

In observations about determiners in BSAfE literature, the focus has been on non-standard article usage (omission, insertion and substitution) and on the non-standard collocation of *this/that* with plural nouns or *these/those* with singular nouns. There are no studies that probe the functional motivations of article choice or that investigate the choice relations amongst demonstrative and possessive determiners and the definite article in BSAfE.

A number of questions arise from the survey of descriptions of Standard English and non-native varieties of English (including BSAfE): Is there corpus evidence for the proposition that the underlying article system of BSAfE differs from that of Standard English? If so, of which semantic features does the article system of BSAfE consist? Is there corpus evidence for the proposition that article choice is conventionalized and that articles are learned as part of larger (lexical) units of language? If so, are there lexical expressions or patterns containing articles that are unique to BSAfE or that are more frequent in BSAfE than in Standard English? Similar questions relating to systemic and/or idiomatic differences can be asked about the demonstrative and possessive determiners. These questions will be addressed in Sections 5.4 and 5.5.

5.4 The articles in the TLEC and LOCNESS

Concordances of *the* and *a* were drawn from the untagged versions of the corpora. Concordance searches of the articles in the untagged versions of the corpora yielded several thousand instances, which necessitated a random sampling of concordance lines for further semantic analysis. A random sample of 400 concordance lines per article per corpus was analysed further with regard to the pragmatic and semantic functions of the article in question. Unlike Minow (2010) who excludes articles in fixed expressions from her frequency counts, article usage in fixed or idiomatic expressions are specifically included in the analysis categories applied to the concordance samples of the articles in this study in order to establish to what extent the Idiom Principle is operative in the use of articles.

Before the results of this analysis are presented, it is informative to compare the overall frequency of the concordance counts to the quantitative results of Sand's (2004) study. Sand's (2004) comparison of article frequencies in a number of ICE⁹⁸ corpora facilitates a comparison of the article usage of the TLEC and LOCNESS (both of which are ICLE corpora) to that of other English corpora. The ICE-corpora Sand (2004) used represent both native English (ICE-GB and ICE-NZ)⁹⁹ and contact varieties of English¹⁰⁰ as well as

⁹⁸ International Corpus of English

⁹⁹ GB = Great Britain, NZ = New Zealand

one learner English corpus (ICLE-GERM)¹⁰¹ to compare article usage across varieties. The ICE-corpora consist of different text types, and Sand (2004:287) reports the frequencies per 1000 words per text type. The table below gives only the frequencies Sand reports for student writing:

Table 5.3 Sand's (2004) frequencies for articles in student writing compared to article frequencies in the TLEC and LOCNESS (per 1000 words in the corpus)

Corpus	<i>the</i>	<i>a/an</i>
ICE-GB	73.9	23.9
ICE-NZ	86.6	25.5
ICE-JAM	78.0	22.3
ICE-EA (K)	67.6	28.9
ICE-IND	62.8	13.5
ICE-SIN	79.7	22.7
ICLE-GERM	60.7	28.4
TLEC	45.3	18.0
LOCNESS	66.4	25.1

The table above shows great variability in the use of articles in student writing across the various corpora. The frequency of the definite article in LOCNESS (66.4) – which, like ICE-GB and ICE-NZ, represents native English – is lower than each of those in the native English corpora and also lower than the average for the contact varieties of English (i.e. 72.0 per 1000), but higher than the frequency in the learner corpus (ICLE-GERM). The frequency of the definite article in the TLEC (45.3) is lower than in any of the corpora given above. The use of the indefinite article in LOCNESS is in line with the average of the table above. The use of the indefinite article in the TLEC is lower than its use in all but one of the corpora (ICE-IND) in the table above.

Although there seems to be great variability in the use of the definite article amongst the various corpora within one text type (student writing), there is some consistency within each corpus across text types in that in any given corpus, spoken text types have fewer definite articles than written text types, with conversations having the least definite

¹⁰⁰ ICE-JAM (Jamaica); ICE-EA(K) (East Africa – Kenya); ICE-IND (India); ICE-SIN (Singapore)

¹⁰¹ International Corpus of Learner English – German component

articles. The same trend is observable in the LSWE Corpus (Biber et al., 1999). The LSWE Corpus does not have a separate student writing component, but Biber et al. (1999:267) report a distribution frequency of less than 50 per 1000 words for the definite article in news and fiction (with news having slightly more definite articles than fiction) and a frequency of roughly 56 per 1000 for academic writing. The frequency of the definite article in the TLEC (45.3/1000) is closest to that of news and fiction writing in the LSWE Corpus. This should not be surprising since the TLEC consists of argumentative essays on general (non-academic) topics. With the regard to the indefinite article, Sand (2004:293) points out that essay topics requiring definitions seem to increase the number of indefinite articles.

Sand's (2004:294) finding that "the substrate hypothesis with regard to article use cannot be substantiated through quantitative analysis" is important for the analysis of article usage in the TLEC, since articles do not occur in its "substrate" language, Setswana. The reason for the relatively low frequency of articles in the TLEC compared to LOCNESS and the ICE-corpora used in Sand's (2004) study should therefore not be sought in L1 influence. The comparatively low article frequencies are probably attributable to a lower frequency of noun phrases requiring articles rather than to the "omission" of articles in noun phrases that would require them in Standard English. Recent studies of BSAfE (Mesthrie, 2006:136; Minow, 2010:199; Siebers, 2012:130) note the relative scarcity of article omission. These observations seem to be confirmed in general by the raw counts of concordance lines of the articles (including zero) flagged as either error or correction extracted from the error-tagged subcorpus of the TLEC.¹⁰²

¹⁰² Roughly 40 000 words of the TLEC have been error-tagged.

Table 5.4 Usage of articles in the error-tagged subcorpus of the TLEC

Article used in TLEC (flagged)	Standard English form expected	Number of occurrences
∅	a	76
a	∅	40
the	a	29
a	the	27
∅	the	164
the	∅	167

The data in the table above confirms Van der Walt and Van Rooy's (2002:120) contention that the norm(s) governing article usage in BSAfE is still unstable. Interchangeability is greatest between *the* and the zero article. It is interesting to note that *the* is omitted where it is expected in Standard English virtually the same number of times that it is inserted where no article is expected. The same parallelism can be observed for *the* and *a*. On the other hand, *a* is omitted in contexts where it is expected in Standard English nearly twice as many times as it is inserted where no article is expected. Now that a quantitative overview of the articles in the TLEC and LOCNESS has been given, concordance samples of the articles will be considered.

5.4.1 The definite article

The framework for analysis (presented in Table 5.5 below) of the random sample of 400 instances of *the* in each corpus is based on the discourse functions of the definite article in Standard English. As pointed out above, the definite article typically indicates that the speaker/writer presumes that the hearer/reader is able to identify the referent (Quirk et al., 1985:265; Biber et al., 1999:263; Payne & Huddleston, 2002:368). The identity of the referent may be retrieved from the shared cultural context of the discourse participants, the immediate situation, or the text itself (Quirk et al., 1985:266-270; Eggins, 2004:34-35). The sources of referent identifying information set out in Section 5.2.4.1 above form the basis of about half of the analysis categories used to classify the concordance lines of the definite article.

In addition to its referent identifying uses, the definite article may also be used in contexts where the speaker/writer does not have the intention of “picking out” a specific instance or group, but to refer to the class (i.e. type) as a whole, i.e. in nominal groups which are denotational rather than referential. The contexts in which such non-particular interpretations of the definite article are possible were set out in Section 5.2.3 above and form part of the analysis categories for the concordances of *the*. There are also lexicalized, conventionalized uses of the definite article in English, for instance, with some proper names (e.g. *the Netherlands*). The definite article and zero article are also highly conventionalized in Standard English, as the following examples illustrate: *go to the movies vs go to town; at the office vs at work*. One of the concordance analysis categories covers the idiomatic usage of the definite article. In order to account for the discourse functions of the definite article in the TLEC, it was necessary to add uses typically associated with the indefinite article to the analysis framework, namely ascription and reference to a particular entity where it would seem that the hearer/reader is not expected to identify the referent, or where the identity of the referent is not retrievable from the context.

In both corpora more than three quarters of the sampled concordance lines of *the* represent usage of the definite article to refer to an **identifiable** instance: 301/400 in the TLEC and 345/400 in the LOCNESS concordance sample. Both corpus samples reflect the prototypical referent identifying function of the definite article, but it should be noted that there are more such uses in LOCNESS than in the TLEC.

Table 5.5. Discourse functions of the definite article in the TLEC and LOCNESS

		TLEC	LOCNESS
Identifying uses of the definite article (TLEC:301) (LOCNESS:345)	Referent identifiable/known in immediate situation (exophoric reference)	5	5
	Referent identifiable/known in larger shared culture (homophoric reference)	53	66
	Referent identified by direct anaphora (anaphoric reference)	35	54
	Referent identified by association with other things mentioned in the discourse (indirect anaphoric reference)	34	33
	Referent identified by modifiers within the nominal group (esphoric reference)	160	178
	Referent identified by predication	14	9
Referring use of the definite article without identification	Referent specific, but not identifiable	3	0
Non-referring (non-particular) use of the definite article (TLEC:96) (LOCNESS:55)	Non-specific, non-identifiable	4	0
	Generic (reference to type rather than instance)	51	32
	Role	2	5
	Ascription	9	0
	Idiomatic	28	17
	Definition	2	1
	SAMPLE TOTAL	400	400

5.4.1.1 Identifying uses of the *in the* TLEC and LOCNESS

The first six concordance analysis categories have been assigned on the basis of identifiability. In examining the uses of *the* to signal that the identity of the referent is retrievable, the focus falls on the epistemological status of the referent from the reader's point of view, i.e. whether the writer presumes that the reader knows (from the situation or from information given in the discourse) **which** thing is intended. As indicated in the sections above, there is a strong correlation between identity ("definiteness") and reference, but the concordance analysis below clearly shows that not all uses of the definite article motivated by identifiability or uniqueness actually entail reference in the sense used by Payne and Huddleston (2002:399).

Since the TLEC and LOCNESS are written corpora, few or no instances of referent **identification based on information in the immediate communicative situation** were expected. However, there were 5 instances of *the* used in **exophoric reference** (Eggins, 2004:34) in each corpus sample. In 3 of the 5 instances in the LOCNESS data, the referent is a specific time which is identified in relation to the time of the utterance. In the other two cases reference is made to objects in a literary text which the writer assumes the reader has read.

(400) That's a name that we've frequently seen in newspapers across America during the past few months. At the time Otakar Kirchner, Richard's biological father <ICLE-US-MRQ-0032.1>

(401) When the hero, Kaliayev was engaged to throw the bomb at the Grand-Duke Serge, [...]<ICLE-BR-SUR-0008.1>

There is one instance in the TLEC sample, where the referent is a place within a narrative.

(402) So he went on with his friends to Mr Manaba's liquor store on the corner. <ICLE-TS-KIMC-0299.1>

In the remaining four instances in the TLEC data, the referent is presumed 'known' on the basis of the essay writing assignment itself. For instance, in the examples below, *the statement* refers to a statement made in the essay topic to which the writer is responding, and in the last example, reference is made to the actual essay.

(403) [...] therefore yes i agree with the statement <ICLE-TS-KIMC-0356.1>

(404) The Following Essay will be based on the argument of How poverty is the cause of the HIV/AIDS epidemic in Africa <ICLE-TS-WITS-0004.1>

So, although both corpus samples have the same amount of referent identification based on the immediate situation in which the discourse takes place, the nature of the kinds of

referents to be identified in this manner differs in LOCNESS and the TLEC. In LOCNESS, the majority of the instances take the time of writing as a reference point to contextualize another point in time. In the majority of the instances in the TLEC, the point of reference is the essay writing assignment itself. In these cases the TLEC writers assume that the reader(s) will be the same person(s) who posed the statement in the essay topic, and the response is similar to the type of reference to preceding discourse found in conversation or in personal written correspondence (e.g. e-mail). In a study of the involved aspects of student writing, Van Rooy and Terblanche (2006:178) point out that the writers of the TLEC assume “more shared background between writer and reader than is the case in native-speaker student writing”, which relates to less nuanced contextualization in writing (also see Van Rooy & Terblanche, 2009).

Since identifying information is retrieved from the immediate communicative situation, the nominal groups in this category, in both corpora, have neither epithetic pronominal modifiers nor postnominal quality specification. There are two cases in the LOCNESS data and one case in the TLEC data where pronominal relative qualities support the situation-based identification.

(405) The Following Essay will be based on the argument of How poverty is the cause of the HIV/AIDS epidemic in Africa <ICLE-TS-WITS-0004.1>

(406) That's a name that we've frequently seen in newspapers across America during the past few months <ICLE-US-MRQ-0032.1>

There are 53 instances in the TLEC sample, and 66 in the LOCNESS data where the identification of the referent relies on the larger context or shared knowledge of writer and reader. With this kind of homophoric reference (Eggins, 2004:34), the identification of the referent is based on its uniqueness within a larger cultural context or shared knowledge framework. In addition to the difference in quantity in this analysis category, the corpora also differ with regard to the type of thing referenced, i.e. in terms of noun denotation. The noun lemmas that most frequently head such nominal groups in this category in the TLEC data are *world* (16x), *government* (9x) and *society* (5x).

(407) Asia is the most populated continent in the world. <ICLE-TS-NOUN-0018.1>

(408) they understand our problem but when coming to soccer they dont. Until the government or members of Fifa pay our players [...]<ICLE-TS-NOUN-0304.1>

(409) The society is faced with one question [...] <ICLE-TS-NOUN-0108.1>

In all four instances with the singular noun *society* in the TLEC sample of *the*, a reformulation without an article would be acceptable in Standard English. Sand (2004: 292) points out that *society* is “of special interest” with regard to article usage because the usage of *society* with and without the definite article is almost evenly distributed in three of the ICE-corpora representing contact varieties, namely ICE-JAM, ICE-IND and ICE-UK, and that one fifth of the instances of *society* with generic meaning in ICE-SIN are preceded by the definite article. A concordance of *society* in the TLEC indicates that the TLEC is in alignment with the contact varieties that use the definite article with roughly half of all instances of the noun *society*. Of the 254 instances of *society* in the TLEC, 57% have *the* as determiner, 24% have no determiner and 19% have determiners other than the definite article. The string *the society* does not occur in the LOCNESS sample of 400 instances of *the*, but a concordance of *society* shows that, of the 212 instances of *society* in LOCNESS, 73% have no determiner; only 4% have the definite article as determiner; and 23% have determiners other than *the*. This demonstrates that some differences in the distribution of articles (and other determiners) are attributable to variability in the conventions that obtain for a certain noun (such as *society*) in a certain speech community. One could argue that the four instances of *the society* in the TLEC concordance sample of *the* should be counted as generic rather than identifying due to largeness of the group of people implied by *society*. However, they are counted under the identifying uses of the definite article because shared cultural knowledge would allow one to demarcate which *society* is intended. However, the fact that 57% of instances of the noun *society* in the TLEC are preceded by the definite article, compared to only 4% in LOCNESS, suggests that the use of the definite article with *society* is conventionalized in BSAfE.

In the LOCNESS sample, the noun lemma that most frequently heads nominal groups in this category is also *world* (5x). Uses of *the* in combination with a proper noun, e.g. *the Gulf*, are included in this category, since a proper name singles out a unique instance in a given context. It is interesting to note that there are more than four times as many nominal groups where *the* is paired with a proper name in the LOCNESS data (31x) compared to the TLEC data (7x) in this analysis category. The use of the proper name makes the uniqueness of the referent so salient that its identifiability (based on uniqueness) need not normally be marked by the definite article, for example *(in) Durban* or *(in) Durban harbour*. However, the use of the definite article with some place names has become conventionalized in standard international English (e.g. *the United States of America*) as well as in South African English, for instance, *the Drakensberg (mountains)*. There are three instances of *the North-West Province* in this category in the TLEC concordance sample. Many of the proper names in the LOCNESS data that combine with the definite article are conventionalized geographic names, e.g. *the United Kingdom* and *the Falklands*. There are also proper names of other things that require the definite article purely for reasons of uniqueness and convention, for example:

(410) Laws Like *the Criminal Law Amendment Act* [...]<ICLE-TS-WITS-0004.1>

The conventionalized combination of proper names and the definite article is not grouped with other fixed idiomatic expressions containing the definite article, such as *in the long run*, which do not necessarily have an ‘identifiable instance’ interpretation. Proper names with a conventionalized definite article belong with the identifying uses of *the* because they pick out a specific referent which is unique (and thus identifiable) in a certain broad cultural context or shared knowledge base.

With regard to referent identification based on information contained in the discourse itself, a distinction is made between nominal groups which have a **direct antecedent** and nominal groups which mention things closely **associated** with the referent, thereby creating a frame of reference in which the referent is identifiable. The example below is of a direct anaphoric relation: *the sport* is an anaphor of *soccer*.

(411) soccer [...] Their talent and passion for the sport [...]<ICLE-TS-NOUN-0264.1>

The definite article is employed more frequently in LOCNESS (54x) than in the TLEC (35x) to indicate identifiability of the referent from direct antecedents. The use of *the* to indicate indirect anaphoric identity retrieval is quantitatively similar for the two corpora: 34 in the TLEC and 33 in LOCNESS. In the examples below, the identity of *the officers* is retrievable from reference to the siege of Mafikeng during the Anglo-Boer (South African) War earlier in the essay, while the mention of *injections* allows for the identification of the composite parts of an injection, such as *the needle*.

(412) the siege of Mafikeng [...] There is also the imperial reserve where the officers were situated. <ICLE-TS-NOUN-0392.1>

(413) Some drugs are ingested into bodies by means of injections. [...] Sometimes the needle may be infected with this dreadful virus of HIV/AIDS. <ICLE-TS-NOUN-0178.1>

Though the things denoted in the nominal groups underlined in the examples above are both identifiable due to their association with other things mentioned earlier in the discourse, they differ with regard to their referential status. The nominal group *the officers* refers to **actual** individuals who participated in an actual event. The nominal group *injections* do not refer to a specific subset of injections administered during a specific event, but to a class of thing. Since the frame-establishing, indirect antecedent (*injections*) has a 'generic' interpretation, *the needle* is to be interpreted as 'non-specific', i.e. as '**any** needle of any injection in any event of intravenous drug administration'. The referential frame in terms of which the identity of *the needle* is to be retrieved is a **potential** scenario, and not an actual event. The adverb *sometimes* and the modal verb *may* contribute to the construal of a potential rather than actual event. However, in spite of its generic indirect antecedent (*injections*) and the non-factuality of the event in which it participates, the nominal group, *the needle*, does not merely specify type. The hearer/reader is expected to imagine a (hypothetical) scenario in which a specific instance of a needle is used (also see Van Rooy, 2008b: 350).

There is also an example of such a hypothetical instance in the LOCNESS data. The example below comes from an essay discussing surrogate motherhood. The reader is

implicitly expected to imagine a **typical**, but **hypothetical scenario** in which an infertile couple engages another woman to bear their child. There are a number of participants closely associated with **any such** scenario (i.e. surrogacy), viz. the surrogate, the mother, the father and the child. The instance that is identifiable within the imagined scenario, and the scenario as such, is typical, but non-actual. The hypothetical instance represents the type.

(414) Due to the fact that she did not bear the child the mother may feel that she does not deserve to “care” for it. <ICLE-US-MRQ-0009.1>

There are 34 instances of associative (indirect anaphoric) identification in the TLEC concordance sample and 33 in the LOCNESS concordance sample. Of these there are 14 ontologically non-actual referents in the TLEC sample and 8 in the LOCNESS sample. However, in a discussion of the discourse functions of determiners, the ontological status of the Thing is secondary to its epistemological status in the discourse. It should be noted that not all of the nominal groups in this analysis category (indirect anaphora) are referring in the strict sense intended by Payne and Huddleston (2002:399). What is at issue here is that whatever is talked about can be identified in terms of a context established earlier in the text. Yet it is interesting to note the more frequent identification of non-specific instances in hypothetical scenarios in the TLEC.

Referents can also be identified by more grammatical means, namely by the information provided within the nominal group itself in the form of modifiers and through identifying predication which establishes an identifying relationship between two definite nominal groups linked by a copula verb. In both corpus concordances, the definite article is most frequently used to indicate that the referent is identifiable and that the identifying information is provided within the nominal group itself. Eggins (2004:34) refers to this kind of referent identification as esphoric reference. In both examples below, additional specifying information is provided by modifiers in the nominal groups. In the first example, the postnominal modifier *of football* is enough to identify which *standard*. In the second example, the first nominal group is the Identifier, and the second the Identified (Halliday & Matthiessen, 2004:227).

(415) The standard of football here is lower as compared to other countries.
<ICLE-TS-NOUN-0225.1>

(416) and these people who are not working are the ones who are involved
in some activities like being prostitutes <ICLE-TS-NOUN-0045.1>

Not all nominal groups in the concordance samples that have modifiers are classified as esphoric reference – only those where the identifying information is primarily found in the modifier rather than in the extralinguistic context or elsewhere in the discourse. This category is the largest analysis category in each of the corpora, suggesting that the identifying information indicated by the definite article is most often to be found within the nominal group itself. The data was subcategorized in terms of the kinds of modifiers that provided the identifying information. Table 5.6 compares the concordance samples with regard to the positional types of modifier used to provide referent identifying information.

Table 5.6 Nominal group internal identification in the concordances of *the* from the TLEC and LOCNESS: Position of identifying information

	TLEC		LOCNESS	
	Raw freq.	% of Cat.	Raw freq.	% of Cat.
Prenominal modifiers only	42	26	32	18
Both prenominal and postnominal modifiers	30	19	34	19
Postnominal modifiers only	88	55	112	63
Total	160	100	178	100

It is interesting to note the similarity between the two corpora with regard to referent identification by means of both prenominal and postnominal modifiers. Table 5.6 shows that in both corpora referent-identifying information is given both pre- and postnominally in 19% of the cases in this analysis category. In both corpora the postnominal position is used more frequently than the prenominal position for referent-identifying qualities, however, it should be noted that the TLEC makes use of slightly fewer postnominal specifiers and slightly more prenominal specifiers than LOCNESS in nominal group internal referent identification. Table 5.7 compares the identity-specifying modifiers with regard to the semantic types of quality realized by the prenominal modifiers and the structural types of postmodifiers.

Table 5.7 Nominal group internal identification in the concordances of *the* from the TLEC and LOCNESS: Quality type and form

	TLEC		LOCNESS	
	Raw freq.	% of Cat.	Raw freq.	% of Cat.
Prenominal relative qualities (incl. extent)	19	11.9	24	13.5
Prenominal classifying qualities	29	18.1	21	11.8
Prenominal epithetic qualities: superlatives	11	6.9	5	2.8
Prenominal epithetic qualities: non-superlative	11	6.9	16	9.0
Appositive nominal group	1	0.6	4	2.2
Postnominal <i>of</i> -phrases	56	35.0	87	48.9
Postnominal prepositional phrases	19	11.9	12	6.7
Postnominal prepositional relative clauses	6	3.8	5	2.8
Postnominal marked relative clauses	13	8.1	9	5.1
Postnominal unmarked relative clauses	9	5.6	10	5.6
Postnominal non-finite clauses	9	5.6	9	5.1
Postnominal finite complement clauses	5	3.1	8	4.5

In Table 5.7 above, the raw frequencies of the various types of modifier per corpus in this category do not add up to the total number of concordance lines assigned to this category in each concordance sample, because some nominal groups contain more than one modifier. In both corpora the structure most frequently employed to provide referent-identifying information is the *of*-phrase, but it should be noted that *of*-phrases are more prevalent in the LOCNESS sample than in the TLEC sample. Conversely, prenominal classifiers, as well as postnominal prepositional phrases (excluding *of*-phrases) and marked relative clauses, are more frequently employed to provide referent-identifying information in the TLEC data than in the LOCNESS data.

What remains to be done is to consider the non-identifying uses of the definite article.

In the TLEC sample there are three instances where reference is made to a specific thing, but where the reader cannot be expected to identify the referent, since it has not previously been introduced into the discourse. Two of these uses of *the* in the TLEC

sample involve an indication of the extent of some phenomenon which the writer might (incorrectly) assume is known to the reader.

(417) Due to the high teenage pregnancy some teenagers are also forced into prostitution <ICLE-TS-NOUN-0146.1>

(418) This is due to the statistics showing the declining of employment opportunity by 2.2 percent. <ICLE-TS-NOUN-0176.1>

In the next example, the referent *the report* is introduced for the first time. There is nothing to suggest that the writer assumes the reader knows which *report* is meant. Here we have reference to a specific, but unknown *report*.

(419) In fact the report says that the majority of 13 million death from infections disease [...]<ICLE-TS-KIMC-0362.1>

With regard to identifying uses of *the*, the corpora are very similar, but there are more identifying uses of *the* in the LOCNESS sample than in the TLEC sample. Of note are the three instances where *the* is used with specific nominal groups of which the referent is not identifiable.

5.4.1.2 Non-particular uses of the in the TLEC and LOCNESS

We now turn to the non-referential, or denotational, class-specifying uses of the definite article. In the random samples of 400 concordance lines per corpus, there are nearly twice as many denotational uses of the definite article in the TLEC data compared to the LOCNESS data: 96x vs 55x.

There are four instances of nominal groups in irrealis or hypothetical contexts that take the definite article in the TLEC concordance sample. There are no corresponding uses in the LOCNESS sample, because in Standard English such nominal groups with non-specific interpretations are indefinite.

(420) Africa has to find ways of making money to cater for the good living standard of its people. <ICLE-TS-NOUN-0284.1>

(421) Women are threaghtned by men that they will divorce them if open the case against man. <ICLE-TS-NOUN-0428.1>

(422) in the streets and shoping complexes looking for a job. They went as far as to the bars where they hope to find one person who can give them some money. <ICLE-TS-NOUN-0140.1>

(423) but he can change if he is willing to do so. He can be send to the parents, who have the similar problem. <ICLE-TS-NOUN-0012.1>

Like the three examples of specific, non-identifiable uses of *the* listed earlier, the first two examples above would require the indefinite article rather than the definite article in Standard English reformulations, as in: *a good living standard for its people*, *a case against [her husband]*. The difference is that these two nominal groups have non-specific interpretations based on the irrealis scope cast by *cater for* and *if* respectively. In Standard English reformulations of the latter two examples above, the underlined nominal groups would have no determiner since they are plural and have non-specific interpretations.

The nominal group *the parents* in the last of the set of examples above is interpreted as non-specific, partly because of the irrealis scope cast by the modal verb *can*. A longer excerpt of the essay in which this nominal group occurs is quoted below to illustrate how the writer moves between reference to type (italics) and reference to instance (underlined).

(424) *Some they* really need help like *the one who kill his parents*. he was in a process of stress.

That everything his parents was saying was the fact. Sometimes when *parents* are angry they turn out saying *the bad things* that *children* feel that *they* have been neglected *they* are nothing. *They* end up in sorrow. the community have to take care of that particular person. This child have to be send to the preist to comfert her to guide him that what he did was wrong but he can change if he is willing to do so.

He can be send **to the parents, who have the similar problem**. So that *they* can advice her that *they* knew it was not his decision to kill his parents. It was only that he kept everything in a negetive way, then he dont have any one to to talk to if he is facing the challeges of life, he regret that if it wa'st his temptation he would have his parents. <ICLE-TS-NOUN-0012.1>

The past tense (not marked for the verb *kill*, but found in *was* in the next clause) suggests a realis interpretation for the referent of the nominal group *the one who kill[ed] his parents*, in other words, that the writer is referring to 'an actual individual who killed his parents' (i.e. an instance) and that this individual belongs to a more general class of 'people who need help rather than imprisonment'. However, it is also possible that no actual individual is intended and that this is a hypothetical example of 'people who need help rather than imprisonment', i.e. the type 'someone (anyone) who kills their parents'. The nominal group *the parents[,] who have the similar problem* illustrates two cases where *the* is used as determiner in a nominal group of which the reader is not expected to identify a referent. Ontologically, the primary nominal group refers to a potential type, rather than actual *parents*, and would have no overt determiner in a standard English reformulation, while the embedded nominal group (*the similar problem*) has a non-specific interpretation requiring *a* instead of *the* as determiner in a standard English reformulation, as in *parents who have a similar problem*.

It is possible to simply describe the four examples above as two instances of article substitution (*the* for *a*) and two instances of article insertion (*the* instead of \emptyset). However, a closer look at the last example showed that the possibility that some of the TLEC writers employ a style in which, hypothetical, non-specific instances, once evoked, become real and therefore specific must be considered. In this regard, Van Rooy (2008b:350-352) notes the use of nominal groups that are highly specified (by modifiers and definite determiners) with verbs that convey an essentially hypothetical scenario in a qualitative analysis of the linear unfolding of the discourse in one TLEC essay. Of these specific nominal groups in clauses with simple present tense, he remarks:

It appears as if the writer continues to create hypothetical experiential gestalts, and in these gestalts, noun referents are created and developed into specific entities (Van Rooy, 2008b:351).

Just as in the essay analysed by Van Rooy (2008b),¹⁰³ the verbs in the essay excerpt above are mostly in the present tense or marked for modality.

There are 9 instances of the definite article as determiner in ascriptive nominal groups in the TLEC sample, but no such uses in the LOCNESS sample. Ascriptive nominal groups typically occur in Subject Predicative role and typically take the indefinite article, but there are cases where the logical uniqueness of the ascription requires a definite article (e.g. *the best*). Standard English reformulations of all but the first of the examples below would have the indefinite article. Being plural, the nominal group in the first example would have no determiner in a Standard English reformulation.

(425) Almost 65% of the pregnant women in the whole Africa are the victims of Aids/HIV. <ICLE-TS-NOUN-0087.1>

(426) South African teams don't have money. It's the question of loyalty. They can play overseas but they must come home to play <ICLE-TS-NOUN-0016.1>

(427) they will finally be the carrier of the disease <ICLE-TS-NOUN-0234.1>

(428) Africa is the continent with dominant population of black people <ICLE-TS-NOUN-0150.1>

(429) he will come back to the society. He will be the right person again. <ICLE-TS-NOUN-0303.1>

The ascriptive nominal groups in predicative position in the examples above all have modifiers that specify the class of thing to which the subject nominal group is assigned. In the first four examples the modifier takes the form of an *of*-phrase. The adjective *right* in the last example above does not, in this context, imply a logically unique entity, as in *the right man for the job*. It is used in the sense of 'good'. The essay writer is referring to a hypothetical, formerly criminal, now rehabilitated individual's return to society, implying *He will be a good person again*.

¹⁰³ <ICLE-TS-NOUN-0078.1>

In the two examples below, the nominal group is a complement of the preposition *as*. Here the *as*-phrases are used to ascribe a social role to the writer.

(430) [...] so as the society we should encourage each other to talk about it.
<ICLE-TS-KIMC-0382.1>

(431) As the concerned member of the society, irrespective of which society
in this world <ICLE-TS-NOUN-0011.1>

In the first example above membership to a specific society is implied and a Standard English formulation would require either the indefinite article or the addition of modifiers to help identify the society in question. In the second example above the supplemental information in the appositive group makes it clear that no specific society is intended. Here a Standard English reformulation of the ascriptive nominal group would have no determiner before *society* (which is non-specific) and an indefinite article as determiner of the nominal group as a whole; viz. *as a concerned member of society*. As noted earlier, the use of the definite article with the noun *society* is variable and partly attributable to conventionalization. These two ascriptive *as*-phrases in the TLEC sample should also be related to the use of *as*-phrases that generally show the writer's sense of belonging to a social group (even if undefined) or to humanity in general as in *we as people* discussed in Chapter 4.

Related to the use of the definite article in ascriptive *as*-phrases is the use of the definite article in nominal groups denoting unique roles within a certain context. In the underlined nominal groups in the examples below, reference is made not to an actual specific individual, but to a specific role and by implication to any individual potentially fulfilling that role at a certain time. It is this role which is actual and identifiable within a given context.

(432) Now women were not allowed to be the head of the house or the leader at work <ICLE-TS-NOUN-0429.1>

The next two examples involve ascriptive nominal groups which characterize the extent of some process and are complements of the preposition *at*. The preposition phrases are adverbials of extent.

(433) [...] the HIV/AIDS virus will be spread all over the country at the faster rate

(434) <ICLE-TS-NOUN-0156.1>

(435) Now our day's we can see that the rate of HIV/AIDS epidemic increased at the higher level, because of poverty. <ICLE-TS-NOUN-0116.1>

The non-particular uses of the definite article discussed so far (with the exception of the two nominal groups denoting 'unique role') occur only the TLEC. The four instances of non-specific nominal groups and the nine instances of ascriptive nominal groups with the definite article suggest a unique discourse function of *the* in BSAfE. It would seem as though the non-actual class representatives in the nominal groups that were first analyzed as 'non-specific' are construed as actual entities, or rather become 'actual' and 'particular' in the unfolding of the discourse. The question is whether such entities are so real in the writer's mind that they become 'known' or 'identifiable', i.e. whether the definite article indicates referent identifiability in these nominal groups. This is not possible to say based on only four examples, but seems to be consistent with the observations of Van Rooy (2008b).

Ascriptive nominal groups assign the Thing to a particular class (type). This class is specified by modifiers in all five examples of ascriptive subject predicatives, in one of the two instances where the ascriptive nominal group is the complement of *as* and in both instances where the ascriptive nominal group is the complement of *at*. Since 8 of the nine ascriptive nominal groups have modifiers, it is postulated that modification motivates the use of the definite article, perhaps indicating that the class to which the subject Thing is ascribed is 'identified' by these qualities.

Nominal groups that represent a type (class, *genus*) and not one or more actual instances have a generic interpretation. For example, *the feminists*, *the homeless*¹⁰⁴ and *the people* in the sense of ‘general populace’. Such generic nominal groups sometimes include an additional classifying element, as in *the Black people*.¹⁰⁵ There are 49 instances of nominal groups with generic interpretation in the TLEC concordance sample of *the*, and 32 in the LOCNESS sample. In both corpora the majority of these generic nominal groups do not contain modifiers (25 in the TLEC, 21 in LOCNESS). There are 16 cases in the TLEC and 11 in the LOCNESS data where the class is further specified by additional prenominal qualities. These premodifiers are accompanied by postmodifiers in four of the instances in the TLEC data and in one instance in the LOCNESS data. The example below illustrates a generic nominal group with both pre- and postnominal modifiers.

(436) Looking at difference places especially the needy one with little knowledge, they get easily influence <ICLE-TS-NOUN-0440.1>

There are 8 instances of generic uses of *the* in the TLEC sample where the classifying information is provided postnominally only. There are no such instances in the LOCNESS sample in this category. In the two examples below, the noun is in the singular form, which is consistent with the generic use of the definite article when the typical representative of a species is intended, as in *the tiger*. Yet, these nouns have plural interpretations, based on the rest of the clause (*they*, *people*) and denote humans and not species of animal.

(437) The student who are studying Computers they are able to use them for their prac <ICLE-TS-NOUN-0444.1>¹⁰⁶

(438) The black in Africa are the most people that have got that disease <ICLE-TS-NOUN-0320.1>¹⁰⁷

¹⁰⁴ <ICLE-TS-NOUN-0439.1> <ICLE-TS-NOUN-0257.1>

¹⁰⁵ <ICLE-TS-NOUN-0075.1>

¹⁰⁶ This example also illustrates the use of the copy pronoun in left dislocation discussed in Chapter 4.

¹⁰⁷ This example also illustrates an idiosyncratic predicative construction with superlative *most*, which will be discussed in more detail in Chapter 6.

According to Payne and Huddleston (2002:403), in definitions, both the definiendum and the definiens are non-referential. The use of the definite article in a definition is in line with the conventions of Standard English as the following examples illustrate:

(439) The European Community is the grouping of 12 sovereign European States whose intention is to develop closer economic and political cooperation. <ICLE-BR-SUR-0029.3>

(440) Society is in actual fact the people who are part of the community in a certain place. <ICLE-TS-NOUN-0416.1>

In the following example of a definition from the TLEC concordance sample, the indefinite article might be expected in a Standard English reformulation. However, the use of a proximal demonstrative before the proper name *Aids*, as well as the fact that *Aids* is the acronym for Acquired Immune-Deficiency Syndrome, suggests that this is not a typical non-referential definition, but an explanation of what the word *Aids* stands for.

(441) This Aids is defined as the acquired immune deficiency syndrome <ICLE-TS-NOUN-0254.1>

5.4.1.3 *The use of the in fixed expressions*

The last analysis category under consideration is the conventionalized use of the definite article. This includes the use of *the* in nominal groups denoting certain human institutions or cultural objects or locations (e.g. *the bank, the law, the mirror, the street(s)*). There are 28 cases in the TLEC sample and 17 cases in the LOCNESS sample assigned to this analysis category. A typical example of this category is *in the long run* which occurs in the concordance samples of both corpora. Only 4 of the 28 instances in this category in the TLEC employ the definite article in ways in which it would not be used in Standard English. In this regard it is interesting to note that Minow (2010:166) remarks that inclusion of the uses of articles in fixed expressions would not have much effect on the article suppliance rates of her study. Three of the four so-called

“non-standard” uses of *the* in fixed expressions in the TLEC entail the use of nouns denoting institutions (cf Sand, 2004: 290-291 and the discussion in Section 5.3 above).

(442) like Mr Colleen Chaoke the one who was vandalising the banks,
escaping from the jail <ICLE-TS-KIMC-0339.1>

(443) I am saying there are viry much for valuable than the ones who comes
from the university. Evenly if they don't have degrees they are valuable
<ICLE-TS-NOUN-0471.1>

(444) This is by means of technology bcease the television and the radio
uses signals. <ICLE-TS-NOUN-0434.1>

The examples above primarily highlight the arbitrary nature of the way in which articles have been conventionalized in Standard English and other varieties of English. One goes *to the shop* and *to the bank*, but *to Ø jail*, *to Ø school*, *to Ø university* and *to Ø hospital*. In English we find both the pattern **Prep the N**¹⁰⁸ (where N denotes a human institution) and the pattern **Prep Ø N** (where N denotes a human institution) but no obvious rule that explains the context in which the one is appropriate and the other is not. Perhaps the time spent at the institution in question influences the article choice, with institutions where longer periods are spent being associated with the zero article. In the last example of the set above, both underlined nominal groups would be without an article in a Standard English reformulation, viz. *television and radio*, but the conventions of Standard English are more complicated than that: one sees something *on Ø television*, but one hears something *on the radio*. In the three examples above an underlying systemic difference, or as Ionin, Ko and Wexler (2004) would have it, eventually settling on the appropriate Parameter Setting for Article Choice, is not at issue. Here the Idiom Principle rather the Open Choice Principle is at work. The examples above provide evidence that the input of the speakers of the non-native variety, rather than the speakers of a native variety, which may be held up as an external norm, provide the input for the conventionalization process (Van Rooy, 2010:7, 2011:189ff). However, it should be born in mind that only a small portion (14%) of the instances of the conventionalized or idiomatic uses of the definite article differs from that

¹⁰⁸ Prep=preposition

of Standard English. The next example is a lexical expression in its own right from which a frame-like pattern (as those given above) cannot be deduced.

(445) Our standard of football is not up to the scratch <ICLE-TS-NOUN-0143.1>

From a psycholinguistic perspective, expressions such as the one above are likely in a second language variety since learners, as a result of more mature cognitive abilities, tend to acquire schemas and learning is less instance-based than in first language acquisition (Van Rooy, 2010:3). It is possible that a schema with the definite article has emerged in BSAfE and that this schema is applied to expressions which would be purely idiomatic (in the sense of 'involving instance-based acquisition') in Standard English. From a sociolinguistic perspective, application of such a schema means that expressions such as the four quoted above will become part of the feature pool of BSAfE and may become conventionalized (Van Rooy, 2010:3, 7-8).

5.4.1.4 Summary of the discourse functions of the definite article

The definite article is typically used to identify referents in the discourse in both corpora. However, there are more referent-identifying uses of *the* in the LOCNESS sample. The initial analysis, taking the factuality of the propositions in which the definite nominal groups occur into account, shows that there are more occurrences of the definite article in nominal groups with non-particular interpretations in the TLEC compared to LOCNESS. However, if the four nominal groups with non-specific interpretations are considered in respect of text production, Van Rooy's (2008b) postulation that the entities are construed as specific, while the events in which they participate are still essentially hypothetical, is supported. The use of the definite article in ascriptive nominal groups seems to be attributable to the specification of the class by means of modifiers. The definite article is used more in idiomatic expressions in the TLEC sample than in the LOCNESS sample. However, there are only four instances in the TLEC sample where the definite article is used in contexts where it would not be used in Standard English. In all four of these cases the nominal group is a complement of the preposition *of*. This suggests that a schema where the definite article follows the preposition *to* is generalized in the acquisition process.

The data sample (400 concordance lines) is large enough to allow the identification of potential emerging patterns, but not large enough to conclude that such structures are indeed features of BSAfE. Textlinguistic analysis involving referent-tracking and due consideration of the construal of events through verbs, as exemplified by Van Rooy's (2008b) analysis of one of the TLEC essays, of more BSAfE texts (spoken and written) is needed to confirm the hypothesis that participants are construed as specific even when the events in which they participate are construed as non-specific (general or hypothetical). Concordance searches (in part-of-speech-tagged corpora) will be able to confirm whether the use of the definite article in ascriptive subject predicatives is triggered by the presence of modifiers. The use of the definite article after *as*, *at* and *to* need to be examined in concordance-based analyses of prepositional complements. These are all areas for future research.

5.4.2 The indefinite article

As was the case with the definite article, the indefinite article is used relatively less frequently in the TLEC than in LOCNESS (18/1000 vs 25/1000). A random sample of 400 concordance lines of the word *a*¹⁰⁹ was categorized in terms of the discourse function of the article for each corpus. Whereas epistemological identity is at the core of the discourse functions of the definite article, specificity is central to distinguishing amongst the discourse functions of the indefinite article. In standard varieties of English the use of the indefinite article as central determiner is restricted to nominal groups of which the (grammatical) head noun denotes particularized, single entities or single units of measurement. In other words, *a/a(n)*, which originates from the word *one*, is used with nouns denoting single instances of things that are construed as countable. At issue is whether reference is made to an actual, specific **instance** of the thing or to a non-specific (even hypothetical) example of the **type**. This distinction is illustrated by the following pair of examples. In the first example, the writer intends to refer to a specific individual. This individual is introduced by the nominal group *a young handsome guy with a big eyes*. In the second example, the writer refers to girls or women in general, i.e. **any** girl or any woman, and does not intend the reader to think of a specific girl.

¹⁰⁹ The alternative form *an* was not concordanced.

(446) last year at the school where i was attending, there came a young Handsome guy with big eyes <ICLE-TS-NOUN-0116.1>

(447) I don't know how a girl or a woman suppose to live in this world.
<ICLE-TS-NOUN-0212.1>

The various nominal groups determined by the indefinite article, but not picking out a specific instance, i.e. the type-evoking indefinite nominal groups, can be further subcategorized according to their uses and interpretations as shown in Table ?? below.

Table 5.8 Discourse functions of the indefinite article in the TLEC and LOCNESS

	TLEC	LOCNESS
Quantification	66	22
Reference to a specific thing	37	67
Ascription	92	127
Co-text-bound non-particular interpretations	141	143
Generic uses	36	26
Idiomatic uses in fixed expressions	28	15

The biggest difference between the two corpora are found in the use of the indefinite article in quantifying constructions and the idiomatic use of the indefinite article in fixed expressions. There are three times as many uses of *a* in quantifying constructions and nearly twice as many idiomatic uses of the indefinite article in the TLEC data compared to the LOCNESS data. This suggests that the Idiom Principle strongly influences the choice of article in BSAfE.

When the 400-line concordance samples of the two corpora are compared, the most obvious difference lies in the use of the indefinite article in **quantifying** constructions. There are 66 quantifying uses in the TLEC data and only 22 in the LOCNESS data. For purposes of this analysis, quantifying constructions with the indefinite article include all uses of *a* in constructions indicating 'amount', 'extent' and 'duration', ranging from the patterns ***a lot of N*** and ***a number of N*** to expressions such as *a long time* and *a high rate*. Amongst quantifying uses in the TLEC, the most frequent pattern is:

a Quant-N of N¹¹⁰ (28 times in the TLEC data, 9 times in the LOCNESS data)

(448) so it Killing a large number of sexual active people. <ICLE-TS-NOUN-0109.1>

(449) most Africans they are experiencing a lot of problems <ICLE-TS-NOUN-0077.1>

(450) when they meet in April at Ellis Park Stadium, they were a lot of people there <ICLE-TS-NOUN-0253.1>

The other frequently occurring pattern involving the indefinite article in quantifying constructions is one in which the quantifying noun is not followed by an *of*-phrase, and where the nominal group functions as clause-rank Complement or Adjunct. This pattern occurs 17 times in the TLEC sample and 6 times in the LOCNESS data.

a Quant-N (17 times in TLEC data, 6 times in LOCNESS data)

(451) experiencing fenacial problems because the medication itself demands a lot <ICLE-TS-NOUN-0021.1>

(452) Those who are rich can always manage to buy and prolong their lives a bit. <ICLE-TS-NOUN-0252.1>

In the TLEC sample, the quantifying noun most often used in the two aforementioned patterns is *lot*, with the pattern **a lot of N** occurring 15 times and **a lot** as complete nominal group occurring 5 times. It should be noted that there are no occurrences of the quantifying expression *a lot* in the LOCNESS sample of 400 concordance lines. The high instance of quantifying uses of the indefinite article in the TLEC sample compared to the LOCNESS sample is largely due to the prevalence of the expression *a lot* in the TLEC. As a quantifying construction, *a lot (of)* will be discussed in more detail in Chapter 6.

¹¹⁰ Quant-N is quantifying noun (i.e. “package” noun), N=noun and its internal modifiers. In other words, this pattern represents both instances where the quantifying noun is modified and instances where the quantifying noun is not modified.

There are nine instances in the TLEC data where *a* is synonymous with 'one' or 'per/for one,' compared to two such instances in the LOCNESS sample.

(453) Lucas Radebe left for Leeds United almost a decade ago <ICLE-TS-NOUN-0088.1>

(454) she charges R250 an hour <ICLE-TS-WITS-0006.1>

The following patterns with *a* in quantifying constructions occur in the TLEC sample, but not in LOCNESS:

***a short/long(er) temporal-N* (5x)**

(455) which may intum help him to stay for a longer period. <ICLE-TS-NOUN-0249.1>

***a Qual risk* (3x)**

(456) This is dangerous because they change partners and there is a high risk of them getting HIV/AIDS. <ICLE-TS-NOUN-0062.1>

***a few N* (3x)**

(457) Saving a few lives is better than nothing. <ICLE-TS-NOUN-0094.1>

Typical examples of the use of *a* in fixed expressions include *as a result*; *play(-s/-ed) a role*; *make a difference*, *cost a fortune*, *earn/make a living*, *in a way* and *take a look*. There are 28 instances of the idiomatic usage of the indefinite article (excluding expressions of quantity, extent and duration) in the TLEC data and 15 instances of such usage in the LOCNESS data. In addition to having more quantifying uses of the indefinite article, the TLEC data also contain some unusual expressions. The expression *as a result* occurs 4 times each in the LOCNESS data and the TLEC data. The TLEC data contain an additional three instances of *as a results*. A concordance of *as a result** in the full TLEC yields 55 lines of *as a result* and 10 lines of *as a results*. The fact that roughly 15% of the uses of *as a result(s)* take the plural form suggests the

expression with the plural noun is conventionalized in BSAfE. The use of plural endings is discussed in more detail in Chapter 6.

There are three instances in the TLEC concordance sample where *a* is used in fixed expressions which take the definite article in Standard English.

(458) The north-west parks and tourism board has taken a lead in showing the world that indeed South Africa [...]<ICLE-TS-NOUN-0029.1>

(459) in a night you might sleep with four men of which is not good in a sense that you may get infected <ICLE-TS-NOUN-0328.1>

(460) This players they have took our country and put it on a map <ICLE-TS-NOUN-0098.1>

There is also one example in the TLEC data where the indefinite article is used where no determiner would be used in Standard English. However, later in the same sentence the TLEC writer uses the expression without the definite article. In this case it would seem as though the use of a surface article is motivated by the modification of the noun.

(461) to have a protected sex you have to wear condoms when you have sex <ICLE-TS-NOUN-0084.1>

Although the non-referential and generic uses of the indefinite article far outnumber the use of the indefinite article in nominal groups with specific reference in both corpora, it should be noted that there are many more specific referential uses in the LOCNESS sample compared to the TLEC sample (67 vs 37). These represent indefinite nominal groups referring to specific instances which are actual in the sense that they exist, whether in reality or in some linguistically construed world.

(462) Themba was given a gun and was told, that he should be the one that demanded the money. <ICLE-TS-KIMC-0299.1>

(463) And suddenly i noticed three guys trying to open a Audi A4 car's doors with a driver (screwe). <ICLE-TS-KIMC-0297.1>

In the TLEC data there are also some uses of the indefinite article where the referent is identifiable through internal elements in the nominal group.

(464) Our continent is facing a problem of unemployment. <ICLE-TS-NOUN-0056.1>

(465) Technology has come to a point where, we no longer control it. <ICLE-TS-NOUN-0465.1>

(466) This is a way Botswana goverment does in education but not in soccer <ICLE-TS-NOUN-0386.1>

In the contexts above, the definite article would be used in Standard English. There is also one instance in the TLEC data in this category where the indefinite article is used in an *of*-complement where Standard English would use the zero article.

(467) attension can be given to them in a particular place at a certain period of a time, by trained people who can teach them a lesson <ICLE-TS-NOUN-0332.1>

In the example above, it is clear from the rest of the clause that the speaker intends to talk about a hypothetical situation (indicated by the modal verb) and a particular but potential (as opposed to actual) place and time associated with such a situation. The indefinite article before time is probably motivated by the expression *at a certain time*. However, the writer added some specification of duration, *period of*, to indicate that it is not an unspecified point in time nor an unspecified stretch of time, but a period long enough to accommodate a lesson. It is also possible that this is simply a production error.

In both corpora, ascription is the second largest usage category of the indefinite article, though it should be noted that there are more such usages in LOCNESS than in the TLEC (127 vs. 92). A typical example of this category is:

(468) This is quite a dodgy issue. <ICLE-TS-NOUN-0425.1>

Apart from the quantitative difference the indefinite article is used in ascriptive nominal groups in the TLEC in the same way it is used in LOCNESS. This native-like usage of the indefinite article in ascriptive nominal groups provides some context in which to view the ascriptive uses of the definite article discussed above. Comparison of the concordance samples shows that both the definite and the indefinite article are used in ascriptive nominal groups in BSAfE. Since most of the instances with the indefinite article also contain modifiers that help to “identify” the class (69/92), the postulation that the use of the definite article is motivated by class-identifying modifiers is not supported.

Typical examples of generic uses of the indefinite article include *a student* as representative of the class *student(s)*, *a man*, *a woman*, *a prisoner*, *a mother*, etc. Here the instance is representative of the type. The type may be subclassified by way of modifiers, e.g.

(469) A minor criminal can go to jail and come back from jail as a very dangerous <ICLE-TS-NOUN-0510.1>

There are 36 generic uses of the indefinite article in the TLEC sample, and 26 in the LOCNESS sample. In the TLEC data there are two generic uses of the indefinite article with countable plural nouns denoting places.

(470) They do not mind about those who lives in a rural areas. <ICLE-TS-NOUN-0086.1>

(471) People in a poor countries tries with their best to get a living. <ICLE-TS-NOUN-0119.1>

Bound non-particular uses of the indefinite article include nominal groups determined by the indefinite article which have non-particular interpretations based on linguistic markers of non-factual contexts. These include non-specific interpretations in clauses with irrealis scope, as well as negative-bound and quantifier-bound non-particular interpretations.

(472) they want money so that they can have a better live like other people
<ICLE-TS-POT-0203.1 [non-specific]

(473) people find themselves without a source of income <ICLE-TS-NOUN-0072.1> [negative-bound]

(474) God have given everyone a brain <ICLE-TS-KIMC-0292.1> [multiple-situation-bound]

In the TLEC data there are three instances in this analysis category where the indefinite article is used as determiner of a noun which would be construed as uncountable in Standard English.

(475) it is going to be very essential that they get a prescribed medication to ease the rate/growth of the virus, <ICLE-TS-NOUN-0516.1>

(476) the mother say to her child that she must go and seek for a food, it means that she must go to sell her body <ICLE-TS-NOUN-0037.1>

(477) When he has a time to relax he look for another woman in Botswana <ICLE-TS-NOUN-0105.1>

The last example is the second instance in the TLEC sample of the use of *a* with *time* where the zero article might be preferred in standard English. Here one could argue that *time* is intended to be quantifiable in the sense of 'a moment' or 'a chance'. A quick concordance of the noun *time* in the TLEC yields 322 concordance lines of which *a time* occurs 6 times. However, of these six only the two examples given in the discussion above can be reformulated without the indefinite article in Standard English. There are also instances of **a Qual time** in the TLEC concordance of *time*, only one of which would be formulated without the definite article in Standard English (*a high time*¹¹¹). In all 12 concordance lines of **a Qual time**, the use of the indefinite article corresponds to its usage in Standard English. This means that there is no real corpus evidence of a trend

¹¹¹ <ICLE-TS-NOUN-0135.1>

to use *a* with time in contexts where time has an uncountable interpretation, although there is some evidence of the general trend to use the indefinite article with nouns that are uncountable.

The concordance analysis shows that differences between the corpora with regard to the use of the indefinite article are slight. The concordance data do not support a claim that a different system underlies the choice of article in BSAfE. In both corpora the indefinite article is used in both specific and non-specific nominal groups. The data suggest that there is some dilution of the 'quantity of one' meaning of the indefinite article in BSAfE. The differences between the corpora can be ascribed to the use of *a/an* in fixed expressions (e.g. in the high frequency of the quantifying expression *a lot* in BSAfE) and production errors.

In the next chapter, which focuses on quantifying constructions, the use of the indefinite article, as well as quantifiers such as *many* and *few*, with uncountable nouns will be examined to determine whether they are indicative of a reconstrual of uncountable things as countable, or conventionalization. The two instances where *a* is used with countable plural nouns and the three instances where *a* is used with uncountable nouns suggest that the indefinite article is used as an indefinite marker in BSAfE without having the inherent meaning of 'one'. The dilution of the inherent quantifying meaning of the indefinite determiner *some* discussed in Chapter 6 relates to this.

5.5 Demonstrative and possessive determiners

In the analysis of the corpus concordances of the plural noun *people* (Chapter 4), it was noted that in cases where a definite determiner is used in *people*-headed nominal groups, the definite article is used more often in LOCNESS than in the TLEC, whereas demonstratives and possessives are selected more often in the TLEC than in LOCNESS. The question arises whether this tendency in the choice of definite determiner pertain only to the noun *people* or whether it is observable across the whole of the corpora under comparison.

The raw frequencies in Table 5.9 below are based on concordances of the lexical items that may potentially realize the definite determinative element drawn from the untagged versions of the corpora. The full concordances of the demonstrative determiners as well as the possessive *her* were manually POS-tagged with the “set” function in the Wordsmith Concordance tool. This was done mainly to distinguish between the pronominal and the determinative functions of the lexical items in question. In the case of the token *that*, non-demonstrative, subordinating functions were also distinguished and obviously excluded from the analysis. Apart from the word class disambiguation, the concordance lines of the coordinated third person singular possessive determiner (*his/her* and *his or her*) were separated from the concordances of *his* and *her* (and counted separately) to avoid counting coordinated possessive determiners twice. Because the TLEC contains many spelling errors (even after some correction procedures) (Van Rooy & Schäfer 2003; Van Rooy, 2005), some allowance for misspellings of the possessive determiners *their* and *its* were deemed necessary. To this end, samples of the TLEC concordances of the tokens *their*, and *there* were manually checked, and based on this sample, the number of instances of *their* that are misspellings of other words (e.g. *there*) and the number of instances of *there* that should actually be spelt *their* were estimated.¹¹² Similar checks were performed on the full concordances of *its* and *it's*.¹¹³

Once the raw frequencies of each definite determiner have been established according to the procedure described above, there are three ways in which these frequencies can be normalized to compare the usage of possessive and demonstrative determiners in the two corpora. The frequency per 100 000 tokens of each token functioning as demonstrative or possessive determiner gives an indication of the distribution of each determiner word across the corpus. The number of occurrences per 1000 definite determiners focuses our attention on the paradigmatic choice relationship between a

¹¹² The concordance of *their* in TLEC yielded 1693 instances. A sample of 388 lines were checked. In this sample, 7 of the instances are misspellings of other words. It is estimated that 31 words in the full concordance may be misspellings of other words.

$$1693 \div 388 = 4.36 \text{ and } 7 \times 4.36 = 30.52 \text{ (rounded up to 31)}$$

The concordance of *there* in the TLEC yielded 1014 instances. A sample of 394 lines were checked. In this sample 5 of the instances should actually be spelt *their*. It is estimated that 13 words in the full concordance may be misspellings of *their*.

$$1014 \div 394 = 2.57 \text{ and } 5 \times 2.57 = 12.85 \text{ (rounded up to 13).}$$

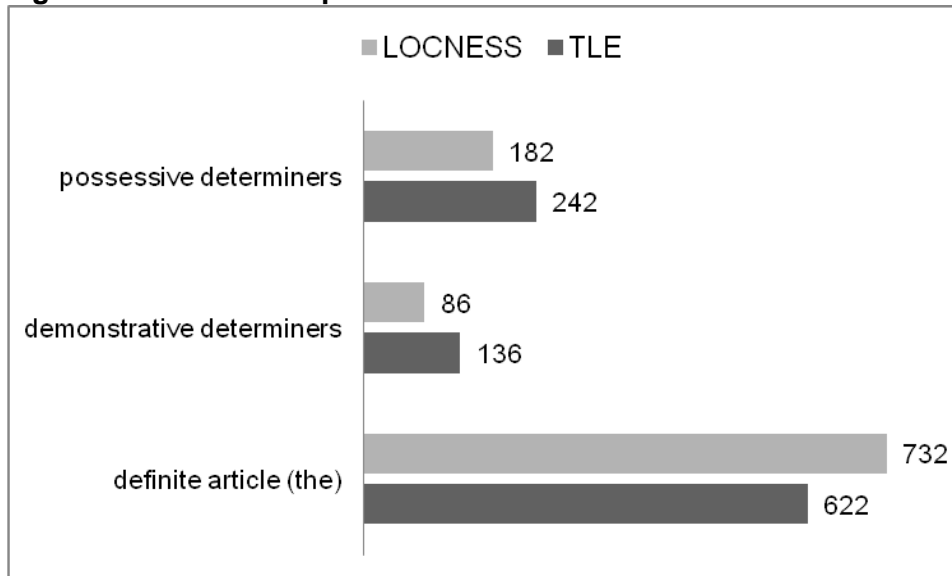
¹¹³ Of the full concordance of *its* (226), only 155 are actually correct spellings of the determiner *its*. There are 13 instances in the concordance of *it's* which should be spelt *its*.

particular lexical item and all the other potential lexical realizations of the definite determiner. The number of occurrences as percentage of either demonstrative or possessive determiners highlights the choice relationship between the closed set of either demonstrative or possessive determiners. The data in Table 5.9 and Figure 5.9 below indicate that this trend in the choice of definite determiner is not restricted to the noun *people*, but indeed applies to the corpora as a whole.

Table 5.9 Frequencies of the main types of definite determiners in the TLEC and LOCNESS

	TLEC			LOCNESS		
	Raw frequency	Normalized frequency per 100 000 in corpus	Proportion of definite determiners in corpus (per 1000)	Raw frequency	Normalized frequency per 100 000 in corpus	Proportion of definite determiners in corpus (per 1000)
Definite article (the)	9085	4514	622	13348	6548	732
Demonstrative determiners	1989	989	136	1568	769	86
Possessive determiners	3527	1752	242	3327	1632	182
TOTALS:	14601	7254	1000	18243	8949	1000

Figure 5.9 Relative frequencies of definite determiners in the TLEC and LOCNESS



In both corpora, the most frequently chosen definite determiner is the definite article. Also, in both corpora, the second most frequent type of definite determiner is the

possessive, followed by the demonstrative. However, considered as a proportion of all the definite determiners used in the corpus (per thousand), the definite article is less frequently preferred as definite determiner in the TLEC than in LOCNESS, whereas the possessive and demonstrative are more frequent in the TLEC compared to LOCNESS. The trend for the choice of definite determiner with the noun *people* noted in Chapter 4 is also evidenced in the rest of the corpus data.

While it is possible to compare the choice of type of definite determiner across nominal groups containing a definite determiner, it is very difficult to compare determiner usage across the board in an unparsed corpus due to the difficulties in accurately determining the amount of nominal groups, including those without overt determiners. So, while the normalized frequencies (per 100 000) in Table 5.9 above indicate that there are more definite determiners in LOCNESS than in the TLEC (8949 vs 7254), it is not possible to determine the actual proportion of nominal groups that are definite, since the corpora are not syntactically parsed. It may well be that there are simply more nominal groups in LOCNESS than in the TLEC. However, the proportional distribution of definite determiner types in definite nominal groups can and deserves to be further investigated since there are obvious differences in preference between the two corpora (as Figure 5.3 shows).

5.5.1 Demonstrative determiners

As pointed out above, the main sources of data for the investigation into the use of demonstrative determiners are concordances from the untagged versions of the TLEC and LOCNESS of each of the words that can function as demonstrative determiner (*this, these, that, those*). However, before dealing with these concordances, data from the error-tagged version of a subcorpus of the TLEC will be considered as a first step in exploring the choice relationship between demonstrative determiners and other determiners, most notably the definite article. From the data in the error-tagged subcorpus of the TLEC the following confusion matrix can be derived:

Table 5.10 Usage of demonstrative determiners in the error-tagged subcorpus of the TLEC

CORRECTION →	this	these	that	those	the	∅
NON-STANDARD FORM ↓						
this		19	0	0	0	9
these	11		0	0	2	3
that	9	0		6	3	1
those	0	5	0		0	1
the	9	8	0	0		
∅	0	4	0	1		

All 97 instances in the concordance where *that* was inserted as correction entails the complementizing or subordinating function of *that* and not the determiner function. Only 30 of the 87 instances of the concordance of *this* as correction involve the determiner function. Of these *this* is suggested as a more idiomatically correct option than *the* in 9 cases and once in the case of *a*. There are 11 cases where the error involved the number features of the proximal determiner, i.e. cases where *these* was used with a non-plural (singular or mass noun). The other 9 cases involve the use of distal *that* in contexts where proximal *this* would be used in Standard English.

(478) a limit of R1000 to be withdrawn a day %%, \$;\$ if you want more than %% the \$this\$ limit you consult your bank <ICLE-TS-NOUN-0144.1>

(479) she %% will \$may\$ just %% came \$come\$ and %% a \$this\$ stranger %% will \$might\$ then rape the poor woman and give her HIV/AIDS. <ICLE-TS-POT-0192.1>

(480) \$and it is sometimes believed\$ that poverty is the main cause of (GNN) %% these \$this\$ disease. This disease is not only (LS) %% incurable \$incurable\$ (QM) %% <ICLE-TS-NOUN-0043.1>

(481) It is very discouraging to hear (GA) %% that \$this\$ kind of statement from a man we all (GVT) %% putted \$put\$ our trust ... <ICLE-TS-NOUN-0014.1>

Of the 45 lines of the concordance of *these* as correction, 36 involve the determiner function. There are 8 cases where *these* are suggested by the L1 editor as replacement for *the*; 19 instances where *this* is used with plural nouns; and 5 instances where distal

those is used in contexts where proximal *these* would be used in Standard English. There are also 4 cases where *these* are inserted as correction where no determiner was used originally.

- (482) The money they use to buy (GNN) the \$these\$ expensive (FS) they \$could\$ have ... <ICLE-TS-KIMC-0360.1>
- (483) All this \$these\$ points are basically the reasons\$ why the province is a good ... <ICLE-TS-NOUN-0036.1>
- (484) (GP) Those \$These\$ 3 factors are (FS) unemployment \$unemployment\$, peer pressure and a <ICLE-TS-NOUN-0212.1>
- (485) They sleep with (WM) these\$ mineworkers for money (QC) they \$sleep\$... <ICLE-TS-NOUN-0007.1>

Seven of the 16 concordances of *those* as correction involve the determiner function. There is one instance where *those* is inserted where no determiner was used originally. The other 6 instances involve the use of *that* with plural nouns.

- (486) Take for example (WM) those\$ people coming from Nigeria (QC) they \$come\$ and ... <ICLE-TS-KIMC-0289.1>
- (487) to see\$ how (LS) their \$the people's\$ huts have been built on (GP) that \$those\$ rocks. <ICLE-TS-NOUN-0041.1>

In the concordance of *the* as correction, there are only five instances where *the* is suggested as a substitution for one of the demonstrative determiners (*these* 2x, *that* 3x), compared to the 17 instances where one of the proximal demonstrative determiners are suggested instead of the definite article. This suggests that the comparatively higher usage of demonstrative determiners in the TLEC is not always due to BSAfE speakers simply preferring to mark the referent as 'identified' rather than 'identifiable'. The possibility that there are more contexts in the TLEC in which a proximal demonstrative determiner rather than the definite article is appropriate should also be considered. This possibility is problematized by the 14 cases (12 of which involve proximals) where demonstratives are used in contexts where no determiner is expected in Standard English.

(488) our economy is very weak because of (WRS) %% this \$0\$ politics.
 <ICLE-TS-KIMC-0359.1>

(489) If (WRS) %% this \$0\$ HIV/AIDS did not exist <ICLE-TS-NOUN-0037.1>

It should be pointed out again that the emphasis of the current investigation is not on the identification of features of BSAfE in terms of non-standardness. It should also be born in mind that the raw frequencies reported in Table 5.10 give no indication of the usage of demonstratives that correspond to their usage in Standard English. Also, the error-tagged data are based on the subjective judgment of a single L1 editor. However, the data are useful in the sense that much of what has been mentioned in the literature on BSAfE and observed in the analysis of the concordance of *people* in the TLEC are confirmed, especially with regard to the interchangeability of the plural and singular forms of the proximal demonstrative determiner. However, the greatest value of the error-tagged data lies in the demonstration that too much should not be made of the choice relationship between the definite article and the demonstratives at the cost of considering the choice relationship between demonstrative determiners and the zero article.

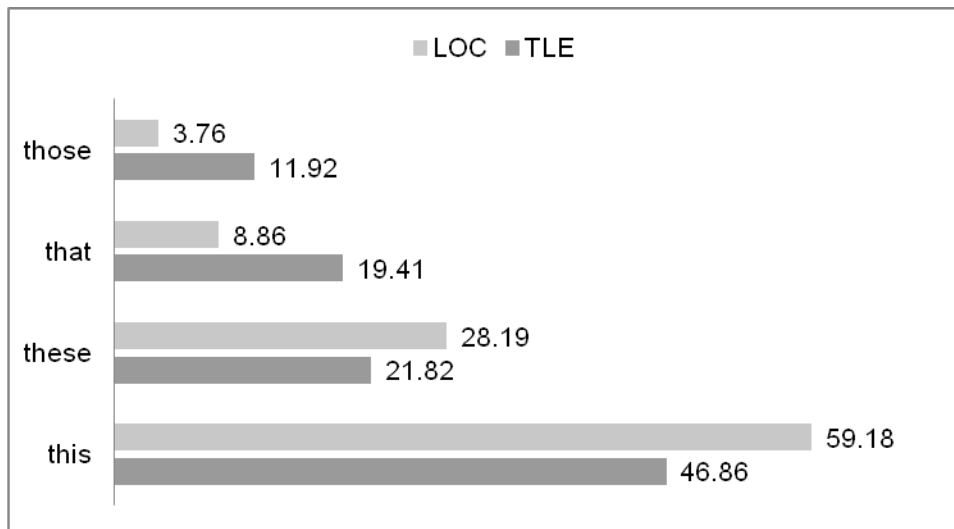
Table 5.11 A comparison of demonstrative determiners in the TLEC and LOCNESS

	Raw frequency as determiner		Frequency as determiner per 100 000 tokens		Frequency per 1000 definite determiners		Frequency per 100 demonstrative determiners	
	TLEC	LOC	TLEC	LOC	TLEC	LOC	TLEC	LOC
this	932	928	463	455	64	51	46.86	59.18
these	434	442	216	217	30	24	21.82	28.19
that	386	139	192	68	26	8	19.41	8.86
those	237	59	118	29	16	3	11.92	3.76
	1989	1568	989	769	136	86	100	100

In terms of normalized frequency per 100 000 tokens, the proximal determiners *this* and *these* show a similar distribution in the two corpora. When *this* and *these* are considered as a proportion of all definite determiners the similarity between the two corpora diminishes. From the perspective of the choice relationship amongst all types of definite determiners, each of the demonstrative determiners is chosen more frequently in the TLEC than in LOCNESS. If one moves further along the systemic network to the point

where the choice has been made to identify the referent by way of a demonstrative determiner, the individual lexical realizations of these choices rank the same for both corpora, with the most frequent lexical realization of demonstrative determiner being *this*, followed by *these* and then *that* and *those*.

Figure 5.10 Relative frequencies of demonstratives in the TLEC and LOCNESS



While the proximal demonstrative determiners outweigh the distal demonstrative determiners in both corpora, it is interesting to note that there are proportionally less proximal and more distal choices in the TLEC (69:31) compared to LOCNESS (87:13). This is also born out in the normalized frequencies per 100 000 tokens. There are nearly three times as many instances of *that* as determiner and four times as many instances of *those* as determiner in the TLEC compared to LOCNESS.

The differences between the two corpora in respect of the use of demonstrative determiners can mainly be attributed to the additional uses of the distal demonstratives in the TLEC.

5.5.2 Possessive determiners

The use of possessive determiners in the corpora is set out in Table 5.12.

Table 5.12 A comparison of possessive determiners in the TLEC and LOCNESS

	Frequency per 100 000 tokens		Frequency per 1000 definite determiners		Frequency per 100 possessive determiners	
	TLEC	LOCNESS	TLEC	LOCNESS	TLEC	LOCNESS
my	93	62	13	7	5	4
our	412	143	57	16	24	9
your	140	44	19	5	8	3
their	832	505	115	56	47	31
its	83	126	12	14	5	8
his	99	621	14	69	6	38
her	61	116	8	13	3	7
his/her	31	15	4	2	2	1
TOTAL	1751	1632	242	182		

Amongst the possessive determiners, the biggest difference in frequency per 100 000 tokens between the corpora is observed with the token *his*,¹¹⁴ which is also identified as a negative keyword in Wordsmith's Keywords tool when the Wordlist of the TLEC is compared to that of LOCNESS . The much higher frequency of *his* in LOCNESS ties in with the high frequency of *he* in LOCNESS. The high frequency of the masculine, third-person, singular determiner and pronoun in LOCNESS can largely be attributed to the large proportion of literary essays which are about a single male protagonist. Fairly large differences in frequency are also observed for the determiners *their*,¹¹⁵ *our*¹¹⁶ and *your*,¹¹⁷ which are more frequent in the TLEC and are identified in Wordsmith as positive keywords. The keyness of these possessive determiners cannot strictly be attributed to essay topic as in the case of *his*.

Out of every thousand nominal groups with definite determiners, 242 in the TLEC and 182 in LOCNESS have possessive determiners. The corpora do not only differ with regard to frequency of possessive determiners, but also with regard to the tokens that

¹¹⁴ Number of occurrences of *his* per 100 000 words in LOCNESS minus number of occurrences in TLEC:

$$621-99=561$$

¹¹⁵ Number of occurrences of *their* per 100 000 words in TLEC minus number of occurrences in LOCNESS:

$$832-505=327$$

¹¹⁶ Number of occurrences of *our* per 100 000 words in TLEC minus number of occurrences in LOCNESS:

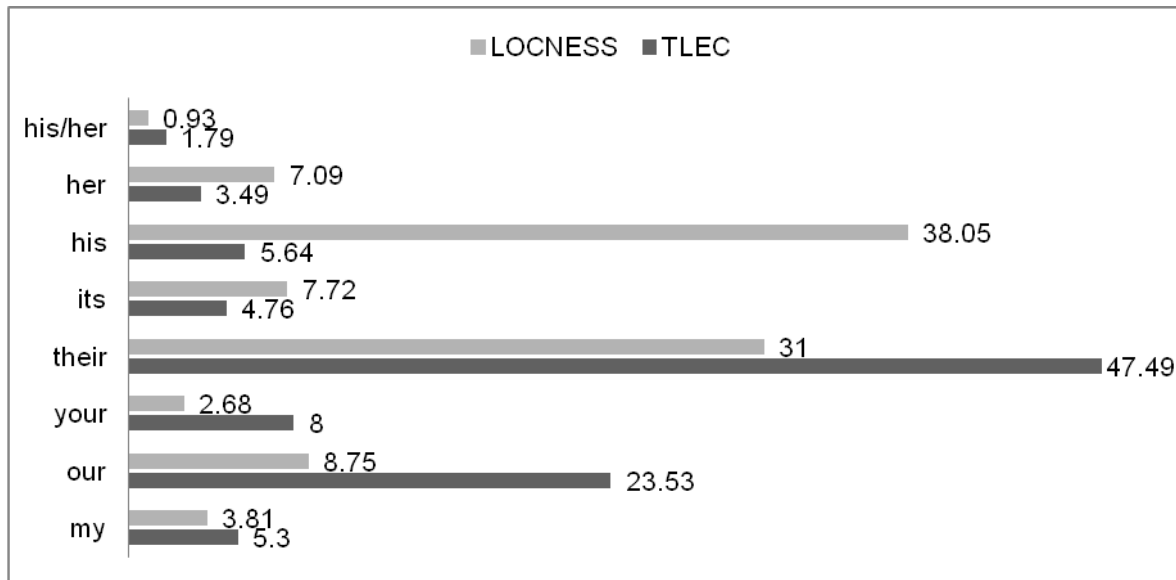
$$412-143=269$$

¹¹⁷ Number of occurrences of *your* per 100 000 words in TLEC minus number of occurrences in LOCNESS:

$$140-44=96$$

realize the possessive determiner. Once the choice has been made to identify the referent of the nominal by means of a possessive determiner, a difference in preference with regard to the person and number of the possessive is observed between the two corpora.

Figure 5.11 Relative frequencies of possessive determiners in the TLEC and LOCNESS



With regard to number, it is interesting to note that the majority of the possessive determiners in the TLEC are plural, while the opposite is true for LOCNESS. Van Rooy (2008a:282) mentions “the general tendency towards more extensive use of the plural in the TLEC, which applies to both pronouns and common nouns”. It seems that this trend is also observable in the possessive determiners. In the TLEC, the most frequently chosen possessive determiner is the third person plural determiner *their*, which accounts for nearly half (47.49%) of all possessive determiners in the corpus, whereas in LOCNESS, the singular third person determiners outweigh the plural form, with the masculine *his* being the most frequently chosen possessive determiner. It has already been pointed out that the very frequent choice of *his* in LOCNESS is probably influenced by essay topics. It should also be noted that the second most frequent possessive determiner in LOCNESS is the plural *their*, and were it not for the many essays on single male protagonists, it might have been the most frequent determiner in this corpus too. In both corpora, the plural form of the first person possessive determiner (*our*) outweighs the singular form (*my*).

The determiner *their* is gender-neutral. With regard to gender, it is interesting to note that there are roughly twice as many coordinated forms of the singular third-person determiner (*his/her*) in the TLEC as in LOCNESS. Gender-specific third person singular possessive determiners are more frequently chosen in LOCNESS than in the TLEC.

In both corpora third person possessive determiners outweigh second and first person possessive determiners combined. However, there are far more third person determiners in LOCNESS (85%) compared to the TLEC (63%), and more first and second person determiners in the TLEC than in LOCNESS. According to Van Rooy (2008a:287), the much higher usage of second person pronouns in the TLEC compared to LOCNESS is indicative of “a more interactive and involved orientation” in the TLEC. In Chapter 4 the use of *our* with *people* was interpreted as an expression of solidarity.

5.6 Conclusion

The concordance analyses in this chapter showed many similarities between the two corpora. In both corpora the definite article is mainly used to mark the referent (including non-actual entities) as identifiable (in more than three quarters of the cases in both corpora). The indefinite article is used in nominal groups with specific referents, as well as in nominal groups with non-specific interpretations in both corpora. There is nearly the same amount of instances where the indefinite article is used in nominal groups with non-specific interpretations in the TLEC and LOCNESS concordance samples. This is the most frequent discourse function of the indefinite article, followed by its use in ascriptive nominal groups in both the TLEC and LOCNESS. In both corpora there are more proximal than distal demonstrative determiners, and in both corpora singular forms of the demonstrative determiners occur more frequently than plural forms. The possessive determiner *their* is the most frequent possessive determiner in both corpora.

The concordance analyses also revealed some interesting differences between the two corpora. There are fewer identifying uses of the definite article and more generic uses of the definite article in the TLEC compared to LOCNESS (301 vs 345 out of 400 and 51 vs 32 out of 400, respectively). There are also more uses of the definite article in fixed

expressions in the TLEC concordance sample (28 vs 17). In the TLEC sample, there are also 9 instances where *the* is used in ascriptive nominal groups. With regard to the indefinite article, the biggest difference between the two corpora lies in the use of *a* in quantifying constructions (66 in the TLEC vs 22 LOCNESS). While there are many more instances of the indefinite article in quantifying constructions in the TLEC sample, there are many more instances of the indefinite article in nominal groups with specific reference in LOCNESS compared to the TLEC (67 vs 37). There are proportionally more distal demonstrative determiners in the TLEC compared to LOCNESS. With regard to possessive determiners, there are more first and second person and plural forms in the TLEC, while there are more third person and singular forms in LOCNESS.

Innovative article usage in New Englishes is attributed to a different conceptual system (Platt, Weber & Ho, 1984). However, the similarities in the discourse functions of the articles in the two concordance samples suggest that article choice in BSAfE cannot be attributed to an underlying conceptual system which is substantially different from that of Standard English. Both BSAfE and Standard English generally use the articles to mark the nominal group as definite or indefinite. Article uses that are unique to BSAfE are mainly due to conventionalization, for instance the use of *the* with nouns denoting institutions and in nominal groups with non-particular interpretations where \emptyset would be used in Standard English. Some of the innovations pertaining to articles, such as the use of *the* in ascriptive nominal groups as in *It's the question of loyalty*,¹¹⁸ probably originated from learner errors. However, it should be kept in mind that the use of the indefinite article in ascriptive nominal groups is the second most frequent function of the indefinite article in both corpora. This suggests that the definite article has not subplanted the indefinite article in ascriptive nominal groups and that its use in such constructions has not stabilized in BSAfE. Analyses of the demonstrative and possessive determiners generally confirm Van Rooy's (2008a) observations.

¹¹⁸ <ICLE-TS-NOUN-0016.1>

CHAPTER 6

Quantification

6.1 Introduction

In the analysis of the concordance of *people* (see Chapter 4), the largest quantitative difference between the two corpora was observed in the use of *people* in *of*-phrases that complement quantifiers such as *some* and *most*. *Most* and *some* are also positive keywords in the TLEC when compared to LOCNESS. The data therefore indicates that the use of quantifiers in the TLEC warrants further investigation.

Words that realize a quantifying function are syntactically multi-functional and typically able to function as quantifying determiners or quantifying pronouns. For purposes of analysis, quantifiers that can fill the determiner slot in the nominal group and/or those that can occur in quantifying *of*-constructions were selected for concordance analysis. The selection of quantifiers, which is mainly based on Payne and Huddleston's (2002:356) list of basic determiners, encompass *all*, *both* (universal), *no*, *none* (negative), *some*, *any* (existential), *either*, *neither* (disjunctive), *each*, *every* (distributive), *enough* (sufficiency), *many*, *much*, *a lot*, *lots*, (multal quantifiers), (*a*) *few*, *little*, *several* (paucal quantifiers), as well as the inflected forms of the multal and paucal quantifiers, viz. *more*, *most*, *fewer*, *fewest*, *less*, *least*.

Many quantifiers (e.g. *any*, *some*, *most*) have determinative as well as pronominal uses. With regard to the pronominal uses, a distinction can be made between quantifiers that serve as an independent clausal subject or complement and quantifiers that (grammatically) head partitive *of*-constructions. With regard to the determinative uses of *all* and *both* a distinction is made between central determiners and predeterminers. These pronominal and determinative uses are illustrated in the following examples of *all* as central determiner, as predeterminer, as logical head of a partitive *of*-construction and as pronominal clause participant respectively:

(490) All trade barriers will be broken down so as to encourage and promote free trade between the countries of Europe, [...] <ICLE-BR-SUR-0025.3>

(491) De Gaulle made use of all his principal powers in his first four years in office [...] <ICLE-BR-SUR-0024.1>

(492) [...] and all of the evidence that the author presents must support the claim. <ICLE-US-MRQ-0026.1>

(493) Voltaire presents much human misery and suffering indicating that all is indeed not well in society; [...] <ICLE-BR-SUR-0004.2>

Payne and Huddleston (2002:364-365) distinguish between the proportional and the non-proportional uses of the quantifiers *some* and *most*. When *some* or *most* means 'not all' it has a proportional meaning. When *some* is used purely as an existential determiner in contexts where the 'not all' meaning is not activated, it is interpreted non-proportionally. Likewise, the superlative sense of *most* is interpreted as non-proportional. Payne and Huddleston provide the following examples to illustrate the proportional and non-proportional uses of *some* and *most*.

(494) Some people misunderstood the question. [proportional] (Payne & Huddleston, 2002:364)

(495) Most students would regard that as unreasonable. [proportional] (Payne & Huddleston, 2002:365)

(496) I saw some children climb over the fence. [non-proportional] (Payne & Huddleston, 2002:364)

(497) It's the Psychology Department that attracts (the) most students. [superlative] (Payne & Huddleston, 2002:365)

In addition to determinative and pronominal uses, some quantifiers (e.g. *much* and *a lot*) have adverbial uses. One could argue that conceptually these adverbial uses still represent a quantificational function, but instead of quantifying things, they quantify the

process, in terms of frequency or intensity, or they indicate the degree of some quality, as can be seen from the examples below:

(498) Jean Capelle, a leading figure in the initiative had very much admired, and tried to emulate the English comprehensive schools. <ICLE-BR-SUR-0032.1>

(499) The real Confederate flag is much different. <ICLE-US-SCU-0001.2>

Some of the words that can function as quantifiers also have non-quantifying functions. For instance, *little* can be used as an adjective denoting size.

(500) A little strip joint is typically viewed as a trashy place with trashy people. <ICLE-US-SCU-0006.2>

Likewise, the words *both*, *either* and *neither* have both quantifying and coordinating syntactic roles. The word *no* can function as a quantifying determiner or a discourse item, i.e. the opposite of *yes*. The examples below illustrate the quantifying as well as non-quantifying uses of *both* and *no*.

(501) Both sides of this issue have strong and weak reasoning behind their arguments. <ICLE-US-MRQ-0037.1>

(502) Watching executions on television will be both well-received and rejected by those who choose to watch these events. <ICLE-US-MRQ-0034.1>

(503) No, to me no murder is right, not even the death penalty. <ICLE-US-MRQ-0016.1>

The non-quantifying uses of words like *little*, *no*, *either*, *neither* and *both* are not relevant to this investigation into quantification. For purposes of this study a distinction is made between quantification in its broadest sense (which includes adverbials, for instance) and nominal quantification, which is restricted to quantifying determiners and quantifying

pronouns. This study focuses on the latter. In the next section some of the claims about quantifiers in literature on non-native varieties of English (including BSAfE) will be summarized. Thereafter a quantitative analysis of the quantifiers in the TLEC and LOCNESS and a more in-depth investigation of the usage patterns of *some* and *most* will follow.

6.2 Quantifiers in non-native varieties of English

Mesthrie and Bhatt (2008:52) note great variability with regard to plural marking in individual World Englishes and go on to summarize a number of general tendencies. Two of these are well-documented in literature on BSAfE, namely “non-use of plural -s where expected in the superstrate” (see De Klerk & Gough, 2002:362; Mesthrie, 2004:971) and “treatment of mass nouns as count nouns” (see Van der Walt & Van Rooy, 2002:124; De Klerk & Gough, 2002:362; De Klerk, 2003a:472; 2003b:230; Minow, 2010:61).

Platt, Weber and Ho (1984:59) note that the collocational restrictions based on the countability of the noun in British and American English do not obtain for New Englishes. In Standard English, quantifying expressions such as *a few* and *many* are used only with countable nouns, while quantifying expressions such as *(a) little* and *much* typically collocate with uncountable nouns. The tendency to use countable nouns with *much* has also been observed in BSAfE (De Klerk, 2003a:473; 2003b:230, Minow, 2010:61). Siebers (2012:136) provides examples of such instances in her BSAfE data:

(504) And there aren't quite much companies that are around (M1) (Siebers, 2012:136).

De Klerk (2003a:472) refers to the “loss of distinction between mass and count nouns” and finds three sets of evidence from her corpus data for this phenomenon, namely (1) “a fair number of plurals for non-count nouns”, (2) “the use of *a* and *the* with non-count nouns”, and (3) “the use of *too much* instead of *too many* with count nouns”. While the use of *a/an* with non-countable nouns might be seen as construing a noun as countable, the same cannot be said for the occurrence of *the* in contexts where Standard English

would have no overt determiner, since *the* can (and often does) occur with non-countable nouns (also see previous Chapter with regard to the use of *the* in nominal groups with non-identifying interpretations in BSAfE). In the analysis of the corpus data in the next section, De Klerk's claims about "[t]he coalescence of mass and count nouns" (2003a:472) will be probed. In this regard, the pluralization of uncountable nouns and the semantic types of nouns used with *much*, *many* and *most* will be examined.

In New Englishes the *a* in quantifying expressions such as *a few*, *a number of* or *a couple of* is omitted or replaced by *some*. In this regard, Platt, Weber and Ho (1984:60) remark that "[t]he use of *a* seems incompatible with *few* and *couple* which are associated with the plural".

(505) This money is given to the girl to buy few articles (from West African English) (Platt, Weber & Ho, 1984:60)

(506) You are expected to say some few words (from Sri Lankan English) (Platt, Weber & Ho, 1984:60)

(507) Some few minutes past nine I leave the office (from West African English) (Platt, Weber & Ho, 1984:60)

The use of *some few* as quantifier in BSAfE is widely reported (De Klerk & Gough, 2002:363; Minow, 2010:63). Van der Walt and Van Rooy (2002: 124) report a 72% acceptance rate of *some few* amongst BSAfE learners. De Klerk (2003b:232) reports 10 instances of *some few* in her corpus of over 540 000 words of spoken Xhosa English. Minow (2010:63) reports only two instances in her database of 85 431 words of spoken BSAfE. She provides the following example:

(508) I think like some few weeks ago, (.) they had a official opening of one of the bars, (..) so. (BSAEF06) (Minow, 2010:63).

De Klerk (2003a:477) reports 56 instances of the use of *each and every* as a unit in her data (540 000+ words of spoken Xhosa English). The use of *other...other* expressions instead of *some...other* in BSAfE is also well-documented (Van der Walt & Van Rooy, 2002:124; De Klerk & Gough, 2002:363; De Klerk, 2003a:470; Mesthrie, 2004:969). The concordance search strings for the data to be investigated in this chapter do not include

other. However, in the analysis of nominal groups with *people* as Thing (see Chapter 4), the *other...other* construction is discussed.

De Klerk (2003a:470) mentions three occurrences of *the most thing* in her corpus of over 540 000 words. The use of *the most thing* for the ‘the thing I [verb] most’ is also mentioned by De Klerk and Gough (2002:363) and Mesthrie (2004:971).

Observations of the use of quantifiers in BSAfE are limited to inventories of non-standard uses of quantifiers that are especially noticeable to the L1 ear. No attempt is made to discover what motivates the (non-standard) quantifying constructions that are supposed to be characteristic of BSAfE. The only notable exception is Mesthrie’s (2006:139-140) observation that “the use the partitive genitive *more of, most of, too much of* for StE ‘more’, ‘most’, ‘too much’” should be attributed to the addition of an extra morpheme rather than to undeletion. He provides the following example from his data:

(509) *Most of the people* are hailing from Malawi (Mesthrie, 2006:140)

No significant attempt has been made to investigate how the function of quantification is realized in BSAfE by means of quantitative and qualitative analyses of corpus data. This is the aim of the remainder of this chapter.

6.3 Quantitative overview of quantifiers in the TLEC and LOCNESS

Concordances of the following quantifying words were drawn from the untagged versions of the TLEC and LOCNESS: *all, both, no, none, some, either, neither, any, each, every, enough, many, much, a lot, lots, (a) few, little, several, more, most, fewer, fewest, less, least*. These search words yielded a total of 5185 concordance lines in the TLEC and 4163 concordance lines in LOCNESS. These concordance lines were coded according to their syntactic roles (i.e. word class). The following syntactic roles were expected in the data: predeterminer, central determiner, pronoun as independent clausal participant, pronoun complemented by *of*-phrase (i.e. part of partitive *of*-construction), adverbial, adjective, discourse item, and coordinator. Since the aim of this analysis is to examine quantification involving the nominal group, the focus will fall on concordance

lines coded as one of the first four syntactic roles (i.e. quantifiers used as determiners of pronouns). Although it could be argued that adverbial uses of potentially quantifying words fulfill the function of quantification, the frequency of such quantification (in the broader sense of the word) in the data will be reported but not analyzed further. Discourse and coordinating functions can obviously not be counted as instances of quantification and they were isolated from the rest of the data. Although their frequencies can be reported, these usages will not be taken into account.¹¹⁹ The use of *little* as attributive adjective describing size is also not counted as quantification.¹²⁰

The analysis categories above can be derived from any descriptive grammar of English and were determined beforehand. However, once the categorization of the concordance lines was underway it became apparent that the categories above do not cover all the usages in the concordances. The concordance data from both corpora included some instances where the search word was a misspelling of another word (e.g. *no* for *know*) or where the search word is part of a pronoun or an adverb that is usually spelt as one word (e.g. *no body* instead of *nobody* and *no where* instead of *nowhere*). There are 136 such misspellings in the TLEC concordance data and 59 in LOCNESS. These were not counted as instances of quantifiers.

Allowance also had to be made for quantifiers that occur as logical heads of predicative elements, for example after copula verbs. In Standard English, this is best exemplified by the pattern *X is enough/sufficient* and the fixed idiomatic expression *X are few and far between*.

(510) It is just not enough to be effective. <ICLE-US-MRQ-0022.1>

(511) From 1958 to 1973, economic growth, full employment and prosperity meant that industrial disputes were fairly few and far between. <ICLE-BR-SUR-0019.1>

¹¹⁹ *No* as discourse item: 16 in the TLEC, 17 in LOCNESS.

Either as coordinator: 18 in the TLEC, 33 in LOCNESS.

Neither as coordinator: 1 in the TLEC, 7 in LOCNESS.

Both in a coordinating construction: 44 in the TLEC, 66 in LOCNESS.

¹²⁰ *Little* as size-describing, attributive adjective: 8 in the TLEC, 24 in LOCNESS.

Predicative quantifiers will be discussed in more detail below since they can be regarded as quantifying constructions. In the TLEC concordance data, an analysis category was added to identify unique, non-standard expressions of quantity or superlativity. This category covers the often-mentioned, under-explained *the most thing*. In fact, 37 of the 38 instances in this category involve the word *most*. These will be addressed in the detailed analysis of the concordance of *most* in the TLEC in Section 6.4.2 below.

Concordance searches of individual words will inevitably yield two (or more) concordance lines for nominal groups containing more than one quantifier, i.e. where quantifiers occur as submodifiers of other quantifiers (e.g. *much more*) and where quantifiers are coordinated (e.g. *each and every*). These should be regarded as multi-word quantifier units that fill a single Numeral slot in the nominal group. In order to avoid counting such constructions twice, they were coded separately in the concordance data. These complex quantifiers will be discussed in more detail later. First, the single-word quantifiers will be examined.

6.3.1 Single-word quantifiers

If one excludes misspellings and obvious non-quantifying uses like those mentioned above, as well as concordance lines containing multi-word quantifiers, one is left with 4774 concordance lines of single-word quantifiers in the TLEC, and 3800 in LOCNESS. These refer to quantification in the wider sense of the word. Per 100 000 words there are more single-word quantifiers in the TLEC than in LOCNESS (2381 vs 1884).

The single-word determinative and pronominal quantifiers in the TLEC and LOCNESS are compared in Table 6.1. The totals show that determinative uses far outweigh pronominal uses in both corpora (see Table 6.2). In terms of raw frequencies there are more quantifying central determiners in the TLEC than in LOCNESS. However, the normalized frequencies per 1000 determinative and pronominal quantifiers indicate that quantifying central determiners are more prevalent in the LOCNESS data than in the TLEC data (735 vs 644).

Table 6.1 Single-word determinative and pronominal quantifiers in the TLEC and LOCNESS

Syntactic role:	Predeterminer		Central determiner		Grammatical head in <i>of</i> -partition		Pronoun as clause participant	
	TLEC	LOCNESS	TLEC	LOCNESS	TLEC	LOCNESS	TLEC	LOCNESS
some	0	0	514	253	186	38	137	29
no	0	0	366	305	0	0	0	0
any	0	0	158	212	3	12	2	3
another	0	0	86	128	0	0	27	26
each	0	0	8	85	1	9	39	42
every	0	0	107	94	0	0	0	0
enough	0	0	77	12	1	3	33	1
either	0	0	1	5	0	0	0	1
all	140	94	135	175	25	57	116	161
both	1	0	9	56	1	4	9	33
neither	0	0	0	3	0	1	0	3
many	0	2	485	356	22	44	17	31
much	0	0	48	77	7	12	17	20
few	0	0	25	43	3	2	10	2
little	0	0	40	29	1	1	9	10
none	0	0	0	0	4	9	1	4
a lot	0	0	1	0	112	31	24	6
lots	0	0	0	0	21	3	0	0
several	0	0	15	48	1	2	0	0
fewer	0	0	2	7	0	0	0	0
fewest	0	0	0	0	0	0	0	0
least	0	0	1	2	0	0	0	1
less	0	0	33	20	0	0	5	3
more	0	0	274	166	4	8	129	32
most	0	0	248	88	350	39	1	3
TOTAL	141	96	2633	2164	742	275	576	411

Table 6.2 Types of determinative and pronominal quantifiers in the TLEC and LOCNESS

Syntactic role:	Predeterminer		Central determiner		Grammatical head in <i>of</i> -partition		Pronoun as clause participant	
	TLEC	LOCNESS	TLEC	LOCNESS	TLEC	LOCNESS	TLEC	LOCNESS
per 1000 quantifiers in the wider sense	30	25	552	569	155	72	121	108
per 1000 determinative or pronominal single-word quantifiers	34	33	643	735	181	93	141	140

The biggest discrepancy between the corpora can be observed with quantifying partitive *of*-constructions. In terms of normalized frequency per 1000 quantifiers (in the wide sense of the word) there are more than twice as many partitive *of*-constructions in the TLEC than in LOCNESS (155 vs 72). The quantifier most frequently used in partitive *of*-constructions in LOCNESS is *all* (57), followed by *many* (44), *most* (39) and *some* (38). In the TLEC the most frequently used quantifier in *of*-partitions is *most* (350), followed by *some* (186) and *a lot* (112). The quantifiers *all* and *many* are used in *of*-partitions only 25 and 22 times respectively in the TLEC. The high frequency of partitive *of*-constructions in the TLEC can be attributed to the high frequency of the quantifier *a lot*, which cannot be used determinatively.¹²¹ The quantifier *a lot* occurs three and a half times more in the TLEC than in LOCNESS (147 vs 42). *A lot* is used in partitive *of*-constructions 112 times in the TLEC and 31 times in LOCNESS. The high frequency of partitive *of*-constructions headed by *most* and *some* (both of which are strong positive keywords in the TLEC when compared to LOCNESS) also plays a role. In fact, in the TLEC, the partitive use of *most* outweighs its determinative use (350 vs 248), whereas in LOCNESS there are more than twice as many determinative uses of *most* compared to

¹²¹ In the TLEC there is one concordance line where the usage of *a lot* is classified as determiner based on the absence of the word *of* in the expression *be given a lot money* <ICLE-TS-NOUN-0226.1>. However, later in the essay the writer uses the same expression with *of*: *be given a lot of money*.

partitive uses (88 vs 39). *Of*-partitions with *most* and *some* will be dealt with in more detail in Section 6.4 below.

A closer look at the 112 concordance lines of *a lot of* in the TLEC reveal that the nominal group *a lot of money* accounts for about a third (33) of all instances of *a lot* in *of*-partitions. One tenth of the concordance lines involve *a lot of people* (13). The thing quantified is uncountable in 69 of the concordance lines (which includes the 33 instances of *a lot of money*), and countable in 43 instances. The use of *a lot of* with uncountable nouns goes some way in explaining the relatively low frequency of *much* in the TLEC.

A cursory glance at the frequencies of each quantifier in its function as determiner (see Table 6.1) indicates that there are certain words like *some*, *many* and *most* that occur very frequently in the TLEC while the load is spread more evenly between the quantifying determiners in LOCNESS. From this the question arises whether these quantifiers convey extended meanings in BSAfE which are conveyed by other words in Standard English. Partridge (2010) finds that fewer lexical verb types are employed in the TLEC in comparison to LOCNESS and that the verbs employed in the TLEC are less specific than those applied in LOCNESS. The possibility that fewer quantifier types and more schematic quantifiers are employed in the TLEC warrants further investigation. As will be seen from the analysis of *some* and *most* in Section 6.4, it is indeed the case that these words have extended meanings in BSAfE which they do not have in Standard English. Given observations about the use of *much* instead of *many* with countable nouns in BSAfE (De Klerk, 2003a:473 & 2003b:230; Minow, 2010:61; Siebers, 2012:136), the possibility that *many* is used with uncountable nouns was also considered.

In Standard English, *many* is restricted to nominal groups headed by countable nouns. This restriction is observed in 99% of the instances of ***many N*** in the TLEC. Only 5 of the 485 instances of *many* as determiner precede uncountable nouns. In the 48 instances where *much* is used as quantifying determiner in the TLEC, there are only three instances where *much* precedes a countable plural noun.

(512) try to find as many information about the danger or the deadly results of HIV/Aids. <ICLE-TS-WITS-0007.1>

(513) In that way South Africa won't be losing so much games as it been losing these "couple of years" <ICLE-TS-KIMC-0291.1>

De Klerk (2003a:472) would describe the examples above as “the use of inappropriate quantifiers”. Even if one only considered the five instances where *many* precedes a non-countable noun and the three instances where *much* precedes a plural countable noun, one would be hard pressed to regard this as evidence of the “**loss** of distinction between mass and count nouns” De Klerk (2003a:472, my emphasis – YVB) (also see Minow, 2010:61; Siebers, 2012:121).

When one considers all the instances of ***many N*** and ***much N***, not only the non-standard ones, such a claim becomes untenable. The analyses of the concordances of *many* and *much* show that there is a clear distinction between countable and uncountable nouns in BSAfE and that it is observed in ways which are very close to that of Standard English. Great caution should be exercised in attributing differences in quantifying constructions to the construal of nouns as countable or uncountable. Reporting a raw number of instances of a (non-standard) usage in a corpus (De Klerk, 2003a,b) without taking into account uses which correspond to those of Standard English leads to the identification of “features”, which are not really fully stable and conventionalized in BSAfE in terms of Van Rooy’s (2011) distinction between innovation and error. What we have here is a few instances of an error rather than a conventionalized innovation.

Now that the determinative and pronominal one-word quantifiers have been considered, we turn to the few cases in the corpora where quantification is part of the syntactic predicate. For example:

(514) It is just not enough to be effective. <ICLE-US-MRQ-0022.1>

(515) From 1958 to 1973, economic growth, full employment and prosperity meant that industrial disputes were fairly few and far between. <ICLE-BR-SUR-0019.1>

There are nine predications of sufficiency (as exemplified in the first example above) and one example of the idiomatic expression of scarcity in LOCNESS. The TLEC contains one example of the formulaic expression *when days are dark, friends are few* (ICLE-TS-KIMC-0348.1). The function of such a predicative quantifier is related to that of a predicative adjective. The expression above can be reworded with an adjective: *friends are scarce*. In the TLEC data, the pattern of the idiomatic expression of scarcity exemplified above is applied to various expressions of ubiquity and plentifulness.

(516) Most of our soccer players have gone to overseas ten years back even now they are still there, players like Lucas Radebe, Mark Fish, John Mosweu players like Hans Fonk and others they are many, and others are stil waiting to go. <ICLE-TS-NOUN-0171.1>

(517) For instance, one day when I got to FNB, so there were so many Botswana students asking for some statements, when it was my turn, the teller told me why we are so many that day, as it seems she was now loosing patience on us. <ICLE-TS-NOUN-0413.1>

(518) Have you ever thought of the effects of retrenchment? One minute you have bulk of money the next minute you are a church mouse. This is because although of the money redeployment is too much, it is too little to help in maintaining the family <ICLE-TS-NOUN-0013.1>

(519) The other reason why poverty was many in Africa is because economically African countries are lagging behind.

The examples above are evidence of an innovative grammatical pattern in BSAfE:

Nomgrp *BE* Quant¹²²

Standard English reformulations of the predications underlined in the examples above would make use of existential *there*:

There are many others still waiting to go.
...why there were so many of us.
... although there is much money...
... why there is (so) much poverty in Africa...

The existential *there*-construction is a marked (non-canonical) syntactic ordering device (Biber et al., 1999:896). In the TLEC examples above, the canonical **Subject Copula Predicative** ordering is used. The existential *there*-construction allows for the quantifier to be placed before the Thing being quantified, either as central determiner or as logical head of an partitive *of*-construction. A parallel can be drawn between attributive vs predicative adjectives and prenominal vs predicative quantifiers. Although, Van Rooy and Terblanche's multidimensional comparative investigations of a number of linguistic features in the TLEC and LOCNESS do not explicitly deal with quantification, they note the more frequent use of *BE* as main verb in the TLEC (2006:171, 2009:247). According to Van Rooy and Terblanche (2006:173):

The use of *BE* as main verb also relates to a more fragmented presentation of information, in comparison to the alternative of more frequent use of attributive adjectives, which are more characteristic of integrated presentation of information.

The TLEC examples of predicative quantifiers show a preference for canonical vs marked syntactic ordering of information, but also result in a more fragmented presentation of quantifying information.

¹²² Nomgrp = nominal group; *BE* = any form of the verb lemma *be*; Quant = Quantifier

6.3.2 Multi-word quantifiers

Excluding misspellings there were a total of 93 complex quantifiers in the TLEC data and 77 in the LOCNESS data (see Tables 6.3 and 6.4). Fewer than 40% of the multi-word quantifiers in LOCNESS are determiners, the majority function as adverbs, with the most frequent being *much more*. In contrast, approximately 60% of the multi-word quantifiers in the TLEC fulfill a determinative function.

Table 6.3 Multi-word quantifiers in the TLEC

	Deter- miner	In of- partition	Pronoun	Subject Predica- tive	Adverbial	Mis- spelling	Total per multi- word quantifier
a lot more	0	0	1	0	0	0	1
all or some	1	0	0	0	0	0	1
any more	0	0	0	0	0	3	3
both or all	1	0	0	0	0	0	1
each and every	31	0	0	0	0	0	31
each or every	1	0	0	0	0	0	1
enough or even too much	1	0	0	0	0	0	1
few or less	1	0	0	0	0	0	1
few or no	1	0	0	0	0	0	1
little more	1	0	0	0	0	0	1
little or no	2	0	0	0	0	0	2
lots and lots	0	2	0	0	0	0	2
lots more	0	0	1	0	0	0	1
many many	2	0	0	0	0	0	2
many more	5	1	5	1	0	0	12
many/most	0	0	1	0	0	0	1
more and more	4	1	0	0	5	0	10
more or less	0	0	0	0	2	0	2
more or several	1	0	0	0	0	0	1
more(most)	0	1	0	0	0	0	1
most if not all	1	1	0	0	0	0	2
much lesser	0	0	0	0	1	0	1
much more	0	0	3	0	4	0	7
no more	1	0	0	0	5	0	6
some few	2	0	0	0	0	0	2
some more	1	0	0	0	0	0	1
some or most	0	0	1	0	0	0	1
Total per syntactic role:	57	6	12	1	17	3	96

Table 6.4 Multi-word quantifiers in LOCNESS

	Deter- miner	With of- partition	Pronoun	Adverbial	Word in fixed ex- pression	Mis- spelling	Total per multi- word quantifier
a few more	1	0	0	0	0	0	1
a little less	1	0	0	0	0	0	1
a lot less	2	0	0	0	0	0	2
any and all	1	1	0	0	0	0	2
any and every	2	0	0	0	0	0	2
any less	0	0	0	1	1	0	2
any more	0	0	1	0	1	2	4
each and every	2	0	0	0	0	0	2
less and less	0	0	0	1	0	0	1
little more	2	0	1	5	0	0	8
little or no	4	0	0	0	0	0	4
many if not most	0	1	0	0	0	0	1
many more	6	0	3	0	0	0	9
more & more	0	0	0	2	0	0	2
more and more	3	0	1	5	0	0	9
more or less	0	0	0	1	0	0	1
much less	0	0	0	1	0	0	1
much more	0	0	1	16	0	0	17
no more	3	0	1	2	0	0	6
several more	1	0	0	0	0	0	1
some more	0	0	0	1	0	0	1
some or all	0	1	0	0	0	0	1
the least little (bit)	1	0	0	0	0	0	1
Total:	29	3	8	35	2	2	77

The most frequent multi-word quantifier in the TLEC data is *each and every* as determiner, which occurs 31 times. It occurs only twice in LOCNESS. De Klerk (2003a:477) regards the 56 instances of *each and every* in her data (540 000+ words of spoken Xhosa English) as a clear indication that “this formulaic expression, probably originating in order to serve as a marker of salience and redundancy, has entrenched itself into XE”. It is not clear what is meant by the marking of “salience and redundancy” to which De Klerk attributes this expression. I take it to mean that “salience” refers to the speaker’s intention to be emphatic. However, if *each and every* is a formulaic expression (i.e. merely based on convention) in BSAfE, one cannot really attribute its use to such an intention. Whether any speaker can *intend* to provide redundant information is questionable. The use of both quantifiers in contexts where either quantifier will do might be perceived as redundant to the native speaker’s ear. In order to uncover the discourse function of *each and every* in BSAfE a closer look at the corpus data is necessary.

The use of *each and every* as multi-word quantifying determiner should be regarded in the context of the use of *each* and *every* as single-word quantifying determiners. In LOCNESS, there are 85 occurrences of *each* as determiner and 94 of *every*. In TLEC, *each* occurs as single-word quantifying determiner only 8 times, compared to the 107 times that *every* occurs. The quantitative data shows that the TLEC writers use *every* as single-word determiner in the same way that the LOCNESS writers do, but suggests that for the TLEC writers the use of *each* might trigger the addition of *and every*. However, if one examines the concordance lines of *each and every* in the TLEC, it becomes clear that they do not represent cases where the determiner *each* could have been used on its own. The instances of *each and every* mainly involve an emphatic type of universal quantification, i.e. as synonym for *all*, in the context of a proposition that is universally valid. According to Payne and Huddleston, *each* and *every* are universal quantifiers “which explicitly indicate that the predication property applies distributively” (2002:362). Using both, instead of either one emphasizes this distributive meaning.

(520) And each and every team has a sponsor no matter how big or small it is

(521) <ICLE-TS-NOUN-0098.1> ('all teams have sponsors' – YVB)

(522) each and every player should work very hard for his country and make his country one of the best country in the world. <ICLE-TS-NOUN-0123.1> ('all players should ...' – YVB).

(523) it is the way they earn money just like each and every worker. <ICLE-TS-NOUN-0008.1> ('... like all workers' – YVB)

(524) Crime does not pay. It is a serious offence that can only destroy your future. For instance in each and every country there is laws and regulation to follow in order to keep our country into a stable point or into an order. <ICLE-TS-KIMC-0341.1> ('all countries have laws' – YVB)

In the examples above, *all* can substitute *each and every*, but it can be argued that the writers chose *each and every* instead of *all* to highlight the role of individual persons or entities. Other instances of *each and every* involve the progression of a phenomenon through time as the following examples illustrate:

(525) Money is very important to ensure that the player stays productive and is eager to improve on his performance each an every match. <ICLE-TS-NOUN-0103.1> (i.e. 'performance improvement, match after match' – YVB).

(526) All the effort have been made. An AIDS awareness campaign after and AIDS awareness campaign, pamphlets were distributed, condoms given to people free of charge. But still, because of the ignorance of our people, we get infected each and every day. <ICLE-TS-NOUN-0129.1>

In the examples above the same propositional information applies to multiple situations which contribute to the progression of a larger process. The use of *each and every*

helps to individualize these situations. A qualitative analysis of the corpus indicates that the use of *each and every* is semantically motivated.

The often-mentioned ***some few*** pattern occurs only twice in the TLEC.

(527) After some few days he says he is still with the squad. <ICLE-TS-NOUN-0329.1>

(528) In some few years ago, Sun city was placed in the world maps, as it hosted music festival <ICLE-TS-NOUN-0031.1>

Platt, Weber and Ho (1985:60) refer to the incompatibility of *a*, which primarily means 'one', with *few*, which is inherently plural, as a possible explanation of the use of *some few* in New Englishes. However, 13 out of the 44 occurrences of *few* in the TLEC (29.5%) are preceded by *a* and only two by *some*. Proportionally *a few* is still used less in the TLEC than in LOCNESS where 23 of the 49 occurrences of *few* (46.9%) are preceded by *a*. Yet, two examples in a 200 000+ corpus are hardly indicative of an emerging feature. In Section 6.4 below, the use of quantifying expressions like *some few* will be related to the extended meaning of *some* in BSAfE. For now, judgment is reserved on whether the expressions should be regarded as an error or a feature or BSAfE.

Table 6.3 shows that 16 of the 27 types of multi-word quantifiers in the TLE (i.e. 59 tokens) arise from coordination and apposition rather than from submodification, e.g. *all or some*, *some or most*, *many/most*. In LOCNESS 10 of the 23 types of multi-word quantifiers involve coordination (25 tokens), while the majority involves submodification, as in *much more*.

6.4 Concordance analyses of key quantifiers: *some* and *most*

In a comparison of the wordlists of the TLEC and LOCNESS in WordSmith 4, *some* and *most* rank 31st and 34th respectively as positive keywords in the TLEC, with keyness

values of 263.13 and 250.65 respectively. Table 6.5 below shows the raw and normalized frequencies of *some* and *most*.

Table 6.5 The frequencies of *some* and *most* in the TLEC and LOCNESS

	TLE raw total	TLE per 100 000 words	LOCNESS raw total	LOCNESS per 100 000 words
<i>some</i>	854	426	324	161
<i>most</i>	733	366	253	125

In order to explain the prevalence of quantifying partitive-*of* constructions in the TLEC, as well as the keyness of *some* and *most*, the concordances of *some* and *most* will be analysed in more detail in the subsections below.

6.4.1 The quantifier *some*

The distribution of *some* across syntactic roles in the TLEC and LOCNESS is shown in Table 6.6 below. In both corpora *some* is used most frequently as determiner, then as logical head of a partitive *of*-construction and then as pronoun.

Table 6.6 A comparison of *some* in the TLEC and LOCNESS

Syntactic role	Raw frequency in TLEC	Raw frequency in LOCNESS
Determiner	515	253
Head in partitive <i>of</i> -construction	186	38
Pronoun	136	29
Part of multi-word quantifier	5	2
Misspelling	12	2
Total	854	324

The keyness of *some* can be attributed in part to its choice as quantifying determiner over other potential quantifiers. It is used in 19.6% of all the determinative uses of single-word determiners in the TLEC data. In the LOCNESS data *some* is chosen as determiner in only 11.7% of the occurrences of single-word quantifying determiners.

Similarly, 25.1% of quantifiers in partitive *of*-constructions in the TLEC contain *some* as head compared to 13.8% in LOCNESS. The raw frequencies of determiners in the corpora (see Table 6.1) show that the load is spread more evenly among the quantifying determiners in LOCNESS, while in the TLEC, *some* carries quite a heavy load, especially in relation to the other existential quantifier *any*. The word *some* (in all of its functions) occurs in 337 of the 519 essays of the TLEC. In 118 of these essays it occurs only once. It occurs twice in 88 essays, three times in 65 essays and four or more times in the remaining 66 essays. In the subsections below the extended non-proportional use of *some* as determiner and the proportional use of *some* in *of*-partitions will be discussed.

A qualitative analysis of the concordance of *some* in the TLEC suggests that the non-proportional meaning of *some* as determiner is extended to contexts in which it would not be used in Standard English. As illustration, the essay containing the most instances of *some* (15) has been quoted in full below:

(529) I can believe with people who say that in our modern world dominated by science and technology and industrialisation there is no longer a place for dreaming and imagination, because this is what is happening in our daily lives. Now these days it seems as everything is now being under the technology ideas. Each and every day some scientists are trying to find some way in which they can use by means of technology. They are doing some researches, and even at some schools some students are encouraged to take some science subjects, starting from the primary level.

I think that in the late centuries there were no auto-bank machines, traffic light, Computers, communication device (eg telephone lines, mobile phones radio and television), but life was good to those people who live by that time. There was a way in which they can communicate and receive information of what is going on. Now these days we have some telephone lines, mobile phones, e-mails, etc. is which we can stay in touch with anyone around the world. And all of these has been developed by a human being, whom we can call a scientist. By radio and television we can receive information of what is happening around

us and around the world. This is by means of technology because the television and the radio uses signals. These signals has been developed by some scientist.

In every day of our lives we use technology, even when we are not aware of this. When we go to the automatic bank machines (ATM) to deposit or to withdraw money, we have used our knowledge of technology to operate that machine (ATM).

Some of the jobs that are advertised in some newspapers by some employees will need only you if you are computer literate. Some big industries in today's life uses some technology to operate some of the machines. And this means that they will not need more employers to work them because some of the work need to be done will be for done machines. It is therefore I agree with those people who say our modern world, is dominated by science and technology and industrialisation and there is no longer place for dreaming and imagination. I'm saying this because if you have dreamed to be in a certain position and that position is operated by means of or technology, your dreams are good for nothing. <ICLE-TS-NOUN-0434.1>

In the essay above most of the determinative uses of *some* occur in contexts where the zero determiner would be used in Standard English (e.g. \emptyset *scientists*, \emptyset *ways*, \emptyset *research*, \emptyset *students*, \emptyset *science subjects*, \emptyset *telephone lines*, \emptyset *scientists*, \emptyset *newspapers*, \emptyset *employers*, \emptyset *technology*). This raises the question: what is the function of *some* in these contexts? In all of the above-mentioned instances *some* indicates 'an actual but unidentified thing'. The things in question may be quite general in nature, but they are actual in the sense that they they exist. Non-proportional *some* is used like the non-singular counterpart of *a/an* instead of the zero determiner to explicitly mark the nominal group as indefinite. The following example also illustrates the use of *some* as determiner indicating 'a particular but unidentified thing'. The noun which is determined by *some* is not marked for plural, but from the plural noun *clinics* (*clinics*) and the rest of the text, it is surmised that a plural meaning is intended for *tent*.

(530) They have built some tent to use as clinics. <ICLE-TS-POT-0204.1>

Like the indefinite determiner *a/an*, *some* can be used in both particular and non-particular contexts. Both examples below represent non-particular nominal groups. In the first example, *some* is inserted into an idiomatic expression *paid peanuts*:

(531) So you cannot play soccer each and every week and after that you've been paid some peanuts. <ICLE-TS-KIMC-0355.1>

(532) Nowadays disease known as AIDS is continuing to erode the quality of life of the people and it is very suicidal to venture out in search of some sexual partners. <ICLE-TS-NOUN-0056.1>

The 18 instances of the use of *some* in the pattern **some other N** in the TLEC, also confirms that *some* is used in BSAfE instead of the zero determiner with plural nouns. Likewise, *some* occurs with the post-determiner *certain*.

(533) Most if not all of the important people who visits South Africa are given a chance to spend some times in SUN city to enjoy the best amenities/beauty that the place provide. Many tourist who came to view wildlife in some other places outside North west province such as Kruga national park prefer to spend their nights at this hotel. <ICLE-TS-NOUN-0031.1>

(534) In some certain circumstances, men ill-treat women. <ICLE-TS-NOUN-0501.1>

The use of *some* with *few* is also explained when *some* is seen as the plural version of *a/an*. Apart from the two instances of *some few* in the TLEC, *some* also occurs with other quantifiers:

(535) they spend some millions of rands in buying the condoms <ICLE-TS-NOUN-0172.1>

(536) They go to some severals areas including towns where they becomes the prostitutes <ICLE-TS-NOUN-0172.1>

(537) From my own point of view I would say the word poverty refers to people who lack some three basic needs of life which are shelter, clothing and food <ICLE-TS-NOUN-0252.1>

(538) You only need your identity card and some small amount of cash to deposit in ones account. <ICLE-TS-NOUN-0154.1>

(539) there will be never be any soccer improvement and for this the goverment would be losing some sums of money to not so committed soccer players. <ICLE-TS-NOUN-0386.1>

In all of the examples above the singular form of the nominal group would require the indefinite article *a* (as opposed to the definite determiner), viz. *a million rand, an area, a basic need, a small amount of cash, a sum of money*. These examples clearly show that terms like “overgeneralisation in the use of quantifiers” or “uniquely characteristic usage of quantifiers” (De Klerk, 2003a:470) do not accurately explain quantifier combinations like *some few* and *some three*. In the examples above *some* is used not as an additional quantifier but as a marker of indefiniteness. Its function is intrinsically (not explicitly) quantifying only in as much as the function of *a/an* is quantifying. Just as *a/an* primarily marks indefinite and secondarily implies ‘one’, *some* primarily marks indefinite and secondarily implies ‘more than one’. This implies that *some few* is not a conventionalized (learnt) expression, but an instantiation of a schema:

some quant N

or

some quant of N¹²³

The concordance data shows that indefiniteness is marked with *some* regardless of whether the nominal group in question has a particular or a non-particular interpretation, which suggests that indefinite/definite distinction in BSAfE does not play second fiddle to the particular/non-particular distinction. It is only in the choice of the determiner where

¹²³ indef = indefinite determiner; quant = quantifier; N = noun and its internal modifiers

BSAfE differs from Standard English: zero determiner vs overt determiner in the form of *some*.

The following example shows that *some* is also used as indefinite marker with uncountable nouns (*technology*).

(540) Some big industries in today's life uses some technology to operate some of the machines. <ICLE-TS-NOUN-0434.1>

In the example above the use of *some* with *technology* does not constitute a reanalysis of *technology* as countable, since the primary function of *some* is not to quantify *technology*, but to mark it as indefinite. The explicit indefinite marking of non-singular (plural and uncountable) nouns by *some* is systematic and semantically motivated enough to be regarded as a feature of BSAfE. It is contended that the use of *some* as indefinite marker in non-singular contexts (i.e. with plural and uncountable nouns) in BSAfE probably arose from the extension of the meaning of non-proportional *some* and that this extension diluted the quantifying meaning of *some* in the same way that the quantifying meaning of *a/an* was diluted in the grammaticalization of *one*. The use of *some* as indefinite marker can be ascribed to undeletion, which involves restoring “an element that is often assumed to be deleted or to have an empty node in generative analyses of English” (Mesthrie, 2006:125). Halliday and Matthiessen (2004:316) point out that in cases where there is no Deictic element in a nominal group, there is still a value in the Deictic system which is realized by a non-overt form in the structure. In other words, the zero determiner is still a realization of the determiner function, but no overt form is required. This is tantamount to an empty node in a generative analysis. Indefinite marker *some* in BSAfE is an overt realization of the determiner in contexts where no surface element is expected in Standard English. This means that referents are not only tracked more explicitly in BSAfE (by for instance demonstrative pronouns), they are also introduced more explicitly by using *some* instead of \emptyset .

6.4.2 The quantifier *most*

Like the quantifier *some*, the quantifier *most* has proportional as well as non-proportional uses (see Section 6.1). The non-proportional use of *most* encompasses cases where *most* is used as superlative of *many* and *much* (Payne & Huddleston, 2002:365). As determiner, *most* can realize both proportional and non-proportional meanings as the following examples illustrate:

(541) High school teachers deserve the most recognition because students in high school seem to be the hardest to deal with in a classroom. <ICLE-US-SCU-0007.2> [superlative, non-proportional]

(542) Most South African soccer players go to other countries, especially European countries. <ICLE-TS-NOUN-0505.1> [proportional]

Table 6.7 below provides a breakdown of the syntactic functions of *most* in the two corpora. The proportional uses of *most* are typically associated with its use in *of*-partitions and as pronoun, while its superlative uses are typically associated with its use as degree adverb.

Table 6.7 A comparison of the *most* in the TLEC and LOCNESS

Syntactic role	Raw frequency in TLEC	Raw frequency in LOCNESS
Determiner	246	90
Head in partitive <i>of</i> -construction	348	39
Pronoun	1	3
Adverbial	88	118
Part of idiomatic expression	0	2
Part of multi-word quantifier	5	1
Non-standard use of <i>most</i>	43	---
Misspelling	2	0
Total	733	253

In both corpora the vast majority of instances where *most* is used as determiner represent the proportional use of *most*. Only two of the 246 instances of *most* in the TLEC have a superlative rather than a proportional (i.e. 'not all') interpretation. These are quoted below:

(543) Each and every University Should be recognised not because it is a university with most Blacks or Whites, but because We are all equal and let our Universities be equal too. <ICLE-TS-NOUN-0435.1>

(544) If such people's bodies became weak, then it would be much easier for HIV/AIDS to affect most people at a time. <ICLE-TS-NOUN-0034.1>

Six of the 88 determinative uses of *most* in LOCNESS are superlative. They are represented by the following example:

(545) Like any other group of ideas or beliefs, the only way to reach the most people possible is through teaching others and then those people teaching other people and so on. <ICLE-US-MRQ-0005.1>

In explaining the frequency of *most* in the TLEC, the possibility that *most*, like *some*, is used to express additional meanings, which it does not express in Standard English, needs to be considered. The excerpt below comes from the TLEC essay containing the most instances of the word *most* in the corpus. The first three paragraphs are quoted in full. Apart from the seven instances of *most* underlined in this excerpt, there are also three instances of *most of the time* in the last two paragraphs of the essay.

(546) Poverty is the most problem in Africa. And it is very hard to control it, since there is and increase of unemployment. Many people are unemployed and most of them are unskilled, and dont want to make anything to at least make a living. For axample by selling small things like fruits, sweats and fat cakes, on the streets.
Most of our working people in Africa, earn little income and most of them are unskilled. They can not even save for future purpose, they

also do not know how to budget. Most of them just persecute children by spending all the income in alcohol and drugs.

Most women spend most of their times on bars and taverns, without money of buying a beer, and they will just wait for men who came to drink, and ask to drink with them after drinking those just take them and go to sleep with them, and they do this to many men. And at last they end up suffering with Aids. <ICLE-TS-NOUN-0317.1>

In the first underlined nominal group in the excerpt above, *most* is used where an adjective like *biggest* or *main* can be used. Here *most* is used in a superlative sense rather than in a proportional sense. In the nominal group *Most women*, it is likely that *most* is not used in the usual proportional sense of the word, i.e. denoting the majority of a designated set of entities, unless the essay writer truly believes that the majority of women in Africa spend most of their time in bars, which is doubtful. The proportion of the set indicated by *most* might not be as large as in Standard English. The following (double-underlined) examples also illustrate uses of *most* which is not equivalent to proportional *most* in Standard English:

(547) Since most countries in Africa are undeveloped, poverty in these countries caused the citizens to find different ways of living including the abusing of sex which resulted in Aids epidemic. Most female citizens ended up being prostitutes because of having no options. <ICLE-TS-NOUN-0051.1>

(548) Most people who had jobs in industries have lost their jobs due to the fact that those industries have implemented machines which can do a man's job. Anyway, by most people I was talking about the mine workers. <ICLE-TS-NOUN-0485.1>

(549) Most people have been infected with this disease and is affecting the economy of Africa because capable, educated and youth are dying before time. <ICLE-TS-NOUN-0182.1>

If the double-underlined nominal groups in the examples above are interpreted as realizing the conventional proportional meaning (i.e. ‘vast majority’), they result in factually unlikely propositions. It is contended that the proportional meaning of *most* is sometimes weakened from meaning ‘majority’ to simply meaning ‘many’ and that superlative *most* is sometimes used instead of adjectives denoting size.

There are 350 occurrences of *most* followed by an *of*-partition in the TLEC, compared to only 39 such cases in LOCNESS. The following patterns with *most* in partitive *of*-constructions can be distinguished in the corpus data:

Table 6.8 Patterns of *most of* in the TLEC

Lexicogrammatical pattern and example	Raw frequency in concordance	
	TLEC	LOCNESS
<i>most of the N</i> ¹²⁴ <u>Most of the families in South Africa</u> are destitute <ICLE-TS-NOUN-0283.1>	178	19
<i>most of pronoun(personal)</i> ¹²⁵ <u>Most of them</u> practice polygamy. <ICLE-TS-NOUN-0150.1>	65	8
<i>most of Ø N</i> <u>Most of tribes</u> believed that the use of condom is the main cause of AIDS because they say during older days they were not there. <ICLE-TS-NOUN-0119.1>	54	1
<i>most of determiner(possessive) N</i> <u>Most of their teaching</u> are theoretical. <ICLE-TS-NOUN-0468.1>	34	6
<i>most of determiner(demonstrative) N</i> <u>Most of these countries</u> are over populated <ICLE-TS-NOUN-0045.1>	18	4
<i>most of wh-clause</i>	0	1
<i>most of are</i>	1	0
TOTAL	350	39

The pattern-based comparison above only shows that *most of* occurs in more or less the same patterns in the TLEC and LOCNESS and that the most frequent pattern in both

¹²⁴ N refers to the lexical noun and any accompanying internal modifiers.

¹²⁵ In notations of patterns, subcategories are indicated in brackets directly after the category.

corpora are **most of the N**. In general, this does not help to explain the high frequency of *most of*-patterns in the TLEC, although there is some indication that the pattern **most of Ø N** warrants further investigation in an attempt to arrive at an explanation. The single instance of this pattern in LOCNESS is probably due to a writing error arising from an incomplete self-correction:

(550) Both sexes are being depersonalized into objects of most of sex.
<ICLE-US-SCU-0001.3>

Mesthrie (2006:139-140) postulates that the use of *most of* in BSAfE in contexts where *most* is expected in Standard English arises from the insertion of an extra morpheme (in this case *of*). The most compelling support for this postulation comes from the 54 instances of the pattern **most of Ø N** in the TLEC, since no argument can be made that the construction arises from the need to mark the Thing being quantified as definite.

(551) In the country that I came from which is Botswana most of people drink too much. <ICLE-TS-NOUN-0119.1>

(552) Most of club owners complain about the standard of soccer in our country. <ICLE-TS-NOUN-0014.1>

Insertion of a definite determiner (the definite article, possessive or demonstrative determiners) together with *of* imbues the Thing with definiteness. A closer look at the 178 TLEC concordance lines with the pattern **most of the N** reveals that in more than half of the cases, the use of the definite article is not motivated by referent identification. In both of the examples below, *of the* can be omitted from the underlined nominal groups. In assessing the data, however, the first example was classified as one in which the insertion of the morphemes *of the* cannot be ascribed to referent identification, whereas in the second example, the addition of the relative clause was taken as providing information which may be interpreted as referent identifying and therefore as possible motivation for the choice of *the*.

(553) Most of the South Africans love soccer and they are using it in order to gain a better living. <ICLE-TS-NOUN-0504.1>

(554) Most of the teams that play in South Africa are not sponsored and they have financial problems. <ICLE-TS-NOUN-0126.1>

Modifiers do not necessarily provide identifying information. In the example below, *rich* specifies a subclass of *people* and is denotational (i.e. type specifying) rather than referent identifying.

(555) Most of the rich people know very well that poverty stricken people are the ones very easy to target. <ICLE-TS-NOUN-0283.1>

In many cases, there is no information in the text that may be interpreted as referent identifying. Cases like the example below illustrate that *of the* is sometimes inserted purely for reasons of convention.

(556) Most of the teenagers and Adults are suffering from this disease. <ICLE-TS-NOUN-0116.1>

The propositional content of the clause in the example above suggests that, as is the case with *most* as determiner, *most* in *of*-partitions do not necessarily mean 'the majority'.

Amongst the instances of the pattern ***most of the N*** that are similar to Standard English uses, there are 20 instances of the idiomatic expression *most of the time*.

(557) Most of the time when international people visited South Africa they were taken here in North West province to see be beauty of this place <ICLE-TS-NOUN-0057.1>

There is enough evidence in the corpus data to confirm Mesthrie's (2006:139-140) proposal that the insertion of the morpheme *of*, or the morphemes *of the*, after quantifiers like *most* is a feature of BSAfE. Analysis of the 178 concordances lines of the pattern ***most of the N*** also indicates that the insertion of *of the* is mostly not motivated by a necessity to mark the quantified Thing as identifiable. The insertion of

morphemes in the quantifying patterns *most of Ø N* and *most of the N* are therefore not functionally motivated but conventionalized. In contrast, most of the *of*-partitions in the pattern *most of pronoun(personal)* are functionally motivated.

The majority of the 65 cases of *most of* followed by a personal pronoun in the TLEC consist of the pattern *most of them* (52). In the majority of cases, *most of them* is anaphorically related to a class of people delineated earlier in the discourse as illustrated in the example below.

(558) People who live in informal settlements are people who are very much under privileged. They do not have TV sets, and radios, most of them are not really educated <ICLE-TS-NOUN-0313.1>

There are also six instances where *most of them* are involved in topicalizing syntactic constructions (i.e. left-dislocation). In three of these *most of them* functions much like the copy pronoun *they* discussed in Chapter 4, except that a quantified subset of the topicalized class of people is denoted.

(559) The people who study at university most of them are now work than those who study at colleges and technikons. <ICLE-TS-NOUN-0444.1>

(560) Many people who are not employed most of them they told themselves that there's nothing that they can do, <ICLE-TS-NOUN-0301.1>

(561) The degrees of our university most of them are of little value <ICLE-TS-NOUN-0500.1>

There are also three cases where *most of them* is the topicalized nominal group which is anaphorically related to an earlier mention of a subset of *people*. This smaller subset is left-dislocated and emphasized by the copy pronoun *they*. The italicized nominal groups in the examples below serve as antecedents of the underlined nominal groups.

(562) There was big war there at Congo that left *most of the people* in a trouble.

Most of them at that moment they dont have house and lack of food which we called poverty it gave power. <ICLE-TS-NOUN-0086.1>

(563) Most of them, they are selling themselves, <ICLE-TS-KIMC-0352.1>

(564) *The players* are not given package most of them they dont have Qualifications <ICLE-TS-NOUN-0225.1>

Literature on BSAfE refers to the use of the expression *the most thing that I [verb]* instead of *the thing that I most [verb]* (De Klerk & Gough, 2002; Mesthrie, 2004). In these expressions the superlative, rather than the proportional meaning, of *most* is activated. In the TLEC there are only two examples of such constructions involving the noun *thing* and two involving the noun *things*, but none of these involve the first person singular pronoun.

(565) The most thing which spread AIDS in Africa is through sex. <ICLE-TS-NOUN-0001.1>

(566) One of the most things they do in order to get money and buy food is through prostitution <ICLE-TS-NOUN-0237.1>

It should also be noted that *thing(s)* is not the only noun involved in such constructions and that *that*-clauses are not the only type of modifier used in such constructions. A closer look at the concordance data indicates that the phenomenon is actually more widespread and complex than has been suggested in the literature.

(567) South Africans are the most people who are doing crime in the whole world why? <ICLE-TS-KIMC-0353.1>

(568) Platinum, asbestos and copper are the most minerals produced in North West. <ICLE-TS-NOUN-0080.1>

(569) I would advise a friend to choose a bank between ABSA and F.M.B Bank. Because they are the most bank that does'nt any problem on

making any problem when opening a saving account. <ICLE-TS-NOUN-0229.1>

The nominal groups underlined in the examples above are Subject Predicatives that fulfill an identifying rather than an attributive function. In SFG terms, they are Identifiers in relational clauses. The predicative nominal groups in the examples above all have the following pattern:

the most NOUN modifier:clause¹²⁶

In the TLEC concordance of *most* there are 7 instances of this pattern occurring as Subject Predicate in identifying copula clauses and one instance where *most* is used pronominally, i.e. where the pattern is ***the most modifier:clause***.

(570) Since the girls aged 12 - 20 are the most who are victimised by poverty that led them to this industry, the chances of kids born with Hiv/Aids are high. <ICLE-TS-NOUN-0072.1>

There are also 5 instances where nominal groups of the pattern ***the most noun modifier:clause*** occur as the Subject (Identified) of a copula clause. Four of these involve the noun *thing(s)*.

(571) The other most thing the tourists like about the North-West is, the freedom they have <ICLE-TS-NOUN-0285.1>

(572) The most problem that soccer facing in this country is the players who go away to other countries to play there. <ICLE-TS-NOUN-0505.1>

Although the two examples above have the same surface pattern, Standard English rewordings of their meanings have different patterns:

The other thing tourists like most about North-West is ...

¹²⁶ "NOUN" refers only to the noun and not its internal modifiers; "modifier:clause" indicates that the modifier following the noun is realized by a clause.

The biggest problem that soccer is facing in this country is ...

In the latter example *most* is used as a superlative adjective of size. In the first example *most* occurs as quantifier in the nominal group in BSAfE where it would occur as clause-level Adverbial in Standard English. This is also true for the Subject Predicative nominal groups mentioned earlier. In the concordance there are an additional 14 instances where *most* occurs in predicative nominal groups. Ten of these have postmodifiers in the form of phrases. Of these ten, half represent the pattern **most noun modifier:phrase** and half represent the pattern **most modifier noun modifier:phrase**. The five examples without prenominal modifiers will be considered first:

(573) Poverty is the most problem in Africa. <ICLE-TS-NOUN-0317.1>

(574) To my opinion I agree that poverty is one most cause of the HIV/Aids epidemic on Africa. <ICLE-TS-NOUN-0320.1>

(575) It is true that poverty is being considered the most cause of it (HIV/AIDS) <ICLE-TS-NOUN-0237.1>

(576) Today our province is the most and number province in harvesting and farming <ICLE-TS-NOUN-0080.1>

(577) Soccer is one of the most and successful sport in this country. <ICLE-TS-NOUN-0505.1>

In the examples above *most* is used to signify 'primacy'. In the first three examples it can be replaced by *main* or *biggest*. In the fourth example, *most* may be substituted by *best*, but it is not certain that this is the meaning intended by the essay writer. It could be that the writer has the amount of agricultural activities in the province (compared to other provinces) in mind. In the last example *the most* signifies 'the largest amount', probably the amount of people playing or supporting soccer compared to other sports.

The examples below have the pattern **most modifier noun modifier:phrase**. In all five examples, the prenominal modifier takes the form of a participial (deverbal) adjective.

(578) Obviously this causes poverty which is the most contributing factor towards the HIV/AIDS epidemic in Africa. <ICLE-TS-KIMC-0380.1>

(579) This is one of the very most contributing factor to the infection of HIV/Aids. <ICLE-TS-WITS-0007.1>

(580) According to the reseach they found youth are most affected one in the world

(581) <ICLE-TS-NOUN-0172.1>

(582) HIV/AIDS is being the most life striking epidemic in Africa. <ICLE-TS-NOUN-0237.1>

(583) Soccer is the most growing sport here in South Africa. <ICLE-TS-KIMC-0369.1>

Standard English reformulations of the first two examples above will involve placement of the information contained in the prenominal modifier after the noun, which in turn would entail *most* functioning as Adverbial in a relative clause rather than as quantifier in a nominal group, viz: *the factor which most contributes to the spread of the HIV/Aids epidemic.* In such reformulations, *most* can be replaced by adverbs indicating primacy, such as *mainly* or *chiefly*. Although similar rewordings are possible for the other three examples, smaller changes on the level of lexical choice can also effect Standard English versions, e.g. *the youth are most affected, the most devastating epidemic, the fastest growing sport.* In the last example *most* is used instead of a more specific superlative adjective such as *fastest*.

(584) When looking in all sporting code in South Africa, soccer is the most supported one. <ICLE-TS-NOUN-0097.1>

(585) we have rural areas producing some products and it helps our provinc to be the top piority and the most touring province. <ICLE-TS-NOUN-0080.1>

In conclusion, the frequent use of *most* in *of*-partitions can be ascribed to the insertion of the morpheme *of*. This explains the high frequency of *most* in *of*-partitions in the TLEC, but not the high frequency with which *most* is chosen as determiner. Based on a qualitative examination of the concordance data it is postulated that in addition to the Standard English proportional meaning of *most* as 'majority', *most* realizes plural meanings in BSAfE (i.e. 'many'). With regard to the superlative use of *most*, the concordance data shows that *most* is sometimes used instead of superlative adjectives denoting size and sometimes used prenominal instead of adverbially.

6.5 Number marking

Both the omission of the plural suffix in plural countable nouns and the addition of the regular plural suffix to uncountable nouns and irregular plural forms are attested in literature on BSAfE (De Klerk & Gough, 2002:362; Van der Walt & Van Rooy, 2002:124; De Klerk, 2003a:472 & 2003b:230; Mesthrie, 2004:971; Minow, 2010:61; Siebers, 2012:136). The use of the plural suffix with countable nouns were checked in the TLEC concordances of *a lot*, *many* and *most*.

Of the 43 instances of the pattern ***a lot of N(countable)***, there are five instances where the countable noun is not marked with a plural suffix. The two instances of the non-marking of *African* as plural occur in the same essay.

(586) Due to the fact that a lot of African are poor they can not afford to send their childrens to school. <ICLE-TS-NOUN-0109.1>

(587) A lot of African are living beyond poverty datum line <ICLE-TS-NOUN-0109.1>

(588) We come to varsity just to come to try to grasp and memorise a lot of fact offered by lectures. <ICLE-TS-NOUN-0469.1>

(589) If there can be a lot of job this will decrease rate of unemployment and obviously poverty decrease. <ICLE-TS-NOUN-0236.1>

(590) This create a lot of problem of not giving successful results after studying. <ICLE-TS-NOUN-0475.1>

The non-marking of plural countable nouns was also examined in the concordance of the quantifier *many* (where *many* functions as a determiner). Of the 485 instances of determinative *many* in the TLEC, 451 determine a plural countable noun. There are 29 instances where *many* precedes the singular form of a countable noun (and a further 5 instances where the noun following *many* is a non-countable noun).

(591) It is true that after finishing matric, many student will like to further their studies at different universities of their choice. <ICLE-TS-NOUN-0490.1>

(592) Many tourist visits North West from all over the world because it is the place to be. <ICLE-TS-NOUN-0096.1>

(593) have one sexual partner and try to find as many information about the danger or the deadly results of HIV/Aids. <ICLE-TS-WITS-0007.1>

The non-marking of plural countable nouns occur in less than 6% of the occurrences of the pattern **many N**, which suggests that even in nominal groups where plurality is already marked by a quantifying determiner such as *many* or a numeral, the plural suffix is regarded as necessary and not as redundant in the vast majority of cases. Only a full-scale analysis of all countable nouns preceded by quantifiers and numerals can conclusively indicate how closely BSAfE follows the norm of Standard English. However, it should be born in mind that *many* is the most frequent quantifying determiner after *some*, and can be taken as indicative of trends involving quantifying determiners and countable nouns. The fact that the plural ending for countable nouns is supplied in 94% of the cases where it is expected in the concordance of *many* (as determiner), suggests that non-suppliance of the plural ending is probably an acquisition error rather than an emerging feature of BSAfE (cf Van Rooy, 2011).

Of the 248 uses of *most* as determiner in the TLEC, the majority precede plural countable nouns (232). There were no uncountable nouns determined by *most*. There are sixteen instances of *most* preceding countable nouns not marked for plural, for example:

(594) Most African are unemployed and there is nothing that helped them to survive from living. <ICLE-TS-POT-0200.1>

In nine of the sixteen instances where the plural suffix is omitted, the noun is *university* or *degree*. These instances come from six essays with the same topic, each of which is represented in the set of examples below.

(595) There is a serious issue concerning most university degrees which are mostly theoretical and do not prepare student for real world. In most university around South-Africa and other countries, they uses theory method for teaching students instead of practical. <ICLE-TS-NOUN-0475.1>

(596) Most university degree are theoretical and do not prepare student for the real world. <ICLE-TS-NOUN-0469.1>

(597) I agree with this topic because it's true that most university degree are theoretical. In most University they offer theory too much than practical <ICLE-TS-NOUN-0464.1>

(598) According to the topic and what I've mentioned above, it is true that most university do not prepare students for the real world. <ICLE-TS-NOUN-0490.1>

(599) Of course one can argue that most university degree is only good at preparing student on graduates on the theory part. <ICLE-TS-NOUN-0476.1>

(600) It is a fact that most degree are more theoretical than practical in universities. <ICLE-TS-NOUN-0455.1>

The phrase *most university degrees* occur 14 times, which suggests that the wording of the essay topic influenced the word choices of the essay writers. This may explain the non-marking of the plural in the noun *university*. However, the phrase *most universities* occur nine times. With regard to the noun *degree*, the wording of the essay topic can be expected to trigger the use of the plural suffix. The remaining three cases of non-marking of plurality involve nouns referring to abstract concepts and institutions: *most subject, most society, most technikon*.¹²⁷ In the concordance of *most*, the plural suffix is supplied in 93% of the contexts in which it is expected after the quantifying determiner *most*. At least ten instances of plural countable nouns without plural suffixes were noticed in the TLEC concordance of *some*.

(601) Some other technikon campus, even those campus that are not doing communication, have campus radio. <ICLE-TS-NOUN-0461.1>

As can be seen from the examples above, omission of the plural suffix does not imply reanalysis of the noun as singular or uncountable. From the context it is clear that the plural meaning is still intended. However, the addition of the plural suffix to uncountable nouns does constitute a reanalysis of the noun in question as countable.

There are five instances in the 514 concordance lines of *some* as determiner in the TLEC where plural suffixes occur with nouns usually construed as uncountable. These include *infrastructures, job creations, researches* (x2) and *managements*.¹²⁸ Of the 43 instances of the pattern **a lot of N(countable)**, there is one instance where a plural suffix is added to a noun normally construed as uncountable (*prostitutions*).

(602) There is a lot of prostitutions in countries in Africa like Botswana, South Africa, Malawi, Namibia, Zimbabwe and others. <ICLE-TS-NOUN-0282.1>

¹²⁷ <ICLE-TS-NOUN-0470.1>, <ICLE-TS-NOUN-0515.1>, <ICLE-TS-NOUN-0471.1>.

¹²⁸ <ICLE-TS-NOUN-0047.1>, <ICLE-TS-NOUN-0099.1>, <ICLE-TS-NOUN-0099.1> & <ICLE-TS-NOUN-0434.1>, <ICLE-TS-NOUN-0385.1>

The wordlist of the TLEC provide further evidence of the addition of the plural suffix to uncountable nouns: *equipments* (10), *monies* (5), *punishments* (5), *researches* (5), *entertainments* (4), *advices* (3), *prostitutions* (3), *counsellings* (2), *employments* (2), *managements* (2), *protections* (2), *contraceptions* (1), *healths* (1), and *ignorances* (1). It is interesting to note that there are more instances of *equipments* than its singular form *equipment* (10 vs 4). The majority of these nouns are abstract nouns which are derived from verbs, e.g. *researches*, *protections*, etc. Equipment and money consist of countable units, e.g. pieces of equipment, rands and cents. While the corpus data demonstrates that re-analysis of abstract uncountable nouns (especially nominalizations) as countable through the addition of plural suffixes is part of the feature pool of BSAfE, this cannot be taken as evidence of the “loss” of the distinction between countable and uncountable nouns in BSAfE. A wider look at quantification in BSAfE suggests that the countable/uncountable distinction is very much intact.

6.6 Conclusion

In Section 6.2 it was pointed out that previous research have listed and exemplified certain characteristic uses of quantifiers in BSAfE without relating these to each other or to the realization of the function of quantification in general. The various concordance analyses in Sections 6.3, 6.4 and 6.5 of this chapter represent an attempt to explain some of the quantification phenomena in BSAfE. The claim that the semantic categories of countable and uncountable are coalesced in BSAfE was investigated and rejected in Sections 6.3 and 6.5. The relatively high frequency of the quantifiers *some* and *most* in the TLEC compared to LOCNESS was explained in terms of the extended meanings of *some* and *most* in BSAfE. Non-proportional *some* is used as an indefinite marker in non-singular contexts where Standard English would use the zero article. This relates to undeletion. It was also shown that proportional *most* sometimes expresses weaker multal meanings, i.e. ‘many’ rather than ‘the majority’ and that superlative *most* is sometimes used instead of superlative size adjectives. The frequency of quantifying *of*-partitions was explained by the insertion of morphemes, a type of anti-deletion.

Conclusion

7.1 Introduction

Studies on BSAfE up to 2006 focused on non-standard features of BSAfE and these were mostly presented as lists of “features” where “features” was a thinly disguised euphemism for “errors” (for example, Gough, 1996; De Klerk, 2003a, 2003b). In 2006 Mesthrie’s explanation of a number of characteristics of BSAfE in terms of anti-deletion appeared. This was the first attempt to relate the features of BSAfE to an underlying tendency in the grammatical system of the variety. The first in-depth corpus-based study of a single aspect of the grammar of BSAfE (the progressive aspect) also appeared in 2006 (Van Rooy). Subsequent corpus studies focusing on a demarcated area of the grammar of BSAfE have focused on the verb and the construal of Processes (Van Rooy, 2006, 2009; Partridge, 2011). The present study set out to complement these studies by focusing on the nominal group and the construal of Things.

The aim of the present study was to contribute to the current understanding of the grammar of BSAfE with reference to the nominal group. Since this study makes use of corpus data, a usage-based approach to grammatical analysis was taken. The broad theoretical framework in terms of which the data have been analyzed was drawn from functional linguistics and cognitive linguistics. The concepts of systemic networks, patterns and constructions were employed to interpret data from the TLEC and LOCNESS.

7.2 Summary of findings

7.2.1 The syntagmatic patterns of the noun *people*

The noun *people* was chosen as focus of the first research question since it has been associated with one of the most discussed features of BSAfE, namely left dislocation

(Mesthrie, 1997:132; De Klerk, 2003a:468; Minow, 2010:188) in previous literature. As a general noun (Mahlberg, 2005:1) it has great potential to be further specified by determiners, quantifiers and modifiers. The noun *people* is also the most frequent noun in both the corpora under investigation, while it is also one of the top keywords in the TLEC compared to LOCNESS. For these reasons, it was expected that concordances of the noun *people* would provide a good basis of comparison of nominal group structure as well as larger patterns featuring a *people*-group.

Comparison of the full concordances of the word *people* from both corpora allowed for the identification of a number of trends that characterize the language of the TLEC. These include emphasis of the humanness of clausal Participants; explicit referent-tracking; linear, rather than syntactically integrated, presentation of information; and evidence of anti-deletion.

Linguistic devices that make the human visible include the avoidance of nominalizations to represent communicative processes in favour of constructions with an overt Recipient; the use of ***those people who*-clause** in contexts where ***those who*-clause** could also have been used and the use of *as people* to elaborate the pronoun *we*. The use of the demonstrative determiner in anaphoric *people*-groups and instances where a *people*-group is left dislocated, or co-referential to a left-dislocated nominal group, are explicit referent-tracking strategies. Linear information presentation is evident in patterns where *people* is followed by a more specific appositive nominal group, as well as in cases where left-dislocation arises from removing a complex nominal group from the main clause. Evidence of anti-deletion is found in the insertion of the definite article in *people*-groups with generic meaning; the insertion of *of* or *of the* after quantifiers (resulting in partitive quantifying constructions); and the undeletion of the noun *people* in ***those people who*-clause**.

There is also evidence of substrate language influence in the rise of an innovative construction involving the pattern *other people* where *other* is used as synonym of *some* rather than as contrastive relative postdeterminer.

With regard to *people*-groups that were marked as definite, a proportionally higher usage of demonstrative and possessive determiners was observed in the TLEC compared to

LOCNESS. In cases where the noun *people* was specified in terms of quantity the proportionally higher usage of partitive *of*-constructions in the TLEC compared to LOCNESS was noted. These observations pointed to the need to investigate determiners and quantifiers, which was done in the subsequent chapters.

7.2.2 Determiner choice

Before the analysis of corpus concordances of the articles, demonstrative determiners and possessive determiners could commence, it was necessary to clarify the following theoretical distinctions: particular/non-particular and definite/indefinite. Though these conceptual distinctions are mentioned in grammatical descriptions of English, the pairs of distinctions are not related to each other in a clear and explicit manner. Chapter 5 contributes to the theory of determiners in general by presenting two system networks in terms of which articles may be chosen depending on which distinction is most salient. It was argued that both sets of distinctions play a role in article choice and that it is too simplistic to say that article choice in Standard English relies on the definite/indefinite distinction, while it relies on the specific/non-specific distinction in another variety of English. Two possible systems were postulated: one in which the referential intent of the speaker (particular/non-particular) has priority over the identity of the referent (definite/indefinite) (cf. Figure 5.5) and one in which identification has priority over referential intent (cf. Figure 5.8).

A comparative analysis of samples of the concordances of the articles revealed similarities and differences between the two corpora. In both corpora the definite article is primarily used to indicate identifiability. However, the corpora differ in that there are proportionally fewer specific and more non-particular uses of the definite article in the TLEC compared to LOCNESS. In both corpora the indefinite article is used for specific reference as well as in nominal groups with non-particular interpretations. However, there are proportionally many more specific uses of the indefinite article in LOCNESS compared to the TLEC. There are also many more instances of the indefinite article in quantifying expressions in the TLEC compared to LOCNESS. The articles (both definite and indefinite) are used to convey the speaker's intent to refer to particular instances (as opposed to non-particular instances) more often in LOCNESS than in the TLEC, while

the TLEC writers make more use of demonstrative determiners to explicitly mark specific reference.

In both corpora there are more proximal than distal demonstrative determiners. However, there are proportionally more distal demonstratives in the TLEC compared to LOCNESS. In both corpora the most frequently chosen possessive determiner is *their*, but while third person possessives are prevalent in LOCNESS, there are more first and second person possessives in the TLEC data. The corpora also differ in respect of the number features of the possessives. There are more plural forms in the TLEC and more singular forms in LOCNESS.

The most important finding of Chapter 5 is that a claim that article choice in BSAfE is based on an underlying system that is different from Standard English would be exaggerated. The corpus data show that articles are used to fulfill the same functions in BSAfE as in Standard English and that they are used to mark the nominal group as definite or indefinite. Article usage patterns that are unique to BSAfE are mainly due to conventionalization. These include the use of *the* in idiomatic expressions and nominal groups with generic interpretations where \emptyset would be used in Standard English, the use of *the* in ascriptive nominal groups where *a/an* would be used in Standard English, and the few examples of the indefinite article with non-singular nouns.

7.2.3 Quantification

In Chapter 6 concordances of the words typically classified as quantifying determiners or quantifying pronouns were analyzed. The analysis showed much more frequent use of quantifying *of*-partitions in the TLEC compared to LOCNESS. This is taken as support for Mesthrie's (2006:139-140) proposal that the insertion of the morpheme *of*, or the morphemes *of the*, after quantifiers is part of the general tendency towards anti-deletion in the grammar of BSAfE.

Since *some* and *most* are positive keywords in the TLEC (when compared to LOCNESS), their full concordances were analyzed in an attempt to explain their high frequency in the TLEC. The analysis showed that *some* functions as a grammatical

marker of indefiniteness in non-singular nominal groups in BSAfE and that its quantifying function is much diluted and secondary to its indefiniteness-marking function in these contexts. The fact that *some* is used as overt indefinite determiner, rather than as quantifier, in BSAfE in contexts where no determiner would be used in Standard English explains the positive keyness of *some* in the TLEC, but also provides an explanation for the use of *some* with quantifiers as in *some few*. The high frequency of *most* in the TLEC is attributable to the fact that proportional *most* sometimes expresses weaker plural meanings, i.e. 'many' rather than 'the majority' and the fact that superlative *most* is sometimes used instead of superlative size adjectives.

In Chapter 6 it was also shown that abstract uncountable nouns (especially nominalizations) are re-analyzed as countable through the addition of plural suffixes in BSAfE. However, this cannot be taken as support for the claim that the countable/uncountable distinction in BSAfE is diminished or lost (De Klerk, 2003a:472). The high suppliance rate (94%) of plural nouns after *many* suggests otherwise. The rare instances of the use of countable nouns with *much*, the use of uncountable nouns with *many* and the omission of the plural suffix on nouns following quantifiers associated with plurality in the corpus data probably only indicate learner errors. They are not frequent enough to be taken as features of BSAfE.

7.3 Conclusion

The analysis indicates that the TLE data contain some errors that may be attributed to second language learning. These errors should not be taken to suggest a different grammatical system for BSAfE. The analysis indicates that BSAfE is less different from Standard English (as represented by LOCNESS) than suggested in previous literature (e.g. Gough, 1996; De Klerk, 1999, 2003a, 2003b). Where differences do exist they pertain to specific constructions. These arise from speakers' gradual extraction of grammatical patterns from usage input and not from an attempt to apply the idealized rules of Standard English.

This study focused on the patterns associated with the most frequent noun (*people*), determiners and quantifiers in order to investigate the specifying functions of the nominal

group. An in-depth corpus study of quality specification in BSAfE is a potential future research area. Lexical specification (denotation) through nouns and adjectives in BSAfE corpora also warrants further investigation.

Bibliography

Biber, D., Conrad, S. & Reppen, R. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman grammar of spoken and written English*. Essex: Longman.

Bouchard, D. 2002. *Adjectives, number and interfaces: why languages vary*. Oxford: Elsevier.

Buthelezi, Q. 1995. South African Black English: lexical and syntactic characteristics. (*In* Mesthrie, R. *ed.* *Language and social history: studies in South African sociolinguistics*. Cape Town: David Philip. p. 242-248.)

Butler, C.S. 2003. *Structure and function: a guide to three major structural-functional theories*. Amsterdam: John Benjamins.

Bybee, J. 2006. From usage to grammar: the minds response to repetition. *Language*, 82(4):711-733.

Bybee, J. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Chomsky, N. 1977. Questions of form and interpretation. (*In* his *essays on form and interpretation*. Amsterdam: North-Holland. p 25-59).

Cole, D.T. 1955. *An introduction to Tswana grammar*. London: Longmans, Green & Co.

Croft, W. 1995. Autonomy and functional linguistics. *Language*, 71(3):490-532.

Croft, W. 2001. *Radical construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, W & Cruse, D.A., 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

Dagneaux, E., Denness, S., Granger, S., & Meunier, F. 1996. *Error tagging manual version 1.1*.

De Klerk, V. 1999. Black South African English: Where to from here? *World Englishes*, 18(3):311-324.

- De Klerk, V. 2003a. Towards a norm in South African Englishes: the case for Xhosa English. *World Englishes*, 22(4):463-481.
- De Klerk, V. 2003b. Xhosa English as an institutionalised variety of English: in search of evidence. *English World-Wide*, 24(2):221-243.
- De Klerk, V. 2004. The use of *actually* in spoken Xhosa English: a corpus study. *World Englishes*, 24(3):275-288.
- De Klerk, V. 2005a. Procedural meanings of “well” in a corpus of Xhosa English. *Journal of Pragmatics*. 37(8):1183-1205.
- De Klerk, V. 2005b. Expressing levels of intensity in Xhosa English. *English World-Wide*, 26(1):77-96.
- De Klerk, V. 2006. *Corpus linguistics and World Englishes*. London: Continuum.
- Deutscher, G. 2005. *The unfolding of language: the evolution of mankind’s greatest invention*. London: Arrow Books (Random House).
- De Klerk, V. & Gough, D. 2002. Black South African English. (In Mesthrie, R. ed. *Language in South Africa*. Cambridge: Cambridge University Press. pp. 356-378.)
- Egins, S. 2004. *An introduction to systemic functional linguistics*. 2nd edition. London: Continuum.
- Evison, J. 2004. What are the basics of analysing a corpus? (In O’Keeffe, A. & McCarthy, M. eds. 2010. *The Routledge handbook of corpus linguistics*. pp. 122-135.)
- Fawcett, 2000. *A theory of syntax for systemic functional linguistics*. Amsterdam: John Benjamins.
- Fawcett, R. P. 2008. *Invitation to systemic functional linguistics through the Cardiff Grammar*. 3rd edition. London: Equinox
- Finn, S.M. 1986. The abuse of English by Vista students. *English usage in South Africa*, 17(1):1-8.
- Givon, T. 1971. Historical syntax and synchronic morphology: an archaeologist's field trip. (In *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society*, April 16–18. Chicago: Chicago Linguistics Society. pp. 394-415)
- Givon, T. 1984. *Syntax: A functional-typological introduction*. Volume I. Amsterdam: John Benjamins.

Givon, T. 1993. *English Grammar: a function-based introduction*. Amsterdam: John Benjamins.

Givon, T. 1995. *Functionalism and grammar*. Amsterdam: John Benjamins.

Givon, T. 2001. *Syntax: An introduction*. Amsterdam: John Benjamins.

Givon, T. 1993. *English grammar: A function-based introduction*. Amsterdam: John Benjamins

Goldberg, A. 1992. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gough, D. 1996. *Black English in South Africa*. (*In De Klerk, V. ed. Focus on South Africa*. Amsterdam: John Benjamins. p. 53-78.)

Granger, S., Dagneaux, E., Meunier, F., Paquot, M. (eds.). 2009. *International Corpus of Learner English Version 2*. Louvain: UCL Presses Universitaires de Louvain.

Halliday, M.A.K. 2002[1970]. *Language structure and language function* (*In Webster, J. ed. On Grammar: Volume 1 in the collected works of M.A.K. Halliday*. London: Continuum. p. 173-195.)

Halliday, M.A.K. & Hasan, R. 1976. *Cohesion in English*. London: Longman.

Halliday, M.A.K. & Matthiessen, C. 1999. *Construing experience through meaning: a language-based approach to cognition*. London: Continuum.

Halliday, M.A.K. & Matthiessen, C. 2004. *An introduction to functional grammar*. 3rd ed. London: Arnold.

Huddleston, R. 1984. *Introduction to the grammar of English*. Cambridge: Cambridge University Press.

Huddleston, R. 2002. *Syntactic overview*. (*In Huddleston, R & Pullum, G.K. eds. The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. p. 43-70)

Huddleston, R. & Pullum, G.K. (eds.). 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hunston, S. & Francis, G. 1999. *Pattern Grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Ionin, T., Ko, H. & Wexler, K. 2004. Article semantics in L2 acquisition: the role of specificity. *Language Acquisition*, 12:3-69.

Kemmer, S. & Barlow, K. 2000. Introduction: a usage-based conception of language. (In Barlow, M. & Kemmer, S. eds. 2000. Usage-based models of language. Stanford, CA: CSLI Publications. p. vii-xxviii).

Kortmann, B, Burrige, K., Mesthrie, R., Sneider, E.W., Upton, C. eds. 2004. A handbook of varieties of English. Volume 2: Morphology and syntax. Berlin: Mouton de Gruyter.

Krüger, C.J.H. 2006. Introduction to the morphology of Setswana. Muenchen: LINCUM.

Langacker, R.W. 1972. Fundamentals of linguistic analysis. New York: Harcourt Brace Jovanovich.

Langacker, R.W. 1987. Nouns and verbs. *Language*, 63(1):53-94.

Langacker, R.W. 2000. A dynamic usage-based model. (In Barlow, M. & Kemmer, S. eds. 2000. Usage-based models of language. Stanford, CA: CSLI Publications. p. 1-63).

Langacker, R.W. 2004. Remarks on nominal grounding. *Functions of Language*, 11(1):77-113.

Langacker, R.W. 2008. Cognitive Grammar: a basic introduction. Oxford: Oxford University Press.

Lindquist, H. 2009. Corpus linguistics and the description of English. Edinburgh: Edinburgh University Press.

Mahlberg, M. 2005. English general nouns: a corpus theoretical approach. Amsterdam: John Benjamins.

Makalela, L. 2004. Making sense of BSAE for linguistic democracy in South Africa. *World Englishes*, 23(3):355-366.

Mesthrie, R. 1997. A sociolinguistic study of topicalization phenomena in South African Black English. (In Schneider, E.W. ed. Englishes around the world: studies in honour of Manfred Görlach. Amsterdam: John Benjamins. p. 119-141.)

Mesthrie, R. 2004. Black South African English: Morphology and syntax. (*In* Kortmann, B. *et al.* eds. A handbook of varieties of English. Volume 2: Morphology and syntax. Berlin: Mouton de Gruyter.)

Mesthrie, R. 2005. Putting back the cart before the horse: The "spelling form" fallacy in second language acquisition studies, with special reference to the treatment of unstressed vowels in Black South African English. *English World-Wide*, 26(2): 127-152.

Mesthrie, R. 2006. Anti-deletions in an L2 grammar: A study of Black South African English mesolect. *English World-Wide*. 27:111-145.

Mesthrie, R. & Bhatt, R.M. 2008. World Englishes: the study of new linguistic varieties. Cambridge: Cambridge University Press.

Minow, V. 2010. Variation in the grammar of Black South African English. Frankfurt am Main: Peter Lang.

O' Donell, M. 2007. UAM Corpus Tool Version 2.6.9. (available from <http://www.wagsoft.com/CorpusTool/>)

Partridge, M. 2010. Specificity in lexical verbs: a corpus-based lexicological study. Vanderbijlpark: NWU. (Dissertation – M.A.).

Partridge, M. 2011. A comparison of lexical specificity in the communication verbs of L1 English and TE student writing. *Southern African Linguistics and Applied Language Studies*, 29(2):135-147.

Payne, J. & Huddleston, R. 2002. Nouns and noun phrases. (*In* Huddleston, R. & Pullum, G.K. eds. The Cambridge grammar of the English language. Cambridge: Cambridge University Press p. 323-524)

Platt, J., Weber, H., & Ho, M.L. 1984. The New Englishes. London: Routledge and Kegan Paul.

Pullum, G.K. & Huddleston, R. 2002a. Adjectives and adverbs. (*In* Huddleston, R. & Pullum, G.K. eds. The Cambridge grammar of the English language. Cambridge: Cambridge University Press p.525-596)

Pullum, G.K. & Huddleston, R. 2002b. Prepositions and preposition phrases. (*In* Huddleston, R. & Pullum, G.K. eds. The Cambridge grammar of the English language. Cambridge: Cambridge University Press p.597-662)

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. A comprehensive grammar of the English language. London: Longman.

- Radford, A. 2004. *English syntax: an introduction*. Cambridge: Cambridge University Press.
- Rayson, P. 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4):519-549.
- Sand, A. 2004. Shared morpho-syntactic features in contact varieties of English: article use. *World Englishes*, 23(2):281-298.
- Scheffler, L. 1978. Common language errors. *English Usage in South Africa*, 9(1):25-27.
- Schmid, H-J. 2000. *English abstract nouns as conceptual shells: from corpus to cognition*. Berlin: Mouton de Gruyter.
- Schmied, J. 1991. *English in Africa: an introduction*. Longman: Longman.
- Schneider, E. 2007. *Postcolonial English: varieties around the world*. Cambridge: Cambridge University Press.
- Scott, M. 2010. *Wordsmith Tools: version 5.0*. Liverpool: Lexical Analysis Software.
- Siebers, L. 2012. *Morphosyntax in Black South African English*. Tübingen: Gunter Narr.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sison, C.P. & Glaz, J. 1995. Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association*, 90(429):366-369).
- Stirling, L. & Huddleston, R. 2002. Deixis and anaphora. (In Huddleston, R. & Pullum, G.K. eds. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press p. 1449-1562)
- Terblanche, L. 2009. A comparative study of nominalisation in L1 and L2 writing and speech. *Southern African Linguistics and Applied Language Studies*, 27(1):55-68.
- Thompson, G. 2004. *Introducing functional grammar*. 2nd edition. London: Arnold.
- Tognini-Bonelli, E. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Tucker, G.H. 1998. *The lexicogrammar of adjectives: a systemic functional approach to lexis*. London: Cassell.

Van der Walt, J.L. & Van Rooy, B. 2002. Towards a norm in South African Englishes. *World Englishes*, 21(1):113-128.

Van Rooy, B. 2004. Black South African English: phonology (In Schneider, E., Burrige, K., Kortmann, B., Mesthrie, R. & Upton, C. eds. *Handbook of varieties of English*. Volume 1: Phonology. p. 943-952).

Van Rooy, B. 2005. The localisation of English resources: spelling checking for South African students. (In Daelemans, W., Du Plessis, T., Snyman, C. & Teck, L. eds. *Multilingualism and electronic language management*. Pretoria: Van Schaik. p. 125-138).

Van Rooy, B. 2006. The extension of the progressive aspect in Black South African English. *World Englishes*, 25(1):37-64.

Van Rooy, B. 2008a. A multidimensional analysis of student writing in Black South African English. *English World-Wide*, 29(3):268-305.

Van Rooy, B. 2008b. An alternative interpretation of tense and aspect in Black South African English. *World Englishes*, 27(3/4):335-358.

Van Rooy, B. 2010. Presidential Address delivered at the City University of Hong Kong, Hong Kong, on December 4, 2008, at the 14th annual meeting of the International Association for World Englishes: Social and linguistic perspectives on variability in World Englishes. *World Englishes*, 29(1):3-20.

Van Rooy, B. 2011. A principled distinction between error and conventionalised innovation in African Englishes. (In Mukherjee, J. & Hundt, M. eds. *Exploring second-language varieties of English and learner Englishes: bridging a paradigm gap*. Amsterdam: John Benjamins. p. 191-209).

Van Rooy, B. (in preparation). Grammatical description of a New Variety: the case of Black South African English.

Van Rooy, B. & Schäfer, L. 2003. An evaluation of three POS taggers for the tagging of the Tswana Learner English Corpus. *Lancaster University Centre for Computer Corpus Research on Language Technical Papers*, 16: 835-844.

Van Rooy, B. & Terblanche, L. 2006. A corpus-based analysis of involved aspects of student writing. *Language Matters*, 37(2):160-182.

Van Rooy, B. & Terblanche, L. 2009. A multi-dimensional analysis of a learner corpus. (In Renouf, A. & Kehoe, A. eds. *Corpus Linguistics: refinements and reassessments*. Amsterdam: Rodopi. p. 239-254).

Van Rooy & Terblanche 2010 Complexity in wordformation processes in New Varieties of South African English. *Southern African Linguistics and Applied Language Studies*, 28(4):357–374.

Van Rooy, B. & Van Huyssteen, G.B. 2000. The vowels of BSAE: current knowledge and future prospects. *South African Journal of Linguistics*, Supplement 38:15-33.

Van Rooy, B. & Van Huyssteen, G.B. 2003. The causative construction in Black South African English: emergent or emerging? Paper presented at the 8th International Cognitive Linguistics Conference, University of La Rioja, Logroño, Spain, 21-25 July.

Vendler, Z. 1957. Verbs and times. *Philosophical Review*, 66:143-160.

Wade, R. 1995. A new English for the new South Africa: restandardisation of South African English. *South African Journal of Linguistics*, Supplement 27:189-202.

Wolf, H-G & Polzenhagen, F. 2009. *World Englishes: A cognitive sociolinguistic approach*. Berlin: Mouton de Gruyter.

Wright, L. 1996. The standardization question in Black South African English (*In De Klerk, V. ed. Focus on South Africa*. Amsterdam: John Benjamins. p. 149-162.)