# Determining an AAC core vocabulary for Zulu-speaking preschool children

**by**

**Jocelyn Ruth Mngomezulu**

**A dissertation submitted in partial fulfilment of the requirements for the degree**

**Master's in Augmentative and Alternative Communication**

**in the Centre for Augmentative and Alternative Communication**

**UNIVERSITY OF PRETORIA**

**FACULTY OF HUMANITIES**

**SUPERVISOR: Dr KM Tönsing**
**CO-SUPERVISORS: Prof S Dada, Ms NB Bokaba**

**June 2017**

# UNIVERSITY OF PRETORIA

## DECLARATION OF ORIGINALITY

This document must be signed and submitted with every essay, report, project, assignment, dissertation and/or thesis.

Full names of student:                          Jocelyn Ruth Mngomezulu

Student number:                                 04578440

**Declaration**

1. I understand what plagiarism is and am aware of the University's policy in this regard.

2. I declare that this dissertation is my own original work. Where other people's work has been used (either from a printed source, Internet or any other source), this has been properly acknowledged and referenced in accordance with departmental requirements.

3. I have not used work previously produced by another student or any other person to hand in as my own.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

SIGNATURE OF STUDENT:

SIGNATURE OF SUPERVISOR:

# ACKNOWLEDGEMENTS

It is the LORD our creator who gave us the gift of language and the words we use to pray and to speak with others. When words fail, the LORD intercedes for us by his Spirit. His name be praised for this gift of words, praise be also for his not making us entirely reliant on words.

Our parents teach us to talk, they teach us to use the words. We learn to build and not tear down. My thanks to my parents for teaching me and encouraging me in everything I do. A special thanks to my mum who taught me to write words down, and that in sensible sentences (most of the time).

My husband listened to many a word being tossed, bounced and even flung about in the process of this study. My thanks to him for his patience, encouragement and joy.

Other teachers along the way help steer us in the path. My thanks to my supervisors in this project who delicately and with immense perseverance carried out this steering.

The children, their parents, the preschools and educators who welcomed me into their spaces and allowed me to collect the words at the heart of this study cannot but be noted.

My thanks also to those who in their different ways pointed me in the direction of this research. Some offered their time and expertise; coaching, encouraging and giving knowledge. Others offered practical problems with their silence that got me to thinking for a solution. This thesis is dedicated to the silent speakers of this fascinating and beautiful language who would, with a borrowed voice, speak their mind.

# ABSTRACT

Persons who cannot rely on natural speech to fulfil their communication need to use augmentative and alternative communication (AAC) to enable them to participate in life activities. For persons who are nonliterate, the vocabulary for the AAC systems must be preselected. A core vocabulary approach proposes that vocabulary used with a high frequency in spoken language samples be included in AAC systems, in order to ensure that communication can take place across a variety of contexts, and to enable the generation of novel utterances. Various studies have recorded the speech of persons without disabilities in order to determine core vocabulary lists in a number of languages.

In South Africa, 22.7% of the population speak Zulu as a home language. There is to date no published Zulu core vocabulary list. This quantitative observational study therefore aimed to determine a core vocabulary for Zulu-speaking preschoolers.

Six children aged between 5;1 (years;months) and 5;9 from three different rural preschools participated in the study. The spontaneous speech of these participants was captured during typical preschool activities using small, body-worn digital voice recorders. A total of 9000 orthographic words were captured and transcribed. Due to the linguistic structure of Zulu, the sample was analysed using a formative tagging system, in order to capture the smallest meaningful parts of the language. By applying inclusion criteria of frequency counts of ≥0.5‰ and use by at least two participants (commonality score), a core vocabulary of 221 formatives was determined, which accounted for 88.9% of the total composite sample. The formatives in this core vocabulary are further described and, where possible, compared to the core vocabulary of previous studies. Implications for the design of Zulu AAC systems are discussed.

**Keywords:** Augmentative and alternative communication, Core vocabulary, Preschool children, Vocabulary selection, Zulu.

# ISIQOQO

Abantu abangakwazi ukukhuluma ngendlela ejwayelekile baye badinge ukusebenzisa i-Augmentative and Alternative Communication (i-AAC) ukuze nabo bezophila benze njengabantu bonke. Kulabo abangakwazi ukufunda nokubhala, kumele bakhethelwe amagama azofakwa ezinsizeni zabo (assistive devices) zokukhuluma. Isu lamagama asetshenziswa njalo njalo liphakamisa ukuthi kumele kufakwe wona ezinsizeni zokukhuluma ikhona inkulumo izophumelela nomaphi futhi ukuze oyisebenzisayo ezokwazi ukuzikhulumela azakhele nemisho emisha. Ucwaningo olahlukahlukene seluye lwaqopha izinkulumo zabantu abazikhulumelayo ukuze kutholakale izinhlu zamagama asetshenziswa njalo njalo ezilimini ezahlukeneyo.

ENingizimu Afrika, ngama-22.7% ezakhamuzi akhuluma isiZulu njengolimi lwebele. Kuze kube yinamhlanje alukho uhla lwamagama esiZulu asetshenziswa njalo njalo oseluke lwashicilelwa. Ngakho-ke lolu cwaningo lokubheka ngezibalo luhlose ukuthola uhlelo lwamagama asetshenziswa njalo njalo yizingane ezikuzinkulisa ezikhuluma isiZulu.

Yizingane eziyisi-6 ezangena kulolucwaningo ezazineminyaka yobudala engaphakathi kwe-5;1 (iminyaka;izinyanga) ukuya eminyakeni yobudala emi-5;9. Ezazisuka ezizinkuliseni zasemakhaya ezintathu ezahlukene. Zazifakwe izikhwanyana ezaziphethe iziqophamazwi, zaqoshwa zizikhulumela zingagqutshwa muntu, zibe zenza zonke izinto ezijwayele ukuzenza umangabe zazikunkulisa. Abhalwa phansi amagama ayekhulunywa izingane aze afika ku-9000. Ngenxa yohlelo lolimi lwesiZulu, kwaphoqa ukuthi kusetshenziswe indlela yokubala amagama eyayizowahlaziya ngokwezakhi zawo. Isakhi ngasinye ukuze sibalwe njengegama elisetshenziswa njalo njalo kwakumele sivele ngokuphindaphinda ngaphezu kuka-0.5‰ (≥0.5‰), siphinde futhi sisetshenziswe izingane okungenani ezimbili. Ngalendlela kwatholakala izakhi ezingu-221 athathwa njengamagama asetshenziswa njalo njalo. Lezozakhi futhi zatholakala ukuthi zibambe inxenye engu-88.9% yayo yonke inkulumo yezingane eyayiqoshiwe. Izakhi ezikuloluhlelo ziyaqhubeka zichazwe, ziqhathaniswe futhi lapho okunokuqhathaniswa khona nohla lwezingcwaningo ezedlule. Umqondo ovezwa izakhi nawo uyahlaziywa mayelana nokusungulwa kwezinsizi zokukhuluma ezisosebenzisa isiZulu.

**Amagama asemqoka:** Izingane zasezinkulisa, Amagama asetshenziswa njalo njalo, *i-Augmentative and Alternative Communication*, isiZulu, Ukhetho lwamagama.

# OPSOMMING

Persone wat nie op natuurlike spraak kan staatmaak om hulle kommunikasiebehoeftes te vervul nie, moet gebruik maak van aanvullende en alternatiewe kommunikasie (AAK) om hulle in staat te stel om aan lewensaktiwiteite deel te neem. Vir persone wat nie geletterd is nie, moet die woordeskat vir die AAK-sisteem vooraf gekies word. 'n Kernwoordeskat-benadering stel voor dat woordeskat, wat met hoë frekwensie in gesproke taal gebruik word in AAK-sisteme ingesluit word, om te verseker dat kommunikasie oor 'n verskeidenheid kontekste kan plaasvind, en om nuwe uitinge te genereer. Verskeie studies het die spraak van persone sonder gestremdheid opgeneem om kernwoordeskatlyste in verkeie tale te bepaal.

In Suid-Afrika besig 22.7% van die bevolking Zulu as huistaal. Daar is tot op hede geen gepubliseerde Zulu kernwoordeskatlys nie. Hierdie kwantitatiewe waarnemingstudie het dus gepoog om 'n kernwoordeskat vir Zulu-sprekende voorskoolse kinders te bepaal.

Ses kinders tussen die ouderdomme van 5;1 (jare;maande) en 5;9 van drie verskillende landelike voorskole het aan die studie deelgeneem. Die spontane spraak van die deelnemers is gedurende tipiese voorskoolse aktiwiteite opgeneem, deur gebruik te maak van klein, digitale bandopnemers wat op die liggaam gedra word. 'n Totaal van 9000 ortografiese woorde is opgeneem en getranskribeer. Weens die linguistiese struktuur van Zulu is die steekproef deur 'n formatiewe koderingsisteem geanaliseer, om die kleinste betekenisvolle dele van die taal vas te lê. Deur insluitingskriteria van frekwensietellings van ≥0.5‰ en gebruik deur ten minste twee deelnemers (algemeenheidstelling) aan te wend, is 'n kernwoordeskat van 221 formatiewe vasgestel, wat 88.9% van die totale saamgestelde steekproef gedek het. Die formatiewe in hierdie kernwoordeskat word verder beskryf, en, waar moontlik, vergelyk met die kernwoordeskat van vorige studies. Implikasies vir die ontwerp van Zulu AAK sisteme word bespreek.


**Sleutelwoorde:** Aanvullende en alternatiewe kommunikasie, Kernwoordeskat, Voorskoolse kinders, Woordeskatseleksie, Zulu.

# CONTENTS

ix

**LIST OF TABLES**

**LIST OF FIGURES**

## LIST OF APPENDICES

# CHAPTER 1

# PROBLEM STATEMENT

## 1.1 Introduction

This chapter orientates the reader to the study. The purpose for conducting the research is outlined in the problem statement. The terminology and abbreviations which are used in the study are described and the contents of the chapters are sketched.

## 1.2 Problem statement and rationale

This research project is concerned with identifying core vocabulary used by typically developing first language (L1) Zulu-speaking preschool children for inclusion in graphic symbol-based Augmentative and Alternative Communication (AAC) systems.

AAC systems allow persons with little or no functional speech (LNFS) to express themselves through different modes. The abilities of the person in need of AAC will determine which of these different modes of communication the person might use. For example, a person who is literate but has LNFS could use an alphabet board to spell the words they wish to convey during an interaction, while those who are preliterate or not literate may be able to use graphic, symbol-based systems. Graphic symbol systems, however, do not typically allow a person to generate their own words but require vocabulary to be preselected.

Deciding what vocabulary to include in a graphic symbol-based aided AAC system is one of the most fundamental aspects to consider for its successful implementation, since the vocabulary determines how and to what extent any system may be used (Trembath, Balandin, & Togher, 2007). For the young AAC user who is not literate, it is often the service providers and parents who make decisions pertaining to vocabulary (Fallon, Light, & Paige, 2001). Selecting appropriate vocabulary is a difficult process as it entails predicting what words the person will need in various situations. Different approaches have been taken to the selection process, such as a developmental, an environmental, a functional, and a core vocabulary approach (Banajee, Dicarlo, & Stricklin, 2003). Each approach has its applications and circumstances to which they

1

are better suited (Beukelman & Mirenda, 2005). Authors suggest the use of multiple methods as a way to ensure a more comprehensive selection (e.g. Beukelman, Jones, & Rowan, 1989; Fried-Oken & More, 1992). This research will focus on core vocabulary for the reasons explained below.

Core vocabulary approaches are based on research evidence that a small set of approximately 250 words make up 80% of spoken communication (e.g. Beukelman, Jones, & Rowan, 1989; Fallon, Light, & Paige, 2001; Trembath, Balandin, & Togher, 2007). This small set of words therefore provides extensive semantic coverage (Baker, Hill, & Devylder, 2000). At the same time, however, core vocabulary accounts for a small portion of an individual's repertoire; although used less frequently than core, the remaining 20% is comprised of a great variety of words which are specific to the context and content of communication, as well as the person using them (Fallon et al., 2001; Trembath et al., 2007). This vocabulary has been termed fringe vocabulary (Trembath et al., 2007). Thus, fringe vocabulary is vast, infrequently used and highly personalised. In contrast, core vocabulary is a relatively small set of words which has been found to remain similar across ages, individuals and environments (Witkowski & Baker, 2012). Arguably, therefore, when included in an AAC system, these core words give the user access to a large portion of communication content while using relatively little space. Given the space and memory restrictions of AAC systems (Beukelman & Mirenda, 2005) and the physical and cognitive access demands that large directories of vocabulary place on AAC users (Light & Lindsay, 1991), it is understandable that core vocabulary approaches have received great interest in the AAC community.

It should however be noted that many of these core vocabulary directories have primarily been developed for English (e.g. Banajee et al., 2003; Beukelman et al., 1989; Fallon et al., 2001; Fried-Oken & More, 1992; Trembath et al., 2007). More recently, research has extended into other languages, for example, Mandarin Chinese (Liu & Sloane, 2006), German (Boenisch & Sachse, 2007) and Korean (Shin & Hill, 2016).

The use of AAC systems is supported in South African education policy documents (Department of Education, 2007, 2010). However, provision of appropriate AAC systems is complicated by multilingualism. The population of South Africa is characterised by a wide

range of languages (Statistics South Africa, 2015). While English is widely used in South Africa, it is home language to only 9% of the population (Census Stats SA, 2011). Zulu is the most widely spoken first language, spoken by 22.7% of the population. There is a greater concentration of first language speakers – 77.8% going up to more than 95% in some areas – in the KwaZulu-Natal province (Census Stats SA, 2011).

Language policy and the use of different languages in schools is a highly contentious and complex issue in South Africa (Bristowe, Oostendorp, & Anthonissen, 2014; Chick & Mckay, 2001; Ngcobo, 2013), though it remains in the school's authority to decide both on a language of instruction as well as on the language policy of the school as a whole (Department of Education, 1997). It must be noted that there exists a strong notion that English is the lingua franca and as such, it is desirable for all South African children to learn and become fluent in English (Bristowe et al., 2014; Chick & Mckay, 2001; Ngcobo, 2013; Prinsloo, 2007). In fact, whether by right or not, proficiency in English paves the way to decent employment and tertiary education (Bristowe et al., 2014). However, to moderate against the use of any one language exclusively, there is also a significant body of research which stands against the abandonment of home language fluency, citing not only lower academic outcomes but also detrimental social and emotional effects (Prinsloo, 2007; Soto & Yu, 2014). Indeed, when describing the functions of communication, Light and colleagues on various occasions spoke not only of the relaying of information and one's own needs and wants, but also the social functions of etiquette and those required for forming social closeness, which are embedded in communication (Light, 1989, 1997; Light & McNaughton, 2014). The language we use is closely linked to constructs of society and identity (Bristowe et al., 2014). The language we use then, while playing a part in identifying each individual and forming part of a social identity, also gives to us the schemas of personal and relational identity (Burgess, 2002). Being so closely linked with identity then, the discussion of language is always a personal one, which may well add further to the complexity of the language policy debate. This, along with South Africa's history of segregation, deeply complicates decision-making with regards to language policy (Ngcobo, 2013).

Nonetheless, multilingual communities are the norm in South Africa (Census Stats SA, 2011; Chick & Mckay, 2001). Furthermore, we understand that children who are bilingual demonstrate better academic and psychosocial outcomes when both languages are supported (de

3

Sousa, Greenop, & Fry, 2010; Soto & Yu, 2014). Given the picture of South African society (Census Stats SA, 2011), we must assume that the children requiring AAC are likely already exposed to more than one language. Therefore, accepting the above, it stands to reason that the support of both (or all) languages known to the child in AAC should be the norm rather than the exception. However, in practice, while AAC systems that give access to a language other than English may be introduced, vocabulary selection and organisation are typically informed by practices and/or research conducted in English.

Finally, then, the question arises as to the appropriateness of simply translating core vocabulary lists from English to another language. Although a core vocabulary list for French has been found comparable to English (Robillard, Mayer-Crittenden, Minor-Corriveau, & Belanger, 2014), fundamental differences between Eastern and Western languages have been found to impact on core vocabulary in Mandarin Chinese (Baker & Chang, 2006; Liu & Sloane, 2006) as well as Korean (Shin & Hill, 2016). This might suggest that translating may not always be a valid option. To date, there is no known publication regarding an investigation into the core vocabulary of African[1] languages to which group Zulu belongs. This research addresses this question and seeks to fill the gap in current evidence to and attempts to establish a core vocabulary for Zulu.

**1.3 Terminology**

Here following is a list of terms that are frequently used in this study, together with the definition that is adopted for this study.

*1.3.1 Aided AAC*

A strategy in which an external aid (something tangible in addition to the person's own body) is used to represent, select or transmit messages (Lloyd et al., 1997).

---

[1] Linguistically, Zulu is classified as S42; a Nguni Language, as part of the Southern Bantu languages in Guthries' classification (Gowlett, 2006). These, in turn, are part of the Bantu language family, which are classified as Southern Bantoid languages, which, in turn, form a branch of the Benue-Congo languages of the Niger-Congo (Nurse & Philippson, 2006). Since the terms 'Bantu' and 'Bantoid' carry derogatory connotations in the South African context, they will not be used in this thesis.

### *1.3.2 Augmentative and alternative communication*

Augmentative and alternative communication (AAC) refers to the research and practice that investigates, attends to and, where necessary, compensates for transient or permanent impairments, restrictions to activity and participation of persons with severe disorders of speech and language production and/or understanding, affecting spoken and written modes of communication (ASHA, 2005).

### *1.3.3 Commonality score*

This refers to the number of participants who use the formative in the total sample. The study involved six participants; therefore, the highest commonality score attainable is 6, which would mean that all the participants involved used this formative at least once.

### *1.3.4 Communication*

Communication is the deliberate sharing of meaning or transmission of messages between people (Tönsing, Alant, & Lloyd, 2005).

### *1.3.5 Core vocabulary*

Words and vocabulary items which are commonly used by a variety of individuals and occur with high frequency (Banajee et al., 2003; Beukelman & Mirenda, 2013; Trembath et al., 2007). In this study, core vocabulary was defined as formatives (see definition in Section 1.3.11) that occurred with a frequency of 0.5‰ or more in the language sample collected, and which had a commonality score (see Section 1.3.3) of more than 1.

### *1.3.6 First language (L1) and second language (L2)*

L1 and L2 refer to the languages learnt first and subsequently by persons who acquire multiple languages sequentially (Kay-Raining Bird, Genesee, & Verhoeven, 2016).

### *1.3.7 Frequency per mille (‰)*

This refers to the measure applied to each formative to determine the frequency with which the formative occurs in the total sample. Frequency per mille (‰) is calculated by dividing the total number of occurrences by the total sample and multiplying by one thousand.

5

### 1.3.8 Fringe Vocabulary

Words and messages which are particular to the individual and context and which tend to occur with lower frequency (Banajee et al., 2003; Beukelman & Mirenda, 2013; Trembath et al., 2007).

### 1.3.9 Grade R

Grade R (referring to 'reception') is the year before children start their first formal year of schooling in South Africa. Grades R to 3 form the foundation phase of the CAPS curriculum; however, this preschool year is not compulsory.

### 1.3.10 Graphic symbols

This refers to symbols which present in a static visual form, having either been drawn or pictured. Such symbols are typically displayed on paper or on a screen. The person using them selects the desired symbols from the array  (Sevcik, Romski, & Wilkinson, 1991).

### 1.3.11 Linguistic formatives

The Zulu language is comprised of a variety of formatives which may be combined according to specific rules of the language to form words according to different parts of speech, moods and context (Doke, 1939; Nyembezi, 1982; Taljaard & Bosch, 1993). This is a common feature among most languages of its kind (Nurse & Philippson, 2006). Formatives have been categorised and described by various grammarians (Doke, 1939; Nyembezi, 1982; Taljaard & Bosch, 1993). Broadly speaking, formatives may be classified as content or structure formatives. Structure formatives are formatives which allow Zulu word and sentence construction, while content formatives refer to the 'things' (content) spoken about. In this study, formatives were classified according to the nine types identified by Doke (1939), with an additional four types added (see sections 2.8.1, 2.8.2 and 3.7 for further details). Formative types were also correlated to the part of speech they helped to form (displayed in Table 4.4, Section 4.5.1).

### 1.3.12 Persons with little or no functional speech (LNFS)

Persons with little or no functional speech are able to speak fewer than 15 intelligible words (Bornman, 2015). This difficulty or inability to produce natural speech typically leads to activity limitations and participation restrictions (Bornman, 2015).

### 1.3.13 Symbol

A symbol is something which represents another thing or message (Lloyd, Fuller, & Arvidson, 1997) Various symbols, including spoken words, sounds, movements, gestures, manual signs, pictures, drawings, and written words, are employed in AAC for communication (Murray & Goldbart, 2009a).

### 1.3.14 System for Analysing Language Transcripts software program

The System for Analysing Language Transcripts (SALT) program (Miller & Iglesias, 2012) is designed to transcribe and analyse language samples. The program provides specific transcription codes that can be used to analyse the language sample.

### 1.3.15 Ukwabelana Corpus

This is an open source morphological tagging system for Zulu developed by Spiegler, Van der Spuy and Flach (2010). It addressed the morphological analysis of Zulu words by first segmenting them into morphemes and then tagging each morpheme with a code which depicts the part of speech being used, as well as the noun class, where applicable.

## 1.4 Notations

In this thesis, linguistic examples in Zulu are italicised. English translations of these examples, as well as all additional English linguistic examples, are given within single quotation marks.

## 1.5 Abbreviations

AAC: Augmentative and alternative communication

LNFS: Little or no functional speech

L1: First language or home language

L2: Second additional language

SALT: Systematic Analysis of Language Transcripts

TNF: Total number of formatives

7

Chapter 1: Problem Statement

TNDF: Total number of different formatives

## 1.6 Outline of chapters

Chapter 1 orientates the reader to the rationale for the study. This is followed by a description of the terminology and abbreviations used in the study. The contents of the chapters in the thesis are sketched.

Chapter 2 describes how graphic symbols are used in AAC, as well as the features pertaining to different types of symbols used in AAC. An overview of approaches to vocabulary selection is given and literature related to the core vocabulary approach in the field of AAC is discussed. The South African linguistic context and specifically the Zulu language are also explored. Corpus research of Zulu is discussed and applied to current AAC vocabulary selection principles. The chapter ends with a rationale for choosing an observational research design to identify Zulu core vocabulary rather than translating English core vocabulary.

Chapter 3 describes the research methodology, beginning with the main aim, sub-aims and the research design utilised. The setting is described, followed by a description of the participants (recruitment, selection criteria and descriptive criteria). Materials and equipment used are described next. The aims, procedures and results of the pilot study are given, and implications for the main study are discussed. The procedures for data collection and analysis are described, including ethical considerations and considerations regarding reliability and validity.

Chapter 4 presents the results of the study. The results are presented in graphs and tables in accordance with the five sub-aims of the study. First, the linguistic formatives used by Zulu-speaking preschoolers are described. Second, the number and types of different formatives used by the preschoolers are described. Third, a core vocabulary is identified by describing the most frequently and commonly used formatives. Fourth, the characteristics of core vocabulary are described by formative type, as well as by content and structure formatives (with subtypes). Fifth, the lexical meaning of the content formatives in the core vocabulary is examined in relation to previous studies.

Chapter 5 presents the discussion of the results of the study. The parameters (total number of formatives, total number of unique formatives) of the total Zulu language sample collected from Zulu-speaking preschoolers are explored and compared to other studies. Similarly, the parameters of the core vocabulary (frequency per mille and commonality scores) identified in this study are compared to those of previous studies. Furthermore, the characteristics of the core vocabulary determined in this study are discussed, and the lexical meanings are compared with English core vocabulary lists. Finally, the implications that a morphologically rich language such as Zulu has for the selection of vocabulary and design for augmentative and alternative communication (AAC) systems are discussed.

Chapter 6 provides an overall summary of the study. The study is critically evaluated and implications for clinical practice are discussed. Recommendations for future research are also provided.

**1.7 Summary**

This chapter orientated the reader to the study. Having outlined the purpose for the study in the problem statement, the terminology and abbreviations which are used in the study were described. The upcoming chapters were then briefly outlined.

**CHAPTER 2**

**LITERATURE REVIEW**

**2.1 Introduction**

This chapter describes how graphic symbols are used in AAC, as well as the features pertaining to different types of symbols used in AAC. An overview of the approaches to vocabulary selection is given, and literature relevant to the core vocabulary approach in the AAC field is reviewed. The South African context and specifically the Zulu language is also explored. Corpus research of Zulu is discussed and applied to current AAC vocabulary selection principles. The chapter ends with a rationale for choosing an observational research design to identify Zulu core vocabulary rather than translating English core vocabulary.

**2.2 Graphic symbols and AAC**

Augmentative and Alternative Communication (AAC) is concerned with the study and use of ways to communicate through additional or alternative means to replace or supplement natural speech (Beukelman & Mirenda, 2013). Necessarily this means using symbols other than the phonemes employed during speech for encoding messages. A variety of symbols have been used in AAC intervention, catering for a diversity of abilities and needs of persons in need of AAC and their communication partners.

At its very essence, a symbol represents a referent (Bornman & Tönsing, 2011). The example below (Figure 2.1) demonstrates a variety of symbols representing the concept of protective clothing typically worn on the feet, a shoe.

10
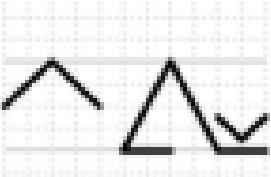
| The spoken word | The written word | Graphic symbols (also termed line drawings) | | South African Sign Language |
|---|---|---|---|---|
| | | Picture communication symbol (PCS)® | Blissymbolics® | |
| *"shoes"* /'ʃuːz/ | **Shoes** | | | |

Figure 2.1. *Various types of symbols used to represent a referent.*

Symbols can be described in terms of their characteristics. In the example above, the spoken word and the manual sign are considered unaided symbols. These are symbols that do not require aids but are produced using only the person's body (Bornman & Tönsing, 2011). The written word and the line drawings – such as Picture Communication Symbols (PCS) and Blissymbols – may be considered aided as they require aids external to the body (Bornman & Tönsing, 2011). Aided symbols are typically selected for use from an array. The exception to this is traditional orthography, which can both be produced (i.e. hand-written) and selected; whether through typing on a keyboard or selecting words and phrases from an array (Beukelman & Mirenda, 2005).

There has been much discussion over the role symbols are expected to play in a communication exchange (e.g. McNaughton, 1993; Murray & Goldbart, 2009; Smith, 2006; Sutton, Soto, & Blockberger, 2002). The bulk of the discussion necessarily compares the symbols of AAC with the symbols of spoken language which they seek to augment or replace, and a further classification of symbols in terms of linguistic characteristics has been proposed. Smith (2006) compared graphic symbols to linguistic symbols of speech, questioning if and how graphic symbols may be expected to fulfil the role of speech for those using them. Smith outlined the fundamental features of linguistic signs as their arbitrary nature (there exists no actual relationship between the speech sound and its referent), their segmentability (linguistic signs may be segmented into meaningless sections and recombined to form new and different

11

meanings, which is sometimes referred to as duality of patterning) and, finally, that linguistic signs are producible by the human body (Smith, 2006). Since graphic symbols are selected from an array rather than produced, and fundamentally attempt to draw resemblance either visually or conceptually to their referent (a feature described as iconicity), they do not share these features of linguistic signs (Smith, 2006). Interestingly, the symbols typically used in AAC do vary in terms of segmentation potential (Smith, 2006). Some symbol systems make use of rule-governed combinations of component parts to create new symbols. The symbols created are therefore segmentable (Smith, 2006). For example, among a host of rules, Blissymbolics uses a circumflex accent (i.e. ^ ) to indicate an action, which can be manipulated further for additional moods and tenses to an action (Blissymbolics Communication International, 2014). The smallest segmentable part, however, is still meaningful unlike the spoken phoneme, and thus Smith (2006) finds that Blissymbols still do not fully comply with the requirement for dual patterning. McNaughton (1993) also made use of this duality of patterning principle when she described Type 1 and Type 2 symbols. Where Type 1 symbols are iconic gestalts that rely on visual matching, for example, the image of a person jumping to represent "jump", Type 2 symbols may rely on phonological or semantic decoding of the composing parts of a symbol to decode the meaning, for example, the written word 'jump' (McNaughton, 1993).

Many widely-used symbol systems (e.g. PCS, Widget) are single-meaning, two-dimensional representations, drawings or pictures, which aim to represent words, phrases or concepts in a visual way (Smith, 2006). These graphic symbols are not segmentable or composed of a small set of components, and therefore, each lexical item available to the person using the system needs to be represented by a unique picture with a unique physical or virtual location (e.g. on a speech generating device or communication board). The person using the system can then select the graphic symbol (for example by pointing at it) or select a sequence of these in order to convey an intended message.

## 2.3 Vocabulary selection and the development of the core vocabulary approach

Since the person using graphic symbols cannot compose or generate their own words, the vocabulary (and thereby the ideas, concepts and messages) represented by these symbols must be selected for inclusion in any AAC array.

Chapter 2: Literature Review

Children who need AAC may have the same vocabulary needs as their typically developing peers (Beukelman, McGinnis, & Morrow, 1991). However, children aged five years typically have an expressive vocabulary of at least 2 000 words and may understand many more words (Lane & Molyneaux, 1992). To provide a graphic symbol-based vocabulary of even 2000 words would pose an overwhelming memory demand on both the child who needs AAC and their support team, as the location and navigation to each symbol would have to be memorised (Light & Lindsay, 1991; Thistle & Wilkinson, 2013) In addition, there may well be physical space restrictions to the AAC system itself (Light & Lindsay, 1991) based on access needs or portability requirements (Fuller & Lloyd, 1997), particularly with regards to non-electronic systems. A process therefore of choosing certain vocabulary from a pool of all possibilities (Yorkston, Dowden, Honsinger, Marriner, & Smith, 1988) must ensue to narrow down the array to a manageable but useful size. Fried-Oken and More (1992) contend that an initial vocabulary should ideally be meaningful, functional, motivating and individualised.

To ensure an individualised vocabulary for preliterate children in need of AAC, therapists, parents and significant others may engage in the task of selecting this vocabulary for them (Banajee et al., 2003; Fallon et al., 2001; Fried-Oken & More, 1992). Moving away from a trial and error approach based on experience and intuition (Fried-Oken & More, 1992; Thistle & Wilkinson, 2015), various methods of selecting vocabulary for graphic symbol-based AAC systems have been described in literature. These include environmental inventories, informant lists and core vocabulary approaches.

Environmental inventories may offer a way in which to gain an understanding of the individual's context and therefore their vocabulary needs. Sigafoos and York (1991) describe the compilation of environmental inventories as a process whereby all the individual's contexts, routines and activities are catalogued and vocabulary pertaining to these is compiled. From this extensive list, a smaller set of words is then selected, leaving the rest for later introduction. However, as demonstrated by Fried-Oken and More (1992), this vocabulary would still be based on what the interventionist thinks is important for these contexts and may not necessarily reflect the vocabulary that a person would actually use. Additionally, many persons who need AAC experience limited opportunity to participate in various contexts  (Beukelman & Mirenda, 2005). An inventory of their environment may therefore be an insufficient source of information.

Chapter 2: Literature Review

In order to structure the selection of vocabulary by informants, literature suggests lists of vocabulary determined by parents/other adults can be obtained using i) a blank sheet method (Morrow, Mirenda, Beukelman, & Yorkston, 1993), whereby informants are requested to write words freely on a blank sheet; ii) checklists (Fallon et al., 2001; Morrow et al., 1993), whereby informants are offered a list of words from which they select vocabulary that they think is applicable; and iii) a categories method where words are given in certain categories such as words pertaining to toys or pertaining to daily routines, and the informant is guided to consider vocabulary along these lines (Fallon et al., 2001). Lists of words are then generated which could be used to create a highly individualised vocabulary for the person who needs AAC.

While ecological inventories and informant lists seem to offer a client-centred way in which to determine vocabulary needs, the time required to do this for every person may become prohibitive to them beginning to use AAC (Fallon et al., 2001). Another problem with this approach may lie in the accuracy of those tasked with prediction. In a study which examined whether professionals who were supporting adults with severe communication impairments would be able to predict the topics and key words of conversations held in the workplace, Balandin and Iacono (1998) found that the professionals were neither able to accurately predict topics nor key vocabulary actually used in the conversations. The participants in Balandin and Iacono's (1998) study were experienced in the field, but this is not always the case, and those to whom the task of vocabulary selection falls may have little or no experience in selecting vocabulary (Banajee et al., 2003; Fallon et al., 2001; Fried-Oken & More, 1992). They may be ill-equipped for the task (Trembath et al., 2007), since the challenges of selecting appropriate vocabulary are marked.

In addition to these challenges, these methods have an inherent bias toward the selection of content words and the expression of needs and wants (Morrow et al., 1993). It has been suggested that this kind of source be used only for fringe vocabularies, while core vocabulary should be obtained from composite lists (Yorkston et al., 1988; see also discussion below).

The third method for selecting vocabulary, therefore, is the use of vocabulary lists generated from transcriptions of speech samples of speakers without disabilities. The use of the most frequently used words from such lists then is based on the conventional wisdom that if a

vocabulary item is used frequently it must be important (Morrow et al., 1993), since it provides significant semantic coverage (Baker et al., 2000). Also, it has been argued that the very characteristics of the frequently used 'core' words offer a framework on which to build further vocabulary and increasingly complex communication (Banajee et al., 2003), as will be demonstrated below.

The attraction of the relatively reduced number of words aside, it is important to consider the nature of the words typically found in core vocabulary lists in contrast with language as a whole. The words which feature on these high frequency lists typically relate to the structure and syntax of the language rather than carrying meaning themselves – the latter is typically the function of fringe vocabulary (Robillard et al., 2014). High frequency words are often prepositions, conjunctions, pronouns, and auxiliary verbs (Trembath et al., 2007), and give structure to language (Banajee et al., 2003). Fringe vocabulary, on the other hand, provides the content words necessary to establish meaning (Robillard et al., 2014). Although studying core vocabulary, many authors argue against the provision of core vocabulary alone but highlight the necessity of core vocabulary as the framework for communication (e.g. Banajee et al., 2003; Robillard et al., 2014; Trembath et al., 2007; Yorkston, Dowden, Honsinger, Marriner, & Smith, 1988).

The provision of core as well as fringe vocabulary, therefore, can ensure access to a system that is linguistically sophisticated (Baker & Chang, 2006) – through the presence core vocabulary the user is able to generate sentences (Liu & Sloane, 2006). Indeed, authors have reported the presence of core together with fringe vocabulary in AAC systems as a feature which directly enhances the frequency with which people who need AAC actually use their AAC system (Banajee et al., 2003; Beukelman et al., 1991).

Such a system, therefore, would reduce the risk of misunderstandings, allowing the user to access subtle shades of meaning achieved with syntactical manipulation (Sutton et al., 2002). To illustrate this phenomenon let us reflect on the words 'tree', 'house' and 'Durban', which are considered fringe vocabulary. Uttered alone, these words are very clear but the meaning of their use is not. When used together with core vocabulary, the meaning becomes clear (core vocabulary underlined); "I went into a tree house in Durban"; alternatively, "It is like the tree at my house in

Durban". Thus, we may note that the use of both core and fringe vocabularies is beneficial in AAC systems, which use preselected vocabulary as a way to enable the user to create novel generative utterances. Extensive work has been done in the AAC field which has led us to understand these vocabularies and how they influence communication and furthermore how vocabulary might be handled in AAC systems.

## 2.4 The core vocabulary approach in AAC

A literature search was conducted on the topic of the core vocabulary approach to AAC vocabulary selection. The keywords 'AAC', 'vocabulary selection' and 'core vocabulary' were used as search terms in the search engine EBSCOhost®, which gave rise to 12 key articles ranging from 1984 to 2016. Ancestry searches and forward citations searches from these articles were conducted to identify relevant articles (Schlosser, Wendt, Bhavnani, & Nail-Chiwetalu, 2006), and a further 39 potential articles were found. The literature reviews and citation lists for these articles were then explored for further publications on the theme and these were sourced. This procedure was repeated until the procedure had exhausted itself and no new publications were identified. The abstracts of the articles were explored and a total of 23 articles were found to be relevant in that they either directly determined a core vocabulary for AAC or spoke to using core vocabulary as a vocabulary selection strategy. Two foundational articles were also included from outside the field of AAC, as they pioneered the investigation of core vocabulary in spoken language sample analysis. The results of the search are presented in Appendix A. Selected articles reveal an historical pattern to the development of this approach, which will be explained below.

Foundational articles referred to research conducted outside of the field of AAC, in fields of language learning and communication disorders. In the sixties, Howes (1966) and Berger (1967), noticing the difference between spoken language and written, began to record and analyse spoken English. Both authors found a pattern of high frequency core and low frequency fringe words. From a large number of total words (250 000 and 25 000 words each); a relatively small number of unique words (9 699 and 2 507 respectively) were found. Approximately half of these unique words were used only once throughout their samples. This phenomenon had been under investigation since the 1930s and has been exploited by various curricula for second language learning on the premise that with a small number of words a person might be able to

express unlimited ideas (Baker & Chang, 2006; Baker et al., 2000), thus giving efficient results for relatively limited amount of lexical learning.

Beukelman and colleagues (1984) then applied the methodology of word frequency studies to adults using communication aids, considering the implications for system design to both enhance communication rate and system learning (Beukelman, Yorkston, Poblete, & Naranjo, 1984). The authors compared participants with each other and found a higher degree of congruence between participants' frequently used words than between their frequently used words and the published word lists at the time. This suggests that the communication samples of persons using communication aids may be a good source for frequent vocabulary for others in need of AAC. Similar to studies outside of the AAC field, the authors also found that a relatively small core vocabulary of between 200 and 500 words would allow a large proportion of the sample to be communicated. The authors suggested that this method would thus increase efficacy of an AAC system, which used whole word retrieval as opposed to typed messages. With a small sample size, the question remained as to whether similar results would be found for different populations.

By the end of that decade and for several decades since, research began investigating the spoken vocabularies of children and adults with the purpose of producing core vocabulary lists (e.g. Banajee et al., 2003; Beukelman et al., 1989; Fallon et al., 2001; Fried-Oken & More, 1992; Marvin, Beukelman, & Bilyeu, 1994; Stuart, Vanderhoof, & Beukelman, 1993). These authors produced composite lists of high frequency words in English. Some of these had an additional interest in comparing core and fringe vocabularies, for example, Marvin and colleagues investigated vocabulary samples taken in different contexts and at different times of day (Marvin et al., 1994). The authors found that, while the core high frequency words remained stable across time and context, the fringe vocabulary was highly variable.

Still others completed work investigating the efficacy of the word lists being suggested; as researchers questioned whether persons using AAC use the words contained in the word lists which had now been compiled by this vocabulary research. Yorkston, Smith and Beukelman (1990) examined the communication samples of ten participants using AAC who were producing their communication through letter-by-letter typing. The authors compared the words produced

17

by the participants to various vocabulary lists being used for AAC vocabulary selection at the time. The authors found high commonality in smaller list sizes but determined that an increased list size correlated with decreased efficacy of vocabulary, thus further demonstrating the importance for individualised vocabulary when planning for the extended vocabulary of a person's AAC systems (Yorkston et al., 1990).

The presence of both core and fringe vocabulary is essential for a generative graphic AAC system, to increase system efficacy (Beukelman et al., 1991; Robillard, Mayer-Crittenden, Minor-Corriveau, & Belanger, 2014; Witkowski & Baker, 2012) and, as already discussed, to give access to language through sentence construction (Banajee et al., 2003; Liu & Sloane, 2006; Trembath et al., 2007) and syntactical manipulation (Sutton et al., 2002). The grounds for core vocabulary to function as the foundation upon which the remainder of the system is built is therefore well established for English.

Having established a substantial amount of evidence in English, primarily from research conducted in North America, research then broadened to other countries and other languages. Trembath et al. (2007) investigated English used by preschoolers in Australia, while others took steps to investigate Korean (Kim, Park, & Min, 2003; Lee, Kim, & Park, 2005; Shin & Hill, 2016) Mandarin Chinese (Chen, Hill, & Yao, 2009; Liu & Sloane, 2006), French (Robillard et al., 2014) and German (Boenisch & Sachse, 2007).

## 2.5 The legitimacy of translation of core vocabulary lists

With such an extensive volume of research already published in the area of core vocabulary, prior to embarking on further observational studies one may question whether existing research might not be generalised. That is to say, could not the existing English core vocabulary lists be translated into new languages, including Zulu?

Indeed, Robillard et al (2014) found similarities between the features of English and French core vocabulary. The idea of semantic primitives (i.e. the notion that a finite set of concepts exists that are universal to all languages; see Baker & Chang, 2006; Liu & Sloane, 2006) might suggest that core vocabulary would be translatable across languages.

Chapter 2: Literature Review

However, such a conclusion would be premature and misleading. The comparisons between French and English revealed similarities in function of vocabulary and word types that were found in the high frequency lists (Robillard et al., 2014). This did not imply that core vocabulary was directly translatable. Furthermore, semantic primitives are conceptually and not linguistically based and thus would be expressed in different ways in different languages. For example, the semantic prime THIS (see Baker & Chang, 2006) cannot be expressed in Zulu without reference to the noun class and thus 12 different words are required to accommodate this semantic prime. In contrast, only one word is needed to express this prime in English. A more in-depth discussion of the characteristics of Zulu can be found in Section 2.6 below.

The first step in any word frequency study is the decision as to what is considered a word; which is the subject of much debate among linguists (Baker & Chang, 2006), and deliberation over concepts which can be expressed in one language with a single word serves only to confuse the decision. Baker and Chang (2006) give the example of French 'regarder' and the English counterpart 'look at', and thus question whether the English is then one word, a phrase or two words. The answer is, to a certain extent, irrelevant if possible at all, but a decision must be made on how frequency counts in a language are handled. For instance, in the Korean studies, Shin and Hill (2016) found that the omission of particles, terms of address and verbal suffixes in prior Korean core vocabulary studies had resulted in a diminished understanding of the Korean word lists which were generated due to the fundamental role which these elements play in the language.

In conclusion, while it may be true that people talk about the same kind of things and there exists a commonality in languages as demonstrated by semantic primes, fundamental differences also exist which would limit the legitimacy of the translation of the results of word frequency studies from one language to another. Rather, the identification of a core vocabulary in a specific language needs to come from the analysis of actual language samples. Since the content of written and spoken language often differs (Liu & Sloane, 2006), a core vocabulary in a specific language aimed at allowing a person in need of AAC to participate in natural face-to-face interactions would require the transcription and analysis of recorded samples of natural conversations.

## 2.6 The Zulu language

While core vocabulary lists and other vocabulary selection tools already mentioned are published and have been well researched as demonstrated in English, a problem is faced by the clinician who is trying to determine a vocabulary for a person who needs AAC but who also needs to communicate in Zulu. To date, there is no known publication regarding an investigation into the core vocabulary of Zulu or any of the African languages in the Nguni family.

The Zulu language is one of the 11 official South African languages. It is spoken as a home language (L1) by approximately 10 million people and as a second language (L2) by more than 15 million people (Lewis, Paul, Simons, & Fennig, 2016). The grouping of languages is the subject of much debate among linguists (Nurse & Philippson, 2006). However, it is widely accepted that Zulu belongs to a family[2] of over 400 languages which are spoken on the African continent (Gowlett, 2006; Nurse & Philippson, 2006; Paulos & Msimang, 1998). In Guthries' classification system, Zulu is classified within the Nguni group of the southern language family designated as S42 (Gowlett, 2006). A total of 95% of all Zulu L1 speakers live in South Africa (Lewis et al., 2016), with the greatest concentration in KwaZulu-Natal, where 77.8% of the population are L1 Zulu speakers (Statistics South Africa, 2011). Nonetheless, code switching and the mixing of languages is common and even inevitable in multilingual societies such as South Africa (Ndimande-Hlongwa & Ndebele, 2014; Slabbert & Finlayson, 2000). With 11 official languages, the population is often exposed to a variety of languages through popular media (Ramsay-Brijball, 1999) and the use of a variety of languages is necessary to be able to communicate with speakers of other languages (Ndimande-Hlongwa & Ndebele, 2014; Slabbert & Finlayson, 2000).

Languages rarely permit strict classification in terms of type, but rather display different features to varying degrees (Kosch, 2006). The Zulu language, like other African languages, is mainly synthetic agglutinative, relying not on word order and auxiliaries but on a series of morphemes (also called formatives; see Section 2.8) including stems, roots and affixes to create

---

[2] We acknowledge that the term 'Bantu language' is used to describe the specific language family that Zulu is a part of. However, since this term has been used outside linguistics for other purposes subsequent to which the term became derogatory and unacceptable even within linguistic use (Paulos & Msimang, 1998), we have, in following the same authors, chosen to use 'language family' instead.

word meaning (Kosch, 2006). Synthetic languages have a higher morpheme-to-word ratio than analytic languages (English being an example of a more analytic language). Zulu is also primarily agglutinating, meaning that morphemes can usually be easily segmented and that their phonological form remains relatively constant (Kosch, 2006). Kosch (2006) gives a clear description of how languages differ in grammatical understanding of the term 'word' and how words are constructed. Kosch compares the English, "*I see them*" to Zulu, "*ngiyababona*" and Northern Sotho, "*ke a ba bôna*"; which all carry the same meaning but differ significantly in structure. It is important to note that although the meanings are comparable, the construction is not. In Kosch's example, the author notes that although the Zulu in the example comprises four meaningful components (i.e. *ngi-* [I], *– ya-* [present tense morpheme], *– ba-* [them], *– bon-* [see], *-a* [final verb suffix]), yet unlike the three distinct words in the English counterpart, none of these components could appear independently but are bound forming a single word (Kosch, 2006;3). On this basis, Kosch (2006) describes orthographic words (words defined by orthographic space) and linguistic words (words defined by unit of meaning). In the example, Kosch describes the Zulu word as a single orthographic and linguistic word; the Sotho as four orthographic but one linguistic word, and the English as three orthographic and three linguistic words.

Agglutinative languages may be written disjunctively, where a single linguistic word is represented by a number of orthographically distinct units. Northern Sotho is one agglutinative language with disjunctive orthography (see the example above). Alternatively, agglutinative languages may be written conjunctively, where the meaningful units of each linguistic word are orthographically joined (Kosch, 2006). The agglutinative morphology of Zulu is written conjunctively (Kosch, 2006; Prinsloo & De Schryver, 2001) as already seen, and is written using the Latin alphabet. Zulu largely follows strict phonetic rules for its spelling, with the exception of the annotation of tone which is not typically used in Zulu written language (Doke, 1939). Zulu uses a system of nine tones with pitches that span just over an octave, the tone either being level or gliding (rising or falling) (Doke, 1939). The absence of tonal annotation can lead to misunderstandings in the written language since the tone is semantic, that is, it lends meaning to words (Doke, 1939). A common folk riddle describes a man who was running late for a function and sent a written message ahead with instructions pertaining to the slaughter of the cow. He

wrote, "*Inkomo ingabulawa ngingakafiki*"; depending on the tone used, the message can either be read, "the cow *may* be killed before I arrive", or "the cow *must not* be killed before I arrive".

Traditionally, core vocabulary studies have relied on the orthographic space to define the linguistic units ('words') that are counted to determine the most frequent vocabulary items. From the examples of English, Northern Sotho and Zulu above, it is clear that this method would give rise to very different types of 'core', depending on the linguistic nature and orthographic conventions of the language. This may therefore not be the most useful way of establishing a core vocabulary in Zulu. In order to consider which linguistic unit should be counted to establish a core vocabulary, the following sections will first give a brief introduction to Zulu grammar. The term 'formative' will then be introduced as the linguistic unit most suited for determining a core vocabulary. Following this, previous Zulu vocabulary studies employing frequency counts will be reviewed, with specific focus on the tagging system developed by Spiegler, Van der Spuy, and Flach, (2010a) and Spiegler et al. (2010b), which was used as the basis for the tagging system developed for this study.

**2.7 Zulu grammar: Noun classes and verbal extensions**

Looking closer at the construction of meaning in the language, Zulu is morphologically complex, with 15 noun classes (Ngcobo, 2013), which have been described and numbered in different ways (e.g. Doke, 1939, whose numbering system is also used by Nyembezi, 1982; and Meinhof, 1932, whose numbering system is also used by Taljaard & Bosch, 1993). As with other languages within this family, the noun class plays a significant role in the structure of the language (Nurse & Philippson, 2006; Pretorius & Bosch, 2003) – it is imperative that the various parts of speech used to qualify the noun agree with the noun according to its class (Spiegler et al., 2010b). From this necessity we see the birth of the concord (Pretorius & Bosch, 2003) which aligns the meaning of the sentence to the substantive (subject). That is to say that through the concordial formative, all words in the sentence must agree with the noun class of the noun which is the subject of that particular sentence.

By way of an example of this concordial agreement, let us consider two comparative examples given in Table 2.1 below.

*Table 2.1*

The Function of Concordial Formatives

|  | Orthographic word 1 | Orthographic word 2 | Orthographic word 3 |
|---|---|---|---|
| Sentence 1 | *buphi* | *ubuhlalu* | *obukhulu* |
| Formatives | {bu-}+{-phi} | {ubu-}+{-hlalu} | {obu-}+{-khulu} |
| Description of formatives | {indicative prefix, class 14}+{-phi} | {class 14}+{-hlalu} | {adjectival concord class 14}+{-khulu} |
| English translation |  | 'Where is the big bead?' |  |
| Sentence 2 | *liphi* | *isaka* | *elikhulu* |
| Formatives | {li-}+{-phi} | {i-}+{-saka} | {eli-}+{-khulu} |
| Description of formatives | {indicative prefix, class 5}+{-phi} | {class 5}+{-saka} | {adjectival concord,class 5}+{-khulu} |
| English tanslation |  | 'Where is the big sack?' |  |

From the sentences given in Table 2.1, it is clear that the morpheme – *hlalu* is the noun root ('bead') which belongs in the noun class 14; *ubu-*, and is always mentioned together with its noun class 14 prefix *ubu-*. Therefore, in this first sentence, the relative stem *-phi* ('where') and the adjective stem *-khulu* ('big') also have to agree with the noun class 14. In the second sentence, the noun stem *-saka*, belongs in the noun class 5; *i(li)-*. Therefore, the concords for both the relative and adjectival stem now refer to the sack and have to take on concordial agreements for noun class 5. Without the concordial agreements it is not possible to infer meaning to the sentence. By way of comparison, word order rules determine meaning in English; for example, a plant pot is different to a pot plant. In Zulu, concords are used to ascribe attributes and ownership, as well as determine the subject and the object of any utterance.

Zulu language therefore is governed by the noun, all objects and persons whether abstract or concrete being assigned to a class, and it is the concordial agreement which lends meaning to all Zulu sentence structures (Taljaard & Bosch, 1993). This structure is given in Zulu grammar textbooks (e.g. Doke, 1939; Nyembezi, 1982; Taljaard & Bosch, 1993) with slight differences only and typically pertaining to the terminology used and the numbering system as already mentioned.

Second to the noun, the verb and its construction in Zulu carries much of the meaning of any utterance (Nyembezi, 1982). The verb in Zulu must be brought into concordial agreement with the substantive it applies to, unless being used imperatively (Doke, 1939), otherwise it cannot be used in the language (Taljaard & Bosch, 1993). The verb additionally carries the negative conjugation as well as seven different moods to the verb, giving tenses, implication and manner (Doke, 1939). Further to this, a defining characteristic of the Zulu verb is the verbal extensions which are suffixed, giving rise to further meanings to the basic verb root (Taljaard & Bosch, 1993). Examples are given in Table 2.2 below in the consideration of the Zulu verb stem -hamba (meaning go or walk) and -shaya (meaning to hit or beat), with selected verbal extensions.

*Table 2.2*

Examples of Verbal Extensions

| Mood | Examples | English translation |
|---|---|---|
| **Infinitive** | *Ukuthatha*<br>{uku-}{-thath-}{-a}<br>{verbal noun class 15}{-thath-}{verb ending}<br>*ukushaya*<br>{uku-}{-shay-}{-a}<br>{verbal noun class 15}{-shay-}{verb ending} | 'to take'<br><br>'to hit' |
| **Indicative mood, continuous** | *lithatha*<br>{li-}{-thath-}{-a}<br>{indicative concord class 5}{-thath-}{verb ending}<br>*lishaya*<br>{li-}{-shay-}{-a}<br>{indicative concord class 5}{-shay-}{verb ending} | 'it takes'<br><br>'it hits' |
| **Indicative mood, causative extension** | *lithathisa*<br>{li-}{-thath-}{-is-}{-a}<br>{indicative concord class 5}{-thath-}{causative extension}{verb ending}<br>*lishayisa*<br>{li-}{-shay-}{-is-}{-a}<br>{indicative concord class 5}{-shay-}{causative extension}{verb ending} | 'it causes (something) to take/ it is contagious'<br>'it causes to hit' or 'collides' |
| **Indicative mood,** | *lithathisisa*<br>{li-}{-thath-}{-isis-}{-a} | 'it takes a lot' |

| | | |
|---|---|---|
| **intensive extension** | {indicative concord class 5}{-thath-}{intensive extension}{verb ending} | 'it gives a thorough beating' |
| | *lishayisisa* | |
| | {li-}{-shay-}{-isis-}{-a} | |
| | {indicative concord class 5}{-shay-}{intensive extension}{verb ending} | |
| **Indicative mood, passive extension** | *lithathwa* | 'it is taken' |
| | {li-}{-thath-}{-iw-}{-a} | |
| | {indicative concord class 5}{-thath-}{passive extension}{verb ending} | |
| | *lishaywa* | 'it is beaten' |
| | {li-}{-shay-}{-iw-}{-a} | |
| | {indicative concord class 5}{-hamb-}{passive extension}{verb ending} | |

## 2.8 Zulu morphology and the function of formatives

From the preceding section, it is clear that Zulu contains a great number of noun and verb roots. However, the root can never be uttered alone, but is combined with other components to encode different meanings within the language (Nyembezi, 1982; Taljaard & Bosch, 1993). The term 'formatives' has been used to describe all the component parts (Kosch, 2006). The construction of meaning in Zulu described in the preceding section by referring to the role of the concord, noun and the verb, may also be described by explaining the use of formatives and/or morphemes. A morpheme refers to the smallest part, and an indivisible unit of semantic or grammatical function which constructs a word (Kosch, 2006). Although initially the term formative was used to describe only morphemes which performed a grammatical function, the usage extended to also include morphemes which contribute lexically to words (Aronhoff, in Kosch, 2006), that is, they contribute content. Thus, in this definition, formatives also include roots for example, which can be seen as the content portion of the word. The term 'formative' is now commonly used as an alternative to morpheme, particularly in the context of certain prefixes of African languages (Kosch, 2006). Indeed, Doke (1939) uses this term and categorises stems, roots, and affixes and distributes these among different formative types which will be described below. The term formative is preferred in this study due to its inclusivity toward lexical content and the ease with which the various formative types may be grouped for further analysis.

Chapter 2: Literature Review

The structure of Zulu allows one to list with certainty the various formatives or 'izakhi' (Nyembezi, 1982, p.43) used to form linguistic words. It is possible to make a distinction between content (lexical) and structure (grammatical) formatives (Kosch, 2006). This is a useful distinction when it comes to AAC since, in the creation of a generative communication system, it will be necessary to involve all structure formatives, to which then the most common content formatives may be added. For an agglutinative language where very few lexical items can stand alone without grammatical context being assigned, this becomes a key part of the AAC approach.

Structure formatives therefore are all the formatives which allow Zulu word and sentence construction, while content formatives refer to the 'things' (content) spoken about. By way of an explanation of this grouping of formatives, let us consider the phrase *isihlalo sami siyaphuka* ('my chair is breaking'). In this phrase, various formatives may be identified (Table 2.3).

*Table 2.3*

Examples of Content and Structure Formatives

| Formative | Formative description | Type of formative |
|---|---|---|
| *isi-* | noun prefix class 7 | structure formative |
| *-hlalo* | noun stem | content formative |
| *s-:* | concord 7 | structure formative |
| *-a-* | possessive formative | structure formative |
| *-mi* | pronoun | content formative |
| *si-* | concord 7 | structure formative |
| *-ya-* | verbal auxiliary | structure formative |
| *-phuka* | verb stem | content formative |

The content formatives therefore are *-hlalo* (noun stem), *-mi* (pronoun), and *-phuka* (verb stem). The content formatives are the portions which are possible to translate with relative ease: (i.e. *hlalo* – 'chair' ; *mi(na)* – 'me' ; *phuka* – 'break'; Dent & Nyembezi, 1995). The structure formatives, on the other hand, enable the relationships between the content formatives to be clarified within the phrase and have no English equivalent (Paulos & Msimang, 1998). Although

content formatives are more closely related to English words and can be translated, neither content nor structure formatives can be used alone, nor do they carry meaning when uttered in isolation.

Thus, structure formatives are those such as concords, adverbial formatives and verbal auxiliaries which are required for the meaningful use of the content formative, which includes various roots and stems such as noun, verb and relative stems. Formatives have been described by different grammarians (Doke, 1939; Nyembezi, 1982; Taljaard & Bosch, 1993) and will be described below according to the types described by Doke (1939).

### 2.8.1 Content (lexical) formatives

*Roots and stems* are the content portion of words, where roots are the ultimate immutable form shorn or all prefixes and suffixes, and a stem is shorn only of its prefix (Doke, 1939). For example, *ngithandile* ('I loved'); identifies the verb stem *–thandile* but the verb root is *-thand-*. These, as lexical items, form the large bulk of language (Nyembezi, 1982). Roots may be found in various parts of speech, such as in the noun, pronoun, relative, verb, and adjective. However, the terms roots and stems are not necessarily that distinguishable among linguists according to Doke (1939), and neither are the Zulu roots or stems always immutable, since through the use of a different tone or pronunciation, a different meaning can be ascribed (Doke, 1939). For instance, the noun *umsebenzi* can mean either 'the work', or 'the worker'; and the verb stem *-hlanza* can mean either 'to vomit' or 'to wash', depending on the tone of pronunciation. Since the distinction between the two is not always clear, rules were established in the frequency counts of this study as to how different roots and stems were to be handled.

*Ideophones* are also described by Doke (1939) as root formatives but they differ from other roots in that they do not employ concords (Nyembezi, 1982). An ideophone can be described as a word that describes a predicate in manner, sound or action (Doke, 1939), and is often onomatopoeic in nature. The closest comparison in English might be the word 'splash' (Taljaard & Bosch, 1993). The ideophone, however, is not necessarily onomatopoeic. It is not found in most European languages (Doke, 1939) and it does not feature in English parts of speech (Taljaard & Bosch, 1993). By way of an example, *khwishi* is an ideophone of blackness and thus

a person might describe a very dark item by saying *limnyama khwishi* (lit. 'it is black', conceptually translated perhaps, 'it is truly black').

*Conjunctive formatives* work by way of joining words together much in the same way as the English conjunction, and mostly feature as isolate words without further devisable portions. The conjunctive formative *na-* however, similar to the English 'and' or 'together with' (Doke, 1939) does require further formative companions for its use. For example, *isinkwa namanzi*, ('bread and water'), or *ngiyamba naye* ('I am walking with her').

### 2.8.2. Structure (grammatical) formatives

*The noun prefix group,* are used to form nouns varying by class. The use of a different noun prefix differs to a concord in that the noun prefix alters the meaning of the noun, whereas the concord serves to point to the noun (Doke, 1939). For example, the root *–nja*, prefixed with the noun prefix for class 9 (*i-*), becomes *inja* ('dog'); while prefixed with class 10 (*izi-*), it becomes *izinja* ('dogs'). Further to this, one might also prefix the same root with the noun prefix for class 14 (*ubu-*) to become *ubunja* ('low morals'); one can only suppose, like those of a dog, not a person.

*Concords,* although closely related to noun prefixes must be rigorously differentiated from them, as their function is to point to the noun as the agent or subject rather than alter it (Doke, 1939). Neither should the concords to be considered a pronoun – such a practice stems from an attempt to correlate the analysis of Zulu with English constructs (Paulos & Msimang, 1998). For example, *izinja ziyakhonkhotha* (dogs are barking); the first underlined part is a noun prefix and, if it were English, alters the noun 'dog' to 'dogs'. The second underlined part is the concord and aligns the verb '*to bark*' with the dogs, ascribing to them the action of barking.

*Verbal auxiliaries* provide verb inflections, tenses, and implications within the rule-bound regular verb system in Zulu (Doke, 1939). Examples are the negative prefix *ka-* or the more modern typical *a-* in the example *kangifuni* / *angifuni* ('I don't want'), and the negative infix *–nga-,* used for instance in the participial mood; *engafuni (*'while he did not want').

*Prefixal formatives* are used to form pronouns, copulatives and adverbs (Doke, 1939). For example, the adverbial formative *ka* adds the manner of action. Added, for instance, to the

adjectival roots *-khulu* ('big') or *-ncane* ('small'), it produces *kakhulu* ('a lot'), *kancane* ('a little').

   *Suffixes* have three basic types. The first type is adjoined to nouns and performs the function of diminution, augmentation or indicating gender. For example, the addition of either the diminutive or augmentative suffix to the noun; *ikhanda* ('head') gives respectively *ikhanjana* ('little head') and *ikhandakazi* ('big head'). Those adjoined to verbs give rise to verbal derivatives; for instance, the verbal extensions described in Table 2.2 and others. Those adjoined to nouns (invariably also coupled with a prefixal inflexion which alters the noun prefix) are locative suffixes; for example, *itafula* ('table') with locative suffixion becomes *etafuleni* ('on/at the table'). (Note the prefixal inflexion – the prefix *i-* becomes *e-*).

   *Enclitic formatives* modify ideas by means of a suffix that is mainly applied to verbs in phrase construction, some of which are interrogative (Doke, 1939). Let us take two examples: i) the suffix *-ke* similar to 'then', used, for example, in *ngena ke* ('Come in then!'); and ii) the interrogative *-ni* applied to a verb *ufuna* ('he wants') becomes *ufunani* ('what does he want').

## 2.9 Published Zulu vocabulary studies

   A handful of studies have been conducted within the field of corpus linguistics and natural language processing employing word frequency counts in Zulu and other languages in this language family. The characteristics of the Zulu language and Zulu orthography described above have presented a number of challenges to those seeking to conduct vocabulary studies either for the purposes of corpus linguistics (Allwood et al., 2010; Allwood & Hendrikse, 2003; Prinsloo & De Schryver, 2001; South African Department of Arts and Culture & Centre for Text Technology [CTexT], 2014) or natural language processing (Pretorius & Bosch, 2003; Prinsloo & De Schryver, 2001; Spiegler et al., 2010b).

   The South African Spoken Language Corpus (SASLC) project (Allwood et al., 2010; Allwood & Hendrikse, 2003) made use of spoken African languages, including Zulu, while the NHCLT Text Corpora (South African Department of Arts and Culture & Centre for Text Technology [CTexT], 2014) made use of written texts available from various government and media websites. The former SASLC reported a spoken Zulu corpus of 25 000 words, while the

latter NHCLT Text Corpora published a substantial Zulu written language corpus of nearly 1.5 million words.

Both of these initiatives used the orthographic space to define unit of analysis in computing results, which gives rise to some issues in statistical significance of the data. The authors did not report any Zulu word frequency results in the paper, but discussing the SASLC progress, authors reported on Xhosa (which is also an Nguni grouped language closely related to Zulu; Gowlett, 2006) high frequency vocabulary (Allwood et al., 2010). The paper tabulates the most frequently occurring words, which on inspection are mainly adverbs and conjunctions. This is understood, and authors report this to be a distortion due to rich morphological variability of the agglutinating language obscuring the frequency of verbal lexical tokens (Allwood et al., 2010). Indeed, the same list shows English-derived words, '*so*' and '*and*[3]' appearing in the top 12 most frequently appearing lexical items of the Xhosa corpus (Allwood et al., 2010).

This is an understandable result for an agglutinating language, if we consider the Zulu example of the lexical token -*ceba;* to tell on someone*;* a commonly used concept among young children. A child might say; "*ngizoku<u>ceba</u>*" (I will tell on you), "*bazoku<u>ceba</u>*" (they will tell on you), "*ngizom<u>ceba</u>*" (I will tell on him), or "*ngim<u>cebile</u>*" (I told on him), "*u<u>ce</u>tshwe ubani?*" (who told on you?). When using an orthographic space when computing the data collected from this child, a word frequency count would reveal five different words, which is accurate when considered under Kosch's (2006) definition of a linguistic word. The commonality shared by each of these words is lost and, when applied to an AAC system where one graphic symbol is used to represent one orthographic word, each of the above words needs a separate representation. Therefore, the reliance on the orthographic space to define the lexical unit sampled in core vocabulary studies would not be helpful to design a generative graphic AAC system.

Acknowledging this problem, Allwood and Hendrikse (2003) make mention of the need to develop a kind of annotation system which will enable the retrieval of significant patterns in the

---

[3] Code switching and borrowing words from other languages is common among South African speakers (Slabbert & Finlayson, 2000).

data. They suggest a practice of morphological tagging but do not elaborate or report any tagged data.

The *Ukwabelana Corpus* (Spiegler et al., 2010a, 2010b) began addressing this question of morphological tagging and analysis which speaks to the agglutinative morphology of Zulu. The Ukwabelana Corpus makes use of written Zulu as a source taken from published fiction and the Zulu bible – it has a total word corpus of nearly 300 000 words. As well as producing a corpus of written Zulu, in this work the authors endeavour to create an automated morphological tagging system which could then be used for different purposes including machine learning (which may be useful for instance in automatic grammar and spell checking). In accompanying material for the corpus, Spiegler and colleagues detail the labels used in the database tagging system which identify both the grammatical function of the formative (coded as a letter of the alphabet) and, where relevant, the noun class agreement (as a numeral), giving rise to 207 different labels or tags (Spiegler et al., 2010a).

By way of illustrating this method, Table 2.4 below applies the Ukwabelana corpus system to the child utterances used above. In this method, the word is segmented into its formative components and tagged with a code (the given code is enclosed in angle brackets – see the second column of the table). It is important to note that the segments into which the words are divided are in no way independent and are only meaningful as part of the whole (Kosch, 2006). Table 2.5 describes the codes which are exhibited in the example.

Table 2.4

*Zulu Words with Ukwabelana Corpus Coding Tags Applied*

| Word | Coding Tags applied | English translation |
|---|---|---|
| ngizoku*ceba* | ngi<i1s>zo<fut>ku<o2s>ceb<vr>a<va> | 'I will tell on you' |
| Bazoku*ceba* | ba<i2>zo<fut>ku<o2s>ceb<vr>a<va> | 'they will tell on you' |
| Ngizom*ceba* | ngi<i1s>zo<fut>m<o1>ceb<vr>a<va> | 'I will tell on him' |
| Ngim*ceb*ile | ngi<i1s>m<o1>ceb<vr>ile<vpl> | 'I told on him' |
| u*ce*tshwe ubani? | u<i2s>ce<vr>tshw<xp>a<va> | 'who told on you?' |

Table 2.5

*Description and Frequency of Tagged Codes in Example*

| Code | Label | Segment tagged | Frequency of occurrence |
|------|-------|----------------|-------------------------|
| <fut> | Futurative | zo | 3 |
| <i1s> | indicative subject prefix class 1 singular | ngi | 3 |
| <i2> | indicative subject prefix class 2 | ba | 1 |
| <i2s> | indicative subject prefix class 2 singular | u | 1 |
| <o1> | object prefix class 1 | m | 2 |
| <o2s> | object prefix class 2 singular | ku | 2 |
| <vr> | verb root | ceb / ce | 5 |
| <va> | verb suffix | a / i / o | 4 |
| <vpl> | perfect verb suffix long form | ile | 1 |
| <xp> | passive extension | tshw | 1 |

A brief examination of the words with the morphological tags applied reveals that although they are orthographically and linguistically distinct words, there are commonalities between these utterances, as well as differences. The Ukwabelana Corpus and its coding system (Spiegler et al., 2010a, 2010b) offer insight into the organic employment of Zulu grammar and morphology to allow for frequency counts of lexical and grammatical units, which reflect the agglutinative nature of the language. Using this method, it is possible to count each meaningful part separately and obtain a true reflection of the frequency of individual concepts conveyed in a language sample. This method of analysis therefore, although deeper than word level, appears to be more capable of producing meaningful units which might be exploited for the purpose of AAC since it permits the segmenting and identification of formatives. By identifying the most frequently and commonly used formatives through this method, a meaningful core vocabulary could be established. Further explanation of this method and adaptations to it are given in Chapter 3, Section 3.7.2.6.

**2.10 Summary**

This chapter described how graphic symbols are used in AAC, as well as vocabulary selection methods used by clinicians seeking to determine vocabulary to provide to preliterate/non-literate persons who need AAC. Research into core vocabulary as one method for

guiding the selection of vocabulary for such systems was reviewed. Reasons why the translation of core vocabulary does not offer a useful method of determining the most frequently used linguistic components in another language were given. As a background for the formative tagging system used in this study as a basis for frequency counts, the most important components of Zulu grammar were briefly described. An explanation of the types of formatives encountered in Zulu followed. Zulu vocabulary studies employing frequency counts were briefly reviewed, with specific focus on the morphological tagging system developed by Spiegler et al. (2010a, 2010b), which was used a basis for the formative tagging system used for this study.

# CHAPTER 3

# METHODOLOGY

## 3.1 Introduction

This chapter describes the research methodology used, beginning with the main aim, sub-aims and research design utilised. The setting is described, followed by a description of the participants (recruitment, selection criteria and descriptive criteria). Materials and equipment are described next. The aims, procedures, results and recommendations of the pilot study are then discussed. The procedures for data collection and analysis are also described, including ethical considerations and considerations regarding reliability and validity.

## 3.2 Aims

### 3.2.1 Main aim

The main aim is to determine the core vocabulary (most frequently and commonly used vocabulary) of typically developing first language Zulu-speaking preschool children during a variety of activities in a preschool setting, as a basis for guiding vocabulary selection for AAC systems.

### 3.2.2. Sub-aims

The following sub-aims were identified:

i. To identify the linguistic formatives[4] used by Zulu-speaking preschoolers;

ii. To determine how many different formatives and formative types are used by Zulu-speaking preschoolers;

iii. To describe the vocabulary in terms of the most frequently and most commonly used formatives and thus determine which formatives, if any, constitute a core vocabulary;

---

[4] As discussed in Chapter 2, the nature of Zulu morphology requires that formatives rather than words be targeted as the unit of analysis for the purposes of this study.

iv.      To further describe the characteristics of core vocabulary by formative types, as well as by content and structure formatives (including subtypes);

v.      To examine the lexical meaning of the content formatives of the core vocabulary in relations to previous studies**.**

### 3.3 Research design

The study used a quantitative non-experimental observational design (McMillan & Schumacher, 2010). This type of design is ideal for researchers wishing to describe and summarise an existing phenomenon (McMillan & Schumacher, 2010), as this design gives access to primary data through observation. The speech of six preschoolers engaged in routine activities in the preschool context was collected using portable voice recorders. The recordings were transcribed verbatim and formatives identified were analysed for the frequency and commonality of use. Drawbacks of an observational design include possible participant reactivity which impacts on reliability of results, and a small sample size impacting on the extent to which results can be generalised.

### 3.4. Stages of the study

Prior to the stages commencing, ethics approval was obtained from the relevant authorities, as explained in Section 3.8. The stages of the study, illustrated in Figure 3.1, consisted of (1) recruitment of participants, (2) screening and selection of participants, and (3) data collection. In order to test the methodology, a pilot study was carried out, which followed the same stages prior to the main study commencing.

35

| Stage 1: Participant recruitment<br><br>Letters of information were emailed to principals of various schools that had a preschool class where Zulu was used as the medium of instruction. The principals were asked for permission to recruit participants from their school. Preschool educators at these schools then identified potential participants. | ⟹ | Stage 2: Screening and selection of participants<br><br>Informed consent was obtained from caregivers. Biographical questionnaires were completed by caregivers. Participants were screened and selected. Participant assent was obtained. | ⟹ | Stage 3: Data Collection and analysis<br><br>Suitable times for recording were set between educator and researcher. The speech of participants was recorded on consecutive days until such time as a sample of 1 500 words had been obtained. Recordings were transcribed and formatives were tagged. Statistical analyses were done. |
|---|---|---|---|---|

Figure 3.1. *Stages of the study.*

## 3.5 Settings



Figure 3.2. *Morning assembly under the tree at Site 2.*

All three preschools (Sites 1-3) which participated were located in a rural area of KwaZulu-Natal under tribal authority. Zulu is spoken by approximately 95% of the population in this area (Census Stats SA, 2011). All three schools used Zulu as the medium of instruction in the foundation phase of learning (preschool to Grade 3). All three preschools were attached to primary schools which went up to Grade 7. Preschool children at each school had varying amounts of involvement with the children from other classes. At Site 2, children spent break

36

times together in a shared playground with children from all the grades, whereas at Sites 1 and 3 the children, although sharing the same space, had separate break times to the other classes. All the children at all sites were provided with a meal during the day. All the schools in the study were fenced and had running water in the yard. None had electricity or landline telephones, but all had ablution facilities for staff and children. The preschools all used Zulu, with English being used sometimes in rhymes and songs. One school also reported children using English on the playground – in the recordings this was observed to be code switching of single words rather than the use of English in complete sentences. Each participating school had only one preschool class. The class size ranged from 14 pupils (Site 3), 17 pupils (Site 2), to 39 pupils (Site 1), all having a single teacher responsible for the group. Site 1 also had a student teacher who was carrying out practical training during the week of this research. At the third site, the teacher was also responsible for the Grade 1 class group during the week of data collection due to staff shortages at the end of term. The children remained in separate groups, however, and were assigned tasks to carry out while the teacher attended to the other class. These children spent less time in structured, teacher-led activities and this may have influenced the children's interaction, leading them to talk more freely – this was seen in data collection that was achieved in a shorter period at this site. None of the schools had formal playground equipment for children but in all preschools the children were observed to play group and team games independently and as led by their teacher, as well as engage in imaginative play with items found in the yard. The preschools all reported using the Grade R curriculum proposed in the National Curriculum and Assessment Policy Statement (CAPS).

## 3.6 Participants

### 3.6.1 Sampling and recruitment

The study was approved by the Research Ethics Committee of the Faculty of Humanities of the University of Pretoria (Appendix B). Permission to approach schools was also granted by the KwaZulu-Natal Department of Education (Appendix C). This study targeted a specific population, namely Zulu-speaking preschool children aged between 5;0 (years;months) and 6;11 who were exposed mainly to Zulu at home and at preschool. Convenience sampling was utilised (McMillan & Schumacher, 2010) to recruit participants who met the selection criteria (see Section 3.6.2). Schools that used Zulu as a medium of instruction in the preschool phase were

therefore identified in rural northern KwaZulu-Natal. The researcher contacted four schools that were geographically nearest to her place of residence within a radius of up to an hour's car drive. School principals of target schools were contacted telephonically and were then provided, via email or hand delivery, with an information letter and a request to provide permission to recruit learners from their schools (see Appendix D). If no response was received within a week of sending/delivering the letter, they were contacted through a second mail or a telephone call. Written permission was granted by three of the four principals who were contacted.

Educators in the preschool phase (of the schools whose principals gave permission) were requested to complete a questionnaire related to the school environment and programme (see Appendix E). The educators were also requested to nominate two children (one boy and one girl) who spoke Zulu at home and who, in their opinion, demonstrated typical speech and overall typical development for participation in the research study. They were provided with a short nomination form (included in the preschool background questionnaire – Appendix E) to assist in this process.

The educators then sent information letters and consent forms in English and in Zulu to parents of the nominated children (see Appendix F). Parents were requested to indicate whether or not they consented for their child to participate in the study. In addition, consenting parents were requested to complete a biographical questionnaire (see Appendix G) which provided basic biographical information about the child, as well as basic information about the physical and social home environment, including languages to which the child was exposed and languages used in the home. A total of seven parents gave consent for their children to participate in the study. The biographical questionnaire was perused to ensure that these children indeed met the selection criteria (see Section 3.6.2). All seven children met the selection criteria.

Once parental consent was obtained, the researcher scheduled a time with the preschool teachers to meet potential participants. At the appointed time, the researcher met each of the potential participants individually. She explained the study to the child using pictures, following the script of Appendix H. The child was then given the opportunity to give or decline assent to take part in the study (Appendix H) by marking the thumbs up picture with his/her own thumb

and paint to indicate assent or the thumbs down to refuse. A total of seven children were approached and six assented to take part in the study.

### 3.6.2 Participant Selection criteria

The selection criteria are provided in Table 3.1

Table 3.1

*Participant Selection Criteria*

| Criteria | Justification | Measure used |
|---|---|---|
| Age 5;0 *(years; months)* - 6;11 | Language used by children of this age typically resembles the structure used by adults (Owens, 2005). Therefore, the language sample gained was expected to show a mature structure but not necessarily the mature vocabulary, but rather the vocabulary which is assumed appropriate for the young AAC user. | Biographical background questionnaire (see Appendix G) |
| Language spoken most frequently at home and in the preschool is Zulu | Zulu is the language targeted in this study. Given the multilingual nature of South African society, with at least 38% of homes bilingual (Census Stats SA, 2011), it can be deduced that many children are introduced to second languages before they reach school. The decision was made not to exclude children who were exposed to other languages. | Preschool background questionnaire (see Appendix E), biographical background questionnaire (see Appendix G) |
| Typical speech and language development as identified by the preschool teacher | Children in need of AAC are assumed to have the same vocabulary needs as their typically developing peers (Beukelman et al., 1991). | Teacher nomination (captured on preschool background questionnaire – Appendix E), as in previous studies (Fallon et al., 2001; Trembath et al., 2007) |
| Children who attended the preschool class for a minimum of two days/week and had attended for at least five months prior to commencement of study | The children needed to be settled and familiar with those in the environment as this relates directly to talkativeness (Miller, Andriacchi, & Nockerts, 2011). The children needed to have also formed social connections with peers and teachers, in order for the recording to include a variety of communication functions. | Preschool background questionnaire (see Appendix E) |

### 3.6.3 Participant description

Individual child participant descriptions are provided in Table 3.2. All except one of the families reported using fire to cook food, while two families reported using camping gas in addition to this. Only one home reported having electricity. Only two children were reported to watch TV and only one family used English as well as Zulu in the home. No other languages were reported. Only two families had anyone that was employed and all except one family relied on social grants to cover their month-to-month costs. The mean age across participants was 5;5 (years; months).

Table 3.2

*Participant Description*

| Part. no. | Site | Gender | Age | Length of attending preschool | Frequency of attendance at preschool | Primary language exposed to over radio | Primary language exposed to over television | Others in the home | | | Monthly income of household |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Relation | Age | Language(s) used | |
| 1 | 1 | F | 5;1 | 5 months | Daily | Zulu | n/a | Brother | 5 | Zulu | R500 – R2000 |
| | | | | | | | | Brother | 16 | Zulu | |
| | | | | | | | | Brother | 17 | Zulu | |
| | | | | | | | | Aunt | 47 | Zulu | |
| | | | | | | | | Granny | 74 | Zulu | |
| 2 | 1 | M | 5;8 | 5 months | Daily | Zulu | Zulu | Brother | 4 | Zulu | > R 6 100 |
| | | | | | | | | Brother | 7 | Zulu | |
| | | | | | | | | Neighbour | 6 | Zulu | |
| | | | | | | | | Neighbour | 8 | Zulu | |
| | | | | | | | | Neighbour | 8 | Zulu | |
| | | | | | | | | Brother | 9 | Zulu | |
| | | | | | | | | Brother | 10 | Zulu | |
| | | | | | | | | Sister | 17 | Zulu | |
| | | | | | | | | Aunt | 30 | Zulu | |
| | | | | | | | | Aunt | 32 | Zulu | |
| | | | | | | | | Father | 32 | Zulu | |
| | | | | | | | | Granny | 58 | Zulu | |
| | | | | | | | | Grandfather | 61 | Zulu | |
| 3 | 2 | F | 5;9 | 6 months | Daily | Zulu | n/a | Sister | 5 | Zulu | R500 – R2000 |
| | | | | | | | | Sister | 10 | Zulu | |
| | | | | | | | | Brother | 24 | Zulu&English | |
| | | | | | | | | Brother | 26 | Zulu&English | |
| | | | | | | | | Mother | 46 | Zulu&English | |
| 4 | 2 | M | 5;3 | 6 months | Daily | Zulu | n/a | Aunt | 10 | Zulu | R500 – R2000 |
| | | | | | | | | Mother | 24 | Zulu&English | |
| | | | | | | | | Granny | 40 | Zulu | |
| 5 | 3 | F | 5;9 | 6 months | Daily | Zulu | n/a | Brother | 4 | Zulu | R500 – R2000 |
| | | | | | | | | Aunt | 27 | Zulu | |
| | | | | | | | | Uncle | 32 | Zulu | |
| | | | | | | | | Uncle | 35 | Zulu | |
| 6 | 3 | M | 5;3 | 6 months | Daily | Zulu | n/a | Uncle | 14 | Zulu | R500 – R2000 |
| | | | | | | | | Cousin | 14 | Zulu | |
| | | | | | | | | Sister | 14 | Zulu | |
| | | | | | | | | Sister | 17 | Zulu | |
| | | | | | | | | Granny | 45 | Zulu | |

Chapter 3: Methodology

## 3.7 Materials and equipment

### 3.7.1 Equipment

#### 3.7.1.1 Voice recorders and microphones

Two digital voice recorders (Olympus, model DM 650) with lapel microphones (Audio Technica ATR 3350) were used to record the speech of participating children. The voice recorders were worn on the body using small nylon waist pouches (see Figures 3.3 and 3.4) and the microphones were attached to the upper body clothing using the clip which comes standard on the microphone.



Figure 3.3. *Small pouch worn by participants during data collection.*



Figure 3.4. *Picture showing the pouch with recording equipment worn by a participating child.*

#### 3.7.1.2 Laptop and headphones

An Acer Aspire 5733Z laptop and Polaroid Neon Studio headphones were used to transcribe and tag the data.

### 3.7.2 Materials

#### 3.7.2.1 Information letters and permission/consent forms

Information letters and consent forms (see Appendix F) were provided in both English and Zulu to the parents/legal guardians of all nominated participants; Zulu and English being the languages which are most widely used in the area. The researcher drew up the letter and consent form in English. These were then translated to Zulu using blind back translations (Peña, 2007). The researcher (who is fluent in Zulu) translated the English letter and consent form into Zulu. A person fluent in written forms of both English and Zulu then blindly translated the Zulu version back into English, and these back translations was compared to the original English version to ensure linguistic equivalence. No changes were required.

#### 3.7.2.2 Pictures and script for providing information to children and assent forms

A script was drafted in Zulu in child-friendly language, which the researcher read out to potential participants to explain the study to them. Each aspect that children were informed about was also illustrated with a picture. A picture-based form was used to gain their assent prior to the commencement of recording. See Appendix H for this information script, pictures and assent form.

#### 3.7.2.3 Preschool background questionnaires

A preschool background questionnaire (Appendix E) was drawn up to obtain information about the preschool environment and programme as well as to identify potential child participants according to the selection criteria. The questionnaire probed for variables such as the numbers of children in the preschool class, whether or not these children interacted with older children from the primary school during the day and what their daily programmes contained. These questionnaires were provided in English, as the teachers completing the forms were qualified educators and proficient in English.

#### 3.7.2.4 Biographical questionnaire

Biographical questionnaires (Appendix G) were given to parents/guardians of the participating children to obtain background information on the home environment as well as basic developmental information and information on language exposure of the child within the

43

home. These questionnaires were also provided in both English and Zulu. The same process of blind back translation was used to ensure linguistic equivalence.

### 3.7.2.5 Systematic Analysis of Language Transcripts (SALT) software

Transcription software, namely the Systematic Analysis of Language Transcripts (SALT) (Miller & Iglesias, 2012) was used. This software was developed to standardise the process of transcribing and analysing language samples. A trained research assistant transcribed the recordings into this programme, following the transcription rules (see Appendix I) established for the study and the conventions of the SALT programme, as well as unique codes applicable to Zulu to allow certain automatic analyses (e.g. frequency counts) to be run by the program. Further explanation of the unique codes that were developed is given in section 3.7.2.6 below and in the Tagging Manual (Appendix J). Additional manual analyses were also done.

### 3.7.2.6 Formative tagging system

In order to establish a meaningful Zulu core vocabulary, a suitable tagging system had to be developed. As discussed in Chapter 2, the structure of the Zulu language prohibits word count conventions developed for European languages (such as English, German and French) to be applied to Zulu. It is common for annotation schemes to be developed to enable research to identify patterns of linguistic phenomena (Allwood & Hendrikse, 2003), which, in the case of this work, is formative frequency patterns. Leech (as cited in Allwood & Hendrikse, 2003) cautions that care must be taken to avoid bias in any annotation scheme additional to what the language ordinarily contains in its traditional orthography. Nonetheless, using additional annotation is now a common practice among corpus linguistic studies (Allwood & Hendrikse, 2003). In this study, the threat of bias was tempered through the use of inter-rater reliability measures to both transcription and tagging phases of data analysis (see Section 3.9.6).

The use of the annotation scheme of morphological tagging developed for the Ukwabelana Corpus (Spiegler et al., 2010a, 2010b) mentioned in Chapter 2 was used as a basis for this method. Further modifications to this scheme were required in order to reap the data which would be meaningful for a generative AAC vocabulary within the core vocabulary approach discussed in Chapter 2. The key aspects in the development of the tagging system employed in this study are described now. The principle guiding all the modifications of the Ukwabelana

44

Corpus coding system for the purpose of this study was therefore that this system needed to capture the smallest possible meaningful components of the language which could then be represented on a graphic system and enable the user to generate novel utterances in Zulu. The amendments needed were identified during the process of applying the coding system to the pilot data, and were confirmed and finalised in consultation with the second co-supervisor. The original Ukwabelana corpus is an open source system and can be found at http://www.cs.bris.ac.uk/Research/MachineLearning/Morphology/resources.jsp The final tagging scheme used in this study is given in Appendix J.

Key changes made to the original tagging scheme are outlined in points (a) to (d) below.

a.  Ukwabelana corpus tagging scheme confers too much detail in the segmentation of verb endings coding as well as grammatical and phonological variations of suffixion. These component parts not always carrying concrete meaning clouded the word count frequency and had to be dissolved. The decision remained as to whether the word to be counted would be the verb root or the stem. For example, the verb stems *-hamba* and *-hambile*, may be translated as *go* and *gone* in English. In previous studies different approaches were taken on grammatical variations of a word. Trembath et al (2007) treated grammatical variations as separate words, whereas Boenisch and Soto (2015) counted grammatical variations of the same word under a single word, reflecting the common practice on AAC displays to group word variations under a single word. The decision was taken to count variations as a single word, coding the verb root (*-hamb-*).

b.  Too little detail was retrieved from frequency counts when tagging the roots only such as the verb roots (<vr>) and noun roots (<nr>). For instance, in the example in Figure 3.5 below, the two verb roots identified are clearly two separate lexical items, *-ceba* ('tell on') and –*shaya* ('hit'). However, using the <vr> tag only, the lexical items would be lost; a word frequency count would only show <vr> occurring twice. Similarly, the variations on future tense between the imminent *-zo-* and far future *-yo-* requires more detailed tagging.

| Ngizom*ceba* | ngi<i1s>zo<fut>m<o1>ceba<vr> | (I will tell on him) |
| Ngiyom*shay*a | ngi<i1s>yo<fut>m<o1>shaya<vr> | (I will hit him) |

Figure 3.5. *Showing tags applied to two different verbs.*

45

The decision was therefore taken to adapt the verb root tag, and all other tags which refer to lexical items (i.e. noun roots, relative stem, adjectival roots, enumerative stem, quantifier roots and all interjections and conjunctions) to demonstrate the part of speech employed (using the tag from Ukwabelana Corpus) and the lexical item immediately following. The resulting tags are illustrated in Figure 3.6 below.

| | | |
|---|---|---|
| Ngizom*ceba* | ngi<i1s>zo<futZO>m<o1>ceba<vrCEBA> | (I will tell on him) |
| Ngiyom*shay*a | ngi<i1s>yo<futYO>m<o1>shaya<vrSHAYA> | (I will hit him) |

Figure 3.6. *Amended tags applied to the same two verbs.*

c. In order to preserve information about the original linguistic and orthographic words within which the formatives appeared, the tags were inserted into the SALT program in such a way that each code could be traced back to the word in which it appeared, by means of the 'word root table/expanded by words' function in SALT. The tags identified in Figure 3.6 above were therefore entered into SALT (using the bar symbol "|" to identify tags as per SALT conventions), as shown in Figure 3.7.

| | | |
|---|---|---|
| Ngizom*ceba* | ngizomceba|i1s ngizomceba|futZO ngizomceba|o1 ngizomceba|vrCEBA | (I will tell on him) |
| Ngiyom*shay*a | ngiyomshaya|i1s ngiyomshaya|futYO ngiyomshaya|o1 ngiyomshaya|vrSHAYA | (I will hit him) |

Figure 3.7. *Formative tagging within the SALT program.*

This was also an important feature to utilise since the tagging as well as all error and spell checking was done by hand; therefore, the presentation of data in this way with its original word to which the tag had been applied meant that it was more easily checked. The output of the SALT program's word root table function (expanded by words) is shown in Figure 3.8. To avoid over-counting of words, only the first column of numerals was used in frequency calculations.

46

| Formatives expanded by words | Total occurrences | Expanded occurrences |
|---|---|---|
| i1s | 2 | |
| NGIZOMCEBA | | 1 |
| NGIYOMSHAYA | | 1 |
| futZO | 1 | |
| NGIZOMCEBA | | 1 |
| futYO | 1 | |
| NGIYOMSHAYA | | 1 |
| o1 | 2 | |
| NGIZOMCEBA | | 1 |
| NGIYOMSHAYA | | 1 |
| vrCEBA | 1 | |
| NGIZOMCEBA | | 1 |
| VRSHAYA | 1 | |
| NGIYOMSHAYA | | 1 |

Figure 3.8. *Output of the SALT word root analysis expanded by words.*

d.   An examination of the grammatical structure of Zulu as outlined somewhat seminally by Doke (1939) and more recently by Taljaard and Bosch (1993) and Nyembezi (1982), in conjunction with the Ukwabelana Corpus morphological tagging system, revealed a shortage of tags to serve the purpose of this study. For instance, additional codes were added for enumerative qualification, interrogative stems, and adverbial formatives. Additionally the treatment of aspect and mode by Ukwabelana Corpus was changed for the present tagging system to be more closely aligned with the grammatical system described by Zulu grammar texts (Doke, 1939; Nyembezi, 1982; Taljaard & Bosch, 1993). Thus, in the place of the codes for <aspect>, and <modal root>, the new tagging system holds greater detail such as remote past, immediate past, present progressive and exclusive aspect.

The final tagging scheme for the study consisted of 285 basic formative tags, which are described in the tagging manual (Appendix J) where they are organised according to various

47

parts of speech in a structure following Zulu grammar textbooks (e.g. Doke, 1939; Nyembezi, 1982; Taljaard & Bosch, 1993).

### 3.7.2.7 Classification of formatives

To aid further analysis, formatives were also classified into content and structure formatives and were also analysed by the nine formative types (Doke, 1939), which were described in Chapter 2, Section 2.8. Four additional categories of formatives were added to the analysis as follows:

The ***miscellaneous category*** was ascribed to words of songs, lines used in ritual games and rhymes chanted in class. Any utterance which was part of these routines was counted as a single word, which is in keeping with the approach used in previous studies (Trembath et al., 2007) so as not to disproportionately count words which were uttered within certain contextual rituals.

***Interjections*** are words or phrases which can involve some kind of emotion and may be used (whether or not with emotion) as an interjecting remark (Paulos & Msimang, 1998). For example, *hhayi* (interjection of dissent/ negation/ surprise) would be tagged hhayi|intjHHAYI.

The ***vocative formative*** is really an interjection but it was extracted due to the distinct characteristic in that it is not used in isolation like the interjections above, but prefixed to names, or nouns, to either call or address. For example, in application to names, the transcript often contained *weCN* (where CN stood in the place of a child's name), and *yewena* (*wena* being the singular pronoun 'you'). The *we-* and *ye-* were therefore tagged and counted as a vocative formative.

Finally, the group termed '***names***' provides somewhat of an anomaly in that four different name categories were ascribed to proper nouns so as to maintain confidentiality in the data while still preserving the information regarding the participants' use of proper nouns. Names of other children were transcribed CN, names of teachers became TN, names of people and characters outside of context were NN, and names of places (e.g. towns) became PN. For example, *ePN* (*to/at PN) would be tagged; ePN|loc ePN|PN.

Apart from classifying formatives by these 13 types, formatives were also described by the part of speech they helped to form. This classification comprised 23 subtypes of formatives as displayed in Table 4.4, Section 4.5.1.

## 3.8 Pilot study

A pilot study was conducted to ascertain whether recruitment procedures, selection criteria and processes, methods of obtaining consent and assent, as well as materials and procedures for data collection were suitable to obtain the data needed to reach the research aims, and to make changes to these should this be required for the main study. The pilot study was conducted with two participants from one participating preschool. Participants were selected according to the selection criteria described in Table 3.1. Procedures outlined in Section 3.9 were followed. The children went on with their regular activities while wearing the equipment (see Figure 3.4). During the pilot, the children were observed to ascertain whether the equipment would interrupt their activities. During the pilot, the equipment was checked at intervals to ascertain its efficacy for the task. The recordings of the pilot study were used to determine whether the equipment rendered sufficiently clear recordings for the purposes of the study. Table 3.3 summarises the procedures and outcomes of the pilot study, as well as changes recommended for the main study. Although some minor changes were recommended, the pilot data collection yielded quality recordings that were representative of the child's speech. Neither did the equipment interfere or interrupt the children's participation in their activities. This data was therefore included in the main study.

Table 3.3

*Aims and Procedures of the Pilot Study*

| Aims | Materials | Procedure | Outcome | Amendments made for main study |
|---|---|---|---|---|
| To establish whether the selected method of recruiting child participants was effective. | Teacher nomination form | Forms with details about selection criteria were sent to the preschool teacher, who nominated children to participate. | Suitable children were nominated – both complied with the selection criteria. | None |
| To establish whether the selected method of obtaining informed consent from participants' parents/guardians was effective. | Parent/guardian information letter and consent form | The teacher sent the forms home to parents and organised their return prior to research commencing. | Forms were still being obtained in the morning of the day on which data collection was due to begin, and data collection was delayed by one hour 30 minutes. | Forms to be sent with first contact to the schools, follow up a week prior to data collection. |
| To ascertain whether child assent procedure was easily comprehended by participating children. | Assent script and procedure | Assent procedure was followed using the script and video recorded. | Children participated without any difficulty. | None |
| To determine whether the biographical questionnaires were clear and easily comprehended. | Home biographical and basic language background questionnaire | Forms were sent to families to be filled out and returned to the school. | The teacher noted that she took guardians' reading ability into consideration when nominating children. All the forms were completed satisfactorily. | The researcher undertook to check the forms during the data collection week and contact the guardians for further clarification when replies were unclear. |
| To determine that the recording equipment does not interfere with the children's activities. | Waist pouches, digital voice recorders and lapel microphones | Recording equipment was worn on waist pouches underneath clothing, the small microphone attached to the collar. The researcher remained at the school to monitor whether children were comfortable with this throughout the day. | Recording equipment was worn throughout the school day, and did not seem to interfere with any activity. One child reported the bag was too loose and this was tightened. The children did not report any further discomfort and teachers did not report observing any of the children displaying discomfort. | None |

| Aims | Materials | Procedure | Outcome | Amendments made for main study |
|------|-----------|-----------|---------|--------------------------------|
| To determine whether recoding equipment allowed for continuous recording of children's speech for the periods of time deemed suitable by the teacher. | Digital voice recorders | It was ensured that the recorder's batteries were fully loaded at the beginning of the day. The recorders were padded inside the pouches to stop accidental pressing of buttons. | Initially, recorders seemed to stop recording or switch off due to jolting and accidental pressing. This typically occurred after approximately one hour of recording. This was resolved by adding padding to the bags and activating the 'hold' setting which prevented this accidental stopping. | Always activate 'hold' setting when doing further recordings. |
| To establish whether the equipment chosen for data collection was capable to produce clear recordings suitable for transcription. | Digital voice recorders Head phones | The transcriber transcribed all the words uttered by the participating child from the recording. | The recordings were clear. Equipment was capable of recording the words of the participating child as well as others in close proximity, but loudness differences made it easy to distinguish whether it was the participant or another person speaking. | Transcribers were instructed to recognise the voice of the child among those of his/her peers and to transcribe only those words spoken by the participating child. |
| To test the tagging methods described for the Ukwabelana Corpus for the purposes of Zulu formative analysis . This open source information can be retrieved from **http://www.cs.bris.ac.uk/** Research/MachineLearning/ Morphology/Resources.jsp | Ukwabelana Corpus morphological tagging method | The researcher began applying the tagging conventions identified for the study to the pilot transcriptions. | The application of the tagging conventions shed light on some of the challenges. The direct use of the tags of Ukwabelana Corpus (Spiegler et al., 2010a, 2010b) was not able to consistently give comprehendible meaningful portions of speech but required some adaptation to be applicable to the study. | Further work was required to create a workable tagging convention which could effectively identify the smallest meaning-carrying units in the language for the purpose of this study The changes made are described in Section 3.7.2.6. The full tagging manual, which resulted and was used in this study, can be found in Appendix J. |

## 3.9 Procedures

### *3.9.1 Ethical issues*

Approval for the study was obtained from the Research Ethics Committee of the Faculty of Humanities, University of Pretoria (Appendix B), as well as from the KwaZulu-Natal Department of Education (see Appendix C).

When human participants are involved, ethical principles demand that 1) participants are not harmed; 2) participants give informed consent voluntarily prior to being included in the study; and 3) the privacy of participants is protected at all times (Leedy & Ormrod, 2014). In this study, these principles were upheld as follows:

The study did not hold direct risk of physical or emotional harm to participants. Recording equipment was fitted in such a way as not to interfere with normal activities in the preschool routine. Participants were asked to report any discomfort, whereupon equipment was adjusted or removed. Recordings of what participants said will not be shared with anybody apart from assistants who transcribed the recordings. Participants were informed of this during the assent process, ensuring that they would neither feel inhibited in their communication nor anxious about the content of what was recorded.

The process of obtaining consent is crucial to ensure that participants are fully aware of their role in the study as well as their autonomy and right to choose to participate (Leedy & Ormrod, 2014). As stated previously, the languages used in this area were primarily Zulu and English. Therefore, it was important that the information forms and questionnaires to parents utilised both English and Zulu without discrimination, to ensure that parents were able to freely choose the language of their preference and understand the content of the information letter without bias, in order to give informed consent. The researcher's contact details were provided so that parents could clarify any questions they had about the study. None of the parents used this option. In addition, for the child participants, information about the study was given to the children in child-friendly language supported with the use of pictures (Appendix G) to focus their attention and aid understanding of each statement being made.

52

Chapter 3: Methodology

Although the lapel microphones did at times pick up the speech of other children in the environment, consent from parents of these children was not obtained. This is in accordance with the research ethics guidelines from the University of Oxford (Central University Research Ethics Committee [CUREC], 2015), which state that when a lapel microphone is used to record the speech of a particular participant, only consent/proxy consent and assent from that particular participant is required, since the speech from others in the environment picked up by the microphone would not be analysed and would also not be able to be linked to specific individuals by the people who listen to/have access to the recording. There is therefore no risk of breach of privacy or of participating 'unwittingly' in the research project.

All participants were assured of their freedom to withdraw from the study at any point without fear of repercussions. Upon withdrawal, all data pertaining to them would be destroyed.

To protect the privacy of participants, identifying information was removed from data in transcription of speech where codes replaced people and place names which could be used to identify participants or participating schools. Recordings and all other data (e.g. questionnaires, consent forms, and assent forms) will be securely stored at the University of Pretoria until the year 2031, whereupon the data will be destroyed.

### 3.9.2 Data collection

Appropriate times for recording were discussed between the teacher and the researcher. The project was explained to the identified children as per the assent procedure (Appendix H) and their assent was sought. The children, having assented to participate, were given small pouches containing portable voice recorders to wear underneath their uniforms and a small microphone was attached to their collars. On the first day of recording, the researcher fitted the participating children with the pouches containing the digital voice recorders and the microphones at the appointed time. On subsequent days, either the researcher or a trained research assistant fitted the recorders at the appointed time. At each fitting, the children were reminded of what to do, that is, not to interfere with the equipment but ask for help if something was wrong. At each fitting, the children were given the opportunity to withdraw from the study should they have wished to do so. Children were also encouraged to carry on their normal activities. The researcher/assistant then removed herself from the view of the children for the

duration of the recording period but remained available in order to be able to assist with the adjustment or removal of equipment as necessary. The recording equipment was then removed at the appointed time by the teacher and given to the researcher/research assistant. Recording continued in this fashion until a total of 1500 words had been collected from each participant. The number of days and the recording time taken per participant is outlined in Table 3.4. Participants were assigned codes in order to preserve confidentiality. These codes contain a numeral which related to the site (1, 2 or 3) and a letter which refers to the gender, that is, girl (g) or boy (b).

Table 3.4

*Participant Data Collection*

| Participant | Recording time | Number of days |
|---|---|---|
| 1g | 3 hours 40 min | 2 |
| 1b | 3 hours 40 min | 2 |
| 2g | 6 hours 30 min | 3 |
| 2b | 6 hours 30 min | 2 |
| 3g | 2 hours | 1 |
| 3b | 2 hours 40 min | 2 |

### 3.9.3 Data preparation and analysis

#### 3.9.3.1 Transcription

The first 20 minutes of the total recording for each participant were not transcribed for analysis, in order to discard data that might have been affected by the novelty effects participants might have experienced due to wearing the recording equipment. A trained research assistant transcribed the recordings verbatim using the conventional orthography of Zulu into the Systematic Analysis of Language Transcripts (SALT) software program (Miller & Iglesias, 2012), following a set of transcription rules and conventions required by the SALT software (see Appendix I).

54

### 3.9.3.2 Tagging/coding

The researcher added the tags to the transcription (in the SALT program) according to the modified formative tagging system (see Section 3.6.2.5, and Appendix J) that was developed based on the tagging scheme for the Ukwabelana Corpus (Spiegler et al., 2010b) as described previously.

### 3.9.3.2 Data analyses

The coded transcriptions were analysed using the functions of the SALT program, as well as additional hand counts and descriptive statistics to determine the total number of formatives (TNF), and the total number of different formatives (TNDF) for individual samples, as well as collectively. The frequency per mille with which each formative occurred within the total sample was also calculated. A core vocabulary was then determined by applying a frequency cut-off as well as a commonality score criterion. A frequency cut off of 0.5‰ (per mille) calculated for each formative was used to identify high frequency and low frequency formatives. This is commensurate with other studies (e.g. Trembath et al 2007). The commonality of use of high frequency formatives across participants was established by counting how many participants used a particular high frequency formative. If the commonality score of a high frequency formative exceeded 1 (i.e. the formative was used by more than one participant), it was included in the core vocabulary. All formatives in the core vocabulary were then classified by types based on the classification by Doke (1939) – see Sections 2.8 and 3.7.2.7. Formatives were also classified as content versus structure formatives and by subtypes relating to parts of speech (see Section 4.5.1). Graphs and tables were used to display proportions and frequency distributions of the different types of formatives in the core vocabulary. The lexical meaning of the content formatives in the core vocabulary was also identified and examined in relations to previous studies.

### 3.9.4 Reliability of data

### 3.9.4.1 Transcriptions

Although the study aimed to investigate a substantial phenomenon, that is audible speech, the reliability of the measurement thereof will still be determined by the measurement instrument and how it is applied (McMillan & Schumacher, 2010). In this case this refers to the recording

55

and transcription of audible speech. In discussing their transcription software (SALT) Miller, Andriacchi, and Nockerts (2011) argue that the predetermined features of transcription rules significantly improve the consistency of transcription, which in turn leads to accurate calculations during the analysis phase. In addition to using the guidelines of the SALT (Miller & Iglesias, 2012), rigorous customised transcription rules were established (Appendix J).

The reliability of the transcription was also ascertained by a second transcriber. This research assistant, who was fluent in written and spoken Zulu, independently transcribed a proportion of the recordings amounting to 20% of the total data for each participant (Ayres & Ledford, 2014). The portions (a fifth from each participant) used for transcription reliability were randomly selected from the total length of the recording as determined by portions of time. For example, one participant's recordings were a total length of 6 hours and 30 minutes over three days' recording. The proportion required for reliability therefore was calculated to be 1 hour and 20 minutes (20% of 6 hours 30 minutes); a randomly selected portion of this participants' recordings, 1 hour and 20 minutes in length, was therefore independently transcribed. The same was done for all participants.

Percentage agreement (McMillan & Schumacher, 2010) between the original transcription and that of the second transcriber was calculated by the following formula (Cucchiarini, 1996):

$$Precentage\ agreement = \frac{number\ of\ agreements}{number\ of\ agreements + number\ of\ disagreements}\ x\ \frac{100}{1}$$

Percentage agreement between 80 and 100% is considered highly reliable (McMillan & Schumacher, 2010). An agreement was counted when a word was transcribed identically by the first and second transcriber. A disagreement was counted when the second transcriber added a word, or omitted a word, or transcribed a word differently to the first transcriber.

This study found a percentage of agreement of 81% for the transcription reliability, which is within the range reported by previous studies. Ambient noise in the class and playground may account for the differences in transcriptions.

56

### 3.9.4.2 Tagging reliability

In respect of the tagging reliability, a portion of 20% of the total data for the participants was re-coded independently by a research assistant in order to ascertain its reliability (Ayres & Ledford, 2014). This research assistant was fluent in written and spoken Zulu, and having a qualification of computer science, was accustomed to the use of codes and tags for computer analyses. Prior to the research assistant independently tagging the required portion of the transcriptions, he was first trained by the researcher in applying the tagging system through the use of the Tagging Manual (Appendix J) and a selected portion of transcription. The portions used in the final reliability tagging phase (a fifth from each participant) were randomly selected using the word count in the raw untagged transcript. That is to say that 300 words per participant were subjected to re-tagging by an independent tagger who applied the tagging conventions given in the tagging manual.

The tagging by the research assistant was compared with the corresponding sections of the tagging done by the researcher. Percentage agreement of the coding was calculated using the same formula mentioned in Section 3.9.6.1. The data sheet produced by the second rater was compared word for word to the same section in the primary data. A disagreement was counted when a different formative tag was used, when a tag was omitted, and when the raters had segmented words differently. The percentage of agreement was 93.8%. In light of McMillan and Schumacher's (2010) suggestion that percentage agreement should be at least 70% or more in order for data to be deemed reliable, 93.8% agreement suggests that data coding was executed reliably.

### 3.9.5 Validity

This design is vulnerable to various threats to validity. Regarding internal validity, the very observation can alter the observed; broadly termed participant reactivity (Leedy & Ormrod, 2014). To reduce this effect, it was important that procedures enabled the children to understand that the nature of the study was not to judge their skill and that their participation was voluntary. Practice runs or acclimatisation periods which are not counted as part of the sample may also allow time for adjustment to the new experience and reduce the novelty effect (McMillan & Schumacher, 2010). For this reason, the first 20 minutes of the recordings were omitted from the transcription and analysis. While research that aims to study real-world phenomenon is useful to

inform real-world intervention (Leedy & Ormrod, 2014), the reduced sample size of observational designs does present a limitation regarding the extent to which findings may be generalised, thus affecting the external validity of the design.

## 3.10 Summary

This chapter described the research methodology used, beginning with the main aim, sub-aims and research design utilised. The setting was described, followed by a description of the participants (recruitment, selection criteria and descriptive criteria). Materials and equipment used were also described. The aims, procedures, results and recommendations of the pilot study were discussed, as were the procedures for data collection and analysis, including ethical considerations and considerations regarding reliability and validity.

# CHAPTER 4

## RESULTS

### 4.1 Introduction

The results of the study are presented and discussed in this chapter. The results are presented in graphs and tables in accordance with the five sub-aims described in Chapter 3. These were: (i) to identify the linguistic formatives used by Zulu-speaking preschoolers, which is in accordance with the requirements outlined in Section 3.7.2.6 of Chapter 3 and fully described in Appendix J; (ii) to determine how many different formatives and formative types are used by Zulu-speaking preschoolers; (iii) to describe the vocabulary in terms of the most frequently and most commonly used formatives and thus determine which formatives, if any, constitute a core vocabulary; (iv) to further describe the characteristics of core vocabulary by formative types, as well as by content and structure formatives (including subtypes); and (v) to examine the lexical meaning of the content formatives of core vocabulary in relation to previous studies.

### 4.2 The determination of linguistic formatives used by Zulu-speaking preschoolers

The study gathered 1 500 spoken words (un-segmented, counted by use of orthographic space) from each of the participants, in total generating a sample total of 9 000 words. The data henceforth, unless otherwise specified, is reported as a collective set.

Prior to segmentation, this sample of 9 000 words yielded 3 203 different words. Of these, 308 were used with a frequency of 0.5‰ (per mille) and above. These 308 words accounted for 57.1% of the total sample.

At first impression, without tagging (as explained in Appendix J) applied, the same pattern of core vocabulary can be observed, whereby a greater number of words may actually contain a smaller number of unique words of which an even smaller group is used with high frequency. However, studies in English (e.g. Beukelman, Jones, & Rowan, 1989; Fallon, Light, & Paige, 2001; Trembath, Balandin, & Togher, 2007) Korean (Lee et al., 2005), and French monolinguals (Robillard et al., 2014) found that core vocabulary accounted for around 80% of the total samples.

On examination, the content of the Zulu list also differs, with conjunctions and interjections ranking highly, as opposed to syntax-related vocabulary typically found in core (Robillard et al., 2014; Trembath et al., 2007). Although other types of words such as verbs, demonstratives and possessives also appear in the list, due to the structure of Zulu, as already discussed, these words contain a cocktail of information specific to the context of the speaker and the object. This particular set of words may be useful when selecting vocabulary for AAC systems that are not intended to promote sentence construction and grammatical skills (e.g. activity- or phrase-based vocabulary for a communication board or an SGD with a limited number of messages) but that are primarily aimed at enhancing functional communication. This list may also be useful for rate enhancement, whereby the most frequently used words can be represented as a whole in the AAC system on a 'quick messages' page and accessed with a single hit rather than compiled from several pieces each time it is used. However, further segmentation of these orthographic words is required in order to lay the foundations of a more generative AAC system for Zulu.

The vocabulary was then tagged with the formative codes described in Section 3.7.2.6, Chapter 3 and Appendix J. This increased the number of vocabulary items, as the words were now segmented into respective formatives used in their construction.

The total sample size rose to 20 137 formatives.

## 4.3 Number of different formatives used

The sample size, with words now segmented, totalled 20 137 formatives (total number of formatives – TNF), and included 1 110 different formatives (total number of different formatives – TNDF). This being a large bulk of unique formatives, for the purpose of analysis it is helpful to group the formatives into their various types. Formative types were described in Section 2.8 of Chapter 2. As explained in Section 3.7.2.6 of Chapter 3, in addition to the nine types identified by Doke (1939), a further four types (interjections, vocative formatives, miscellaneous and names) were added, resulting in 13 types of formatives. The formatives found in the sample are grouped by these types and are shown in Table 4.1 ranked according to the TNDF occurring in each group.

Table 4.1

*Total Number of Different Formatives (TNDF) and Total Number of Formatives (TNF) Grouped by Formative Type*

| Formative type | TNDF | TNF |
|---|---|---|
| roots/stems | 766 | 6 007 |
| miscellaneous | 121 | 161 |
| ideophones | 64 | 119 |
| interjections | 43 | 701 |
| prefixal formatives | 27 | 2 826 |
| verbal auxiliaries | 20 | 1 913 |
| concords | 19 | 5 288 |
| noun prefixes | 14 | 1 658 |
| conjunctions | 13 | 176 |
| enclitic formatives | 11 | 304 |
| vocative formatives | 6 | 353 |
| names | 4 | 625 |
| suffixes | 2 | 6 |
| **TOTAL** | 1 110 | 20 137 |

Roots and stems, which are the primary lexical (content) component of the word, form the largest bulk of the formatives, with 766 different formatives, containing noun and verb stems, and relative, adjective and adverbial roots.

The miscellaneous category was ascribed to words of songs, lines used in ritual games and rhymes chanted in class, which were clustered together and counted as a single entry so as not to disproportionately count formatives which were uttered within certain contextual rituals. This contributed a large portion (121) of different formatives.

Ideophones are described by Doke (1939) as root formatives but differ from other roots fundamentally in that they do not employ concords (Nyembezi, 1982). These words function to describe predicates in manner, sound or action (Doke, 1939). As such, they are highly varied. The sample proffered 64 different ideophones.

Chapter 4: Results

Interjections are words or phrases used as an interjecting remark (Paulos & Msimang, 1998). Forty-three different interjections were used in the total sample; for example, *mi* ('here take') and *hhayi* (interjection of dissent; negation; surprise).

Prefixal formatives are used to form pronouns, copulatives and adverbs; 27 different prefixal formatives were found in the total sample.

Verbal auxiliaries provide verb inflections, tenses and implications. The total sample provided 20 different verbal auxiliaries.

Concords are formatives which function to align all words in the sentence to agree with the noun class of the noun which is the subject of that particular sentence; 19 different concords were found in the total sample.

The noun prefix group, which are used to form nouns varying by class, gave 14 different kinds in the total sample.

Thirteen different conjunctive formatives (including words used as conjunctions according to the tagging manual [Appendix J], for example, *kanti* ['just so,' 'in fact'] and *uma* ['if,' 'when']) were used in the total sample.

Enclitic formatives which modify ideas in phrase construction, such as ideas expressed through verbs and interrogative expressions, proffered 11 different formatives.

The vocative formative is really an interjection but is often prefixed to names and nouns to call or address. For example, *weCN* (where CN stands for child name). There were six different formatives in this group.

The group termed 'names' with four different tags (i.e. CN –child name; TN – teacher name; NN – name or character not present; PN – place name) were ascribed to proper nouns so as to maintain confidentiality in the data while still preserving the information regarding the participants' use of proper nouns. Four different formatives were therefore found in the total sample.

62

Finally, suffixes are used to form various alterations to adjectives, nouns, verbs and relatives. There were two different formatives in this group.

Regarding the total number of formatives (TNF) by type, it is interesting to note from Table 4.1 that roots and stems were the group occurring most frequently in the sample – this was also the group containing the highest number of different formatives (766). In contrast, concords were the group occurring with the second highest frequency, although there were only 19 different concords. The noun class system of Zulu (as described in Section 2.7) demands the frequent use of concords. Since only a finite number of concords exist, the number of different concords remained limited.

## 4.4 Most frequently and commonly used formatives

The frequency of occurrence per mille (‰) was calculated for each individual formative identified in the sample, by dividing the total number of occurrences of this formative by the TNF and multiplying by one thousand (see formula below)

$$\frac{total\ number\ of\ occurances}{TNF}\ x1000 = frequency\ ‰$$

The whole list of formatives identified through the tagging system was then ranked by frequency of occurrence. The pre-determined frequency cut off of 0.5‰ was applied, yielding a list of 223 formatives. These 223 formatives accounted for a total of 17 934 formatives used, thus accounting for 89.06% of the total sample.

Further analysis of this core list in terms of commonality revealed that of these 223 formatives, 127 were used by all participants, giving a commonality score of six. Forty-two were used by five participants, 29 were used by four of the children, 15 were used by three of the children, eight were used by only two participants, and two were used by only one participant. Table 4.2 demonstrates the commonality distribution described.

Table 4.2

*Commonality Distribution Among High Frequency Formatives*

| Number of different formatives | Commonality score |
|---|---|
| 127 | 6 |
| 42 | 5 |
| 29 | 4 |
| 15 | 3 |
| 8 | 2 |
| 2 | 1 |

The commonality score of 2 and above was applied as an inclusion criteria as the formatives with a lower commonality score were particular to a single participant and, as such, could be considered personal core, rather than common. Both frequency and commonality inclusion/exclusion criterion are arbitrary cut off points, but have been used in previous studies to establish a set of core vocabulary (Banajee et al., 2003; Beukelman et al., 1989; Fallon et al., 2001; Fried-Oken & More, 1992; Robillard et al., 2014; Trembath et al., 2007).

This commonality criteria then applied, a core vocabulary of 221 formatives was determined. These 221 formatives accounted for 17 900 of the total sample, or 88.9% of the total sample.

A list of all 221 formatives arranged from most to least frequently occurring formative can be found in Appendix K.

**4.5 Further description of the core vocabulary**

Table 4.3 describes the formatives comprising the core vocabulary by formative type.

64

Table 4.3

*TNDF and TNF in Core Vocabulary Grouped by Formative Type*

| Formative type | Example (formative underlined) | TNDF | TNF | % in total sample |
|---|---|---|---|---|
| roots/stems | *ngizoku<u>ceba</u>* | 120 | 4 248 | 21.1 |
| prefixal formatives | *<u>lo</u>muntu* | 24 | 2 811 | 13.9 |
| verbal auxiliaries | *ngi<u>ya</u>dlala* | 18 | 1 907 | 9.5 |
| concord | *<u>ngi</u>cela* | 17 | 5 278 | 26.2 |
| interjection | *<u>yebo</u>* | 13 | 593 | 2.9 |
| noun prefixes | *<u>i</u>gama* | 12 | 1 648 | 8.2 |
| enclitic formative | *ungishaye<u>lani</u>* | 5 | 285 | 1.4 |
| conjunction | *<u>uma</u>* | 5 | 149 | 0.7 |
| names | *CN*[a] | 3 | 621 | 3.1 |
| vocative formative | *weCN* | 3 | 348 | 1.7 |
| miscellaneous | *wenkela_wenke_awenke*[b] | 1 | 12 | 0.05 |
| **TOTAL** | | 221 | 17 900 | 88.8911 |

[a] This code stands for 'child name' in order to protect the privacy of the children referred to in the sample
[b] These words formed part of a song/rhyme and were therefore assigned one code.

As shown in Table 4.3, roots and stems account for the largest group of different formatives (with 120 different formatives) within the core vocabulary identified. However, their collective frequency of use within the total sample is lower (at 21.1%) compared to the smaller group of concordial formatives, which number only 17 different formatives but collectively account for over 26% of the total sample. Twenty-four different prefixal formatives make up 13.9% of the sample. Eighteen different verbal auxiliary formatives comprise just over 9.5% of the total sample, and just 12 noun prefixes account for 9% of the total sample. Thirteen different interjections account for only 2.9% of the sample. Three different vocative formatives account for nearly 2% of the total sample; five different enclitic formatives account for 1.4%; and five different conjunctions account for just 0.7% of the sample. Figure 4.1 gives a visual representation of these proportions of formative use within core vocabulary, while Figure 4.2 demonstrates the number of different formatives occurring in the core within each group of formatives.
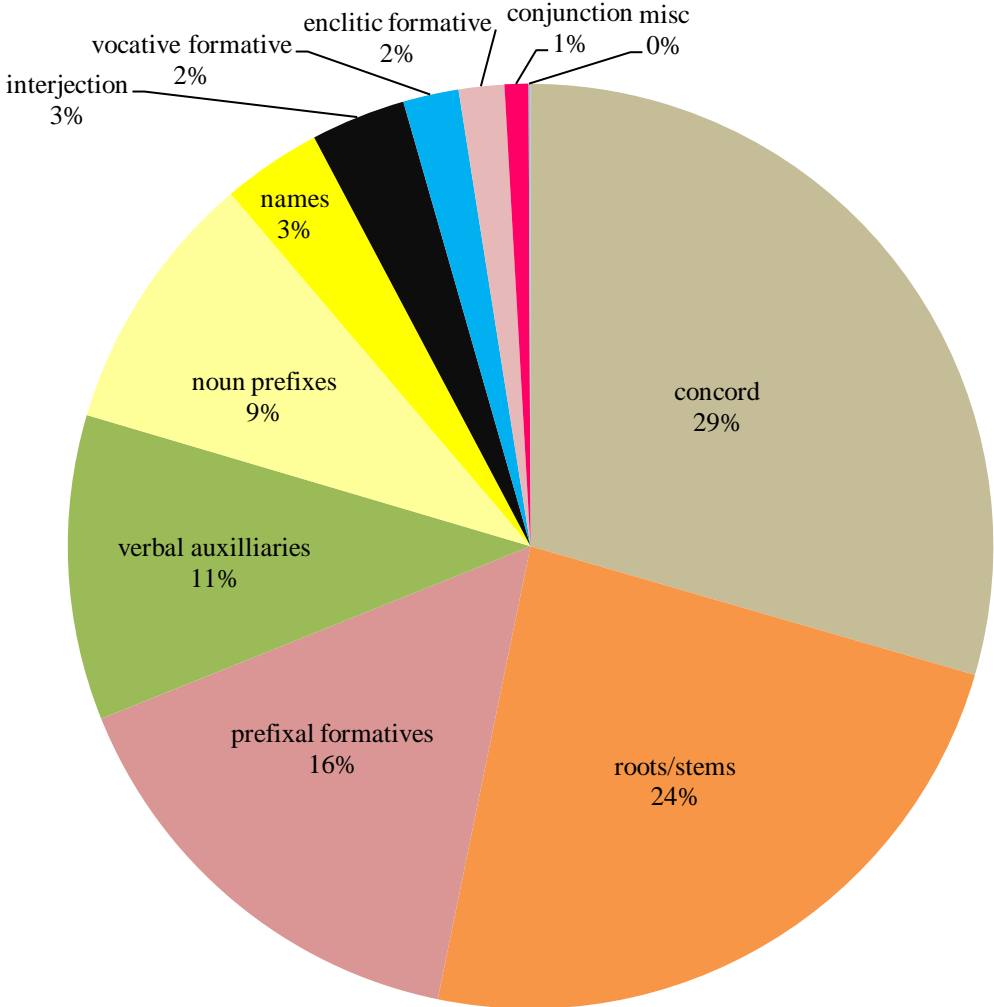
Figure 4.1. *Proportions of formative types within core vocabulary.*
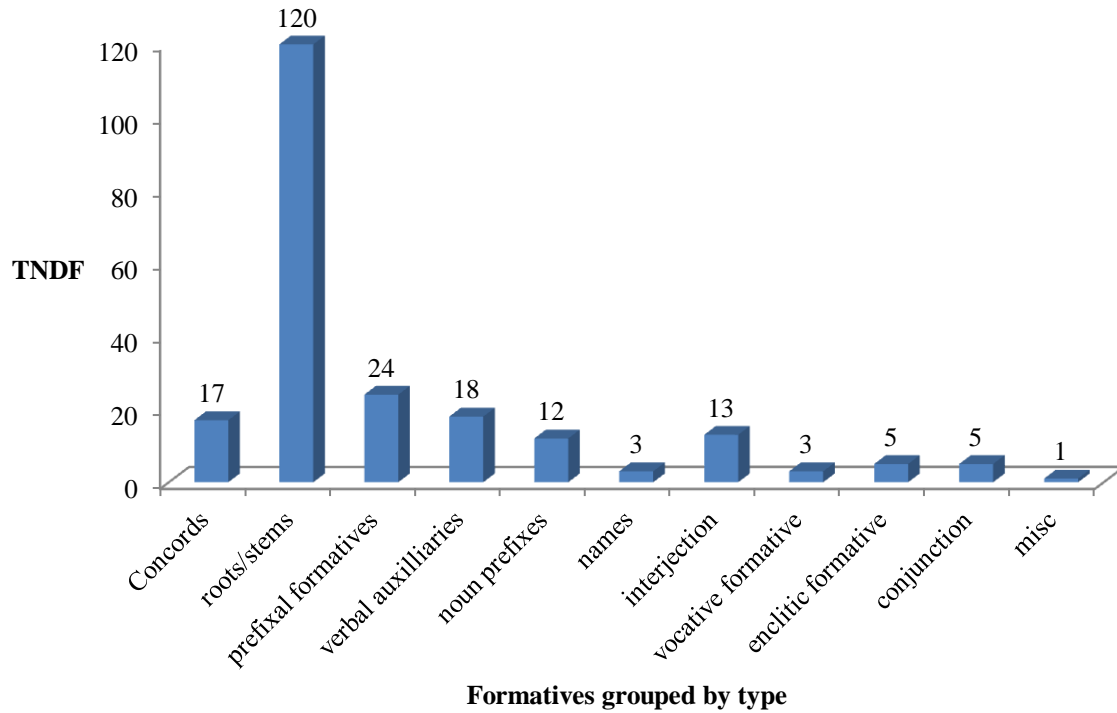
Figure 4.2. *Number of different formatives (grouped by type) within core vocabulary.*

From the two figures, it is clear that concordial formatives as a group occur most often – comprising 29% of the core vocabulary, although there are only 17 different ones. Roots and stems make up 24% of the core vocabulary but are comprised of 120 different formatives.

### 4.5.1 Content and structure formatives

For further analysis of this core, it is helpful for the formatives to be grouped into two broad categories of content and structure. As described in Chapter 2, Sections 2.7 and 2.7.1, the content formatives are formatives that carry lexical meaning, for instance various roots and stems such as noun, verb and relative stems. Structure formatives (as described in Sections 2.7 and 2.7.2) are those which permit the meaningful use of the content formatives by relating them to one another. These are formatives such as concords, adverbial formatives and verbal auxiliaries. Figure 4.3 demonstrates the proportions of structure to core formatives, while Figure 4.4 shows the number of different formatives in the respective categories.

Figure 4.3. *Proportion of structure and content formatives within core vocabulary.*
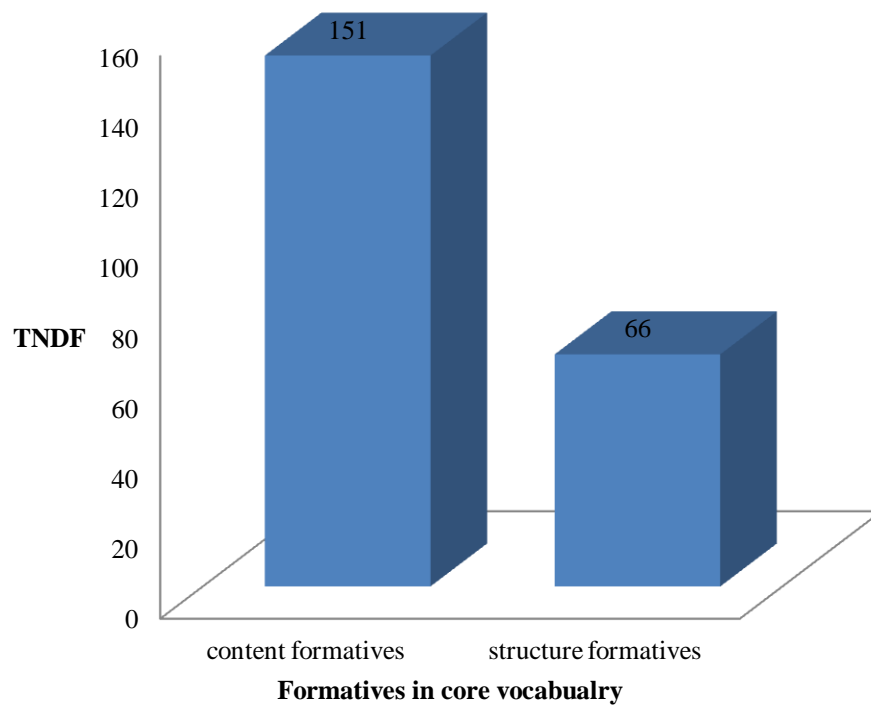


Figure 4.4. *Number of different structure and content formatives within core vocabulary.*

68

Chapter 4: Results

Sixty-six different structure formatives which are directly related to language structure comprise the bulk of the core vocabulary (61% of core formatives), while 151 different content formatives make up the smaller proportion of the core, at 39% of all core formatives.

Table 4.4 shows the content of these two broad categories framed within the parts of speech in which they occur, as described by Zulu grammar texts (Doke, 1939; Nyembezi, 1982; Taljaard & Bosch, 1993). The overarching part of speech (second column in the table) is mirrored in the tagging manual, as the overall organisation of the tagging system (Appendix J).

The two broad categories of formatives (content and structure) found in the core vocabulary will now be considered individually. The formatives within each category (framed now by parts of speech) will be described according to the number of different formatives found per group within the core vocabulary, the number of times these 'core formatives' were used by the children, and the frequency (expressed as a percentage) with which formatives within each category were used within the total sample, as calculated by the following formula

$$\frac{total\ number\ of\ occurances\ of\ formative\ in\ total\ sample}{total\ number\ of\ formatives\ in\ total\ sample}\ x100$$

69

Table 4.4

*Core Vocabulary Formatives Grouped as They Relate to Grammatical Parts of Speech*

| | Overarching part of speech | Formative | TNDF | TNF | Frequency with which these formatives appeared in the total sample (%) |
|---|---|---|---|---|---|
| Content formatives | Substantive | noun roots | 30 | 809 | 4.0 |
| | | pronouns | 14 | 1 450 | 7.2 |
| | Qualificative | relative roots | 4 | 192 | 0.9 |
| | | adjective roots | 11 | 215 | 1.1 |
| | | enumerative roots | 1 | 73 | 0.4 |
| | Predicative | verb roots | 62 | 2 525 | 12.5 |
| | Descriptive | adverb roots | 11 | 423 | 2.1 |
| | Conjunction | conjunctions | 5 | 149 | 0.7 |
| | Interjection | interjections | 13 | 593 | 2.9 |
| **Subtotal** | | | **151** | **6 429** | **31.8%** |
| **Structure formatives** | Substantive | noun prefixes | 12 | 1 648 | 8.2 |
| | | demonstrative formatives | 3 | 354 | 1.7 |
| | | presentative formatives | 2 | 145 | 0.7 |
| | Qualificative | relative concords | 1 | 297 | 1.5 |
| | | adjective concords | 1 | 111 | 0.5 |
| | | enumerative concords | 1 | 46 | 0.2 |
| | | possessive concords | 1 | 394 | 1.9 |
| | Predicative | verb auxiliaries | 20 | 1 986 | 9.9 |
| | | verb concords | 13 | 4 430 | 21.8 |
| | Descriptive | adverbial formatives | 5 | 810 | 4.0 |
| | Interrogative | enclitic formatives (interrogation) | 3 | 183 | 0.9 |
| | Phrase | enclitic formatives (vocative) | 4 | 434 | 2.2 |
| **Subtotal** | | | **66** | **10 838** | **53.8%** |
| **Names** | | proper names | 3 | 621 | 3.1 |
| **Miscellaneous** | | rhymes, games and songs | 1 | 12 | 0.1 |
| **TOTAL** | | | 221 | 17 900 | 88.9 % |

*4.5.1.1 Subtypes of structure formatives of the core vocabulary*

Figure 4.5 gives a visual representation of the different subtypes of structure formatives identified in the core vocabulary – classified by the part of speech they help to form.
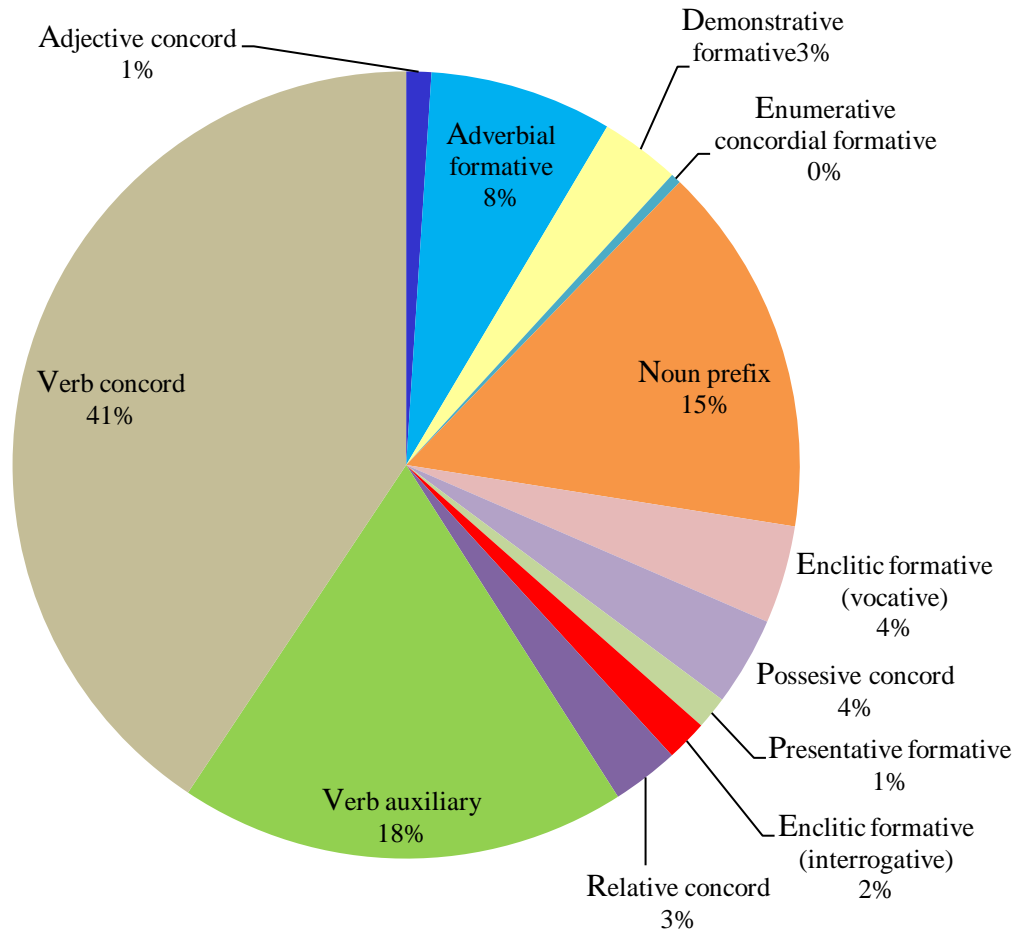


Figure 4.5. *Proportional distribution of core vocabulary structure formatives.*
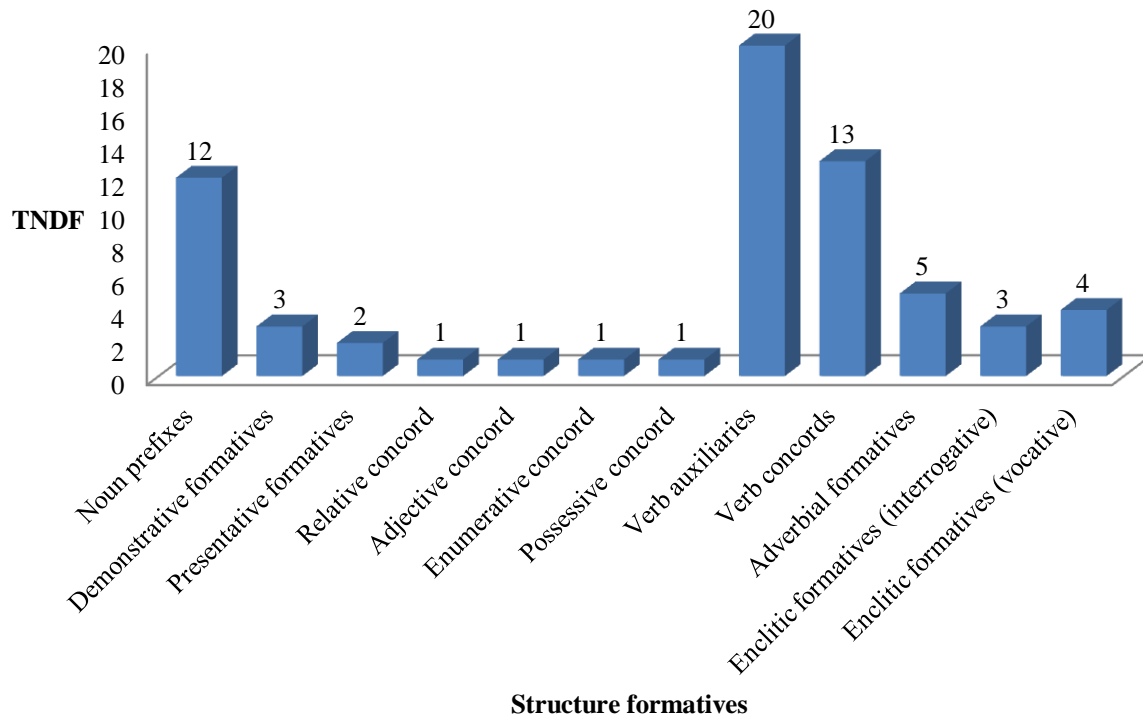
Figure 4.6. *Number of different structure formatives used within core vocabulary.*

Figure 4.6 shows that there were 20 different verb auxiliary formatives, 13 different verb concords[5], and 12 different noun prefix formatives. It is interesting to note that these three subtypes of formatives also accounted for the largest proportion of structural formatives found in the core vocabulary, accounting for 41%, 18% and 15% of the core respectively (as indicated in Figure 4.5).

*4.5.1.2 Subtypes of content formatives of core vocabulary*

Figure 4.7 shows the proportional distribution of different subtypes of formatives, while Figure 4.8 shows the number of different content formatives per subtype that were found in the core vocabulary.

---

[5] With the formatives now categorised by parts of speech, the concords were categorised as adjective, relative, enumerative, possessive and verb concords. The first four types consisted of one concord each and the verb concords of 13 different ones, making up the initial 17 concords reported within core vocabulary.

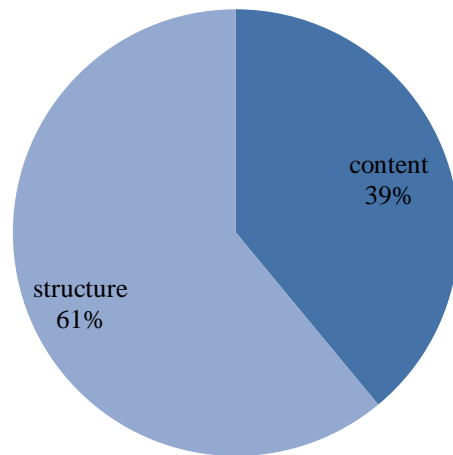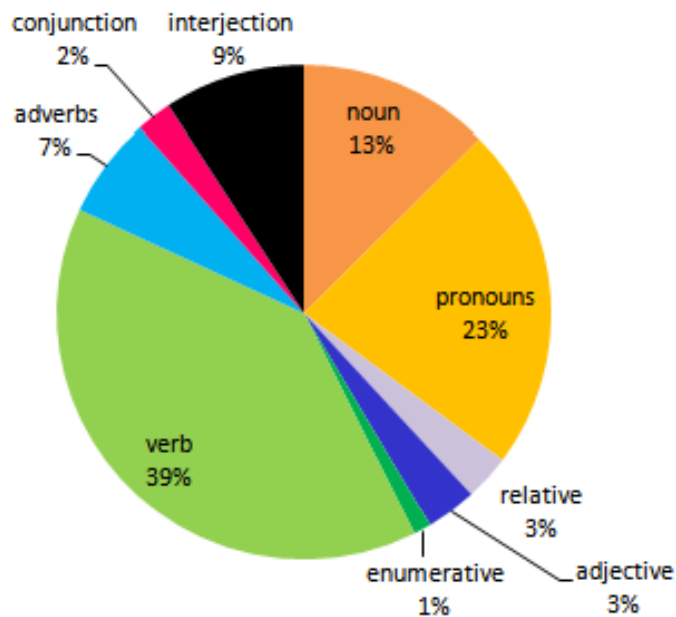Figure 4.7. *Proportional representation of content formatives within core vocabulary.*
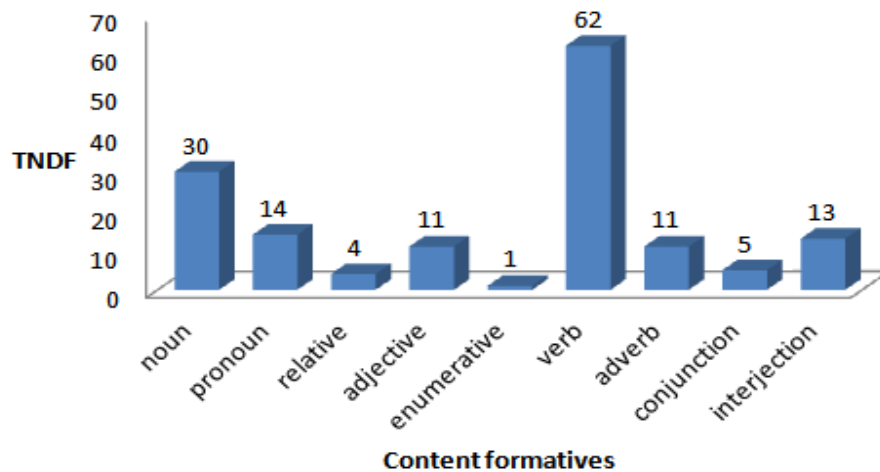


Figure 4.8. *Number of different content formatives used within core vocabulary.*

Most content formatives were verb roots (39% of content formatives) with 62 different verb roots being used. Thirty different noun roots accounted for 13% of the content formatives, and just 14 different pronouns accounted for 23% of the content formatives used in the sample. Thirteen different interjections accounted for 9% of content formatives, 11 different adjective

73

roots made up 3%, 11 adverbs made up 7%, five different conjunctions accounted for only 2%, four relative stems accounted for 3%, and just one enumerative stem accounted for 1% of the content formatives used.

The full list of formatives can be found with English translation of content formatives (according to the dictionary by Dent & Nyembezi, 1995) in Appendix K. With each formative, Appendix K also gives the Zulu word most frequently utilising this formative from the composite sample, as well as an English translation of that Zulu word.

## 4.6 The lexical meaning of content formatives in the core vocabulary

The fundamental differences between Zulu and English language structure complicate comparisons between English core vocabulary lists and Zulu core vocabulary identified in this study. Structure formatives, which comprise the bulk of the Zulu core vocabulary, are not comparable to English in a word-for-word manner, and circumscribing the meaning of these formatives can be misleading (Paulos & Msimang, 1998). However, it may be useful to compare the lexical meanings which can be observed in the content formatives to English core vocabulary lists. Further examination of the top 30 content formatives is given in Table 4.5. The Table also shows whether words with equivalent meanings could be found in four English core vocabulary lists.

Table 4.5

*Top Thirty Most Frequently Occurring Content Formatives of the Current Study Compared to Other Core Word Lists*

| | | | | | | | | | Presence of similar word(s) in other core vocabulary lists | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ranking | Formative | Tag expounded | Number of occurrences | Composite frequency ‰ | Commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list | Fried-Oken & More (1992) | Beukelman, Jones & Rowan (1989) | Trembath et al (2007) | Boenisch & Soto (2015) |
| 5th | pr1s | Pronoun first person singular | 438 | 21.75 | 6 | I, me, myself | *Mina, na<u>mi</u>, okwa<u>mi</u>* | Me, and me, it's mine | ✓ | ✓ | ✓ | ✓ |
| 15th | intjYEBO | Interjection YEBO | 288 | 14.30 | 6 | yes | *Yebo* | Yes | ✓ | ✓ | ✓ | ✓ |
| 18th | pr2s | Pronoun second person singular | 270 | 13.41 | 6 | you | *Wena, na<u>we</u>, wak<u>ho</u>* | You, and you, yours | ✓ | ✓ | ✓ | ✓ |
| 24th | vrITHI | Verb root ITHI | 225 | 11.17 | 6 | say; intend; think | *Awu<u>thi</u>* | Say, let me, | ✓ | ✓ | ✓ | ✓ |
| 29th | advLA | Adverbial formative LA | 165 | 8.19 | 6 | here | *Woza <u>la</u>* | Come here | ✓ | ✓ | ✓ | ✓ |
| 32nd | pr1 | Pronoun class 1 | 135 | 6.70 | 5 | he; him; himself; she; her; herself | *Yena, na<u>ye</u>, la<u>khe</u>* | He/she, and him/her, his/hers | ✓ | ✓ | ✓ | ✓ |
| 35th | pr9 | Pronoun class 9 | 124 | 6.16 | 6 | it; itself | *Yona, akusi<u>yo</u>* | It, its it not it | ✓ | ✓ | ✓ | ✓ |
| 38th | pr15 | Pronoun class 15 | 120 | 5.96 | 6 | it; itself | *Kona, nak<u>ho</u>* | It, and it | ✓ | ✓ | ✓ | ✓ |

| Ranking | Formative | Tag expounded | Number of occurrences | Composite frequency ‰ | Commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list | Fried-Oken & More (1992) | Beukelman, Jones & Rowan (1989) | Trembath et al (2007) | Boenisch & Soto (2015) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 39th | PRS | Presentative formative | 120 | 5.96 | 6 | here 'they' are/ here 'it' is / here 'he' is etc | *Nakhu* | Here it is | | | | |
| 40th | vrBUKA | Verb root BUKA | 117 | 5.81 | 6 | look at; watch; admire | *Buka* | Look (imperative) | ✓ | ✓ | ✓ | ✓ |
| 42nd | vrBONA | Verb root BONA | 114 | 5.66 | 6 | see; understand; give regards | *Uyabona* | Do you see? | ✓ | ✓ | ✓ | ✓ |
| 44th | vrUKUYA | Verb root UKUYA | 106 | 5.26 | 6 | go to | *Uyaphi* | Where are you going? | ✓ | ✓ | ✓ | ✓ |
| 46th | vrFUNA | Verb root FUNA | 105 | 5.21 | 6 | search; x; desire | *Ufuna* | You want | ✓ | ✓ | ✓ | ✓ |
| 49th | vrDLALA | Verb root DLALA | 98 | 4.87 | 6 | play; dance; frolic | *Ngiyadlala* | I am playing/joking | ✓ | ✓ | ✓ | ✓ |
| 50th | vrHAMBA | Verb root HAMBA | 97 | 4.82 | 6 | walk; go; travel | *Hamba* | Go (imperative) | ✓ | ✓ | ✓ | ✓ |
| 56th | nrUMISI | Noun root UMISI | 81 | 4.02 | 6 | lady teacher | *uMisi* | Miss | | | ✓ | ✓ |
| 57th | rPHI | Relative root PHI | 81 | 4.02 | 6 | where | *Kuphi* | Where | ✓ | ✓ | ✓ | ✓ |
| 59th | vrIPHA | Verb root IPHA | 78 | 3.87 | 6 | give | *Ungiphe* | You give me | ✓ | ✓ | ✓ | ✓ |
| 60th | drNJE | Demonstrative relative NJE | 76 | 3.77 | 6 | such as this; like this | *Nje, Kanje* | ~just~ | | | | |

76

| Ranking | Formative | Tag expounded | Number of occurrences | Composite frequency ‰ | Commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list | Fried-Oken & More (1992) | Beukelman, Jones & Rowan (1989) | Trembath et al (2007) | Boenisch & Soto (2015) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 61st | vrCELA | Verb root CELA | 76 | 3.77 | 5 | request; negotiate for a wife; be almost | *Ngicela* | Please | ✓ | ✓ | ✓ | ✓ |
| 62nd | pr3 | Pronoun class 3 | 76 | 3.77 | 4 | it; itself | *Wona, ngawo, akuwona* | It, with it, it is not it | ✓ | ✓ | ✓ | ✓ |
| 64th | eNI | Enumerative root NI | 73 | 3.63 | 6 | what; of what sort | *Ini, wenzani* | What, what are you doing | ✓ | ✓ | ✓ | ✓ |
| 65th | advKHONA | Adverb root KHONA | 71 | 3.53 | 6 | of place; here; there | *Khona, akekho* | There, not there. | ✓ | ✓ | ✓ | ✓ |
| 67th | vrBHALA | Verb root BHALA | 70 | 3.48 | 6 | write; write an examination | *Ukubhala* | To write | | | | ✓ |
| 68th | vrCEBA | Verb root CEBA | 70 | 3.48 | 5 | report someone; invent; plot against; be rich | *Ngizokuceba* | I will tell on you | | ✓ | ✓ | ✓ |
| 70th | pr5 | Pronoun class 5 | 64 | 3.18 | 5 | it; itself | *Lona,lilodwa* | It, there is one | ✓ | ✓ | ✓ | ✓ |
| 72nd | vrWOZA | Verb root WOZA | 60 | 2.98 | 6 | come | *Woza* | Come (imperative) | ✓ | ✓ | ✓ | ✓ |
| 73rd | vrAZI | Verb root AZI | 58 | 2.88 | 6 | know | *Angikwazi* | I don't know it/how | ✓ | ✓ | ✓ | ✓ |
| 75th | vrSHAYA | Verb root SHAYA | 56 | 2.78 | 6 | strike; punish; play (as an instrument) | *Ngizokushaya* | I will hit you | | | | ✓ |
| 77th | nrINTO | Noun root INTO | 54 | 2.68 | 6 | thing; object | *Lento* | This thing | ✓ | ✓ | ✓ | ✓ |

It is important to note that these are not the top 30 formatives but in fact the 30th content formative is in fact the 77th formative of core vocabulary as ranked by frequency.

Only six of the top 30 content formatives are not reflected in all four comparisons published English core vocabulary lists. Of these six, one formative (verb root *ceba* 'tell') is reflected in three lists, one formative (noun root *uMisi* 'Miss' or 'teacher') is reflected in two lists, two formatives (verb root *bhala* 'write' and verb root *shaya* 'hit') are reflected in only one list. Two formatives (presentative formative tagged PRS e.g. *nakhu* 'here it is' and demonstrative relative stem *nje* 'such as this;' 'like this') are not reflected in the comparison lists.

## 4.7 Conclusion

The results of the study were presented and discussed in this chapter. The results were presented through the use of graphs and tables in accordance with the five sub-aims described previously in Section 3.7.2.6 of Chapter 3. First, the linguistic formatives used by Zulu-speaking preschoolers were described. Second, the number and types of different formatives used by the preschoolers were described. Third, a core vocabulary was identified by describing the most frequently and commonly used formatives. Fourth, the characteristics of core vocabulary were described by formative types, as well as by content and structure formatives (including subtypes). Fifth, the lexical meaning of the content formatives in the core vocabulary was examined in relation to previous studies.

# CHAPTER 5

# DISCUSSION

## 5.1 Introduction

The results of the study are discussed in this chapter. The parameters (total number of formatives, number of unique formatives) of the total Zulu vocabulary identified from the spoken language of Zulu-speaking preschoolers are explored and compared to other studies. Similarly, the parameters of the core vocabulary (frequency and commonality scores) identified in this study are compared to that of previous studies. Further, the lexical meaning of the content formatives in the core vocabulary determined in this study is discussed and compared with English core vocabulary lists. Finally, the implications that a morphologically rich language like Zulu has for the selection of vocabulary and design for augmentative and alternative communication (AAC) systems are discussed.

## 5.2 The vocabulary and core vocabulary parameters

The composite sample collected for this study comprised 20 137 formatives (TNF), of which 1 110 were unique (TNDF). The ratio of total number of formatives to number of unique formatives is about 18:1. The proportion varies but this pattern of unique words outnumbering the total number of words by a considerable margin compares to that of English vocabulary studies, for instance Trembath et al.'s (2007) composite sample of 18 000 words contained 1 411 unique words (ratio ~13:1); while Fallon et al's (2001) composite sample of 5 000 words contained 641 unique words (ratio ~ 8:1). Among these English studies it appears that the ratio may be linked to sample size, where the number of unique words does not increase proportionally to an increased total sample. This would once again corroborate the existence of a core vocabulary of a small set of words that are used repeatedly (Trembath et al., 2007). In Korean, however, the proportion differed somewhat, with Shin and Hill's (2016) composite sample of 16 944 words having 3 669 unique words (ratio ~ 5:1). This may be due to differences between the languages themselves, though it may also result from a difference in what the authors regarded as a unit of analysis within the study. For instance, Shin and Hill (2016) describe the inclusion in their analysis of morphological information described as suffixes in an

79

elaborate verb system as well as various particles through which syntactic relationships are expressed.

In this study, 221 formatives accounted for 88.9% of total sample. The proportional representation of core to fringe vocabulary is similar to other studies in other languages with monolingual subjects. For instance, Trembath et al. (2007) determined a core of 263 words, which accounted for 79.8% of the total sample in Australian preschool children speaking English. Boenisch and Soto's (2015) monolingual English participants used 200 words for 80% of their recorded communication, while Robillard et al's (2014) monolingual French participants made use of 216 words for 80.15% of their communication. Studies in Mandarin Chinese reported a larger list of lexemes in the core vocabulary. Liu and Sloane (2006) reported 1 000 lexemes which accounted for over 90% of the communication of their Mandarin Chinese participants, while Shin and Hill (2016) reported 219 Korean words accounting for just 60.82% of their composite sample. Again it is conceivable that these disparities may be accounted for by differences in linguistic characteristics as well as perhaps the definition of the unit counted to arrive at frequency scores used in establishing a core vocabulary for AAC. Baker and Chang (2006) recommend consistent handling of linguistic decisions around word definition, without which word frequency studies would be meaningless. This may also explain the difference in core presentation within Korean studies, since Lee, Kim and Park (2005) reported 276 core vocabulary items accounting for over 81% of the sample, compared to Shin and Hill's 219 words accounting for 60.82% (Shin & Hill, 2016). In Zulu, the disparity between the determination of core with and without morphological tagging has already been discussed, and the results shown.

While some observations about the influence of gender and site can be made, the sample size of this study prohibits any definite conclusions based on general patterns of commonality. For example, only girls used the verb -gqoka ('to dress/wear'), while only boys used the verb –yekela ('leave /let go'). With a larger sample size, one could also comment on communication styles influencing core vocabulary, for instance all the girls and just one boy said sorry, while all the boys and just one girl used the interjection O of astonishment. However, the sample is too small to draw any conclusions or indeed to make a comparison.

Chapter 5: Discussion

Comparisons by site are also prohibited by the sample size, although one may comment, for instance, that both the English and Zulu way of counting is reflected in the core, with Site 2 participants using the adjectival root *-ne* ('four'), while the participants from the other sites used the English root 'four'. A number of other English words are also found in the core (see Appendix K), for example, interjection 'OK,' 'toilet' and *uMisi* 'Miss'. Code switching and the mixing of languages is common and even inevitable in multilingual societies (Ndimande-Hlongwa & Ndebele, 2014) such as South Africa, and has been the focus of considerable study (Ramsay-Brijball, 1999). Various descriptions arise around terminology; whether the switch occurs at a sentence level (inter-sentenial) or deeper within, by the addition of morphemes or lexical items (intra-sentenial; Poplack, as cited in Ndimande-Hlongwa & Ndebele, 2014). This intra-sentenial language mixing appears in the Zulu core vocabulary determined here. For example, in the children's use of the English word 'crayon' as both a noun *u*crayon and a verb *uku*crayon*a,* the children employ the structure of Zulu and modify the English lexical item 'crayon' to function like a Zulu noun or verb root. Of course, the very nature of language by definition is changeable (Besio & Chinato, 1996; Smith, 2006). As such, the point at which a word becomes adopted and is no longer distinguished as foreign is a subject for another debate. Others examine and postulate the reasons for this mixing of languages by speakers. For some time it was thought that it was indicative of a poor command over either language (Lipski, as cited in Ndimande-Hlongwa & Ndebele, 2014), though more recently, literature argues for a host of reasons, such as to enhance the listener's understanding and maximise communication in a multilingual environment (Slabbert & Finlayson, 2000), as well as to avoid words which would be considered taboo in one of the languages (Ramsay-Brijball, 1999). Ndimande-Hlongwa and Ndebele (2014) assert that these reasons cannot be considered in isolation, but the speaker who is code switching is influenced by all simultaneously to differing degrees. Authors now conclude that code switching is a linguistic tool the use of which shows an indication of mastery and its use managing a delicate socio-cultural scenario (Ndimande-Hlongwa & Ndebele, 2014; Slabbert & Finlayson, 2000). The difference between the research sites here is interesting in that Site 2, which showed the children using the Zulu root *-ne* ('four') is the more rural community with fewer links with commerce, the three formal employers being the primary school, a primary healthcare clinic and a local luxury beach lodge, while the other sites (where children used the English word 'four') have easier access to moderately developed towns with nationwide chain

81

stores such as SPAR and various banking institutions. Conversely, in the biographical questionnaire, only participants from Site 2 reported using English as well as Zulu in the home (Table 3.2). Slabbert and Finlayson (2000) pose urbanisation and socio-political factors as the premise for code switching, while Ramsay-Brijball (1999) points to TV and radio media channels as well as the economic climate as other driving factors. This is interesting to note, but again the sample is too small and the data insufficient to draw any conclusions.

Limitations of sample size and a minimum commonality score of 2 as a cut-off criterion for inclusion into core meant that expressions peculiar to a particular site were also included in the core, which may not necessarily be representative of the vocabulary used by children in general, but rather a reflection of the activities and context of that site on that day. For example, at Site 1 the children had read a story about a crocodile which ate a woman, and the noun root crocodile *'nrINGWENYA'* appears in the core vocabulary. At Site 2 the children spent one lesson counting, and although all these instances of counting were treated as a single unit, the string *'kunye-kubili-kuthathu-kune'* ('one_two_three_four') appeared in the core list with a commonality score of 2. Trembath et al (2007) also noted this, finding a high incidence of the use of the words *aeroplane* and *swing* among the children whose preschool was near an airport and had a swing in the playground. A larger sample size (higher number of participants) may reduce the effects of commonality in future research. In other studies with much larger participant sample sizes, commonality criteria were more stringent. For example, in the study conducted by Robillard et al. (2014), at least 50% of the 57 participants had to use a word for it to be included in the core vocabulary. However, there are no guidelines to indicate what an acceptable commonality criterion would be in establishing a core vocabulary. With regard to the current study, although the criterion of a commonality of 2 may have been lenient, there was nonetheless a general tendency for formatives with higher frequency scores to also have higher commonality scores (see Appendix K). This confirms the assumption held in the AAC field that frequently used words are important (Morrow et al., 1993), with a positive linear relationship between frequency and commonality (Beukelman et al., 1989). It furthermore demonstrates the legitimacy of the use of a commonality score (as used in studies of English; e.g. Fallon et al., 2001; Trembath et al., 2007) in this study.

Chapter 5: Discussion

Acknowledging that the small sample size affects the way in which the commonality score might be interpreted, the frequency measure may also be discussed as another arbitrary cut off. To qualify as core vocabulary, a formative had to occur with a minimum frequency of 0.5‰. Thirteen formatives appeared with frequency of 0.49‰ accounting for 130 total formatives used in the sample, each of these formatives was used 10 times in the composite sample only marginally less than those included in the core vocabulary. These 13 formatives were roots of different kinds; six verbs, two adjectives, three nouns, one relative, and one quantifier root. The group just above this cut off and included in core vocabulary with a frequency of 0.54‰ were 11 different formatives accounting for 110 of the total formatives in the sample. Each of these formatives occurred 11 times in the composite sample. This group contained the other quantifier root among verb, adjective and adverb roots, a grammar formative, and an interjection. Without further evidence to the justification for its use (although it is used in other word frequency studies in the AAC field, such as Robillard et al., 2014; Trembath et al., 2007), it is not possible to say with certainty that the cut off of 0.5‰ is not an arbitrary one. A different frequency measure may be a way to overcome the arbitrary frequency cut off of 0.5‰ in future studies. One study determined vocabulary by coverage when it calculated the relative coverage that any group of words would give over the total vocabulary sample (Liu & Sloane, 2006). This is somewhat similar to the reports in other research of the coverage of core vocabulary determined (Beukelman et al., 1989; Boenisch & Soto, 2015) within the study such as this one, which also reports 221 formative, accounting for 88.9% of the sample. However, this measure was used by Lui and Sloane (2006) to determine the core rather than to describe it in that they set out with a predetermined coverage amount to be achieved. The most frequently used vocabulary items that covered 90% of conversations were included in the vocabulary list. The authors do not state why they chose this coverage amount, and the vocabulary is reported only in Mandarin Chinese. Other research has used a grouped frequency measure whereby the number of unique words within groups of frequency are plotted on a graph to demonstrate visually where a core/fringe cut off might lie (Shin & Hill, 2016). Interestingly, in this study in Korean, through using this way to analyse the data distribution between high and low frequency words, the researchers selected a frequency cut off which was lower than the typical 0.5 ‰, which in turn resulted in the determination of 627 high frequency words, a number much higher than those reported for English. With commonality scores then applied to the 627 high frequency words, the number of

words dropped to 219 and the coverage offered by this 'core' vocabulary was 60.82% (Shin & Hill, 2016). Despite the lower frequency cut off, the Korean core vocabulary list offers much lower coverage than lists determined in other studies. Although similarities do certainly exist, this suggests that Robillard et al.'s (2014) assertion that regardless of the language, the number of core words and percentage coverage given by these words are comparable requires some further qualification. Given that Robillard et al. (2014) were considering English and French, which are both Indo-European languages (Cowgill & Jasanoff, 2017), it may be that other conclusions may be drawn from comparisons of languages which share fewer linguistic commonalities.

## 5.3 The content of the Zulu core vocabulary determined

### *5.3.1 Types of formatives found in core vocabulary – Structure and content formatives*

This manner of frequency counts at the formative level reaps very different results to previous word frequency counts done on languages within this language family. Where Allwood et al. (2010) reported that conjunctions and interjections constituted the words with the highest frequencies, now, the larger proportion by far are the concordial formatives and roots and stems, comprising 26.2% and 21.1% respectively. Comparatively, although still appearing in core vocabulary, interjections contributed only 2.9% and conjunctions less than 1%. As demonstrated in Chapter 2 (Table 2.4 and 2.5), in previous word frequency studies, orthographic space was used to define the boundaries of the unit of analysis. When remaining undivided, orthographic words in Zulu are highly variable, and the many common features among words are lost. On each appearance in the corpus these words were therefore counted as entirely different words.

An examination of the figures and tables presented in Section 4.5 reveals patterns of frequency of the various formative types within core vocabulary that reflect the structure and nature of the Zulu language. Domination of concords (as shown in frequency of use but with a very restricted number of formatives) demonstrated the previously mentioned domination of the noun and noun class structure in this language and all languages in this language family (Doke, 1939; Paulos & Msimang, 1998; Taljaard & Bosch, 1993). In contrast to Robillard et al.'s French study, which found only one noun and no object names in any of the groups of data collected (Robillard et al., 2014), this study determined a total of 30 noun roots in the core vocabulary.

84

This finding, however, correlates with the Australian study, which also determined 30 nouns (Trembath et al., 2007).

Also reflective of the nature of Zulu, is the verb and associated formatives in the core vocabulary which comprise both a large number of formatives and which were used with a high frequency. This is in keeping with Doke's assertion that no utterance in Zulu is complete without it either containing or being a predicate (Doke, 1939). Other studies also report high incidences of verbs in core vocabulary. In English studies, Boenisch and Soto (2015) report verbs comprising 26 and 28% for their two groups of participants, while Fallon et al. (2001) report verbs comprising 29%. Korean word frequency studies also report verbs featuring prominently at 21.92% of core vocabulary (Shin & Hill, 2016). In all of these studies which reported proportions of word types, the verb category took the highest proportion overall. It is interesting then to note that in the Zulu core vocabulary, the concords rank higher. This correlates with the domination of the noun class system in Zulu as already demonstrated.

Similar to studies in other languages, the major proportion of the core vocabulary comprised structure formatives. Conversely, the content formatives contained a great number of verbs, adverbs, and a few nouns. Interestingly, all forms of Zulu structure appeared in the data collected from participants, the majority of which were included in the core vocabulary of the participants. This correlates with the participant selection criteria assertion that children of this age have gained the mature structure of language (Owens, 2005; Suzman, 1990).

Sixty-six different structure formatives which are directly related to language structure comprise the bulk of the core vocabulary (61%), while 151 different content formatives make up the smaller proportion of the core at 39%. This is similar to patterns found in other languages where the larger proportion of core vocabulary is comprised of structure vocabulary, and the smaller proportion of core vocabulary consisting of content words, while the number of unique structure words in the core vocabulary is typically low and the number of unique content words is typically high (Boenisch & Soto, 2015; Fallon et al., 2001; Robillard et al., 2014; Trembath et al., 2007).

Most of the structure formatives are closed class items (i.e. consisting of a finite number of items/formatives), such and the noun prefixes and concords, which are directly controlled by the

85

noun class system and parts of speech rules (Nyembezi, 1982; Taljaard & Bosch, 1993). Should all structure formatives be accounted for, this set of words would remain stable regardless of sample size increasing. Boenisch and Soto (2015) observed this phenomenon when considering their top 20, 50, 100 and 200 words. They noted that the proportion of function (or structure) vocabulary decreased as more words were added to the list (Boenisch & Soto, 2015). In the Zulu results, some structure formatives (nine unique formatives) fell outside of core vocabulary in the data set, due to lower frequency counts, which may or may not indicate a limitation of sample size and/or that the frequency and commonality criteria used to identify a core vocabulary were too strict. On closer inspection, none of the excluded formatives represented an entire part of speech but rather pertained to a noun class variation (for instance, the noun prefixes 4 and 14 were excluded but all others were not), which relates to fewer nouns belonging to that class being spoken about by participants. This raises confidence in the inclusion criterion but points to the limited sample size.

In contrast, content formatives come generally (but not entirely) from a group of open-class concepts, such as verbs and nouns (Boenisch & Soto, 2015). In Zulu, while there are a great number of verbs and nouns which exist in an open class (Nyembezi, 1982) there is a certain proportion of Zulu content formatives that is limited in number. According to Zulu grammarians, the enumerative, adjective, and adverbial formatives of manner and location are among a closed class of lexical items which can be listed in their entirety (Doke, 1939; Nyembezi, 1982; Taljaard & Bosch, 1993). It is interesting to note then that although being similar to structure formatives in that they originate from a closed class, this does not appear to have bearing on their frequency of use. Within the content category, these closed class lexical items account for lower proportions of the vocabulary; enumerative accounts for just 1%, adjectives for 3% and adverbs for 7% (Figure 4.7).

### 5.3.2 Lexical meaning of core content formatives as compared to other studies

The fundamental differences between Zulu and English language structure do complicate comparisons between the languages. While not all 151 content formatives of Zulu core vocabulary could be compared for the sake of space, the overlap in the top 30 is remarkable, with only six of the top 30 formatives identified in this study not appearing in all of the comparison lists of English core vocabulary. Differences in the manner of expression may account for the

86

discrepancies found. For instance, the English word 'some', which ranks quite highly in compared English lists, is not directly translatable into Zulu but may appear in the verbal applied extension (tag: xa) *ngihlephulele*, meaning 'break it for me'; where an English-speaking child might say, "I want some". This formative tagged 'xa' ranks 48th in the Zulu core vocabulary (see Appendix K). This demonstrates the different linguistic way of expressing the same function. Among the six formatives not reflecting in English, the demonstrative relative *nje* (tagged as drNJE) and presentative formative (tagged as PRS) do not directly translate into English, and as such do not appear in any of the lists compared.

## 5.4 Implications of a morphologically rich language for AAC vocabulary selection and AAC system design

When considering the use of formatives, particularly structure formatives within a graphic symbol-based AAC system, it is necessary to again take into account the role of the graphic symbols in relation to the natural language they attempt to represent. This discussion can be framed effectively by contrasting Type 1 and Type 2 symbols (McNaughton, 1993).

Many graphic symbol libraries used in AAC have a 1:1 relationship between word and symbol. Regardless of the language they represent, symbols from symbol sets such as PCS in and of themselves operate in a system whereby there is a direct relationship between each symbol and its referent (typically a word) without any segmentability, and are thus Type 1 (McNaughton, 1993). These symbols attempt to harness iconicity; that is, the perceived visual relationship between symbol and referent, to enhance learning. Authors have hypothesised that the more apparent the relationship between the symbol and its referent, the more readily the symbols can be used and remembered (Bornman, Alant, & Du Preez, 2009; Dada, Huguet, & Bornman, 2013; Fuller & Lloyd, 1997). However, others have argued that this is only a short-term benefit that is based on a perception, which is influenced by background as much as it is by preference (Stephenson, 2009). Nonetheless, Type 1 symbols are effective (Beukelman & Mirenda, 2013) and are widely used in South Africa (Bornman, Bryen, Kershaw, & Ledwaba, 2011; Dada, Murphy, & Tönsing, 2017).

Many of the graphic symbol sets used in South Africa were designed by speakers of English. PCS symbols, for example, were designed by Roxanna Johnson (Fuller, Lloyd, &

Stratton, 1997), while Widgit symbols are based on the original Rebus symbols designed by Woodcock and colleagues (Fuller et al., 1997). English is largely an analytical language and uses word order to convey meaning. When attempting to represent an analytical language through Type 1 symbols, it is possible to convey meaning through a crude form of 'syntax', by sequencing the Type 1 symbols in a word order of the language that is to be represented. However, a synthetic agglutinative language (like Zulu) does not convey meaning primarily by word order but rather, through morphological changes.

A number of difficulties might be encountered when attempting to represent a synthetic language (that is, a language conveying meaning through morphological changes rather than word order) using a Type 1 set approach such as PCS. An argument has already been made as to why meaning should be represented at the level of formatives rather than at the level of orthographic or linguistic words for a Zulu AAC system when a level of generativity is required. If an AAC system were designed based on formatives, a person using the system would need to employ the rules which govern the natural language in terms of the construction of meaning through the use of these formatives. For instance, selection of the formatives as follows; relative + possessive-concord-5 + pronoun-2 would produce *elabo* ('theirs') while relative + possessive-concord-5+ pronoun-1-singular would produce *elami* ('mine') but pronoun-2 and pronoun-1-singular alone would give *bona* ('they/them/themselves'), and *mina* ('I/me/myself') respectively. Since the formatives cannot be used on their own, the graphic symbols that represent individual formatives would be classified as Type 2 symbols (McNaughton, 1993), since a set of grammatical rules would be needed to combine these symbols to convey meaning. This may be similar to some of the Type 2 symbols employed in Bliss, where, for example, the addition of a circumflex accent (i.e. ^ ) is used to indicate an action, as described in Chapter 2 (Blissymbolics Communication International, 2014).

Furthermore, although it may be possible to depict the content formatives of the Zulu core vocabulary in a way that visually or conceptually resembles the referent, this is hardly possible for structure formatives, which carry little or no meaning on their own. They also do not always present with the same orthographic or phonetic realisation but many have phonologically, grammatically or lexically conditioned allomorphs which are determined by the context in which they are being used (Kosch, 2006). Although one may argue that English structure words are also

not 'picture producers', words such as pronouns and prepositions may be depicted using translucent symbols. Although such symbols may not be guessable on first exposure, there is a logical relationship between the symbol and the referent, which makes memorising the symbol easier once learnt (Dada et al., 2013). For example, in PCS, prepositions are depicted mostly using a line and a dot. Once it is clear that the dot indicates the position being described (in relation to the line), it becomes easy to remember that a dot under a line means 'under', while a black dot over a line means 'above.' To find transparent symbols for the representation of structure formatives such as 'relative' in Zulu may be much more difficult, and it is likely that such symbols would be arbitrary, with neither a logical link between symbol and referent nor an orthographic word with which it could be labelled. This would have an influence on learnability (Fuller & Lloyd, 1997; Luftig & Bersani, 1985). Thus, in its use, the complexity of the graphic system increases, since a large proportion of (66 of the 221 core) formatives included in the core vocabulary of the system would now be represented by more arbitrary graphic symbols that need to be combined following a set of rules in order to create meaning.

Although this may seem to place a huge learning demand on the person using the system, it should be remembered that, in addition to the effect of the iconicity of a symbol, the individual, the referent itself and the instruction received are all variables which impact on an individual's ability to gain and maintain a relationship between symbol and referent (Bornman, Alant, & Du Preez, 2009; Dada, Huguet, & Bornman, 2013; Fuller & Lloyd, 1997). In a longitudinal study of children's development of Zulu grammar, Suzman  (1990) surmised that the children acquired this very complicated system of morphology, tone and syntax in spoken Zulu through context and adult models who demonstrated very little simplification of the language when conversing with children. It is therefore possible for children to acquire a complex linguistic system, given enough exposure and models. Authors have also sought to address the question of teaching the use of grammatical morphemes (e.g. the plural '-s' and the present continuous verb tense ending '-ing') to children who use AAC (Binger, 2008; Binger, Maguire-Marshall, & Kent-Walsh, 2011). These morphemes, like formatives, carry little meaning on their own and cannot easily be represented by graphic symbols that visually or conceptually resemble their referent. In both instances, the authors reported the participants readily learnt the concept but found a particular kind of intervention (the use of contrasts), to be key to the participants maintaining the skill. The importance of fluent adult models gives further credence to the importance of full support of

89

language within AAC interventions, together with the use of systems more closely aligned with natural language to enable the child who uses AAC to learn to access and use the system.

## 5.5 Conclusion

The results of the study were discussed in this chapter. The parameters (total number of formatives, number of unique formatives) of the total Zulu vocabulary identified from spoken language of Zulu-speaking preschoolers were explored and compared to other studies. Similarly, the parameters of the core vocabulary (frequency and commonality scores) identified in this study were compared to that of previous studies. Furthermore, the lexical meaning of the content formatives of core vocabulary determined in this study was discussed and compared with English core vocabulary lists. Finally, the implications that a morphologically rich language like Zulu has for the selection of vocabulary and design for augmentative and alternative communication (AAC) systems were discussed.

# CHAPTER 6

# CONCLUSION

## 6.1 Introduction

In this chapter, a summary of the study is provided. The study is critically evaluated and implications for clinical practice are discussed. Recommendations for future research are provided.

## 6.2 Summary of the study

Non-literate persons who use AAC typically rely on pre-selected vocabulary in their communication systems. One of the methods for selecting vocabulary, particularly for systems which attempt to offer greater access to a generative system of language, is the core vocabulary approach.

The main aim of the study was to identify the core vocabulary of Zulu-speaking preschool children to the purpose of informing AAC communication system design for people who use Zulu.

Six children from three different preschools participated in this study. Each of the children wore voice recorders while continuing to participate in their typical preschool routines. Each child contributed 1 500 words, creating a composite list of 9 000 words.

Due to the linguistic structure of the Zulu language combined with the need to reap a vocabulary which would lend itself to a generative AAC system, the words collected were then segmented into formatives. The subsequent composite list collected comprised 20 137 formatives.

A core vocabulary of 221 different formatives was identified from this list through the use of frequency rankings and commonality scores. These 221 formatives accounted for 88.9% of the composite sample. This proportion is similar to that determined in English and other

languages (e.g. Beukelman, Yorkston, Poblete, & Naranjo, 1984; Boenisch & Soto, 2015; Liu & Sloane, 2006; Trembath, Balandin, & Togher, 2007).

Similarly to studies in other languages, the major proportion of the core vocabulary comprised of structure vocabulary (as counted by frequency of occurrence of structure versus content formatives). The content vocabulary, conversely, contained a great number of verbs, adverbs and a few nouns. Notwithstanding the fundamental differences between languages which prohibit comparisons at a word level, a remarkable overlap of content was nonetheless observed when comparing the top 30 Zulu content formatives to previous studies in English.

The study showed that, through a method of counting formative frequency, it was possible to determine a Zulu core vocabulary, which can be used as a starting point in designing a more generative Zulu AAC system than would be possible when representing only orthographic Zulu words on an AAC system. The potential challenges with such a system raise questions for further research. These concern, for instance, the design of appropriate graphic symbols to represent structure formatives and also the learnability demands of a system with more abstract (and, in themselves minimally meaningful) units represented.

**6.3 Critical evaluation of the study**

This study was the first to attempt identifying a Zulu core vocabulary for the purpose of informing AAC system design. To this end, a method of tagging formatives was custom developed. The tagging system may be useful for determining core vocabulary in other Nguni languages (e.g. Xhosa, Swati or Ndebele), while it will also be useful for other South African and African languages (e.g. the Sotho-Tswana languages) of similar linguistic grouping. The study further validates the use of frequency counts of smaller units (such as morphemes/formatives) than linguistic words for agglutinative languages.

The observational design allowed for the recording of speech produced by preschoolers during naturally occurring activities. Allowing for a 20-minute warm-up period after children were fitted with the recording equipment reduced participant reactivity to the equipment. Recording proceeded during typical preschool activities continuously throughout the day, minimising the risk that vocabulary collected would only reflect that appropriate to one type of activity, for example, circle time or break time only. This increased the external validity of the

92

data and made it more likely that the core vocabulary identified would indeed be useful for children with good Zulu language comprehension who would need and have the capacity to learn a linguistic AAC system.

High inter-rater reliability of both the transcriptions and the formative tagging (data analysis) strengthened the confidence in the internal validity of the results.

Although three different sites were used for the study, the small sample size of only six participants does give rise to limitations regarding the extent to which the core vocabulary can be regarded as completely representative. It should also be kept in mind that the limited variation of the settings (all were rural preschool settings) where data was collected, the limited time span across which data was collected (one to three days, only during the preschool day) and the limited age range of the participants (five to six years) further reduces generalisability of the results.

Despite the 20-minute warm-up period (during which the children wore the recording equipment and continued with their activities but the data of which was not analysed), participant reactivity remained an unavoidable factor within observational design affecting the external validity.

Since no real guidelines exist regarding the establishment of in- and exclusion criteria regarding the frequency and commonality scores used to identify core vocabulary, the commonality score of $\geq 2$ and the frequency count of $\geq 0.5‰$ as criteria for the inclusion of formatives into core vocabulary have to be regarded as somewhat arbitrary. Alternative measures such as grouped frequency counts may represent more defensible criteria for the determination of a core vocabulary.

## 6.4 Clinical implications

One of the main outcomes of the study was the determination of a core vocabulary of 221 formatives for Zulu. Within a clinical setting these formatives could be used to develop a generative communication tool for preschoolers who use graphic AAC systems.

Chapter 6: Conclusion

The system would necessarily function by means of representing the various core formatives graphically with the possibility to add fringe particular to the client. Many decisions remain to be made as to how this might be achieved, including in what way the formatives might be symbolised, and design decisions such as layout, colour coding and feedback systems. Effective teaching methods would also need to be determined. All these factors would impact on the potential effectiveness and ease of learning of the AAC system, which would be of fundamental interest to the clinician and researcher alike.

The development of the formative tagging method may be used as a basis for core vocabulary determination in other Nguni and linguistically related languages. Such core vocabularies can then lead to the design of not only monolingual AAC systems in the commonly used Southern African languages but can also be used to design AAC systems that give access to more than one language. Once again, there are many further design and symbolisation decisions to be made, and appropriate teaching methods will need to be determined. Nevertheless, the tagging system developed has the potential to play a critical role in the design of AAC systems that give access to various and multiple Southern African languages.

## 6.5 Recommendations for further research

The study could be replicated with an increased sample size to enhance the extent to which the results might be generalised.

The study could be replicated among different populations to examine possible influences of age, gender and geographical areas (for instance rural, urban and peri-urban populations) on Zulu core vocabulary.

An extension of the study into other populations and increased sample sizes could also allow the analysis of the impact of multilingualism and code switching on Zulu core vocabulary.

Participatory design research could be employed to devise a graphic symbol system appropriate for representing a formative-based Zulu core vocabulary. Stakeholders, including persons in need of AAC, families and service providers could give input on appropriate symbols for this goal. Further research is also needed in determining possible appropriate layout formats for this vocabulary, which would be matched to the cognitive and physical skills of persons in

94

need of the system. The effect of various teaching methods to promote the acquisition and use of such a system would also need to be explored by further studies.

The formative tagging method developed may also have application in studies of child language development and childhood language disorders in Zulu and linguistically related languages. The lack of appropriate language assessment materials in many of the official South African languages has been repeatedly noted (Kathard et al., 2011; Pascoe & Norman, 2011). Natural language sampling has been used as a method to assess expressive language (Binger, Ragsdale, & Bustos, 2016; Kovacs & Hill, in press). The use of the tagging method developed may be explored as a method to assist in conducting such analyses.

## 6.6. Conclusion

In this chapter, a summary of the study was provided. The study was critically evaluated and implications for clinical practice were discussed. Recommendations for future research were provided.

# REFERENCES

Allwood, J., Hammarström, H., Hendrikse, A., Ngcobo, M. N., Pretorius, L., & Merwe, M. Van Der. (2010). Work on Spoken (Multimodal) Language Corpora in South Africa. In *Proceedings of the Seventh conference on International Language resources and Evaluation (LREC'10), Valetta, May 19-21*. European Language Resources Association (ELRA).

Allwood, J., & Hendrikse, A. P. (2003). Spoken language corpora for the nine official African languages of South Africa. *Southern African Linguistics and Applied Language Studies*, *21*(4), 189–201. doi:10.2989/16073610309486343

American Speech-Language-Hearing Association. (2005). Roles and Responsibilities of Speech-Language Pathologists With Respect to Augmentative and Alternative Communication: Position Statement. doi:10.1044/policy.PS2005-00113/

Ayres, K., & Ledford, J. R. (2014). Dependent measures and measurement systems. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology* (pp. 124–153). New York, NY: Routledge.

Baker, B., & Chang, S.-K. (2006). A Mandarin language system in augmentative and alternative communication (AAC). *International Journal of Computer Processing of Languages*, *19*(4), 225–237. doi:10.1142/S0219427906001438

Baker, B., Hill, K., & Devylder, R. (2000). Core vocabulary is the same across environments. In *California State University at Northridge Conference* (pp. 3–8).

Balandin, S., & Iacono, T. (1998). A few well chosen words. *AAC Augmentative and Alternative Communication*, *14*(September), 147–161. doi: 0743-4618/98/1403-0147-0147

Banajee, M., Dicarlo, C., & Stricklin, S. B. (2003). Core Vocabulary Determination for Toddlers. *Augmentative and Alternative Communication*, *19*(2), 67–73. doi:10.1080/0743461031000112034

96

References

Berger, K. (1967). The most common words used in conversations. *Journal of Communication Disorders*, *1*(3), 201–214.

Besio, S., & Chinato, M. G. (1996). A semiotic analysis of the possibilities and limits of Blissymbols. In *Augmentative and Alternative Communication: European Perspectives* (pp. 182–192). London: Whurr.

Beukelman, D. R., Jones, R., & Rowan, M. (1989). Frequency of word usage by non-disabled peers in integrated preschool classrooms. *Augmentative and Alternative Communication*, (5), 243–248.

Beukelman, D. R., McGinnis, J., & Morrow, D. (1991). Vocabulary selection in Augmentative and Alternative Communication. *Augmentative and Alternative Communication*, *7*(3), 171–185. doi: 10.1080/07434619112331275883

Beukelman, D. R., & Mirenda, P. (2005). *Augmentative and Alternative Communication; Supporting children and adults with complex communication needs* (3rd ed.). Baltimore, MD: Paul H Brookes.

Beukelman, D. R., & Mirenda, P. (2013). *Augmentative and Alternative Communication: Supporting children and adults with complex communication needs* (4th ed.). Baltimore, MD: Paul H. Brookes.

Beukelman, D. R., Yorkston, K. M., Poblete, M., & Naranjo, C. (1984). Frequency of word occurence in communication samples produced by adult communication aid users. *Journal of Speech and Hearing Disorders*, *49*, 360–367.

Binger, C. (2008). Grammatical morpheme intervention issues for students who use AAC. *Perspectives on Augmentative and Alternative Communication*, *17*(2), 62–68. doi:10.1044/aac17.2.62

Binger, C., Maguire-Marshall, M., & Kent-Walsh, J. (2011). Using aided AAC models, recasts, and constrative targets to teach grammatical morphemes to children who use AAC. *Journal of Speech Language and Hearing Research*, *54*(1), 160–176. doi:10.1044/1092-4388(2010/09-0163)

References

Binger, C., Ragsdale, J., & Bustos, A. (2016). Language sampling for preschoolers with severe speech impairments. *American Journal of Speech-Language Pathology*, *25*, 493–507. doi:10.1044/2016

Blissymbolics Communication International. (2014). Complete Authorised Lexicon. Retrieved March 10, 2017, from http://www.symbolnet.org/bliss/BCI-AV_2014-12-05_%28en+deriv+es+sv%29_symbols-draft-3.pdf

Boenisch, J., & Sachse, S. (2007). Sprachforderung von anfang an zum einsatz von kern- und fandvokabular in derfruhen Forderung [Language development from the beginning for the use of core and fringe vocabulary In: Augmentative and Alternative Communication]. *Unterstutzte Kommunikation*, *3*, 23–27.

Boenisch, J., & Soto, G. (2015). The Oral Core Vocabulary of Typically Developing English-Speaking School-Aged Children : Implications for AAC Practice. *Augmentative and Alternative Communication*, *31*(1), 77–84. doi:10.3109/07434618.2014.1001521

Bornman, J. (2015). AAC Terminology. In E. Moolman & M. Herold (Eds.), *AAC Resource Manual* (2nd ed., p. 13.1-13.16). Pretoria: Centre for Augmentative and Alternative Communication.

Bornman, J., Alant, E., & Du Preez, A. (2009). Translucency and learnability of Blissymbols in Setswana-speaking children: an exploration. *Augmentative and Alternative Communication*, *25*(December), 287–298. doi:10.3109/07434610903392456

Bornman, J., Bryen, D. N., Kershaw, P., & Ledwaba, G. (2011). Reducing the risk of being a victim of crime in South Africa: You can tell and be heard! *Augmentative and Alternative Communication*, *27*(2), 117–130. doi:10.3109/07434618.2011.566696

Bornman, J., & Tönsing, K. M. (2011). Augmentative and alternative communication. In E. Landsberg, D. Kruger, & E. Swart (Eds.), *Addressing barriers to learning: a South African perspective* (2nd ed, pp. 187–210). Pretoria: Van Schaik.

Bristowe, A., Oostendorp, M., & Anthonissen, C. (2014). Language and youth identity in a multilingual setting : A multimodal repertoire approach. *Southern African Linguistics and*

98

*Applied Language Studies*, *32*(2), 229–245. doi:10.2989/16073614.2014.992644

Burgess, S. (2002). *SA Tribes: who we are, how we live, what we want from life in teh new South Africa*. Durban: New Africa Books.

Census Stats SA. (2011). Census in Brief. Pretoria: Statistics SA 2012.

Central University Research Ethics Committee [CUREC]. (2015). Research on teachers and teaching in educational settings for typically-developing students. Oxford: University of Oxford. Retrieved from http://www.admin.ox.ac.uk/curec/resources/informed-consent/%0ACUREC

Chen, M., Hill, K., & Yao, T. (2009). Preliminary Vocabulary Frequency Findings for Mandarin Chinese AAC Treatments. In *Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology* (p. 30). ACM.

Chick, K., & Mckay, S. (2001). Teaching English in multiethnic schools in the Durban area : The promotion of multilingualism or monolingualism? *Southern African Linguistics and Applied Language Studies*, *19*, 163–178. doi:10.2989/16073610109486285

Cowgill, W., & Jasanoff, J. . (2017). Indo-European Languages. Retrieved May 26, 2017, from https://www.britannica.com/topic/Indo-European-languages

Crestani, C. M., Clendon, S. A., & Hemsley, B. (2010). Words needed for sharing a story : Implications for vocabulary selection in augmentative and alternative communication. *Journal of Intellectual and Developmental Disability*, *35*(4), 268–278. doi:10.3109/13668250.2010.513966

Cucchiarini, C. (1996). Assessing transcription agreement: Methodological aspects. *Clinical Linguistics & Phonetics*, *10*(2), 131–155. doi:10.3109/02699209608985167

Dada, S., Huguet, A., & Bornman, J. (2013). The iconicity of picture communication symbols for children with English additional language and mild intellectual disability. *Augmentative and Alternative Communication*, *29*(4), 360–73. doi:10.3109/07434618.2013.849753

References

Dada, S., Murphy, Y., & Tönsing, K. M. (2017). *Augmentative and alternative communication implementation: A descriptive study of the perceptions of South African speech-language therapists*. submitted for publication.

de Sousa, D. S., Greenop, K., & Fry, J. (2010). The effects of phonological awareness of Zulu-speaking children learning to spell in English: A study of cross-language transfer. *British Journal of Educational Psychology*, *80*(4), 517–533. doi:10.1348/000709910x496429

Dent, G. R., & Nyembezi, C. L. S. (1995). *Scholar's Zulu Dictionary* (3rd ed.). Pietermaritzburg: Shuter & Shooter.

Department of Education. Language in Education Policy (1997). Pretoria: Government Press.

Department of Education. (2007). Guidelines to ensure quality education and support in special schools and special school resource centres. Pretoria, South Africa: Department of Education.

Department of Education. (2010). Guidelines for inclusive teaching and learning. Pretoria, South Africa: Department of Education.

Doke, C. M. (1939). *Textbook of Zulu grammar* (3rd ed.). London: Longmans Green and Co.

Fallon, K. a., Light, J. C., & Paige, T. K. (2001). Enhancing Vocabulary Selection for Preschoolers Who Require Augmentative and Alternative Communication (AAC). *American Journal of Speech-Language Pathology*, *10*(1), 81–94. doi:10.1044/1058-0360(2001/010)

Fried-Oken, M., & More, L. (1992). An initial vocabulary for nonspeaking preschool children based on developmental and environmental language sources. *Augmentative and Alternative Communication*, *8*(1), 41–56.

Fuller, D. R., & Lloyd, L. L. (1997). Symbol selection. In L. L. Lloyd, D. R. Fuller, & H. . Arvidson (Eds.), *Augmentative and alternative communication: a handbook of principles and practices* (pp. 214–225). Boston: Allyn & Bacon.

References

Fuller, D. R., Lloyd, L. L., & Stratton, M. M. (1997). Aided AAC symbols. In L. L. Lloyd, D. R. Fuller, & H. H. Arvidson (Eds.), *Augmentative and alternative communication: a handbook of principles and practices* (pp. 48–79). Needham Heights, MA: Allyn & Bacon.

Gowlett, D. (2006). Zone S. In D. Nurse & G. Philippson (Eds.), *The Bantu Languages* (illustrate, pp. 609–651). London: Routledge.

Howes, D. (1966). A word count of spoken English. *Journal of Verbal Learning and Verbal Behaviour*, *5*(6), 572–606.

Kathard, H., Pascoe, M., Ramma, L., Jordaan, H., Moonsamy, S., Wium, A.-M., … Khan, N. (2011). How can speech-language therapists and audiologists enhance language and literacy outcomes in South Africa? (And why we urgently need to ). *South African Journal of Communication Disorders*, *58*, 59–71. doi: 10.4102/sajcd.v58i2.27

Kay-Raining Bird, E., Genesee, F., & Verhoeven, L. (2016). Bilingualism in children with developmental disorders: a narrative review. *Journal of Communication Disorders*, *63*(2016), 1–14. doi:10.1016/j.jcomdis.2016.07.003

Kim, Y. T., Park, H. J., & Min, H. K. (2003). School-aged children and adult's core vocabulary for the development of an augmentative and alternative communication tool. *Communication Sciences & Disorders*, *8*(2), 93–110.

Kosch, I. M. (2006). *Topics in morphology in the African lanuage context*. Pretoria: UNISA.

Kovacs, T., & Hill, K. (n.d.). Language samples from children who use speech generating devices: Making sense of small samples and utterance length. *American Journal of Speech-Language Pathology*.

Lane, V. W., & Molyneaux, D. (1992). *The dynamics of communiative development*. Englewood Cliffs, NJ: Prentice-Hall.

Lee, Y. M., Kim, Y. T., & Park, E. E. (2005). A preliminary study for the core and fringe AAC vocabulary used by elementary school students. *Communication Sciences & Disorders*, *10*(1), 134–152.

101

References

Leedy, P. D., & Ormrod, J. E. (2014). *Practical Research: Planning and Design* (10th ed.). Harlow, Essex: Pearson.

Lewis, M., Paul, G. F., Simons, F., & Fennig, C. D. (2016). Ethonologue: Languages of the world, Nineteenth edition. Retrieved June 21, 2016, from http://www.ethnologue.com/language/zul

Light, J. (1989). Toward a definition of communication competence for individuals using Augmentative and Alternative Communication systems. *Augmentative and Alternative Communication*, *5*(2), 137–144. doi: 10.1080/07434618912331275126

Light, J. (1997). " Communication Is the Essence of Human Life ": Reflections on Communicative Competence. *Augmentative and Alternative Communication*, *13*(June), 61–70. doi:10.1080/07434619712331277848

Light, J., & Lindsay, P. (1991). Cognitive science and augmentative and alternative communication. *Augmentative and Alternative Communication*, *7*(3), 186–203. doi: 10.1080/07434619112331275893

Light, J., & McNaughton, D. (2014). Communicative competence for individuals who require Augmentative and Alternative Communication: A new definition for a new era of communication? *Augmentative and Alternative Communication*, *30*(1), 1–18. doi:10.3109/07434618.2014.885080

Liu, C., & Sloane, Z. (2006). Developing a Core Vocabulary for a Mandarin Chinese AAC System Using Word Frequency Data. *International Journal of Computer Processing of Oriental Languages*, *19*(4), 285–300. doi: 10.1142/S0219427906001530

Lloyd, L. L., Fuller, D. R., & Arvidson, H. H. (1997). Glossary. In L. L. Lloyd, D. R. Fuller, & H. H. Arvidson (Eds.), *Augmentative and alternative communication: a handbook of principles and practices* (pp. 522–543). Needham Heights, MA: Allyn & Bacon.

Luftig, R., & Bersani, H. (1985). An investigation of two variables influencing Blissymbol learnability with nonhandicapped adults. *Augmentative and Alternative Communication*, *1*(1), 32–37. doi:10.1080/07434618512331273501

References

Marvin, C. A., Beukelman, D., & Bilyeu, D. (1994). Vocabulary-use patterns in preschool children: Effects of context and time sampling. *Augmentative and Alternative Communication*, *10*(4), 224–236.

McMillan, J., & Schumacher, S. (2010). *Research in Education: Evidence Based Enquiry* (7th ed.). Cape Town, South Africa: Pearson.

McNaughton, S. (1993). Graphic representational systems and literacy learning. *Topics in Language Disorders*, *13*(2), 58–75. doi: 10.1097/00011363-199302000-00007

Meinhof, C. (1932). *Introduction to the phonology of Bantu languages*. Berlin: Dietrich Reimer/Ernst Vohsen.

Miller, J., Andriacchi, K., & Nockerts, A. (2011). *Assessing language production using SALT software*. Middleton, WI: SALT Software LLC.

Miller, J., & Iglesias, A. (2012). Systematic Analysis of Lanuage Transcripts (SALT), Student version 2012 [Computer software]. Middleton, WI: SALT Software. LLC.

Morrow, D., Mirenda, P., Beukelman, D., & Yorkston, K. M. (1993). Vocabulary selection for augmentative communication systems: A comparison of three techniques. *American Journal of Speech-Language Pathology*, *2*(2), 19–30. doi: 10.1044/1058-0360.0202.19

Murray, J., & Goldbart, J. (2009a). Augmentative and Alternative communication: a review of current issues. *Pediatrics and Child Health*, *19*(10), 464–468. doi:10.1016/j.paed.2009.05.003

Murray, J., & Goldbart, J. (2009b). Cognitive and language acquisition in typical and aided language learning: A review of recent evidence from an aided communication perspective. *Child Language Teaching and Therapy*, *25*, 31–58. doi:10.1177/0265659008098660

Ndimande-Hlongwa, N., & Ndebele, H. (2014). Digging deep into IsiZulu-English code-switching in a peri-urban context Digging deep into IsiZulu-English code-switching in a peri- urban context. *Language Matters*, *45*(2), 237–256. doi:10.1080/10228195.2014.910248

103

References

Ngcobo, S. (2013). Educators' attitudes towards the role of isiZulu in education : Additive rather than exclusive. *Southern African Linguistics and Applied Language Studies*, *31*(2), 185–205.doi:10.2989/16073614.2013.815838

Nurse, D., & Philippson, G. (2006). Introduction. In D. Nurse & G. Philippson (Eds.), *The Bantu Languages* (illustrate, pp. 1–12). London: Routledge.

Nyembezi, C. L. S. (1982). *Uhlelo lwesiZulu* (4th ed.). Pietermaritzburg: Shuter & Shooter.

Owens, R. E. (2005). *Language development: an introduction* (6th ed.). Boston: Pearson.

Pascoe, M., & Norman, V. (2011). Contextually relevant resources in speech-language therapy and audiology in South Africa - are there any? *The South African Journal of Communication Disorders*, *58*, 2–5. doi:10.7196/SAJCD.256

Paulos, G., & Msimang, C. T. (1998). *A linguistic analysis of Zulu* (1st ed.). Cape Town, South Africa: Via Afrika.

Peña, E. D. (2007). Lost in Translation: Methodological Considerations in Cross-Cultural Research. *Child Development*, *78*(4), 1255–1264.

Pretorius, L., & Bosch, S. E. (2003). Computational aids for Zulu natural language processing. *Southern African Linguistics and Applied Language Studies*, *21*(4), 267–282. doi:10.2989/16073610309486348

Prinsloo, D. (2007). The right to mother tongue education : a multidisciplinary , normative perspective. *Southern African Linguistics and Applied Language Studies*, *25*(1), 27–43. doi:10.2989/16073610709486444

Prinsloo, D. J., & De Schryver, G. M. (2001). Corpus applications for the African languages , with special reference to research , teaching , learning and software. *Southern African Linguistics and Applied Language Studies*, *19*(1–2), 111–131. doi:10.2989/16073610109486279

Ramsay-Brijball, M. (1999). Understanding Zulu-English code-switching : a psycho-social

perspective. *South African Journal of Linguistics*, *17*(2), 161–172.
doi:10.1080/10118063.1999.9724532

Robillard, M., Mayer-Crittenden, C., Minor-Corriveau, M., & Belanger, R. (2014). Monolingual
and bilingual children with and without primary language impairment : Core vocabulary
comparison. *Augmentative and Alternative Communication*, *30*(3), 267–278.
doi:10.3109/07434618.2014.921240

Schlosser, R. W., Wendt, O., Bhavnani, S., & Nail-Chiwetalu, B. (2006). Use of
information-seeking strategies for developing systematic reviews and engaging in
evidence-based practice: the application of traditional and comprehensive pearl growing.
*International Journal of Language & Communication Disorders*, *41*(5), 567–582.
doi:10.1080/13682820600742190

Sevcik, R., Romski, M. A., & Wilkinson, K. (1991). Roles of graphic symbols in the language
acquisition process for persons with severe cognitive disabilities. *Augmentative and
Alternative Communication*, *7*(September), 161–170. doi:10.1080/07434619112331275873

Shin, S., & Hill, K. (2016). Korean word frequency and commonality study for augmentative and
alternative communication. *International Journal of Language and Communication
Disorders*, *0*(0), 1–15. doi:10.1111/1460-6984.12218

Sigafoos, J., & York, J. (1991). Using ecological inventories to promote functional
communication. In J. Reichle, J. York, & J. Sigafoos (Eds.), *Implementing augmentative
and alternative communication: Strategies for learners with severe disabilities* (pp. 61–70).
Baltimore: Paul H Brookes.

Slabbert, S., & Finlayson, R. (2000). "l'm a cleva!": the linguistic makeup of identity in a South
African urban environment. *International Journal of the Sociology of Language*, *144*(1),
119–135. doi:10.1515/ijsl.2000.144.119

Smith, M. (2006). Speech , language and aided communication : Connections and questions in a
developmental context. *Disability and Rehabilitation*, *28*(3), 151–157.
doi:10.1080/09638280500077747

References

Soto, G., & Yu, B. (2014). Considerations for the Provision of Services to Bilingual Children Who Use Augmentative and Alternative Communication. *Augmentative & Alternative Communication*, *30*(October 2013), 83–92. doi:10.3109/07434618.2013.878751

South African Department of Arts and Culture, & Centre for Text Technology (CTexT). (2014). NCHLT Text Corpora. Centre for Text Technology (CTexT) North-West University, South Africa. Retrieved from http://www.nwu.ac.za/ctext

Spiegler, S., Van der Spuy, A., & Flach, P. A. (2010a). *Additional material for the Ukwabelana Zulu corpus.* Retrieved from http://www.cs.bris.ac.uk/Research/MachineLearning/Morphology/Resources

Spiegler, S., Van der Spuy, A., & Flach, P. A. (2010b). Ukwabelana - An open-source morphological Zulu corpus. *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, (August), 1020–1028. doi:10.1109/ANZIIS.1995.705747

Statistics South Africa. (2015). South Africa. Retrieved October 1, 2015, from www.statssa.gov.za/?page_id595

Stephenson, J. (2009). Iconicity in the Development of Picture skills : Typical Development and Implications for Individuals with Severe Intellectual Disabilities. *Augmentative & Alternative Communication*, *25*(3), 187–201. doi:10.1080/07434610903031133

Stuart, S., Vanderhoof, D., & Beukelman, D. (1993). Topic and vocabulary use patterns of elderly women. *Augmentative and Alternative Communication*, *9*(2), 95–110. doi:0743-4618/93/0902-0095

Sutton, A., Soto, G., & Blockberger, S. (2002). Grammatical issues in graphic symbol communication. *Augmentative and Alternative Communication*, *18*, 192–204. doi: 10.1080/07434610212331281271

Suzman, S. M. (1990). *Language acquisition in Zulu*. Witswatersrand.

Taljaard, P. C., & Bosch, S. E. (1993). *Handbook of isiZulu* (2nd ed.). Pretoria: J.L. van Schaik

106

References

(Pty) Ltd.

Thistle, J. J., & Wilkinson, K. M. (2013). Working memory demands of aided augmentative and alternative communication for individuals with developmental disabilities. *Augmentative and Alternative Communication*, *29*(3), 235–245. doi:10.3109/07434618.2013.815800

Thistle, J. J., & Wilkinson, K. M. (2015). Building evidence-based practice in AAC display design for young children: current practices and future directions. *Augmentative and Alternative Communication*, *31*(2), 124–136. doi:10.3109/07434618.2015.1035798

Tönsing, K. M., Alant, E., & Lloyd, L. L. (2005). Augmentative and alternative communication. In E. Alant & L. L. Lloyd (Eds.), *Augmentative and alternative communication and severe disabilities: Beyond poverty* (pp. 30–67). London: Whurr.

Trembath, D., Balandin, S., & Togher, L. (2007). Vocabulary selection for Australian children who use augmentative and alternative communication. *Journal of Intellectual and Developmental Disability*, *32*(4), 291–301. doi:10.1080/13668250701689298

Witkowski, D., & Baker, B. (2012). Addressing the Content Vocabulary With Core: Theory and Practice for Nonliterate or Emerging Literate Students. *Perspectives on Augmentative and Alternative Communication*, *21*, 74–81. doi:10.1044/aac21.3.74

Yorkston, K. M., Dowden, P. A., Honsinger, M. J., Marriner, N., & Smith, K. (1988). A comparison of standard and user vocabulary lists. *Augmentative & Alternative Communication*, *4*(4), 189–210. doi:10.1080/07434618812331274807

Yorkston, K. M., Smith, K., & Beukelman, D. (1990). Extended Communication Samples of Augmented Communicators I: A comparison of individualised versus standard single-word vocabularies. *Journal of Policy and Practice in Intellectual Disabilities*, *55*(2), 217–224.

Appendix A

# APPENDIX A

# TABLE OF STUDIES

This table is not intended to give a full critique of the studies, but to provide an overview of some of the research into word use in spoken language and on the theme of core vocabulary as it has applied to AAC. It is organised chronologically from the earliest to the latest studies.

| Author and year | Title | Aim | Design | Language Studied and Country | Participants | Procedures | Findings regarding vocabulary |
|---|---|---|---|---|---|---|---|
| **Howes (1966)** | A word count of spoken English. | To present an updated reliable corpus of spoken English | Quantitative observational | English; Boston, Massachusetts | 20 sophomore students and 21 patients (cerebral defects excluded, patients mainly admitted for bone injuries and peripheral nerve conditions) | In total 250 000 words were collected from recordings of free conversation.  40 of the participants were interviewed once, collecting 5 000 words from each person. 1 participant was interviewed 10 times in a study on the stability of word frequency data for an individual. | Of the 250 000 words collected, a total number of 9 699 different words were found.  Of these, 4 097 were uttered only once, reflecting a smaller list of higher frequency words. The full table of words and their frequencies is published. |
| **Berger (1968)** | The most common words used in conversations. | To catalogue vocabulary in unguarded informal conversation | Quantitative observational | English; North-East Ohio | No formal recruitment of participants, authors reported an eavesdropping technique | In total 25 000 words were collected from 2 418 sentences collected through eavesdropping in a variety of contexts | Of the 25 000 words, 2 507 different words were found, 'nearly half' of which occurred only once. The authors give a comparison of word frequency compared with word syllable length and sentence length. Their findings reflect the tendency in spoken language to use a small list of words with high frequency A full appendix of words and their frequencies is published. |

108

Appendix A

| Author and year | Title | Aim | Design | Language Studied and Country | Participants | Procedures | Findings regarding vocabulary |
|---|---|---|---|---|---|---|---|
| **Beukelman, Yorkston, Poblete, & Naranjo (1984)** | Frequency of word occurrence in communication samples produced by adult communication aid users | To collect, analyse and describe vocabulary and characteristics of "natural" communication samples from language in-tact non-speaking adults | Quantitative observational | English; North America | 5 non-speaking adults aged 16 – 25, who relied on typed output for communication | The communication aids used produced strips of typed messages; these were torn off after each message and collected for 14 days. Communication samples ranged from 3 770 and 10 195 words. | The number of different words for each participant ranged from 957 to 1 313. In order to compare participants, the authors investigated the congruence between lists of most frequently appearing words for each participant. For example the top 50 words showed a mean of 78.8% congruence and the top 500 words a mean of 84.2% congruence. A list of the top 500 words is published. |
| **Beukelman, Jones, & Rowan, (1989).** | Frequency of word usage by non-disabled peers in integrated preschool classrooms. | To collect spoken language samples from non-disabled peers and described the vocabulary in terms of frequency and commonality of words used. To consider how the information might be used in the preschool curriculum and AAC interventions for the peers who need AAC. | Quantitative observational | English, North America | 6 non-disabled preschool children aged between 3y8mo and 4y9mo from 3 different classrooms participated over 2 to 8 days. | A dictaphone was placed near to the participant while researchers also observed and compiled hand written transcriptions of the child's spoken output. This procedure was used with each participant until 3 000 words had been collected from each child. The recordings were then reconciled with the hand written transcriptions. The words collected were then analysed for total words, total unique words, word frequency and word commonality. | The number of different words from each participant ranged between 404 and 468 unique words uttered. A total of 250 words are reported having occurred with a frequency of 5‰. These 250 words represent 85% of the total sample. The full list of these words is published. |

109

Appendix A

| Author and year | Title | Aim | Design | Language Studied and Country | Participants | Procedures | Findings regarding vocabulary |
|---|---|---|---|---|---|---|---|
| **Yorkston, Smith, & Beukelman, (1990)** | Extended communication samples of augmented communicators I: A comparison of individualised versus standard single-word vocabulary lists. | To compare relative benefits of word lists individualised for specific AAC users versus standard vocabulary lists for the purpose of providing acceleration techniques. | Quantitative observational | English; North America | 10 AAC users aged between 13 and 30 participated over 14 days 2 of the 10 also submitted a journal record of activities during the time studied | The participants produced their communication samples by letter typing and the printed output was collected and analysed for single word frequencies. These lists were then compared to 6 vocabulary lists either typically used in AAC devices or as a source for linguistically intact augmented communicators | The lists did show overlap, but decreasing percentages thereof with increasing list size, suggesting inefficiency of standard word lists to predict words required by AAC users and leading possibly to the spending of time learning "unnecessary" words. Authors suggest the use of standard lists balanced with individualised considerations for acceleration vocabularies. |
| **Fried-Oken & More (1992)** | An initial vocabulary for non-speaking preschool children based on developmental and environmental language sources | To create an age appropriate list of single words which would serve the acquisition of expressive language and linguistic interaction for young children | Quantitative survey and observational | English; North America | a. 15 Parents and 15 clinicians of 15 Children young non-speaking children  b. 30 speaking children matched for age with the 15 non-speaking children. | The adults were asked to list 100 most important single words that their child would say if they could talk.  The 30 children were recorded while engaged in a play activity and in a conversation about home activities, producing a sample of 1 000 words each. | A great degree of variance was found in the word lists produced by adults. The spoken samples giving 30 000 words, of which in total 1 600 words were different. Authors however highlight the use of the same activity for spoken sample collection and question if this influenced the small number of different words. Authors also note that no one word was found to be common to all word lists, and highlight the variance of word lists and thus the individual nature of vocabulary. |
| **Stuart, Vanderhoof & Beukelman, (1993)** | Topic and vocabulary use patterns of elderly women. | To collect communication samples from normally communicating elderly individuals and to analyse the | Quantitative observational | English; North America | Five retired women aged 63 – 79 years | The women were given portable voice recorders which they were free to turn on and off as they wished while going about their day until they had completed an hour of recording. From this data, 3 000 words per | Findings were analysed both for topics and vocabulary used. Vocabulary was selected with a cut off of .5 per 1 000 frequency giving 268 words. Proportionally the first 250 words could account for 77.8% of all words used. In commonality measures, the top 99 words were used by all participants, down to a |

110

| Author and year | Title | Aim | Design | Language Studied and Country | Participants | Procedures | Findings regarding vocabulary |
|---|---|---|---|---|---|---|---|
| | | samples for frequency and commonality of topics and consider these in the light of messages for those requiring AAC | | | | participant was taken. | mean commonality score of 3.68 for the final 226-250th word. Authors concluded that the core words are much the same for the elderly, as for AAC users, preschoolers, and middle schoolers as reported in previous studies. |
| **Marvin , Beukelman & Bilyeu (1994)** | Vocabulary-use patterns in preschool children: Effects of context and time sampling. | To compare word use patterns between home and preschool settings, to compare the commonality of word usage appear with different contexts and over time, and to describe content and structure words within different contexts and over time | Quantitative observational | English; North America | 10 preschool children from three different preschools aged between 4 and 5y2mo participated in the research conducted both at preschool and in home contexts. | Portable voice activated recorders were worn by the participating children during data collection which lasted 2 to 2.5 hours at school then at home. The children were divided into two groups, the first of whom were recorded at home within two weeks of the preschool recording and the second group recorded on the same day. The recordings were transcribed verbatim and 2 000 word samples from each child were computer analysed for number of different words, frequency of occurrence, the different context and group samples were compared in composite lists to compare commonality and differences in home and school vocabulary | An analysis of type token ration revealed the diversity of vocabulary remained very similar across contexts. The word frequency analysis showed the most frequently occurring words were identical across contexts and that a core of 250 words was stable across contexts. An analysis of the different words by context revealed that one quarter to one third of the words used by the children occurred in one context and not the other. The analysis of lists of words used in only one contexts vs. those used in both contexts revealed structure words accounting for the majority of cross-context words, and content vocabulary accounting for single-context vocabulary. The most frequently occurring words in both contexts (divided into structure and content words) are published. |

111

Appendix A

| Author and year | Title | Aim | Design | Language Studied and Country | Participants | Procedures | Findings regarding vocabulary |
|---|---|---|---|---|---|---|---|
| **Balandin & Iacono (1998)** | A few well chosen words | To discover if professionals supporting adult AAC users can accurately predict topics and key words of conversation in the workplace, whether there was commonality between the professional's predictions. | Quantative survey | English; Australia | 10 professionals working with people with communication impairment | The professionals were asked to fill out a survey predicting topics of meal-break conversation and keywords use in conversation around those topics on named days of the week. The predictions of the professionals were compared to 250 actual 15-minute meal-break conversations previously collected by the first author. | There was some commonality of topics and keywords between professionals, but prediction lacked accuracy particularly when pertaining to the keywords of the vocabulary used in the conversations. |
| **Fallon, Light & Paige (2001)** | Enhancing vocabulary selection for preschoolers who require AAC | To develop a more effective and efficient approach to vocabulary selection through two distinct studies | Quantitative; a.= observational b.= survey | English; North America | a. 5 children aged 3.9 and 4.9 years<br><br>b. 45adult service providers acting in teams of 3 for 15 different identified children who were AAC users. | a. The children's speech while engaged in a variety of routine preschool activities was recorded using portable voice recorders until a total of 1 000 words was collected per child. The sample was analysed for total word count, frequency and commonality of word usage. The sample was also analysed for semantic content<br><br>b. Based on the semantic categories identified in a. and on already published word lists, a vocabulary | a. The top 250 most frequently occurring words based on cumulative occurrence showed 45% structure words, and 55% content words. Semantic categories most frequently occurring were; verbs, prepositions, descriptors and pronouns, adverbs and contractions.<br><br>b. The vocabulary selection questionnaire was found elicit the identification of a vocabulary between 174 and 819 words, with approximately 45% of each child's total vocabulary being duplicated by the different informants. The vocabulary selection questionnaire was found to be efficient (taking between 36 and 41 |

112

Appendix A

| Author and year | Title | Aim | Design | Language Studied and Country | Participants | Procedures | Findings regarding vocabulary |
|---|---|---|---|---|---|---|---|
| | | | | | | questionnaire was developed at tested by the team of 45 adults. | minutes on average to complete) and satisfactory (with 96% reporting they were likely to use the questionnaire again) by informants who trialed it. |
| **Banajee, Dicarlo & Stricklin (2003)** | Core vocabulary determination for Toddlers | To identify core vocabulary for toddlers and determine whether this changed depending on the activity and to analyse the functions served by these common words | Quantitative observational | English; North America | 50 toddlers aged between 24 and 36 months | The toddlers wore voice activated recorders during one child directed- and one adult directed activity over 3 days, collecting 6 sets of data which were transcribed to collect the first 150 utterances per participant. | Authors present the list of words with commonality scores between 6 (maximum possible) and 4 and summarise their respective frequency of use calculated as a percentage. Authors present a total of 23 words. Notably there are no nouns on this list, but there are demonstratives, verbs, pronouns, prepositions and articles. Semantically the words represent agents, objects, labeling, and actions. Pragmatically the words serve the purpose of attracting attention, maintaining attention indicating recurrence and terminating interaction. |
| **Kim, Park & Min (2003)** | School-Aged Children and Adult's Core Vocabulary for the Development of an Augmentative and Alternative Communication Tool. | To create a database of vocabulary taken from the general population. As a first step to develop an AAC tool. | Quantative observational | Korean | 24 school aged children and 20 adults | Spontaneous speech samples were collected from a variety of situations common to daily life. 11 092 vocabulary items were collected | The researchers reported a 53% difference between common core of school aged and adult participants; the ratio of structure to content words was 34.8%:65.2%. Researchers reported a high frequency pattern showing the first 25 words representing 34%, and the first 50 words representing 58% of the total sample. |
| **Lee, Kim & Park (2005)** | A preliminary study for the core and fringe AAC vocabulary used by elementary school students. | To analyse core and fringe vocabulary as used by nondisabled children in elementary | Quantative observational | Korean | 91 elementary school students | 29 580 spontaneous words were collected from the 91 participants in various school situations (8 different classes, as well as lunch and play times). The words were analysed for frequency of | Researchers reported the number of different words to be 276 which accounted for 81.13 % of the total sample collected. Researchers reported differences in vocabularies used by the different children in different contexts. The authors noted that the pronoun |

113

| Author and year | Title | Aim | Design | Language Studied and Country | Participants | Procedures | Findings regarding vocabulary |
|---|---|---|---|---|---|---|---|
| | | | | | | occurrence, total number of words as well as number of different words. | category was exceptionally high, in total accounting for 4 689 words. The authors concluded the fringe vocabulary to be a total of 2 894 words, accounting for approximately 9% of the total sample, nonetheless it is noted that although not a high proportion, the fringe vocabulary played an important role in the communication exchange. |
| **Liu & Sloane, (2006)** | Developing a core vocabulary for a Mandarin Chinese AAC system using word frequency data | To determine a vocabulary list for Mandarin Chinese for the purpose of AAC system | Quantitative descriptive | Mandarin Chinese | Authors used a number of existing computational linguistic corpora, amounting to 44,404 tokens. | The authors set out 3 major elements to vocabulary selection: core vocabulary, semantic primes and user expectation. The data set was preprocessed to remove unnecessary information leaving 43 968 lexemes. The data was analysed for any patterns (such as common phrases and combination patterns), was tagged with parts of speech and lexeme frequency analysis. | The researchers report that core vocabulary is representative in Mandarin Chinese, finding the top 1000 lexemes representing over 90% of the total word appearances. The top 100 most frequently used words are listed (in Mandarin); authors report that this finding differs to that found when analysing written Mandarin Chinese. Authors demonstrated that standard practice in AAC development can be applied to Asian languages, but draw attention to necessary linguistic and technical challenges to be overcome when applying these practices to a different language. |
| **Trembath, Balandin & Togher (2007)** | Vocabulary selection for Australian Children who use augmentative and alternative communication | To identify words most frequently and commonly used by Australian children, and thus to inform AAC vocabulary selection for the peers. | Quantitative observational | English; Australia | 6 children (3 boys and 3 girls) aged between 3 and 5. | The participants wore voice recorders and their spontaneous spoken communication was recorded as they went about their preschool activities. Recording was staggered over a number of days to collect a variety of contexts. In total 3 000 words were collected from each participant (18 000 total sample). The sample was | Of the 18 000 total sample 1 411 different words were found. Authors reported a core vocabulary of 263 words which were determined to account for 79.8% of the communication sample with an average commonality score of 5.14.<br><br>Authors noted the influence of current events and context on the vocabulary used by the children which extended into core with the word "Spiderman" due to the coinciding of research with the release |

114

| Author and year | Title | Aim | Design | Language Studied and Country | Participants | Procedures | Findings regarding vocabulary |
|---|---|---|---|---|---|---|---|
| | | | | | | analysed for the number of different words in total sample, the frequency of use for each word as well as the proportion of participants to use each word. | of the popular film and related marketing paraphernalia which influenced the children's play and conversation. Authors caution overreliance on word lists, but support the use of multiple sources for vocabulary determination. |
| Chen, Hill & Yao (2009) | Preliminary vocabulary frequency findings for Mandarin Chinese AAC treatments. | To demonstrate reliable transcription and word segmentation methods for Mandarin Chinese and to identify vocabulary frequency of native Mandarin Chinese in dyadic conversation. | Quantitative observational | Mandarin; Chinese | 12 adults (aged between 26 and 67) | Participants wore portable voice recorders while carrying out conversation over 1 hour. The middle 20 min of recording was transcribed in MC and exported into a word list per participant. Participants produced between 1 567 and 3 225 words each. This was analysed for frequency of each word individually and cumulatively. A cumulative percentage per word was generated and words were ranked. | On average 57.2 words made up 60% of the total words used by participants. Authors reported the top ten high frequency words which made up between 22 and 29% of the vocabulary. Authors report a high level of overlap between participants. Authors discuss the difference in previous automated analyses of MC which have generated lists of characters based on units of meaning rather than individual characters required to generate a grammatically correct utterance. |
| Crestani, Clendon & Hemsley (2010) | Words needed for sharing a story: Implications for vocabulary selection in augmentative and alternative communication. | To examine vocabulary used in a variety of narrative tasks, to inform vocabulary selection for AAC. | Quantative observational | English; Australia | 18 children between the ages of 5;0 and 7;2 with language development within normal range (a battery of tests was applied). | The children's speech was recorded while they participated in a number of structured narrative tasks; a story retelling (fictional narrative), a personal narrative, and a story telling based on a picture sequence (script narrative). The transcriptions were analysed for frequently occurring content and structure words. The vocabulary was analysed as a complete set as well as | The participants produced 145 samples, in total 6 679 words, of which 908 were unique. 173 words could account for 80% of the sample. The top 50 words overall are reported. The authors discuss the content of the various word lists generated for each activity at length, and compare these to already published lists. The authors conclude that the children use a rich vocabulary which is essential to them being able to share information. |

115

Appendix A

| Author and year | Title | Aim | Design | Language Studied and Country | Participants | Procedures | Findings regarding vocabulary |
|---|---|---|---|---|---|---|---|
| | | | | | | grouped related to the activity, the sample produced by the younger group of participants was also compared to the older. | |
| **Robillard, Mayer-Crittenden, Minor-Corriveau & Belanger (2015)** | Monolingual and Bilingual children with and without primary language impairment: Core vocabulary comparison | To examine frequently used vocabulary items by monolingual, bilingual and children with primary language impairment (PLI), with the aim of informing vocabulary selection for AAC of children in bilingual environments | Quantative observational | French; Canada | 57 children in all, between the ages of 53 and 77 months. Divided into groups; monolingual, bilingual English dominant, bilingual French dominant, and language impaired (all bilingual). | The participants wore voice recorders while at school, language samples were collected from participants over a single day each. The vocabulary sampled was examined within the respective groups for frequency and commonality, core vocabulary was determined using a frequency of at least 0.5 per 1 000 and commonality of at least 50% of participants. Each group produced a list of core vocabulary which was then compared to the other groups. | The authors reported that there was no significant difference of French core vocabulary among the groups studied. |
| **Boenisch & Soto (2015)** | The oral core vocabulary of typically developing English-speaking school-aged children: Implications for AAC practice. | To identify the most frequently used words of typically developing school aged children and their peers who are using English as a second language (ESL). | Quantative observational | English; North America | 30 children aged between 7 and 14 years, 22 of whom were first language English speakers and 8 were ESL speakers. | The children wore portable voice recorders while carrying out two different activities in their normal school day. On average 105 minutes sample was collected from each participant. The samples were analysed for total number of words, total number of different words as well as frequency of word use between participants. Words | Authors report a high degree of overlap in the high frequency vocabularies between first language and ESL participants. Patterns of core vocabularies found in other studies were comparable to this in terms of word types found in the high frequency lists as well as the proportions of unique words to total sample. |

116

Appendix A

| Author and year | Title | Aim | Design | Language Studied and Country | Participants | Procedures | Findings regarding vocabulary |
|---|---|---|---|---|---|---|---|
| **Shin & Hill (2016)** | Korean word frequency and commonality study for augmentative and alternative communication | To produce a Korean core vocabulary list based on the speech of Korean adults; to discuss characteristics of these words; and to discover whether grouped frequency distribution data analysis would support the identification of high and low frequency vocabulary. | Quantative observational | Korean; Seoul dialect | 12 monolingual native Korean speakers, 6 men and 6 women between ages of 18 and 65 who were healthy and without cognitive and sensory impairments (a battery of tests was carried out) | were further classified by their various parts of speech. The participants engaged with the researcher in a dyadic conversation which was recorded and later transcribed. The researcher prompted the conversation using open questions but topics were not controlled. The data collected then was analysed for the number of different words as well as the word frequency ratio (%). The word frequency rate was calculated from this for each word. The data was subjected to grouped frequency distribution to produce different distribution patterns for high and low frequency words in order to ascertain a core. The list was further reduced using word commonality scores. | A total of 3 669 different words from a total vocabulary of 16 944 words were collected. Authors reported the grouped frequency distribution data at intervals of .1, .2 and .3 per mil. A clear pattern was discernible at .2 and .3 per mil grouped frequency and authors selected .2 per mil list. This gave a list of 627 words which accounted for 75.83% of total sample. After applying the commonality measure of 6 to this list, the final core vocabulary list identified was 219 words. The proportions of various parts of speech of this list are given. Authors discuss the linguistic differences in Korean which influence vocabulary studies which seek to inform AAC practice. |

117

# APPENDIX B

# ETHICS APPROVAL (UNIVERSITY OF PRETORIA)

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Research Ethics Committee

12 April 2016

Dear Prof Bornman

| Project: | Determining a core vocabulary for Zulu-speaking pre-schoolers: implications for AAC system design |
|---|---|
| Researcher: | JR Mngomezulu |
| Supervisor: | Dr K Tönsing |
| Department: | Centre for Augmentative and Alternative Communication |
| Reference: | 04578440 (GW20160308HS) |

Thank you for the application that was submitted for ethical consideration.

I am pleased to inform you that the above application was **approved** by the **Research Ethics Committee** on 7 April 2016, conditional to written permission being granted by:

• Department of Education.

Please note that data collection may not commence prior to the Department of Education giving permission and subject to final approval by this committee. To facilitate the administrative process, please respond to Ms Tracey Andrew at tracey.andrew@up.ac.za or Room HB 7-27, at your earliest possible convenience.

Sincerely

**Prof Maxi Schoeman**
**Deputy Dean: Postgraduate Studies and Ethics**
**Faculty of Humanities**
**UNIVERSITY OF PRETORIA**
**e-mail:tracey.andrew@up.ac.za**

Kindly note that your original signed approval certificate will be sent to your supervisor via the Head of Department. Please liaise with your supervisor.

118

**APPENDIX C**

**ETHICS APPROVAL (KWAZULU NATAL DEPARTMENT OF EDUCATION)**

# education

Department:
Education
**PROVINCE OF KWAZULU-NATAL**

Enquiries: Phindile Duma          Tel: 033 392 1004          Ref.:2/4/8/747

Mrs J Mngomezulu
18 Plantation Road
Pietermaritzburg
3201

Dear Mrs Mngomezulu

**PERMISSION TO CONDUCT RESEARCH IN THE KZN DoE INSTITUTIONS**

Your application to conduct research entitled: **"DETERMINING A CORE VOCABULARY FOR ZULU-SPEAKING PRESCHOOL CHILDREN"**, in the KwaZulu-Natal Department of Education Institutions has been approved. The conditions of the approval are as follows:

1. The researcher will make all the arrangements concerning the research and interviews.
2. The researcher must ensure that Educator and learning programmes are not interrupted.
3. Interviews are not conducted during the time of writing examinations in schools.
4. Learners, Educators, Schools and Institutions are not identifiable in any way from the results of the research.
5. A copy of this letter is submitted to District Managers, Principals and Heads of Institutions where the Intended research and interviews are to be conducted.
6. The period of investigation is limited to the period from 18 March 2016 to 30 June 2017.
7. Your research and interviews will be limited to the schools you have proposed and approved by the Head of Department. Please note that Principals, Educators, Departmental Officials and Learners are under no obligation to participate or assist you in your investigation.
8. Should you wish to extend the period of your survey at the school(s), please contact Miss Connie Kehologile at the contact numbers below
9. Upon completion of the research, a brief summary of the findings, recommendations or a full report / dissertation / thesis must be submitted to the research office of the Department. Please address it to The Office of the HOD, Private Bag X9137, Pietermaritzburg, 3200.
10. Please note that your research and interviews will be limited to schools and institutions in KwaZulu-Natal Department of Education.

(Please See List of Schools Attached)

**Nkosinathi S.P. Sishi, PhD**
**Head of Department: Education**
**Date: 18 March 2016**

KWAZULU-NATAL DEPARTMENT OF EDUCATION

POSTAL:        Private Bag X 9137, Pietermaritzburg, 3200, KwaZulu-Natal, Republic of South Africa   ...dedicated to service and performance
PHYSICAL:      247 Burger Street, Anton Lembede House, Pietermaritzburg, 3201. Tel. 033 392 1004 beyond the call of duty
EMAIL ADDRESS: kehologile.connie@kzndoe.gov.za / Phindile.Duma@kzndoe.gov.za
CALL CENTRE: 0860 596 363; Fax: 033 392 1203 WEBSITE:        WWW.kzneducation.gov.za

**APPENDIX D**

**PRINCIPAL INFORMATION LETTER AND PERMISSION REPLY**

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

**Faculty of Humanities**

April 2016

Dear _____,

Re: Permission to access preschool of _____ to participate in a research study

My name is Jocelyn Mngomezulu. I am currently doing a Master's degree in Augmentative and Alternative Communication (AAC) at the University of Pretoria. The title of my study is "Determining a core vocabulary for Zulu-speaking preschool children." The aim of the study is to find out what words Zulu-speaking preschool children use in their typical routines at the preschool. For this reason I would like to record and analyse the speech of some preschool children.

I have been granted permission by the KZN Department of Education to access particular preschools in order to carry out the above research. Please see attached copy of this permission letter.

I would be much obliged if you would permit me to include the preschool at

_____ in this study.

**Rationale for the study**
Children in South Africa who cannot rely on natural speech for their communication are sometimes given an augmentative and alternative communication (AAC) system to use. These may include picture symbols on paper or in a file or a speech device. Choosing the right words for these communication systems is one of the most important steps to ensure the system works well and allows the child to communicate in different situations. However, at present we do not have guidelines to tell us which words to choose for an AAC system for a child who will be using Zulu to communicate. This study is a first step to develop such guidelines.

**What will be expected of the school?**
I will require the help of the teachers to nominate two (2) children who could possibly participate in the study. Teachers will also be asked to complete a one-page background questionnaire about the preschool, its facilities and the programme. I will also require the help of the teacher to send information and consent forms to the parents of the nominated children in order to gain their consent. If parents consent for their children to participate, and children are happy to do so, they will be fitted with a small voice recorder which will allow the researcher to record the words the child uses during their time at the preschool. The researcher or her assistant will fit and remove the voice recorder, remaining nearby throughout the recording time, but will not interfere with the normal routine of the preschool. The routines and programme of the preschool will not be altered

1

120

or interfered with during the child's participation. Recording will continue on consecutive days until a total of 3000 words have been collected for each participating child. I expect that this will not take longer than 5 days.

**What will be expected of the children participating in the study?**
- The children will meet me with the teacher present. I will ask them if they assent to participate. I will use pictures to explain what the study involves and they will also respond using pictures as well as verbally.
- If the child agrees, he/she will be required to wear small bag around their waist containing a small voice recorder and a microphone clipped to their collar or shirt during their normal routines at preschool, on consecutive days for no longer than an estimated 5 days.
- The children will be instructed not to adjust the recorders. They will also be told to tell an adult if they need help, or if they wish to stop participating. The adult will either assist them or remove the voice recorder. Children may stop the procedures at any time without any negative consequences to them.

**The following ethical principles will be upheld within this study:**
- Permission has been obtained from the KwaZulu Natal Department of Education (KZNDoE)
- Written consent from all participants' parents and written assent from the participants themselves will be obtained prior to conducting the study
- All participants will be made aware of their right to withdraw from the study at any point in time without any negative consequences to themselves.
- The recordings which are made during the study will be accessed only by the researcher, her supervisors and the research assistants.
- All information will be kept confidential from those external to the study. Any identifying information will be removed from the transcription (e.g. names of people and places will not be transcribed). No individual or school names will be mentioned in any published data.

**Who will have access to the results of the study?**
The research will be stored in both hard copy and electronic format at the University of Pretoria in the Centre for Augmentative and Alternative Communication for 15 years. The data obtained from the research will be used for writing a Master's dissertation, writing a scientific paper and for presentation at professional conferences and seminars. A summary of the results will be made available for any interested staff or parents. Transcriptions (from which all identifying information has been removed) may be used for secondary data analysis. Voice recordings will only be used for further analysis if consent from the parents and assent from the children has been obtained again.

**What are the risks and the benefits?**
At no time during the participation in the research will the children be at risk of any harm. The children will not miss out on any of their daily programmes through participating in this research. Potential benefits of this study may include extending research within the field of AAC and providing empirical evidence to help guide practitioners when designing AAC systems for children who need AAC and who use Zulu as their language of communication.

Please feel free to contact me or my supervisors if you have any questions about this study. I look forward to receiving your response.

Kind regards,

_____  07 April 2016
Jocelyn Mngomezulu  _____
▮▮▮▮▮▮▮@rocketmail.com  Date
Cell: ▮▮▮▮▮▮▮

_____  07 April 2016
Dr. Kerstin Tönsing  _____
Supervisor  Date
▮▮▮▮▮▮@up.ac.za
Tel ▮▮▮▮▮▮

_____  07 April 2016
Prof Shakila Dada  _____
Co-supervisor  Date
▮▮▮▮▮▮@up.ac.za Tel
▮▮▮▮▮▮

Appendix D

## Principal permission: Reply Slip

**Name of principal**: _____

**Name of School**: _____

**Project title:  Determining a core vocabulary for Zulu-speaking preschoolers.**

**Researcher:**          Mrs. JR Mngomezulu
                         Master's Student
                         Centre for AAC
                         Tel: ███████

I,_____,
                     Name and surname

*(please tick box that applies)*

☐     give permission to Mrs Jocelyn Mngomezulu to recruit children from the preschool named above for possible participation in the study entitled *"Determining a core vocabulary for Zulu-speaking preschoolers"*. This permission is voluntary and I understand that I may it at any time. I understand that participating children will be audio-recorded. I understand that the data will stored for 15 years at the CAAC and that all data will be treated confidentially.  I understand that the data maybe re-used for analysis. I understand that the data may be used for a scientific article and for conference presentations. I understand that all information used and obtained in this study will be treated as confidential.

                              OR

☐     do not give permission to Mrs Jocelyn Mngomezulu to recruit children from the preschool named above for possible participation in the study entitled *"Determining a core vocabulary for Zulu-speaking preschoolers"*.

_____
Principal Signature

_____
Date

|  |
|---|
| School stamp |

4

123

**APPENDIX E**

**PRESCHOOL BACKGROUND QUESTIONNAIRE AND NOMINATION FORM**

*Please complete the following questionnaire. Please write on the spaces provided or tick the appropriate block.*

**General Information**

Name of preschool     _____

Name of person completing the form     _____

Position     _____

Is Zulu the only language used in the preschool classroom?

    Yes ☐

    No ☐

If no, what other languages are used?     _____

When are these languages typically used?     _____

Is Zulu the only language used in the preschool playground?

    Yes ☐

    No ☐

If no, what other languages are used?     _____

**Information about the facilities of the preschool**

How many classrooms does the preschool have?     _____

Does the preschool have electricity?

    Yes ☐

    No ☐

Does the preschool have running water?

124

Yes ☐

No ☐

If yes, is this water in the yard or inside the buildings?

Inside tap ☐

In yard ☐

What kind of toilets does the preschool have?

indoor ☐

outdoor (pit) ☐

outdoor (flush) ☐

portaloo ☐

don't have ☐

How many toilets are available to staff? _____

How many toilets are available to children? _____

Is the preschool fenced?

Yes ☐

No ☐

What type of roof is on the preschool buildings?

corrugated iron ☐

tiles ☐

grass ☐

other ☐

*if other please specify* _____

Does the preschool have access to the following communication technologies?

A landline telephone ☐

A fax machine ☐

Internet ☐

**Information about the children and the programme**

125

Appendix E

How many children are enrolled at the preschool?

Do the parents pay fees?

Yes ☐
No ☐

If yes, how much do they pay? _____ monthly / annually

What are the ages of the children?

From _____ (youngest), to _____ (eldest).

Are the children arranged into groups?

Yes ☐
No ☐

If yes, what is the grouping arrangement?

| Class/group name | Age of children in group (range) | Number of children in group | Number of educators in group | Number of assistants in group |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

If grouped, do these groups play together of join together during the day?

Yes ☐
No ☐

Please describe: _____

Does your preschool follow a particular programme or timetable?

Yes ☐
No ☐

If yes, please describe: _____

126

Appendix E

_____

_____

_____

Does your preschool use a curriculum?

|        |      |   |
|--------|------|---|
| Yes    | Yes  | ☐ |
| No     | No   | ☐ |

If yes, please indicate what this is: _____

_____

**Nomination of participants**

The goal of the study is to obtain a realistic sample of vocabulary used by Zulu speaking preschoolers. Would you kindly nominate one girl and one boy from your class who speaks Zulu as a first language and who, in your opinion, speaks how you would expect a child of his age to be speaking? (That is, the speech of the child is typical or normal). The child should also demonstrate overall typical development. The child must be between 5 years and 6 years 11 months old. He of she must have been attending this preschool for at least 6 months prior to now, and must attend for at least 2 days a week.

| Name of child | Age | Gender | Group |
|---------------|-----|--------|-------|
|               |     |        |       |
|               |     |        |       |

Please send an information letter, consent form and questionnaire (see attached) to the parents/legal guardian of each of the children you have nominated.

Thank you for taking the time to complete the questionnaire.

I appreciate your help.

127

**APPENDIX F**

**PARENT INFORMATION LETTER AND CONSENT REPLY SLIP**

**Faculty of Humanities**

Dear Sir / Madam

Re: Requesting permission for your child to participate in a research study

I, Jocelyn Mngomezulu, am currently doing a Master's degree in Augmentative and Alternative Communication (AAC). As a part of my studies, I want to analyse the vocabulary use of Zulu speaking preschool children in their typical routines at their preschool. In order to do this I will need to record the speech of 2 children while they engage in their normal routine activities at preschool.

I would therefore like to ask for your permission for your child to participate in this study.

**Why I want to do this study**
Children in South Africa who cannot rely on natural speech for their communication are sometimes given an augmentative and alternative communication (AAC) system to use. Choosing the right words for these communication systems is one of the most important steps to make sure the system works well and allows the child to communicate in different situations. The reason I want to find out what Zulu speaking children say is so that we can begin to understand what kind of words to prepare for children who need a Zulu AAC system.

**What is expected of your child?**
Should your give permission for your child to participate in the study, and should your child be willing to participate in the study, the following will be expected of him/her:
- To wear a small bag containing a digital voice recorder around his/her waist and a small microphone, which will be attached to his collar or shirt. Your child will be asked to wear this during the normal preschool activities.

- To take part in the preschool activities as he/she normally does.

- To tell an adult if he/she does not want to continue wearing the equipment

- To tell an adult if the recorder is bothering him/her

**What are my child's rights?**
Nobody will force you or your child to take part in the study. You or your child may stop taking part in the study at any time. All the recordings and other information about your child will be immediately destroyed. There will not be any negative consequences for yourself or your child if you or your child decide to stop taking part.

Centre for Augmentative and Alternative
Communication, Room 2-36, Com path
Building, Lynnwood Road
University of Pretoria, Private Bag X20
Hatfield 0028, South Africa
Tel +27 (0)12 420 2001
Fax +27 (0) 86 5100841
Email saak@up.ac.za
www.caac.up.ac.za

Fakulteit Geesteswetenskappe
Lefapha la Bomotho

128

All information about your child will be kept strictly confidential. Only I and my assistant will know your child's name. No personal information about your child will be made known when I write or speak about the results of the study.

**What will happen after I collect the information?**
Only I and the research assistants will look at the information you give us and that which your child says. All the information I get will be kept safe at the University of Pretoria for 15 years. I hope to write an article about it for a scientific journal, and also present the results at conferences.

If you want to find out about the results of the study, you are welcome to contact me and I will send the results to you.

**What are the risks and benefits?**
At no time during the participation in the research will your child be at any risk of any harm. Your child will continue to take part in the normal preschool activities as always so he/she will not miss out on valuable class time. The results of this study may help us to know what words we can program onto AAC systems to help Zulu children who cannot speak.

Please would you

1. indicate whether you do/do not give permission for your child to participate on the reply slip below and return this to your child's preschool

2. if you do give permission for your child to take part, please will you fill in the attached questionnaire and return it together with the reply slip to your child's preschool

Please feel free to contact me or my supervisor if you have any questions about this study. I look forward to receiving your response.

Kind regards,

_____         ___May_____2016___
Jocelyn Mngomezulu                              Date
████████████████@rocketmail.com
Cell:█████████

_____         3 May 2016

_____         _____
Dr. Kerstin Tönsing                              Date
Supervisor
Tel:█████████
████████@up.ac.za

---

**Faculty of Humanities**
**Fakulteit Geesteswetenskappe**
**Lefapha la Bomotho**

129

Appendix F

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

**Faculty of Humanities**

**Participant Informed Consent: Consent Reply Slip**

**Name of Child**: _____

**Name of Parent / guardian:**  _____

**Project title:**  **Determining a core vocabulary for Zulu speaking preschoolers**

**Researcher:**          Mrs. Jocelyn Mngomezulu
                        Master's Student
                        Centre for AAC
                        Tel: ██████████

I, _____ ,
                        Name and surname

☐      consent to participate in the study entitled *"Determining a core vocabulary for Zulu speaking preschoolers"* conducted by Mrs. Mngomezulu. This consent is voluntary and I understand that I may withdraw from the study at any time. I understand that the data will stored for 15 years at the CAAC and that all data will be treated confidentially.  I understand that the data maybe re-used for analysis. I understand that the data may be used for a scientific article and for conference presentations. I understand that all information used and obtained in this study will be treated as confidential.

                        OR

☐      do not give consent to participate in the study entitled *"Determining a core vocabulary for Zulu speaking preschoolers"* conducted by Mrs. Mngomezulu.


_____          _____
Parent / Guardian Signature                Date

130

Appendix F

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

**Faculty of Humanities**

Mzali,

Kucelwa imvumo yakho ukuthi umtwana wakho angisize kuleResearch engizoyichaza ngezansi.

Mina Jocelyn Mngomezulu ngiphakathi nokufundela iMasters ephathelene nezindlela ezahlukahlukene zokukhulumisana umangabe kungakhulumeki ngendlela ejwayelekile. Ngisho ukuthi uma umuntu ekhubazekile engakwazi ukukhuluma. Njengoba ngifunda kumele ngenze leResearch okungukuthola ulwazi olusha ngokuhlola izinto. Mina ngifisa ukuthola ulwazi ngokuthi ingane ekhuluma isiZulu ijwayele ukusebenzisa maphi amagama. Ukuze ngenze lokhu kuzomele ngiqophe inkulumo yezingane ezimbili esikoleni somtanakho ngenkathi zisaqhubeka nezijwayele ukukwenza khona.

Ngakho ke ngokuzithoba ngicela imvumo yakho ukuthi umtwana wakho u_____ (igama) angenele, ngiqophe inkulumo yakhe.

**Kungani ngifisa ukwenza leResearch**
Umangabe ingane ingakwazi ukukhuluma ngesinye isikhathi ziyanikezwa izinto zokuthi zikwazi ukusho ezikushoyo yize zingakhulumi. Lokhu kungaba yincwadi noma umshini kodwa kuba namagama phakathi bese umtwana akhombise lokhu akushoyo. Ngenkathi umtwana enikezwa lento kumele kuhlelwe ukuthi leyonto izofakwa maphi amagama okungukuthi phela azosetshenziswa umtwana ngenkathi asho lokhu akushoyo. Lesi yisinyathelo esibalulekile ngoba amagama akhethiwe yiwona azonqumela umtwana ukuthi angathini futhi yini angeke ayisho. Isizathu esenza ukuthi ngifise ukuthola ukuthi ingane ekhuluma isiZulu ithini yingoba ngifuna ukuqonda ukuthi yimaphi amagama ayoba wusizo nabalulekile ukuthi akhethelwe wona umtwana ongakwazi ukukhuluma kodwa ozosebenzisa into emphathela amagama.

**Umtwana wakho kulindelweni kuyena**
Uma uzovuma ukuthi umtwana wakho angenelele, futhi uma umtwana wakho naye evuma ukungenelela, kuzolindelwa okulandelayo:

- Uzogqokiswa isikhwanyana esincane ngathi esemali esizobe siphethe umshini wokuqopha amazwi. Bese kuba ne-Microphone ezofaselwa empahleni njengasesiphikeni sakhe ukuze kuqopheke lokhu azokusho. Umtwana uzocelwa ukuthi aqhubeke nokwenza lokhu ajwayele ukukwenza nezinye izingane nothisha wakhe ngokwenjwayelo abe efake lomshini wokuqopha.

- Ukuthi aqhubeke nosuku lweskole ngokwenjwayelo

- Ukuthi umtwana atshele umuntu omdala uma ngabe esefuna ukuzikhipha lezinto afakwe zona angamane azikhiphele yena.

131

- Umntwana ukhululekile futhi ukuthi atshele umuntu omdala uma ngabe kukhona into emhluphayo kulezinto afakwe zona.

## Athini amalungelo omtwana

Akekho oyokuphoqa noma aphoqe umtwana wakho ukuthi avume ukungenela leResearch. Futhi ningahoxisa noma yinini. Uma kwenzeka nihoxisa, ngenkathi nikhetha ukuhoxisa okuqoshiwe nayo yonke imniningwane iyosuswa ilahlwe. Futhi ayikho into embi eyokwenzeka uma ngikhetha ukuhoxisa.

Noma kuqhutshekwa, yonke imniningwane ngomtwana wakho iyogcinwa iyimfihlo. Kuyoba yimina nalona ongisizayo kuphela oyokwazi imniningwane negama lomtwana. Ayikho imniningwane ezovezwa noma mangabe ngibhala noma ngikhuluma ngeResearch engiyenzile. Kodwa imininingwane nokuqoshiwe kuyobe sekuhlanganiswe ndawonye kungasabonakali ukuthi ubani owayethini nowayenjani futhi.

## Kuyokwenzekani uma sengiqedile ukuqopha?

Kuyoba yimi nalona ozongisiza oyobheka imininingwane yenu nalokhu okuzoqoshwa okuyinkulumo yomtwana wakho. Konke okumayelana naleResearch kuzogcinwa kubekwe kahle eUniversity of Pretoria CAAC kube yiminyaka ewu15. Kuyisifiso sami ukuthi ngibhale indaba emfushane eyofundwa abanye abasebenza ngalabantwana abangakhulumi futhi ngiphinde ngihambe inkomfa ngifike ngisho ukuthi umsebenzi uhambe kanjani kwatholakala ukuthini.

Uma kade uthanda nawe ukwaziswa ngemiphumelo aleResearch, ukhululekile ukuthi ungithinte ngiyothumelela okuyimphumelo.

## Kuyosiza ngani futhi kuyingozi ngani ukungenelela?

Angisoze ngalimaza umtwana wakho ayikho futhi into ezomlimaza uma engelelela. Umtwana wakho futhi uyoqhubeka nokufunda nokwenza izinto asuke eyozenza esikholeni ngakho angeke ngithi kukhona okuzomeqa okwesikole. Imiphumela siyethemba ukuthi iyosisiza ukwazi ukuthi yimaphi amagama okumele siwafake emishinini nasezincwadini zalezizingane ezingakwazi ukukhuluma ukuze nazo zisho ezikushoyo njengezingane ezinye.

Ngokuzithoba ngicela ukuthi

1.  Usho ukuthi uyavuma noma awuvumi ukuthi umtwana wakho angenelele. Uzosebenzisa lesisiqephu esingezansi ungihambisele sona esikoleni somtwana.

2.  Uma uvuma bengicela futhi ukuthi ugcwalise lemibuzo efakiwe ubuyise konke nesiqephu sokuthi uyavuma ukuyise esikoleni somtwana.

Ngicela ukuthi ukhululeke ungithinte noma usupervisor wami uma unemibuzo ngaleReseach. Ngiyabonga.

Appendix F

Yimina,

_____
Jocelyn Mngomezulu
Jocelyn.mngomezulu@rocketmail.com
Cell: 0726951651

_____
Date

_____
Dr. Kerstin Tönsing
Supervisor
Tel: 012 420 4729
Kerstin.tonsing@up.ac.za

3 May 2016
_____
Date

**Faculty of Humanities**
**Fakulteit Geesteswetenskappe**
**Lefapha la Bomotho**

133

Appendix G

**APPENDIX G**

**BIOGRAPHICAL QUESTIONNAIRE**

*Please complete the following questionnaire. Please write on the spaces provided or tick the appropriate block.*

**Identifying information**

| | |
|---|---|
| Name of Child | |
| Date of Birth | |
| Gender | |
| Person filling in the form | |
| Relationship to child | |
| Contact details (*phone number*) | |

**General information about the child**

Does the child speak Zulu as a home language?

Yes ☐
No ☐

At what age did he/she start speaking?

_____

Does the child speak any other languages?

Yes ☐
No ☐

If yes, what language (s) does he/ she speak?

_____

Does the child have any physical / sensory disabilities?                    *tick if there is a problem*

vision ☐
hearing ☐
walking ☐

134

Do you think your child is developing well at the moment?

Yes ☐

No ☐

**Information about the child's home environment**

Siblings and other children living together with the child at home

| Gender | Relationship to the participating child | Age | Language most spoken by this child | Other languages spoken | Tick those that spend most time with the child |
|---|---|---|---|---|---|
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

Adults living together with the child at home

| Gender | Relationship to the participating child | Age | Language most used | Other languages spoken | Tick those that spend most time with the child |
|---|---|---|---|---|---|
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

What language (s) is (are) used in conversations at home?

*Please list all*

_____

_____

_____

Which language is spoken the most?

_____

What language does the child most often hear spoken on the radio (if you play the radio in your home)?

_____

Does the child watch television?

| | |
|---|---|
| Yes | |
| No | |

If yes, what language do you think that they are most exposed to through the TV?

_____

**Information about the family home and surroundings**

Do you have access to tap water?

| | |
|---|---|
| Yes | |
| No | |

If yes, where is the tap?

| | |
|---|---|
| in the yard | |
| inside the house | |
| a community tap | |

Does your home have a 'jojo' tank?

| | |
|---|---|
| Yes | |
| No | |

What do you use to cook your food?

| | |
|---|---|
| gas 'cadac' | |
| fire | |
| gas stove | |
| electric stove | |
| other | |

*if other please specify* _____

Does your home have electricity?

Yes ☐
No ☐

How many adults in your home are employed?

_____

Please indicate the source (s) of income that your family has

a family member sends money home ☐
self-employment ☐
someone at home works ☐
social grant ☐

Which of these above contributes the most to your family's income?

_____

Please indicate how much money you think your household has for spending and saving every month?

R 500 - R 2, 000 ☐
R 2, 100 - R 4, 000 ☐
R 4, 100 - R 6, 000 ☐
more than R 6, 100 ☐

Thank you for taking the time to fill out this questionnaire

137

## APPENDIX H

## CHILD ASSENT SCRIPT AND FORM

| Photo of researcher |

Hello, my name is Jocelyn.

I want to ask you whether you will help me to find out what children like to talk about. If you say yes, this is what we will do:

First, I will put this *(show)* little bag with a recorder in it around your waist and put this *(show)* microphone onto your shirt.

Then, you will go and carry on with your school work and play time too while you are wearing the bag and microphone.

I will keep the things you say safe. No one except me and the people who are helping me to find out what children like to say will listen to what you said.

If the recorder is bothering you, don't try and fix it yourself. Instead please tell me or the teacher and we will help you to fix the problem.

If you want to stop at any time, I want you to please tell me or the teacher or point to this picture of the stop sign and I will take the recorder off for you.

138

Appendix H

**Training question for the assent form**

| | |
|---|---|
|  | **Is your name _____(say the child's name)?**  |
|  | **Are you a _____boy/girl? (say the wrong one)**  |
|  | **Is your school very far from your house?**  |

139

## Child Assent Form

| | |
|---|---|
|  | *Do you understand everything I explained to you?* |
|  | *Do you understand that it is your choice to help me today?* |
|  | *Do you understand that you can stop anytime you want to?* |
|  | *Do you understand that I will be recording your talking today?* |
|  | *Do you have any questions?* |
|  | *Are you happy with the way your questions were answered?* |
|  | *Do you want to wear the recorder today?* |

140

Photo of

researcher

here

Ya_____(name) igama lami uJocelyn.

Ngifuna ukucela ungisize. Ngifuna ukwazi ukuthi izingane zithanda ukukhuluma ngani. Uma uvuma ukungisiza sizokwenza kanjena:

Nasi lesisikhwama esinomshini wokuqopha (show). Uzosifaka lesikhwama okhalo bese ngifaka nayi nale-microphone esikibheni sakho.

Uzobese wena uyahamba uyoqhubeka nokufunda nokudlala ube ufake lesisikhwama sami ne-microphone.

Konke ozokusho ngizokubeka kahle. Kuzolalela mina nalaba abangisizayo. Akekho omunye oyolalela ukuthi utheni.

Uma umshini wami ukuhlupha ungazami ukuwulungisa kodwa uze kimi noma uye kuthisha umtshele. Sizobese siyakusiza silungise umshini noma isikhwama singasakuhluphi.

Uma ufuna ukuyeka futhi uvele ungitshele noma utshele uthisha. Uma uthanda ukhombe lesisithombe sokuthi "wo". Uzobese uyakhunyulwa isikhwama nezinto zokuqopha. Angeke ngize ngithethe.

**Child Assent Form**

| | |
|---|---|
|  | Uzwe kahle konke engikutshele khona? <br>   |
|  | Uyaqonda ukuthi kuzosho wena ukuthi uyavuma noma awuvumi ukungisiza? <br>   |
|  | Uzwile ukuthi ungayeka noma yinini uma ungasafuni? <br>   |
|  | Uzwile ukuthi ngithe ngizoqopha ukuthi utheni? <br>   |
|  | Uyafuna ukubuza? <br>   |
|  | Ngikuphendule kahle imibuzo yakho? <br>   |
|  | Uyavuma yini ukuthi ngikufake umshini wami? <br>   |

142

**Training question for the assent form**

| | |
|---|---|
|  | u_____igama lakho? (say the child's name)? <br><br>          |
|  | Ungumfana / uyintombazane? (say the wrong one) <br><br>          |
|  | Kukude kakhulu ekhaya uma usuka esikoleni? <br><br>          |

## APPENDIX I

## TRANSCRIPTION RULES

Coding conventions suggested by the SALT (Miller & Iglesias, 2012) were followed, for example the transcription of unintelligible words using a certain code, verbatim transcription of any foreign language, annotation of singing and chanting as such, and the verbatim transcription of spoken language errors together with the correct equivalent. Additional rules were made to which were particular to the study. The table below details all the rules which were used.

| Code or rule | Explanation | Example |
|---|---|---|
| c | At the beginning of the line, c indicates the child utterance. It is a requirement of the statistical programme being used to count words | |
| punctuation | The end of an utterance must always be punctuated. A period or question mark may be used, or if the child simply stopped talking unexpectedly a '>' sign may be used. If it so happened that the child was interrupted, a '^' sign may be used. | woza lana.<br>uthini?<br>we><br>angf^ |
| c somethings = something | In raw transcript, an equals sign below the line beginning c, indicates a note on the line above. The = sign excludes the subsequent text from analysis thus allowing annotation within the transcript. | c mina wena asambe = thatha, indicates the mina being said is not the pronoun, but means 'take this' |
| [chant] or [sing] | This indicates an utterance which had been sung or chanted such as a rhyme or as part of a game. These words are not treated to morphological coding, but are treated as single words/whole words. | |
| [cs] | notes a code switch (usually to English), not always annotated but was particularly useful in transcription stage, the [cs] note is eliminated from the tagged transcript as words are treated equally in the tagging system, but the [cs] helps with being able to understand the transcript. | uya[cs]crayona<br>mina awu[cs]six awami |
| (+) | A plus sign indicates time elapsed on the recording, to be added at the top of each page or after a period of silence. | +02:04:54 |
| xx | Represents an incomprehensible word uttered by the child, it will not be tagged, but will be deleted from tagged transcript where it stands alone, | |
| xxing | The xx remains if it forms part of a word the rest of which is comprehendible | njalo wena angixx = it is evident that the child was starting to say that they didn't... something, therefore the angxx will be coded. |

144

| word_word | The use of word_word shows an utterance which will not be tagged (this could be words from a song, or a chant, or counting) | inj_encane_inamazeze_iyesabeka |
|---|---|---|
| CN, TN, PN | People's names as well as place names are not transcribed, but are represented by CN (child name-used for class and playmates), TN (teacher name), and PN (place name). | weCN, ngizoya ePN nomama' |
| utterances of agreement and assent | in addition to saying "yebo" the child may also agree without words. These are to be written as follows; ehhe (open lips voiced) mh (closed lips voiced) | |
| utterances of disagreement and refusal | in addition to saying "cha" the child may also disagree without using words. These are to be written as follows; eheh (open lips voiced), ehehna (open lips voiced with added emphasis) , mhmh (closed lips voiced). | |
| utterances of interrogation | Sometimes the children request information without using formal questions, commonly after being asked a question they will request for the question to be repeated by saying "eh?" this is to be written in this way whether the child used a open lipped voiced 'eh' or closed lipped 'mm?' with question tone. | |
| % | the use of the percent symbol marks an ideophone | lathi %bhu phansi |

145

**APPENDIX J**

**TAGGING MANUAL**

Table J1

*General Tagging Principles Explained*

| General principle | Example |
|---|---|
| When tagging, although separating portions of words according to their parts, the words remain intact and are given multiple tags: this assists in analysis. | The word *uyahamba* in this system can be divided : *u* (indicative 1) *ya* (present tense) *hamba* (verb root) but will be written *uyahamba*/i1 *uyahamba*/pres *uyahamba*/vrHAMBA |
| All concords are represented by a letter or letters (which refers to the part of speech) and a number (which refers to the noun class to which it concords). In the analysis stage certain concords are grouped so as to count the classes together, retaining also the use of the part of speech, such as relative concord (r) or adjectival concord (a) as a single unit. | a1 refers to the adverbial concord to noun class 1; for example, *omkhulu* would be divided as *om* (a1) *khulu* (adjective root or 'ar'), whereas *elikhulu* is *eli* (a5) *khulu* (ar). |
| All stems and roots are given a category according to their part of speech and then the stem/root itself is also written in the tag | *hamba* is a verb, it will be written and tagged as the verb root *hamba*(vrHAMBA)<br><br>*indaba* is a noun, it will be segmented and tagged: *in-* (noun prefix class 9) –*daba*(noun root) and written: *indaba*\|n9 *indaba*\|nrINDABA. |
| Verbs are tagged in the imperative form | *hamba* may appear elsewhere as *bahambile*, but will be tagged, *bahambile*\|i2 *bahambile*\|vrHAMBA |
| All nouns are tagged in their singular form regardless of their presentation (excepting where there is no singular, such as *amanga* and *amasi*) | *izikhwama* = *izi*(n10) *khwama*(nr) = *izikhwama*\|n10 *izikhwama*\|nrISIKHWAMA |
| Words were written as they are said, including verbal contractions | *ikhwama* (sometimes writtten *i'khwama*) is a contracted form of *izikhwama* |
| Contractions are nonetheless, tagged as if they were complete | *ikhwama* = *i*(n10) *khwama*(nr) = *ikhwama*\|n10 *ikhwama*\|nrisikhwama |

146

| English words are included and are treated the same way | c *thatha one = thatha*\|vrTHATHA *one*\|arONE because in English 'one' is an adjective. |
|---|---|
| When English words are included in the noun classes, these are tagged, but the noun root is written in English only, that is , it is shorn of its given noun class | *ngifuna amachips = ngifuna*\|i1s *ngifuna*\|vrFUNA *amachips*\|n6 *amachips*\|nrCHIPS |

147

Table J2 below gives a complete list of all tags used for all the formatives coded in this study. For ease of reference, the formatives and their tags are organized by grammatical function/parts of speech. The major categories are colour coded and include: **questions, substantives, predicatives, qualificatives, descriptives, conjugations** and **interjections.** The second part of the appendix lists some of the roots in the different categories.

Table J2

*List of Tags by Category*

| QUESTIONS IMIBUZO | With the exception of *pho, na* and in this system *eh*, which are by nature interrogative, the Zulu interrogative is formed from other parts of speech. | | |
|---|---|---|---|
| | **Tag** | **Description** | **Example** | **Comments** |
| | \|intPHO | The interrogative interjection is added to the start/end of a question, or is stated alone. | *Pho uthini? Ngenzeni pho?* | pho uthini = pho\|intPHO uthini\|… |
| | \|intNA | The interrogative stem turns a statement into a question. | *Uyahamba na?* | uyahamba na = uyahamba\|i2s uyahamba\|pres uyahamba\|vrHAMBA na\|intNA? |
| | \|intEH | This interrogative utterance usually means 'what did you say?' | The child has been told something, and they respond, "*eh*?", and the statement is repeated. | eh = eh\|intEH? |
| | \|rPHI | The relative stem forms locative questions ('where'). | *Likuphi? Uyaphi? Uwakhephi amanzi?* | uwakhephi = uwakhephi\|s2s uwakhephi\|o6 uwakhephi\|vrIKHA uwakhephi\|rPHI |
| | \|ePHI | The enumerative stem which forms categorical questions ('which')? | *Usho maphi?* | This is dealt with under enumerative qualificatives. |
| | \|eNI | The enumerative stem is used to form many substance questions ('what' and 'why'). | *Usho ngani?* | ngani= ngani\|advNGA ngani\|eNI |
| | \|intWORD | Some questions were coded as stems, listed here. | *-leni,-lani, yini, ngaki, nini,* | The decision to do this was based on the difficulty in locating any obvious formatives for these. |

Appendix J

| | | | | |
|---|---|---|---|---|
| **SUBSTANTIVE** *USOBIZO* | Substantive (usobizo) is comprised of nouns and pronouns. | | | |
| **NOUNS** *AMABIZO* | Nouns are made up of noun classes and noun roots, nouns can also be made from stems of other parts of speech, for the purposes of this system, nouns are tagged as they present. Additional tags applying to nouns are diminutive and reduplication. | | | |

| Tag | Definition | Example | | Comments |
|---|---|---|---|---|
| nr | noun root / isiqu sebizo | ubhontshisi = ubhontshisi\|n3 ubhontshisi\|nrUBHONTSHISI | | NR is as open group |
| n1-15 | noun class / isiqalo sebizo | | | n1-15 is a closed group and is listed below |
| dim | dimunitive / ukunciphisa | isandlana = isandlana\|n7 isandlana\|nrISANDLA isandlana\|dim | | |
| red | reduplication / ukuphinda | iqandaqanda = iqandaqanda\|n5 iqandaqanda\|nrIQANDA iqandaqanda\|red | | The tag is only used when the word is intentionally reduplicated, for instance *isiphukuphuku*, *isiphicaphicwano* and *uNkulunkulu* are not reduplication |

| Code | Noun class | Example | | Translation |
|---|---|---|---|---|
| n1 | umu | umuntu = umuntu\|n1 umuntu\|nrUMUNTU | | person |
| n1a | u | uCN = uCN\|n1a uCN\|CN | | child name |
| n2 | aba | abafana - abafana\|n2 abafana\|nrUMFANA | | boys |
| n2a | o | oCN = oCN\|n1b oCN\|CN | | child name and his associates |
| n3 | umu | umuthi = umuthi\|n3 umuthi\|nrUMUTHI | | tree |
| n4 | imi | imithi = imithi\|n4 imithi\|nrUMUTHI | | trees |
| n5 | i(li) | itshe = itshe\|n5 itshe\|nrITSHE | | stone |
| n6 | ama | amatshe = amatshe\|n5 amatshe\|nrITSHE | | stones |
| n7 | isi | isitsha = isitsha\|n7 isitsha\|nrISITSHA | | container |
| n9 | i(n) | inja = inja\|n9 inja\|nrINJA | | dog |
| n10 | izi | izinja = izinja\|n10 izinja\|nrINJA | | dogs |
| n11 | u(lu) | uphondo = uphondo\|n11 uphondo\|nrUPHONDO | | horn |
| n14 | ubu | ubuhlalu = ubuhlalu\|n14 ubuhlalu\|nrUBUHLALU | | beads |
| n15 | uku | ukudla = ukudla\|n15 ukudla\|nrUKUDLA | | These are verbal nouns, also presenting as verbal infinitive. |

Appendix J

| PRONOUNS IZABIZWANA | There are true pronouns (izabizwana zoqobo) and various descriptive pronouns (izabizwana ezihlungayo) which are demonstrative, presentative, enumerative and descriptive. In this system, descriptive pronouns are dealt with using the relative prefix, rather than having a category on their own. | | | |
|---|---|---|---|---|

| True Pronoun *Isabizwana soqobo* | **Tag** | **Definition** | **Example** | **Comments** |
|---|---|---|---|---|
| | pr1-15 | pr refers to pronoun, where the number refers to the noun class. The letters s and p refer to singular and plural. | | pr 1 - 15 is a closed group and is listed below |
| | stab | stabiliser -*na* | *Kimi* ('to me') is made up of two parts, *ki*(locative) *mi*(pr1s), but it can also present as *kimina*, where it also has a stabiliser -na added on. | |
| | **Tag** | **Component** | **Example** | **Comments** |
| | pr1 | *ye* | *yena* | Pronouns are coded as in the following examples: *yena = yena/pr1 yena/stab* *mina = mina/pr1s mina/stab* *thina = thina/pr1p thina/stab* etc. |
| | pr1s | *mi* | *mina* | |
| | pr1p | *thi* | *thina* | |
| | pr2 | *bo* | *bona* | |
| | pr2s | *we* | *wena* | |
| | pr2p | *ni* | *nina* | |
| | pr3 | *wo* | *wona* | |
| | pr4 | *yo* | *yona* | |
| | pr5 | *lo* | *lona* | |
| | pr6 | *wo* | *wona* | |
| | pr7 | *so* | *sona* | |
| | pr9 | *yo* | *yona* | |
| | pr10 | *zo* | *zona* | |
| | pr11 | *lo* | *lona* | |
| | pr14 | *bo* | *bona* | |
| | pr15 | *ko / kho* | *kona/khona* | |

150

| Demonstrative Pronoun *Isabizwana Sokukhomba* | Tag | Definition | Example | Comments |
|---|---|---|---|---|
| | d | demonstrative of 1st position | *lo, leli, lawa, lokhu* | This is a closed group and is listed below. They perform functions performed in English by the words 'this here' (position 1), 'that there' (position 2) and 'that there yonder'(position 3). Each demonstrative pronoun has at minimum two codes – one to indicate the position (1st, 2nd, or 3rd) and one to indicate the noun class. They may additionally have the codes |stab for the stabilizer and/or |ke for the suffix. All these parts are coded individually as shown in examples below (colour coding added to facilitate readability). |
| | dB | demonstrative of 2nd position | *loyo, lelo, lawo, lokho* | |
| | dC | demonstrative of 3rd position | *loya, leliya, lawaya, lokhuya* | |
| | d1-15 | d refers to demonstrative, and the number refers to the noun class | | |
| | stab | stabiliser -*na* (optional ) | *lo / lona ; le / lena; lawaya / lawayana ; leliya / leliyana* | |
| | ke | suffix -*ke* (optional) | *lawo / lawoke ; lelo/ leloke ; lokho / lokhoke* | |
| | | d | dB | dC |
| | d1 | *lona/ lo = lo\|d lo\|d1* | *loyo = loyo\|dB loyo\|d1* | *loya = loya\|dC loya\|d1* |
| | d2 | *laba = laba\|d laba\|d2* | *labo/laboke = labo\|dB labo\|d2* | *labaya / labayana = labayana\|dC labayana\|d2* |
| | d3 | *lona = lona\|d lona\|d3 lona\|stab* | *loyo / loyoke = loyo\|dB loyo\|d3* | *loya / loyana = loyana\|dC loyana\|d3 loyana\|stab* |
| | d4 | *le / lena = lena\|d lena\|d4 lena\|stab* | *leyo / leyoke = leyoke\|dB leyoke\|d4 leyoke\|ke* | *leya / leyana = leya\|dC leya\|d4* |
| | d5 | *leli* | *lelo* | *leliya* |
| | d6 | *lawa* | *lawo* | *lawaya* |
| | d7 | *lesi* | *leso* | *lesiya* |
| | d9 | *le* | *leyo* | *leya* |
| | d10 | *lezi* | *lezo* | *leziya* |
| | d11 | *lolu* | *lolo* | *loluya* |
| | d14 | *lobu* | *lobo* | *lobuya* |
| | d15 | *loku* | *lokho* | *lokhuya* |

<table>
<tr><td rowspan="2">Presentative<br>Pronoun<br><em>Isabizwana<br>Sendawo</em></td><td colspan="4">The dictionary defines this group as "locative demonstrative copulative," in this system they are named presentative pronouns (<em>isabizwana sendawo</em>).</td></tr>
<tr><td><strong>Tag</strong></td><td><strong>Description</strong></td><td><strong>Example</strong></td><td><strong>Comments</strong></td></tr>
</table>

| Tag | Description | Example | Comments |
|---|---|---|---|
| prs | presentative of 1st position | nangu, nali, nawa, nakhu | This is a closed group and is listed below. Each presentative pronoun has at minimum two codes – one to indicate the position (1st, 2nd, or 3rd) and one to indicate the noun class. They may additionally have the codes 'stab' for the stabilizer and/or 'ke' for the suffix. All these parts are coded individually as shown in examples below (colour coding added to facilitate readability). |
| prsB | presentative of 2nd position | nangoke, nalo, nawo, nakhoke | |
| prsC | presentative of 3rd position | nanguya, naliya, nawaya, nakhuya | |
| prs1-15 | prs refers to presentative, and the number refers to the noun class | | |
| stab | stabiliser | naliyana, nanguyana, nawayana | |
| ke | suffix ke | nangoke, naloke, nasoke | |
| | prs | prsB | prsC |
| prs1 | nangu = nangu\|prs nangu\|prs1 | nangoke = nangoke\|prsB nangoke\|prs1 nangoke\|ke | nanguyana = nanguyana\|prsC nanguyana\|prs1 nanguyana\|stab |
| prs2 | naba = naba\|prs naba\|prs2 | nabo = nabo\|prsB nabo\|prs2 | nabaya = nabaya\|prsC nabaya\|prs2 |
| prs3 | nawu = nawu\|prs nawu\|prs3 | nawo /nawoke = nawoke\|prsB nawoke\|prs3 nawoke\|ke | nawuya +- na = nawuya\|prsC nawuya\|prs3 |
| prs4 | nayi = nayi\|prs nayi\|prs4 | nayoke = nayoke\|prsB nayoke\|prs4 nayoke\|ke | nayiya = nayiya\|prsC nayiya\|prs4 |
| prs5 | nali | nalo +-ke | naliya |
| prs6 | nawa / nanka | nawo+- ke | nawaya |
| prs7 | nasi | naso | nasiya |
| prs9 | nayi | nayo | nayiya |
| prs10 | nazi | nazo | naziya |
| prs11 | nalu | nalo +-ke | naluya |
| prs14 | nabu | nabo | nabuya |
| prs15 | nakhu | nakhoke | nakhuya |

| Enumerative Pronoun *Isabizwana Senani* | The enumerative pronoun is composed of the pronoun (pr1-15) and a quantifier root (qr) which are either DWA or NKE. | | | |
|---|---|---|---|---|
| | **Tag** | **Description** | **Example** | **Comments** |
| | pr1-15 | already described | | |
| | qrDWA | quantifier root meaning 'alone' | lodwa, sodwa | |
| | qrNKE | quantifier root meaning 'all' | lonke, sonke | |
| | | | yedwa = yedwa\|pr1 yedwa\|qrDWA | |
| | | | bodwa = bodwa\|pr2 bodwa\|qrDWA | |
| | | | bonke = bonke\|pr2 bonke\|qrNKE | |
| | qrBILI/ THATHU /NE / HLANU | quantifier root 'two', 'three', 'four',and 'five' | | These roots are coded under adjective root. |
| | prBO | pronoun *bo* used for pr1-2p, when arWORD follows | bobabili | |
| | prYO | pronoun *yo* used for 3 to 4, when arWORD follows | yombili | |
| | prO | pronoun o used for 5 to 6, when arWORD follows | omabili | |
| | prZO | pronoun zo used for 7 to 15, when arWORD follows | zombili | |

| Descriptive Pronoun Isabizwana sesicasiso | The predicative pronoun is used when the description comes prior to the noun, or if the noun is absent. Examples are *amhlophe amadada adlayo* ('they are white, the ducks that are eating') *awami adla kahle* ('mine are eating well') This is currently coded with the relative (r) and then possessive concord (z) and possessive stem, such as awami = awami\|r6 awami\|z6 awami\|posspr1s |
|---|---|

| PREDICATIVE ISILANDISO | Predicative (isilandiso) includes verbs and copulative parts of speech | | | |
|---|---|---|---|---|
| VERBS IZENZO | Verbs present with different endings depending on mood, tense and verbal extensions. In this coding the verb root is tagged and the only changes to the end of the verb which are tagged in addition are plural imperative (yempoqo), and verbal extensions (impambosi zesenzo). | | | |
| | **Tag** | **Description** | **Example** | **Comments** |
| | vr | verb root / isenzo | ngiyahamba = ngiyahamba\|i1s ngiyahamba\|pres ngiyahamba\|vrHAMBA | The verb root is coded in the imperative mood. |
| | plimp | plural imperative/impoqo | hamba**ni** | |
| | xa | applied extension/ yokwenzela | hamb**el**a = hambela\|vrHAMBA hambela\|xa | The vr is always coded, as well as the extension afterward. |
| | xr | reciprocal extension/yokwenzana | hamb**isan**a = hambisana\|vrHAMBA hamb**isan**a\|xc hambisana\|xr | Sometimes more than one extension is used. |
| | xc | caustative extension/yokwenzisa | hamb**is**a | Examples to the left are shown here mostly without coding for ease of reading, but the relevant formatives are **colour coded** in the word. |
| | xn | neuter extension /yokwenzeka | hamb**ek**a | |
| | xp | passive extension / yokwenziwa | han**jw**a | |
| | xi | intensive extension /yokwenzisisa | hamb**isis**a | |
| | red | reduplication of the verb/ yokwenzayenza | hamba**hamba** = hambahamba\|vrHAMBA hambahamba\|red | Herewith two examples with tags applied: ngiyahamba = ngiyahamba\|i1s ngiyahamba\|pres ngiyahamba\|vrHAMBA / sengihamba = sengihamba\|EXCLasp sengihamba\|i1s sengihamba\|vrHAMBA |
| | pres | present | ngi**ya**hamba, ziyahamba, luyahamba | |
| | futZO | near future | ngi**zo**hamba, zizohamba, luzohamba | |
| | futYO | remote future | ngi**yo**hamba, ziyohamba, luyohamba | |
| | immpast | immediate past | **be**ngihamba, bezihamba, beluhamba | |
| | rempast1-15 | remote past concord 1 – 15 | **nga**hamba, zahamba, lwahamba | |
| | pot | potential | ngi**nga**hamba, zingahamba, lungahamba | |
| | opt | optative | u**bo**hamba, zibohamba, | |
| | infinite | infinite mood is counted under noun class 15 'uku-' | ukuhamba = ukuhamba\|n15 ukuhamba\|vrHAMBA | |
| | cont 1 -15 | contingent | **bengo**hamba, bezohamba, belohamba | |
| | hort | hortative | angihambe, azihambe, aluhambe | |
| | EXCLasp | exclusive aspect | sengihamba, zingakahambi, | |
| | PROGasp | progressive aspect | ngisahamba, zisahamba, lungasahambi | |

154

Appendix J

| Verb concords *izivumelwano zesenzo* | Concords allow the use of the verb and serve in pointing to the substantive, the object of the sentence. These are in the following categories: negative, indicative, object, subjunctive, past auxiliary modifier, and participial subject prefix groups. | | | |
|---|---|---|---|---|
| | **Tag** | **Description** | **Example** | **Comments** |
| The Negative | negPRE | negative prefix | akafuni | |
| | negIN | negative infix | engafuni | |
| | g1-15 | negative subject concord 1 to 15 | | g1-15 is a closed group and is listed below |
| | **tag** | **component** | **example** | |
| | g1 | ka | akafuni | Examples are shown here without coding, for ease of reading.  Here following two examples with tags applied: |
| | g1s | ngi | angifuni | |
| | g1p | si | asifuni | angifuni = angifuni\|negPRE angifuni\|g1s |
| | g2 | ba | abafuni | angifuni\|vrFUNA   \ |
| | g2s | wu | awufuni | engafuni = engafuni\|p1 engafuni\|negIN engafuni\|vrFUNA |
| | g2p | ni | anifuni | |
| | g3 | wu | awufuni | |
| | g4 | yi | ayifuni | |
| | g5 | li | alifuni | |
| | g6 | wa | awafuni | |
| | g7 | si | asifuni | |
| | g9 | yi | ayifuni | |
| | g10 | zi | azifuni | |
| | g11 | lu | alufuni | |
| | g14 | bu | abufuni | |
| | g15 | ku | akufuni | |

Appendix J

| Indicative subject prefix | Tag | Description | Example | Comments |
|---|---|---|---|---|
| | i1 - 15 | indicative subject concord classes 1 to 15 | | i1-15 is a closed group and is listed below |
| | **Tag** | **Component** | **Example** | **Comments** |
| | i1 | u | uyathanda | Examples are shown here without coding, for ease of reading.  Here following two examples with tags applied: |
| | i1s | ngi | ngiyathanda | |
| | i1p | si | siyathanda | kuyathanda = kuyathanda\|i15 kuyathanda\|pres |
| | i2 | ba | bayathanda | kuyathanda\|vrTHANDA    /    lithandile = liyathanda\|i5 |
| | i2s | u | uyathanda | lithandile\|vrTHANDA |
| | i2p | i | niyathanda | |
| | i3 | u | uyathanda | |
| | i4 | i | iyathanda | |
| | i5 | li | liyathanda | |
| | i6 | a | ayathanda | |
| | i7 | si | siyathanda | |
| | i9 | i | iyathanda | |
| | i10 | zi | ziyathanda | |
| | i11 | lu | luyathanda | |
| | i14 | bu | buyathanda | |
| | i15 | ku | kuyathanda | |

156

| Object prefix | Tag | Description | Example | Comments |
|---|---|---|---|---|
| | refl | reflective object prefix *isakhi sokuzenza* | ukuzibamba = ukuzibamba\|n15 kuzibamba\|refl ukuzibamba\|vrBAMBA | |
| | o1-15 | object prefix classes 1 to 15 | | o1 -15 is a closed group and is listed below |
| | **Tag** | **Component** | **Example** | **Comments** |
| | o1 | m | ngizombamba | EXAMPLES shown here without coding, for ease of reading.  Here following two examples with tags applied: |
| | o1s | ngi | uzongibamba | |
| | o1p | si | uzosibamba | |
| | o2 | ba | uzobabamba | uzombamba = uzombamba\|i2s uzombamba\|futZO |
| | o2s | ku | ngizokubamba | uzombamba\|o1 uzombamba\|vrBAMBA    / |
| | o2p | ni | ngizonibamba | wangibamba = wangibamba\|sp1 wangibamba\|o1s |
| | o3 | wu | ngizowubamba | wangibamba\|vrBAMBA |
| | o4 | yi | ngizoyibamba | |
| | o5 | li | uzolibamba | |
| | o6 | wa | uzowabamba | |
| | o7 | si | uzosibamba | |
| | o9 | yi | uzoyibamba | |
| | o10 | zi | uzozibamba | |
| | o11 | lu | uzolubamba | |
| | o14 | bu | ngizobubamba | |
| | o15 | ku | ngizokubamba | |

157

| Subjunctive subject prefix | Tag | Description | Example | Comments |
|---|---|---|---|---|
| | s1-15 | subjunctive prefix classes 1 to 15 | | s1-15 is a closed group and is listed below with examples |
| | Tag | Component | Example | |
| | s1 | a | athole | EXAMPLES shown here without coding, for ease of reading. Here following two examples with tags applied: |
| | s1s | ngi | ngithole | |
| | s1p | si | sithole | akucebe = akucebe|s1 akucebe|o2s akucebe|vrCEBA    / |
| | s2 | ba | bathole | kugcine = kugcine|s15 kugcine|vrgcina |
| | s2s | u | uthole | |
| | s2p | ni | nithole | |
| | s3 | u | uthole | |
| | s4 | i | ithole | |
| | s5 | li | lithole | |
| | s6 | a | athole | |
| | s7 | si | sithole | |
| | s9 | i | ithole | |
| | s10 | zi | zithole | |
| | s11 | lu | luthole | |
| | s14 | bu | buthole | |
| | s15 | ku | kuthole | |

Appendix J

<table>
<tr><td rowspan="3">Past auxiliary modifier prefix</td><td><strong>Tag</strong></td><td><strong>Description</strong></td><td><strong>Example</strong></td><td><strong>Comments</strong></td></tr>
<tr><td>sp1-15</td><td>past auxiliary modified subjunctive classes 1 to 15</td><td></td><td>sp1-15 is a closed group and is listed below with examples</td></tr>
<tr><td><strong>Tag</strong></td><td><strong>Component</strong></td><td><strong>Example</strong></td><td></td></tr>
<tr><td></td><td>sp1</td><td>wa</td><td>wawa</td><td rowspan="16">EXAMPLES shown here without coding, for ease of reading. Here following two examples with tags applied:<br><br>ngawa = ngawa|sp1s ngawa|vrIWA   /   lwafuna = lwafuna|sp11 lwafuna|vrFUNA</td></tr>
<tr><td></td><td>sp1s</td><td>nga</td><td>ngawa</td></tr>
<tr><td></td><td>sp1p</td><td>sa</td><td>sawa</td></tr>
<tr><td></td><td>sp2</td><td>ba</td><td>bawa</td></tr>
<tr><td></td><td>sp2s</td><td>wa</td><td>wawa</td></tr>
<tr><td></td><td>sp2p</td><td>na</td><td>nawa</td></tr>
<tr><td></td><td>sp3</td><td>wa</td><td>wawa</td></tr>
<tr><td></td><td>sp4</td><td>ya</td><td>yawa</td></tr>
<tr><td></td><td>sp5</td><td>la</td><td>lawa</td></tr>
<tr><td></td><td>sp6</td><td>a</td><td>awa</td></tr>
<tr><td></td><td>sp7</td><td>sa</td><td>sawa</td></tr>
<tr><td></td><td>sp9</td><td>ya</td><td>yawa</td></tr>
<tr><td></td><td>sp10</td><td>za</td><td>zawa</td></tr>
<tr><td></td><td>sp11</td><td>lwa</td><td>lwawa</td></tr>
<tr><td></td><td>sp14</td><td>ba</td><td>bawa</td></tr>
<tr><td></td><td>sp15</td><td>kwa</td><td>kwawa</td></tr>
</table>

159

Appendix J

| Participial subject prefix | Tag | Description | Example | Comments |
|---|---|---|---|---|
| | p1-15 | participial subject classes 1 to 15 | | corresponds closely to indicative, but differs radically in pronunciation |
| | **Tag** | **Component** | **Example** | p1-15 is a closed group and is listed below |
| | p1 | e | ekhala | EXAMPLES shown here without coding, for ease of reading. Here following two examples with tags applied: |
| | p1s | ngi | ngikhala | |
| | p1p | si | sikhala | ekhala = ekhala\|p1 ekhala\|vrKHALA / kukhala = |
| | p2 | be | bekhala | kukhalalwa\|p15 kukhalwa\|vrKHALA kukhalwa\|xp |
| | p2s | u | ukhala | |
| | p2p | ni | nikhala | |
| | p3 | u | ukhala | |
| | p4 | i | ikhala | |
| | p5 | li | likhala | |
| | p6 | e | ekhala | |
| | p7 | si | sikhala | |
| | p9 | i | ikhala | |
| | p10 | zi | zikhala | |
| | p11 | lu | lukhala | |
| | p14 | bu | bukhala | |
| | p15 | ku | kukhala | |

| *COPULATIVE* *isibanjalo* | This refers to the use of a noun as a verb, not always notable in orthography, since it is often only indicated by tone, except in certain classes. It can sometimes be determined from the sentence structure. | | | |
|---|---|---|---|---|
| | **Tag** | **Description** | **Example** | **Comments** |
| | cop | copulative | yinja = yinja\|cop yinja\|nrINJA | |
| | | | yimi = yimi\|cop yimi\|pr1s | |
| | | | nguwe = nguwe\|cop nguwe\|pr2s | |

| | | | |
|---|---|---|---|
| **QUALIFICATIVE** *ISICHASISO* | Qualificative (isichasiso) is comprised of relative, adjective, enumerative and possessive parts of speech | | |
| **RELATIVE** *ISIBALULI* | Relatives are made up of a relative concord (|r#) and a relative stem (|rWORD) | | |
| | **Tag** | **Description** | **Example** | **Comments** |
| | rWORD | relative stem | banzi, duma, lula, ngcono, phi, buhlungu, luhlaza, lukhuni, manzi, bomvu, nzima… | This doesn't always have to be preceded with a relative concord, for example, *lingcono* uses the indicative verbal concord = lingcono|i5 lingcono|rNGCONO |
| | r 1 – 15 | relative concords | | r# are a closed group and are listed below |
| | drWORD | demonstrative relative | nje, njalo, ngaka, njani | |
| | rsf | relative suffix | ohambayo = ohambayo|r1 ohambayo|vrHAMBA ohambayo|rsf | |
| | **Tag** | **Component** | **Example** | |
| | r1 | o | ongcono | EXAMPLES shown here without coding, for ease of reading. Here following two examples with tags applied:<br><br>ongcono = ongcono|r1 ongcono|rNGCONO  /<br>okwakungcono = okwakungcono|r15 okwakungcono|rempast15 okwakungcono|i15 okwakungcono|rNGCONO |
| | r1s | engi | engingcono | |
| | r1p | esi | esingcono | |
| | r2 | aba | abangcono | |
| | r2s | o | ongcono | |
| | r2p | eni | eningcono | |
| | r3 | o | ongcono | |
| | r4 | e | engcono | |
| | r5 | eli | elingcono | |
| | r6 | a | angcono | |
| | r7 | esi | esingcono | |
| | r9 | e | engcono | |
| | r10 | ezi | ezingcono | |
| | r11 | olu | olungcono | |
| | r14 | obu | obungcono | |
| | r15 | oku | okungcono | |

| ADJECTIVE *ISIPHAWULO* | Adjectives are made up of adjectival concords (a#) and an adjective stem (ar) | | | |
|---|---|---|---|---|
| | **Tag** | **Description** | **Example** | **Comments** |
| | arWORD | adjective root | omkhulu = omkhulu\|a1 omkhulu\|arKHULU | There are a limited number of adjective roots in Zulu, namely: *-khulu, -ncane, -ningi, -dala, -hle, -bi, -ngaka, -de, -fushane, -bili, -thathu, -ne, -hlanu, - nye,-sha* |
| | | | lidala= lidala\|i5 lidala\|arDALA | |
| | | | mathathu = mathathu\|e6 mathathu\|arTHATHU | |
| | | | eliblue = eliblue\|a5 eliblue\|arBLUE | |
| | a 1 – 15 | adjectival concord classes 1 to 15 | | The adjective root isn't always preceded by an adjectival concord e.g. Likhulu uses an indicative verbal concord = likhulu\|i5 likhulu\|arKHULU a# is a closed group and is listed below |
| | **Tag** | **Component** | **Example** | **Comments** |
| | a1 | om | omkhulu = omkhulu\|a1 omkhulu\|arKHULU | EXAMPLES shown here without coding, but clearly there are other parts to the word which are also requiring codes. Two e.g.s: omuhle = omuhle\|a1 omuhle\|arHLE / esikhulu = esikhulu\|a7 esikhulu\|arKHULU |
| | a2 | aba | abakhulu | |
| | a3 | om | omkhulu | |
| | a4 | emi | emikhulu | |
| | a5 | eli | elikhulu | |
| | a6 | ama | amakhulu | |
| | a7 | esi | esikhulu | |
| | a9 | e (m)(n) | enkulu / embi | |
| | a10 | ezi (m) (n) | ezinkulu | |
| | a11 | olu | olukhulu | |
| | a14 | obu | obukhulu | |
| | a15 | oku | okukhulu | |

Appendix J

| ENUMERATIVE INANI | The enumerative is made of two parts, the enumerative concord (e1-15), and the enumerative root (e) | | | |
|---|---|---|---|---|
| | **Tag** | **Description** | **Example** | **Comments** |
| | eWORD | enumerative root | muphi = muphi\|e1 muphi\|ePHI | There are only four enumerative roots; namely : -ni, -nye, -phi, -mbe. The eNI is used with adv and vr to form interrogative. |
| | | | linye = linye\|e5 lunye\|eNYE | |
| | | | ngani = ngani\|advNGA ngani\|eNI | |
| | | | uthini = uthini\|i2s uthini\|vrITHI uthini\|eNI | |
| | e 1 – 15 | enumerative concord classes 1 to 15 | | e# is a closed group listed below with examples |
| | **Tag** | **Component** | **Example** | |
| | e1 | mu | muphi = muphi\|e1 muphi\|ePHI | This root *phi* differs from the root *-phi* used for location, this *-phi* means 'which', for example, *ngithathe luphi lona* (ePHI) ('which one should I take') versus *luphi ngiluthathe* (rPHI) (where is it so I will take it') |
| | e2 | ba | baphi, bani | |
| | e3 | mu | muphi, muni | |
| | e4 | mi | miphi, mini | |
| | e5 | li | liphi, linye | |
| | e6 | ma | maphi, mani | |
| | e7 | si | siphi, simbe | |
| | e9 | yi | yiphi, yinye | |
| | e10 | zi | ziphi, zini | |
| | e11 | lu | luphi, lunye | |
| | e14 | bu | buphi, bunye | |
| | e15 | ku | kuphi, kunye | |

Appendix J

| POSSESSIVE *ONGUMNINI* | The Zulu possessive is formed by the concord which agrees with the item which is possessed followed by the possessive stem (or possessive pronoun) which indicates the owner. These are both closed groups which are listed side by side below. | | | |
|---|---|---|---|---|
| | **Tag** | **Description** | **Comments** | |
| | z 1 – 15 | possessive concords (the item being owned) | | |
| | posspr 1 to 15 | possessive stems (the owner) | | |
| | **Tag** | **Possessive concord** | **Tag** | **Possessive stem** |
| | z1 | wa | posspr1 | -khe |
| | z1s | ka | posspr1s | -mi |
| | z1p | baka | posspr1p | -ithu |
| | z2 | ba | posspr2 | -bo |
| | | | posspr2s | -kho |
| | | | posspr2p | -inu |
| | z3 | wa / ka | posspr3 | -wo |
| | z4 | ya / ka | posspr4 | -yo |
| | z5 | la | posspr5 | -lo |
| | z6 | a | posspr6 | -wo |
| | z7 | sa | posspr7 | -so |
| | z9 | ya | posspr9 | - yo |
| | z10 | za | posspr10 | -zo |
| | z11 | lwa | posspr11 | -lo |
| | z14 | ba | posspr14 | -bo |
| | z15 | kwa | posspr15 | -ko / - kho |
| | | | | |

164

| | | | |
|---|---|---|---|
| **DESCRIPTIVE** *ISIKHANYISO* | Descriptive (*isikhanyiso*) is comprised of adverbs (isandiso) and ideophones (isenzukuthi) | | |
| **ADVERB** *ISANDISO* | Adverbs are made up of adverbial formatives as well as adverbial roots; all adverbs are coded with \|advWORD format; adverbs of place are coded as locative (\|loc); | | |

| Tag | Description | Example | Comments |
|---|---|---|---|
| loc | locative | endlini = endlini\|loc endlini\|n9 endlini\|nrINDLU | Sometimes the noun class is no longer visible, nonetheless it is coded. Locative form changes depending on the noun class which follows, but the code remains unaltered |
| | | ezindlini = ezindlini\|loc ezindlini\|n10 ezindlini\|nrINDLU | |
| | | kuCN = kuCN\|loc kuCN\|n1a kuCN\|CN | |
| | | kwelikhulu = kwelikhulu\|loc kwelikhulu\|a5 kwelikhulu\|arKHULU | |
| advWORD | | adv refers to the part of speech of adverb, WORD refers to the adverbial root or formative, written according to the word to which is refers: some examples below. | |
| **ADVERBIAL FORMATIVES** | | | |
| advKA | adverbial formative KA- | kakhulu; kade, kancane …etc | |
| | | kahle = kahle\|advKA kahle\|arHLE | |
| | | kabi = kabi\|advKA kabi\|arBI | |
| | | kanje = advKA kanje\|drNJE | |
| advNA | adverbial formative NA- (with/and/have) | unamanga = unamanga\|i2s unamanga\|advNA unamanga\|n6 unamanga\|nrAMANGA | |
| | | nesitsha = nesitsha\|advNA nesitsha\|n7 nesitsha\|nrISITSHA | |
| | | nami = nami\|advNA nami\|pr1s | |
| advNGA | adverbial formative NGA- (by means of) | ngaphezulu | |
| | | ngemoto | |
| advNJENGA | adverbial formative NJENGA - | Njengami | |
| | | njengokuthi | |
| | | ingangendlu | |

165

| ADVERBIAL FORMATIVES *cont*: | | | |
|---|---|---|---|
| advKUNA | | mkhulu kunami | |
| | | lide kunaye | |
| **ADVERBIAL ROOTS:  ( examples)** | | | |
| advLAPHA | adverbial root LAPHA | *lapha / la / lana* --all coded as \|advLAPHA +- \|stab | |
| advKHONA | adverbial root – KHONA | abekho = abekho\|negPRE abekho\|g2 abekho\|advKHONA | |
| | | bakhona = bakhona\|i2 bakhona\|advKHONA | |
| advNKATHI | adverbial root - NKATHI | ngenkathi | |
| advPHELA | adverbial root PHELA | | |
| advEDUZE | adverbial root -duze | eduze\|loc eduze\|advEDUZE | |
| advOUT | adverbial word OUT | out (English); again, English words are coded in as much of the same way as allows | |

| IDEOPHONE *ISENZUKUTHI* | Ideophones are uttered alone, when made into words, these are coded as the words they have made, and no longer as ideophones | | | |
|---|---|---|---|---|
| | **Tag** | **Description** | **Example** | **Comments** |
| | % | ideophone | cwaka; giyo; hlephu | ideophone code is also used for the letter sounds; a ba ca etc, and imitations of animal sounds, such as *Wu, kukhonkotha inja.* |
| | | | | |

| CONJUNCTION *ISIHLANGANISO* | Conjunctions are joining words, also used at the commencement of an utterance. | | | |
|---|---|---|---|---|
| | **Tag** | **Description** | **Example** | **Comments** |
| | cjWORD | conjugation | noma, uma, nokho, kodwa, selokhu, ngakhoke | Conjunctions are coded together with the WORD eg \|cjWORD |

Appendix J

| INTERJECTION *ISIBABAZO* | Interjections are comprised of Vocative, interjection and social or *ubuntu* parts of speech | | | |
|---|---|---|---|---|
| | These words tend to stand alone, or be joined to nouns and pronouns, particularly vocatives and social interjections | | | |
| | **Tag** | **Description** | **Example** | **Comments** |
| | intjWORD | interjection | yebo, cha, maye, ya, ohho, yazi, | In transcription, an agreement had to be reached on how to write utterances of agreement and disagreement, since these are often used in the place of *yebo* and *cha* among young children. Therefore, these are all words expressing 'yes': *yebo, ehhe, mh, ya, ehhena, oo,* These are all words expressing 'no': *cha, mhmh, eheh, ehehna, ayi, hhayi* All are coded individually, for example, *ehhe/intEHHE* |
| | voc | vocative | weCN = weCN\|voc weCN\|CN | |
| | | | yenina = yenina\|voc yenina\|pr2p yenina\|stab | |
| | vocAWE | vocative term | awema = awema\|vocAWE awema\|nrumama | |
| | vocBO | vocative term | bo = bo\|vocBO | |
| | vocCOSI | vocative term | cosi = cosi\|vocCOSI | |
| | vocMANI | vocative term | mani = vocMANI | |
| | vocYEYI | vocative term | yeyi = yeyi\|vocYEYI ; yeynina = yeynina\|vocYEYI yeynina\|pr2p yeynina\|stab | |
| | ubuntuYA | social and greetings | YaMzala = yamzala\|ubuntuYA yamzala\|nrumzala | |
| | ubuntuSHO | social and greetings | sho = ubuntuSHO | |
| | ubuntuSAWUBONA | social and greetings | sawubona = ubuntuSAWUBONA | |
| | ubuntuSANIBONA | social and greetings | ubuntuSANIBONA | |
| | ubuntuSORRY | social and greetings | ubuntuSORRY | |

168

Table J3 below gives an example of tags applied to various roots. These are presented according to the parts of speech as used above.

Table J3

*Examples of Roots and Respective Tags*

| QUESTION stems | SUBSTANTIVE ; Usobizo | |
|---|---|---|
| Interrogative stems | Noun roots | Noun |
| intEH | nrADDIDAS | LOKHUZA |
| intKOMBE | nrAMADONGA | |
| intLANI | nrAMAFU | |
| intLENI | nrAMAHLONI | |
| intMH | nrAMAJITHA | |
| intNGAKI | nrAMANGA | |
| intPHO | nrAMANYALA | |
| intYINI | nrAMANZI | |
| | nrAMASI | |
| | nrAMASIBOMU | |
| | nrAMASIMBA | |
| | nrAMATHE | |
| | nrAMBULANCE | |
| | nrBABY | |
| | nrBATTERY | |
| | nrBIRTHDAY | |
| | nrBOOTS | |
| | nrBOY | |
| | nrBREAK | |
| | nrBUM | |
| | nrCARROT | |
| | nrCHIPS | |
| | nrCHOCOLATE | |
| | nrCLASS | |
| | nrCOMPUTER | |
| | nrCORNER | |
| | nrCRAYON | |
| | nrCRECHE | |
| | nrCROCODILE | |
| | nrCUTEX | |
| | nrDESK | |
| | nrFILE | |
| | nrFISH | |
| | nrGARAGE | |
| | nrGLOW | |

| PREDICATIVE ; Isilandiso | |
| --- | --- |
| Verb root | Verb root |
| vrABSENT | vrGQIBA |
| vrAZI | vrGQOKA |
| vrBABA | vrGUBHA |
| vrBALA | vrGULA |
| vrBALEKA | vrGUNDA |
| vrBAMBA | vrGUQA |
| vrBAMBELELA | vrGUQUBALA |
| vrBANDA | vrGUQULA |
| vrBANGA | vrGWINYA |
| vrBEKA | vrGXOBA |
| vrBESE | vrHALELA |
| vrBHABHAZA | vrHAMBA |
| vrBHAJWA | vrHLALA |
| vrBHALA | vrHLANGANYELA |
| vrBHAXA | vrHLANYA |
| vrBHAYIZA | vrHLEHLA |
| vrBHEDA | vrHLEKA |
| vrBHEKA | vrHLEPHULA |
| vrBHIDLIKA | vrHLIBA |
| vrBHIDLIZA | vrHLOMA |
| vrBHIMBA | vrHLUBA |
| vrBHOSHA | vrHLUKANA |
| vrBHUKUDA | vrHLULA |
| vrBIZA | vrHLULEKA |
| vrBOLA | vrIDLA |
| vrBOLEKA | vrIKHA |
| vrBONA | vrILWA |
| vrBONGA | vrIMA |
| vrBUKA | vrINYA |
| vrBUKELA | vrIPHA |
| vrBULALA | vrISHO |
| vrBUSISA | vrITHI |
| vrBUTHA | vrIWA |
| vrBUYA | vrIZWA |
| vrBUZA | vrJABULA |
| vrCABANGA | vrJIKA |
| vrCAPHUNA | vrJWAYELE |
| vrCEBA | vrKADE |
| vrCELA | vrKALA |
| vrCHAMA | vrKE |

170

| QUALIFICATIVE Isichasiso | | |
|---|---|---|
| Relative | Demonstrative relative | Adjectival root |
| rBANDA | drNGAKA | arBI |
| rBOMVU | drNJALO | arBILI |
| rBUHLUNGU | drNJANI | arBLACK |
| rLUHLAZA | drNJE | arBLUE |
| rMANZI | drNJEYA | arBROWN |
| rMHLOPHE | | arDALA |
| rMNANDI | | arDE |
| rMNYAMA | | arEIGHT |
| rNCANE | | arFISHANE |
| rNGCONO | | arFIVE |
| rNZIMA | | arFOUR |
| rPHI | | arGOOD |
| rPHUZU | | arHLANU |
| | | arHLE |
| | | arHOLY |
| | | arHUNDRED |
| | | arKHULU |
| | | arLATE |
| | | arNCANE |
| | | arNE |
| | | arNINE |
| | | arNINGI |
| | | arNYE |
| | | arONE |
| | | arPINK |
| | | arPLUS |
| | | arPURPLE |
| | | arRIGHT |
| | | arSEVEN |
| | | arSIX |
| | | arSIXTEEN |
| | | arSTRAIGHT |
| | | arTEN |
| | | arTHATHU |
| | | arTHREE |
| | | arTWENTY |
| | | arTWO |
| | | arYELLOW |
| | | arZERO |

171

Appendix J

| DESCRIPTIVE ; Isikhanyiso | | | |
|---|---|---|---|
| Locative | Adverbial formative | Adverb | Ideophone |
| loc | advKA | advEDUZE | %A |
| locNGA | advNA | advEMUVA | %AYEYE |
| | advNGA | advFUTHI | %BA |
| | advNJENGA | advIZOLO | %CA |
| | advNGANGA | advKANYEKANYE | %CHAPHA |
| | advKUNA | advKHONA | %CHASHA |
| | | advKUDALA | %COSI |
| | | advKUDE | %COSOLOLO |
| | | advKUPHELA | %CWAKALALA |
| | | advKUQALA | %DA |
| | | advKUSASA | %DU |
| | | advLAPHA | %DWI |
| | | advLAPHANA | %E |
| | | advLAPHAYA | %ENA |
| | | advLAPHO | %FA |
| | | advLE | %GA |
| | | advMANJE | %GIYO |
| | | advNAMUHLA | %GUNQULUZA |
| | | advNGAPHA | %HA |
| | | advNGEMPELA | %HHAWU |
| | | advNKATHI | %HHESHI |
| | | advOUT | %HLEPHU |
| | | advPHAKATHI | %I |
| | | advPHAMBILI | %ISH |
| | | advPHANDLE | %ISHU |
| | | advPHANSI | %JA |
| | | advPHELA | %KELE |
| | | advPHEZULU | %KULUKULU |
| | | advUP | %LA |
| | | | %MA |
| | | | %MMMMMM |
| | | | %NA |
| | | | %NTSHE |
| | | | %NYAWU |
| | | | %O |
| | | | %PA |
| | | | %PE |
| | | | %QA |
| | | | %QEQEQE |

172

| CONJUNCTION ; Isihlanganiso | INTERJECTION ; Isibabazo | | |
|---|---|---|---|
| Conjunction | Interjection | Vocative | Social |
| cjANGITHI | intjA | voc | ubuntuBYE |
| cjBESE | intjAMEN | vocAWE | |
| cjFUTHI | intjAWEMA | vocBO | ubuntuSAWUBONA |
| cjHHAYI | intjAWU | vocCOSI | ubuntuSHO |
| cjINGANI | intjCHA | vocMANI | ubuntuSORRY |
| cjKANTI | intjCOSI | vocYEYI | ubuntuYA |
| cjKODWA | intjE | | |
| cjKONJE | intjEHEH | | |
| cjNGOBA | intjEHEHNA | | |
| cjNJENGOBA | intjEHHE | | |
| cjNOMA | intjEHHENA | | |
| cjSENGATHI | intjEYI | | |
| cjUMA | intjFUSEGE | | |
| | intjHA | | |
| | intjHHA | | |
| | intjHHAWU | | |
| | intjHHAYI | | |
| | intjHHE | | |
| | intjHHO | | |
| | intjHHOWU | | |
| | intjHOWU | | |
| | intjIYE | | |
| | intjLANI | | |
| | intjMH | | |
| | intjMHMH | | |
| | intjMINA | | |
| | intjO | | |
| | intjOHHE | | |
| | intjOHHO | | |
| | intjOK | | |
| | intjOO | | |
| | intjYA | | |
| | intjYAZI | | |
| | intjYEBO | | |
| | intjYES | | |
| | intjYEYI | | |

## APPENDIX K

## LIST OF CORE FORMATIVES

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **c1s** | Concord first person singular | 1057 | 52.49[6] | 6 | | *Ngicela* | Please |
| **c2s** | Concord second person singular | 809 | 40.17 | 6 | | *Unamanga* | You are lying |
| **CN** | Child name | 561 | 27.86 | 6 | | *WeCN* | Hey CN |
| **c1** | Concord class 1 | 513 | 25.48 | 6 | | *Uphi?* | Where is he/she |
| **c15** | Concord class 15 | 440 | 21.85 | 6 | | *Kugcine* | In the end it.. |
| **pr1s** | Pronoun first person singular | 438 | 21.75 | 6 | I, me, myself | *Mina, nami, okwami* | Me, and me, it's mine |
| **PRES** | Present tense | 425 | 21.11 | 6 | | *Ngiyadlala* | I am playing/joking |
| **z** | Possessive concord | 394 | 19.57 | 6 | | *Okwami* | It's mine |
| **advNA** | Adverbial formative NA | 376 | 18.67 | 6 | | *Unamanga* | You are lying |
| **n1a** | Noun class 1a | 343 | 17.03 | 6 | | *Umisi, uCN* | Miss, CN |
| **c9** | Concord class 9 | 338 | 16.79 | 6 | | *Ini? ingaka* | What? It is this big |
| **VOC** | Vocative formative | 300 | 14.90 | 6 | | *WeCN* | Hey CN |
| **R** | Relative formative | 297 | 14.75 | 6 | | *Okwami* | It's mine |

---

[6] Although now rounded to the 2nd decimal place, during data analysis frequency ranking to six decimal places was used.

174

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **c1p** | Concord first person plural | 296 | 14.70 | 6 | | *Sizokuceba* | We will tell on you |
| **intjYEBO** | Interjection YEBO | 288 | 14.30 | 6 | yes | *Yebo* | Yes |
| **futZO** | Future formative ZO | 287 | 14.25 | 6 | | *Ngizokuceba* | I will tell on you |
| **D** | Demonstrative formative | 281 | 13.95 | 6 | | *Lo (ie lomuntu)* | Him |
| **pr2s** | Pronoun second person singular | 270 | 13.41 | 6 | you | *Wena, nawe, wakho* | You, and you, yours |
| **n5** | Noun prefix class 5 | 257 | 12.76 | 6 | | *Igama* | Name |
| **LOC** | Locative | 247 | 12.27 | 6 | | *Ekhaya* | At home |
| **c5** | Concord class 5 | 237 | 11.77 | 6 | | *Liphi* | Where is it |
| **negPRE** | Negative prefix | 237 | 11.77 | 6 | | *Angifuni* | I don't want |
| **c2** | Concord class 2 | 236 | 11.72 | 6 | | *Bayangena* | They are entering |
| **vrITHI** | Verb root ITHI | 225 | 11.17 | 6 | say; intend; think | *Awuthi* | Say, let me, |
| **n9** | Noun prefix class 9 | 211 | 10.48 | 6 | | *Imali* | Money |
| **n6** | Noun prefix class 6 | 179 | 8.89 | 6 | | *Amanga* | Lies |
| **hort** | Hortative | 174 | 8.64 | 6 | | *As'hambe* | Let's go |
| **n15** | Noun prefix class 15 | 171 | 8.50 | 6 | | *Ukuthi, ukudla* | To , food |
| **advLA** | Adverbial formative LA | 165 | 8.19 | 6 | here | *Woza la* | Come here |
| **n3** | Noun prefix class 3 | 144 | 7.15 | 6 | | *Umsindo* | Noise |
| **exclASP** | Exclusive aspect formative | 137 | 6.80 | 6 | | *Sengiqedile* | I have finished now |
| **pr1** | Pronoun class 1 | 135 | 6.70 | 5 | he; him; himself; she; her; herself | *Yena, naye, lakhe* | He/she, and him/her, his/hers |

175

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| n7 | Noun class prefix class 7 | 132 | 6.56 | 6 | | *Isikhwama* | Bag |
| intEH | Interrogative EH | 128 | 6.36 | 6 | | *Eh?* | Huh? |
| pr9 | Pronoun class 9 | 124 | 6.16 | 6 | it; itself | *Yona, akusiyo* | It, its not it |
| c7 | Concord class 7 | 122 | 6.06 | 6 | | *Siphi* | Where is it |
| c3 | Concord class 3 | 121 | 6.01 | 6 | | *Uphi* | Where is it |
| pr15 | Pronoun class 15 | 120 | 5.96 | 6 | it; itself | *Kona, nakho* | It, and it |
| PRS | Presentative formative | 120 | 5.96 | 6 | here 'they' are/ here 'it' is / here 'he' is etc | *Nakhu* | Here it is |
| vrBUKA | Verb root BUKA | 117 | 5.81 | 6 | look at; watch; admire | *Buka* | Look |
| Sp | Past tense concordial formative | 116 | 5.76 | 6 | | *Wathi* | He/she said |
| vrBONA | Verb root BONA | 114 | 5.66 | 6 | see; understand; give regards | *Uyabona* | Do you see? |
| A | Adjectival formative | 111 | 5.51 | 6 | | *Omunye* | Another |
| vrUKUYA | Verb root UKUYA | 106 | 5.26 | 6 | go to | *Uyaphi* | Where are you going? |
| c10 | Concord class 10 | 105 | 5.21 | 6 | | *Ziyafana* | They are the same |
| vrFUNA | Verb root FUNA | 105 | 5.21 | 6 | search; want; desire | *Ufuna* | You want |
| n10 | Noun prefix class 10 | 101 | 5.02 | 6 | | *Izingodo* | Logs |
| XA | Verb extension ; applied | 99 | 4.92 | 6 | | *Uzongithengela* | She/he will buy for me |
| vrDLALA | Verb root DLALA | 98 | 4.87 | 6 | play; dance; frolic | *Ngiyadlala* | I am playing/joking |
| vrHAMBA | Verb root HAMBA | 97 | 4.82 | 6 | walk; go; travel | *Hamba* | Go (imperative) |

176

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **PLIMP** | Plural imperative | 94 | 4.67 | 6 | | _Asihambeni_ | Let's all go |
| **c6** | Concord class 6 | 89 | 4.42 | 6 | | _Akhona_ | There are some |
| **advNGA** | Adverbial formative NGA | 88 | 4.37 | 6 | | _Ngezandla_ | Using hands |
| **KE** | Suffix KE | 86 | 4.27 | 6 | | _Wozake_ | Come then |
| **negIN** | Negative infix | 82 | 4.07 | 6 | | _Ungamtsheli_ | Don't tell her/him |
| **nrUMISI** | Noun root UMISI | 81 | 4.02 | 6 | lady teacher | _uMisi_ | Miss |
| **rPHI** | Relative root PHI | 81 | 4.02 | 6 | where | _Kuphi_ | Where |
| **advKA** | Adverbial formative KA | 79 | 3.92 | 6 | | _Kanje_ | Like this/ in this way |
| **vrIPHA** | Verb root IPHA | 78 | 3.87 | 6 | give | _Ungiphe_ | You give me |
| **drNJE** | Demonstrative relative NJE | 76 | 3.77 | 6 | such as this; like this | _Nje, Kanje_ | ~just |
| **vrCELA** | Verb root CELA | 76 | 3.77 | 5 | request; negotiate for a wife; be almost | _Ngicela_ | Please |
| **pr3** | Pronoun class 3 | 76 | 3.77 | 4 | it; itself | _Wona, ngawo, akuwona_ | It, with it, it is not it |
| **XC** | Verb extension ; causative | 74 | 3.67 | 6 | | _Buyisa_ | Bring it back |
| **eNI** | Enumerative root NI | 73 | 3.63 | 6 | what; of what sort | _Ini, wenzani_ | What, what are you doing |
| **advKHONA** | Adverb root KHONA | 71 | 3.53 | 6 | of place; here; there | _Khona, akekho_ | There, not there. |
| **intjEHHE** | Interjection EHHE | 70 | 3.48 | 6 | | _Ehhe_ | ~yes |
| **vrBHALA** | Verb root BHALA | 70 | 3.48 | 6 | write; write an examina-tion | _Ukubhala_ | To write |

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **vrCEBA** | Verb root CEBA | 70 | 3.48 | 5 | report someone; invent; plot against; be rich | *Ngizoku<u>ceba</u>* | I will tell on you |
| **c2p** | Concord second person plural | 67 | 3.33 | 6 | | *<u>Ni</u>yamsaba* | I am scared of him |
| **pr5** | Pronoun class 5 | 64 | 3.18 | 5 | it; itself | *<u>Lo</u>na,li<u>lo</u>dwa* | It, there is one |
| **COP** | Copulative formative | 63 | 3.13 | 6 | | *<u>Yi</u>mi* | It is me |
| **vrWOZA** | Verb root WOZA | 60 | 2.98 | 6 | come | *Woza* | Come (imperative) |
| **vrAZI** | Verb root AZI | 58 | 2.88 | 6 | know | *Angikw<u>azi</u>* | I don't know it/how |
| **pr2p** | Pronoun second person plural | 56 | 2.78 | 6 | | *<u>Ni</u>na, ye<u>nina</u>, okw<u>enu</u>* | You, hey you, it is yours |
| **vrSHAYA** | Verb root SHAYA | 56 | 2.78 | 6 | strike; punish; play (as an instrument) | *Ngizoku<u>shaya</u>* | I will hit you |
| **progASP** | Progressive aspect formative | 55 | 2.73 | 6 | | *Ngi<u>sa</u>ya* | I am still going |
| **nrINTO** | Noun root INTO | 54 | 2.68 | 6 | thing; object | *Len<u>to</u>* | This thing |
| **vrFAKA** | Verb root FAKA | 53 | 2.63 | 6 | put in; put on; put around; commencement of udder to fill with milk | *<u>Faka</u>* | Put |
| **nrAMANGA** | Noun root AMANGA | 52 | 2.58 | 6 | lie; untruth | *Una<u>manga</u>* | You are lying |
| **nrISIKHWAMA** | Noun root ISIKHWAMA | 52 | 2.58 | 6 | small bag; pocket; purse; fund | *<u>Isikhwama</u>* | Bag |
| **nrUBANI** | Noun root UBANI | 49 | 2.43 | 6 | who? | *Ubani* | Who |
| **n1** | Noun prefix class 1 | 48 | 2.38 | 6 | | *<u>Umu</u>ntu* | Person |

178

Appendix K

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **vrTHATHA** | Verb root THATHA | 48 | 2.38 | 6 | take; marry | *Thatha* | Take |
| **NN** | Name non-present | 47 | 2.33 | 6 | | *uNN* | NN |
| **cjUMA** | Conjunction UMA | 46 | 2.28 | 6 | if; when | *Uma* | If, when |
| **e** | Enumerative formative | 46 | 2.28 | 6 | | *Muhle* | He/she is pretty |
| **vrENZA** | Verb root ENZA | 46 | 2.28 | 6 | do; make | *Wenzani* | What are you doing |
| **vrZWANI** | Verb root ZWANI | 46 | 2.28 | 6 | hear; listen; taste; smell; feel; sense; live; be alive | *Uyezwa?* | Do you understand/hear/feel etc |
| **dC** | Demonstrative formative position C | 44 | 2.19 | 6 | | *Loya (muntu)* | That (person) |
| **vrIDLA** | Verb root IDLA | 44 | 2.19 | 6 | eat; confiscate; eat into; cost | *Asidleni* | Let's all eat |
| **advPHELA** | Adverb PHELA | 43 | 2.14 | 6 | indeed; truly | *Phela* | |
| **vrHLALA** | Verb root HLALA | 43 | 2.14 | 6 | sit; stay; remain | *Hlala* | Sit (imperative) |
| **nrUMUNTU** | Noun root UMUNTU | 42 | 2.09 | 6 | human being; African; one with human feelings; blunt instrument (as knife) | *Umuntu* | Person |
| **advLAPHA** | Adverb root LAPHA | 42 | 2.09 | 5 | here | *Lapha* | Here |
| **intjMINA** | Interjection MI / MINA | 41 | 2.04 | 6 | here take | *Mina, (thatha)* | ~take this |
| **intYINI** | Interrogative YINI | 41 | 2.04 | 6 | what is it; is it so | *Yini ?* | What is it |
| **vrNGEKE** | Verb root NGEKE | 41 | 2.04 | 6 | never | *Ngeke* | Never |

179

Appendix K

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **pr1p** | Pronoun first person plural | 40 | 1.99 | 6 | we; us | *Thina, sonke, zethu* | Us, all of us, ours |
| **nrIKHAYA** | Noun root IKHAYA | 40 | 1.99 | 6 | home | *Ekhaya* | At home |
| **pr7** | Pronoun class 7 | 39 | 1.94 | 3 | it; itself | *Sona, akusiso,elaso* | (It), it is not it, it belongs to it |
| **vrNIKEZA** | Verb root NIKEZA | 38 | 1.89 | 6 | give; tell off; pass on; hand over | *Nika* | Give (imperative) |
| **vocYEYI** | Vocative YEYI | 37 | 1.84 | 2 | | *Yeyi* | ~oi |
| **vrLETHA** | Verb root LETHA | 36 | 1.79 | 5 | bring | *Letha* | Bring |
| **cjNGOBA** | Conjunction NGOBA | 35 | 1.74 | 6 | because | *Ngoba* | Because |
| **vrISHO** | Verb root ISHO | 33 | 1.64 | 6 | say; mean | *Ngisho* | I mean |
| **vrTSHELA** | Verb root TSHELA | 33 | 1.64 | 6 | tell | *Ngizomtshela* | I will tell him/her |
| **vrUKUBA** | Verb root UKUBA | 33 | 1.64 | 6 | to be; if; in order that; because | *Ukuba* | To be |
| **intjEHEHNA** | Interjection EHEHNA | 32 | 1.59 | 6 | | *Ehehna* | ~no |
| **vrQALA** | Verb root QALA | 32 | 1.59 | 6 | begin; commence; annoy | *Okokuqala* | The first one, first of all |
| **arNE** | Adjective root NE | 32 | 1.59 | 2 | four | *Kune* | There are four |
| **arNYE** | Arjective root NYE | 31 | 1.54 | 6 | some; other | *Omunye* | Another |
| **nrINGANE** | Noun root INGANE | 31 | 1.54 | 6 | child | *Ingane* | Baby/ child |
| **intjMH** | Interjection MH | 31 | 1.54 | 5 | | *Mh* | ~yes |
| **Db** | Demonstrative formative position B | 29 | 1.44 | 6 | | *Lowoke, lokhoke* | That (person), that (thing) |

180

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **vrKADE** | Verb root KADE | 29 | 1.44 | 6 | of action just completed | *Kade ngisadlala* | I was still playing |
| **vrQEDA** | Verb root QEDA | 29 | 1.44 | 6 | finish; complete | *Sengiqedile* | I am finished now |
| **n2a** | Noun prefix class 2a | 28 | 1.39 | 6 | | *Omama, oCN* | Mothers, CN and his fellows |
| **vrYEKA** | Verb root YEKA | 28 | 1.39 | 6 | leave off; stop; let go | *Ongiyeke* | Leave me alone |
| **XP** | Verbal extension passive | 28 | 1.39 | 6 | | *Uhlulwa* | You are defeated by… |
| **intjA** | Interjection A | 28 | 1.39 | 5 | | *A! (yini ke le)* | |
| **vrBEKA** | Verb root BEKA | 28 | 1.39 | 3 | put; place; select pup | *Uyabeka* | You are putting |
| **vrSUKA** | Verb root SUKA | 27 | 1.34 | 5 | go off; commence; originate | *Suka* | Go away |
| **pr6** | Pronoun class 6 | 27 | 1.34 | 4 | it; itself | *Wona, anginawo* | It, I don't have it |
| **pr2** | Pronoun class 2 | 27 | 1.34 | 3 | them; themselves | *(Bona), nabo, yibo* | (Them), and them, it is them |
| **cjANGITHI** | Conjunction ANGITHI | 26 | 1.29 | 6 | | *Angithi* | |
| **vrTHANDA** | Verb root THANDA | 26 | 1.29 | 6 | desire; like; love; be inclined; wind; plait | *Angilithandi* | I don't like it |
| **nrINDLU** | Noun root INDLU | 26 | 1.29 | 5 | house; hut; room; dwelling place | *Indlu* | House |
| **vrPHATHA** | Verb root PHATHA | 26 | 1.29 | 5 | hold; handle; control; administer | *Ngiphethe* | I am holding |
| **vrVELE** | Verb root VELE | 26 | 1.29 | 5 | do merely; do originally | *Vele* | |

181

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **cjKANTI** | Conjunction KANTI | 25 | 1.24 | 6 | just so; in fact; (intj) what of it; what if so; who can stop me | _Kanti_ | |
| **PRSc** | Presentative formative position C | 25 | 1.24 | 6 | | _Nanguya, naliyana_ | There he is, there it is |
| **vrTHOLA** | Verb root THOLA | 25 | 1.24 | 6 | find; get; adopt | _Ngisazoku<u>thola</u>_ | I will still get you |
| **arHLE** | Adjective root HLE | 25 | 1.24 | 5 | good; beautiful; pretty | _Ka<u>hle</u>_ | Do well |
| **vrTHENGA** | Verb root THENGA | 25 | 1.24 | 5 | buy | _Uku<u>thenga</u>_ | To buy |
| **vrKHULUMA** | Verb root KHULUMA | 25 | 1.24 | 4 | talk; speak | _Into yoku<u>khuluma</u>_ | A thing for talking |
| **arKHULU** | Adjective root KHULU | 24 | 1.19 | 5 | large; great | _Om<u>khulu</u>_ | A big (person) |
| **nrIMALI** | Noun root IMALI | 24 | 1.19 | 5 | money | _I<u>mali</u>_ | Money |
| **arNCANE** | Adjective root NCANE | 23 | 1.14 | 6 | small; few; young | _Om<u>ncane</u>_ | A small (person) |
| **vrIMA** | Verb root IMA | 23 | 1.14 | 6 | stand; stop | _<u>Ima</u>_ | Stop |
| **vrPHUMA** | Verb root PHUMA | 23 | 1.14 | 6 | come out; go out; lose colour | _Baya<u>phuma</u>_ | They are exiting |
| **nrCRAYON** | Noun root CRAYON | 23 | 1.14 | 4 | crayon | _U<u>crayoni</u>_ | Crayon |
| **pr10** | Pronoun class 10 | 23 | 1.14 | 3 | it; itself | _(Zona), na<u>zo</u>_ | (It), and it |
| **XN** | Verbal extension Neutral | 22 | 1.09 | 6 | | _Kuyavul<u>eka</u>_ | It is opening |
| **intjO** | Interjection O | 22 | 1.09 | 4 | | _<u>O</u>! (anibuke lona)_ | |
| **nrIMOTO** | Noun root IMOTO | 22 | 1.09 | 4 | motor-car | _I<u>moto</u>_ | Car |

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **nrINDABA** | Noun root INDABA | 21 | 1.04 | 6 | affair; topic for discussion; story | *Anginandaba* | I don't care |
| **vrVULA** | Verb root VULA | 21 | 1.04 | 6 | open; commence | *Vula* | Open |
| **nrUMAMA** | Noun root UMAMA | 21 | 1.04 | 5 | my / our mother | *Umama* | Mother |
| **nrIGAMA** | Noun root IGAMA | 21 | 1.04 | 4 | name; song; fame | *Igama* | Name |
| **nrIFONI** | Noun root IFONI | 21 | 1.04 | 3 | phone | *Kunefoni* | There is a phone |
| **vrCRAYONA** | Verb root CRAYONA | 21 | 1.04 | 3 | 'the act of crayoning' c. colouring | *Ngiyacrayona* | I am crayoning |
| **drNJANI** | Demonstrative relative NJANI | 20 | 0.99 | 6 | of what kind; of what sort | *Kunjani?* | What is it like |
| **vrGIJIMA** | Verb root GIJIMA | 20 | 0.99 | 6 | run; flow | *Gijima* | Run (imperative) |
| **advLAPHO** | Adverb root LAPHO | 20 | 0.99 | 5 | there | *Laphoke* | There |
| **nrISO** | Noun root ISO | 20 | 0.99 | 5 | eye | *Ezamehlo* | They are for the eyes |
| **XR** | Verbal extension; reciprocal | 20 | 0.99 | 5 | | *Uyasabisana* | He is scary |
| **locNGA** | Locative formative NGA | 20 | 0.99 | 4 | | *Ngapha* | This side |
| **nrUMLUNGU** | Noun root UMLUNGU | 20 | 0.99 | 4 | white person; European | *Umlungu* | White person |
| **nrITAFULA** | Noun root ITAFULA | 20 | 0.99 | 2 | table | *Itafula* | Table |
| **REFL** | Reflexive formative | 19 | 0.94 | 6 | | *Ngiyazibhalela,* | I am writing for myself |
| **advFUTHI** | Adverb FUTHI | 19 | 0.94 | 5 | again; once more; perpetually | *Futhi* | Again, and |

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **vrKHIPHA** | Verb root KHIPHA | 19 | 0.94 | 5 | take out; pull out | *Ngiyikhiphile* | I have taken it out |
| **nrUMSIZI** | Noun root UMSIZI | 19 | 0.94 | 4 | helper; pencil | *Umsizi* | Pencil |
| **vrBUYA** | Verb root BUYA | 18 | 0.89 | 6 | return; go back; turn inwards | *Ngiyabuya* | I am coming back |
| **nrUMSINDO** | Noun root UMSINDO | 18 | 0.89 | 5 | noise | *Umsindo* | Noise |
| **vrNGENA** | Verb root NGENA | 18 | 0.89 | 5 | come in | *Bayangena* | They are entering |
| **n2** | Noun prefix class 2 | 17 | 0.84 | 6 | | *Abantu* | People |
| **cjKODWA** | Conjunction KODWA | 17 | 0.84 | 4 | but | *Kodwa* | But |
| **intjEHEH** | Interjection EHEH | 17 | 0.84 | 4 | | *Eheh* | ~no |
| **n11** | Noun prefix class 11 | 17 | 0.84 | 3 | | *Unxantathu* | Triangle |
| **arTHATHU** | Adjective root THATHU | 17 | 0.84 | 2 | three | *Kuthathu* | There are three |
| **RSF** | Relative suffix formative | 16 | 0.79 | 5 | | *Afanayo* | That are the same |
| **advMANJE** | Adverb MANJE | 16 | 0.79 | 4 | now; at the present time | *Manje* | Now |
| **vrGQOKA** | Verb root GQOKA | 16 | 0.79 | 3 | wear; be clothed | *Abaligqoki* | They do not wear it |
| **nrINGWENYA** | Noun root INGWENYA | 16 | 0.79 | 2 | crocodile | *Sengwenya* | Of the crocodile |
| **intjCHA** | Interjection CHA | 15 | 0.74 | 6 | no | *Cha* | No |
| **drNJALO** | Demonstrative relative NJALO | 15 | 0.74 | 5 | like that | *Njalo* | like that/ always |
| **nrCHIPS** | Noun root CHIPS | 15 | 0.74 | 5 | chips | *Amachips* | Chips |
| **nrUKUDLA** | Noun root UKUDLA | 15 | 0.74 | 5 | food; eating | *Ukudla* | Food |

184

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **vrBAMBA** | Verb root BAMBA | 15 | 0.74 | 5 | catch; grasp; hold; overtake; take on (as seedling); delay; surprise one doing mischief | *Ngibambe* | Hold me/ while I held |
| **vrIWA** | Verb root IWA | 15 | 0.74 | 5 | fall | *Liwe* | It fell |
| **vrPHILA** | Verb root PHILA | 15 | 0.74 | 5 | live; be in good health | *Siyaphila* | We are fine |
| **vrVALA** | Verb root VALA | 15 | 0.74 | 5 | shut; close | *Soyalivala* | He/she is closing it now |
| **nrIKHEKHE** | Noun root IKHEKHE | 15 | 0.74 | 4 | cake | *Ikhekhe* | Cake |
| **nrIPHEPHA** | Noun root IPHEPHA | 15 | 0.74 | 4 | paper | *Iphepha* | Paper |
| **nrUGOGO** | Noun root UGOGO | 15 | 0.74 | 4 | grand-mother | *Ugogo* | Grandmother |
| **futYO** | Future tense formative YO | 14 | 0.70 | 5 | | *Bayothatha* | They are going to take |
| **intLANI** | Interrogative formative LANI | 14 | 0.70 | 5 | | *Ukuthelelani?* | Why are you pouring on it |
| **nrISANDLA** | Noun root ISANDLA | 14 | 0.70 | 5 | hand; hand-writing; assistant | *Izandla* | Hands |
| **arONE** | Adjective root ONE | 14 | 0.70 | 4 | one | *One* | One |
| **ubuntuSORRY** | social SORRY | 14 | 0.70 | 4 | sorry | *Sorry* | Sorry |
| **nrTOILET** | Noun root TOILET | 14 | 0.70 | 3 | toilet | *Etoilet* | Toilet |
| **arFOUR** | Adjective root FOUR | 14 | 0.70 | 2 | four | *Four* | Four |
| **vrCHAMA** | Verb root CHAMA | 14 | 0.70 | 2 | urinate | *Ukuchama* | To urinate |
| **vrFANA** | Verb root FANA | 13 | 0.65 | 6 | be like; resemble | *Lifana* | It is like…. |

185

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|-----------|---------------|------------------------|------------------------|-------------------|---------------------------------------------------|---------------------------------------------------------------------------------------------------|----------------------------------------------------------------|
| POT | Potential formative | 13 | 0.65 | 5 | | *Ngingakushaya* | I could hit you |
| vrBIZA | Verb root BIZA | 13 | 0.65 | 5 | call; be expensive; price; be luring | *Bayanibiza* | They are calling you |
| vrCULA | Verb root CULA | 13 | 0.65 | 5 | sing | *Yokucula* | For singing |
| vrFIKA | Verb root FIKA | 13 | 0.65 | 5 | come; arrive; reach | *Ufike* | You arrive… |
| arBI | Adjective root BI | 13 | 0.65 | 4 | ugly; bad; evil | *Kabi* | In a bad manner/ poorly |
| nrCLASS | Noun root CLASS | 13 | 0.65 | 4 | class | *Eklasini* | In class |
| vrBOLEKA | Verb root BOLEKA | 13 | 0.65 | 4 | borrow; lend | *Awungiboleke* | Won't you lend me |
| PN | Place name | 13 | 0.65 | 3 | | *OwasePN* | (Person) from PN |
| advLAPHAYA | Adverb root LAPHAYA | 12 | 0.60 | 5 | yonder | *Laphaya* | Over there |
| intjHHAYI | Interjection HHAYI | 12 | 0.60 | 5 | | *Hhayi* | ~no |
| advNAMUHLA | Adverb NAMUHLA | 12 | 0.60 | 4 | today | *Namuhla* | Today |
| intjOK | Interjection OK | 12 | 0.60 | 4 | ok | *Ok* | Ok |
| vrBHEKA | Verb root BHEKA | 12 | 0.60 | 4 | look; observe; go towards | *Ubheke* | While you look / he is going toward |
| vrSHESHA | Verb root SHESHA | 12 | 0.60 | 4 | make haste; be quick; glide along | *Shesha* | Hurry |
| advLE | Adverbial formative LE | 12 | 0.60 | 3 | far away | *Le, ngalena* | Far away, that side far away |
| advPHANSI | Adverb root PHANSI | 11 | 0.55 | 6 | beneath; below; down; lower end | *Phansi* | Down |
| qrNKE | Quantative root NKE | 11 | 0.55 | 5 | (all) | *Sonke* | All of us |

Appendix K

| Formative | Tag expounded | number of occurrences | composite frequency ‰ | commonality score | Translation of the content formative (dictionary) | Formative in its most frequently used word(s) from composite list (relevant formative underlined) | Translation of the frequently used word(s) from composite list |
|---|---|---|---|---|---|---|---|
| **arFIVE** | Adjective root FIVE | 11 | 0.55 | 4 | five | *Five* | Five |
| **intjHHAWU** | Interjection HHAWU | 11 | 0.55 | 4 | | *Hhawu* | |
| **Rempast** | Remote past concordial formative | 11 | 0.55 | 4 | | *Ngangihamba* | I was going |
| **vocMANI** | Vocative MANI | 11 | 0.55 | 4 | | *Mani* | |
| **arBILI** | Adjective root BILI | 11 | 0.55 | 3 | two | *Mabili* | There are two |
| **vrSABA** | Verb root SABA | 11 | 0.55 | 3 | fear; be afraid | *Uyasabisana* | He is scary |
| **vrWASHA** | Verb root WASHA | 11 | 0.55 | 3 | do laundry work | *Siyawasha* | We do the laundry |
| **vrYEKELA** | Verb root YEKELA | 11 | 0.55 | 3 | leave off; let alone | *Muyekele* | Leave him |
| **KUNYE_** **KUBILI_** **KUTHATHU_** **KUNE** | Miscellaneous String | 12 | 0.60 | 2 | | | one_two_three_four |