A CORPUS-BASED INVESTIGATION OF XHOSA ENGLISH IN THE CLASSROOM SETTING

A thesis submitted in partial fulfilment of the

requirements for the degree of

MASTER OF ARTS

of

RHODES UNIVERSITY

by

CANDICE LEE PLATT

January 2004

Supervisor: Professor V.A. de Klerk

ABSTRACT

This study is an investigation of Xhosa English as used by teachers in the Grahamstown area of the Eastern Cape. The aims of the study were firstly, to compile a 20 000 word mini-corpus of the spoken English of Xhosa mother-tongue teachers in Grahamstown, and to use this data to describe the characteristics of Xhosa English used in the classroom context; and secondly, to assess the usefulness of a corpus-based approach to a study of this nature.

The English of five Xhosa mother-tongue teachers was investigated. These teachers were recorded while teaching in English and the data was then transcribed for analysis. The data was analysed using Wordsmith Tools to investigate patterns in the teachers' language. Grammatical, lexical and discourse patterns were explored based on the findings of other researchers' investigations of Black South African English and Xhosa English.

In general, many of the patterns reported in the literature were found in the data, but to a lesser extent than reported in literature which gave quantitative information. Some features not described elsewhere were also found. The corpus-based approach was found to be useful within the limits of pattern-matching.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

My Supervisor, Professor V.A. de Klerk for inspiring me to begin this project, for her unfailing enthusiasm for the topic, and her patience and continued support even in trying circumstances.

The staff of the Department of English Language and Linguistics at Rhodes University, for their helpful advice, kindness, and concern, and for always believing in me.

Jon Leiper for sharing the Master's journey with me and making even the darkest moments bright.

Michelle Götte-Meyer and Neil Meyer for the use of their car to get to the schools and for their love, support and encouragement throughout this project.

My close friends for keeping me going and putting up with me. Your love, prayers, and support has helped more than I could say.

Emmanuel Lamprecht for all that you did for this project, from coming to the schools with me to writing programs and fighting with printers. Thank you especially, for never giving up on me, for dealing so well with my tears and frustrations, and for doing all that you could to make things easier for me while still encouraging me to grow. I could never have done this without you.

My extended family for always encouraging me to do my best and for always being proud of me.

My parents and my sister for being my *Chariots of Fire*, cock crow, starter's gun, watering tables, and much needed "physio" tents all along the way. Thanks for waiting for me to come in even after the sun had set and everyone was packing up to go! Thank you for believing in me, encouraging me, and always giving me so much love and support.

CHAPTER ONE: INTRODUCTION

Biber, Conrad, and Reppen (1998:1) point out that linguistic researchers tend to explore language either by looking at "what is theoretically possible in language" or by exploring language in use. This study, which looks at a specific register of a variety of South African English, is an example of an exploration of language in use. A detailed description of a language variety could not, however, be readily and accurately achieved without the use of corpus linguistics. Corpus linguistics is a vital approach to this exploration of language in use, as it allows for large amounts of language data to be stored and analysed (Biber et al. 1998). In this first chapter, I will begin by looking briefly at corpus linguistics - what it is, major events in its development, and its usefulness for this study. I will then introduce the variety of South African English which is the focus of my research. Thirdly, the aims of my thesis will be presented and, lastly, an overview of the thesis will be given in the final section of this chapter.

1.1 CORPORA

The term "corpus" comes from the Latin word meaning "body", and is used in corpus linguistics to refer to a collection of texts - standardly made up of one million running words (Nelson 1996) - to be used for linguistic research (Curzan 2000). As early as 1755, Johnson made use of corpus-based techniques to obtain data for his dictionary of English (Biber et al. 1998). Although such analyses of texts "by hand" date back to the early 17th century (Kennedy 1998), various scholars have criticised early corpus linguistics because of the considerable difficulty in the accurate analysis of large collections of texts (Abercrombie 1963; Chomsky 1964; both cited in McEnery and Wilson 1997). For such reasons, corpus-based linguistics is often believed to have effectively begun only in the early 1960s (Kennedy 1998). Two important events at this time led to rapid development in the field of corpus linguistics. The first was the Survey of English Usage (SEU) initiated by Randolph Quirk to serve as a basis for "the systematic description of spoken and written English" (Aijmer and Altenberg 1991:1). Secondly,

computers able to "store, scan and classify large masses of material" (Aijmer and Altenberg 1991:1) aided the advance of corpus linguistics enormously. As a result, modern corpora are standardly obtained in machine-readable form, or are easily typed or scanned into a computer database (Rundell and Stock 1992). Thus, huge bodies of text can be stored, allowing a variety of extracts to be used for analysis.

Since these crucial developments, various important corpora have been assembled. A brief summary is presented here, but the development of these corpora will be explored in detail in Chapter 2. Technological advancements have had a highly significant impact on the development of corpus linguistics. While the SEU (Survey of English Usage) corpus - begun in 1959 - is seen as "the last major pre-electronic corpus" (Kennedy 1998:23), the Brown Corpus - assembled in the early 1960s - is viewed as pioneering work in computer-based corpus linguistics (Meyer 2002). Published shortly before this work, Chomsky's *Syntactic Structures* (1957) presented a resolute and highly influential view that the aim of linguistic research should be to explore and describe what language users know - their competence - rather than their imperfect linguistic behaviours - their performance (Kennedy 1998). As a result, the Brown Corpus was developed in considerable opposition to this trend in linguistic thought at the time (Kennedy 1998). Soon after this corpus of written American English had been created, a British English corpus was also assembled - closely conforming to the Brown Corpus model (Kennedy 1998). This Lancaster-Oslo/Bergen (LOB) corpus was completed almost ten years after the Brown Corpus and was, as a result, able to take advantage of the development in computers that had taken place after the completion of the Brown Corpus (Kennedy 1998). Kennedy (1998) points out that the Brown Corpus has also been used as a model for more modern corpora of written varieties of Indian, New Zealand, and Australian English. Computer-based corpora have also been developed for a variety of specific purposes such as: in lexicography (Biber et al. 1998; Kennedy 1998); to explore developments or changes in language over time (Biber et al. 1998; Kennedy 1998); to investigate learner-languages (Biber et al. 1998); and for the characterisation of registers or dialects (Biber et al. 1998). As such, corpus

linguistics is central to this study, where the characterisation of a variety of English is the chief aim.

While the above-mentioned earlier corpora were patterned after the Brown and LOB corpora and sized, like them, at about one million words, technological advances altered this state of affairs in about the 1980s (Kennedy 1998). An important development is the CoBUILD Corpus (Collins Birmingham University International Language Database) (Kennedy 1998) - created in the 1980s and made up of about 20 million words (Meyer 2002). This Corpus was used to produce a new English dictionary (Kennedy 1998), but has since become part of a far bigger project - the Bank of English Corpus (Meyer 2002). As a result of rapid advances in computer technology, major corpus projects in the 1990s were able to make use of corpora of one hundred million words or more (Kennedy 1998). One such project was the British National Corpus (BNC) - compiled between 1991 and 1995 and consisting of 100 million words, 10 million of which were spoken words (Kennedy 1998).

On an even larger scale is the International Corpus of English (ICE) - a project exploring parallel corpora which sample the English used in a number of English-speaking countries involved in the project. A sub-project, for which computer corpora of the essay-writing of advanced students are being compiled, is termed the ICLE (the International Corpus of Learner English) (Greenbaum 1996). The ICLE includes, at present, only a small selection of eleven mother-tongue backgrounds (Granger 1996), and excludes any African language backgrounds. However, Schmied (1996) suggests that new opportunities for corpus linguistics could arise out of an extension of corpus work into English second language (ESL) contexts. This has obvious relevance to this study which explores a variety of English spoken as a second language in South Africa.

While the SEU (Kennedy 1998) and ICE make use of both written and spoken data (Greenbaum 1996), the focus of corpus work in the past has more often been on written texts (McCarthy 1998). This is largely due to the relative

ease of capturing written data (McCarthy 1998) as well as the enormous investments of time and money required to prepare spoken data for analysis (Kennedy 1998). However, a focus on written language is problematic, not only because text sampling tends to be opportunistic, but also because these data are often not very representative. McCarthy (1998) highlights the importance, too, of spoken data for the analysis of actual language. This is in contrast to the oft-made assumption that spoken language parallels the written form. In this study, then, I intend to make use of spoken data to get a clear picture of actual language used. In the next section, I will introduce the English variety that is explored and described in this study.

1.2 BLACK SOUTH AFRICAN ENGLISH

There are, essentially, three main difficulties in defining Black South African English (BSAE): the limited descriptions of the variety which exist (Gough 1996; Smit 2000); the prescriptive way in which the variety is generally defined in terms of its "deviation" from standard varieties (Gough 1996); and the tendency to focus on written English as a base for description. The fact that an extremely limited number of descriptions of BSAE exist is an obvious hurdle to its characterisation and the chief problem in attempting to define this variety. The second problem, as De Klerk and Gough (2002:356) point out, is that this variety has often been prescriptively defined in terms of its "deviat[ion] from the norm". Gough (1996:57) highlights that the English of black South Africans has traditionally been described only in terms of its "deviance" from standard English - displayed in titles such as "Common Language Errors" (Scheffler 1978) and "The abuse of English..." (Finn 1986). In contrast, it has been argued (Kachru 1990) that new varieties of English, which have developed out of the linguistic demands of "multilingual non-western states", also deserve to be made legitimate.

A third problem in defining and characterising BSAE is the tendency to focus on written English as a standard. This is evident in the fact that current corpora, (as discussed in section 1.1, above), are composed mainly of written bodies of language (Barnbrook 1996). These last two factors - viewing

BSAE as a "deviation" from standard English, and a focus on written text - perpetuate linguistic imperialism by upholding prescriptive values (Phillipson 1992). An important aspect of this study, then, is its shift of attention to spoken rather than written data. It is evident, in addition, that there is a definite need for an adequate description of BSAE. Furthermore, Gough (1996) states that variation within BSAE has not been researched and it is obvious that there is an urgent need for research of this nature, a view strongly supported by De Klerk (2002). For this reason, this study explores not only BSAE as a variety, but narrows the focus to look at the English spoken by Xhosa mother-tongue speakers - i.e. Xhosa English (XE). This scope is further narrowed by a focus on the English used by Xhosa mother-tongue teachers in classrooms. This allows for a more in-depth look at a specific register of XE and also allows for later, genre-based comparisons with other registers within XE.

To properly understand why the term BSAE arose, it is important to look briefly at the historical context in South Africa. Various authors (Gough 1996; Wright 1996) highlight the influence of the apartheid era on the acquisition of English by black South Africans. Although, as Nwaila (1993) states, teachers in "black" schools felt that they taught in standard English, and pupils felt that they were learning standard English, in reality this was not the case. Most black South Africans experienced what Macdonald (1990:1) terms a sudden "deep end" transition from mother-tongue instruction to all-English medium in the fifth year of schooling. An important related factor is that the vast majority of teachers in black schools are not English mother-tongue speakers, and are themselves products of the Bantu Education system (Gough 1996) and, therefore, their language displays characteristics of black English (Buthelezi 1995; Gough 1994). As a result, Black South African English (BSAE), as a variety, has been passed down through generations of black South African pupils (Gough 1996). The impact on pedagogy becomes apparent when one bears in mind the prescriptive view of BSAE as "deviant" from the standard, rather than a legitimate variety. This study aims to look more closely at BSAE in an effort to move away from the more traditional, prescriptive view to a descriptive view of this variety.

There is, at present, a move to describe BSAE in detail, based on the 9 mother-tongue sources of this variety. Important to this study, is De Klerk's (2002) work towards a description of a specific mother-tongue variety i.e. a corpus of spoken Xhosa English (XE). My study will fit into this broader picture by exploring XE in a specific context viz. in the classroom setting. This mini-corpus of XE will then later form part of De Klerk's (2002) larger corpus and be useful for comparison with similar data-bases of other (BSAE) varieties. This context will be more closely examined in later sections in the thesis. At this point, however, the aims of this study can be addressed.

1.3 THE RESEARCH GOALS

To ensure a closely defined area of investigation, the main aim of this study is to make use of corpus linguistics to carefully describe Xhosa English (XE) as used by teachers in the classroom setting. To achieve this aim, my goals are as follows:

1) To use a mini-corpus of approximately 20 000 running words of classroom discourse in English - collected from a representative sample of Xhosa mother-tongue teachers in Grahamstown - to describe, rather than explain, the characteristics of XE as a legitimate variety of English in terms of selected lexical and syntactic features.
2) To assess the usefulness of a corpus-based approach in a small-scale, pilot study such as this one, by exploring the efficacy and advantages of this method.

This study should give useful insights into the characteristic features of XE as used in the classroom. A description of current patterns in classroom XE - as used by teachers, arguably the educated "experts" in their speech communities - may improve perceptions of this variety as a variety in its own right, rather than a deviation from the "standard" (Gough 1996).

1.4 OUTLINE OF THESIS

A review of the relevant literature will be the focus of the second and third chapters of this thesis. As the central theoretical areas of concern - viz. corpus linguistics and BSAE - are not intrinsically related, they will be addressed in separate chapters. Chapter 2 reviews the literature relating to corpus linguistics, while BSAE is the subject of concern in Chapter 3. The methodology used for this study is discussed in Chapter 4, and the findings and discussion are the subject of the fifth chapter. The final chapter (Chapter 6) highlights the main usefulness of this study and discusses some of its inherent weaknesses while bringing together the major threads and looking ahead to future applications of the study.

CHAPTER TWO: CORPUS LINGUISTICS

In this chapter, I will discuss issues central to corpus linguistics - building on ideas introduced in the first chapter of this thesis. I will begin by expanding on the definitions of corpora given in chapter one and describing what is meant by corpus linguistics. The second and third sections of this chapter contain a more detailed overview of the history and development of corpus linguistics than that given in the introduction. Specific focus is given to the development of spoken corpora - in the fourth section - as this has particular importance to my study. I will then explore and summarise the potential advantages of corpus-based approaches in the fifth section. Finally, I look at the limitations of corpus-based approaches.

2.1 DEFINING CORPORA AND CORPUS LINGUISTICS

2.1.1 Corpora

In the introduction (section 1.1), I pointed out that the term "corpus" comes from the Latin word for "body" (Curzan 2000). Nevertheless, a linguistic corpus is not easily defined, and there is variation in the scope of language which could be included in a corpus (Meyer 2002). Generally, however, linguists favour the view of a corpus as a body or collection of texts which are available for analysis (Curzan 2000; Meyer 2002:xi). It is important to note, though, that "texts" need not necessarily describe only written language. Kennedy (1998:1) points out that, "in the language sciences a corpus is a body of written text *or transcribed speech* which can serve as a basis for linguistic analysis and description [my italics]". Nelson (1996) adds that a corpus is traditionally seen to be made up of one million running words, as evidenced in the size of, for example, the Brown Corpus (Aijmer and Altenberg 1991; Kennedy 1998; Meyer 2002) and the Survey of English Usage (SEU) (Aijmer and Altenberg 1991). This view can be contested, however, in the light of developments which have taken place in corpus linguistics - explored in sections 2.2. and 2.3 below. A basic criterion, then, is

that a corpus is a body of text - composed of spoken or written language - which can be used for linguistic analysis.

A further specification is presented by Barnbrook (1996:23-24) who argues that the term "corpus" refers to a particular "collection of computer-readable language" deliberately designed and selected with definite research goals in mind rather than "random collections of texts held in text archives". Kennedy (1998:4) expands on this point by explaining that, "[w]hereas a corpus designed for linguistic analysis is normally a systematic, planned and structured compilation of text, an archive is a text repository, often huge and opportunistically collected, and normally not structured". As such, corpora are specifically designed collections of text rather than a random accumulation of texts.

In line with Barnbrook's (1996:23) view that corpora are set up in "computer-readable language", Meyer (2002:xii), too, highlights that modern definitions of corpora tend to include reference to the fact that the body of text be available "in computer-readable form for computer analysis". Kennedy (1998:2) warns, however, that not all dealings with texts and computers fall under "corpus linguistics".

Although, traditionally, corpora were defined as being made up of one million running words (Nelson 1996), this criterion is becoming less and less important as technological advances make larger corpora increasingly practicable. Large corpora may, in fact, be highly desirable in many cases. Indeed, as De Klerk (2002:33) points out, "the larger number of variables to be covered [by a corpus], the larger a corpus must become" to be statistically significant. Smaller corpora, such as this one, are, however, not without merit. De Klerk (2002:33) argues that corpora which are fairly "homogeneous" - containing, therefore, "fewer variables" - can give a representative picture of a speech community without their becoming too "unmanageable" (De Klerk 2002:34) or being reduced to a collection of ideolects. As this study focuses on a small, fairly homogenous group of speakers, there are very few variables (as is discussed in the Methodology

Chapter). Despite its small size, then, this mini-corpus should be able to provide an acceptable preliminary description of the English language used by the community of Xhosa mother-tongue teachers in the Grahamstown area of the Eastern Cape.

In summary, a corpus is a body of text - spoken or written language - which can be used for linguistic analysis. It is important that the corpus be collected in a structured and systematic way for a specific purpose and presented in computer-readable format. In addition, the size of the corpus should be dependent on the purpose(s) for which that corpus has been designed.

2.1.2 Corpus Linguistics

Kennedy (1998:7) views corpus linguistics as "based on bodies of text as the domain of study and as the source of evidence for linguistic description and argumentation". Meyer (2002), in line with Kennedy (1998), points out that corpus linguistics should not be viewed as a paradigm within linguistics, but should rather be seen as a methodology for linguistic description. In this methodology, as Kennedy (1998:7) argues, the research involves the "quantification of the distribution of linguistic items". It is important to note, then, that corpus linguistics focuses on "performance rather than competence" and "on observation of language in use leading to theory rather than vice versa" (Kennedy 1998:7). As such, corpus linguistics describes a methodology for linguistic description which is based on the study of bodies of text. Importantly, corpus linguistics looks to language in use rather than any theoretical postulation of "competence". As a result, corpus linguistics enables researchers to build theory from real language in use rather than adapting actual language to a theoretical premise. The brief summary of the history and development of corpus linguistics given in the introductory chapter will be expanded on in sections 2.2 and 2.3 below.

2.2 THE HISTORY OF EARLY CORPUS LINGUISTICS

I pointed out in section 1.1 that analyses of texts "by hand" date back to the early 17th century (Kennedy 1998). A variety of important work with corpora has been done since that time. Kennedy (1998) writes, for example, about Cruden's biblical concordance which was first published in 1736. Since the publication of this work, various similar concordances have been built on the writings of "established authors" such as Shakespeare (Kennedy 1998:14).

Biber, Conrad and Reppen (1998) highlight the fact that the field of lexicography - unlike other areas of linguistic interest - has long been influenced by empirical, corpus-based approaches. A famous example, as mentioned in section 1.1, is Johnson's *A Dictionary of the English Language* (1755). Johnson made use of a collected corpus of written texts to gather authentic examples of English words in use for his dictionary (Biber et al. 1998). Kennedy (1998:114) draws attention to the fact that Johnson based his work on the effort of his "predecessors over the previous 150 years". The *Oxford English Dictionary* (OED), too, was based on a corpus of written texts collected in the late nineteenth century (Biber et al. 1998). The first volume of the OED was published in 1884, and the final - twelfth - volume in 1928 (Kennedy 1998).

Jespersen's seven volume work, *A Modern English Grammar on Historical Principles* (1909-49), also relied on early corpus linguistics. Jespersen discusses a variety of linguistic structures based on a corpus "representing the canon of English literature" from authors such as Chaucer, Shakespeare, and Austin (Meyer 2002:xii). An important consideration in these early uses of a corpus-based approach is that dictionaries generally had a largely prescriptive rather than descriptive function (Biber et al. 1998). In addition, written rather than spoken texts were used almost exclusively until the 1970s (Biber et al. 1998).

Despite the significance of works emanating from these first steps, early, manual use of corpus linguistics has been criticized by various scholars as

being insufficiently accurate in dealing with large collections of text (Abercrombie 1963; Chomsky 1964; both cited in McEnery and Wilson 1997). Kennedy (1998) highlights, therefore, that corpus linguistics is often viewed as essentially beginning only in the early 1960s. As mentioned in the Introduction, tremendous growth in corpus linguistics took place at this time as a result of two important developments. Firstly, the Survey of English Usage (SEU), initiated by Randolf Quirk, which aimed to describe both written and spoken English in a systematic way (Aijmer and Altenberg 1991) and, secondly, developments in computer technology. Aijmer and Altenberg (1991) point out that an advantage of Quirk's early corpus (the SEU) was its capacity to consist of approximately fifty percent spoken text - something which is commonly avoided in modern, computer-based corpora because of the practical difficulties which spoken texts present. This issue is addressed in greater detail in the section on spoken corpora (section 2.4), below. The second important occurrence was the development of computers able to cope with large quantities of data (Aijmer and Altenberg 1991). Because of these developments, corpora today are either obtained in a computer-readable form, or scanned or typed into computer databases (Rundell and Stock 1992). This means, then, that linguistic analysis can fairly readily be performed on enormous bodies of text. In the next section, I will explore the developments which came out of these early strides in corpus linguistics.

2.3 THE DEVELOPMENT OF CORPUS LINGUISTICS

The development of corpus linguistics has been greatly influenced by technological advancements. As mentioned in section 1.1, while the SEU (Survey of English Usage) - initiated in 1959 - is described as "the last major pre-electronic corpus" (Kennedy 1998:23), the Brown corpus is seen as pioneering work in computer-based corpus linguistics (Meyer 2002). The Brown corpus was developed by Francis and Kučera at Brown University in the early 1960s (Aijmer and Altenberg 1991) and became available in 1964 (Kennedy 1998). The corpus is made up of 2000-word samples - totalling one million words - of "edited written American English" (Meyer 2002:143). Meyer (2002:xii) points out that the Brown corpus can be viewed as a

"balanced corpus" as it is made up of samples from different genres of written English - allowing for a study of specific genres and for cross-genre comparisons.

Although the Brown corpus is now seen to have spearheaded work in corpus linguistics, it was assembled in substantial conflict with linguistic thought at the time (Kennedy 1998). Chomsky's highly influential work, *Syntactic Structures* (1957), was published only a few years before work began on the Brown corpus (Kennedy 1998). Hockey (2000) highlights that Chomsky's perspective held tremendous sway in North America for over two decades. Chomsky's prevailing view was that linguistic research should aim to explore and describe the linguistic competence - rather than imperfect performance - of language users (Kennedy 1998). Chomsky (in Hockey 2000:85) argued that a corpus "can never provide evidence for all possible linguistic features". As a result, researchers turned instead to "native speakers' intuition" and invented sentences (Hockey 2000:85). Once the Brown corpus was completed, however, it proved to have significant influence on linguistic research, despite such powerful opposition.

About a decade after the Brown corpus of written American English was completed, the Lancaster-Oslo/Bergen (LOB) of "edited written British English" (Meyer 2002:147) was begun - closely conforming to the model used by the Brown corpus (Kennedy 1998). Compiled between 1970 and 1978, the LOB corpus is composed of written texts all published in 1961 (Kennedy 1998; Aijmer and Altenberg 1991). Like the Brown corpus, the LOB corpus is also a balanced corpus consisting of about a million words divided into 2000-word samples (Kennedy 1998; Meyer 2002), but the LOB corpus was able to take advantage of development in computers that had taken place in the period after the Brown corpus had been published (Kennedy 1998).

As mentioned in the Introduction, many modern corpora of written English have been modelled on the Brown corpus (Kennedy 1998). Kennedy (1998) cites corpora of Indian, New Zealand and Australian English as examples. These corpora, as well as being modelled after the Brown and LOB corpora,

were also sized at about one million words, but this standard was altered in the 1980s when technological advances made far larger corpora possible (Kennedy 1998).

2.4 SPOKEN CORPORA

McCarthy (1998) directs attention to the fact that investigations into spoken language have taken place in the fields of anthropology and dialectology, with some of the earliest uses of tape-recorded data-collection being used in explorations of child language-acquisition. He does say, however, that the various obstacles involved in compiling spoken corpora make them far more difficult to assemble than written corpora (McCarthy 1998). This view is supported by the team working on the Corpus of Spoken Israeli Hebrew (CoSIH) who argue that the difficulties involved in putting together spoken corpora result in a skewing in the collection of data towards written data (Izre'el 2002). They argue, in fact, that corpora which make use of both spoken and written texts tend to favour written texts, with as much as 66 percent of the corpus being written (Izre'el 2002). Logically, the CoSIH team contend, this unbalanced representation, does not reflect language use in everyday life (Izre'el 2002).

In section 2.2 above, I mentioned the important contribution made to spoken corpus linguistics by the SEU (Survey of English Usage corpus) (Aijmer and Altenberg 1991). This contribution is important in two ways: firstly, in terms of the SEU's balance between spoken and written texts and, secondly, in terms of its focus on language in use rather than linguistic competence. In contrast to the typical focus on written language mentioned above, Aijmer and Altenberg (1991) point out that the SEU - composed of approximately one million running words of British English - consists of about 50 percent spoken language. Peppé (1995:187) argues that the "pioneering aspect of the ... SEU was the inclusion of informal conversational material". Peppé (1995:187) further points out that the tape-recorder made a significant contribution to corpus work by making possible the capturing of "the most natural form of language and the one that is most overwhelmingly dominant

for any individual in whatever walk of life" (Svartvik and Quirk 1980:9 in Peppé 1995:187). This view is in contrast to those who, like Chomsky, felt that the focus of linguistic interest should be on language competence rather than performance (Kennedy 1998; Hockey 2000).

While an obvious advantage to spoken corpora - for linguists interested in language in use - is the "naturalness" of the data obtained, present technology requires that conversations be transcribed into a written form for analysis (Peppé 1995:187). Barnbrook (1996:29) asserts that some form of "transcription convention" is needed to convert spoken language into a computer-analysable form as the speech-recognition technology currently available does not yet allow for spoken sounds to be "processed as language" (Barnbrook 1996:33). Various authors (e.g. Barnbrook 1996; Meyer 2002; Peppé 1995) write about both the practical and the methodological difficulties involved in the translation of spoken text into a written form. A significant practical difficulty in spoken corpora is that of transcribing tape-recorded, spoken language into a written form. Meyer (2002:70) admits that transcription is a "time-consuming task". The time taken to transcribe spoken language into a written form, Meyer (2002:71) argues, "runs a continuum". He maintains that "multi-party dialogues with numerous overlaps" are the most time-consuming, while "scripted monologues" fall at the opposite end of the continuum (Meyer 2002:71). Meyer (2002) claims that the transcription, marking-up, and proof-reading of a 2000-word section of a corpus can take - from one end of his proposed continuum to the other - anything from five to twenty hours to complete. A second practical problem - which relates to the time taken for transcription - is the financial cost of compiling a corpus of spoken language (Meyer 2002). Besides the practical difficulties, such as the time and cost involved in transcription, various methodological issues are also important reasons for the imbalance in spoken to written corpora.

Methodological issues include the "highly artificial" nature of representing "an exclusively oral form of language ... in written form" (Meyer 2002:71). Knowles (1995:208) questions, for example, the custom of presenting spoken

texts in a manner determined by "Western writing systems". Peppé (1995:187) writes that, in the conversion of spoken language into a written form, "some form of selectivity and interpretation is bound to be involved". Work on the SEU was, therefore, subject to the "'total accountability' principle of corpus-based study" (Peppé 1995:187). As such, phoneticians worked in pairs to transcribe and check all tape-recorded material (Peppé 1995). Peppé (1995) further explains that the SEU is made of texts of approximately 5000 words each - taken from language-in-use situations for which the contexts have been clearly recorded.

Although the SEU made an important contribution to spoken corpus linguistics, this influence was vastly extended by a move to make the SEU corpus internationally accessible (Leech, Myers, and Thomas 1995). The machine-readable form of the SEU was "made available by the Survey of Spoken English at Lund University" in 1975 and "became known as the London-Lund Corpus (LLC)" (Leech et al. 1995:188).

Another significant contribution to spoken corpus linguistics was made by the CoBUILD corpus (Collins Birmingham University International Language Database) (Kennedy 1998) composed of approximately 20 million words (Meyer 2002), over 10 million of which were made up of a "wide variety of spoken English" (Payne 1995:203). An even larger project, begun in 1990, is the International Corpus of English (ICE) which aims to allow for comparison between corpora of the national English variety of more than 16 countries (Nelson 1995). Although this project aims for corpora which reflect both spoken and written varieties of English, 600 000 of each million word sample collected is from spoken language.

2.5 ADVANTAGES OF A CORPUS-BASED APPROACH

There are several advantages to the use of corpora in linguistic research. These advantages can be grouped around two main themes, viz. those which relate directly to corpus linguistics, and those which relate more generally to the use of computers in linguistic analysis. Addressing those advantages

which relate most directly to corpus linguistics, there is, firstly, the advantage of being able to explore language in use, rather than focusing on the idealised "competence" of speakers (Kennedy 1998:7). A second, related, advantage is, as Kennedy (1998:4) points out, that a corpus not only allows for the identification of linguistic patterns which make up languages, but also permits a researcher to "map out" how these linguistic systems are used. Lawler and Dry (1998) highlight, however, that encoding or "markup" greatly increases the usefulness of a corpus by making it possible for a researcher to focus on specific features of interest. Kennedy (1998) points out that many modern corpora have been designed with the specific purpose of describing various patterns in a language. Such an understanding of the systems of a language has many uses including machine-processing of the language as well as a better conception of how best to teach the language (Kennedy 1998).

Barnbrook (1998:11) looks at four important benefits to the use of computers in linguistic research. Firstly, the "speed, accuracy and consistency of processing", secondly, "the ability to perform further processing on the results". Barnbrook (1998:11) points, thirdly, to the "ease with which data [can] be manipulated, selected, sorted and formatted". Finally, the "lack of any human bias" is presented as an advantage to using computers in linguistic research (ibid.).

2.6 LIMITATIONS OF A CORPUS-BASED APPROACH

Despite the advantages to using corpora for linguistic research, there are certain drawbacks which should be considered. The disadvantages, too, can be grouped around the two main themes addressed in section 2.5 above, (corpus linguistics and the use of computers in linguistic analysis). Kennedy (1998) describes two important limitations to using corpus linguistics. Firstly, Kennedy (1998) argues, there is a danger in thinking that any use of computers and texts in combination constitutes a corpus linguistic methodology. As discussed in section 2.1 of this chapter, it is important to clearly define corpora in any research undertaken in this field. Kennedy's

(1998:4) second point is that a corpus should have a clear "representative function" and not simply constitute - as expressed earlier - a linguistic archive. He argues that "[i]n many respects it is ... the use to which the body of textual material is put, rather than its design features, which define what a corpus is" (ibid.). The factors which are important to consider when designing a corpus are addressed in the Methodology Chapter of this thesis (section 4.1, below).

Four disadvantages to the use of computers in linguistic analysis are presented by Barnbrook (1998). Firstly, Barnbrook (1998) writes about the effort involved in organising texts, which are seldom available in computer-readable form, for processing. This is especially true, of course, in the case of spoken texts as was discussed in some detail in section 2.4 above. Barnbrook (1998) mentions, too, the difficulties arising out of having to adjust the process of analysis according to the operations of the computer. Thirdly, the "extra work involved in program development and testing" is presented as a weakness in the use of computers in linguistic analysis (Barnbrook 1998:12). Barnbrook's (1998:12) final concern is "the computer's lack of normal human background knowledge". This, naturally, requires that the researcher be alert to possible widening or narrowing of focus on the computer's part.

CHAPTER THREE: BLACK SOUTH AFRICAN ENGLISH

This chapter explores theoretical concepts central to this thesis, but not directly related to corpus linguistics. As discussed briefly in the introduction, Black South African English (BSAE) is difficult to define. This was shown to be due to the paucity of descriptions of the variety; the view of BSAE as a "deviation" from standard English; and the focus on written rather than spoken English in descriptions of BSAE. The introduction also touched on the historical context in South Africa which led to the development of BSAE. In this chapter, I will look at both of these aspects - defining BSAE and its historical development - in greater detail as well as describing BSAE as it is standardly classified. I will begin with an overview of the historical context of black education in South Africa - to give a greater understanding of how BSAE evolved - and then move on to defining BSAE. The third section of this chapter looks at BSAE in the classroom. The final section briefly refers to the characterisation of BSAE.

3.1 HISTORICAL CONTEXT

As BSAE is generally believed to have developed out of the teaching of English to black South Africans (De Klerk forthcoming), this section follows the chronological development of black education in South Africa, focusing, where appropriate, on the events and conditions in the (Eastern) Cape area. I have made fairly extensive use of the work of Behr (1970; 1984), who writes in depth on the development of education in South Africa. Brown (1992), too, gives a quite detailed account of the sociolinguistic history of South Africa and much of his report is also relevant here.

In looking at the historical development of BSAE, the first period of interest is that up to the unification of South Africa in 1910. Brown (1992) points to the influence which missionaries have had over education in South Africa - a fact supported by De Klerk and Gough (2002). In fact, Behr (1984:173) argues that, up until "well into the 19th century", missionary societies in the Cape Colony both funded and ran schools for black South Africans as, "[t]he

Government itself made little provision for the schools for Non-Whites other than the slaves". In 1839, the mission schools came under the charge of the newly formed Department of Education (Behr 1984) and, from 1841 onwards, state-aid was made available to the mission schools, establishing "the system of state-aided mission schools ... as the pattern of Black education until the 1950s" (Behr 1984:173). Government's subsidy of mission-schools cannot be seen as entirely altruistically motivated, however, when one considers the views of some Government officials in South Africa in the mid nineteenth century. For example, Behr (1984:174) argues that the governor of the Cape in 1855, Sir George Grey, saw the education of black people in South Africa as "the most important factor in the peaceful subjugation of the Blacks" and managed to convince the British government to subsidize missionary schools to enable them to train black people and "fit them to act as interpreters, evangelists and schoolmasters *amongst their own people* [emphasis mine]" (Grey in Behr 1984:174). This view, that black people should be treated differently to, (and kept separated from), whites, was shared by Dr. Thomas Muir, Superintendent-General of Education (1892-1915), who felt that the courses offered to black pupils - similar in content to those offered to white pupils - were "altogether too bookish and unpractical" (in Behr 1984:174). According to Behr (1984:174), Muir further maintained that "little or no attention was given to the teaching of the vernacular, and far too much time was devoted to the teaching of English as a subject". As such, the seeds of the "apartheid" philosophy were already planted in South Africa's early history while still under British rule.

The passing of the South Africa Act (1910) by the British government, meant that South Africa became a "self-governing dominion within the British Empire" (Leach 1989:35) - the Union of South Africa (Behr 1984). Leach (1989:35) says that the passing of the South Africa Act meant that: "[a]t a stroke the British government washed its hands of all responsibilities for the country's non-white population". Besides education, all matters pertaining to black South Africans were handed over to the Minister of Native Affairs (Behr 1984). Missionaries were, however, still extensively involved in black education in South Africa, as shown in the fact that, in about 1925, there were

2 702 mission schools, enrolling 215 956 pupils, compared to 68 state schools, with enrolment figures of 7 710 pupils (Behr 1970). The Report of the Interdepartmental Committee on Black Education, dated 12 July 1935, showed that, at that stage, each province had its own provincial council as well as an "advisory board" which ensured liaison between the education departments and schools (Behr 1984:176). The Committee reported that, although state-aided mission schools were necessary at the time, steps should be taken to ensure that the state would eventually take over all responsibility for black education (Behr 1984). From around 1935, after a minimum of four years of teaching in the vernacular, English was used as the medium of instruction in black schools (Gough 1996). During this time, mother-tongue speakers of English taught English to black children at the mission schools (De Klerk and Gough 2002). Brown (1992) points out, however, that apartheid ideologies were critical of the education system and language policies of the missionaries.

The coming to power of the Afrikaner Nationalist government in 1948 was, according to Walters (1996:212), "undoubtedly the single most important event" for black education in South Africa. Walters (ibid.). points out that the government took almost total control over "all aspects of education for black South Africans". In 1949, the newly elected Nationalist government set up the Eiselen Commission (Walters 1996), under Dr. W.W.M Eiselen (Behr 1984). This Commission was tasked with establishing plans "to provide 'education for Blacks as an independent race, in which their past and present, their inherent racial qualities, their distinctive character and aptitude, and their needs under ever-changing conditions' were taken into consideration" (Rose and Tunmer 1975:244 in Behr 1984:179). Thus, as Behr (ibid.) argues, "[t]he Commission ...began with the premise that a distinction should be drawn between White and Black education". The Eiselen Commission's report was brought out in 1951 and "its main recommendations were embodied in the *Black Education Act*, 1953" (Behr 1984:179) - described as "the watershed in the control of educational services for Blacks in this country" (Ackerman 1975:22-38 in Behr 1984:181). This Act resulted in an extension of the vernacular as medium of instruction for a further four years -

with English and Afrikaans taking over this role as "dual media-of-instruction" (Wright 1996:150). Hartshorne (1995) explains that the passing of this act was met with particularly strong opposition in the Cape where many teachers either resigned or were dismissed for failing to implement the new policy. Walters (1996:212-213) explains the situation in greater detail by saying that, as a result of the report from the Eiselen Commission, English instruction in black schools was to be delayed to "successively higher school standards", coupled with "an equal demand for the use of Afrikaans as medium". Walters (ibid.) cites Hartshorne (1992:197) in this regard:

> In the primary schools the Eiselen recommendations on mother tongue medium were adopted in full, and without exception schools were required to extend the mother tongue medium, class by class, year by year, starting with Std. 3 [Walters adds here "the fifth year of school"] in 1956 through to Std. 6 in 1959, when the Std. 6 public examination was, for the first time, written in one or other vernacular language instead of in English. A new spirit of doctrinaire inflexibility had entered into a domain which had been treated flexibly in the past.

In addition, in 1955, the South African Bureau of Racial Affairs (SABRA) - a body "inspired by the Broederbond" and with "connections with the National party" (Lazar 1986 in Brown 1992:82) - published a pamphlet admonishing the mission education system's language policies and teacher-training programmes (Brown 1992). The move to "return to the use of the vernacular as medium of instruction in African schools" (Brown 1992:82) saw the transfer of missionary teachers from "African schools". Hartshorne (1992:197 in Walters 1996:213) states that, "those major ex-mission institutions which had not been closed down during the transfer of control to the State, were put under the control of white, Afrikaner Nationalist principals and teachers". Hartshorne (ibid.) adds that these principals and teachers were "committed to the ideologies of apartheid and Christian National Education" and that Afrikaans "soon became the dominant language in black education, especially at the levels of management, control and administration and

teacher training". Wright (1996:150) points out that this meant that "mother-tongue English teachers in black schools were now a rarity". As a result, English was used less and less in everyday life and become instead a "language of the classroom" (Wright 1996:150). Walters (1996:213) argues that it was not long before the impact of these policy and practices was made evident "in the form of declining levels of competence in English amongst the products of 'Bantu Education'".

An enormous population growth in the black community in South Africa, coupled with the government's "deliberate under-provision of educational infrastructure and supplies" (Wright 1996:151) meant that the education system for black learners was put under serious strain. Wright (1996:151) emphasizes that this system was then brought to virtual collapse by "liberal forces...hoping to use the schools as a power base in the political struggle". As mentioned earlier, Wright (1996) says that as a result of these factors, black South Africans had very little contact with mother-tongue speakers of English. Instead, teachers, who were themselves products of the black education system, were responsible for teaching English to the pupils in their classes. Walters (1996:213-215) points to efforts to counter this trend. He speaks of L.W. Lanham's work with primary school teachers in Soweto in the early 1960s - where "attempts [were made] to improve the teaching of English to their pupils" - as well as the founding of the Institute for the Study of English in Africa (ISEA) at Rhodes University, Grahamstown, in 1964. Walters (ibid.) also talks in some detail of the Molteno Project which was set in place to obtain "an analysis of the problems connected with the use of English as medium in African schools and the recommendation of methods for preparing lower primary children for its use for the fifth year of schooling upwards" (Rodseth 1978:1 in Walters 1996:214). Despite these interventions, however, for most black pupils, learning English with little or no contact with native tongue speakers resulted in "non-standard" patterns of English becoming entrenched in the English of black South Africans (Wright 1996:151). Wright (ibid.) also stresses the increased importance placed on written English texts in the absence of contact with mother-tongue speakers of English. In addition, deviances in the language produced by students,

which would normally be corrected by teachers, become "habitual" through regular, unchecked use (Wright 1996:151).

De Klerk and Gough (2002), too, comment on the role of the apartheid government in the development of BSAE. They point out that the government's education policy of mother-tongue instruction in black schools, while ostensibly encouraging the "inalienable human right to preserve separate identity" (De Klerk and Gough 2002:356), was actually "a way of reinforcing separateness" (ibid.). They also highlight the suspicion with which this policy was regarded by black South Africans and the subsequent "demand for the forbidden" by black speakers wanting access to the opportunities which a knowledge of English afforded (De Klerk and Gough 2002:357). The well-known "Soweto uprising" in 1976 resulted in an adjustment of government's view of language in education (ibid.). The Education and Training Act, 1979, repealing "all existing legislation relating to Black education promulgated between 1953 and 1978" (Behr 1984:200) includes a clause which required mother-tongue instruction for the first four years of school, but thereafter, as De Klerk and Gough (2002) state, gave permission for schools, together with parents, to "choose their own medium of instruction (MOI)" from one of the official languages of South Africa (Behr 1984:200). Hartshorne (1995:313) points out that, when the first draft of the Bill, 1978, leading to the Education and Training Act, 1979, was introduced by government, "it was clear that its intention was to maintain six years of mother-tongue instruction". He highlights, though, that "a storm of protest" resulted in government conceding to mother-tongue instruction being mandatory only up until, and including, the fourth year of school (ibid.). Once schools were given the choice of MOI, however, English was "the overwhelming choice" (De Klerk and Gough 2002:357).

Macdonald (1990:1 in Gough 1996:54) speaks of the "sudden 'deep end' transition to English medium of instruction which takes place, for the majority of black South Africans, in the fifth year of schooling. Gough (1996) argues that both the teaching materials used and the lack of mother-tongue English speakers as teachers in black schools mean that this transition is not

supported for either teachers or learners. The teaching of English in black schools is, therefore, "in the hands of non-native speakers of the language, who have often been unqualified or under-qualified for the task" (Buthelezi 1995:242). Furthermore, the change-over to English as medium of instruction corresponds with an enormous student drop-out rate (Macdonald 1990 in Gough 1996). Those students who do stay in school are taught by speakers of BSAE who often teach in their mother-tongue despite English being the official medium of instruction (Buthelezi 1995 and Gough 1994 in Gough 1996). This view is also supported by Amuzu (1992:129), who, in his study of the role of English in education in South Africa, looks at "the way English is being used as a medium of instruction in certain parts of the country". Amuzu (1992) states that, in his experience, English may officially be the medium of instruction, but in practice this is not the case. From my own very limited experience in ex-DET[1] schools in the Grahamstown area, a number of teachers code-switch between English and isi-Xhosa in their teaching, but some seem to use isi-Xhosa the majority of the time. From informal conversations with teachers, however, many feel that it is more important that the students are able to grasp the material being taught than to stick rigidly to English as LOLT (Language of Learning and Teaching).

Gough (1996:55) speaks of the changes that have taken place in South Africa since the desegregation of state education in 1991. Since this time, many former students from the Department of Education and Training (DET) have moved across to "white" schools. He argues that the resulting increase in contact with mother-tongue speakers brings with it an increased pressure to speak English in a more standard way (ibid.). Gough (1996) states, too, that parents of black children often place pressure on them to speak English at home as well as in the more formal educational context. Gough (1996:55) mentions Schlebush's (1994) experiences in Model C (i.e. formerly white) schools which show that some language shift is occurring in these speakers with a "decrease in competence in the mother tongue" becoming a factor. This observation makes the current trend in studying BSAE that much more

[1] The Department of Education and Training (DET) was the new name for "Bantu Education" from about 1980 (Hartshorne 1995:312).

important, as increasing contact with mother tongue English speakers may well play a role in changing the characteristics of BSAE over the next few decades. Many theorists argue, however, that the reverse may well be true - that BSAE may have an increasing influence on "standard" South African English. This issue will be explored in greater detail in the following section which looks at defining BSAE.

3.2 DEFINING BLACK SOUTH AFRICAN ENGLISH

Although certain - mainly white (Gough 1996) and Indian (Mesthrie 1992) - varieties of English spoken in South Africa have been studied and described, the variety spoken by black South Africans is only recently being explored (Gough 1996). A major reason for the scarcity of descriptions of BSAE (and other "non-standard" varieties of English in South Africa), Van der Walt (1999:46) argues, is that many linguists see the language used by black South Africans as "incomplete or deficient subsystems of English" (e.g. Lanham 1990; Titlestad 1993; Wright 1996). Van der Walt (1999:46) emphasizes that so-called "errors" which occur widely in a linguistic community - to the extent that they have become "features of a specific code" - should be seen as "features" rather than errors. Van der Walt (1999:46) further points out that until the varieties of English spoken in South Africa are recognized as "full-blown systems of communication for certain groups of speakers", where consistently and widely-used "errors" are seen, instead, as "features" of a variety, no descriptions can be started. Since Van der Walt's (1999) strongly argued assertions, however, much work has begun in South Africa to address this issue. For example, De Klerk's (2002) work at Rhodes University on a corpus of Xhosa English as well as work done by Wissing at Potchefstroom University on a corpus of learner Tswana English (De Klerk 2002:22).

A succinct definition of BSAE is given by Van Rooy (n.d.) who uses the term "Black South African English" to refer to "the second language variety of English spoken by the large majority of South Africans who speak one or another Bantu language as first language". This definition, while clear and

straightforward, can be fleshed out in a number of ways. Firstly, an important consideration in any definition of BSAE is the fact that speakers of Bantu languages in South Africa have different proficiencies in English - ranging from those who speak English as a "second first language" (De Klerk 1996) to those with "minimal competence" (Van Rooy n.d.). This, naturally, means that, even if clear norms for the variety are established in the first place, not all speakers of BSAE will conform to all the norms of the variety. Indeed, participants in this study - teachers who are expected to use English as the language of learning and teaching (LOLT) - should have a relatively high English proficiency. As such, their language could be expected to lean more toward standard English. This will be explored in greater detail in the Methodology Chapter as well as the Findings and Discussion section.

Secondly, Gough (1996) reveals that the typical context of acquisition and use of English for black learners shares some characteristics with new Englishes described by other authors (he lists Sridhar and Sridhar 1992 and Williams 1987). Other authors have also expressed the idea that the grammatical features typical of BSAE are common to other new Englishes both in general and in Africa (Platt, Weber and Ho 1984; Schmied 1991; De Klerk and Gough 2002). Van der Walt and Van Rooy (2002), too, explore the idea that BSAE is "approaching the status of a New English" in that it is "an institutionalised second language variety" that is taught in many township schools. Wright (1996:150) approaches this point rather more pessimistically when he describes the teaching of English in black schools in South Africa: "generations of black South Africans have been forced to learn English from each other in situations which offered limited occasion for its use and little contact with native-speaker norms". As such, Wright argues that BSAE results almost exclusively from the education system in South Africa as speakers of BSAE had little or no contact with other varieties of English (see also Mugoya 1991 in Gough 1996:54). (This issue was discussed in detail in section 3.1, above). In whichever light the situation is viewed, however, it is important to bear in mind that BSAE is currently perpetuated in many schools in South Africa. It is, therefore, deeply entrenched in the education system. The variety of BSAE taught in schools, though, may well differ to some

degree from that more generally described. This point is discussed in greater detail in Chapter Five.

The variation which exists within BSAE is a third issue to be taken into account when defining BSAE. Gough (1996 after Schmied 1991) explains that, while variation in South African Indian English, for example, has been explored (Mesthrie 1992), variation in BSAE has not been researched to the same extent. He points out that such variation results from influencing factors such as: the continuum of proficiency in English among speakers of BSAE (as discussed above); the role of different mother tongues (mentioned below); and regional features. Gough (1996) argues that there is a need for further research in the area of variation within BSAE. De Klerk (2002; forthcoming) supports this idea by presenting a strong case for looking at variation within BSAE. In her argument, De Klerk (2002) points out that speakers of BSAE make up 9 indigenous languages in South Africa. While these 9 languages can be grouped into 4 language families, she stresses that "it is natural to assume that different varieties of English have evolved along slightly different lines in different areas, based on a shared [mother tongue] and shared local values and traditions" (De Klerk 2002:35). For this reason, De Klerk (2002) argues, an ideal solution would be a large corpus of BSAE made up of nine equally sized smaller corpora from each of the 9 indigenous languages. This corpus would then allow researchers to find those characteristics shared by different mother-tongue speakers and those which can be linked to a specific mother-tongue background (De Klerk 2002).

A fourth issue, which bears consideration when defining Black South African English, is that dealt with in some detail by Wright (1996): standardization. Wright (1996) argues that BSAE needs to be standardized if South African English is to be used as a lingua franca - both within South Africa and for international communication. English is seen by many black South Africans as a passport to "liberation" and "success", Wright (1996:152-3) asserts. As such, though, BSAE should "observe norms very close to those of international standard English" which Wright (1996:150) believes "educated Afrikaans, coloured and Indian varieties" of South African English already do.

Corpus studies by various researchers (e.g. De Klerk 2002; Van Rooy n.d.) show that there is an interest in describing BSAE in terms of the mother-tongue languages of its speakers. While such close explorations of potentially different varieties of BSAE may not result in answers to Wright's (1996) standardization question, they should, nevertheless, give a clearer indication of the extent to which BSAE can be viewed as a single variety of South African English.

Simply put, then, BSAE is that variety of English which is spoken by mother-tongue speakers of one of the indigenous languages of South Africa. Speakers of BSAE vary in their competence in English and, as such, the degree to which BSAE differs from Standard English (SE) varies from BSAE speaker to BSAE speaker. BSAE - arguably a new English - is perpetuated in schools. There is considerable variation in BSAE, not only because of different English competencies of the speakers, but also because of their range of mother-tongue influences on the variety. It has been argued that BSAE needs to be standardized if it is to be intelligible both within South Africa and internationally. So, while BSAE may be succinctly definable at a surface level, it would be a mistake to think that there is not more depth to this variety.

3.3 BLACK SOUTH AFRICAN ENGLISH IN THE CLASSROOM

The preceding sections in this chapter have looked at the historical development of BSAE (section 3.1) and ways in which this variety can be defined (section 3.2). This section draws on each of these previous sections to form a picture of the issues around BSAE which apply specifically to the classrooms in South Africa today.

In section 3.2, above, I mentioned that there are four important considerations which need to be borne in mind when defining BSAE, viz.: BSAE speakers' different levels of English competency; the characteristics which BSAE shares with new Englishes elsewhere; the variation within BSAE; and issues around standardization of English in South Africa. In the

classroom situation, I believe, these issues condense to the importance of standardization and accepted norms. This is especially true in an educational context where one variety of English is necessarily held up as the "model for teaching and learning" (Van der Walt and Van Rooy 2002:113) and where teachers of English serve as models for their students (Mafisa 2002). Van der Walt and Van Rooy (2002:113) discuss this issue in the light of research into the norms of BSAE. They point out that, on the one hand, studies of BSAE usually focus on "its character as a second language in an educational context" - looking at it in terms of its "deviance from standard English" (Gough 1996:57). On the other hand, there are arguments that South African English (SAE) should be adjusted to recognize "some of the emerging norms of BSAE" (Van der Walt and Van Rooy 2002:113). The difficulty of reconciling "standard" English with the English used by the majority of South Africans is especially pertinent in the South African classroom, particularly in relation to issues of language assessment (Van der Walt 2001). Van der Walt (2001) argues that the central problem is that assessment necessarily requires a norm - a factor which is seldom considered.

Van der Walt (2001:2) emphasizes that any discussion of a norm in a language needs to be considered in terms of "three sociolinguistic facts": firstly, languages are constantly changing; secondly, most languages consist of a variety of dialects; and finally, for most languages, the standard variety derives from the dialect of the elite group. The standard language of a society, as Kaschula and Anthonissen (1995:4) explain, is "a dialect which is more highly valued than other dialects of the same language". In South Africa, the dialect of white, mother-tongue speakers was, arguably the standard dialect, although various authors (Wade 1997; De Klerk 2002) argue that BSAE is increasingly gaining prestige. The growing status of BSAE has led to more determined efforts to recognize the norms which are emerging in the variety (Van der Walt and Van Rooy 2002). Van der Walt and Van Rooy (2002) present a convincing argument that South African English is in the process of becoming a 'new' English - or a non-native institutionalized variety of English (NIVE) (Williams 1987:162).

Gill (1999:216 cited in Van der Walt and Van Rooy 2002:116) suggests three stages in the development of any new English (Platt, Weber and Ho 1984) norm. These three phases are: the exonormative phase, the liberation phase, and the endonormative phase. The exonormative phase, as its name suggests, is evident when there is an external norm for a language which is clearly recognized and applied by teachers of English. Since teachers are familiar with the norm and confident in its use, they apply it consistently and learners move towards this norm albeit with typical learner errors along the way.

The second phase identified by Gill (1999 in Van der Walt and Van Rooy 2002) is the liberation phase. The important characteristic of this phase, Gill (ibid.) argues, is that there is uncertainty about the norm which results in its being inconsistently used by teachers. A natural result is that learners, too, display confused and inconsistent use of the norms of the language.

The final phase in the establishment of a new English, is the endonormative phase (Gill 1999 in Van der Walt and Van Rooy 2002). This phase is again one of certainty, but, unlike the exonormative phase, the norm in the endonormative phase is internal. This means that, teachers confidently apply the "local educated" norm which is internally consistent. Gill (ibid.) argues that learners in this phase "exhibit the most consistent linguistic behaviour" as the English variety which they acquire is "already adjusted to their environment" and input from teachers is consistent.

Van der Walt and Van Rooy (2002) argue that BSAE is, at the moment, in the liberation phase of development on its way to becoming a new English. They base this argument on a study which looked at the attitudes and practices of a group of sixty black English language teachers in the Potchefstroom area as well as a grammaticality judgement task performed by 670 BSAE speakers studying English at the third-year level at a tertiary institution. The teachers in Van der Walt's and Van Rooy's (2002:125) study were found to see a norm as important, while still accepting some grammatical features of BSAE (ibid.). At the same time, the teachers were found to be inconsistent in their

application of the norm - often rejecting standard English constructions and accepting non-standard usages (ibid.). The learners, too, displayed general confusion and a "high level of acceptance of non-standard features" (Van der Walt's and Van Rooy's 2002:125). This evidence taken together led Van der Walt and Van Rooy (2002:125) to conclude that South African English is now in the liberation phase of development.

While it appears that the status of BSAE is improving in South Africa, there is still a level of uncertainty about the norms of the variety. Van der Walt and Van Rooy (2002) argue convincingly that this uncertainty is evidence that South African English is in the process of becoming a new English. If this is true, it seems likely that teachers model English inconsistently, which may impair student performance (Mafisa 2002). This issue is addressed in the Findings and Discussion Chapter of this thesis.

3.4 DESCRIBING BLACK SOUTH AFRICAN ENGLISH

Several authors have discussed characteristics of BSAE (Wade 1997; Van der Walt and Van Rooy 2002; De Klerk and Gough 2002; De Klerk forthcoming). Many of these descriptions have some basis in Gough (1996), who gives a detailed description of the features typical of BSAE. I have, therefore, used the classifications of BSAE used by Gough (1996) and De Klerk and Gough (2002) as a point of reference for my findings in this study. To avoid unnecessary repetition, descriptions of BSAE - from Gough (1996) and others - are given alongside my findings in Chapter Five.

CHAPTER FOUR: METHODOLOGY

The main aim of this study, as presented in the introductory chapter, was to use corpus linguistics to describe the variety of Xhosa English (XE) used by Grahamstown-based teachers in the classroom. While the historical development of corpus linguistics was discussed in the second chapter of this thesis, in this chapter I will look in some detail at the practical considerations involved in designing a corpus. I will also describe how I collected my data and prepared it for analysis. I will then explain the steps involved in the analysis of the data. Limitations of the study will be addressed in the final section of this chapter.

4.1 CORPUS DESIGN

The Wellington Corpus of Spoken New Zealand English (WSC), completed in 1994, has provided invaluable information regarding the steps to take, as well as problems to anticipate, in the construction of a corpus. In addition, Rundell and Stock (1992) highlight three important considerations in the design of a suitable corpus: the size of the corpus; the "boundedness" of the corpus; and corpus structure and design. The relevance of these factors to this study is discussed below.

4.1.1 The size of the corpus

A standard corpus for the International Corpus of English (ICE) consists of approximately one million words obtained from 500 texts of about 2 000 words each (Nelson cited in Greenbaum 1996). McCarthy (1998:128), however, points out that the optimal corpus size is a "largely irresolvable" issue. It has been suggested, for example, that corpora of a relatively small size, (20 000 - 30 000 words), can give useful insights if carefully designed with a particular goal in mind (Carter & McCarthy 1995 in McCarthy 1998).

As this study is a pilot exploration, the target was to obtain approximately 20 000 words of spoken XE which will ultimately form part of a full corpus of

Xhosa spoken English currently being compiled by De Klerk (2002). De Klerk's corpus consists, at present, of about 540 000 words of "naturalistic spoken data" - approximately 73% private and public dialogue and 23% unscripted monologue (De Klerk 2002; De Klerk forthcoming). These data are from 299 speakers - all of whom are Xhosa speakers of direct Xhosa descent (ibid.). The contributors to De Klerk's corpus are also all over the age of 15 years and have had either eight years of formal English tuition or have used English in their daily lives for twenty or more years (ibid.). A final specification is that all contributors must reside in the Eastern Cape province (De Klerk 2002; De Klerk forthcoming).

4.1.2 The "boundedness" of the corpus

This study will focus on spoken texts obtained within a limited time frame, of approximately 15 months. As such, the mini-corpus will be bound by time, giving a synchronic description of a sub-genre of XE. This will allow for a detailed description of the characteristics of this variety, allowing for comparison across other varieties of BSAE in due course, as well as with other varieties of English world-wide. Schmied (1996) points out that corpus linguistics is able to be enriched by "overdifferentiation", i.e. by making as many differentations as possible. That is to say, by adapting ICE subcategories to "more community-specific texts", richer information can be obtained. In the case of this thesis, which focuses specifically on the language used by teachers, this enriching function can be served.

4.1.3 The structure and composition of the corpus

The mini-corpus for this study is made up of secondary school classroom dialogue of Xhosa mother-tongue speakers speaking English. For various reasons, (explored in section 4.2.2 below), I have concentrated my analysis on the speech of teachers in the classroom, although all English spoken in the classroom by Xhosa mother-tongue speakers is, in principle, relevant to this study. There are various reasons for my decision to focus on the spoken rather than written language of teachers. An important consideration is the

view that teachers are both "models of English usage" and "gatekeepers who determine the standards of education and language usage" (Van der Walt and Van Rooy 2002:115). As such, a study of the spoken language of teachers affords one a valuable glimpse into the real-life experience of learners at ex-DET schools in the Grahamstown area. Secondly, while studies of written BSAE doubtless produce patterns characteristic of the variety (see, for example, Gough 1996; Greenbaum and Mbali 2002; Mafisa 2002), it makes intuitive sense to consider that the written mode allows language users more time for reflection and editing; this is in contrast to the relative speed of spoken language. Furthermore, as Fromkin and Rodman (1993:379) explain, written language is likely to be more formal and "conservative" than spoken language and language users are apt to be more concerned with following the "prescriptive" rules of language when writing than in speech. As such, spoken data are likely to be less edited than written data would be. An obvious reason, then, is that spoken language is "primary" (De Klerk 2002:22) and, as such, reflects language use in everyday life (Isre'el 2002) and is "the most natural" and most "dominant" form of language used by any individual (Svartvik and Quirk 1980:9 in Peppé 1995:187).

The focus of this corpus on the spoken English of Xhosa mother-tongue teachers allows, ultimately, for a genre-based approach (McCarthy 1998:8) to the corpus, by focusing not only on a targeted population of speakers, but also on a particular environment and context. This also avoids the earlier tendency of gathering vast quantities of "broadcast" speech which, although more easily collected, are arguably not representative of the target population and may also have been scripted and edited by mother-tongue speakers - losing all but the phonetic characteristics of the variety (McCarthy 1998). Subsequent data collection could focus on other genres such as business encounters, professional discourse, etc., eventually allowing for clear genre-based descriptions (Atkins, Clear, and Ostler 1992).

4.2 DATA COLLECTION

4.2.1 Participants

To obtain sufficient spoken text to make up a mini-corpus of approximately 20 000 words, I tape-recorded classroom dialogues at three of the six ex-DET schools in the Grahamstown area of the Eastern Cape. The five participants whose language is used for analysis are, in all cases, speakers of Xhosa as a mother-tongue, or first language. All of them were born and raised in the Eastern Cape - two in Grahamstown itself. Four of the participants are female and one, male. While this means that my sample does not represent males and females equally, it does seem to be in line with teacher demographics at the schools which I visited. In a small pilot study of this nature, is would be impossible to attempt to obtain data from every Xhosa mother-tongue teacher in the Grahamstown area. While a larger group of participants may, possibly, have given a more reliable reflection of the language of these teachers, it may also have resulted in more superficial data. I have not aimed for statistical representivity in this study, but have chosen, instead, to look at the language of fewer teachers in greater detail.

Importantly, in terms of both ethical considerations and for accuracy, permission to use the data, as well as relevant social and educational background information was obtained from each participant. As these data are to form part of De Klerk's (2002) corpus of spoken Xhosa English, the permission slips used are the same as those used by De Klerk (please see Appendix A for a copy of the permission slip used). Table 1, below, summarises the information provided by the participants on the permission slips. To maintain confidentiality, I have labelled each teacher with a two-letter code: M or F to mark the gender of the speaker and then A, B, C, or D as appropriate to differentiate them from each other. Abbreviations used in Table 1, are as follows:

BORN Place of birth. These include the Eastern Cape (E.C.), Grahamstown (GHT), and King William's Town (KWT).

RAISED Place where participant was raised. These include those towns listed above as well as Uitenhage (UITEN.).

HOME Home Language.

OTHER Additional Language(s) in which the participant is fluent.

AGE-GRP Age-group of the participant.

EDU Highest educational qualification.

ENG SUB. The number of years in which the participant studied English as a subject.

ENG MOI Whether or not the participant was taught through the medium of English - here they could choose *yes*, *no* or *sometimes* (SOME).

ENG D. The amount of English which participants use on a daily basis; choices were: *None*, *A little*, *A lot*, or *All the time*.

SUBJ. The subject taught by the participant, including English (ENG), Business Economics (B ECON.) and Biology. (This information is not part of that obtained from the permission slip as it is not generally relevant to all participants of De Klerk's larger study).

	MA	FA	FB	FC	FD
BORN	GHT	E. C.	KWT	GHT	PEDDIE
RAISED	GHT	UITEN.	KWT	GHT	PEDDIE
HOME	XHOSA	XHOSA	XHOSA	XHOSA	XHOSA
OTHER	ENG	ENG, AFR	ENG	ENG	ENG, AFR, ZU
GENDER	MALE	FEMALE	FEMALE	FEMALE	FEMALE
AGE-GRP	30 - 34	35 - 39	25 - 29	30-34	25 - 29
EDU	BA ED	BA	STD[1]	HDE	DEGREE
ENG SUB.	16 yrs	1 yr	15 yrs	16 yrs	9 yrs
ENG MOI	YES	YES	YES	SOME	YES
ENG D.	A LOT	A LITTLE	A LOT	A LOT	A LOT
SUBJ.	ENG	B ECON.	ENG	B. ECON.	BIOLOGY

TABLE 4.1: BACKGROUND INFORMATION ABOUT PARTICIPANTS

[1] STD: Secondary Teacher's Diploma - A two-year teacher's training course for black teachers (Behr 1884:192).

4.2.2 Fieldwork

The tape-recordings of teacher-talk in classrooms were collected over a period of 15 months - from February 2001 until May 2002 - with most of the recordings being made in August and September 2001. During this time, I visited three ex-DET Senior Secondary Schools in the Grahamstown area (Please see Appendix B for more detail). To guarantee that I would be able to obtain a useful data set from five to six teachers, I sat in on the lessons of twelve different teachers at these schools. These teachers taught a range of subjects including Biology, Economics, English, Geography and Mathematics.

I was given permission to attend normal school lessons at each school. During the data-collection phase of my study, my main concern was to ensure that I collected the approximately 6-7 hours of useful talk required for a data-set of 20 000 words. As a safe-guard, I recorded about fifteen hours of classroom interaction. The content taught in the classes, although important to the analysis phase of this project, was relatively unimportant to me during my data-collection, as long as the teachers taught in English - the official medium of instruction for all content subjects at the schools. At each school, I sat in on the lessons and operated a small hand-held tape-recorder to record the classes. (I explain my reasons for this choice of recording equipment later in this section). I also made notes on my observations which helped enormously in the transcription and analysis of the data.

Although, as explained in section 4.1.3, above, all talk in the classroom from Xhosa mother-tongue speakers would be relevant to this study, I chose to focus on the speech of the teachers for several reasons. Firstly, as Mafisa (2002:16) argues, "the teacher of English serves as a model to his/her learners". This means that the competence of the teacher impacts directly and powerfully on the success of the learners (Strevens 1977 in Mafisa ibid.). Secondly, XE is likely to be far more entrenched in the teachers than in the pupils as the teachers have not only been using English for longer than their pupils have - the teachers are all over 25 years of age and were all taught, at least partially, through English - but they also all make use of, and model,

English in their daily lives. I also felt that the relatively small range of the tape-recorder would make it more practicable to focus on the teachers' language rather than attempting to move around the class with the tape-recorder.

Despite its limitations, I feel that the hand-held tape-recorder which I used for data-collection was appropriate to the context of this study. Although this equipment was not very sensitive, I nevertheless felt that it was less intrusive in the context than more invasive, technologically advanced equipment - such as a video camera, for example - would have been. The students were obviously intrigued by my presence and by the tape-recorder and, while the teachers did not appear as concerned, I felt that it was important to interfere with the normal group dynamics in the classrooms as little as possible. As with many fieldwork data collection techniques, I felt that I had to weigh up the disadvantages of limited recording equipment against the already significant observer's paradox of my standing out as a very white face in the group. There is some evidence in the data that my presence did have some influence on the group, but I feel that this impact was reduced by my making myself as inconspicuous as possible under the circumstances.

An obvious area of difficulty for this study, then, was that of my being an outsider to the group under investigation. Schmied (1996:186) warns against what he terms the "sociolinguistic paradox". By this he means that while corpus writers need to obtain natural speech in context, their presence, as outsiders in a speech community, is likely to shift natural speech towards more formal, prestigious forms of the language. De Kadt (1993:317), for example, points out the difficulty in finding natural ESL in South Africa, where it is often seen as "rude" to use English, which excludes elders in the group who do not have the necessary English language skills. This problem is likely to be of less significance in the English classroom context as the teachers are, at least in theory, behaving as they would normally do. In other words, while natural ESL data may be difficult to obtain since speakers would be more likely to "naturally" speak their mother tongues, in the classroom

situation it is natural for teachers to speak in English - even if they do code-switch as well.

In total, I observed about twenty-four lessons - lasting from 30 to 45 minutes each. From these recordings, I was able to select a core set of data for transcription and analysis. I based my data-selection on some key considerations. Firstly, I felt that it would be important to look at the language used by English teachers as this affords an opportunity to explore the explicit teaching that is taking place. For this reason, I chose two English teachers for my sample. Both MA and FB are English teachers: MA from a fairly large school with better facilities, and FB from a poorer, smaller school. I felt that it would be useful to include data from both of these teachers as their teaching contexts are quite different and I thought that it might be useful to compare their use and teaching of English. A second consideration in my selection of a data-set was that I wanted to look at the teaching of content subjects which were as diverse as possible. For this reason, I decided to look at a commerce subject - Business Economics - and a Science subject - Biology. I chose FA and FC as the two Business Economics teachers for my sample. As was the case with the English teachers, FA and FC also teach in very different teaching environments: FA at a reasonably well-equipped, fairly large school, and FC at the same poorer school at which FB teaches. Here, too, I thought that this could allow for interesting comparisons to be made. An additional reason for my including FC was that she code-switched a fair amount in her teaching (as I explain in the Findings and Discussion Chapter of this thesis). I thought that it would be interesting, then, to look briefly at code-switching as it occurs in the classroom while still choosing data with enough English usage to make it useful to my study. FD's inclusion is on the basis of her teaching Biology as well as the fact that she provided very "dense" data. What I mean by this is that FD spoke far more in her lessons than any of the other teachers that I observed. I thought, therefore, that it could be interesting to compare her English usage with that of the other teachers in my sample. Finally, practical considerations also played a role in my selection of data for transcription and analysis. Some of the recordings - where games were being played outside the classroom or where the teacher

walked around a lot during the lesson, for example - were quite indistinct. There were also a number of classes where very little teacher-talk took place such as when a teacher would set work or a test for the students to write, or where students spent most of a lesson copying from the board. These recordings, as well as those where teachers spoke mainly in isi-Xhosa, were not useful for my study and, as such, were not included. A summary of the data used in this study is presented in Table 2, below. (File numbers - labelled FILE # - are for reference purposes only and are based on the tapes used for recording).

FILE #	WORD COUNT	TEACHER	DATE	GRADE	CONTEXT
001	2 154	MA	01/08/01	12	English: poetry lesson
002	2 049	MA	01/08/01	12	English: poetry lesson
003	2 431	MA	01/08/01	11	English: lesson on "Applying for a job"
004	1 567	MA	11/09/01	11	English: test review
005	1 836	FA	12/09/01	11	Business Economics: public relations
006	1 563	FA	12/09/01	11	Business Economics: public relations
007	2 094	FB	12/09/01	10	English: dictionaries
008	1 654	FB	12/09/01	10	English: dictionaries
009	2 346	FB	12/09/01	10	English: dictionaries
010	755	FB	12/09/01	10	English: dictionaries
011	1 737	FC	11/09/01	10	Business Economics: functions
012	4 005	FD	12/09/01	11	Biology: plant cells

TABLE 4.2: SUMMARY OF RECORDED DATA USED IN THIS STUDY

By planning for the possibility that some of the data which I collected would prove unsuitable for my study, I feel that I have been able to accomplish what I set out to do. The entire mini-corpus consists of 24 191 words: 8201 from

MA, 3399 from FA, 6849 from FB, 1737 from FC, and 4005 from FD. Naturally, the number of words transcribed for each of my selected teachers could only be calculated at the end of the transcription process. While I did attempt to record equal amounts of teaching time from each teacher, this was not always possible. There is also a fair amount of variation in the number of English words spoken by any teacher during a lesson as can be seen by comparing one lesson from FC (at 1737 words) with one from FD (at 4005 words), for example. Despite this variation, the data obtained from my final sample group of five teachers do, I am sure, give some idea of XE as used in the ex-DET schools in the Grahamstown area. In addition, by focusing closely on the language of five teachers, I have been able to comment on specifics of the context and speakers to a greater degree than would be possible in larger corpora studies.

4.2.3 Preparation of data for analysis

My first step was to transcribe the recorded teacher-talk. Cook (1995), however, warns of the problems involved with the conversion of spoken text to a written form - as used in a corpus. The chief areas of difficulty are, firstly, that spoken language is usually context-embedded. A related second problem is that, because contextual elements are far less likely to be linguistically realised in spoken language, the possibility of variations in interpretation increases substantially (Cook 1995). I have tried to overcome both of these potential problems by limiting the context in which recordings took place to the classroom. In addition to this set context, I have tried to make contextual notes visible in the transcriptions to make the context clearer. This is where my fieldnotes were helpful, as I was reminded of unusual happenings or sounds which would otherwise be uninterpretable. Cook (1995) points to the uneconomical amount of time and effort required to convert spoken data into a written corpus as an additional disadvantage of transcription. I have found this to be true, but feel that direct engagement with the data has been the best way to become familiar with it.

The difficulties involved in converting spoken data to written text were discussed in section 2.4, above. In my data, to reflect the spoken data in written form and for ease of reference, I divided the transcribed teacher-talk into individual utterances by entering an explicit new line after each utterance. While the divisions are necessarily subjective, it is the only reasonable way in which the structure of the spoken data could be realised in the written form without losing all important information about the speech structure. The only non-subjective way in which this could have been done would have been to divide the data based on changes in speaker. This would, however, have made it virtually impossible to locate any extracts discussed, because, in the classroom context, teachers have long sections of monologue.

After the data were transcribed into a computer-readable form, it was marked up using tags approximately following normal HTML conventions (Vine 1999). These conventions, for the most part, follow those used in the Wellington Corpus of Spoken New Zealand English (WSC) which is, in turn, modelled on those used in the International Corpus of English (ICE). The mark-up of the data - made possible by a custom-written AWK program - consisted of numbering utterances for ease of reference (please see Appendix C for an illustration of the program's operation). Each utterance was given a code, for example: <XEC:#001:MA:001>. This example shows that the utterance is part of a Xhosa English corpus (XEC), is from the first file in my data (#001), is produced by participant MA and is the first utterance (001) in that file. File numberings are based purely on the tapes used for recordings and, again, allow for ease of reference. This mark-up allows for easy location of utterances within the data set.

Any notes or comments which I wanted to include in my transcriptions were written in angle brackets to allow them to be useful in looking at the data, while still being able to be ignored by Wordsmith Tools (as will be discussed in section 4.3, below). This included comments such as "<writing on board>" or "<interruption>" as well as any utterances in which the teacher spoke in isi-Xhosa (e.g. "<xhosa> <ne?> <xhosa>"). Where teachers used isi-Xhosa,

each Xhosa section was marked to reflect its length: short phrases or single utterances were marked as "<xhosa> <comment> <xhosa>", whereas sections thought to consist of two or more utterances were marked "<xhosa> <explanation> <xhosa>". While this approach is subjective, especially in view of the fact that I do not speak Xhosa, it enabled me to capture the relevant information without having to resort to timing individual Xhosa sections. Such a process would be highly error-prone owing to the relatively large number of short <comment>-type sections.

Finally, any of the students' utterances which were audible and, therefore, transcribed were marked with brace brackets. This is to distinguish them from my comments and, again, to allow them to be ignored by Wordsmith Tools in the analysis stage. The students' utterances - as with my observer's notes in angle brackets - give a clearer sense of context, but needed to be excluded from the data for analysis. The overall intention was that this mini-corpus could later form part of a larger corpus and be useful for comparison with similar data-bases. A sample of transcribed and marked up data from each teacher is available in Appendix D.

4.3 DATA ANALYSIS

Wordsmith Tools was used on the prepared data to identify and describe selected grammatical, lexical and discourse characteristics of Xhosa English (XE) evident in my data. The analysis was based on Gough's (1996) list of typically cited features of BSAE. While I have also drawn on more recent studies in my analysis (such as Van der Walt and Van Rooy 2002; De Klerk and Gough 2002; and De Klerk forthcoming), I chose to use Gough's (1996) list as a starting point as it is commonly used as a basis for descriptions of BSAE (as evident in the above sources, too). I feel, therefore, that Gough's (1996) description is a useful point of comparison which can fairly easily facilitate comparisons with other varieties of BSAE at a later stage.

I used a two-pronged approach in my data analysis - both using Wordsmith tools. On the one hand, I wanted to check for word-based patterns (as

explained below) which Gough (1996) and others have cited as typical features of BSAE. On the other hand, my analysis involved checking for broader grammatical and discourse patterns evident in my data. I will begin by explaining the more complex process that I used to identify word-based patterns, and then explain the steps I took to look for broader patterns in my data.

Patterns such as the use or absence of articles, the marking of count and non-count nouns, and concord-marking I have termed "word-based" patterns. This is because, by looking at specific word classes - nouns and verbs in the above examples - I was able to discover details about how these word classes are dealt with by speakers of XE in my sample. The first stage of this part of the analysis involved my using Wordsmith Tools to generate a list of all of the words in my mini-corpus ordered according to their frequency of occurrence. The result was a list of 1939 different words ranging in frequency from 1474 appearances ("the") to words which occurred only once in the data (746 words in my mini-corpus occur only once). I felt that dealing with the entire set of words for this section of the analysis would be cumbersome and impractical, but nevertheless wanted to have representative data. I decided, therefore to focus on words which occurred 10 or more times in the data. There were 352 such words in my frequency list. These 352 words were used in the word-based section of my analysis.

Once I had established which words to include in my word-based analysis, I set about making word lists to be searched by Wordsmith Tools' concordance. I divided the list of 352 words from the Wordsmith generated frequency list into word lists of articles, nouns, pronouns, prepositions, and verbs - repeating words which could conceivably fall into more than one category. The word "form", for example, was included in both the noun and the verb word-lists to ensure that no instances of the word were missed. I ran Wordsmith Tools' concordance on these word-lists to reveal the patterns of each occurrence of each of the words. When Wordsmith Tools' concordance was run on my noun word-list for example, 3727 occurrences of the words were revealed. These data, then, allowed me to look for the use and

absence of articles in noun phrases, the marking of count and non-count nouns, and noun phrases (un)marked for number. This process was particularly important in looking for the absence of features such as articles or number markings. Wordsmith Tools' concordance was then run on each of the word-lists, generating five files which could be more closely examined.

The files generated by Wordsmith Tools' concordance from the word-lists that I had made were then carefully checked for any "false finds". Occurrences of words, such as "form" described above, which had been entered on more than one word-list had to be deleted if inappropriate i.e. "form" as a noun had to be deleted from the verb list and vice versa. This meant that I had completed files consisting only of appropriate entries.

The fourth step in the process was to go through each utterance in each file - a total of 15 288 utterances at this pared-down stage - to look for "matches" for Gough's (1996) list of features. The files were then divided into sub-files of utterances which I felt matched Gough's (1996) classification. As Van Rooy (n.d.) points out, this process is made "a bit more tricky" by the fact that the data had not been marked for grammatical features. He also points out, though, that "there is a school of thought that holds that mark-up obscures important linguistic facts as much as it highlights others" (Van Rooy n.d.). Although analysis at this level was rather painstaking, then, I feel that the data obtained in this way is rich and detailed. Finally, patterns of frequency were revealed in the data by comparing the trends in the sub-files with those listed by Gough (1996). Not all of the grammatical features listed by Gough (1996) were explored, however, as is explained in Chapter 5.

The second way in which I analysed my data was by making more conventional use of Wordsmith Tools. Here, I searched for specific collocations in the entire mini-corpus. This was easily done when searching for specific patterns such as the use of the progressive, structures of comparison, or "can be able to" used as a modal verb. These data were much more easily found and analysed than those found by the method I described first in this section.

An important consideration in the classification of my data was the interpretation of grammatical patterns used within context. For example, MA devoted part of a lesson to a discussion of the extension of the progressive - a commonly cited feature of BSAE (Gough 1996). Wordsmith Tools, naturally, found a large occurrence in these data of this use of the progressive. A clear knowledge of the context in which these utterances occurred was necessary in order to interpret them appropriately.

4.4 LIMITATIONS OF THE STUDY

The small scope of this study is an obvious limitation as it precludes any broad generalisations about BSAE. A second limitation is that the teachers in my sample are not equally represented in the data, as I explained in section 4.2.2 above. I nevertheless feel that I have managed to achieve what I aimed to do. The mini-corpus compiled for this study consists of 24 191 words where I aimed for 20 000 words. I aimed to look at the English used by five or six Xhosa mother-tongue teachers in the Grahamstown area and have obtained useful data from five teachers specifically selected from a larger set as described, above. In this way, I feel that I have compiled as representative a mini-corpus as possible within the scope of a small-scale pilot study. I also believe that the small scope of this study, while a weakness in terms of generalisability, is also a strength in terms of the detail in which I have been able to analyse my data. This level of detail may not have been possible in a larger corpus study.

CHAPTER FIVE: FINDINGS AND DISCUSSION

In this chapter, I explore trends evident in my data. Again, emphasis is placed on the fact that these trends are based on a very small data set and, as such, conclusions drawn must be viewed tentatively at best. I do, however, feel that this pilot-study points to areas of interest which may encourage further research. I will begin by demonstrating why I have focused my analysis as I have done. I will then look at each area of focus in turn. Finally, I will briefly summarise the patterns that I have found in my data.

5.1 AREAS OF FOCUS IN MY ANALYSIS

As I mentioned in section 3.4 at the end of Chapter 3, I used the classification of Black South African English (BSAE) put forward by Gough (1996) and Gough and De Klerk (2002) as a guideline for my analysis. This allowed for a close analysis of the features evident in my data in terms of standardly agreed on characteristics of BSAE. However, I have not used all of Gough's subsections in my analysis. For example, while Gough's (1996) discussion of BSAE includes information on the phonology and stress and intonation patterns seen to be typical of BSAE, these areas are beyond the scope of this study. A highly detailed, phonetic transcription would be needed for such an analysis - making it far too labour-intensive for a study of this size. Instead, I have chosen to focus chiefly on the grammatical features evident in my data as well as the vocabulary used. I have also looked at discourse patterns in my data which, I argue, are specifically related to the classroom context. I have also drawn on various other authors in my discussion, notably Williams (1987), Van der Walt (2001), and Van der Walt and Van Rooy (2002). As discussed in Chapter 3 (section 3.2) of this thesis, Williams (1987) looks at the features of new Englishes (which she terms NIVEs), and Van der Walt (2001) and Van der Walt and Van Rooy (2002) look at aspects of BSAE both in relation to its status as a new English and also in terms of the teaching context.

Throughout this section, I will make use of tables to briefly summarise any evidence from the data in each subsection. In each table, utterances are referenced as described in section 4.2.3 above. These reference codes reveal each extract's exact location within the mini-corpus. Most importantly for my discussion, the codes also reveal the speaker in each case. Wherever possible, I have given at least one example from each teacher in the summarising table. Since much of my discussion is based on the number of utterances in which features are found, I feel that it is important to present a summary of the number of utterances per teacher, here. As code-switching is also briefly discussed in my findings, the number of utterances in which Xhosa is used is also included in this table. As described in Chapter 4 (section 4.2.3), Xhosa utterances were marked with "<comment>" or "<explanation>" depending on their length - the table reflects this distinction.

TEACHER	TOTAL WORDS	TOTAL UTTERANCES	XHOSA UTTERANCES	
			<comment>	<explanation>
MA	8 201	1172	9	4
FA	3 399	518	4	1
FB	6 849	757	27	0
FC	1 737	207	22	41
FD	4 005	330	0	4

TABLE 5.1: SUMMARY OF NUMBER OF WORDS AND UTTERANCES OF PARTICIPANTS

5.2 FINDINGS

In the Introductory Chapter of this thesis, I pointed out that my aim in this study is to describe rather than explain the language patterns which I find in my data. Despite this, I have ventured possible explanations for my observations in many cases, as I feel that this adds a richness to my analysis.

As discussed in the Methodology Chapter (section 4.3), some of these findings are based on word-lists generated from the most frequently used 352

words in my corpus, while others come out of a broader concordance run on the whole data set. Since the relatively small size of my mini-corpus prevents me from making any broad generalisations about my data, I have taken the liberty of commenting on the teachers individually in many cases during my analysis. This is because I feel that the language use of individuals would not, in general, fall under the spotlight but can, nevertheless, reveal some interesting insights into why certain trends were evident. I think that this more in-depth awareness of the speakers would not be possible in more typical, larger studies. As such, I think that it is important to take advantage of the small sample size to explore issues which would otherwise not be discussed.

5.2.1 Grammatical Features

(i) Non-count and count nouns

Leech, Deuchar and Hoogenraad (1982) describe count nouns as those which are able to be counted and, as such, have a plural form. Mass nouns, in contrast, refer to qualities or substances which cannot be counted and, thus, usually do not have a plural form (ibid.). In addition, Leech et al. (1982:45) point out that the indefinite article is "a marker of count nouns". Simply put, a count noun is one which can take a plural marker, whereas a non-count or mass noun cannot (Finegan and Besnier 1989).

Gough (1996) and De Klerk and Gough (2002) show that the treatment of count and non-count (mass) nouns in BSAE is different from that in Standard English (SE). In his attitude study, Van der Walt (2001) found that over 65% of his informants accepted non-count nouns marked as count nouns. Williams (1987) shows that a loss of distinction between count and non-count nouns is a common feature of many new Englishes, and argues that this may be a result of regularisation - where the linguistic system exhibits fewer exceptions to general rules (Long 1982 in Williams 1987).

My data set of 3298 noun occurrences revealed only 5 examples of non-standard marking of count nouns. This is, obviously, a very low incidence

(about 0.15% of the data). These examples are presented in Table 5.2, below. I did, however, find 31 instances of mass nouns marked as count nouns. Some examples of these are presented in Table 5.3.

	EXTRACT	REFERENCE
1	another fears?	<XEC:#003:MA:354>
2	yes but he he still experiences fears	<XEC:#003:MA:409>
3	name the type of tissues we discussed yesterday	<XEC:#012:FD:009>
4	any type of tissues we discussed yesterday	<XEC:#012:FD:010>
5	yesterday we managed to talk about three kind of tissues	<XEC:#012:FD:010>

TABLE 5.2: UNUSUAL MARKING OF COUNT NOUNS

The non-standard marking of count nouns as shown here, was displayed by only two of the five teachers. It is worth considering that there is a high level of consistency in these few examples. What I mean by this is that both of MA's utterances centre around the word *fear* and similarly, the examples from FD's utterances are all of the "type of X" structure - where X is a count noun. (Interestingly, FD's utterances marked the non-count noun as count in each case. In other words, "types of tissue" is rendered as *type of tissues*). This seems to show that, although these two teachers have used a feature of BSAE, they have each used it in only one context and with very specific nouns. Non-standard marking of count nouns is thus not highly spread in my data, although it is a feature commonly associated with BSAE.

This picture changes slightly when one looks at the marking of mass (i.e. non-count) nouns as count nouns. Although this feature is not commonly found in my data (0.9%), it still occurs more often than the marking of count nouns as non-count nouns. A sample of the 31 occurrences of this feature is given in Table 5.3, below.

RHODES UNIVERSITY LIBRARY

	EXTRACT	REFERENCE
1	from the mother nature	<XEC:#003:MA:229>
2	but once you've a good will it means you have their trust	<XEC:#006:FA:001>
3	so next to abortion there is a information about abortion	<XEC:#008:FB:064>
4	so I have got the head word stompie next to it there is an information about the word stompie	<XEC:#009:FB:377>
5	as to what is a data and what is a an information	<XEC:#011:FC:047>
6	do you see a difference between a data and information?	<XEC:#011:FC:081>
7	and there is a lignin or cellulose a lot of cellulose or a lot of lignin at the corners of the sclerenchyma tissue	<XEC:#012:FD:400>

TABLE 5.3: NON-COUNT NOUNS MARKED AS COUNT NOUNS

Unlike the data summarised in Table 5.2, here we see that all five teachers mark non-count nouns as count nouns at least once. MA does so twice, FA once, FB three times and FD twice. FC, however, produces 24 examples of this feature - using *a data* 16 times and *a(n) information* seven times. This particular teacher, FC, made use of the collocates *data* and *information* many times during her teaching, but was inconsistent in her marking of count as non-count nouns. Interestingly, FC marks mass and count nouns in a standard way 34 times, again with the words *data*, (29 times), and *information* (five times). Examples 5 and 6 shown in Table 5.3, above, demonstrate FC's inconsistent marking of mass and count nouns. This may signal an unawareness of this issue, although Van der Walt and Van Rooy's (2002) argument that BSAE is in the liberation phase of developing into a new English seems to make more sense of this issue. While this aspect of count as mass noun marking was far more common than the unusual marking of count nouns, it is nevertheless still a small percentage of the total data. Although there is evidence for non-standard markings of count and non-count

nouns, then, my data does not reveal that this is a very widely occurring phenomenon.

(ii) Articles

The use of articles in standard English (SE) clearly distinguishes between those nouns that are "agreed upon by speaker and listener" - where the definite article, *the*, is used - and those where no particular referent is implied - where the indefinite article, *a/an*, is used (Fromkin and Rodman 1993:156-7). This means that, for SE speakers, the sentence, "I saw the boy", does not have the same meaning as "I saw a boy" (ibid.). Similarly, as mentioned above, Leech et al. (1982:45) point out that the indefinite article is "a marker of count nouns". As such, a study of the use (or absence) of articles in my data revealed some interesting trends. For example, evidence for mass nouns being marked as count nouns, discussed in section (i), above, comes out of a study of the use of articles in my data. I made use of my generated list of the most frequently used nouns in my data to explore the use (or absence) of articles in my data.

Williams (1987 after Platt et al. 1984) argues that new Englishes often adapt the article system of native speaker (NS) English to reflect a specific / non-specific distinction, rather than the definite / indefinite distinction used by NSs. Van der Walt and Van Rooy (2002:120) found that teachers in their study accepted both the insertion and deletion of articles - showing a high level of uncertainty about the norm. De Klerk and Gough (2002) point out that typically cited features of BSAE include the absence of articles where speakers of (SE) would have used an article. In my data, I found 35 such cases - again sourced from a concordance of 3298 utterances containing nouns. Not all of these cases, however, were entirely clear-cut. I find it interesting that an Economics teacher (FA) did not display this feature typical of BSAE, while the English teachers in the sample (MA and FB) omitted articles 14 and 4 times, respectively. FC omitted 5 articles where speakers of SE would not have done so. This feature was found 12 times in FD's speech, although many of these are not absolute cases, as I discuss below. Table

5.4, shows example utterances from each of the teachers. I have used a ^ symbol (after Gough 1996:61) where I feel that SE would make use of an article.

	EXTRACT	REFERENCE
1	usually the soldiers of Umkhonto we Sizwe the military wing of the ANCs were fighting with ^ then South African Defence Force	<XEC:#002:MA:110>
2	by what would a small child even ^ unborn child be afraid of being cursed?	<XEC:#003:MA:384>
3	you just fill in ^ pronoun	<XEC:#004:MA:266>
4	so this n is ^ abbreviation of noun	<XEC:#008:FB:083>
5	and the mere fact that each and every one has got a dictionary it means that ^ dictionary is a useful tool is it not so?	<XEC:#009:FB:012>
6	when we talk of ^ part of speech what are we talking about?	<XEC:#010:FB:057>
7	okay so now we are busy with ^ administrative function	<XEC:#011:FC:020>
8	that is ^ internal source of data that means that information that you can be able to obtain it inside the business	<XEC:#011:FC:091>
9	those tissues are said to be permanent tissues and in the case of permanent tissues we have ^ epidermis we have i-sclerenchyma the collenchyma the parenchyma and the chlorenchyma the xylem and the phloem	<XEC:#012:FD:141>
10	there is a tissue in the plant which is able to transport i-water molecules from ^ roots to ^ leaves that tissue is known as the xylem tissue	<XEC:#012:FD:203>
11	but ^ phloem tissue conduct those food substances from one organ to another organ	<XEC:#012:FD:303>

TABLE 5.4: ABSENT ARTICLES

I feel that examples 1, 2, 4, 6, 7, 8, and 9 are unambiguous cases of omitted articles. A clearer context disambiguates examples 3 and 5: in 3, MA is talking about pronouns generally, rather than simply using the word *pronoun*; and in 5, FB's intonation indicates that she uses the *that* immediately preceding *dictionary* as a complementizer (Radford 1997), rather than an article. Both 10 and 11 could, perhaps, be classed as standard usages of absent articles. I do feel, however, that speakers of SE would be inclined to use articles in both of these cases. Further evidence for this comes from the way in which FD deals with the words *leaves*, *roots*, and *phloem tissue* in the rest of my data. She uses the word *leaves* 19 times in total: once without an article; thirteen times following *the* and five times preceded by the prefix *i-* (please see my discussion of the use of the *i*-prefix, below). Similarly, of the 13 times FD uses the word *roots*, she only leaves out the article 3 times. FD's use of *phloem tissue* - 15 times in total - is also interesting. She says *the phloem tissue* 10 times, *i-phloem tissue* and *a phloem tissue* once each, and only omits an article 3 times. This seems to suggest that FD is more likely to use an article than not in these cases. This inconsistency ties in with Van der Walt and Van Rooy's (2002) argument that South African English is in the process of becoming a new English.

I began section (ii), by saying that there was much to be gained from exploring the use or absence of articles in my data. Even greater insight into this structure may be gained by exploring another intriguing observation from my data. Although arguably of lexical rather than grammatical significance, I would like to discuss the use of the nominal prefixes *i-* and *u-* here, as I feel that this ties in closely with article use. The use of the prefixes *i-* and *u-* has, to my knowledge, not been addressed outside the realm of code-switching. I found in my data, however, that there was a fairly large occurrence of both of these prefixes and of *i-* in particular. A concordance of my entire mini-corpus revealed a total of 140 instances of the prefix *i-* and 21 of *u-*. Why I mention this here is that the use of these prefixes seems to be strongly related to article use in SE. A comparison made between the examples given in Table 5.4, above, and those shown in Table 5.5, below, for instance, illustrates this point. For this illustration, I will use the notation 5.4:6 to refer to Table 5.4,

example six. A comparison of 5.4:6 and 5.5:6 (both FB), then, reveals that where there is a missing article in 5.4:6 (*when we talk of ^ part of speech*), the prefix *i-* is used in 5.5:6 (*when you talk of i-part of speech*). Similarly, contrasting 5.4:11 with 5.5:14 (both FD) shows that *but ^ phloem tissue conduct* can also be realised as *so i-phloem tissue it conduct.* (One example in my data, however, does not fit the pattern neatly: *so the i-xylem tissue it helps* (<XEC:#012:FD:210>).

It appears, then, that the *i-* prefix acts as a pseudo-article in the majority of instances where it occurs. This could be an interesting aspect to investigate in other, larger, BSAE corpora. It seems intuitively likely that these prefixes are transferred from isi-Xhosa. The noun class system in isi-Xhosa includes a class of borrowed words - class 9 - with the prefix *i-* and a class of nouns as derivations of verbs - class 11 - with the prefix *u-* (Bantele, Davey, Lanham, Mahlasela, Mathiso and Riordan 1969:169). It would be particularly useful if mother-tongue speakers of the BSAE varieties in question were to explore the use of these prefixes as they would doubtless see far more in the data than I could hope to find. Table 5.5, below shows some examples - from each of the five teachers this time - of the use of the pseudo-article prefix *i-*. Table 5.6, which looks at some examples of the use of the *u-* prefix, follows a discussion of Table 5.5.

	EXTRACT	REFERENCE
1	you you realise that that what all all these things mentioned here are what we call i-childhood fears	<XEC:#002:MA:043>
2	i-businesses what do they do for our community?	<XEC:#005:FA:262>
3	when we say i-firm has got a favourable image how does it look like?	<XEC:#005:FA:321>
4	all right those are the things that we think about when we talk about i-labour	<XEC:#006:FA:092>

5	so i-labour <click> i mean i- i-public relations function it is also responsible now for things that the business is able to solve nè?	<XEC:#006:FA:094>
6	what is it that you are talking about when you talk of i-part of speech parts of the speech?	<XEC:#009:FB:065>
7	secondly what else do we use i-dictionary for?	<XEC:#010:FB:033>
8	okay again we look for i-abbreviation	<XEC:#010:FB:039>
9	okay again we use the dictionary for i-pronunciation	<XEC:#010:FB:084>
10	<xhosa> <comment> <xhosa> i- i- i- i- i- i-data eh represents a collect collection of facts or figures etcetera that means it can be described as a raw data	<XEC:#011:FC:186>
11	whereas if they want you to get i-information as to as to what did the reaction of our consumers towards their product that means that information can be obtained from outside the business	<XEC:#011:FC:228>
12	about i-information about i-aids <xhosa> <explanation> <xhosa> nè?	<XEC:#011:FC:244>
13	sclerenchyma tissue it does not allow i-water molecules to pass through	<XEC:#012:FD:113>
14	so i-phloem tissue it conduct the food substances like sugars starch and other substances	<XEC:#012:FD:301>
15	i-sieve tube cells as you will notice in your drawing that in the case of the sieve tube cell there is no nucleus because at maturity the sieve tube cell they have a tendency to lose their nuclei but the companion cells they keep their nucleus up until maturity	<XEC:#012:FD:320>

16	during photosynthesis i-leaves they make starch they make sugar and other food substances	<XEC:#012:FD:346>

TABLE 5.5: THE USE OF THE PREFIX *i-* INSTEAD OF AN ARTICLE

Although the example from MA (1 in total) is not ideal, the others, I feel, do illustrate my point quite well. MA is an English teacher so, while we see some examples of missing articles from him (14), it may be that he is more careful in his modelling of English. The use of the prefix *i-* is, I think, more "obvious" a departure from SE than a missing article. This argument is refuted, though, by strong counter-evidence from the second English teacher in the group. FB uses *i-* a total of 56 times and 18 of the twenty-one uses of the prefix *u-* are made by this teacher. This may be because FB code-switched a fair amount in her speech, but only for very short comments. A second possible reason for this disparity could be the different levels of formality evident in the lessons of these two teachers. FB tended to be more "informal" in her dealings with the pupils than MA was, possibly due - at least in part - to the different sizes of the classes. In general FB's classes consisted of about 15 - 20 learners whereas MA's were far larger. MA teaches at a far larger school, too.

Both FA and FC make only limited use of the *i-* prefix, (16 times each). Furthermore, they do not attach the prefix to a large range of nouns, but use only 10 between them - examples of the most common usages are given in Table 5.5. I find this interesting as, while FA uses relatively little code-switching, FC code-switches far more than the other teachers in my sample do. FD - a biology teacher - makes frequent use of the *i-* prefix which she uses 51 times on a range of nouns. She does code-switch a little, but not to the extent that FB and FC do. Intuitively, it would seem to make sense that speakers who code-switch more are more likely to transfer features of their mother tongue across to the "matrix language" (Myers-Scotton in Eastman 1992:20). As such, I would have expected to find that the highest proportion of *i-* prefixes could be found in the data from FB and FC. Contrary to this expectation, the teacher who code-switches most, FC, makes very little use of

this mother-tongue feature. It may be, then, that the use of the pseudo-article prefix *i-* is not proportional to the extent to which the mother-tongue of the speaker is used. MA, however, code-switches very little and also makes negligible use of the *i-* prefix, which supports my initial hypothesis. My data set is naturally too small to make any judgements on this issue, but I think that it would be an interesting area to explore further.

While it seems that *i-* is used as a prefix where NSs would commonly use an article, *u-*, it seems, is used where an article would not be used in SE. As I have said, the *u-* prefix appears 21 times in the entire mini-corpus. MA uses it once, FB 18 times, and FC twice. The examples given in Table 5.6, below, show that, in each case, *u-* is used where NSs would be unlikely to use an article.

	EXTRACT	REFERENCE
1	change u-growl first	<XEC:#004:MA:092>
2	I want you to look at the pronunciation of u-refuse	<XEC:#007:FB:219>
3	this is how to pronounce u-abortion you see?	<XEC:#008:FB:096>
4	meaning that when you you pronounce u-refuse you will pronounce it as refuse	<XEC:#008:FB:174>
5	then <cough> u-word processing <cough> sorry word processing that means it can be processed processing of text in readable format	<XEC:#011:FC:256>

TABLE 5.6: USE OF THE *U-* PREFIX

(iii) Resumptive pronouns

Another characteristic of BSAE which Gough (1996:61) describes as "typically cited" is the extensive use of resumptive pronouns. Other authors have called this phenomenon "pronoun copying" (Wade 1998) or "double subjects" (Bamiro 1995). Van der Walt and Van Rooy (2002:120) call this

feature "one of the most prototypical features of BSAE, and probably one of which teachers are more consciously aware". This feature was rejected by 70% of the teachers in their study, an indication, they argue, that "general acceptance of BSAE as a standard has not taken place" (ibid.). Wade (1998:6) suggests that pronoun copying in BSAE may function as "a relatively unmarked way" of showing a referent change - especially in light of the relatively minor role played by intonation in the discourse of BSAE speakers (Lanham 1984 in Wade 1998). Williams (1987), however, argues that pronoun copies increase the salience of the subject when they are used in non-relative clauses such as those cases found in my data. I found 108 examples of this phenomenon in my data, (from the set of 2170 utterances with pronouns, i.e. almost 5% of the pronouns in my sample were used in this way). Examples from this data set are given in Table 5.7, below.

	EXTRACT	REFERENCE
1	some of you you are afraid of cats maybe or dogs	<XEC:#003:MA:046>
2	the child he is asking forgive-forgiveness for the lies that she gonna tell when she is brought into this world	<XEC:#003:MA:299>
3	those businesses they do something you know other firms like Rhodes they pay for your things	<XEC:#005:FA:288>
4	because the workers they do have their own misconceptions and transfer those misconceptions to in the public and people will have another attitude about your business	<XEC:#006:FA:165>
5	the head word it is the word pretend abortion is the head word	<XEC:#009:FB:172>
6	those words they are highlighted they are written	<XEC:#010:FB:137>
7	a data it is a raw information	<XEC:#011:FC:048>

8	eh data processing it's whereby a data it will be it's whereby the raw data will be processed that means it's it's going to be a a combination of isoneumerical eh and fixed data	<XEC:#011:FC:241>
9	a pitted vessel it is formed by those thickened cell walls forming pits	<XEC:#012:FD:279>
10	companion cells and also the sieve tube cells they form a phloem tissue	<XEC:#012:FD:299>

TABLE 5.7: RESUMPTIVE PRONOUNS

Only four of the teachers in my sample display this feature, although some used it far more than other did: MA used pronoun copying eight times; FA did not use this pattern at all; FB copied pronouns 30 times, FC 14 times, and FD a total of 56 times. The fairly wide-spread occurrence of this feature in my data seems to tie in with the literature, but the incidence in this data set is not nearly as high as reported (e.g. Van der Walt and Van Rooy 2002; De Klerk forthcoming).

(iv) Gender

Gender conflation is another commonly cited feature of BSAE (Gough 1996:61) and is often thought to result from "the lack of parallel gender markings in African languages" (De Klerk forthcoming). Van der Walt (2001:5) found that gender conflation in pronouns was accepted by 50 - 64% of the participants in his study. My data, however, seems, in general, to contradict this view. I found 5 examples which could be viewed as gender conflations, but closer inspection gives a different impression as discussed below. Table 5.8, below, gives all of the examples found in my data:

	EXTRACT	REFERENCE
1	what are his fears there? or her fears? let us suppose that he is a he or that he is she	<XEC:#002:MA:062> <XEC:#002:MA:063> <XEC:#002:MA:064> <XEC:#002:MA:065>
2	even though he is not yet born or she is not yet born it is afraid of being tortured	<XEC:#002:MA:136>
3	it has not yet learn how to pray	<XEC:#003:MA:023>
4	what does it hope for?	<XEC:#003:MA:217>
5	then what do I mean when I say he's so she's so sweet?	<XEC:#007:FB:274>

TABLE 5.8: GENDER "CONFLATION"

The first four examples in this table are all drawn from the same context: the teacher here is leading the class in a discussion of Louis Macneice's poem "Prayer before Birth". As the child's gender is not made clear in the poem, MA consciously makes use of both *he* and *she* throughout. For this reason, extracts 1 and 2 show MA alternating between the two. Examples 3 and 4 are included for their use of *it* to refer to the unborn child (something seen in example 2, too). Their inclusion is somewhat debatable and may well be pushing the boundaries a little. The fifth example in Table 5.8 (5) does, at first glance, suggest gender conflation in that the teacher was talking about a female pupil at the time and used the pronoun *he*. However, as is clearly shown in this extract, the teacher immediately realises her mistake and self-corrects. While this extract could be viewed as a prototypical example of gender conflation, I feel that this is more of a "slip of the tongue" type error than a feature of FB's speech. De Klerk (forthcoming) suggests that a frequent switching of pronouns in speech may indicate a high level of uncertainty as to the "correct" pronoun required in the situation. None of the occurrences found in my data seem to fit this pattern. It is interesting, then, how few examples there are of this "typical" feature in my data. It seems, in fact, that there is a conscious effort made to speak in a gender-neutral way as evinced by the 54 examples of *he* and *she* or *his / him* and *her* occurring

together in the data. There is also an example of *son* and *daughter* occurring together for the same reason (<XEC:#001:MA:061>). It must be borne in mind, though, that 48 of the 54 examples are from MA - 44 from his discussion of Louis Macniece's poem - and only 4 from another teacher (FC). Some of these examples are provided in Table 5.9, below.

	EXTRACT	REFERENCE
1	do not write any people's name with- without having asked his or her permission	<XEC:#001:MA:319>
2	he always or she always gives that she is not yet born	<XEC:#002:MA:009>
3	let not the blood-sucking bat the rat stoat club-footed ghoul come near him or her	<XEC:#003:MA:046>
4	what are his or her fears?	<XEC:#003:MA:074>
5	what kind of sin will he or she commit?	<XEC:#003:MA:263>
6	when he or she goes to interview a person maybe for instance ah Mandela	<XEC:#011:FC:052>
7	eh the journalist does not write it in full but sh- she or he uses what is known as shorthand	<XEC:#011:FC:054>

TABLE 5.9: GENDER-SENSITIVE PRONOUN USE

These examples show an awareness of the gender of pronouns rather than the confusion often wrought by the gender of pronouns to speakers of BSAE. It is important to note, here, that the context in which the data were obtained probably played a role in the limited data available for this feature. It seems natural to assume, for example, that everyday conversations would yield far richer data in terms of gendered pronouns than a Biology or Economics lesson would. At this stage, my data seems to contradict the generally held view that BSAE is characterised by gender conflation in pronouns, but the sample size is too small for any definite conclusions to be drawn.

(v) Number

The idea that BSAE commonly exhibits nouns not always marked for number (Gough 1996:61) is supported 47 times in my data. While this evidence does come out of a corpus of 3298 entries, and thus occurs only about 1.4% of the time, it does not appear to be merely the result of "error". Instead, I am reminded of Van der Walt's (1999) argument - as discussed in the Black South African English Chapter of this thesis - that one needs to consider the status of so-called "errors" when they are widely used in a variety. The evidence from my data seems to suggest that where plurality is clearly marked once in an utterance, the plural marker is not repeated on the noun. This makes sense in terms of "the selective production of redundant markers" common in new Englishes (Williams 1987:174). In other words, where the grammatical information of plurality is conveyed in some other way, the noun itself does not receive an "additional" marking. A few examples of nouns not marked for number are given in Table 5.10, below.

	EXTRACT	REFERENCE
1	they might they might have advertised many posts there or many job	<XEC:#001:MA:402>
2	so these are the fear in stanza two	<XEC:#002:MA:186>
3	like i said that these word they are two	<XEC:#007:FB:138>
4	what what are the other thing in a dictionary?	<XEC:#009:FB:043>
5	some of the meaning can be figurative some can be literal	<XEC:#009:FB:350>
6	so those xylem vessel they overlap each other	<XEC:#012:FD:235>
7	i-sieve tube cells as you will notice in your drawing that in the case of the sieve tube cell there is no nucleus because at maturity the sieve tube cell they have a tendency to lose their nuclei but the companion cells they keep their nucleus up until maturity	<XEC:#012:FD:320>

TABLE 10: PLURAL NOUNS NOT MARKED FOR NUMBER

Plurality in the example extracts above is shown in (at least) one of three ways: by the plural demonstrative used before the noun (1, 3, 5, and 6); by the plural verb form used (2, 3, and 4); or by the plural form of the copy pronoun used in the utterance (3, 6, and 7). As such, the plural marker on the nouns could be seen as "unnecessary" duplication. Only MA, FB and FD exhibited this feature in their speech: MA and FB eight times each, and FD 31 times in total. As such, this feature is not at all widely spread in my data and, in fact, seems to be typical of a particular teacher rather than the group as a whole.

(vi) Progressive

Another typically cited grammatical feature of BSAE (Gough 1996; De Klerk and Gough 2002) is the extension of the progressive tense to verbs not expressing an action in progress. Van der Walt (2001:5) found that between 65% and 74% of the subjects in his study accepted sentences with this feature. Wade (1998 after Platt et al. 1984) suggests that the extension of the progressive tense may result from an over-emphasis of this verb form in English teaching. In contrast, Williams (1987) suggests that extension of the progressive is part of the regularisation process which takes place in new Englishes. In all, my data revealed 46 examples of this feature. Since a total of 4252 verbs were used, this means that only about 1% of the verbs in my data set were in the progressive form in an extended way. Some example extracts are presented in Table 5.11, below.

	EXTRACT	REFERENCE
1	what I said here is that due to should be requiring a subject a noun	<XEC:#004:MA:212>
2	at the beginning of a sentence I would prefer using owing to rather than due to	<XEC:#004:MA:213>
3	you know that in our businesses we are using those types of?	<XEC:#005:FA:182>

4	every day we are complaining about that they come down and sit down	<XEC:#006:FA:046>
5	which one do you think is having a literal meaning from the two sentences?	<XEC:#009:FB:321>
6	it does not matter even if you are six in a group or five or what as long as i will be having four groups in this class	<XEC:#010:FB:007>
7	if I am saying they are not living cells that means those cells are dead cells this xylem vessel and the tracheid they are living cells	<XEC:#012:FD:286>

TABLE 5.11: EXTENDED USE OF PROGRESSIVE VERB FORMS

Not all of the teachers in my sample displayed this typical feature of BSAE. Of those teachers who did display this feature, MA used it 19 times, FA 7 times, FB 18 times and FD only twice. The low occurrence of this feature in my data, too, seems to contradict the commonly reported findings. Once more, I think that it can be argued that the teachers are "model" language users which means that they are likely to be higher on the English proficiency continuum than the average speaker of BSAE. Nevertheless, this feature does occur and it is fairly wide-spread in the data.

(vii) Concord

De Klerk and Gough (2002) give the simplification of verbal concord as another grammatical feature commonly found in BSAE. Van der Walt and Van Rooy (2002) report that, while the teachers in their study accepted almost all the standard English forms of verbal concord, they also accepted a large percentage of non-standard forms. They state that this shows that the teachers in their study still see SE as a norm, but that confusion may arise in learners out of the teachers' inconsistent treatment of verbal concord. I found 12 examples of the simplification of verbal concord in my data - a mere 0.3% of the 4252 verbs used. Some examples of this feature are shown in Table 5.12, below.

	EXTRACT	REFERENCE
1	but his hope ironically does not come from the human beings it come from mother nature	<XEC:#002:MA:257> <XEC:#002:MA:258>
2	so that is why we say because of the growth in the power of of of public opinion that have led now to the importance of the public relations	<XEC:#005:FA:108>
3	it does only those things that the government say or it does them up to a limit according to the?	<XEC:#006:FA:274>
4	I have not I I have not seen their hands up mean that they don't know	<XEC:#009:FB:088>
5	what did we said about the function that we are involved with?	<XEC:#011:FC:028>
6	then it mean that we can proceed	<XEC:#011:FC:233>
7	and also the tracheids is like that because of those tapering ends they are able to overlap each other and water molecules can be conducted from this vessel to another vessel until they reach the top of the tree	<XEC:#012:FD:236>
8	but phloem tissue conduct those food substances from one organ to another organ	<XEC:#012:FD:303>

TABLE 5.12: SIMPLIFICATION OF VERBAL CONCORD

In total, MA and FA displayed this feature twice each, FB and FC, once each and FD six times. Williams (1987) says that, while simplification of the verbal concord may be seen as minimising redundancy, it actually makes the listening process more complex. She suggests, therefore (after Slobin 1977), that this feature may, instead, be a result of phonological simplification, where final consonants and consonant clusters are deleted.

(viii) Tense simplification

Gough (1996) and De Klerk and Gough (2002) list tense simplification and the frequent lack of past tense markers as separate characteristics of BSAE. As both of these factors imply a simplification of the verb, however, I have combined them into a single section. My data revealed only 1 clear example of tense simplification (from MA). Table 5.13 gives this example. The table also gives all 5 occurrences of absent past tense markers found in my data: 3 from MA and 2 from FB. Two utterances - one each from MA and FA - fall, potentially, into both categories and are also included in the tense simplification table, below. In this table I have added, in square brackets, the parts of the verb phrase which I believe are missing and marked the simplified modal with a trailing [M]; verbs which have not been marked for tense are underlined in the table. In total, tense simplification applied to less than 0.2% of the verbs in the data set. As discussed in the section on the gender of pronouns (iv) above, I think that tense simplification may well be far more evident in "normal" conversations than in the classroom situation, as much of the teacher-talk relates to the present.

	EXTRACT	REFERENCE
1	so if you made me as one of your references I can I can be telephoned at any time and be asked if I know you well	<XEC:#001:MA:314>
2	even you when you were still children you are afraid to to go into deserted house empty houses when people no longer live there	<XEC:#002:MA:177>
3	it has not yet learn how to how to pray	<XEC:#003:MA:023>
4	[do] not [be] afraid of her she is my visitor	<XEC:#003:MA:027>
5	to fail to take take heed of the advice that he might [have] be[en] given by old men by experienced people	<XEC:#003:MA:371>
6	is it the first time that you [have] come across that word?	<XEC:#005:FA:151>

7	so that's how abortion is pronounce	<XEC:#007:FB:215>
8	is there anyone who has not yet understand what I've been saying	<XEC:#008:FB:252>

TABLE 5.13: TENSE SIMPLIFICATION

The modal verb in extract 1 has not been inflected to show the past tense evident in the rest of the utterance. Extracts 2, 3, 7, and 8 are clear examples of absent past tense markers - again evident from the context in each case. The fourth example (4) shows a simplified verb form. Extracts 5 and 6 illustrate a combination of a lack of past tense markers and, as a result, a simplified verb form. Platt et al. (1977 in Williams 1987:176) argues that, phonologically, an unmarked past tense is most likely in verbs ending in consonant + d/t clusters, slightly less likely in verbs with vowel change, and least likely in verbs ending in consonant + *ed*. My small sample matched Platt's (ibid.) finding perfectly as my data revealed a missing past tense marker on three verbs ending in C+*t*, one verb with vowel change, and no verbs ending C+*ed*. The past time reference was made clear by the context for each of the missing past tense markers in my data. Williams (1984:177) would argue, then, that this clear context made past tense markers on the verbs "semantically redundant".

(ix) Prepositions

Gough (1996), De Klerk and Gough (2002), and Van Rooy (n.d.) all point out the extent to which BSAE is believed to be marked by "non-standard" preposition usage. Van der Walt and Van Rooy (2002) found that teachers in their study almost always accepted standard English use of prepositions, but also corrected less than 20% of the "non-standard" constructions in their test. In a study of the written language of low-achieving first year law students, Greenbaum and Mbali (2002) found that the students, who were mostly BSAE speakers, had particular difficulty in their use of prepositions - a serious problem in the field of law, as Greenbaum and Mbali (ibid.) stress. De Klerk (forthcoming) also points to a "distinctive use of prepositions" in BSAE.

The most commonly used prepositions in my data are: *of* (702 occurrences), *in* (326), *for* (220), *with* (110), and *about* (106). A concordance run on all prepositions occurring more than 10 times in my data resulted in 1364 occurrences. Of these, I classified 61 (or around 4.5%) as being "unusual" usage - either different from (underlined) or extra to (in italics) SE usage. Some examples are presented in Table 5.14 below.

	EXTRACT	REFERENCE
1	but they have to tell them so that they could know by which place can they place you	<XEC:#001:MA:273>
2	because it is very used to be to get hooked to drugs	<XEC:#003:MA:150>
3	what slogan do you remember about coca-cola?	<XEC:#005:FA:374>
4	because the workers they do have their own misconceptions and transfer those misconceptions to in the public and people will have another attitude about your business	<XEC:#006:FA:165>
5	the words that are written on top of the page of the dictionary	<XEC:#009:FB:126>
6	that's the first word in that page	<XEC:#009:FB:137>
7	before before we continue with our lesson who can just remind me *as to* what we were busy with?	<XEC:#011:FC:007>
8	into eh available format that means now <xhosa> <explanation> <xhosa> I can be able to to buy and to read it because it has been processed to the language whereby everybody can be able to understand it	<XEC:#011:FC:296>
9	so the i-xylem tissue it helps for the process of taking water molecule from the roots to the leaves	<XEC:#012:FD:210>

10	and because of those thickening vessels these vessels are called in different names because of the shape formed by those perforations and also the thickened wall	<XEC:#012:FD:260>

TABLE 5.14: UNUSUAL USE OF PREPOSITIONS

These few examples from the set of non-standard prepositions illustrate, I think, that the prepositions used by speakers of BSAE can be quite different to those used by SE speakers. It is worth noting, however, that I classed only a small number of prepositions in my data as unusual (approximately 4.5%). A large proportion (about 95.5%), then, could be classed as standard use. This finding is not only in contrast to most authors' views on preposition usage in BSAE (Gough 1996), but also seems to support Van Rooy's (n.d.) findings. Van Rooy (n.d.) concluded that about 88% of prepositions in his data set were used in a SE way. Of course, my data set is much smaller than Van Rooy's and broad generalisations are not possible from such a small set of data. It is, however, interesting that there is such a high proportion of standard usage in a feature that is often seen as highly typical of BSAE.

(x) Intensifiers

Although the use of *too* and *very much* as intensifiers is seen as a feature typical of BSAE (Gough 1996; Van der Walt 2001; Van Rooy n.d.), I did not find this to be the case in my data at all. There are no cases of *too* in my data and *very much* is only used once, (by FB), and that in a standard way: *I love your company very much* (<XEC:#009:FB:358>). Other intensifiers, such as *slight* and *very* are used in a fairly standard way, too. The size of my mini-corpus may, of course, be a factor in my findings here. However, I think that the context and topics probably played a greater role.

(xi) Generalisations of "being"

I found only two examples in my data where *being* acts as a participle. They were: *than to be guided by you being a murderer* (<XEC:#003:MA:253>) and

being the last one (<XEC:#012:FD:092>). Once again, although this feature is cited as typical (Gough 1996), there is scant evidence in my data to support this claim.

(xii) "Can be able to" as modal verb

My data revealed 16 instances of *can be able to* used as a modal verb - a characteristic common to BSAE (Gough 1996:63). Close inspection reveals that this feature is not widely spread throughout the sample. FA used this feature only once and FD, twice. The remaining 13 uses of this string were all made by FC. Although, as I have mentioned, my data set is too small to make any generalisations, it does seem as if this word-string is not consistently widely spread. The examples from FA and FD as well as a selection of examples from FC are presented in Table 5.15, below.

	EXTRACT	REFERENCE
1	we want to win the public so that they can be able I mean we can be able to serve them	<XEC:#005:FA:185>
2	only the person who has written it can be able to understand that data	<XEC:#011:FC:080>
3	that is internal source of data that means that information that you can be able to obtain it inside the business	<XEC:#011:FC:091>
4	that means that raw data now it has been processed in a text whereby you can be able to read it and understand	<XEC:#011:FC:257>
5	but as you are processing it means that readable format that means everybody who is able to read there can be able to read that because it has been processed into the language where everybody who can be able to read can be able to read it	<XEC:#011:FC:297>

6	they can be able to stand because of the cellulose wall and also of the lignin which is in the cell wall	<XEC:#012:FD:244>
7	so those food substances are manufactured by i-leaves and they are they are transported using the phloem tissue to the root so that the roots of the carrot can be able to store food substances	<XEC:#012:FD:351>

TABLE 5.15: "CAN BE ABLE TO" USED AS MODAL VERB

While Williams (1987) does not name the string *can be able to* as one commonly found in new Englishes, she does discuss the reduction of ambiguity and the principle of maximum salience as common underlying forces in the language of new Englishes. These two factors could play a role in the tautologous use of *can* and *be able to* in BSAE.

(xiii) "Nè?" as tag question

Gough (1996:62) lists "nè? as an invariant tag question" as a common characteristic of BSAE. The tag *nè*?, De Klerk (forthcoming:14) points out, is borrowed from Afrikaans and is "very commonly used in South African English to perform the same function as the tag question '*isn't that so?*' inviting the assent and participation of the listener". Williams (1987) points out that the use of tags, unlike those used by NSs, is one of the most obvious features of NIVEs. She says, however, that the phrase "invariant tag question" is "something of a misnomer, since the tag generally appears either invariably without a verb, or has a limited set of variants but does not agree with the verb in the main clause" as required in SE (ibid.). Williams (ibid.) suggests that the limited variation of tag questions in new Englishes is likely to be as a result of regularisation, i.e. "the reduction of irregularities" (Williams 1987:169).

In my data, the tag-question *nè?* is used 75 times in total - by four of the teachers - (11 times by FA, 32 times by FB, 22 times by FC, and 10 times by

FD). It is interesting to note that it is the male teacher who does not make use of this tag question whereas all the female teachers in the sample do so quite extensively. Gender studies researchers have often found that tag questions are more commonly and more widely used by females (see, for example, Cameron (1985); Lakoff (1975 in Coates 1993); and Preistler (1986)). Although this topic is outside the scope of this thesis, it may be interesting to explore this issue at a later stage.

Although it may be ideolectal rather than generally applicable, the tag *is it not so?* is used 20 times by FB. At no stage does FB abbreviate this phrase to *not so?*, despite the frequency with which she uses it. This is debatably not a prototypical tag question, but it does, I think, call into to question the notion that *nè?* alone is the tag question used by BSAE speakers. This, too, is something which may be worth exploring at a later stage.

5.2.2 Vocabulary

For these data, I looked at the entire mini-corpus as lexical items like these are easily dealt with by Wordsmith Tools' concordance. Much of the vocabulary in my data - particularly that from FD - is of a technical, field-specific sort which is doubtless uncommon in everyday conversations. Otherwise, the examples of "non-standard" meanings for SE words given by Gough (1996:64) are absent from my data. However, some words or phrases often found in BSAE by other authors were found in my data. Of particular interest are those words or phrases found by De Klerk (forthcoming) in her corpus of XE. For example, De Klerk (ibid.) found many examples of *maybe* "used to express conditional modality" in her corpus. *Maybe* was used 40 times in my corpus, twice in combination with *if* (<XEC:#001:MA:081> and <XEC:#007:FB:197>) - a common combination in De Klerk's corpus. It is often mentioned that BSAE speakers make use of the phrase *X's first time*, rather than *the first time* or *my first time* (Gough 1996:63; De Klerk and Gough 2002; Van der Walt and Van Rooy 2002:124). While De Klerk (forthcoming) found a number of occurrences of this expression in her corpus of XE, I found that there were only three occurrences of *first time* in my data,

all preceded by *the* as used in SE. An important reason for the very low incidence of *X's first time* in my data is probably the context in which the data were recorded. A structure such as *X's first time* is probably more likely to occur in "normal" conversations than in the teaching context.

De Klerk (forthcoming) also found evidence of the phrases *each and every* and *by all means* in her data. *Each and every* was used by three of the teachers in my sample - FB, FC, and FD - and occurred 4 times in my data. *By all means*, on the other hand was entirely absent from my corpus. The use of *than* in structures of comparison was also found in De Klerk's (ibid.) data. This feature is often cited as typical of BSAE (Gough 1996). Of the 12 occurrences of *than* in my data, only one revealed an unusual comparison structure: *meaning that ah abundant* <writing on board> *will be in that page because a comes first than o is it not so?* (<XEC:#009:FB:122>). I found little evidence in my data of the use of *-self* (occurring 3 times) and *-selves* (twice), all used in a standard way, in contrast to the fairly large incidence of these forms in De Klerk's (forthcoming) corpus.

5.2.3 Discourse Features

There are a number of discourse features in my data which, I feel, are typical of the classroom context. Two particularly interesting discourse patterns are explored in this section. For analysis of these data, I ran the concordance tool on the entire data set, rather than looking at my specifically targeted word-lists used for some findings in section 5.2.1.

(i) Student-Teacher interaction

In a discussion of teacher-talk, Cazden (1988:160) says that, in the classroom, "teachers talk two-thirds of the time ... initiate almost all interactions ... and interrupt but are not interrupted". In my data, there were a total of 1001 student utterances - 356 by single students, and 645 by two or more students - and 2984 utterances by teachers. On average, then, this means that teachers spoke for about 75% of the total utterances in the

classes and students, around 25%. The breakdown of student utterances into those produced by single students and those by groups of students shows that, while 36% of the student utterances are given by one student alone, around 64% of the student utterances were produced by two or more students - usually the whole class. Almost exclusively, any student participation which took place in the classes that I observed was teacher-initiated. The "interruptions" in my data, as I discuss below, result, in general, from the teaching-style of the teachers. As such, overlaps very often form part of the group-dynamic and learning environment set up by the teacher in the classroom.

(ii) Questions

An important feature of teacher-talk is the type and number of questions asked by the teachers. This gives an indication of the kind of student-teacher interaction which is evident in the classroom (Barnes 1969 cited in Stubbs 1976). In total, there were 825 questions asked by the teachers in my data set. Since the teachers produced a total of 2984 utterances, that means that about 28% of the teachers' utterances were in the form of a question. A summary of the ratio of questions produced by each teacher is presented in Table 5.16. This is followed by a classification of question type in Table 5.17.

TEACHER	TOTAL NUMBER OF UTTERANCES	TOTAL NUMBER OF QUESTIONS	RATIO OF QUESTIONS: UTTERANCES
MA	1 172	230	1 : 5 (20%)
FA	518	276	1 : 2 (53%)
FB	757	194	1 : 4 (26%)
FC	207	71	1 : 3 (35%)
FD	330	54	1 : 6 (16%)

TABLE 5.16: RATIO OF QUESTIONS TO UTTERANCES PER TEACHER

I thought that it would be useful to look at the specific types of questions used, rather than simply counting the number of questions used by each teacher. I have, therefore, designed a classification system for my question data. This classification system is based on four sources: Barnes (1969 in Stubbs 1976); Cazden (1988); Makhobotloane (1992); and the data themselves. I begin by briefly presenting relevant terms and concepts from these sources and then present a summary of the terms used in my classification system. Finally, I give a table showing the types of questions used by each teacher.

Firstly, Barnes (1969 cited in Stubbs 1976:80) gives a summary of four broad areas of questions used by teachers in the classroom: factual, reasoning, open and social question. He admits to this being a fairly vague and subjective classification, but says that a study of teachers' questions is nevertheless a useful study in understanding classroom discourse. Barnes (ibid.) argues that many questions have a function different to their apparent surface function. He terms such questions "pseudo questions" and gives the example of a teacher asking a pupil's opinion about something - the surface view - while really having "a particular view in mind" - the real function of the question.

Secondly, Cazden (1988:160), who draws on Stubbs, speaks of various types of "metacommunicative talk" in the classroom - the function of which is to "monitor and control the classroom communication system". A few of these are useful in looking more specifically at questions used by teachers in the classroom, namely those utterances which control the amount of talk used by students, check student understanding, or ask for closer definitions from students.

Makhobotloane (1992:106), the third source used in the development of my classification system, uses what she terms an "eclectic system of classification of ... questions" in her study. She speaks of echoic questions (after Kearsley 1976), which consist of: questions asking students to repeat what they have said; questions which check for student comprehension;

"clarification requests", which help the teacher to follow what the student has said; and "confirmation checks" which ensure that students have understood the teacher's utterance in the way intended by the teacher. Kearsley (1976 in Makhobotloane 1992:106) also makes use of so called "epistemic questions" - consisting of evaluative questions "that serve to establish the addressee's knowledge and inference skills" (ibid.) and "referential questions" which ask the speaker to "provide contextual information about something" (ibid.).

Finally, the data from my teachers played a central role in determining categories for the questions used by the teachers. Table 5.17 below gives my definition or description of each category, followed by examples (in italics) from my data - one example from each teacher, where possible.

TYPE	DESCRIPTION AND EXAMPLES
display	these relate to Barnes' (1969 in Stubbs 1976) "factual" questions and Kearsley's (1976) "evaluative" questions; they serve to give students a chance to "display" their knowledge <XEC:#001:MA:361> *where can you get a testimonial?* <XEC:#005:FA:053> *who is the public?* <XEC:#007:FB:198> *what part of speech is abortion?* <XEC:#011:FC:103> *where can I get that information?* <XEC:#012:FD:368> *what is the function of the sclerenchyma?*
checks	used in Cazden's (1988) classification and named after Kearsley's (1976) "confirmation checks", these question types are used for the teacher to check that the students have understood what s/he has said <XEC:#005:FA:295> *you understand?* <XEC:#007:FB:100> *is it clear?* <XEC:#011:FC:031> *what have we what have I just said?* <XEC:#012:FD:348> *okay are there any questions?*

sentence	this category evolved from a study of my data, rather than any other source; here, the teachers begin a sentence use a rising intonation typical of questions, without completing the sentence; students usually interpret this as a question <XEC:#001:MA:334> *ja you must use simple words and avoid?* <XEC:#005:FA:041> *talking about those relations between the business and the?* <XEC:#009:FB:218> *and that is the spelling of?* <XEC:#011:FC:208> *internal source of data and?* <XEC:#012:FD:054> *in the deep of woods of plants are the?*
nè?	again, this is a category which comes out of my data; this question also serves as a type of confirmation check or yes / no question <XEC:#005:FA:123> *as we know businesses are increasing nè?* <XEC:#007:FB:276> *she's nice nè?* <XEC:#011:FC:250> *that means all the words alphabetically nè?* <XEC:#012:FD:118> *so these walls are said to be impermeable nè?*
more	related to Barnes' (1969 in Stubbs 1976) "reasoning" questions, Cazden's (1988) classification, and Kearsley's (1976) "referential" questions, these questions involve the teachers asking the students to give them more detailed information about something <XEC:#001:MA:123> *why maybe a telephone company?* <XEC:#006:FA:214> *how can you create that?* <XEC:#007:FB:014> *what else is there in the dictionary?* <XEC:#012:FD:014> *and what?*

yes / no	this is a standard question requiring that students answer yes or no <XEC:#001:MA:359> *do you know what a testimonial is?* <XEC:#005:FA:023> *is she right?* <XEC:#007:FB:120> *do you see those words?* <XEC:#011:FC:145> *do you see the difference?* <XEC:#012:FD:392> *are there intercellular air-spaces between the sclerenchyma cells?*
Xhosa	I've used this label for questions which were asked in Xhosa; although both the intonation used by the teacher and the response of the students reveal that these are questions, I am not able to classify these questions closely since I do not have a knowledge of Xhosa
echoic	I have borrowed this term from Kearsley (1976), but have limited it to only those questions in which the student is asked to repeat him/herself <XEC:#001:MA:349> *sorry?* <XEC:#005:FA:029> *hey?* <XEC:#008:FB:215> *what pleasant what?* <XEC:#011:FC:157> *pardon?* <XEC:#012:FD:038> *can you please come again?*
vocative	in these questions, again labelled from my data rather than any of the other sources, the teacher has used a student's name in a question <XEC:#002:MA:242> *what could it be <name>?* <XEC:#008:FB:210> *<name>?*
odd	these are all questions which do not fit neatly into any of the other categories <XEC:#001:MA:266> *the arthritis ability?* <XEC:#005:FA:372> *coca-cola?* <XEC:#010:FB:092> *<xhosa> <comment> <xhosa> written the word there?* <XEC:#011:FC:017> *<to me> you don't mind if I use Xhosa?*

reported	reported rather than actual questions are included here <XEC:#001:MA:224> *so within the interview maybe you will be asked who told you about that this job?* <XEC:#001:MA:287> *if I am if I was asked I'm I'm looking for employment at another school and they ask me what salary do you expect?*
self-check	questions in which the teacher asks the equivalent of "are you sure?" to guide students in their answers <XEC:#001:MA:048> *does it say so?* <XEC:#005:FA:027> *you are all?*

TABLE 5.17: CLASSIFICATION OF TEACHERS' QUESTIONS WITH EXAMPLES

Table 5.18, below, gives a summary of these question types as used by each teacher.

QUESTION TYPE	TEACHERS					TOTALS
	MA	FA	FB	FC	FD	
display	142	76	67	6	17	308
checks	0	63	38	8	12	121
sentence	12	82	16	4	3	117
nè?	0	11	31	22	10	74
more	28	22	8	0	5	63
yes / no	17	6	18	2	1	44
Xhosa	3	3	3	22	3	34
echoic	12	9	6	1	3	31
vocative	5	0	1	0	0	6
odd	1	1	1	3	0	6
reported	5	0	0	0	0	5
self-check	2	1	0	0	0	3
TOTALS	230	276	194	71	54	825

TABLE 5.18: QUESTION TYPES PER TEACHER

The different teachers in my sample used questions in different ways and for different ends. Most of these are clear from the contexts shown above, but the use of sentence-completion type questions deserves closer comment, I feel. While many of the questions from each teacher were of the "display" type, the sentence-completion questions seem to fulfil a very different function. For example, FA used a large number of sentence-completion questions to encourage class participation. As a result there is a fairly high degree of overlap between student and teacher utterances in the data from her classes. Of the 389 overlaps in my data, 150 (or 38.5%) of these were from FA. Data from MA revealed 76 occurrences (19.5%), FB had 97 overlaps (close to 25%), FC 50 (nearly 13%), and FD only 16 occurrences of overlaps (4%).

The extract below shows how FA models the kind of response pattern that she is looking for in her sentence-completion type questions. (The student responses, as explained above, are in brace brackets).

<XEC:#005:FA:033>	okay good people when you are talking about public relations i-public relations if you remember from standard six you talked about those people who give their <writing on board> opinion in order to improve the?
<XEC:#005:FA:034>	the image of a firm
<XEC:#005:FA:035>	okay?
<XEC:#005:SS:036>	{yes}

The above extract, as can be seen from the line numbering, occurred early on in the lesson that I was observing. Later on in the same lesson, this pattern is repeated with the students playing a far more active role as shown in the extracts below.

<XEC:#005:FA:182>	you know that in our businesses we are using those types of?
<XEC:#005:SS:183>	[{communication media}

<XEC:#005:FA:184> [communication media in order to communicate with the public because we want to win what?

<XEC:#005:FA:185> we want to win the public so that they can be able I mean we can be able to serve them

<XEC:#005:FA:186> okay?

<XEC:#005:SS:187> {yes}

This type of sentence-completion question is favoured by FA, and used, to a far smaller extent, by all the teachers in my sample. The response of the students shows that they are familiar with what their teacher expects from them. The more typical open question seen in utterance <XEC:#005:FA:184> is similarly "correctly" interpreted by the students: they realise that their teacher's style is to build a question intonation into her speech. This seems to ensure that the students are far more active participants in the lesson than they otherwise would be. The extract below shows this feature of participation very clearly.

<XEC:#005:FA:245> that is why we say because of the growth in the?

<XEC:#005:FA:246> [social responsibility

<XEC:#005:SS:247> [{social responsibility}

<XEC:#005:FA:248> that has also led to the?

<XEC:#005:FA:249> [importance of public relations

<XEC:#005:SS:250> [{importance of public relations}

My data revealed very few generally applicable trends in the teachers' use of different question types. While most teachers (MA, FA, FB) seemed to favour display-type questions, other question types seemed to be a result of individual preference and teaching-style rather than general teacher practice. FA's use of sentence-completion questions, as discussed above, is one such example. While the content of the lessons would always play a role in the types of questions used, the teachers' use of "more"-type questions, in particular, seems to be influenced by context. It makes intuitive sense that, for example, a poetry lesson would allow far more room for expanded answers from students than would a biology lesson where new information is

being given. As FD's biology lesson involved her largely presenting new information to the class, it makes sense that she used very few questions in her discussion - almost half of these being some sort of check for understanding. Another case in which context plays a role in the type of questions used is in the case of "reported" questions. This structure (as described above) is unlikely, I think, in any but the context in which it occurred - a discussion of applying for a job.

It is interesting to note the relatively limited use of students' names in questions. This may be a reflection of the group- rather than individual-answer apparently expected by many teachers. This, too, then can be related to individual teaching styles. By way of illustration, MA asked the most display questions by a large margin, and also made particular use of student names in his questions. The reasons for the teachers' relatively scant use of student's names in questions could be an interesting area for future studies.

A high occurrence of Xhosa questions and the use of the *nè?* tag in FC's speech ties in, I feel, with her relatively high use of isi-Xhosa during her teaching. The other teachers in my sample, although making quite frequent use of the *nè?* tag, used relatively few Xhosa questions (only 3 each).

The ways in which the teachers in my study use questions seem to tie in closely to the context. While there is a common context, in that all the data is collected from teacher-talk in the classroom, there are also variations in terms of class size, content taught, and the student-interactions set up by each teacher in his or her class. Future studies could explore the used of questions by Xhosa mother-tongue teachers in greater depth. It may also be interesting to compare the patterns found in my data with those found in data from other BSAE varieties.

5.3 SUMMARY AND CONCLUSION

While Gough's (1996) classification of grammatical features of BSAE was useful as a basis for my data analysis, other authors provided greater depth and context to my discussion. Van der Walt (2001) and Van der Walt and Van Rooy (2002) had looked at BSAE in an educational context, so their findings were used as a point of comparison with my classroom data. I found Van der Walt and Van Rooy's (2002) well-constructed argument, using teachers' and learners' perceptions and use of the "norms" of BSAE as evidence that South African English is in the liberation phase of new English development, particularly useful. I also made use of Williams' (1987) exploration of new Englishes (or NIVEs, as she terms them). Williams' (1987) study proved useful in my interpretation of certain trends in my data. It was also useful to compare my findings with De Klerk's (forthcoming) study, as her focus, too, as I have explained, is on spoken Xhosa English.

In general, I found that, although many of the features reported by Gough (1996) as "typically cited" are evident to some extent in my data, the incidence of occurrence is far lower than Gough (1996) and De Klerk and Gough (2002) seem to suggest, although neither of these sources quantifies the extent to which the characteristics listed are evident in BSAE. In this regard, I found Van der Walt and Van Rooy's (2002) basic quantification of features in their data useful for comparative purposes. Williams (1987), while not focusing on the frequency of occurrence of features of new Englishes, did point to a useful number of characteristics which are also evident in BSAE.

The relatively low frequency of occurrence of many "typical" features of BSAE in my data may, as I have argued, result from the context in which my data was obtained. While Gough's (1996) list of features is compiled from student writing, my data comes from the speech of teachers. The difference between speech and writing must, of course, play a role. I have also argued that the classroom situation requires that the teachers be models of language use for their students. This is in contrast to the data used to compile Gough's (1996) list of typical features.

Evidence from my data, although not widely generalizable, seems to lie in favour of Van der Walt and Van Rooy's (2002) hypothesis. I, too, found that there was clear evidence of inconsistencies in my participants' use of English. Where inconsistent evidence came from different teachers, various contextual factors could be suggested to account for the variation in findings. By way of explanation, I mentioned, for example, contextual factors such as the size of the class or the content covered in a lesson as well as factors relating to the different personalities of the teachers. Where individual teachers were inconsistent in their usages, however, Van der Walt and Van Rooy's (2002) argument seems to be most convincing. It seems as if, in many cases, the teachers in my study were aware of the "norms" of standard English - as evinced by the relatively high frequency of "standard" usages for the great majority of the features explored in this thesis. Almost all of the features studied were used in a "non-standard" form less than 5% of the time. Thus, it appears that teachers in my sample were more likely to use "standard" utterances than not.

A focus on features suggested by other authors naturally means that there are many features in my data which have not been explored at all. However, I feel that the most useful way to investigate the data was to study the known features of the variety. A strong familiarity with my data meant that high frequency constructs not described in the literature - such as the use of the *i-* and *u-* prefixes - were not overlooked.

I have mentioned the limitations of my study, and the fact that its small size prevents me from being able to make firm generalisations about BSAE in schools. I do, however, feel that this study has raised a number of issues which could be explored at a later date. An important area of interest would, naturally, be a comparison of teacher-talk from speakers of different language backgrounds. My data reflects what is common among a small group of Xhosa mother-tongue speakers in Grahamstown, but it would be interesting to compare and contrast this with data from other speakers. Other areas which could be explored later are discussed in the Concluding Remarks of this thesis.

CHAPTER SIX: CONCLUDING REMARKS

I have mentioned, throughout my thesis, the potential for various issues to be explored further. The greatest scope in this regard, I think, is the potential for comparative studies such as: an exploration of XE as used in contexts other than the classroom; a comparison of this data set with a similar study of another variety of BSAE; or the comparison of the language of XE teachers in Grahamstown with that of mother-tongue English teachers, for example.

I think that comparisons made with studies of XE in other contexts may help to clarify whether the XE used by teachers is significantly different from that used in other contexts. A study of teacher-talk in another variety of BSAE could reveal some interesting information. Both this, and a study of mother-tongue English teacher-talk could reveal patterns of teachers' language use.

More specific features revealed in my data could also be usefully sought in other BSAE studies. My participants use of the *i-* and *u-* prefixes is one such example, as is its possible link to code-switching behaviour in speakers.

It could be interesting to look at the production "errors" in BSAE speech. I only mentioned this briefly in my Findings and Discussion, (in section 5.2.1(iv)), as it is beyond the scope of this study, but I feel that an exploration of this phenomenon may give some interesting insights.

On the whole, the corpus-based approach used in this study was useful. However, it must be remembered that analysis based on pattern-finding necessarily relies on the researcher having a clear idea of potential patterns in the data. As such, the patterns must be inspired by the data. This requires that the researcher has a strong familiarity with the data.

APPENDIX A: BACKGROUND INFORMATION AND CONSENT FORM

Background Information/Consent Forms

Circle answer where appropriate

1. a. Were you born in South Africa? Yes No

b. Please specify town or region .....

c. Where did you grow up? (specify town or region) .....

2.a. Which language did you speak first in your home? .....

b. Which language did your mother first speak at home?.....

c. Which language did your father first speak at home?.....

d. Which other language(s) do you speak fluently?.....

3. Please ***circle the appropriate answer***

a. Gender: Female Male

b. Age group

16 - 19	20 - 24	25 - 29	30 - 34
35 - 39	40 - 44	45 - 49	50 - 54
55 - 59	60 - 64	65 - 69	70 - 74
75 - 79	80 - 84	85 - 89	older

c. Which ethnic group(s) do you identify with? (You may circle more than one)

Xhosa	Zulu	Pedi	Venda
Sotho	Tswana	Tsonga	Ndebele
Tsonga	Other (please specify).....		

4. What is the highest educational qualification you have obtained? .....

a. About how many years did you study English as a subject?.....

b. Were you taught through the medium of English? Yes No Sometimes

c. How much English do you use daily?
None A little A lot All the time

5. Are you currently: employed unemployed a student

I give permission for the recording of my voice to be included in a corpus of South African English (which may be released on CD) to be used for linguistic research and teaching purposes.

Signed .....

Rhodes Corpus of Spoken South African English

Background information (continued)

PLEASE NOTE THE FOLLOWING INFORMATION FOR EVERY RECORDING

Date recorded: ...../...../.....

Place recorded: .....

Number of people present: .....

(Fill in additional sheets providing background information on all contributors)

Private or public:.....

Audience or not (if so state approximate size):.....

Domain (e.g. home, business, school, Television):.....

Topic:.....

Distance (i.e. telephone) or direct (i.e. face-to-face).....

Spontaneous or prepared (scripted).....

Formal/Neutral/ Informal:.....

Dialogue or Monologue:.....

Any other relevant information:.....

.....

.....

.....

.....

APPENDIX B: DETAILS OF SCHOOLS

Details about the ex-DET schools in Grahamstown are almost unobtainable. Mr. B. Fargher, manager of GADRA Education in Grahamstown, was most helpful in giving me the following information. There are 6 ex-DET schools in "Grahamstown East", and one school which has a mixed enrolment of Xhosa mother-tongue and Coloured pupils. A total of 539 pupils from these schools wrote their matric exam last year, 440 from the 6 ex-DET schools mentioned. The numbers of matrics at each school give only a small indication of pupil numbers, Mr. Fargher told me, as there is a large drop-out rate at most of these schools and most students do not get as far as matric.

The Regional Department of Education was unable to give me any specific statistics about the numbers of teachers at the schools or the number of pupils. Kevin Parry from Statistics South Africa was helpful, but statistics from him seem, in his words, "a little underestimated": a total of 4 teachers in Grahamstown, 3 male and 1 female.

What information I have, then, is from the schools themselves and that, too, is not too detailed. The teachers in my sample come from two different schools, which can be labelled A and B. School A has about 1100 pupils and 30 teaching staff. The teaching staff are all Xhosa mother-tongue speakers. I am told that there are "more" female than male teacher, although specific figures could not be obtained. School A seems relatively well-equipped and the poverty level seems lower than at school B. School B has about 800 pupils and 25 members of staff - 24 of whom are Xhosa mother-tongue speakers. There are 12 male and 13 female teachers at this school. School B is much less well equipped - there are usually not enough desks or chairs for each pupil, for example - and students seem to be very poor. Teachers at school B told me that, although the students do go home for "lunch", there is often no food at home for them to eat.

APPENDIX C: DETAILS OF AWK

AWK is a simple programming language most useful for text processing. It is often used to report on specific data in large text files, based on patterns in the text. Such patterns are specified in "regular expressions" by the programmer. Each line matching a regular expression will cause instructions associated with that regular expression to be executed. An instruction might be, for example, to print the line matching the regular expression.

AWK was first designed and described in Aho, Weinberger, and Kernighan (1988) whose initials make up its name. The version of AWK used here is GNU Awk, developed as part of the GNU Project. The complete manual for GNU Awk is available online at:

http://www.gnu.org/software/gawk/manual

GNU Awk is available at no charge for most current computer systems. The manual explains how and where to obtain it.

The utterance-numbering of my transcribed data was done by an AWK program written for me by Emmanuel Lamprecht. I had transcribed my data, using simply "T" at the beginning of a teacher's utterance and "S" or "SS" at the beginning of student/s utterances. Each new utterance was preceded by an explicit new line. The AWK program converted my data into a clearly coded form as can be seen in Appendix D.

APPENDIX D: SAMPLES OF TRANSCRIBED DATA

SAMPLE OF TRANSCRIBED DATA: MA

<XEC:#003:MA:001> page four seven
<XEC:#003:MA:002> prayer before birth by louis macniece
<XEC:#003:00:003> <general noise>
<XEC:#003:S:004> <unclear>
<XEC:#003:S:005> <sneeze>
<XEC:#003:MA:006> this is a poem by louis macniece
<XEC:#003:MA:007> and it is about a prayer by an unborn child
<XEC:#003:MA:008> if you analyse the title prayer before birth
<XEC:#003:MA:009> so the child is not yet unborn
<XEC:#003:MA:010> but she or he is praying from where?
<XEC:#003:S:011> {from his mother's womb}
<XEC:#003:MA:012> from the mother's?
<XEC:#003:SS:013> [{womb}
<XEC:#003:MA:014> [womb ja
<XEC:#003:MA:015> so expressing his his or her <unclear> mother's womb
<XEC:#003:MA:016> the speaker an unborn child
<XEC:#003:SS:017> <general noise>
<XEC:#003:S:018> <cough>
<XEC:#003:MA:019> what is he praying about?
<XEC:#003:MA:020> you will just read the poem
<XEC:#003:MA:021> you see that what is the child praying about?
<XEC:#003:MA:022> it was an unborn child
<XEC:#003:MA:023> it has not yet learn how to how to pray
<XEC:#003:MA:024> but this one in the words of this poem is praying
<XEC:#003:MA:025> can anyone of you read the poem?
<XEC:#003:MA:026> please
<XEC:#003:MA:027> not afraid of her she is my visitor
<XEC:#003:MA:028> <xhosa> <comment> <xhosa>
<XEC:#003:S:029> <reading>
<XEC:#003:MA:030> thank you <name>
<XEC:#003:MA:031> i will read it once again
<XEC:#003:MA:032> by the time I read it I want you to focus on what the child's fears are
<XEC:#003:MA:033> what the child is afraid
<XEC:#003:MA:034> if you listen in in each of the opening lines of these stanzas
<XEC:#003:MA:035> hear me console me provide me forgive me help me carry me fill me
<XEC:#003:MA:036> so the child is always appealing for something
<XEC:#003:MA:037> either to be heard to be consoled to be comforted to be provided with something
<XEC:#003:MA:038> to be forgiven to be rehearsed be given practise in something rehearse
<XEC:#003:MA:039> to be heard again or to be filled with something
<XEC:#003:MA:040> focus on that on what his fears are
<XEC:#003:MA:041> what the child is fearing something
<XEC:#003:MA:042> <reading>
<XEC:#003:MA:043> the child's fears there
<XEC:#003:MA:044> what are the child's fears there?
<XEC:#003:MA:045> ... what does the child fear in stanza number one?
<XEC:#003:MA:046> let not the blood-sucking bat the rat stoat club-footed ghoul come near him or her
<XEC:#003:MA:047> what are his fears?
<XEC:#003:MA:048> <name>

<XEC:#003:S:049> <unclear>

<XEC:#003:MA:050> ja animals
<XEC:#003:MA:051> you all know even you during your childhood you had these fears unfounded fears som- sometimes of certain creatures
<XEC:#003:MA:052> some of you you are afraid of cats maybe or dogs
<XEC:#003:MA:053> so this child is afraid of the bat the blood-sucking bat
<XEC:#003:MA:054> maybe she was told or he was told that there is an a blood-sucking bat
<XEC:#003:MA:055> he is afraid of the rat or the stoat
<XEC:#003:MA:056> the stoat is a small brown furry long-bodied animal which was found in europe
<XEC:#003:MA:057> and it feeds on other animals
<XEC:#003:MA:058> so he was told that there is an animal that is a stoat
<XEC:#003:MA:059> a club-footed ghoul
<XEC:#003:MA:060> an imagined spirit
<XEC:#003:MA:061> an imagined monster that could be ghoul
<XEC:#003:MA:062> something that is called ghoul
<XEC:#003:MA:063> an imagined spirit
<XEC:#003:MA:064> so this child has this childhood fears of certain animals and also certain imagined animals or spirits
<XEC:#003:MA:065> also in our culture we are afraid of certain spirits that we are told that they exist
<XEC:#003:MA:066> stanza number two
<XEC:#003:MA:067> <reading>
<XEC:#003:MA:068> stanza number stanza number one he appeals to be heard
<XEC:#003:MA:069> in stanza number two she or he wants to be consoled to be comforted
<XEC:#003:MA:070> <reading>
<XEC:#003:MA:071> that is his fear
<XEC:#003:MA:072> <reading>
<XEC:#003:MA:073> again let's identify the fears of the child in stanza number two
<XEC:#003:MA:074> what are his or her fears?
<XEC:#003:MA:075> this stanza number two they are written there in black and white
<XEC:#003:MA:076> ... the first one

<XEC:#003:S:077> <unclear>

<XEC:#003:MA:078> his his his fears
<XEC:#003:MA:079> what is afraid of?
<XEC:#003:MA:080> look at stanza number two
<XEC:#003:MA:081> the human race
<XEC:#003:MA:082> what particularly is he afraid of from human race?

<XEC:#003:S:083> <unclear>

<XEC:#003:MA:084> sorry?

<XEC:#003:S:085> {prison}

<XEC:#003:MA:086> prison
<XEC:#003:MA:087> what makes you think that?
<XEC:#003:MA:088> .. just call the word that makes you think that

---

<XEC:#003:S:089> {because of tall walls}

<XEC:#003:MA:090> tall walls ja they could be referring to prison because prison is really about tall walls
<XEC:#003:MA:091> he is afraid of being confined
<XEC:#003:MA:092> so he <unclear> confinement
<XEC:#003:MA:093> prison

<XEC:#003:SS:094> <noise>

<XEC:#003:MA:095> we accepted what you have said
<XEC:#003:MA:096> another fear?

SAMPLE OF TRANSCRIBED DATA: FA

<XEC:#005:FA:001> okay good people today amongst us we've got a visitor she is candice and she's from rhodes okay?

<XEC:#005:SS:002> {yes}

<XEC:#005:FA:003> she's going to be with us today
<XEC:#005:FA:004> just be free don't be scared

<XEC:#005:SS:005> [<unclear>

<XEC:#005:FA:006> [all right
<XEC:#005:FA:007> okay otherwise today we are going to start on a new thing
<XEC:#005:FA:008> we are going to talk

<XEC:#005:00:009> <interruption>

<XEC:#005:FA:010> okay so so we are going to start on a new chapter <unclear> public relations <xhosa> <comment> <xhosa>

<XEC:#005:S:011> <unclear>

<XEC:#005:FA:012> public relations okay?
<XEC:#005:FA:013> let us just recap or remind ourselves
<XEC:#005:FA:014> what do we know about public relation function?
<XEC:#005:FA:015> we have done public relations in standard six standard seven standard eight standard nine

<XEC:#005:SS:016> {yes}

<XEC:#005:FA:017> ... come people don't waste time
<XEC:#005:FA:018> what do you remember?
<XEC:#005:FA:019> don't don't be shy
<XEC:#005:FA:020> what do you remember about

<XEC:#005:S:021> <unclear>

<XEC:#005:FA:022> those in the power of public opinion
<XEC:#005:FA:023> is she right now?

<XEC:#005:SS:024> {yes}

<XEC:#005:FA:025> is she right?

<XEC:#005:SS:026> {yes}

<XEC:#005:FA:027> you are all?

<XEC:#005:SS:028> {wrong}

<XEC:#005:FA:029> hey?

<XEC:#005:SS:030> <unclear>

<XEC:#005:FA:031> you are all?

<XEC:#005:SS:032> <general noise>

<XEC:#005:FA:033> okay good people when you are talking about public relations i-public relations if you remember from standard six you talked about those people who give their <writing on board> opinion in order to improve the?
<XEC:#005:FA:034> the image of a firm
<XEC:#005:FA:035> okay?

<XEC:#005:SS:036> {yes}

<XEC:#005:FA:037> in order to improve the image of that firm

<XEC:#005:00:038> <writing on board>

<XEC:#005:FA:039> who do you think now is it when we are talking <writing on board>

<XEC:#005:S:040> [{shh}

<XEC:#005:FA:041> <writing on board> [about i-public relations again we are talking about those relations between the business and the?
<XEC:#005:FA:042> and who?
<XEC:#005:FA:043> we are talking about good relations between the business and?

<XEC:#005:SS:044> {the public}

<XEC:#005:FA:045> the public
<XEC:#005:FA:046> okay?

<XEC:#005:SS:047> {yes}

<XEC:#005:FA:048> <writing on board> as the name implies public relations okay?
<XEC:#005:FA:049> so we are talking about the good relations between who?
<XEC:#005:FA:050> the public <writing on board> and the business
<XEC:#005:FA:051> okay?

<XEC:#005:SS:052> {yes}

<XEC:#005:FA:053> who is the public?
<XEC:#005:FA:054> when i'm saying we are talking about good relations between the public and the?
<XEC:#005:FA:055> and the business
<XEC:#005:FA:056> who is the public?

<XEC:#005:S:057> {consumers}

<XEC:#005:FA:058> the?

<XEC:#005:S:059> {consumers}

<XEC:#005:FA:060> consumers very good
<XEC:#005:FA:061> who else?
<XEC:#005:FA:062> <writing on board> we are talking about consumers

<XEC:#005:00:063> <teacher walking>

<XEC:#005:FA:064> there a lot of people that are involved there
<XEC:#005:FA:065> consumers who else?
<XEC:#005:FA:066> ... come people don't <unclear>
<XEC:#005:FA:067> don't make me <unclear>
<XEC:#005:FA:068> mm?
<XEC:#005:FA:069> okay we're talking about the consumers <writing on board> talking about the government
<XEC:#005:FA:070> okay?
<XEC:#005:FA:071> all right
<XEC:#005:FA:072> so now we are ta- going to talk about [those relations

<XEC:#005:00:073> [<whistling outside>

<XEC:#005:FA:074> those those good relations between?
<XEC:#005:FA:075> the business and the public
<XEC:#005:FA:076> that is what means we talk about public relations
<XEC:#005:FA:077> okay?
<XEC:#005:FA:078> there are certain factors that have lead to the importance of public relations
<XEC:#005:FA:079> so let us start with those <writing on board> factors <xhosa> <ne?> <xhosa>

SAMPLE OF TRANSCRIBED DATA: FB

<XEC:#009:FB:001> please please please people shhhhh

<XEC:#009:SS:002> {shhh}

<XEC:#009:FB:003> when the bell rings please you have to go to the classroom
<XEC:#009:FB:004> that's the meaning of the bell that's go to your classroom
<XEC:#009:FB:005> so today i want us to have a look at the dictionary
<XEC:#009:FB:006> i know that some of you have got

<XEC:#009:SS:007> {shh}

<XEC:#009:FB:008> dictionaries of your own is it not so?

<XEC:#009:SS:009> {yes}

<XEC:#009:FB:010> it is not a bad time that you are bringing dictionaries
that's what i believe

<XEC:#009:SS:011> <unclear>

<XEC:#009:FB:012> and the mere fact that each and every one has got a
dictionary it means that dictionary is a useful tool is it
not so?

<XEC:#009:SS:013> {yes}

<XEC:#009:FB:014> so what i would like to know is that when you buy a
dictionary there is a reason for buying it <xhosa> <ne?>
<xhosa>

<XEC:#009:SS:015> [{yes}

<XEC:#009:FB:016> [i want to know what is it that you buy a dictionary for
<XEC:#009:FB:017> do you buy it so that you can use it as an ornament in your
home?

<XEC:#009:SS:018> {no}

<XEC:#009:FB:019> so then tell me what do you use it for?

<XEC:#009:S:020> {so that maybe i'm looking for a particularly word that i
don't know}

<XEC:#009:FB:021> ja good boy
<XEC:#009:FB:022> he's using i-dictionary book to search for words

<XEC:#009:00:023> <writing on board>

<XEC:#009:FB:024> the meaning of the words
<XEC:#009:FB:025> why else do we use i-dictionary for?

<XEC:#009:SS:026> {how to spell words}

<XEC:#009:FB:027> ja for the spelling of the words
<XEC:#009:FB:028> if you are not sure of the spelling just consult your
dictionary <writing on board>
<XEC:#009:FB:029> what else do we use i-dictionary for?

<XEC:#009:S:030> {for understanding}

<XEC:#009:FB:031> meaning?

<XEC:#009:SS:032> {meaning}

<XEC:#009:FB:033> that's the meaning
<XEC:#009:FB:034> if you don't understand the meaning of the word

<XEC:#009:SS:035> <laugh> <noise>

<XEC:#009:FB:036> you will have to use your dictionary
<XEC:#009:FB:037> why else do we use i-dictionary for?

<XEC:#009:S:038> {ah alphabetically the words}

<XEC:#009:FB:039> <laugh> so i-dictionary what is he is telling us that if you have not used i-dictionary or maybe it's the first time to use i-dictionary you must know that it is written in the alphabetical order meaning that from a to?
<XEC:#009:FB:040> [z

<XEC:#009:SS:041> [{z}

<XEC:#009:FB:042> what else can you tell me about the dictionary what other thing are there in a dictionary?
<XEC:#009:FB:043> what what are the other thing in a dictionary?

<XEC:#009:SS:044> <unclear>

<XEC:#009:FB:045> <xhosa> <comment> <xhosa> i-pronunciation

<XEC:#009:00:046> <writing on board>

<XEC:#009:FB:047> how to pronounce the word how to spell the word
<XEC:#009:FB:048> what else is there in a dictionary?
<XEC:#009:FB:049> ... mm?

<XEC:#009:S:050> <cough>

<XEC:#009:FB:051> we have abbreviation

<XEC:#009:00:052> <writing on board>

<XEC:#009:FB:053> just look at the back of your dictionary or in the front it depends
<XEC:#009:FB:054> the other ab- the abbreviations are written at the back the other dictionaries in the other they are written in front
<XEC:#009:FB:055> there's a lot of things there
<XEC:#009:FB:056> the countries of the world they are written there
<XEC:#009:FB:057> ah the abbreviations
<XEC:#009:FB:058> <cough> there is a lot of there are a lot of things that you can find in your dictionary
<XEC:#009:FB:059> and the <unclear> concerns the abbreviations and?
<XEC:#009:FB:060> <writing on board> the parts of speech
<XEC:#009:FB:061> the part of speech of the word
<XEC:#009:FB:062> when you talk of the part of speech can you remind me what do we talk about?

<XEC:#009:S:063> <cough>

<XEC:#009:FB:064> will you please remind me when you talk about part of speech what are you talking about?
<XEC:#009:FB:065> what is it that you are talking about when you talk of i-part of speech parts of the speech?
<XEC:#009:FB:066> ah how many part parts of speech do we have?

<XEC:#009:SS:067> <unclear>

<XEC:#009:FB:068> two four?

<XEC:#009:SS:069> [{four}
<XEC:#009:SS:070> [{two}
<XEC:#009:SS:071> {eight}

<XEC:#009:FB:072> eight

SAMPLE OF TRANSCRIBED DATA: FC

<XEC:#011:FC:042> not that you are saying the other functions are not crucial but these three are the most crucial ones <xhosa> <ne?> <xhosa>

<XEC:#011:SS:043> {yes}

<XEC:#011:FC:044> okay and then ... so we are going to continue as to the difference between a data and information

<XEC:#011:SS:045> <unclear>

<XEC:#011:FC:046> eh i'm going to continue with the difference between a data and information
<XEC:#011:FC:047> as to what is a data and what is a an information
<XEC:#011:FC:048> ... eh a data it is a raw information
<XEC:#011:FC:049> that means you cannot relate this as a <unclear>
<XEC:#011:FC:050> an information is a processed data
<XEC:#011:FC:051> for instance lets take a eh um a journalist
<XEC:#011:FC:052> when he or she goes to interview a person maybe for instance ah [mandela

<XEC:#011:S:053> <cough>

<XEC:#011:FC:054> eh the journalist does not write it in full but sh- she or he uses what is known as shorthand
<XEC:#011:FC:055> that means we use known abbreviations for <unclear> <xhosa> <ne?> <xhosa>

<XEC:#011:SS:056> {yes}

<XEC:#011:FC:057> but if for the for instance if for instance i give you that [that information

<XEC:#011:SS:058> [<laugh>

<XEC:#011:FC:059> that data which has been used maybe you cannot read because it has been <unclear> in sh- written in in in shorthand
<XEC:#011:FC:060> that is it is a raw data
<XEC:#011:FC:061> that means you cannot re-read it whereby you can understand it
<XEC:#011:FC:062> but then once that information has been processed that means you see it when it has been <unclear>
<XEC:#011:FC:063> that data has been transformed i-whereby you should have to look at it and be able to read and understand it
<XEC:#011:FC:064> that means <unclear> that means it has been processed
<XEC:#011:FC:065> whereby you will be able to read and understand it
<XEC:#011:FC:066> <xhosa> <question?> <xhosa>

<XEC:#011:SS:067> {yes}

<XEC:#011:FC:068> <xhosa> <explanation> <xhosa> <xhosa> <ne?> <xhosa>

<XEC:#011:SS:069> {yes}

<XEC:#011:FC:070> <xhosa> <explanation> <xhosa> <xhosa> <ne?> <xhosa>

<XEC:#011:SS:071> {yes}

<XEC:#011:FC:072> <xhosa> <explanation> <xhosa> understand?

<XEC:#011:SS:073> {yes}

<XEC:#011:FC:074> <xhosa> <explanation> <xhosa>

<XEC:#011:S:075> <unclear>

<XEC:#011:FC:076> <xhosa> <explanation> <xhosa>

<XEC:#011:FC:077> because n- now we can be able to read it and understand it
<XEC:#011:FC:078> whereas before it was a data it was long
<XEC:#011:FC:079> n- now we couldn't read it so that we could understand it
<XEC:#011:FC:080> only the person who has written it can be able to
understand that data
<XEC:#011:FC:081> do you see a difference between a data and information?

<XEC:#011:SS:082> {yes}

<XEC:#011:FC:083> <xhosa> <explanation> <xhosa>

<XEC:#011:SS:084> {yes <unclear>}

<XEC:#011:FC:085> <xhosa> <question?> <xhosa>

<XEC:#011:SS:086> {yes}

<XEC:#011:FC:087> a <cough> a data can be collected in various places
<XEC:#011:FC:088> that means <xhosa> <explanation> <xhosa>
<XEC:#011:FC:089> you can get it inside a business or outside a business
<XEC:#011:FC:090> it depends as to what is the kind of a data you want to get
<unclear> and outside the business
<XEC:#011:FC:091> that is internal source of data that means that information
that you can be able to obtain it inside the business
<XEC:#011:FC:092> that it <unclear> information that has to do with the
<unclear>
<XEC:#011:FC:093> that is a internal data
<XEC:#011:FC:094> but if potentially you want anything about <name of school>
<XEC:#011:FC:095> that means you cannot go and get it outside but you have to
get information

<XEC:#011:SS:096> {inside}

<XEC:#011:FC:097> inside <name of school> high school
<XEC:#011:FC:098> <xhosa> <question?> <xhosa> internal [source of data
---
<XEC:#011:SS:099> [{source of data}

<XEC:#011:FC:100> <xhosa> <explanation> <xhosa>
<XEC:#011:FC:101> i-information about <name of other school> <xhosa>
<question?> <xhosa>
<XEC:#011:FC:102> what is happening at <name of other school> high school?
<XEC:#011:FC:103> where can i get that information?

<XEC:#011:SS:104> <unclear>

<XEC:#011:FC:105> <xhosa> <question?> <xhosa>

<XEC:#011:S:106> {from outside}

<XEC:#011:FC:107> outside <name of school> <xhosa> <ne?> <xhosa>

<XEC:#011:SS:108> {yes}

<XEC:#011:FC:109> <xhosa> <comment> <xhosa> <xhosa> <ne?> <xhosa>

<XEC:#011:SS:110> {yes}

<XEC:#011:FC:111> so <unclear> there are two ways of collecting data
<XEC:#011:FC:112> you can collect the data [inside the business
---
<XEC:#011:SS:113> [{inside}

<XEC:#011:FC:114> or [outside the business

<XEC:#011:SS:115> [{outside}

<XEC:#011:FC:116> internal sources of data and [external sources of data

SAMPLE OF TRANSCRIBED DATA: FD

<XEC:#012:00:001> <general noise as class settles in>

<XEC:#012:FD:002> yesterday we were busy with plant tissues <xhosa> <ne?>
<xhosa>
<XEC:#012:FD:003> yesterday we were busy dealing with plant tissues
<XEC:#012:FD:004> <unclear> the plant tissue
<XEC:#012:FD:005> what is it a tissue?

<XEC:#012:SS:006> <unclear>

<XEC:#012:FD:007> a tissue is a groups of cell having the same function and
structure
<XEC:#012:FD:008> we discussed the different kinds of tissue
<XEC:#012:FD:009> name the type of tissues we discussed yesterday
<XEC:#012:FD:010> any type of tissues we discussed yesterday
<XEC:#012:FD:011> yesterday we managed to talk about three kind of tissues
<XEC:#012:FD:012> one the meristematic tissue
<XEC:#012:FD:013> two the epidermal tissue
<XEC:#012:FD:014> and what?

<XEC:#012:S:015> <unclear>

<XEC:#012:FD:016> the epidermis is also permanent tissue
<XEC:#012:FD:017> you have meristematic tissue and you also have permanent
tissue whereby the epidermal tissue is one of the permanent
tissues <xhosa> <ne?> <xhosa>

<XEC:#012:SS:018> [{yes}

<XEC:#012:FD:019> [so yesterday we discovered the epidermis which is part of
meristematic tissue
<XEC:#012:FD:020> what is the third one?

<XEC:#012:SS:021> <unclear>

<XEC:#012:FD:022> parenchymal tissue
<XEC:#012:FD:023> give two types of meristematic tissue
<XEC:#012:FD:024> ... name two types of

<XEC:#012:S:025> <unclear>

<XEC:#012:FD:026> we discussed apical meristem and?

<XEC:#012:S:027> <unclear>

<XEC:#012:FD:028> and also the secondary meristem
<XEC:#012:FD:029> what is another name for an apical meristem?

<XEC:#012:S:030> {primary meristem}

<XEC:#012:FD:031> it is also known as a primary st- meristem
<XEC:#012:FD:032> what are the functions of the epidermal tissue of the
epidermis?
<XEC:#012:FD:033> ... what are the functions of the epidermis?
<XEC:#012:FD:034> ... mm one

<XEC:#012:S:035> <unclear>

<XEC:#012:FD:036> for?

<XEC:#012:S:037> <unclear>

<XEC:#012:FD:038> can you please come again?

<XEC:#012:S:039> <unclear>

<XEC:#012:FD:040> for growth in length and in thickness no

<XEC:#012:FD:041> it is the meristematic tissue
<XEC:#012:FD:042> it is the meristematic tissue which is responsible for growth of plants in length and also in thickness
<XEC:#012:FD:043> i'm asking the function of the epidermis or epidermal tissue
<XEC:#012:FD:044> ... mm

<XEC:#012:S:045> {<cough> they are usually colourless}

<XEC:#012:FD:046> they are usually?

<XEC:#012:S:047> {colourless}

<XEC:#012:FD:048> they are usually colourless
<XEC:#012:FD:049> the function of the epidermis
<XEC:#012:FD:050> the function of the epidermis

<XEC:#012:S:051> <cough>

<XEC:#012:FD:052> ... what is the main function of the epidermis?

<XEC:#012:S:053> {it increase <unclear>}

<XEC:#012:FD:054> in the deep of woods of plants are the?
<XEC:#012:FD:055> meristematic cells or the
<XEC:#012:FD:056> meristematic tissue
<XEC:#012:FD:057> i'm talking about the epidermis
<XEC:#012:FD:058> and yesterday i said the epidermal tissue of roots is different from the epidermal tissue of stems and also the leaves because the function of the tissue in those three different organs is different
<XEC:#012:FD:059> but there is a main function of the epidermis
<XEC:#012:FD:060> epidermal tissue it covers and protects the inner tissues
<XEC:#012:FD:061> the epidermis it covers and also it protects other tissues
<XEC:#012:FD:062> ... it protects and also it covers other tissues
<XEC:#012:FD:063> we talked about the meri- the parenchymal tissue
<XEC:#012:FD:064> what are the functions of the parenchymal tissue?

<XEC:#012:S:065> <unclear>

<XEC:#012:FD:066> they store food substances such as starch and sugars?

<XEC:#012:S:067> <unclear>

<XEC:#012:FD:068> it contains what?

<XEC:#012:S:069> <unclear>

<XEC:#012:FD:070> they contain chloroplasts

<XEC:#012:S:071> {yes}

<XEC:#012:FD:072> not the parenchymal tissue
<XEC:#012:FD:073> the parenchymal tissue having chloroplasts is given another name
<XEC:#012:FD:074> what is it called?
<XEC:#012:FD:075> the parenchymal tissue having chloroplasts is given another term
<XEC:#012:FD:076> what is it called?

<XEC:#012:S:077> {chlorenchyma}

<XEC:#012:FD:078> it is called a chlorenchyma
<XEC:#012:FD:079> i am asking you questions the the functions of the parenchymal tissue
<XEC:#012:FD:080> one is the storage function whereby the parenchyma stores food substances such as sugars and starch
<XEC:#012:FD:081> second one

REFERENCES

Ackerman, G.J. (1975). Education for Blacks in historical perspective. South African Pedagogical Society (Publication Series No. 3).

Amuzu, K. (1992). "The Role of English in Education in the New South Africa". English Usage in Southern Africa. 23:129-141.

Atkins, S., Clear, J., and Ostler, N. (1992). "Corpus design criteria". Literacy and Linguistic Computing. 7:1:1-16.

Aijmer, K. and Altenberg, B. (Eds.). (1991). English Corpus Linguistics: Studies in Honour of Jan Svartvik. London and New York: Longman.

Bamiro, E.O. (1995). "Syntactic variation in West African English". World Englishes. 14:2:189-204.

Bantele, S.V., Davey, A.S., Lanham, L.W., Mahlasela, B., Mathiso, M., and Riordan, J. (1969). Lumko: Xhosa Self-Instruction Course. Lumko.

Barnbrook, G. (1996). Language and Computers: A Practical Introduction to the Computer Analysis of Language. Edinburgh: Edinburgh University Press.

Behr, A.L. (1970). "Bantu Education". Standard Encyclopedia of South Africa vol.2. Cape Town: Nasou.

Behr, A.L. (1984). New Perspectives in South African Education: A review of education in South Africa, 1652-1984 2nd edition. Durban/Pretoria: Butterworth Publishers (PTY) LTD.

Biber, D., Conrad, S., and Reppen, R. (1998). Corpus Linguistics: Investigating language structure and use. Cambridge: Cambridge University Press.

Brown, D. (1992). "Language and Social History in South Africa: A task still to be undertaken" in Herbert, R. (Ed.). Language and Society in Africa: The Theory and Practice of Sociolinguistics. p. 71-92. Wits: Witwatersrand University Press.

Buthelezi, Q. (1995). "South African Black English: lexical and syntactic characteristics" in Mesthrie, R. (Ed.). A Reader in South African Sociolinguistics. p. 242-250. Cape Town: David Philip.

Cameron (1985). Feminism and Linguistic Theory. London: Macmillan.

Cazden, C.B. (1988). Classroom Discourse: The Language of Teaching and Learning. Portsmouth, N.H.: Heinemann Educational Books, Inc.

Coates, J. (1993). Women, Men and Language: A sociolinguistic account of gender differences in language (2nd edition). London: Longman.

Cook, G. (1995). "Theoretical issues: transcribing the untranscribable" in Leech, G., Myers, G., Thomas, J. (Eds.). (1995). Spoken English on Computer: Transcription, mark-up and application. p. 35-53. New York: Longman.

Curzan, A. (2000). "English historical corpora in the classroom". Journal of English Linguistics. 28:1:77-90.

De Kadt, E. (1993). "Attitudes towards English in South Africa". World Englishes. 12:311-24.

De Klerk, V.A. (1996). "Use of and attitudes to English in a multilingual university". English World-wide. 17:1:111-127.

De Klerk, V.A. (2002). "Starting with Xhosa English...towards a spoken corpus". International Journal of Corpus Linguistics. 7:1:21-42.

De Klerk, V. (forthcoming). "Towards a norm in South African Englishes: the case for Xhosa English". World Englishes.

De Klerk, V. and Gough, D. (2002). "Black South African English" in Mesthrie, R. (Ed.). Language in Social History (2nd edition). p. 356-378. Cape Town: David Philip.

Eastman, C. (Ed). (1992). Code-switching. Cleveland/Philadelphia/Adelaide: Multilingual Matters Ltd.

Finegan, E. and Besnier, N. (1989). Language: Its structure and use. New York: Harcourt Brace Jovanovich, Inc.

Finn, S.M. (1986). "The abuse of English by Vista Students". English Usage in South Africa. 17:1:1-8.

Fromkin, V. and Rodman, R. (1993). An Introduction to Language (5th edition). Fort Worth: Harcourt Brace Jovanovice College Publishers.

Gough, D. (1994). "English People and People's English" in Reckwitz, E., Venarini, L. and Wegener, C. (Eds). The African Past and Contemporary Culture. p. 202-235. Essen: Die Blaue Eule.

Gough, D. (1996). "Black English in South Africa" in De Klerk, V. (Ed.). Varieties of English around the World: Focus on South Africa. p. 52-77. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Granger, S. (1996). "Learner English around the World" in Greenbaum, S. (Ed.). (1996). Comparing English Worldwide: The International Corpus of English. p. 13-24. Oxford: Clarendon Press.

Greenbaum, L. and Mbali, C. (2002). "An analysis of language problems identified in writing by low achieving first-year students, with some suggestions for remediation". Southern African Linguistics and Applied Language Studies. 20:233-244.

Greenbaum, S. (Ed.). (1996). Comparing English Worldwide: The International Corpus of English. Oxford: Clarendon Press.

Hartshorne, K. (1992). Crisis and Challenge: Black Education 1910-1990. Cape Town: Oxford University Press.

Hartshorne, K. (1995). "Language policy in African Education: a background to the future" in Mesthrie, R. (Ed.). Language and Social History: Studies in South African Sociolinguistics. p.306-318. Cape Town: David Philip.

Hockey, S. (2000). Electronic Texts in the Humanities: Principles and Practice. Oxford: Oxford University Press.

Izre'el, S. (2002). http://spinoza.tau.ac.il/hci/dep/semitic/cosih.html 15/01/02

Kachru, B. B. (1990). "World Englishes and applied linguistics" in Halliday, M.A.K., Gibbons, J., and Nicholas, H. (Eds.). Learning, Keeping and Using language. p. 203-230. Amsterdam: John Benjamins.

Kachru, B.B. (1992). (Ed.). The Other Tongue: English across Cultures (2nd edition). Urbana/Chicago: University of Illinois Press.

Kaschula, R. and Anthonissen, C. (1995). Communicating Across Cultures in South Africa: towards a critical language awareness. Johannesburg: Witwatersrand University Press.

Kearsley, G.P. (1976). "Questions and question-asking in verbal discourse: A cross-disciplinary review". Journal of Psycholinguistic Research. 5:355-375.

Kennedy, G. (1998). An Introduction to Corpus Linguistics. London and New York: Longman.

Knowles, G. (1995). "Converting a corpus into a relational database: SEC becomes MARSEC" in Leech, G., Myers, G., and Thomas, J. (Eds.). Spoken English on Computer: Transcription, mark-up and application. p208-219. New York: Longman Publishing.

Lanham, L. (1990). "Stress and Intonation and the Intelligibility of South African Black English" in Ramsaran, S. (Ed.). Studies in the Pronunciation of English: A Commemorative Volume in Honour of A.C. Gimson. p243-260. London: Routledge.

Lawler, J.M. and Dry, H.A. (Eds). (1998). Using Computers in Linguistics: A Practical Guide. London: Routledge.

Leach, G. (1989). South Africa: no easy path to peace. London: Methuen Mandarin.

Leech, G., Deuchar, M., and Hoogenraad, R. (1982). English Grammar for Today: A new introduction. London: The Macmillan Press Ltd.

Leech, G., Myers, G., and Thomas, J. (Eds.). (1995). Spoken English on Computer: Transcription, mark-up and application. New York: Longman Publishing.

Macdonald, C. (1990). Crossing the threshold into standard three. Report, Soling 16. Pretoria. Pretoria: HSRC.

Mafisa, P.J. (2002). "Grammatical Competence of ESL Teachers". Per Linguam. 18:1:15-26.

Makhobotloane, F.M. (1992). "Ratios and Types of Questions in the Communicative Class: Is there a predictable relationship?". English Usage in Southern Africa vol.23:105-128.

McEnery, T. and Wilson, A. (1997). Corpus Linguistics: Part one. Early Corpus Linguistics and the Chomskyan Revolution. Edinburgh: Edinburgh University Press.

Mc Carthy, M. (1998). Spoken Language and Applied Linguistics. Cambridge: Cambridge University Press.

Mesthrie, R. (1992). English in Language Shift: The History, Structure and Sociolinguistics of South African Indian English. Johannesburg: Wits University Press.

Mesthrie, R. (1995). (Ed.). Language and Social History: Studies in South African Sociolinguistics. Cape Town: David Philip.

Mesthrie, R. (2002). (Ed.). Language in Social History (2nd edition). Cape Town: David Philip.

Meyer, C.F. (2002). English Corpus Linguistics: An introduction. Cambridge: Cambridge University Press.

Mugoya, C. (1991). A study into the use of English across the curriculum in Transkei. Unpublished Honours essay. University of Transkei.

Nelson, G. (1995). "The International Corpus of English: mark-up for spoken language" in Leech, G., Myers, G., and Thomas, J. (Eds.). Spoken English on Computer: Transcription, mark-up and application. p 220-223. New York: Longman Publishing.

Nelson, G. (1996). "The Design of the Corpus" in Greenbaum, S. (Ed.). (1996). Comparing English Worldwide: The International Corpus of English. Oxford: Clarendon Press.

Nwaila, C. (1993). "English in South Africa: The Dilemma Facing Teachers". ELTIC Reporter. 17:2:3-10.

Payne, J. (1995). "The COBUILD spoken corpus: transcription conventions" in Leech, G., Myers, G., and Thomas, J. (Eds.). Spoken English on Computer: Transcription, mark-up and application. p 203-207. New York: Longman Publishing.

Peppé, S. (1995). "The Survey of English Usage and the London-Lund Corpus: computerising manual prosodic transcription" in Leech, G., Myers, G., and Thomas, J. (Eds.). Spoken English on Computer: Transcription, mark-up and application. p 187-202. New York: Longman Publishing.

Phillipson, R. (1992). Linguistic Imperialism. Oxford: Oxford University Press.

Platt, J., Weber, H., and Ho, M. (1984). The New Englishes. London: Routledge and Kegan Paul.

Preistler (1986). Linguistic Sex Roles in Conversation. Berlin: Mouton de Gruyter.

Radford, A. (1997). Syntax: A Minimalist Introduction. Cambridge: Cambridge University Press.

Rissanen, M. (2000). "The world of English historical corpora". Journal of English Linguistics. 28:1:7-28.

Rodseth, J.V. (1978). The Molteno Project Report: Mother-Tongue Reading Instruction and English Language Teaching in African Primary Schools. Ed by Driver, D.J. and de Villiers, A. Unpublished Report. Grahamstown ISEA: Rhodes University.

Rose, B. and Tunmer, R. (1975). Documents in South African education. Johannesburg: Ad Donker.

Rundell, M., and Stock, P. (1992). "The corpus revolution". 3-part article in English Today. 8:2:9-17, 8:3:21-32 and 8:4:45-51.

Scheffler, L. (1978). "Common Language Errors". English Usage in South Africa. 9:1:25-7.

Schlebush, A. (1994). Non-racial schooling in selected Cape Schools: Language, attitudes and language learning. Unpublished MPhil.: University of Cape Town.

Schmied, J. (1991). English in Africa: An Introduction. London: Longman.

Schmied, J. (1996). "Second-Language Corpora" in Greenbaum, S. (Ed.). Comparing English Worldwide: The International Corpus of English. p. 182-196. Oxford: Clarendon Press.

Smit, U. (2000). "Investigating BSAE in the new millennium - on the potential and problems of language attitutes research". South African Journal of Linguistics Supplement. 38:133-148.

Sridhar, K. and Sridhar, S. (1992). "Bridging the paradigm gap: second language acquisition theory and indigenized varieties of English" in Kachru, B. (Ed.). The Other Tongue: English across Cultures. p. 91-107. Urbana: University of Illinois Press.

Strevens, P. (1977). New Orientations in the Teaching of English. Oxford: Oxford University Press.

Stubbs, M. (1976). Language, schools and classrooms. London: Methuen and Co. Ltd.

Svartvik, J. and Quirk, R. (Eds). (1980). A Corpus of English Conversation. Lund: Lund University Press.

Titlestad, P. (1993). "President's Address". The English Academy Review. 10:191-197.

Van der Walt, C. (1999). "Justifying Their Existence: South African Varieties of English". The English Academy Review. 15:42-51.

Van der Walt, J.L. (2001). "Some implications of new English forms for language assessment". SAALT Journal for Language Teaching. 35:1:1-11.

Van der Walt, J.L. and Van Rooy, B. (2002). "Towards a norm in South African Englishes". World Englishes. 21:1:113-128.

Van Rooy, B. (n.d.). "Myths about Black South African English: evidence from a corpus". Unpublished manuscript.

Vine, B. (1999). Guide to the New Zealand Component of the ICE. Wellington: Victoria University of Wellington.

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.

Wade, R. (1998). "Arguments for Black South African English as a distinct `new' English". http://www.und.ac.za/und/ling/archive/wade-03.html

Walters, P. (1996). "Issues in English Teaching in Black Primary Schools" in De Klerk, V. (Ed.). Varieties of English around the World: Focus on South Africa. p211-230. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Williams, J. (1987). "Non-native varieties of English: a special case of language acquisition". English World-Wide. 8:2:161-191.

Wright, L. (1996). "The Standardisation Question in Black South African English" in De Klerk, V. (Ed.). Varieties of English around the World: Focus on South Africa. p149-162. Amsterdam/Philadelphia: John Benjamins Publishing Company.

BIBLIOGRAPHY

Abercrombie, M., Hickman, C.J., and Johnson, M.L. (1980). (Eds.). The Penguin Dictionary of Biology (7th edition). Harmondsworth: Penguin Books Ltd.

Brinton, L. J. (2000). The Structure of Modern English: A linguistic introduction. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Herbert, R. (Ed.). Language and Society in Africa: The Theory and Practice of Sociolinguistics. Wits: Witwatersrand University Press.

Holmes, J. (1996). "The New Zealand Spoken Component of ICE: Some Methodological Challenges" in Greenbaum, S. (Ed.). Comparing English Worldwide: The International Corpus of English. p. 163-181. Oxford: Clarendon Press.

Houghton-Hawksley, H.S. and Eaton, A.B.S. (1982). The Wild Wave: an anthology of poetry. London: John Murray Publishers Ltd.

Hunston, S. and Francis, G. (2000). Pattern Grammar: A corpus-driven approach to the lexical grammar of English. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Knowles, G., Wichmann, A., and Alderson, P. (Eds.). (1996). Working with Speech: Perspectives on research into the Lancaster / IBM Spoken English Corpus. London: Longman.

Ladefoged, P. (1975). A Course in Phonetics. New York: Harcourt Brace Javanovich, inc.

Scott, M. and Oxford University Press (1998). Wordsmith Tools Manual Version 3.0

Unsworth, L. (Ed.). (2000). Researching Language in Schools and Communities: Functional Linguistic Perspectives. London and Washington: Cassell.

Webb, V. (1996). "English and Language Planning in South Africa: the Flip-side" in De Klerk, V. (Ed.). Varieties of English around the World: Focus on South Africa. p175-190. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Young, S. and Bloothooft, G. (Eds.). (1997). Corpus-based Methods in Language and Speech Processing. Dordrecht/Boston/London: Kluwer Academic Publishers.

RHODES UNIVERSITY LIBRARY