Determining the core vocabulary used by Sepedi-speaking preschool children during regular preschool-based activities

by

Ngwanamashiane Rahab Blondie Mothapo

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: Prof Kerstin Tönsing

CO-SUPERVISOR: Miss Refilwe Morwane

April 2019

The financial assistance of the National Research Foundation (NRF) in the form of a grant holder linked bursary (#TTK150617119597) towards the research project is hereby acknowledged. Opinions and conclusions expressed in this report arrived at are those of the author and are not to be attributed to the NRF.

UNIVERSITY OF PRETORIA

DECLARATION OF ORIGINALITY

This document must be signed and submitted with each

essay, project, dissertation and or thesis.

Full names of student: Ngwanamashiane Rahab Blondie Mothapo

Student number: 12301656

Declaration

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

2. I declare that this project 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

Firstly, I would like to express my sincerest gratitude to *God* for presenting this dream to me at the beginning, carrying me through this journey and bringing me to the final submission. I have seen Him carrying out his promise to prosper me and never to disappoint me. I would like to acknowledge the following people as well:

My amazing supervisor, *Prof Kerstin Tönsing*. You have played a pivotal part in this process. Thank you for your tireless efforts, words of encouragement and academic support. This paper is both my pride and yours. I also acknowledge my co-supervisor, *Miss Refilwe Morwane,* for all her creative ideas and suggestions.

The two people who conceived me with love, my parents, *Simon* and *Inger Mothapo*. Thank you for raising me to be the woman I am. Thank you for the amazing upbringing and for accepting me for who I am. Thank you for the unwavering support you've given me from the start of this journey.

My three sisters, *Lerato, Mosa* and *Victoria Mothapo*. Thank you for being my role models, for supporting me, loving me unconditionally and for being present. You mean the world to me.

A word of appreciation goes to the father of my unborn children, the man who stood by me throughout this journey, *Chibuike (Ike) Okorie*. Thank you for continuously supporting my dreams and visions. Thank you for staying up late with me, compromising so much for my happiness, willingly giving me insightful inputs and always reminding me of my strength in my times of weakness. All your love is sincerely appreciated.

Thank you to my colleague and friend, *Divhanani Sithi*, for her willingness to assist me throughout all the stages of my research, for her patience with me through all the ups and downs. Thank you for consistently being there to listen and to calm the storms. I pray that God blesses you beyond measure.

A word of gratitude also goes to my funders, my language editor *(Barbara Bradley)*, research assistants, the principals of the schools who welcomed me into their premises with warmth and the parents of the children who took part in the study. I pray that our nation can be as supportive as you were to me. I strongly believe that our ability to generate research heavily relies on people as flexible and supportive as you were. You did not know me prior to this study, but you clearly saw my vision and supported it and for that I say, Ke a leboga.

This dissertation is dedicated to all AAC users and developers who tirelessly continue to fight for the voice of AAC users. Let us not get weary. And lastly to all researchers in the field of South African languages, let us strive for research that is context-appropriate.

You took common courtesy to an uncommon level. I am so humbled and grateful.

ABSTRACT

Augmentative and alternative communication (AAC) systems can benefit persons whose speech is too limited to meet their communication needs. Some AAC systems require words and phrases to be preselected and represented on the system, to allow the person using the system to select them for expression. In order to be usable, such AAC systems require the selection of a limited set of words from possibilities.

Core vocabulary consists of words from a variety of word classes that appear most frequently in natural conversations. It has been suggested that the inclusion of these words on the AAC system can give people using the system access to the production of many different phrases and sentences, thereby enhancing their expressive power. Since core vocabulary is language-specific, the study's aim was to determine the core vocabulary of Sepedi-speaking preschoolers, in order to inform the vocabulary selection for AAC systems for children in need of AAC from a Sepedi language background.

The speech of six preschool children without disabilities was recorded using small body-worn audio recording devices. Children were recorded during their regular preschool day. The recordings were transcribed, coded and analysed. The composite transcript consisted of 17 579 words, of which 1 023 were different words. The core vocabulary was determined by identifying all words that were used with a minimal frequency of 0.5‰, and that were used by at least half of the participants. The Sepedi core vocabulary consisted of 226 words that accounted for 88.1% of the composite sample.

As in other studies, the core vocabulary consisted of a smaller number of words that represent a large proportion of the composite sample. The core word list determined in this study can be used as one resource among others to assist with vocabulary selection for children from a Sepedi language background who may require AAC.

Keywords: Sepedi, Augmentative and Alternative Communication (AAC), Core Vocabulary, Preschool Children, Vocabulary Selection.

KAKARETŠO

Disistimi tša *Augmentative and Alternative Communication* (AAC) di thuša batho bao ba palelago ke go ntšha di hlokwa tša bona ka polelo. Disistimi tša go thuša batho go ipolelela di hloka gore mafoko le dika di kgethwe ebile di bonagatšwe mo go sistimi yeo gore motho yo a yago go šomiša sistimi yeo a kgona go di kgetha ge a bolela. Gore di sistimi tše di šome, go hlokega gore go kgethwe mantšu a ma nnyane gare ga mantšu a dikete ao a bego gona. Mantšu a tlotlontšo ya motheo *(core vocabulary)* ke mantšu ao a tšwago mafapheng a go fapafapana a polelo ebile ke ona a tšwelelago kudu ge motho a bolela. Go akanywa gore go tsentšhwa ga mantšu ao mo go disistimi tše, go ka fa motho yo a šomišago sistimi yeo mokgwa wa go ntšha mafoko le mantšu a go fapana, ka tsela yeo ele go mo fa maatla a go ikemela.

Mantšu a tlotlontšu ya motheo a ya ka polelo ye nngwe le ye nngwe. Nyakišišo ye, e duma go ka laetša mantšu a tlotlontšu ya motheo a bana bao ba šešogo ba tsena sekolo bao ba bolelago Sepedi. Seo se tla thuša ka kgetho ya mantšu a go ya mo go sistimi ya bana ba go hloka AAC ba tšwago magaeng ao a šomišago Sepedi. Polelo ya bana ba tshela ba mphato wa R ba go hloka bogole, e rekhodilwe ka di rekhoda tše di nnyane, tša go aparwa mo mmeleng. Bana bao ba rekhodilwe gare ga matšatši a sekolo ao a tlwaelegilego a sekolo. Ditaba tšeo di rekhodilwego di ile tša ngwalwa fase, tša fiwa dikhoudi tša ba tša tsitsinkelwa. Sampolo ye kgolo ya bana ka moka e ile ya ba le mantšu a 17 579, le mantšu a go fapana a 1 023. Mantšu Mantšu a tlotlontšu ya motheo a khethilwe ka go lebelela mantšu ka moka ao a bilego le frikwensi ya bo nnyane bja 0.5% ebile a šomišitšwego ke bana ba bararo go ya godimo.

Mantšu a tlotlontšu ya motheo a Sepedi a be a le 226 a lekana 88.1% ya sampolo ye kgolo. Bjalo ka di nyakišišo tše dingwe, Mantšu a tlotlontšu ya motheo a dirwa ke nomoro e nnyane ya mantšu ao a emetšego karolo e kgolo ya sampolo e kgolo ya mantsu. Mantšu a tlotlontšu ya motheo ao a tšweletšago ke nyakišišo ye, a tla šomišwa bjalo ka mokgwa o tee wa go thuša go kgethela bana bao ba šomišago distimi tša AAC ba go tšwa magaeng a go šomiša Sepedi.

Mantšu a bohlokwa: *Augmentative and Alternative Communication* (AAC), kgetho ya mantšu, Mantšu a tlotlontšu ya motheo, Sepedi le bana ba mphato wa R.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

PROBLEM STATEMENT

1.1 Introduction

This chapter hopes to provide an introduction to the subject matter and an overview of the structure of the study. The chapter commences with an outline of the problem statement and rationale of the study. The terminology and abbreviations found within the study are also discussed. Lastly, a brief summary of each chapter of the study is provided.

1.2 The problem statement and rationale

Individuals whose speech is too limited or not sufficiently functional to meet their communication needs may benefit from augmentative and alternative communication (AAC) to supplement or substitute their limited or absent natural speech. AAC consists of the use of symbols, strategies and techniques to meet communication requirements (Tönsing, Alant, & Lloyd, 2005). Various authors have suggested that direct relationships exist between aspects of language development and the use of AAC (Gerber & Kraat, 1992 ; Renner, 2003; Trudeau, Sutton, Dagenais, De Broek & Morford, 2007). In the case of children who have a good understanding of spoken language, an alignment between the spoken language and the AAC system would therefore seem desirable.

Over 90% of the population in South Africa is bi- or multilingual (du Plessis, 2006). This refers to the use of two or several languages (Robillard, Mayer-Crittenden, Minor-Corriveau, & Belanger, 2014). According to the 2011 census, English is the first language of only 9.6% of the South African population (Statistics South Africa, 2012). It should, however, be noted that most AAC systems are English-based. Although English is acquired as a second language by many children and is the language of instruction in many schools (Department of Education, 2010), experts in language development have outlined that children need to be proficient in their first language within the first six years of life in order to acquire a second language successfully (Kuhl, 2010). It follows that children in need of AAC should have access to AAC systems that encourage proficiency in their home language.

Children who are not yet literate often make use of AAC systems that use graphic symbols to represent words and phrases. Such systems require that these words and phrases

are preselected and represented on the AAC system so that the children using them can select them to express themselves. To be usable, such AAC systems require the selection of a limited number of words from many possibilities (Boenisch & Soto, 2015). This can be difficult and time-consuming and children who use AAC often cannot actively participate in the selection of these words owing to their limitations in communication.

Educators, parents and therapists have raised concerns about selecting appropriate vocabulary that caters for the communication needs of children in need of AAC. Several authors (Fallon, Light, & Paige, 2001; Fried-Oken & More, 1992) have suggested that for a comprehensive and appropriate vocabulary to be selected, multiple sources should be consulted. These include parent/caregiver interviews, making use of published vocabulary lists (Banajee, Dicarlo, & Stricklin, 2003; Beukelman,Yorkston, Poblete, & Naranjoo, 1984) and to vocabulary studies of children without disabilities (Beukelman, Jones, & Rowan, 1989; Burroughs, 1957).

It has been suggested that the inclusion of core vocabulary words on the AAC system can give people using the system access to the production of many different phrases and sentences, thereby enhancing their expressive power. Core vocabulary comprises of words from a variety of word classes and includes the words that appear most frequently in natural conversations (Witkowski & Baker, 2012). Research has been done on the core vocabulary in various languages, for example, for English- and French-speaking school-aged children in countries other than South Africa (Trembath, Balandin, & Togher, 2007; Robillard et al., 2014). Although some similarities may be found between the French and English core vocabulary lists, there are also differences. These are to be expected, since core vocabulary typically consists mainly of function words that relate to the grammar and syntax of the language. Since languages differ in their grammatical structure, it would follow that it is unlikely that the translation of core vocabulary found in one language would be a true reflection of the core vocabulary in another language.

It is therefore important that language-specific core vocabulary studies are conducted. To date, a study on the core vocabulary of isiZulu preschoolers in South Africa has been conducted (Mngomezulu, 2017). The orthographic conventions of isiZulu (a Nguni language that has a conjunctive rather than disjunctive orthography), as well as the linguistic structure of the language (primarily synthetic agglutinating) necessitated analysis on a morphological

level, rather than using orthographic words as units of analysis. These factors contributed to the limited overlap between English core vocabulary lists determined in various other studies and the isiZulu core vocabulary list, and once again underlined the need for language-specific studies. Somewhat more overlap with English was found in a study conducted on Afrikaans-speaking preschoolers (Hattingh, 2018), which may be explained by the fact that both English and Afrikaans are West Germanic languages. To date no studies have been conducted on core vocabulary in other South African languages, specifically no studies in any language belonging to the Sesotho language group, such as Sepedi.

The aim of this study is therefore to establish a core vocabulary list in Sepedi by determining the most frequently and commonly used words spoken by Sepedi-speaking children without disabilities in Lebowakgomo, Limpopo. This list could be one resource that could guide vocabulary selection for Sepedi-speaking preschoolers who require AAC. The study also has the potential to contribute to research on core vocabularies and their use in the design of AAC systems.

1.3 Terminology

The terminology most frequently used and deemed important in the study will be defined in order to scaffold readers' understanding of specific concepts used. The terms are presented in alphabetical order.

1.3.1 Augmentative and alternative communication

AAC is introduced to supplement or substitute natural speech. AAC comprises the use of symbols, strategies and techniques to meet communication requirements (Tönsing, Alant, & Lloyd, 2005). Such symbols, strategies and techniques can include line drawings displayed on communication boards and books, speech generating devices, objects, gestures and manual signs – these can all be used to assist a person whose speech is limited to meet their expressive communication needs (American Speech-Language-Hearing Association, 2019).

1.3.2 Children in need of augmentative and alternative communication

Children in need of AAC are children who have difficulty or an inability to produce natural speech (von Tetzchner & Basil, 2011). This usually leads to activity limitations and participation restrictions (Bornman & Donohue, 2013).

1.3.3 Code switching

Code switching refers to the process of alternating between two or more languages in the same context or a single conversation. It is generally observed when words from another language are inserted in a sentence. Inter-sentential switching refers to a type of code switching that occurs at phrasal, sentence, or discourse level, whereas intra-sentential switching refers to code-switching within a sentence level by means of adding lexical units of another language into a sentence that is spoken in the first language (Zirker, 2007). This is not to be confused with loan words (see Section 1.3.15).

1.3.4 Commonality score

The number of participants who used a particular word in the total sample is referred to as the commonality score. The study involved six participants; therefore, the highest commonality score attainable was 6. Any word with a commonality score of 6 was used by all the participants involved at least once.

1.3.5 Content words

Content words are words that carry meaning. Word such as nouns, verbs, adverbs and adjectives are considered to be content words (Trembath et al., 2007). These words can often be used in isolation for labelling and carry meaning on their own. However, content words on their own are typically inadequate in conveying more complex messages, as syntactical constructions require the use of structure words (Sutton, Soto, & Blockberger, 2002).

1.3.6 Core vocabulary

In the field of AAC, core vocabulary has come to be defined as the words that appear most frequently and commonly in natural conversations (Witkowski & Baker, 2012). Various core vocabulary studies conducted in different languages (van Tilborg & Deckers, 2016) have found that core vocabularies typically consist of a limited number of words (when compared to the total number of different words in a spoken sample) that cover a large proportion of the conversational samples. These words also remain relatively constant across different communication environments (Beukelman & Mirenda, 2013; Boenisch & Soto, 2015). Core vocabulary in this study was defined according to its frequency and commonality. Words that occurred with a minimal frequency of 0.5‰ and a commonality score of at least 3 in the composite sample were considered core words.

1.3.7 Frequency per mille (‰)

Frequency per mille (‰) is calculated by dividing the total number of word occurrences of a particular word by the total number of words in the sample obtained and multiplying by 1000. Frequency per mille refers to the measurement used to calculate the frequency of occurrence of a certain word out of 1000.

1.3.8 Fringe vocabulary

These are words and messages that are particular to the individual and context and that tend to occur with lower frequency and commonality in spoken samples (Banajee et al., 2003; Beukelman & Mirenda, 2013; Trembath et al., 2007). In the current study, all words that occurred with a frequency of less than 0.5 per 1000 words (0.5‰) and/or had a commonality score of less than 3 were designated as fringe words.

1.3.9 Grade R

Grade R (also called 'reception') is a part of the foundation phase of basic education in South Africa. It is the year before children start their first formal year of schooling. Although this year of schooling is not compulsory, there is a national Grade R curriculum contained in the Curriculum and Assessment Policy Statement (Atmore, van Niekerk, & Ashley-Cooper, 2012).

1.3.10 Graphic symbols

These are symbols that are static in nature and can be represented in the form of pictures consisting of line drawings and pictorial representations. There are numerous commercially and freely available symbol collections/libraries that are used in the field of AAC to encode messages. These graphic symbol collections share some similarities with other symbolic systems such as speech and writing, but most of them lack some pertinent characteristics of language, such as arbitrariness and duality of patterning (Smith, 2006).

1.3.11 Grammatical variation/inflection form

This refers to the process of identifying the word root part of a word and the inflected aspect (which can generally be replaced by another morpheme to form another word) (Quirk, Greenbaum, Leech, & Svartoik, 1985). An example of a grammatical variation is where morphemes are used to indicate the past tense form of a verb (*ngwala,* which means write in

the present tense, to *ngwadile*, which means wrote in the past tense). It usually does not cause the word to change its word class according to the parts of speech (Payne, 1997).

1.3.12 Heteronyms

Heteronyms are words that are written in the same way but are pronounced differently and have different meanings. Sepedi contains various heteronyms, for example the word *ke* may be used in two different sentences. For example in the sentence *Molato ke eng?* (What is the problem?), the word *ke* is used as a copulative particle. In contrast, in the sentence *Ke sa sepela* (I am leaving) the word *ke* is used as a subject concord first person singular (with a pronomial function).

1.3.13 Home language

Home language in this study refers to the language to which the participant is most often exposed at home. It is a language spoken mostly by the family members during everyday interactions at home.

1.3.14 Lemma

The lemma is the dictionary or citation form of a word. In English, for example, the lemma of *run, runs, ran* and *running* is *run,* since this is the form that would be found as the main word in the dictionary. The lemma is often the most uninflected form of the word, but this does not mean it necessarily consist of only one morpheme. In Sepedi, for example, the lemmas of verbs are the positive (affirmative) form of imperfect (present) tense verbs, and these consist of a root and a suffix *-a*. For example, the verb *bala* consist of *bal-* (root) and *-a* (suffix).

1.3.15 Loan words/lexical borrowing

A loan word is a word that has been taken from one language and incorporated into another language. In the process, the phonological and morphological form of the word are typically modified to make it part of the target language (Mojela, 2010). Many words in the English language, for example, were originally borrowed from Latin (for example, ‘emphasise’) but are now accepted parts of the language. The Sepedi language similarly contains words originally borrowed from other languages such as Afrikaans, English and isiZulu. An example is *lelekere* (sweet) borrowed from the Afrikaans ‘lekker’.

1.3.16 Parts of speech

These are also referred to as 'word classes'. Parts of speech refer to a set of word groups that are used to classify words according to their syntactical functioning in the language. The words are classified according to their purpose in language (Croft, 2000). Such word groups include verbs, nouns, concords and conjunctions. Grouping of words into parts of speech is a grammatical classification (Mojapelo, 2007).

1.3.17 Polysemous words

Polysemous words are words that have related, yet slightly different meanings, and in some cases polysemes belong to different parts of speech (Faaß, Heid, Taljard, & Prinsloo, 2009; Noruzi, 2006). Many function words in Sepedi are polysemous, for example *sa* may be used as concord or aspectual prefix.

1.3.18 Root/root word

Roots are the central morphemes of words that carry the main lexical meaning of the word. Roots consist of only one morpheme and cannot be divided into smaller meaningful units (Howard, 2003). In Sepedi, for example, adjectives consist of a class prefix and an adjectival root. For example, *mogolo* (one inflection of the adjective 'large' or 'big') consists of the class prefix *mo* and the adjectival root *golo*. Note that *golo* is not a word on its own, but a bound morpheme.

1.3.19 Structure words

Structure words are words that generally assist verbs and nouns. In English, these words include prepositions, articles, and pronouns (Banajee et al., 2003). In Sepedi, many of the disjunctively written bound morphemes are function words, for example concords and particles (Faaß et al., 2009). Structure words typically contribute to the grammatical correctness of sentences. These words generally provide syntactic purpose while lacking semantic meaning on their own.

1.3.20 System for Analysing Language Transcripts

The System for Analysing Language Transcripts (SALT) is a software program designed to assist in language sample analysis and it automatically generates descriptive statistics and counts based on the transcripts provided. SALT can be used for purposes such

as obtaining word counts, determining the total number of words (TNW), the number of different words (NDW), as well as the type-token ratio (TTR) (Miller & Iglesias, 2012).

1.3.21 Type token ratio

The TTR is calculated by considering the NDW divided by the TNW (Kettunen, 2014).

1.4 Notations

Linguistic examples in Sepedi are provided in italics in the current paper. English translations of these examples are given in parentheses or in single quotation marks.

1.5 Abbreviations

AAC: Augmentative and alternative communication
CN: Child's name
CS: Code switch
NDW: Number of different words
PCS: Picture communication symbols
PN: Place name (substituted for a proper name describing a location in the speech sample)
SALT: Systematic Analysis of Language Transcripts
TN: Teacher's name
TNW: Total number of words
TTR: Type token ratio

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. An outline of the contents of the chapters is included.

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. Furthermore, the South African linguistic context is described. The Sepedi language is described in terms of its linguistic structure and the different types of parts of speech that exist in the language.

Chapter 3 deals with 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 included. The aims, procedures and results of the pilot study are given, and implications for the main study are outlined. The procedures for data collection and analysis are described, including ethical considerations and considerations regarding reliability and validity.

In Chapter 4, the results of the study are presented in graphs and tables in accordance with the four sub-aims of the study. First, the complete speech sample obtained from preschoolers who speak Sepedi is described. Second, the words most frequently and commonly used by Sepedi-speaking children without disabilities during regular preschool activities (core vocabulary) are identified. Third, the core vocabulary is described by part-of-speech categories as well as by differentiating structure (grammatical) and the content (lexical) vocabulary is determined. Last, the core vocabulary identified is compared to that found in other studies.

In Chapter 5 the results of the study are discussed in the light of previous literature. The parameters of the language sample collected are explored and compared to other studies. Similarly, the parameters and characteristics of the core vocabulary identified in this study are compared to those of previous studies. The lexical meanings of the core words identified are compared with English core vocabulary lists and the isiZulu list.

Chapter 6 provides an overall summary of the study. The study is critically evaluated and the implications that a morphologically rich language such as Sepedi has for the selection of vocabulary and design for AAC systems are discussed. Recommendations for future research are also provided.

1.7 Summary

This chapter orientated the reader to the study. Having outlined the purpose of the study in the problem statement, the terminology and abbreviations that are used in the study were described. Each of the upcoming chapters was then briefly introduced.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter will provide background to the study by discussing the field of AAC, with specific focus on graphic symbol-based AAC systems, vocabulary selection methods, core and fringe vocabulary and the limitations of translating core vocabulary. The chapter will also provide an overview of the Sepedi language and its structure, including the implications this structure has for a core vocabulary study. The different parts of speech of the language will be described.

2.2 Augmentative and alternative communication

Children who are unable to meet their daily communication needs through speech require communication support. AAC is implemented to either supplement speech (when some intelligible speech is present) or to become the main method of communication (when speech is absent (Beukelman & Mirenda, 2013). The primary purpose of any AAC intervention is to compensate for impairments in speech language production and comprehension (Sigafoos, Ganz, O'Reilly, Lancioni, & Schlosser, 2007). In the case of individuals whose speech will remain limited, AAC techniques should result in generative, functional communication in all communication contexts with a variety of communication partners (Mirenda, 2003).

AAC is defined as any device, system or method that can be used to supplement or replace the speech of an individual, and assist him/her to communicate effectively (Beukelman & Mirenda, 2013). AAC is introduced when a child experiences severe delays in the development of speech and communication. Indicators such as a moderate to severe expressive speech/language disorder, minimal improvement in expressive language with therapy and/or the individual exhibiting frustration due to his/her inability to communicate messages effectively may be signs that some form of AAC should be introduced (Sigafoos & Drasgow, 2001).

AAC systems can be categorised as aided or unaided systems. Unaided systems do not require methods that are external to the body. These usually involve the use of symbols

such as manual signs, pantomime, gestures, motor behaviour and eye gaze (Mirenda, 2003). Aided systems consist of devices and aids that are external to the body of the individuals who use them (e.g. communication books, picture exchange communication systems, and SGDs).

These are categorised into low and high technology. Low technology aided systems require little or no technology and usually do not need a power source. Light technology aided systems are usually SGDs that are battery-operated and have a static (non-changing) display (Wilkinson & Hennig, 2007). Examples of these are the Big Mack[1] and Go-Talk.[2] High technology devices are systems generally requiring a rechargeable battery source and having a dynamic (i.e. computerised, changing screen) display, which use highly advanced technology. These include software programs or applications that can be loaded onto mainstream hardware, for example an Apple iPad with an AAC app such as Proloquo2Go[3] or GoTalk Now[4]. Customised off-the-shelf solutions (hardware with software loaded) are also available from some companies, for example, the Indi 7[5] and the Nova Chat 5[6] (van der Sandt-Koenderman, Wiegers, & Hardy, 2005).

Most people use a combination of communication techniques, strategies, and modalities, depending on factors such as the context and preference of the communication partner (Beukelman & Mirenda, 2013).

On aided AAC systems, meaning can be represented through a variety of symbols, such as traditional orthography, photographs or graphic symbols. The linguistic potential of these symbols varies. Alphabetic orthographies, for example, display the same dual structure as spoken language. This means that a limited number of meaningless elements (i.e., the letters of the alphabet) can be combined in various ways to give access to the expression of an unlimited number of meanings. However, other types of symbols used on aided AAC

[1] BigMack is a product of Inclusive Technology, Riverside Court, Huddersfield Road, Delph. Oldham. http://www.inclusive.co.uk.
[2] Go-Talk is a product of Spectronics Inclusive Learning Technologies, Unit E1 Commercial Court 130 Kingston Road Underwood, and Australia. spectronics.com.au https://www.attainmentcompany.com/gotalk-now
[3] Proloquo2Go is a product of Assistive Ware, Laurierstraat 193, and 1016 PL Amsterdam, Netherlands. https://www.assistiveware.com
[4] GoTalk Now is a product of Widgit Software, 26 Queen St, Cubbington Leamington Spa Warwickshire, UK. https://www.widgit.com/products/third_party/gotalknow/index.htm
[5] The Indi 7 is a product of Tobii dynavox, https://www.tobiidynavox.com/en-us/devices/multi-access-devices/indi-7-snap-communicator-italian
[6] The Nova Chat 5 is a product of Saltillo, 2143 Township Road 112 Millersburg, OH 44654-9410 USA https://saltillo.com/products/print/nova-chat-5

systems may not allow the people using them to generate their own novel messages to the same extent (Smith & Grove, 2003). Constructing such types of AAC systems usually entails the preselection of vocabulary that is then made available to the person using the system.

2.3 Graphic symbol-based augmentative and alternative communication systems

The old adage, 'a picture is worth a thousand words', expresses the view that pictures can be used to convey meaning (Patel, Schooley, & Wilner, 2007). The use of formalised collections of picture-like symbols or line drawings (commonly known as graphic symbols) as a method to represent meaning on aided AAC systems has been a common practice in the field of AAC for a few decades (Lloyd, Quist, & Windsor, 1990). Graphic symbols are particularly suitable for individuals who are not conventionally literate (e.g. for young children who are not yet able to use the alphabet to spell) (Trudeau et al., 2007).

Graphic symbols are used as a method to represent meaning on an AAC system. There are three methods in which graphic symbols are used to represent language. These are the use of single-meaning pictures (including graphic symbols), the use of pictures that are composed of a number of semantic sub-elements, and semantic compaction (Hill & Romich, 2002). Semantic compaction or iconic encoding is based on the polysemy that is typical of pictures – that is, one picture can have multiple meanings or evoke multiple associations. This approach therefore uses variable sequences of a limited set of graphic symbols (also called 'icons') to represent different meanings. This is an added advantage regarding the generativity of varying utterances in communication, but may place high learning demands on the person using the system (Light, Drager, McCarthy, Mellot, Millar, Parrish, & Wellevier, 2004).

Blissymbols include symbols that consist of a number of semantic elements and thus have the potential for segmentation and exist within a rule-governed system. The majority of Blissymbols are composed of 120 key symbols. These key symbols are combined in different ways to make new symbols, thereby enabling any concept to be repreesnted. However, unlike most orthographies, the key symbols or elements composing Blissymbols have a semantic meaning and are predominantly non-alphabetic (Smith, 2006).

Single-meaning symbols cannot be subdivided and only represent a single concept (Smith, 2006). Most commercially available graphic symbol collections (e.g., Picture

Communication Symbols, Widgit symbols, and SymbolStix) consist predominantly of single-meaning symbols, where one symbol represents one concept – in English this usually translates to one orthographic word. Such single-meaning graphic symbols typically incorporate a degree of visual resemblance between referent and symbol if possible; for example, the graphic symbol for 'car' depicts the visual features of a car. More abstract referents cannot be pictured by visual resemblance, and the nature of these symbols is typically either translucent (when the meaning of the symbol is not guessable but a logical link between symbol and referent becomes apparent once the relationship is explained) or opaque (when the relationship between a symbol and a referent is not easily guessed and only determined when the referent is known) (Fuller & Lloyd, 1991).

Most commercially available graphic symbol libraries consist of a large vocabulary. For example, the Picture Communication Symbols classic collection consist of 4 500 symbols (Mayer-Johnson, 2018) and includes symbols representing both content words (i.e., words carrying semantic meaning, such as nouns and verbs) as well as structure words (those carrying less semantic meaning but required to make sentences, e.g., prepositions) (Shi, Werker & Cutler, 2006). Although many of these systems typically cannot fully mirror the grammar of a spoken language, symbols may be combined to create 'sentences,' thereby increasing the expressive power of the person using them beyond the use of single words.

When single-meaning graphic symbols are used in an AAC system, each symbol must occupy a unique location in the system. In order to use the symbol, one must memorise the location of the symbol and navigate to it. The bigger the symbol vocabulary, the bigger the memory demands on the person using the system (Light & Lindsay, 1991). For this reason, it is typically necessary to select only a limited number of symbols for inclusion in the system, in order to prevent memory overload (Light & Lindsay, 1991). At the same time, the vocabulary should be selected in order to convey essential messages and to allow for the development of language skills (Beukelman & Mirenda, 2013).

2.4 Vocabulary selection methods

Practitioners are faced with difficult decisions when vocabulary needs to be preselected and included in AAC systems. Robillard and colleagues (2014) point out that speaking 5 year children have an expressive vocabulary of an estimated 2 100 words and 2 600 words by 6 years. This vocabulary is also diverse and contains different parts of

speech. When preliterate children require aided forms of AAC, their caregiver or AAC interventionist typically chooses, programmes and organises words for them (Robillard et al., 2014; Trembath et al., 2007). Vocabulary selection is still seen as difficult process, as limited guidelines for selection have been documented in literature (Fallon et al., 2001).

There are many factors to be considered in this process, including the child's preference, the setting in which the system is to be used, the child's cognitive abilities, ease of use and communication functions supported by the system. This process is further complicated by the heterogeneity of children who can benefit from AAC devices (Thistle & Wilkinson, 2015). Various methods of vocabulary selection have been documented in the literature.

Environmental inventories have been shown to provide important information on the person's vocabulary needs and his/her context. An inventory is compiled by assessing the environments and activities relevant to the individual. The words and phrases selected are those that are usually required in those specific environments and activities (Beukelman & Mirenda, 2013).

Vocabulary selection can also be done through the use of informants (Trembath, Balandin, & Dark, 2006). Informants are generally people who know the person in need of AAC well. They can suggest vocabulary through the blank page method by writing down lexical items that they feel should be included in an AAC system on a blank page (Fried-Oken & More, 1992). Informants may also select vocabulary they deem appropriate from a variety of category-based words (Fallon et al., 2001). Informants may be caregivers, teachers or therapists who interact with the person in need of AAC (Witkowski & Baker, 2012). Other methods include open-ended surveys, which are also used to select vocabulary for an AAC system. People who are surveyed are often asked to list words that are frequently used. Surveys should not be used in isolation, as respondents typically provide content words (e.g., *ball, swing, bread*) rather than structure words (such as *you, and, to*) even though structure words are used more frequently in everyday language (Beukelman & Mirenda, 2013; Robillard et al., 2014). Bornman and Bryen (2013) report other methods, such as the use of diary studies and focus groups of non-speaking individuals.

One method is to identify the words that speaking individuals use most frequently and commonly. Such core vocabulary lists may provide a good indication of the words that are most important to include on an AAC system. The underlying assumption is that the communication needs of persons who require AAC are the same as those of persons who use speech. Compiling core word lists based on data obtained from speech samples has been one method for identifying vocabulary to be included on AAC systems (Beukelman, McGinnis, & Morrow, 1991).

2.5 Core and fringe vocabulary

Core vocabulary refers to the words most frequently and commonly used among individuals (Beukelman et al., 1989). Although the words that an individual person uses most frequently may depend to some extent on features such as personal interest and environments (Beukelman et al., 1989), it has been shown that, across individuals, there is often a high degree of commonality in the words they use most frequently. Across a number of different studies in English, for example, it has been established that approximately 200 to 400 words represent about 80% of spoken language used by individuals of various ages (van Tilborg & Deckers, 2016). Core vocabulary word lists often include many structure words such as pronouns, conjunctions, prepositions, auxillary verbs, determinants and adverbs (Trembath et al., 2007).

Various authors emphasize the essence of adding frequently used vocabulary in AAC systems (Balandin & Iacono, 1999; Fallon et al., 2001; Trembath et al., 2007). Authors assume that if these core words are provided to a child on a system, he/she could express the majority of messages, as these words are said to be the framework of a language. They are essential in forming sentences and connecting utterances (Robillard et al., 2014). The inclusion of core vocabulary on AAC systems is specifically advocated for children with good language comprehension, since it is argued that core words allow for the expression of novel sentences, and a small vocabulary can be compensated for by circumlocution (Liu & Sloane, 2006). However, other authors stipulate that core word lists should be used with caution because an effective AAC system must contain not only frequently used words as identified by research, but also individualised vocabulary, also known as fringe vocabulary (Robillard et al., 2014; Balandin & Iacono, 1999).

In contrast to core vocabulary, fringe vocabulary items are words that are large in number, context-dependent, and highly individualised. These words are particular to a person's environment and interest. Fringe vocabulary is known for having a low degree of commonality among users and being content-specific (Trembath et al., 2007).
Fringe vocabulary changes rapidly and consists almost exclusively of content words (i.e. nouns, verbs and adjectives).

For accurate communication of messages, children with complex communication needs would require the inclusion of both core and fringe vocabularies AAC systems that. Consequently a balance is important between the two types of vocabulary (Robillard et al., 2014). Research highlights the need for professionals involved in vocabulary predictions and selection to consult a number of sources and informants in order to gain understanding and a working knowledge of particular situations and to validate that vocabulary items are socially appropriate in a given situation (Balandin & Iacono, 1999). Particularly, both core vocabulary lists and the use of informants and ecological inventories to determine fringe vocabulary may be helpful to establish a useful set of vocabulary items to be included on the AAC system.

2.6 Core vocabulary in different languages

Since core vocabulary has been found to include many structure words, various authors have suggested that it would be difficult to translate core vocabulary from one language to another (Trembath et al., 2007, Mngomezulu, 2017), especially if languages differ considerably in their grammatical and morphological structure. To date, few studies seem to have attempted to compare core vocabulary findings across different languages. Although van Tilborg and Deckers (2016) compared the words found on 15 core vocabulary lists generated from persons of different ages, with and without disabilities, across different languages (among others, German, English and French, as well as a language spoken by Taiwanese children, which was not specified), no language-specific comparisons were conducted. In order to obtain an impression of the similarity and differences between core vocabulary lists across different languages, a composite English word list compiled by Hattingh (2018) from five English core vocabulary lists was obtained as a basis for comparison with a Korean, German and isiZulu core vocabulary list.

Hattingh (2018) consulted lists based on core vocabulary studies compiled by Mngomezulu (2017) as well as van Tilborg and Deckers (2016). She then selected English

core vocabulary lists as identified through these studies that complied with the following criteria: (a) a published list of the 100 most frequently occurring words was available, (b) lists were based on data obtained from at least five participants, (c) lists were based on data obtained from transcribing spontaneous natural conversations rather than researcher-initiated activities, and (d) vocabulary lists were published in 1980 or later.

Five English core vocabulary lists were identified as a basis for a composite list. A brief overview of the studies that generated these lists is given in Table 2.1

Table 2.1

English Studies Selected to Generate Composite List

Authors	Population sampled	Location
Beukelman et al., (1989)	Six children without disabilities between 3; 8 and 4; 9 (years, months).	United States of America
(Stuart, Beukelman, & King, 1997)	Elderly individuals (64-74 years old and 75-85 years old)	United States of America
Trembathet al., (2007)	Children of 3-5 years old who speak English only.	Australia
Boenisch and Soto (2015)	7 – 14-year-old children, native speakers of English and children who spoke English as a second language (ESL); the students belonged to a wide range of ethnic and cultural backgrounds and came from varying socio-economic backgrounds, as is typical in urban school districts throughout the US.	United States of America (San Francisco Bay Area)

The top 100 words from each list were used. In total 183 unique words were identified, of which 86 occurred on at least three of the five lists. These 86 words were used as a basis for comparison with core vocabulary lists in other languages.

2.6.1. Comparison of the English composite list with core vocabulary lists developed for other languages

In order to identify core vocabulary lists in other languages, the researcher consulted the list of core vocabulary studies identified by Mngomezulu (2017) and identified six studies

that aimed at identifying core vocabulary in a language other than English – three studies that sampled Korean, two that sampled Mandarin, and one that sampled French (Lee, Kim, & Park, 2005; Shin & Hill, 2016; Kim, Park & Min, 2003; Liu & Sloane, 2006; Chen, Hill, & Yao, 2009; Robillard et al., 2014). The author furthermore consulted an article by van Tilborg and Deckers (2016), and identified an additional study conducted regarding German (Boenisch & Soto, 2015) and two additional studies conducted with Taiwanese children (language not specified) from their reference list (Chen, Chen, & Chen, 2013; Chen, Ko, Hsu, Lin, Chi, & Huang, 2011). She then traced these references to identify whether they included core vocabulary lists that complied with the criteria as set out in Table 2.2.

Table 2.2

Selection Criteria for Comparison Articles

Criterion	Justification
Vocabulary lists were accessible through the researcher's institution's data base	This was a pre-requisite for access to articles for comparison.
Vocabulary lists provided a published list of at least the top 100 most frequently used vocabulary items.	This was a requirement to facilitate comparisons.
Vocabulary lists used for the comparison needed to have English translations where possible or needed to be in a language that allowed the researcher to translate the list herself or via a translation service.	This was required as a basis for a comparison between the English composite list and the core vocabulary list in the other languages.
Studies needed to have been conducted in 1980 or later.	This was done in order to optimise the relevancy of the vocabulary used.

The researcher identified only three studies (targeting German, Korean and isiZulu respectively) that complied with the selection criteria. A summary of these studies is provided in Table 2.3.

Table 2.3

Table with Summary of Studies

Authors and year	**Participants**	**Data collection and analysis**	**Operational definition of core vocabulary**	**Results**
Boenisch and Soto (2015)	Forty-four learners attending Grades 2-10 in special schools for children with intellectual disabilities in Germany.	The children wore portable voice recorders while engaged in various school activities including classroom activities, break time and mealtime at school. In total 125 454 words were collected, averaging 2 850 per participant. The samples were analysed for the total number of words, total number of different words as well as frequency of word use. Words were further classified by their various parts of speech.	Core vocabulary refers to the most frequently used words that together cover 80% of all conversations.	From the total corpus, the top 201 most frequently used words made up 80% of the words transcribed and were designated as core vocabulary. More than 80% of the top 100 most frequently used words overlapped with those used by German learners without disabilities in Grade 2.
Shin and Hill (2016)	Twelve monolingual native Korean speakers, six men and six women between the ages of 18 and 65, who were healthy and had no cognitive or sensory deficits.	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 a word commonality score.	Words occurring with a minimal frequency of 0.2‰ and a minimal commonality score of 50% (i.e., at least six of the 12 participants used the word).	In total 3 669 different words from a total vocabulary of 16 944 words were collected. This gave a list of 627 words that were used with a frequency of 0.2‰ or more. After applying the commonality criterion of 6 (50%) to this list, the final core vocabulary list identified comprised 219 words covering 60.82% of the total number of words used in the conversations.

Authors and year	Participants	Data collection and analysis	Operational definition of core vocabulary	Results
Mngomezulu (2017)	Six preschool children (aged 5; 0 to 6; 11) speaking isiZulu as their home language	The children were fitted with voice recorders while carrying out all the activities in their normal preschool day. These recorded 1 500 words per participant. After transcription, the orthographic words were further divided into formatives (akin to morphemes in English). The samples were analysed for the total number of formatives, total number of different formatives as well as frequency of formative use. Formatives were further classified by their various parts of speech.	Formatives (morphemes) that were used with a minimal frequency of 0.5‰ and used by at least two of the six participants were designated as core formatives.	The transcribed composite sample yielded a total of 20 137 formatives. Of these, 221 were used with a minimal frequency of 0.5‰ and by at least two participants. These 221 formatives covered 88.9% of the formatives used in the total sample.

The top 100 most frequently occurring items in each of the lists were identified to simplify comparisons. English translations of vocabulary items in the isiZulu and Korean lists (of those items where translation was possible) were provided by the authors of these respective studies. Mngomezulu (2017) followed a rigorous process when translating the formatives according to the dictionary (Dent & Nyembezi, 1995). The German list was translated by a German speaker and cross-checked against the online Langenscheidt German-English Dictionary (https://en.langenscheidt.com/german-english/).

The English translations of each core vocabulary list were alphabetised and compared to the English composite list identified by Hattingh (2018). The number of vocabulary items that had an equivalent in the English list were then determined.

When translating vocabulary items from one language to another, several challenges can arise that can make a word-for-word, one-on-one translation difficult. First, one word may have several different translations. This can be due to homonymy and polysemy in the original word. The Shared Distributed Asymmetrical Model illustrates how translation pairs generally do not share meaning completely. For example, the Sepedi word *nyaka* can be translated as 'search', 'look for', or 'want' in English. The model also acknowledges that not all words in one language may be translated by means of a single word in other languages, but may require circumlocution instead (de Groot, 2013). The second challenge occurs when the original word or vocabulary unit is translated not by a single word or item, but by a phrase in the target language. These factors complicate the comparison between word lists in different languages somewhat. The English translations of all three core vocabulary lists all contained examples of numerous English words being provided for one original vocabulary item and also of English phrases translating the original item. To enable a comparison between this translated list and the English composite list by Hattingh (2018), these challenges were dealt with in the following way:

1. If the original vocabulary item had numerous translations, each of these was considered in the comparison. If at least one of the translations was present in the composite list, the word was considered to have an equivalent in the English list. For example, the word *bolela* means 'speak/tell/talk'. The word was considered to have an equivalent if any of the English translations was present on the English lists.

2. When the original vocabulary item was translated by an English phrase, all words making up the phrase had to be present in the English composite list for the word to be regarded as having an equivalent. For example, *tše* means 'these ones'. The English list had to contain both 'these' and 'ones' for the word to be considered as having an equivalent.

In addition, some words/vocabulary items in the Korean and isiZulu list were not translatable, since they marked an aspect of grammar that does not have an equivalent in English. Understandably, these items therefore had no equivalent in the English composite list.

Appendix A gives a complete list of English translations of the 100 top items of each list, and indicates which of these had equivalents in the English composite list. Table 2.4 illustrates the total number of words per top 100 in each of the three lists that had equivalent words in the English composite list.

Table 2.4

Number of Common Words Identified in Comparing the Top 100 Core Vocabulary Items of Lists in Different Languages and an English Composite List

Language	Number of words in top 100 that had an equivalent in the English Composite list
isiZulu	31
Korean	45
German	63

It was found that 31 vocabulary items in the isiZulu top 100 list, 45 items in the Korean top 100 list and 63 items in the German top 100 list had an equivalent in the English composite list of 84 items. The German and English lists had the highest overlap. In part, this may be attributed to the fact that they belong to the same language family (Western Germanic languages) and therefore share semantic and syntactic similarities. Korean and isiZulu, in turn, belong to different language families when compared to English, and their grammatical and semantic features also differ considerably from English. This may partly account for the results described above. However, the fact that each list was obtained from a specific

population with a specific age, in a specific geography, and during different types of natural conversations (e.g. name and context) can also contribute to the uniqueness of the lists. As is evident from the comparison of English core vocabulary lists (Hattingh, 2018), these factors do have an influence even if the language across lists remains the same, and no core word list should ever be regarded as 'universal' for a specific language.

These results do support the notion that core vocabulary should be determined from language samples from specific populations and in specific languages, rather than being translated. In addition, the orthographic word as unit of analysis used in a core vocabulary study must be carefully considered. Although the orthographic word (as defined by the space that appears between the series of letters that make up a written word in languages with alphabetic orthographies) as unit of analysis may be useful for some languages, this is by no means an ideal basis for universal grammatical description. For example, in one language one may need a sentence (consisting of a number of words) to express something, whereas in another one may need only one word to express the same information (Kosch, 2006). In the studies compared, isiZulu had formatives as the unit of analysis. Other languages, such as German, French, and Korean, analysed words, as was done in English studies.

Since it is clear from the above discussion that core vocabulary is not necessarily translatable, the need for a Sepedi core vocabulary list developed from transcriptions of natural conversations becomes clear. In addition, the linguistic structure of Sepedi needs to be considered to determine a useful unit of analysis. In the sections following, general background information about the Sepedi language will be given first. This will be followed by a consideration of the Sepedi language structure and particularly its morphological typology. The Sepedi writing convention will also be discussed. The implications of the Sepedi language structure and its orthography for the establishment of a core vocabulary will lastly be considered.

2.7 The Sepedi language

2.7.1 Historical background

A collective set of people from the Limpopo area in South Africa, speaking various dialects of the Sotho language, came to be known as the Bapedi (Pedi people) (Mokwana, 2009). Their language, Sepedi (Sesotho sa Leboa as previously known) is spoken by 9, 1% of

the South African population, constituting about 12 million citizens. It is the fifth most frequently spoken home language in South Africa (Statistics South Africa, 2012).

The Sepedi language falls in the Sotho language cluster together with SeTswana and SeSotho. Sepedi has several distinct varieties, including Sepulana and Selobedu (Prah, 2007). The Sotho language cluster falls within a linguistic grouping that has historically been referred to as Bantu languages. However, since this term was used under the previous South African apartheid regime to refer to African people in a derogatory way, it will not be used in this study. The term 'African languages 'is more socially and politically accepted in South Africa but linguistically may be perceived as misleading as one may assume it refers to all languages spoken in South Africa, whereas not all of them belong to the same language family' (Kosch, 2006). [7]

In 1948, when the National Party Afrikaner elite came into power, it carried on its back the notion of achieving cultural and linguistic supremacy as a collective memory to its historical baggage of cultural rivalry against the English. Under the apartheid regime, two languages of the white minority were holding sway over and above indigenous African languages. These languages were English and Afrikaans (Prah, 2007).

Forced removals and "grand apartheid" began in 1960 in order to make the so-called non-white population "invisible". People who spoke different languages were geographically and physically separated into territorial units. Specific rules also banned them from accessing other homelands during certain hours of the day. Cultural visibility of a group was only allowed in the particular Bantustan/homeland. It was on the basis of these homelands that quasi-independence developed. Quasi-independence refers to not being influenced or controlled by others and thinking for oneself (Kosar, 2011). By the 1970 and 1980s each group/tribe had its own homeland, where languages and cultures were embraced. The Pedi people's homeland was Lebowa (Prah, 2007).

Each Bantustan was therefore regarded as and treated like a unique "nation" in order to eliminate the potential unity of the African majority. Therefore, the languages and cultures

[7] The term "Bantu language" is acknowledged as it was used to refer to a specified language family. The term has derogatory connotations and was deemed inappropriate – it is thus no longer used in society and in linguistic contexts (Kiango, 2005). The authors have resorted to using "African language" instead.

of the African-language-speaking citizens were relegated to a minority status, even though these citizens comprised about 75% of the South African population (Prah, 2007).

The dawn of democracy in 1994 resulted in significant changes for languages in South Africa. The Bill of Rights stated that "Everyone has the right to use any language and to participate in the cultural life of their choice, but no one exercising these rights may do so in a manner inconsistent with any provision of the Bill of Rights." The Constitution, which came into force in 1996, recognised (in Section 6 of the *Founding Provisions*) 11 official languages of the Republic, namely SePedi, SeSotho, SeTswana, siSwati, TshiVenda, XiTsonga, Afrikaans, English, isiNdebele, isiXhosa and isiZulu. The new democratic South Africa has given all people the right to reside in any province. As a result, census data of 2011 (Statistics South Africa 2011) indicates that although most Sepedi-speaking people are found in Limpopo, the rest of the Pedi people are dispersed across the country as a result of job opportunities, marriages and better service provision, to name a few (Prah, 2007).

2.7.2 The Sepedi language structure

Morphology (morphemes and words) and syntax (clauses, phrases and sentences) constitute the grammatical units of a language. Grammar refers to the rules that govern the formation and usage of the units of language.

The Sepedi language is rich in morphemes (Kosch, 2006). These are identified as the smallest possible sequences of sounds that can be associated with a particular meaning or constant function and occur regularly in more than one word. It is for this reason that the morpheme is described as the smallest meaning-bearing unit of grammatical analysis. In Sepedi, morphemes are used to formulate words, and may also be used to change the part of speech to which the word belongs (e.g., the addition of a morpheme to a verb lemma is used to formulate a noun) and to show functions such as tenses and moods. As in other African languages, morphemes are used to distinguish nouns in an elaborate noun class system of 18 classes (van Wyk, Groenewald, Prinsloo, Kock, Taljard; 1992). For example, the morpheme *mo-* can appear in words such as *mollo* (fire), *monwana* (finger) or *mohlare* (tree). The difference in the meaning of these three orthographic words is determined by the second morpheme in each word. The morpheme *mo-* has the function of marking the word a singular inanimate noun of Class 3. All singular nouns in this class form the plural by replacing *mo-*

with *me*-. The elements *mo*- and *me*- can therefore be assumed to be morphemes. Therefore knowledge of Sepedi noun classes becomes imperative in categorising nouns to determine the plural form.

Morphemes can be identified by comparing lists of partially similar forms and 'picking out' recurring sequences of sounds that bear the same semantic content or perform the same grammatical functions. For example, –*rek*- is a morpheme. It occurs repeatedly in the words and has the same basic meaning, 'buy' in each word. It appears in words such as *moreki* (buyer), *morekisi* (seller) etc. This method of discovering morphemes has its own downfalls. Sometimes a morpheme may differ slightly in shape from the morpheme to which it is semantically related. For example,when considering *go ithekela* (to buy for oneself), one needs to have knowledge of the phonological rules of the language to realise that -*thek*- is just a different manifestation of the same morpheme, -*rek*-. One needs to understand that the concept of morpheme is known as a concrete and an abstract unit (i.e. the morpheme can be altered in its original form, for example from -*rek*- to -*ithek*-) (Kosch, 2006). Because Sepedi is rich in morphemes and most morphemes remain constant in form, it is classified as a mainly agglutinative language morphologically. This is further explained in the next section.

2.7.2.1 Morphological typology of Sepedi

Languages are classified according to various principles in typology. These include ranges of sounds (phonetic typology), ways in which sounds and sound features are distributed into phonological systems and syllable structures (phonological typology), grammatical systems to mark syntactic relationships and sentence structure, for example, by word order, word class membership and word structure patterns. The four classical types that are distinguished in morphological typology are isolating, agglutinating (or agglutinative), fusional and polysynthetic. There is probably no language that corresponds perfectly to any single type. Nevertheless, languages may be slotted into one the classical types depending on their greater affinity for one specific type. The African languages (such as Sepedi), for example, are classified as mainly agglutinating, although they also display some fusional and isolating features. Therefore the classical types should be regarded as mere abstract constructs with no clear-cut limits. Knowledge of morphological typology enables a reader to understand how words are formed.

African languages are said to be agglutinative. The term 'agglutinating' was derived from the Latin 'gluten,' meaning 'glue'. In the context of language, agglutination refers to the process through which various affixes are 'glued' on or simply 'stuck' onto other morphemes in a sequence, without the affix changing its structure (Kosch, 2006). This implies that ideally there are no allomorphs (i.e. alternative realisations of morphemes) in such languages. However, Sepedi is not an entirely agglutinative language. Allomorphs do exist, for example, in reflexive verbs (verbs that indicate that an action was done to oneself). In these verbs, the verb stem changes in phonology as a result of the operation of plosivation (occlusivation, sound strengthening) (Kosch, 2006). See also the example of *ithekela* given in Section 2.7.2 above.

2.7.2.2 Sepedi orthography

A word can be used to refer to two concepts. One can distinguish a word as either orthographic or linguistic (Kosch, 2006). An orthographic word refers to a unit separated by spaces from other units in written sentences. In contrast, a linguistic word refers to a unit that functions as member of a word category and has its own particular meaning. The writing convention in African languages whereby single linguistic words may be represented by a number of orthographically separated units is known as the disjunctive writing system. This system is adopted by all Sotho languages (i.e. Sepedi, SeTswana and SeSotho). Therefore, linguistic and orthographic words in Sepedi do not always map onto each other in a one-on-one fashion. For example, linguistically, *ke a se tseba* (I know it) is described as one word, but orthographically, it consists of four words. In essence, the linguistic word is often represented by several orthographic words, as a result of the disjunctive writing system followed in Sotho languages. One consequence of this disjunctive orthography is that many orthographic words consist of only one single morpheme. However, some morphemes are written conjunctively within a word, such as the noun class prefix, and verbal endings. For example, the word *mohlare* (tree) consist of two morphemes, namely the class prefix *mo-* and the noun root *-hlare*.

2.8. Selecting a unit of analysis for a Sepedi core vocabulary study

Many core vocabulary studies in English have used the orthographic space to define the linguistic units ('words') that are counted to determine the most frequent vocabulary items. In contrast, Mngomezulu (2017) chose to count morphemes, or what is termed *izakhi* in isiZulu (Nyembezi, 1982, p. 43), translated as formative. This decision was based

primarily on the conjunctive orthography of isiZulu, which results in each orthographic word comprising multiple morphemes. This complicates the establishment of a core vocabulary based only on orthographic words, as one orthographic word may relate to a number of concepts and require numerous pictures to be represented. The drawback of a morphological analysis is that the established core units are rather abstract, as many morphemes are not meaningful and only gain meaning in combination with other morphemes. This level of abstraction may make it difficult to represent the units using graphic symbols, and may also require metalinguistic skills on the part of the person using the system to combine these abstract units.

Since Sepedi has a disjunctive orthography, the use of the orthographic space as a definer of the units of analysis would not necessarily have the same drawbacks as it has in isiZulu. In this study, orthographic words were therefore used as the primary unit of analysis. This meant that the units identified in the analysis linguistically consisted of primarily one, but sometimes more than one morpheme, as discussed in Section 2.7.2.2. Examples of orthographic words that contain more than one morpheme are verbs, nouns, and adjectives. For example, the verbs *rêka, rêke,* and *rêkile* each consist of a verb root (*rêk*-) and various suffixes or endings. In order to preserve some of the commonality that occurred within these units, the following grammatical variations of orthographic words were counted together (although note was taken of different surface structures):

- Nouns: Plural forms, locative forms and diminutives were all counted under the lemma (singular form). For example, the words *ntlo, dintlo, ntloana, ntlong* were all counted under *ntlo*.
- Verbs: Tenses, moods and other inflections were all counted under the lemma (imperfect positive form). For example, the words *reka, rekela, rekeleng, nthekela* were all counted under *reka*.
- Adjectives: Different inflections formed by the root and the class prefix were all counted under the root. For example, *golo, bogolo, megolo, mogolo* were all counted under *golo*.

A more detailed breakdown of the way these units were counted is given in the coding rules in Appendix L.

The decision to count these word forms together is precedented by similar decisions made in some English studies (e.g. Boenisch & Soto, 2015). The argument for such an approach is that it is unlikely that an AAC system with limited space for vocabulary would include various inflections of the same word. However, no other inflections were counted together, and words that contained derivational rather than inflectional morphemes (i.e., morphemes that affect meaning and possibly word class significantly) were always counted separately and not coded. Therefore, words such as *reka* (buy), *moreki* (buyer), and *morekisi* (seller) were always counted as different words.

The decision to count orthographic words did present the challenge of heteronyms and polysemous words. Although these two phenomena differ slightly from each other, they both describe words that are spelled the same (though not necessarily pronounced the same) but that have slightly or completely different meanings, and are often also classified as different parts of speech. For example *ka* may be a preposition, such as in the sentence *serurubele se tšwa ka lefasetere* (the butterfly goes out through the window) or it may be a potential morpheme such as in the sentence *o ka dira eng?* (what can you do?). In order to avoid confounding counts of heteronyms and polysemous words, arbitrary codes were given to the heteronyms and polysemes to allow them to be counted separately by the SALT software program.

2.9 Parts of speech in Sepedi

Core vocabularies identified in other languages have been further described by classifying vocabulary items into parts of speech (e.g., Boenisch & Soto, 2015; Mngomezulu, 2017). Various grammar books describe the parts of speech found in Sepedi (e.g., Poulos & Lourens, 1994; van Wyk et al., 1992). Dictionaries, too, often classify the entered word as a particular part of speech, for example, the Oxford *Pukuntšu ya Sekolo* dictionary (de Schryver, 2007). Furthermore, an automatic Sepedi 'part-of-speech tagger' demonstration (de Pauw & de Schryver, 2007) is available online (at https://www.aflat.org/sothotag) – this tagger allows the user to enter text in Sepedi online and run an analysis, resulting in each orthographic word being assigned a specific part of speech. Grammar books (e.g. Poulos & Lourens, 1994) assign parts of speech to linguistic rather than orthographic words – in Sepedi, there are various orthographic words that are morphemes (as described in Section

2.9.1.). Dictionaries and part-of-speech taggers, in contrast, assign parts of speech to orthographic words. Since this study identified a core vocabulary by using orthographic words as the level of analysis, the parts of speech relevant to orthographic words are described below. The description will commence with parts of speech that are designated as content vocabulary (words that carry semantic meaning), namely nouns, verbs, adjectives and adverbs. Thereafter, parts of speech designated as structure vocabulary will be considered, namely concords, pronouns, prepositions and locative particles, as well as verbal prefixes and suffixes (e.g., negative morphemes, future morphemes, hortative particle, etc.).

2.9.1 Nouns

As mentioned above, Sepedi (like other African languages of the same language family) has an elaborate noun class system consisting of 15 main classes (van Wyk et al., 1992), numbered according to an internationally recognised system. Nouns consist of a class prefix and a lemma (sometimes also called root; see van Wyk et al., 1992, p. 6). For classes 1–10, the uneven classes contain singular nouns, whereas the evenly numbered classes contain plural nouns. For example:

Class 1: *Mosadi* (*mo-* = class prefix; *-sadi* = noun stem) – ‘woman’

Class 2: *Basadi* (*ba-* = class prefix; *-sadi* = noun stem) – ‘women’

Further inflections of nouns include the diminutive, denoted by adding the suffix *–ana* or *–nyana* to the noun. Various sound changes typically occur to the original noun ending when adding a diminutive suffix. For example:

mošemane (boy) – *mošemanyana* (small boy)

The last inflection of the noun that will be considered in this study is addition of the locative suffix *–ng*, resulting in a noun that has been transformed into a location (also called a locativised noun) (van Wyk et al., 1992). For example:

sekôlô ([the] school) – *sekôlông* (to/at [the] school

It is worth noting that, while nouns in Classes 1–10 and 14 typically describe objects, people and also abstract nouns (specifically Class 14), nouns in Classes 16–18 would be translated as prepositions in English. For example:

Class 16: *Fase* (*fa-* = class prefix; *-se* = noun stem) – ‘under’

Furthermore, since what grammar books describe as noun class 15 consists of a disjunctively written infinite prefix and a verbal stem, these constructions were not designated as nouns in this study, but as infinite prefix and verb. For example:

go kitima (*go* = infinite prefix; *kitima* = verbal stem) – 'to run'.

2.9.2 Verbs

Verbs consist of a root and one or more prefixes and suffixes that are written conjunctively as part of the verb (e.g., suffixes denoting tenses or moods, and prefixes denoting the object concord of the first person singular). Herewith some examples of verb roots and prefixes/suffixes:

Bala ('*bal-'* = verb lemma; *-a* = imperfect form ending) – 'read(s)'

Rêkilê (*'rêk-'*= verb root; *-il-* = perfect tense suffix; *-ê* = perfect form ending) – 'bought'

Ntseba (*n-* = object concord first person singular; *-tseb-* = verb root; *-a* = imperfect form ending) – 'know(s) me'

2.9.3 Adjectives

Adjectives typically consist of an adjectival root preceded by the class prefix. Examples of adjectives are

Mogolo (*mo-* = class prefix of noun class 1; *-golo* = adjectival root) – 'big/large'

Monna yô mogolo – 'the large man'

Megolo (*me-* = class prefix noun class 4; *-golo* = adjectival root) – 'big/large;

Metse yê megolo – 'the big villages'.

2.9.4 Adverbs

Adverbs in Sepedi describe or qualify verbs. Many adverbs consist of only one morpheme, for example *kudu* ('very') and are regarded as true adverbs. Some consist of the morpheme *ga-* placed before an adjectival root/stem or a noun, and are considered derived adverbs. For example:

Gabotse (*ga-* = morpheme; *-botse* = noun) – 'well'.

Neither true nor derived adverbs have any inflection.

2.9.5 Concords

Concords are linked to the noun class of the noun they refer to, but they should not be confused with pronouns. Subject concords link the noun that acts as the subject of the sentence to the verb, for example,

Monna	*o*	*šoma*	*serapêng.*
Noun class 1	subject concord class 1	verb	locativised noun
(The man		works	in the garden.)

Object concords take the place of a noun acting as the object of a sentence, for example, the object *dipuku* ('books') in the first sentence is replaced by the object concord *di* in the second sentence in the example below:

Monna o rêka puku. (The man buys books.)
Monna o a ***di*** *rêka.* (The man buys them.)

Possessive concords are used in possessive constructions. In the examples below, *wa* and *tša* are possessive concords:

mmotoro wa monna (the car of the man)
diaparô tša bana (the clothes of the children).

There are specific subject, object, and possessive concords for each of the 15 noun classes.

2.9.6 Prepositions and locative particles

Prepositions and locative particles in Sepedi are used to show the relationship between nouns and other words in a sentence. They are used in front of nouns. For example,

Mosadi	*o*	*tsena*	*ka*	*gae.*
Noun class 1	subject concord	verb	preposition	noun
(The woman		enters	into	the house)

Ngwana	*o*	*dutše*	*godimo*	*ga*	*tafola.*
Noun	subject concord	verb	noun	locative particle	noun
(The child		is sitting	on top	of	the table).

2.9.7 Verbal prefixes and suffixes

Some verbal prefixes and suffixes are written disjunctively from the rest of the verb and were identified separately. Each of these words modifies the meaning of the verb in the sentence to some extent. These include negative morphemes, aspectual prefixes, copulative particles and infinite prefixes, to name a few. For example, the infinite prefix *go* transforms the verb into an infinitive:

Monna	*o*	*ya*	*go*	*robala.*
noun	concord	verb	infinite prefix	verb
(The man		is going		to sleep.)

2.9.8 Interjections

Interjections are words that express abrupt remarks or words used to express immediate feelings. An example of an interjection is *ah*. This interjection is often used in conversations to show the emotion of surprise.

2.10 Summary

This chapter provided background to the study by briefly discussing AAC in general, followed by more specific information pertaining to graphic symbol-based AAC systems. The vocabulary selection methods explained in literature were looked at, followed by an explanation of core and fringe vocabulary. The translatability of core vocabulary lists was considered by comparing lists in different languages, and it was concluded that core vocabulary is not translatable between languages. An overview of the Sepedi language and its structure was provided, with consideration of the implications for a core vocabulary study. The different parts of speech of the language were also described.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter of the thesis focuses on the research methodology used in the study. This chapter commences with the main aim, sub-aims and the research design, as well as a summary of the phases of the study. This will be followed by a description of the setting, the description of the participants (including recruitment, sampling and selection criteria), materials and equipment. The aims, procedures, results and recommendations of the pilot study will be reported on. The following sections will detail the procedures for data collection, data analysis, ethical considerations and issues concerning reliability and validity.

3.2 Aims

3.2.1 Main aim

The main aim of the study was to determine the words most frequently and commonly used (core vocabulary) by Sepedi-speaking preschool children without disabilities during regular preschool-based activities.

3.2.2 Sub-aims

The sub-aims of the study were:

i. To identify the words and the number of different words Sepedi-speaking children without disabilities use during regular preschool activities.
ii. To determine the words most frequently and commonly used (core vocabulary) by Sepedi-speaking children without disabilities during regular preschool activities
iii. To describe the Sepedi core vocabulary by parts of speech as well as by differentiating structure (grammatical) and content (lexical) vocabulary
iv. To compare the core vocabulary identified with that found in other studies.

3.3 Research design and phases

A quantitative descriptive observational design was used (McMillan & Schumacher, 2010). Six children aged 5; 0 to 6; 11 who speak Sepedi as a first language were recorded

with body-worn recorders and small microphones during their regular preschool activities. The design allowed the researcher to observe participants in their natural environment, thereby increasing objectivity (Cresswell, 2009). The researcher did not exert control over the variables and the observation in a natural environment. This can increase the external validity of the findings. Disadvantages of the design include that participants may react to being observed and act atypically, reducing the validity of the data. The data collected in studies using this design is also typically limited to a small number of participants, resulting in narrow superficial datasets of the population (Cresswell, 2009).

3.3.1 Stages of the study

Ethics approval from the Research Ethics Committee, Faculty of Humanities, University of Pretoria, was obtained by the researcher prior to carrying out the study (see Appendix B). The stages of the study are illustrated in Figure 3.1. These include (1) recruitment of participants, (2) screening and selection of participants and (3) data collection. The methodology suggested was tested using a pilot study, which was carried out before the actual study was done.

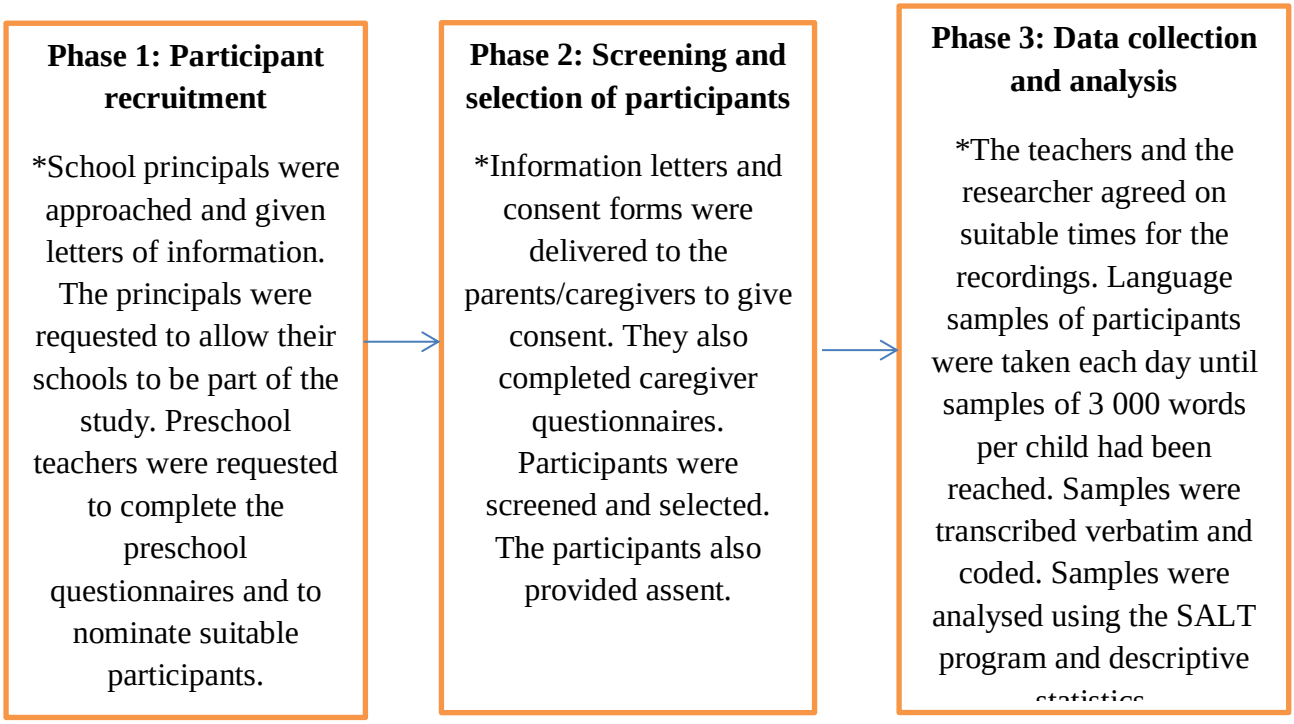


Figure 3.1. Stages of the study.

3.3.2 Setting

The three preschools from where participants were recruited were located in a semi-rural area southeast of Polokwane (the capital of Limpopo) around Lebowakgomo.

Chapter 3: Methodology

Lebowakgomo is situated in the district municipality of Capricorn, one of the five district municipalities of Limpopo. In this region 52,9% of the population use Sepedi as home language (Statistics South Africa, 2012). The schools are further situated in the local municipal area of Lepelle-Nkumpi. The maps provided in Figure 3.2 show the location.

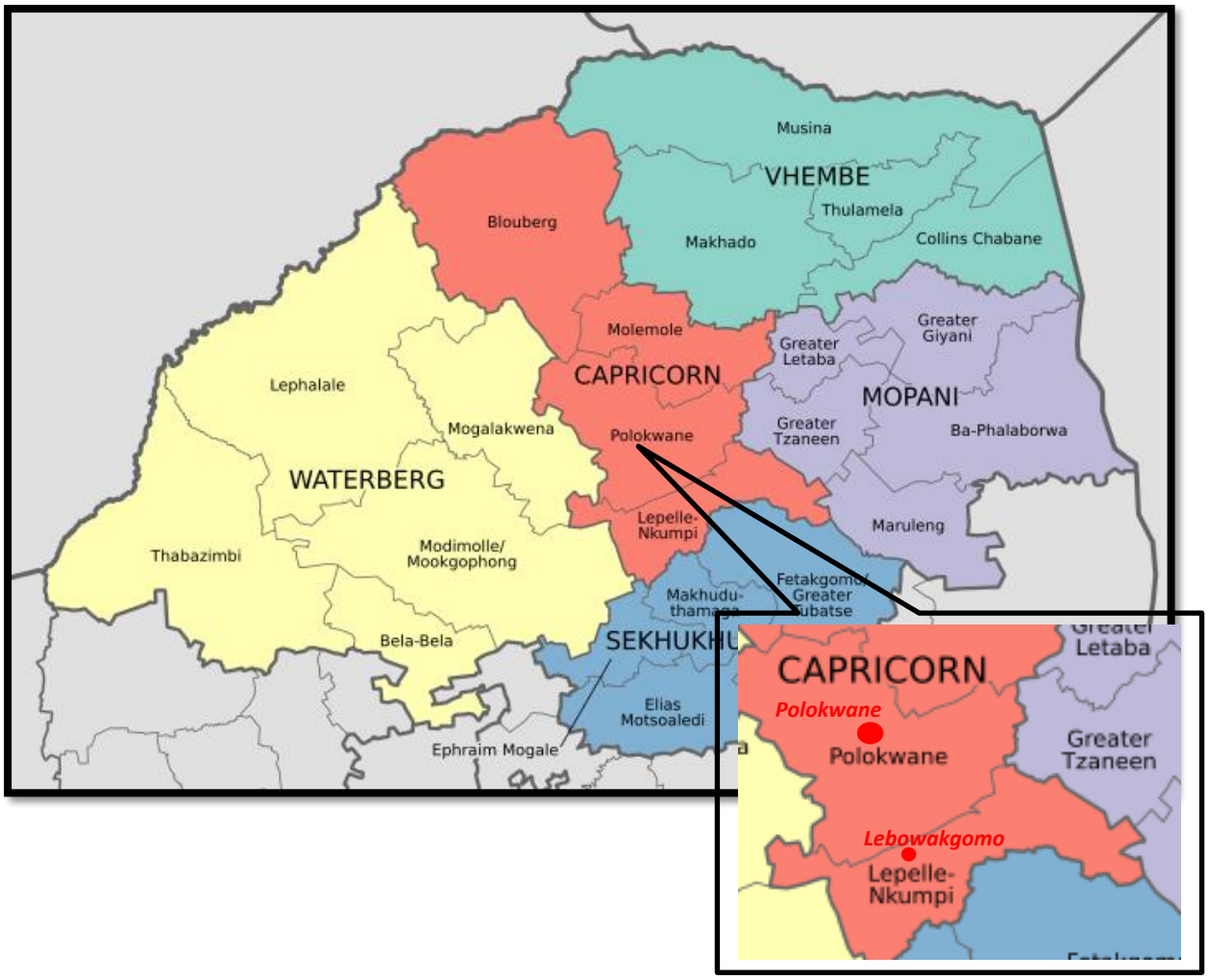


Figure 3.2 Map showing allocated area for research sites. Maps copied from Wikipedia (https://en.wikipedia.org/wiki/List_of_municipalities_in_Limpopo) under the free and open license terms of use.

The three preschools that were selected used Sepedi as the primary language of instruction. However, English was also used by teachers in the Grade R classrooms during activity time, morning rings and reading time. The number of children in the three Grade R classes from which participants were drawn varied from one site to another. Site 1 had 37 children, Site 2 had 62 children, and Site 3 had 72 children in the Grade R classroom. Two of

the three schools (Sites 1 and 2) each had one teacher responsible for the whole class, with no assistants in the classroom. The school at Site 3 had one teacher with five assistants. The children in the Grade R classrooms interacted with other children from the preschool (aged between zero and five years) each morning at the morning assemblies. They did not have contact with the other children during the rest of the day.

The schools followed the Grade R curriculum of the National Curriculum Framework. The schools followed routine programmes daily, which included mealtimes at the schools. Many teacher-facilitated activities took place on most of the recording days. This may have resulted in the lengthy recording times for most of the participants (as seen on the data collection time, indicating how long each participant took to reach the targeted 3 000 words), as the children did not initiate much speech during these activities. Although the children engaged in a lot of rote chanting, reciting and imitated speech, such utterances were not included in the data analysis. Only more voluntary and self-initiated speech was included. Most of the speech was obtained from break times, playtimes, toilet routines and mealtimes.

All three preschools had running water, electricity, playgrounds, indoor water and toilet facilities in place. All the preschools had separate toilet facilities for the teachers and the schools were fenced. One of the preschools had a landline telephone, a fax machine and internet connection at the school. One of the schools had none of the services listed. The last school had a landline telephone facility only.

3.4. Participants

3.4.1 Recruitment and sampling

The researcher obtained ethics approval from the Research Ethics Committee of the Faculty of Humanities of the University of Pretoria (Appendix B) prior to carrying out the study. Convenience sampling was employed to recruit participants. Principals of three preschools where Sepedi was the language of instruction were approached. The three preschools were located within a radius of 40 km of the researcher's residence.

The principals were provided with an information letter containing all the details of the study, as well as a permission form (see Appendix C) to give or decline permission to conduct the study at their preschool. All three principals gave permission. Educators from

each of the three preschools (whose principals had given permission) were requested to identify two children with no speech and language impairments to be considered as possible participants in the study and nominate them on the preschool background questionnaire (Appendix D). Educators were requested to provide the children's parents/legal guardians with information letters in both Sepedi and English (Appendix E) describing all aspects of the study. Two parents/legal guardians at each school were approached and had the opportunity to give or decline consent for their child to participate by completing a written consent form (see Appendix E). All the parents consented and were asked to fill in a caregiver questionnaire (Appendix F).

Participant assent was obtained prior to any study procedures. The researcher met each potential participant and explained the study to him/her in Sepedi, in child-friendly terms according to a script (Appendix G). Visual aids were used to enhance understanding. Potential participants had the opportunity to provide or decline assent both verbally and by marking their thumb print on a child-friendly form (see Appendix H). Six children were asked and all six provided assent.

3.4.2 Participant Selection Criteria

The selection criteria to select participants are provided below (Table 3.1).

Table 3.1

Participant Selection Criteria

Criterion	**Justification**	**Measure used**
Participant should be between the ages of 5; 0 and 6; 11.	Children of this age have relatively mature speech and language skills (Owens & Leonard, 2002)	Caregiver Questionnaire (Appendix F)
The participant should have no speech and language impairments and there should be no	Since the aim was to determine the vocabulary used by children without speech and language delays/disorders, no concerns in this area of development should be present. Also, since other	Caregiver Questionnaire (Appendix F)

Criterion	Justification	Measure used
other developmental concerns	developmental concerns may have concomitant effects on speech and language, these should not be present (Wallace, Berkman, Watson, Coyne-Beasley, Wood, Cullen, & Lohr, 2015)	
The child should have been in the preschool for at least three months and was required to attend at least three days per week	The participant should be comfortable and familiar with the environment to reduce novelty effects (Johnson, Arago, Shaik, & Palma-Rivas, 2000); Trembath et al., 2007)	Caregiver Questionnaire (Appendix F) Teacher nomination (as indicated on Preschool Background Questionnaire (Appendix D)
The child should have Sepedi as home language and the language of instruction at school	This is the language targeted in the study and therefore the chance of obtaining speech samples that include code-switching and mixing should be minimised (Bosma & Bloma, 2018).	Preschool Background Questionnaire (Appendix D) Caregiver Questionnaire (Appendix F)

3.4.3 Participant description

Descriptions of the selected participants are provided in Table 3.2.

Chapter 3: Methodology

Table 3.2

Participant Description

Part no.	Site	Gender	Age	Frequency of attendance at school	Primary language(s) used at home	Language exposure via television	Language exposure via radio	Other individuals living in the household			Monthly income of household
								Relation to child	Age	Language(s) spoken by individual	
1	1	M	6; 8	Daily	Sepedi	Sepedi	n/a	Sister	18	Sepedi	< R6 700
								Brother	13	Sepedi	
								Mother	43	Sepedi	
2	1	M	6; 8	Daily	Sepedi	Sepedi	Sepedi	Mother	36	Sepedi	>R6700
								Father	47	Sepedi	
3	2	F	6; 7	Daily	Sepedi	English	English	Brother	19	Sepedi and English	< R6 700
								Brother	17	Sepedi and English	
								Grandmother	67	Sepedi	
								Grandfather	80	Sepedi	
								Mother	28	Sepedi and English	
								Uncle	24	Sepedi	
4	2	F	6; 7	Daily	Sepedi	English	English Sepedi	Sister	10 months	None yet	>R6700
								Mother	33	Sepedi and English	
								Grandmother	65	Sepedi and English	
								Aunt	28	Sepedi and English	
5	3	M	5; 10	Daily	Sepedi	Sepedi	n/a	Brother	9	Sepedi	
								Mother	31	Sepedi	< R6 700
								Aunt	21	Sepedi	
								Uncle	27	Sepedi	
								Grandmother	84	Sepedi	
6	3	F	5; 3	Daily	Sepedi	Sepedi	Sepedi	Grandmother	56	Sepedi	< R6 700
								Uncle	34	Sepedi	
								Uncle	38	Sepedi	

From the table it is clear that children's ages ranged from 5; 3 to 6; 8. The mean age of all the participants was 6; 3 (*SD* = 7 months). While Sepedi was the primary language spoken at home, two participants were also exposed to English through family members and through radio/television. Four children seemed to be exposed exclusively to Sepedi in the home environment. Four parents reported that their monthly household income was < R6 700 (less than the minimum taxable income for the year in which the study was conducted), while the other two reported that they had a household income of > R6 700. In addition, caregivers were asked about amenities in the home as another indicator of socio-economic status. All of the caregivers reported that they had access to running water and electricity in the house. Four of the caregivers reported that they had indoor toilet facilities, whereas two reported that they did not have these facilities.

3.5. Materials and equipment

3.5.1 Equipment

3.5.1.1 Voice recorders and microphones

Small digital voice recorders (Olympus, Model DM 650) with lapel microphones (Audio Technica ATR 3350) were used to collect speech samples. The voice recorders were inserted into custom-made body-worn pouches that were able to fit around the participants' waists and the microphones were attached to the top part of the participants' jerseys/shirts using the microphone clip.

3.5.1.2 Laptop and headphones

The audio files were loaded from the recorders onto a Lenovo laptop (IdeaPad 110 N3060). Headphones were used to listen to the playback of the audio files during transcription.

3.5.2 Materials

3.5.2.1 Information letters and permission/ consent forms

Information letters and permission forms (see Appendix C) were sent out to the principals to explain the crucial aspects of the study. The letters and forms were drafted in English by the researcher and then she translated them into Sepedi. The letters were then checked by an educator who is fluent in Sepedi. Letters and forms were provided in both

English and Sepedi for the principal to select the preferred language. In South Africa, many individuals conduct their education and especially higher education in English, and literacy skills in English are therefore often better than literacy skills in their home language. There are often few educational opportunities that are conducted in African languages specifically (Rooy, 2006; de Klerk, 2016).

The parent information letters and consent forms (Appendix E) were distributed by the principals to the parents. These letters explained the fundamentals of the study and outlined the rights of the children. The letters and forms were also drafted in English and Sepedi by the researcher, and the Sepedi version was verified by an educator who is fluent in Sepedi. They were provided in both English and Sepedi in order to accommodate the parents' language needs and literacy level. The parents were given an opportunity to provide informed consent for the child to take part in the study.

3.5.2.2 Assent script and pictures for obtaining assent from participants

The researcher drafted a script in child-friendly language explaining all aspects of the study in Sepedi (see Appendix G). She included a picture illustrating each aspect. She also compiled an assent form accompanied by pictures to enable the children to provide or decline assent (see Appendix H). For the sake of the reader's convenience, the assent script and form are given in both Sepedi and English.

3.5.2.3 Caregiver questionnaire

The parent/caregiver questionnaire was devised to obtain background information relating to specific selection criteria, including the child's age, development, and the use of Sepedi as the main language in the home, as well as the length and frequency of the child's attendance at the current preschool. Questionnaires were drafted in English and Sepedi. These were checked by an educator who is fluent in both languages. They were sent to the parents in both English and Sepedi (see Appendix F). Additional information about the child's language exposure via family members and the media, the household income (above or below the cut-off for taxable income) and information about services accessible to the family was also sought in order to describe the participants in more detail. Questions were asked regarding the availability of water, electricity, toilet facilities and average income in the households. Since these factors may have an influence on how language is used, they were deemed important so that the results could be understood in context (Perkins, Finegood, &

Swain, 2013). The questionnaire was first devised in English, and then translated by the researcher (who is fluent in Sepedi) and corrected by an educator fluent in both written Sepedi and written English.

3.5.2.4 Preschool background questionnaires

A preschool background questionnaire (Appendix D) was devised to gather information about the language used at the school to establish whether the school met the selection criteria of having Sepedi as the main language of instruction, and also to gather descriptive information about the number of children in the preschool classes, exposure and use of other languages (e.g. in child-to-child interactions), and the daily program and curriculum followed in the preschool. The facilities/services available at each school were also noted for descriptive purposes. This form was drafted in English, and then translated by the researcher. It was then checked by an educator who is fluent in both languages. This form was provided to teachers in both Sepedi and English. On this form, teachers were given the selection criteria and asked to nominate a boy and a girl from their class for inclusion in the study.

3.5.2.5 Teacher instructions and daily feedback form

Teachers at the various preschools were provided with written instructions on how to operate and monitor voice recorders during the day (see Appendix I). The instructions also informed teachers of what to do when they suspected a problem during the day. A daily feedback form was also provided and was used by the researcher each day to conduct a brief feedback session at the end of each day of recording, during which the teacher informed the researcher how the recording went and if any problems had been experienced during recordings (see Appendix J).

Systematic Analysis of Language Transcript

The SALT software is a system that manages the process of analysing transcribed language samples. It was loaded onto a Windows computer and was used to automate some aspects of the analysis, such as frequency counts (e.g. number of words and number of different words) (Miller & Iglesias, 2012).

3.5.2.6 Transcription and coding rules

The researcher compiled a set of rules to guide the transcription process of the voice recordings. These rules were based on those provided by Trembath et al. (2007) and Mngomezulu (2017). The rules included conventions required for the SALT program and other rules that ensured that transcriptions were done in a consistent manner. The list of transcription rules is given in Appendix K.

Coding rules were also compiled to guide the researcher in terms of how to add codes that enabled the correct counting of inflected nouns, verbs and adjectives, as well as the counting of heteronyms and polysemous words. The list of coding rules is given in Appendix L.

3.6. Pilot study

A pilot study was conducted prior to the actual study with one participant to ensure that the recruitment and sampling methods, the participant selection criteria, the process of obtaining consent and assent, the equipment and material, and the procedures to collect data were workable and successful in obtaining the target data. The pilot study also served to test the data analysis procedures to allow for any changes to be made to obtain better results in the actual study. The pilot study was conducted with one participant from one of the sites (a site that was also included in the main study) who fitted the selection criteria. Recruitment, consent and assent processes were followed as described in Section 3.4.1, and the participant complied with all the selection criteria described in Table 3.1. Procedures outlined in Section 3.7 were followed. The child carried on with his routine activities while the recorder was fitted on the recording days. The child was monitored to ensure that the equipment did not interfere with his daily routine activities at school. The equipment was checked by the teacher at two-hourly intervals to ensure that the recordings were not interrupted at any point. The pilot study recordings were used to give the researcher an idea of the functioning of the equipment in terms of providing clear and audible recordings. Table 3.3 shows a summary of the procedures, outcomes and the changes suggested for the main study.

The pilot data provided the researcher with valuable information on the process of the study and provided clear recordings, which were sufficient for testing the data analysis process. However, some threats to the internal validity of the study were experienced during the pilot investigation (see Table 3.3 point 6). Therefore, the researcher decided not to include the data from the pilot study in the main study because of the threats to the accuracy

of the data that was collected. All recommendations suggested were implemented in the main study.

Table 3.3

Pilot Study Aims, Materials, Procedures, Results and Recommendations of the Pilot Study

Aims	**Materials**	**Procedure**	**Outcome**	**Amendments made for main study**
1. To test whether the selected method of recruiting participants would be successful.	Teacher information letter and nomination form	Information letters and details about selection criteria were delivered to the preschool teacher, who nominated a child to participate.	The teacher selected an appropriate participant.	None
2. To determine whether informed consent could effectively be obtained from participants' parents/guardians.	Parent/guardian information letter and consent form	The teacher sent the form home to the parent and organised the return prior to the data collection commencing.	The form was returned to the school before data collection commenced.	None
3. To ensure that the procedure of obtaining child assent was effective.	Assent script and form	The script was presented orally, and the prospective participant was shown the pictures to scaffold understanding. He was then asked a series of questions to ascertain understanding and request assent, and was asked to answer by ticking the correct option using a marker.	The participant showed understanding of the process and gave assent to participate.	None

Aims	Materials	Procedure	Outcome	Amendments made for main study
4. To establish whether the caregiver questionnaire was clear and easily comprehended.	Caregiver questionnaire	The educator at the school sent the questionnaire to the pilot participant's family. The family returned the completed questionnaire.	The parent told the teacher that she did not understand the questions about the other family members.	The researcher explained the section to all the teachers before they handed the questionnaires to the caregivers. Caregivers could therefore request clarification from the teachers.
5. To determine whether the preschool background questionnaire was concise and clear for the teachers to understand.	Preschool background questionnaire	The researcher sat with the teacher and principal to fill in the questionnaires regarding the preschool background.	The questionnaire was well understood and filled in.	None
6. To determine that the recording equipment did not interfere with the children's activities.	Waist pouches, digital voice recorders and lapel microphones	The recording equipment was placed inside a waist pouch over a jersey/shirt, and the small lapel microphone was attached to the child's collar. The researcher remained at the school to monitor whether the child was comfortable with this throughout the day.	The participant seemed comfortable with the equipment and managed to continue with his activities.	None
7. To determine whether recoding equipment	Digital voice recorders	It was ensured that the recorder was fully charged at	The recorder remained switched on from the time	None

Aims	Materials	Procedure	Outcome	Amendments made for main study
allowed for continuous recording of the child's speech for the periods of time deemed suitable by the teacher.		the beginning of the day. The recorder was padded inside the pouch to prevent accidental pressing of buttons.	it was fitted until the time that is was removed at the end of the preschool day.	
8. To ensure that the chosen equipment would be effective in achieving intelligible voice recordings and quality transcription data.	Digital voice recorders Lenovo laptop Headphones	The researcher transferred the audio files from the recorders to the laptop, and listened to the recordings via headphones to transcribe all the words uttered by the participating child from the recording.	The sample yielded 0.4% of unintelligible words. This indicates that most of the recording was intelligible.	None
9. To determine whether the child's spoken words that were transcribed seemed to represent a natural speech sample that was not overly influenced by the presence of the recorder.	Transcript generated	The transcript was scrutinised to note any topics of conversation that seemed to be referring to the recorder or seemed to be otherwise different from typical preschool interactions.	It became apparent from the transcript that the teacher felt she needed to prompt the child to speak more than other classmates and to give him more turns than the others. For example, the child was instructed to lead the morning ring, he was expected to answer all the questions asked in class and asked to lead the nursery rhymes.	Before data collection for the main study, each teacher was requested not to change her behaviour towards the target children but to behave as naturally as possible, not giving the target children special attention. The importance of having children behave as naturally as possible was explained to each teacher.

Aims	Materials	Procedure	Outcome	Amendments made for main study
10. To determine that transcription rules were comprehensive to guide the transcription process and to ensure that the appropriate unit of analysis was applied.	Transcription rules	The researcher applied the transcription rules during the transcription process to obtain a good quality transcription.	The transcription rules adequately guided the process, and each rule was applied as expected.	None
11. To determine if the coding system was reliable to guide the coding process	Coding rules	The coding rules were applied after the transcription process to obtain a good quality coded transcription.	The coding rules adequately identified the different words. However, more coding was required for heteronyms.	More coding rules were added to distinguish heteronyms.
12. To determine whether the analysis process would yield the desired outcomes as outlined in the aims of the study.	SALT program	The researcher selected the analysis options on the SALT program, which generated a Standard Measures Report and a Word Root Table.	The analysis options generated reports that provided the total count of words, the number of different words, the number of occurrences of each different word and word variation found in the sample.	None

3.7. Procedures

3.7.1 Ethical issues

The ethical principles for conducting research highlighted in the Belmont Report (1978) were reviewed and were upheld to protect participants in the study. The researcher firstly applied for and obtained clearance for the study from the Research Ethics Committee of the Faculty of Humanities, University of Pretoria (Appendix B) and requested permission from the school principals (see Appendix C).

The process of obtaining consent is crucial in ensuring 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). Respect for people (autonomy) was upheld by providing participants with the opportunity to take part in the study voluntarily. Parents/legal guardians and children were fully informed of all aspects of the study before voluntary consent (from parents/legal guardians) and voluntary assent (from children) were obtained.

The information forms and questionnaires sent out to parents (see Appendix E) utilised both English and Sepedi without discrimination, to ensure that parents were able to choose the language they preferred freely. The researcher thus ensured that parents understood the content of the information letter without bias/deception, and provided voluntary and informed consent. For the child participants, information about the study was given to the children in child-friendly language (supported by the use of pictures) to focus their attention and aid understanding of each statement being made. Participants and their parents/legal guardians were notified that they could withdraw from the study at any stage of the process without any negative consequences. Should they have chosen to withdraw; the child's data would immediately have been destroyed.

The study did not carry risk of physical or psychological harm, since it entailed natural observations rather than treatment. Recording equipment was fitted in such a way as not to interfere with normal activities in the preschool routine and not to harm participants physically. The participants were reminded that they should approach the teacher if the recording equipment caused any discomfort or if they would like it to be removed. Teachers were instructed to adjust or remove the equipment at the request of the participant. In this way, non-maleficence (not causing harm) was ensured.

Although taking part in the study did not directly benefit the participants, the data gathered is beneficial for the design of Sepedi AAC systems (principle of beneficence). Since participants did not derive any benefit over and above children not participating, there was no risk that the principle of justice would be violated – participants were not be advantaged over those not participating in the study.

Confidentiality was upheld by not disclosing the participants' identities to any third party. Only the researcher, supervisor and assistants who transcribed the recordings had access to the recordings. Transcriptions and questionnaires were de-identified prior to storage. Names of teachers and classmates mentioned in the recordings were replaced by a code. The data will be stored safely at the Centre for Augmentative and Alternative Communication at the University of Pretoria for 15 years.

3.7.2 Data collection

Appropriate recording days and times were discussed with the teachers at each preschool. On days agreed upon, the researcher arrived at the designated site (i.e. selected preschool). On the first day of recording, she explained the procedures and the reasons for them to each participant individually, following the assent script (see Appendix G). The researcher explained that she wanted to learn about the words that children use to talk to their friends and teachers and she asked each child if they would be willing to help her. She explained all the procedures (e.g. wearing a pouch around the waist with a recorder and a microphone on the collar). She also explained that the children should not touch or play with the pouch while they were wearing it and that they could tell their teacher if the pouch or microphone was annoying them. She gave each participant the opportunity to give or decline assent to take part in the study and reminded them that they were allowed to withdraw at any point with no negative outcomes. After children had provided assent, she fitted the voice recorder in the body-worn pouch and the lapel microphone on each participant and ensured that the recorder was switched on. The participants were then permitted to return to their respective classrooms. Teachers were given specific instructions (see Appendix I) regarding the recording equipment. This included that children should wear the equipment for as long a time as possible without causing harm or discomfort. Teachers were told that they were free to switch off recorders and possibly also remove them at any time that they felt it would be unsafe, unsuitable, or inappropriate for the children to wear the recording equipment/be

recorded. They were also requested to monitor the children and to ensure that the children were safe and comfortable at all times. They were asked to behave as they would normally in the classroom and not to alter their behaviour towards the children in an attempt to make the child "talk more". The researcher only came in the mornings to place the recorders on the children and in the afternoons to remove them each day. The researcher made telephonic contact with the teachers twice during the course of each day of recording to find out if any problems were being encountered. Teachers were also encouraged to report any problems. Only one teacher reported a problem on one recording day when the recorder had been switched off. The researcher told the teacher over the telephone how to rectify this. The researcher also obtained information after each recording day to find out if any problems had been encountered with the teachers (see Appendix J). Two children from each of three schools (amounting to a total of six) took part in the main study. Each participant was recorded until 3 000 orthographic words had been reached.

The number of days and recording hours required to reach the 3 000 word mark for each child are given in Table 3.4 below.

Table 3.4

Total Time Taken to Record 3 000 Orthographic Words from Each Participant

Participant number	**Total number of days recorded**	**Total time recorded**
1	2	08 h 14 min
2	4	21 h 26 min
3	4	16 h 04 min
4	2	07 h 47 min
5	4	16 h 33 min
6	3	12 h 42 min

3.7.3 Transcriptions

Transcriptions were done by following a set of predetermined transcription rules (see Appendix K), in order to ensure that the necessary conventions of the SALT program were followed and to ensure consistency in the transcriptions. The first 20 minutes of recording were not analysed, to counter any novelty effects. All conversations with the researcher and

discussions about the recording equipment and the process were discarded. Language samples were transcribed into the (SALT) program (Miller & Chapman, 1990). The researcher and four research assistants transcribed the samples verbatim (Miller & Chapman, 1990). The research assistants were trained individually on how to apply the transcription rules when transcribing the voice recordings. The research assistants had the following qualifications respectively: (1) matric qualification, (2) diploma in local government finance, (3) BA degree in language and editing and (4) B-Tech in language practice.

Individual files were created for each participant. These files were then combined into one composite file.

3.7.4 Coding

Coding rules were predetermined by the researcher to ensure consistency in coding the samples. These rules enabled the researcher to (1) identify code switching in the samples; (2) trace inflected forms of nouns, verbs and adjectives back to the root/lemma; and (3) distinguish between heteronyms (which are words that are spelled the same, have different meanings but sound differently). These transcription rules are important, as they prevent skewing of the conversational samples and minimise over- or undercounting.

The researcher coded the transcriptions (once these had been checked for accuracy – see Section 3.7.7) according to the pre-developed coding rules. The list of coding rules is given in Appendix L.

3.7.5 Data analysis

The researcher used the SALT program to perform relevant analyses, such as determining the TNW across all transcripts, as well as the NDW. The TTRs were also calculated by the SALT program. The TTR describes the degree of lexical variation and is calculated by dividing the NDW by the TNW. The number of times each different word occurred in the sample was also determined. The SALT program's Word Root Tables and the Standard Measures Report provide these functions. The Word Root Tables also allow inflected forms of verbs and nouns to be traced back to their lemma and adjectives to their root.

The frequency with which every different word occurred in the sample was then calculated by dividing the number of times a word was used in relation to the total number of words, expressed as *per mille* (‰).The information for the calculation was obtained from the word root table from the SALT program. The formula below was used:

$$Frequency\ ‰ = \frac{Total\ number\ of\ occurences}{Total\ number\ of\ words} \text{X} \frac{1000}{1}.$$

Analyses were then conducted to determine the core vocabulary. First, all words (in the case of nouns and verbs the lemma, in the case of adjectives the root, and in all other cases the orthographically distinct form with heteronyms separated) were inspected using the frequency criterion of ≥ 0.5‰. Once the words that met the criterion had been defined, the commonality score of each of these words was determined. The commonality score indicated how many participants used a particular word, and this was determined by searching the word root tables of each participant's transcript to determine the presence or absence of the word in each participant's sample. The number of participants using a word equated to the score (i.e., if one participant used the word, the commonalty score = 1). Only words that met both the frequency criterion and the commonalty criterion (commonality score of ≥ 3) were included in the Sepedi core vocabulary. These criteria for determining a core vocabulary have been used in previous studies (Trembath et al., 2007; Boenisch & Soto, 2015; Mngomezulu, 2017).

Each word presented in the core vocabulary was classified into parts of speech. This was done by locating each core word in the Oxford Pukuntšu ya Sekolo dictionary (de Schryver et al., 2007). The classification provided in the dictionary was used to categorise the words into word classes. Where necessary, the grammar books by Poulos and Lourens (1994) and van Wyk and colleagues (1992) were also consulted. The Sepedi 'part-of-speech tagger' demonstration (de Pauw & de Schryver, 2007) available online at https://www.aflat.org/sothotag and described in Chapter 2, Section 2.9, was also consulted at times. When the word was a code switch to English, the English section of the Oxford Pukuntšu ya Sekolo dictionary was consulted. The classification was undertaken by the student and verified by the supervisor and co-supervisor.

3.7.6 Reliability of transcription

To ensure reliability of transcription, similar measures as those implemented by Romski, Sevcik, Adamson, Cheslock, Smith, Barker (2010) were followed. Transcription reliability was determined by cross-checking the first transcript with the voice recording and correcting it for each participant. The transcripts were checked by a different person (researcher or research assistant) from the one who had transcribed the data before. The researcher checked the reliability of four transcripts. The other two were checked by two assistants (one with a Bachelor degree in Speech Language Pathology and Audiology and one with a post-graduate certificate in Education respectively). Prior to checking the transcripts, the research assistants were briefed on the transcription rules. They were also provided with a copy of the transcription rules.

3.7.7 Reliability of coding

Coding reliability is usually assessed by intercoder reliability, which refers to the extent to which two or more independent coders code the data in the same way (Freelon, 2013). In order to determine coding reliability, the 3 000-word transcription of each participant was divided into five equal parts (based on word count – first 600 words, next 600 words, etc.) One part (amounting to one fifth or 20% of each participant's transcription) was randomly selected to be coded a second time by an independent coder, a research assistant with a Bachelor degree in Speech-Language Pathology and Audiology. This person followed the same list of coding rules (see Appendix L). The coding by the first and second coder was then compared, and the percentage of agreement was calculated using the following formula:

$$\text{Percentage agreement} = \frac{\text{Agreements}}{\text{Agreements + disagreements}} \times 100.$$

The results per participant and for the total sample are provided in Table 3.5.

Table 3.5

Percentage Agreement of Coding of 20% of Each Participant's Transcript

Participants	1	2	3	4	5	6	Total
Percentage agreement	93.8	94.2	92.2	93.7	95.5	95.3	94.1

The percentage of agreement between coders ranged from 92.2% to 95.5% per participant, with an average of 94.1(SD = 0.89). Since percentage agreement of 80 and above can be regarded as satisfactory (McMillan & Schumacher, 2010), this represents good agreement. The coding was therefore conducted consistently and reliably.

3.7.9 Validity

The research design used in this study is susceptible to threats to validity. Regarding internal validity, observation in its own capacity poses a threat, as the participants may behave differently when observed from how they would typically behave (often referred to as participant reactivity) (Leedy & Ormrod, 2014). In order to minimise reactivity, the researcher explained to the teachers that the main aim of the study was not to assess any skills but rather to obtain an objective view of their typical behaviour. Children were also told that the aim was to help the researcher and not to test their knowledge or abilities. Moreover, the first 20 minutes of the samples were not considered in analysis, to reduce novelty effects. In addition, all references by participants to recording equipment were omitted from transcripts. A sample of 3 000 words was collected per participant in order to get a representative sample that was not limited to interactions on a single day or during a single activity. Furthermore, recording took place during the whole preschool day, rather than during specific times only. Other measures that were used to strengthen the internal validity of the study included using the same procedures for all participants and having the same examiner who provided the same instructions and preparation to the participants (McMillan & Schumacher, 2010).

The external validity of the study was affected by the small sample size (McMillian & Schumacher, 2010), which affected the ability to generalise the results of the study to the rest of the population. Although an attempt was made to introduce some variation by accessing three different sites, and including both boys and girls in the sample, the generalisability of any study including only six participants remains limited.

3.8 Summary

This chapter aimed to provide a description of the research methodology followed. It outlined the main aims and sub-aims and provided a description of the research design followed. This was followed by a description of the setting and the participants, including the recruitment, selection criteria and descriptive criteria. The materials and equipment used were also outlined. This was followed by the pilot study data, where the aims, procedures, results

and recommendations were explained. The last part of the chapter outlined the procedures for data analysis, ethical considerations and ensuring the reliability and validity of the study.

Chapter 4

RESULTS

4.1 Introduction

This chapter presents the results of the study. The results are represented in graphs and tables as guided by the sub-aims outlined in Chapter 3. The sub-aims were: (i) to identify the words and the number of different words Sepedi-speaking children without disabilities use during regular preschool activities, (ii) to determine the words most frequently and commonly used by Sepedi-speaking children without disabilities during regular preschool activities (core vocabulary), (iii) to describe the core vocabulary by parts of speech as well as by differentiating structure (grammatical) and content (lexical) vocabulary, and (iv) to compare the core vocabulary identified with that found in other studies.

4.2 Description of the sample

The researcher obtained 3 000 words (TNW) from each participant. When unintelligible words, phrases and utterances were subtracted from the total number of words, the remaining (intelligible) number of words collected per participant varied from 2 885 to 2 961. The TNW, NDW and TTR, i.e. the ratio of NDW to TNW, per participant is given in Table 4.1.

Table 4.1

TNW, NDW, and TTR per Participant

Participant Number	**TNW[a]**	**NDW**	**TTR**
P1	2 943	377	0.13
P2	2 932	400	0.14
P3	2 885	429	0.15
P4	2 961	489	0.17
P5	2 897	411	0.14
P6	2 951	438	0.15
M	2 928.2	424	0.15
SD	30.6	38.5	0.01
Total	17 569	1 023	0.06

[a]After removal of unintelligible words, phrases and utterances

The composite transcript contained 17 569 words (TNW). The NDW of the composite sample amounted to 1 023. As indicated in the coding rules, in the NDW counts, heteronyms and polysemous words were counted separately, since they have different meanings (even though they have identical orthographic representations). Inflected forms of verbs and nouns were traced back to their lemma and counted together with the lemma form (e.g., plurals and singular forms of the same noun were counted as one word) when counting NDW. Similarly, inflected forms of adjectives were counted by the root of the adjective. To protect the privacy of the participants, no proper nouns were transcribed, but were replaced with the code CN (child name), TN (teacher name), and PN (place name). All child names were therefore counted together as one word, although in reality different child names were mentioned. Similarly, all teacher names and all place names were counted together. The NDW count is thus somewhat compromised. The overall TTR is 0.06.

4.3 Core and fringe vocabulary

In order to establish the core vocabulary, two criteria were applied. First, the word had to occur in the sample with a frequency of 0.5‰ (i.e., one per 2 000) or more. Second, the word had to have a commonality score of at least 3 (i.e., the word had to have been used by at least 50% of the participants). These criteria, although somewhat arbitrary, have been used in previous studies (Mngomezulu, 2017; Banajee et al., 2003). The frequency of occurrence of each unique word in the sample was calculated by dividing the total number of occurrences of this word by the TNW in the composite sample and multiplying by 1000 (Mngomezulu, 2017). The formula used was as follows:

$$\frac{\text{Total number of occurrence}}{\text{Total number of words}} \times 1000 = \text{frequency per mille.}$$

Words were then arranged in an Excel file by their frequency score, to identify all those that had a frequency score above 0.5‰, which amounted to 241 words complying with this criterion. These words were then inspected for commonality, and it was found that 226 met the criterion of ≥3, while 15 words had a score of ≤ 2. Therefore, a core vocabulary of 226 words was established. This core word list with accompanying frequency and commonality scores is provided in Appendix M. The appendix also contains information about inflected forms of core words that were found in the sample. More detail pertaining to the commonality scores of the core vocabulary included is given in Table 4.2.

Table 4.2

Commonality Distribution among Core Words

Number of different words	Commonality score
84	6
46	5
46	4
50	3

The frequency counts of the 226 words were summed, and amounted to 881.21‰, or 88.1%. This constitutes the coverage of the core vocabulary – meaning that 88.1% of the words used during conversations were core words. Core words therefore 'covered' 88.1% of the conversations.

The remaining 797 words were designated as fringe words. Although these words were considerably higher in number when considering NDW, their coverage only amounted to 11.9%. The NDW and coverage of core versus fringe vocabulary are displayed in Figures 4.1 and 4.2 respectively.

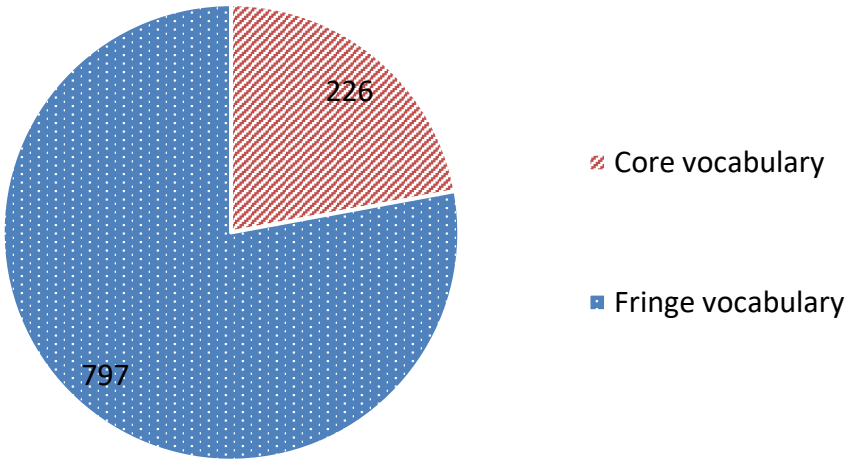


Figure 4.1. Number of different words constituting the core and fringe vocabulary.

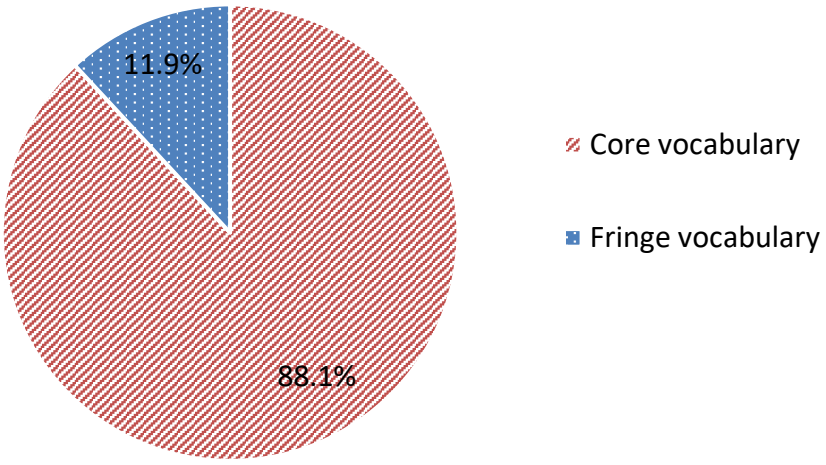


Figure 4.2. Coverage of different words constituting the core and fringe vocabulary.

4.4 Further description of the core vocabulary

4.4.1 Classification into content and structure words

In an attempt to analyse the core vocabulary further, the sample was classified in terms of content versus function words. The content words are words that carry meaning. These words include verbs, nouns, adverbs, and adjectives (Shi et al., 2006). Structure words (also sometimes referred to as function words) are words that create structural relationships in which content words may exist/fit. These words include prepositions, conjunctions, and particles. Figure 4.3 gives an indication of the number of different content and structure words constituting the core vocabulary.

Chapter 4: Results

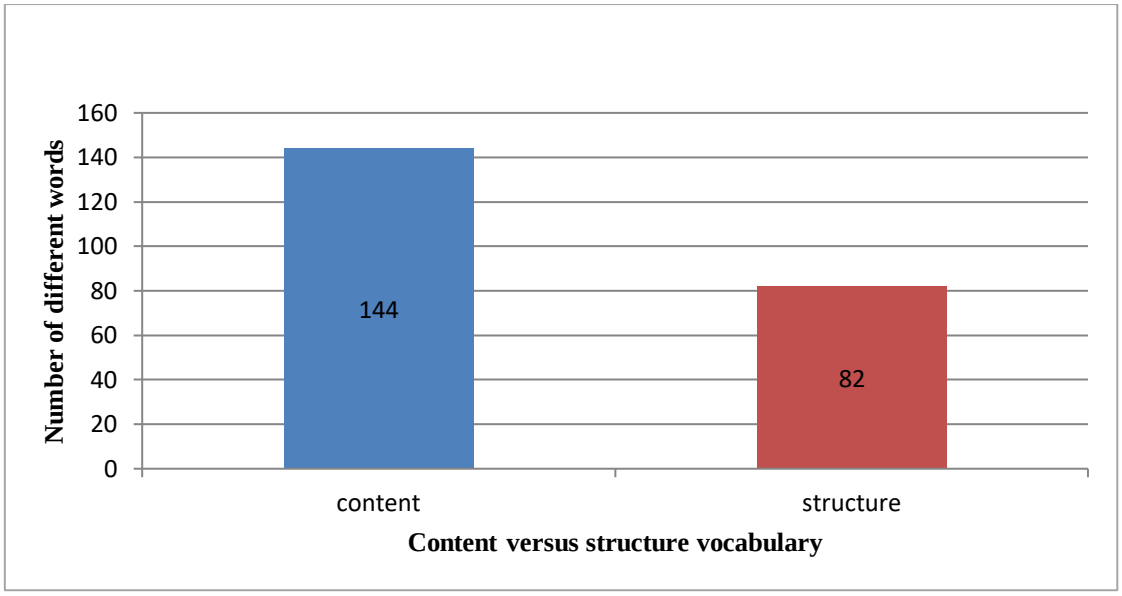


Figure 4.3. Number of different content and structure words within core vocabulary.

A hundred and forty-four different content words constituted 64% of the core vocabulary, while 82 structure words accounted for the remaining 36% of the core vocabulary.

The percentage of content versus structure words within the top (most frequently used) 20, 50, 100, 200 and total number of core words was also determined. The comparison is provided in Figure 4.4.

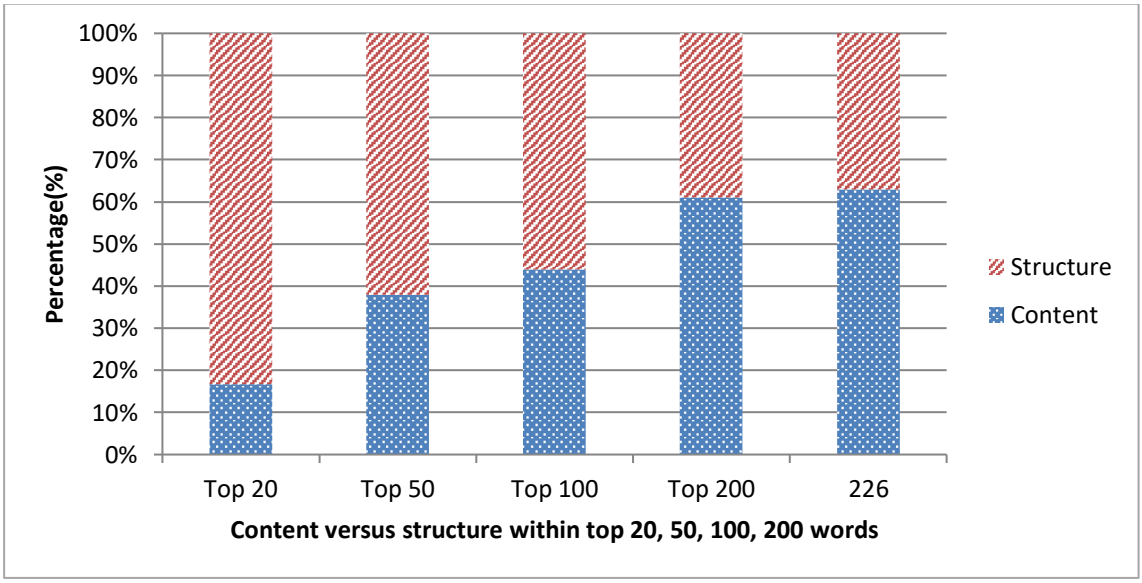


Figure 4.4. Percentage of content versus structure words within the top 20, 50, 100, 200 and total core vocabulary words.

Figure 4.4 shows that the percentage of content words increases as the number of words increase. For example, within the top 20 most frequently used words, structure words constitute 85% of the words, whereas content words represent 15%. However, when considering all 226 core words, the bulk of these words are content words (64%).

The coverage (frequency of use) of content core versus structure core words was also compared. Figure 4.5 shows the coverage of content core, structure core, and fringe vocabulary.

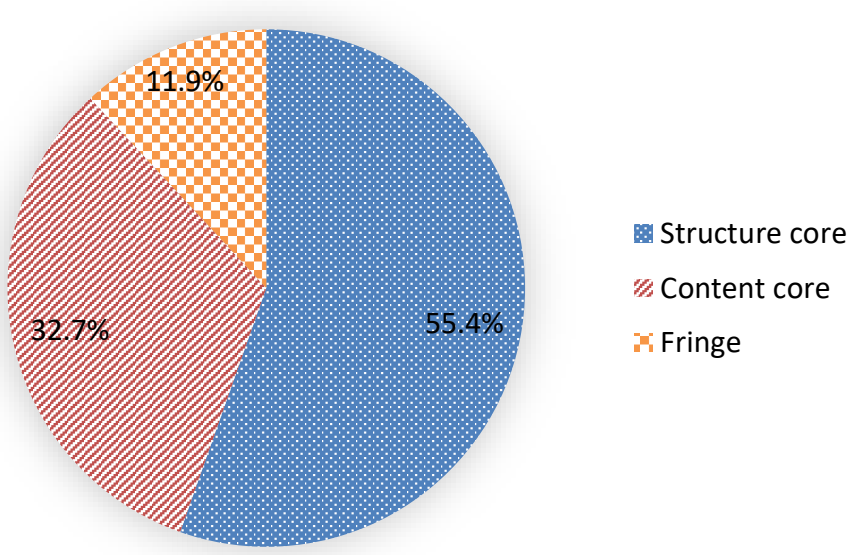


Figure 4.5. Coverage of different words constituting the content core vocabulary, structure core vocabulary and fringe vocabulary.

From Figure 4.5, it is clear that structure core words were used with a frequency of 55.4%, while content core words were used with a frequency of 32.7% in the sample. Therefore, although there were a higher number of different content words in the core vocabulary (n = 144), their coverage was lower than that of the 82 structure core words. The remaining 11.9% of the sample consisted of fringe words.

4.4.2 Classification by parts of speech

To describe the core vocabulary further, core words were classified into parts of speech. This was done by locating each core word in the Oxford Pukuntšu ya Sekolo dictionary (de Schryver, 2007). The classification provided in the dictionary was used to categorise the words into word classes. Where necessary, the grammar books by Poulos and

Lourens (1994) and van Wyk and colleagues (1992) were also consulted. The Sepedi 'part-of-speech tagger' demonstration (de Pauw & de Schryver, 2007) available online at https://www.aflat.org/sothotag and described in Chapter 2, Section 2.9, was also consulted at times. When the word was a code switch to English, the English section of the dictionary was consulted. The classification was undertaken by the student and verified by the supervisor and co-supervisor. The number of different core words falling into the different parts of speech, the proportion of each part of speech category within the total core vocabulary, as well as the number and frequency with which each part of speech category in the core vocabulary appeared in the sample were calculated. The results are given in Table 4.3 below. The table is arranged from most to least frequently occurring part of speech.

Table 4.3

Parts of Speech Occurring in the Core Vocabulary

Parts of speech	**NDW**	**Proportion in core (in terms of NDW)**	**Number of occurrences in sample**	**Frequency of occurrences ‰**
Concords	18	8%	4324	246.1
Verbs	83	36.7%	3214	182.9
Nouns	49	21.7%	2173	123.7
Pronouns	24	10.6%	1717	97.7
Interjections	14	6.2%	673	38.3
Negative morphemes	3	1.3%	519	29.5
Conjunctions	7	3.1%	518	29.5
Copulative particle	1	0.4%	444	25.3
Present tense morpheme	1	0.4%	416	23.7
Prepositions	1	0.4%	278	15.8
Future morphemes	3	1.3%	276	15.7
Adverbs	6	2.6%	238	13.5
Demonstrative particle	1	0.4%	170	9.7
Adjectives	6	2.6%	123	7.0
Hortative particle	1	0.4%	120	6.8
Locative particles	3	1.3%	106	6.0
Potential morphemes	1	0.4%	70	3.9
Aspectual prefixes	2	0.9%	55	3.1
Infinite prefix	1	0.4%	28	1.6
Question particles	1	0.4%	23	1.3
Total	226	100%	15 485	881.4

Chapter 4: Results

The NDW categorised by parts of speech is also shown in Figure 4.6. Figure 4.7, in turn, shows the frequency with which each category was used in the total sample.

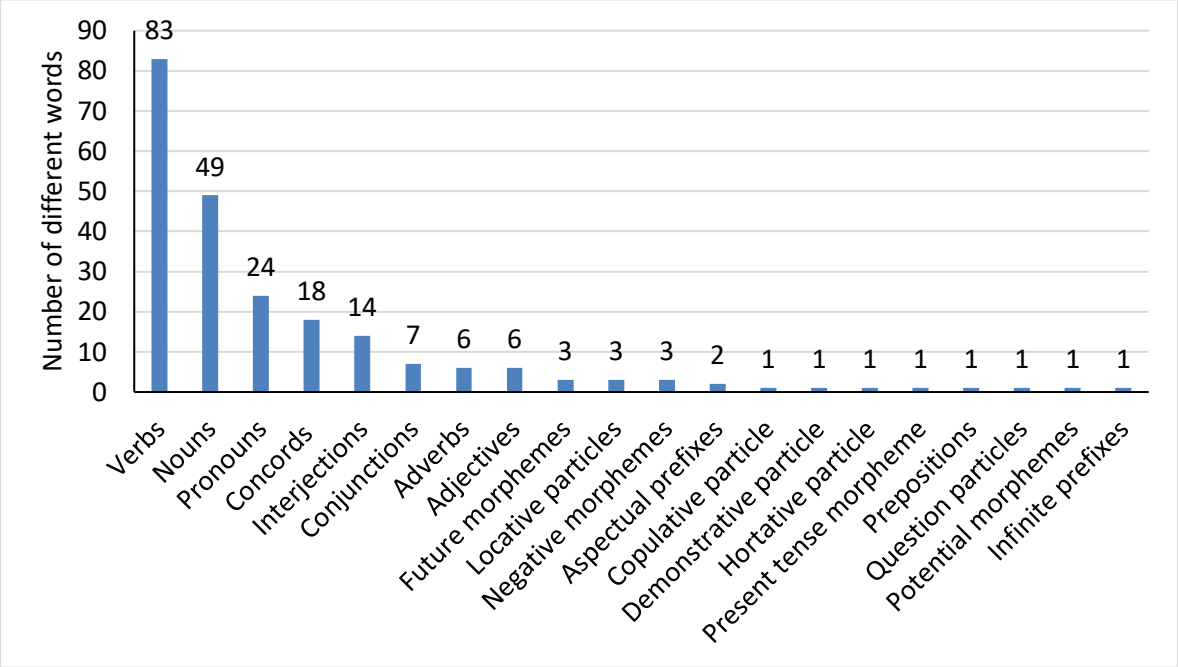


Figure 4.6. NDW constituting the different parts of speech within the core vocabulary.

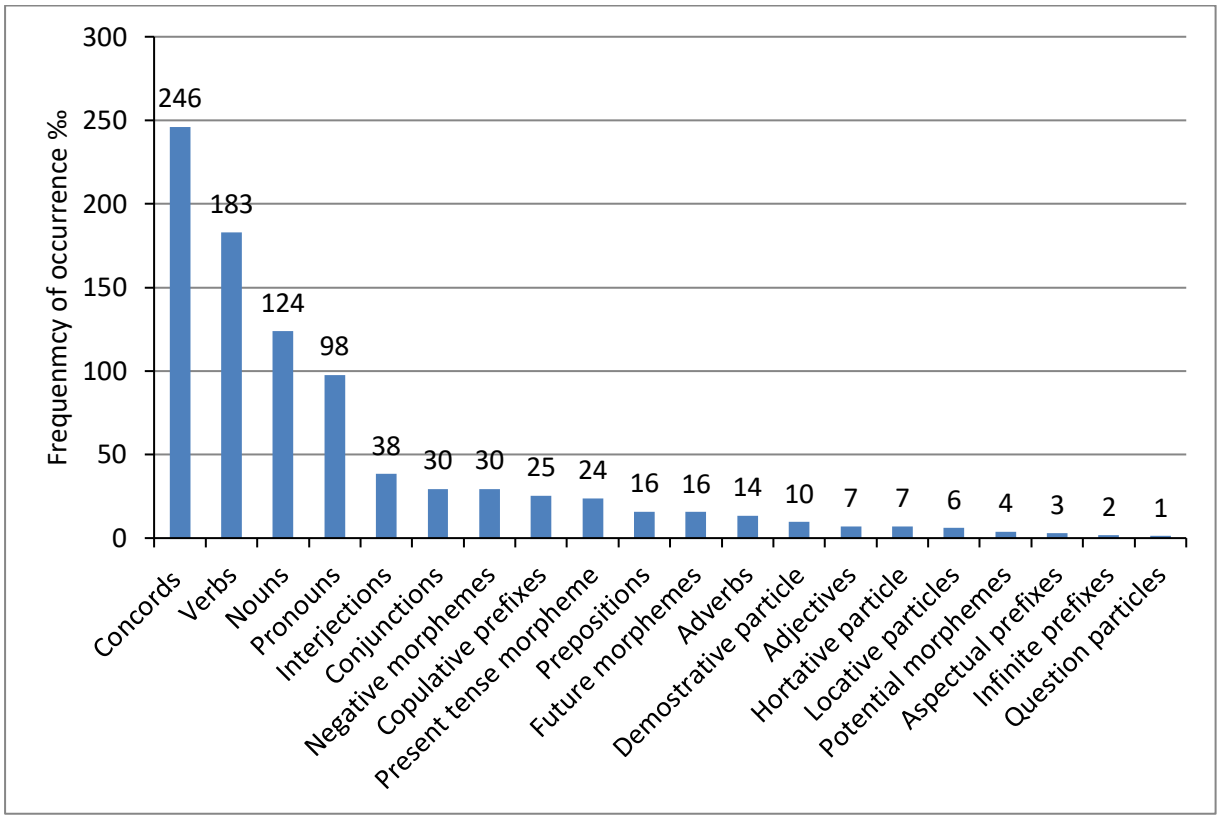


Figure 4.7. Frequency of occurrence (in ‰) of different parts of speech in the core vocabulary.

Table 4.3 and Figure 4.6 clearly show that the core vocabulary contained a high number of different verbs (83 different verbs), nouns (49 different nouns), and pronouns (24 different pronouns). In contrast, the core vocabulary only contained one of each of the following parts of speech: copulative particles, demonstrative particles, hortative particles, present tense morphemes, prepositions, question particles, potential morphemes and infinite prefixes.

When considering frequency of use (see Table 4.3 and Figure 4.7), it appears that verbs, nouns and pronouns also featured relatively highly, being the second to fourth most frequently used parts of speech. However, concords were the most frequently used part of speech overall. Taken together, these four parts of speech were used with a frequency of 650.4‰ during the participants' recorded conversations. It is clear that the number of different words contained per part of speech does not necessarily predict the frequency of use of that part of speech. For example, 83 verbs covered 182.3‰ of the sample, whereas 18 concords covered 246.1‰.

Results pertaining to the 20 different parts of speech will now be described in more detail. For the sake of this description, certain parts of speech will be grouped together. The preposition and locative particle will be grouped together. Certain verbal prefixes and suffixes will also be grouped, namely the future morpheme, present tense morpheme, potential morpheme, infinite prefix, aspectual prefix, copulative particle, question particle, demonstrative particle and hortative particle.

4.4.2.1 Verbs

The core vocabulary contained 83 different verbs, which were used with a frequency of 182.9‰. These words, which form part of the content core vocabulary, appeared 3 214 times in the sample. All core verbs appeared in the imperfect tense. All core verbs except for seven occurred in other inflected forms as well. Fourteen other inflected forms were found. The most frequently occurring inflection other than the imperfect tense was the negative form of the imperfect tense, with 64 verbs appearing in this form. Figure 4.8 shows how many core verbs were used in different inflected forms. For ease of reference the affirmative and negative forms of a particular inflection (when both occurred in the sample) are grouped together.

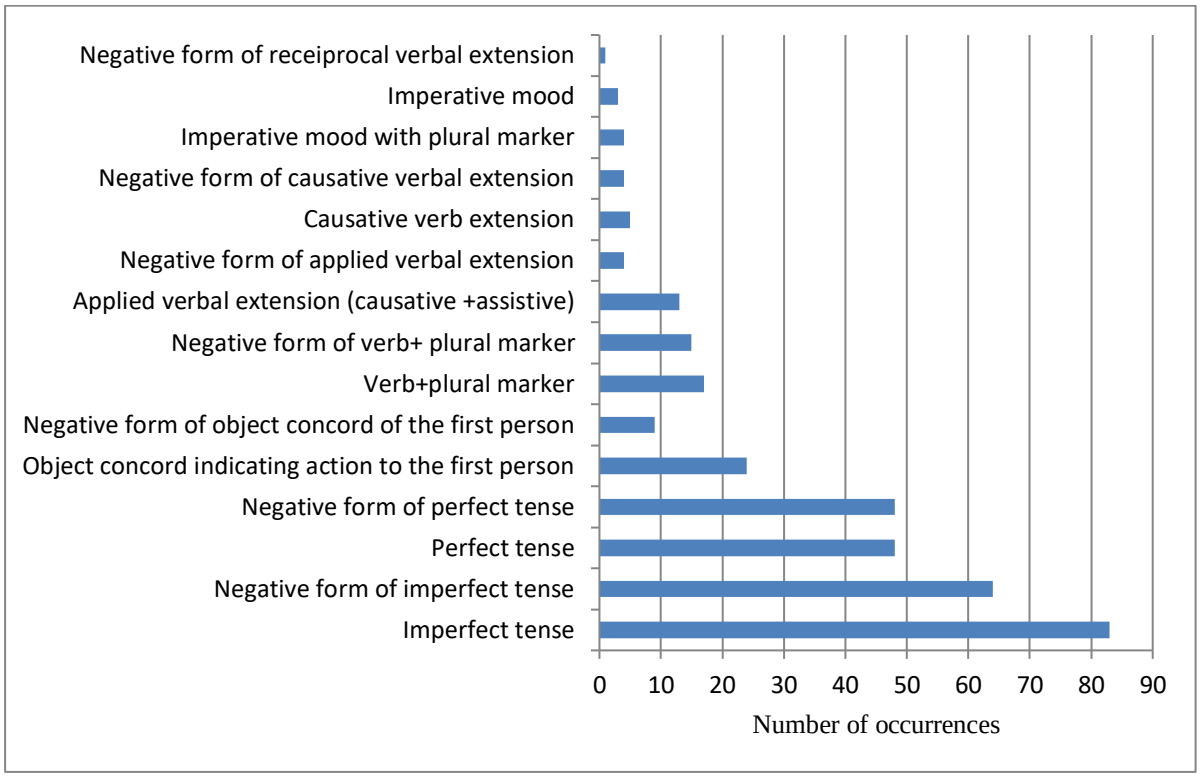


Figure 4.8. Number of occurrences of core verb roots versus different inflected forms.

The most frequently used verb was *ya* (which means go), occurring 276 times and being used by all the participants. *Ya* was the 15th most frequently used word in the core vocabulary list. The verb was used 194 times in the imperfect indicative (lemma) form. It was also used in eight other inflected forms, namely the negative form of the imperfect indicative, perfect tense, negative form of the perfect tense, negative form of object concord indicating action to the first person, verb and plural marker, negative form of the verb and plural marker, imperative mood and the imperative mood with the plural marker.

4.4.2.2 Nouns

There were 49 nouns in the core vocabulary. These were used with a frequency of 123.7‰. Nouns are content words. Nouns occurred 2 173 times in the sample. The two most frequent nouns in the sample were CN (which referred to the children's names), occurring 760 times, and TN (referring to teacher names), occurring 118 times. These nouns were used by all the participants. CN was the second most frequently used word in the core vocabulary

list. It is, however, clear that both CN and TN were over-counted, since these codes really represent different names. The most frequently used noun apart from CN and TN was *eng*, which functions similar to the English question word 'what.'

The core vocabulary contained singular nouns and nouns that do not have a plural form (e.g., *fase*), as well as plural nouns, and locativised singular nouns. No diminutives were found in the core vocabulary. A total of 48 nouns occurred in the singular form. These forms of nouns occurred 1 969 times. Nineteen plural nouns were found in the core vocabulary (of these, 18 appeared in the singular form as well). One noun, *dijo*, appeared only in the plural form. Plural nouns that formed part of the core vocabulary appeared 179 times in the sample. Five nouns appeared in the locativised form, and these locativised nouns were used 25 times in the sample.

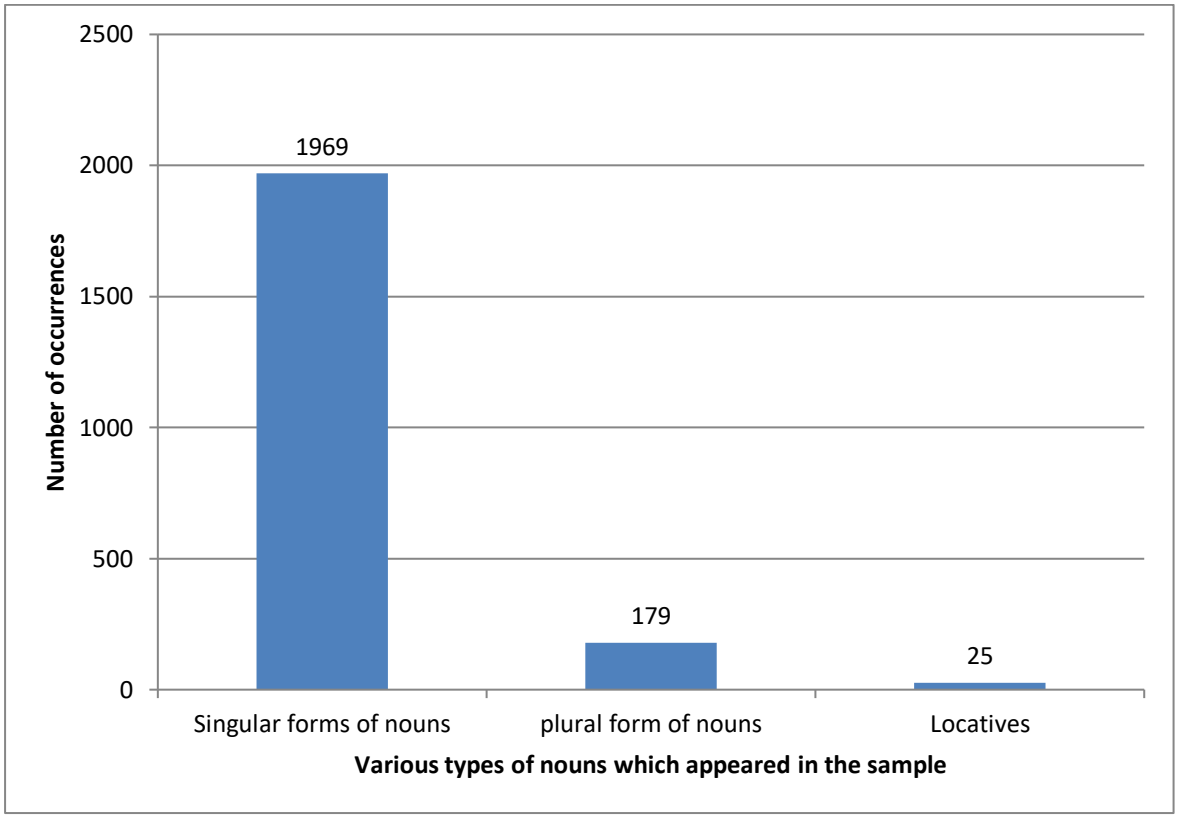


Figure 4.9. Number of occurrences of singular forms of nouns versus different inflected forms.

4.4.2.3 Concords

Eighteen different concords were found in the core vocabulary list. These were used with a frequency of 246.1‰. These words are structure words. Concords occurred 4 324

times in the sample. In this study, no distinctions were made between subject, object and possessive concords, as many concords can function as more than one type of concord (e.g. *o* can be a subject or an object concord). The most frequently used concord was *o,* which appeared 1 046 times and was used by all the participants. *O* was the most frequently used word in the core vocabulary list.

4.4.2.4 Pronouns

There were 24 different pronouns in the core vocabulary that were used with a frequency of 97.7‰. These words are structure words. Pronouns occurred 1 717 times in the sample. The most frequently used pronoun was *nna,* which translates to 'I'—this appeared 381 times and was used by all the participants. *Nna* was the seventh most frequently used word in the core vocabulary list.

4.4.2.5 Interjections

Fourteen different interjections were identified in the core lists; these were used with a frequency of 38.3‰. These words are structure words. Interjections occurred 673 times in the sample. The most frequently used interjection was *ah,* which appeared 125 times and was used by all the participants. *Ah* was the 28th most frequently used word in the core vocabulary list. This interjection was used to show the emotion of surprise.

4.4.2.6 Conjunctions

Seven different conjunctions were used with a frequency of 29.5‰. These words are structure words. Conjunctions occurred 518 times in the sample. The most frequently used concord was *le,* which translates to 'and'/'together with'. It appeared 335 times and was used by all the participants. *Le* was the eighth most frequently used word in the core vocabulary list.

4.4.2.7 Adverbs

Six different adverbs were found in the core lists; these were used with a frequency of 13.5‰. These words are content words. Adverbs occurred 238 times in the sample. The most frequently used adverb was *so* (which was used as *like this*). The Sepedi language includes some loan words from English and Afrikaans (Nong, de Schryver, & Prinsloo, 2002). The word *so* is borrowed from the Afrikaans word 'so' which means *in that way/such* (Oup Elt, 2017). Since it is pronounced in a different way than the Afrikaans word, it is not considered

a code switch. The word appeared 88 times and was used by all the participants. *So* was the 41st most frequently used word in the core vocabulary list.

4.4.2.8 Adjectives

Six different adjectives were used with a frequency of 7.0‰. These words are content words. Adjectives occurred 123 times in the sample. Adjectives were counted under their uninflected root. The most frequently used adjective root was *ngwe* (other), which appeared 43 times and was used by all the participants. *Ngwe* was the 70th most frequently used word in the core vocabulary list. Adjectives are typically inflected according to the noun class of the noun they describe. All six adjectives were found in various inflected forms. The number of times different inflected forms of the adjectival root *ngwe* appeared in the sample is illustrated in Figure 4.10 below.

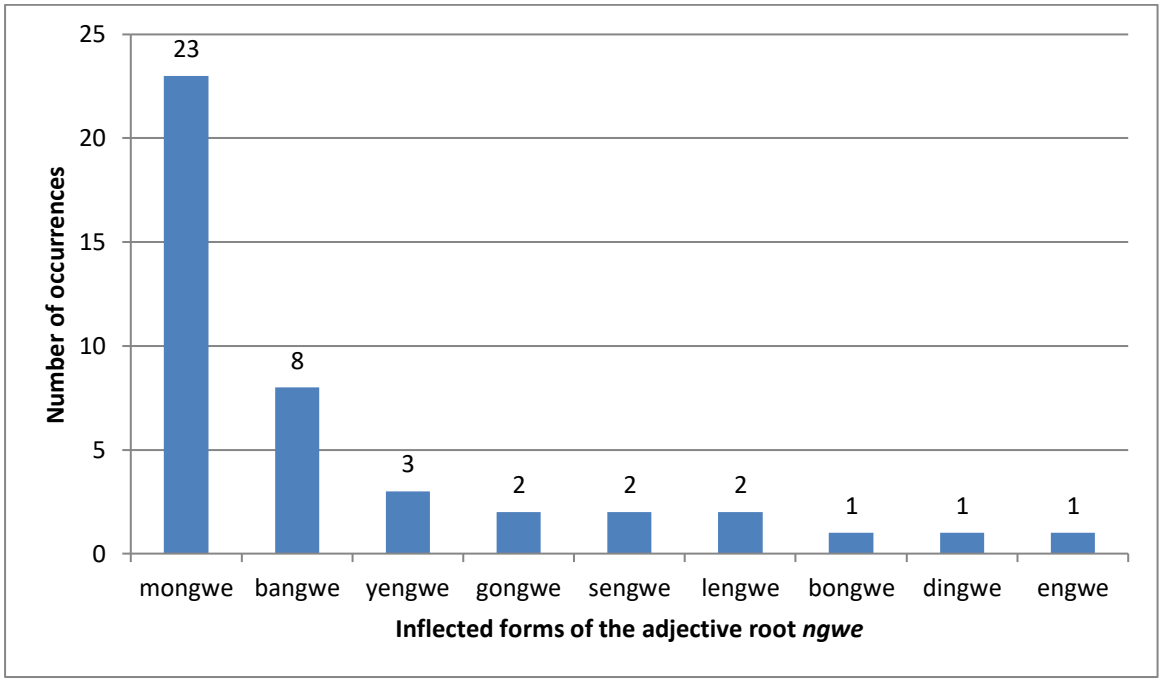


Figure 4.10. Number of occurrences of the inflected forms of the adjective root *ngwe*.

4.4.2.9 Verbal prefixes and suffixes

In Sepedi, various verbal prefixes and suffixes are written disjunctively from the rest of the verb and were therefore counted separately in this study. These included two different negative morphemes, two aspectual prefixes, and one of each of the following: copulative particle, demonstrative particle, hortative particle, present tense morpheme, future morpheme, question particles, potential morpheme and infinite prefix. They appeared with a frequency

ranging between 1.3 and 21.7‰, and accounted for an overall frequency of 12.1% in the sample. Each of these words modifies the meaning of the verb in the sentence to some extent. For example, the infinite prefix *go* transforms the verb into an infinitive, for example, *go robala* (to sleep).

4.4.2.10 Prepositions and locative particles

Prepositions and locative particles were also identified in the core vocabulary. There were one proposition and three locative particles. The preposition was used with a frequency of 15.8‰ and the locative particles were used with a frequency of 6.0‰. These parts of speech form part of structure words. The preposition was *ka*, which means 'with' or 'about'. It appeared 278 times in the sample. *Ka* was the 14th most frequently used word in the core vocabulary list. The locative particles appeared 106 times in the sample. The locative particle that appeared with the highest frequency was *kua*, as it appeared 47 times. *Kua* was the 69th most frequently used word in the core vocabulary list. It means 'there'.

4.4.2.11 Code switches

Although not technically a part of speech, code switches are also described here, as a number of them occurred in the sample. Code switches were identified when words from another language were used in the utterances. These words were used without any phonological or morphological changes. The code switches found in the core were *why*, *jersey* and *toilet*. The frequency with which these occurred in the sample was 15‰.

4.5 Comparison of core vocabulary to core vocabularies established in other languages

To explore the similarities and differences between the current core vocabulary list and those found in other languages further, comparison of the Sepedi core vocabulary to an isiZulu core vocabulary (Mngomezulu, 2017) and also to three English core vocabulary lists (Boenisch & Soto, 2015; Trembath et al., 2007) was undertaken.

4.5.1 Comparison to isiZulu core list (Mngomezulu, 2017)

Since the isiZulu core vocabulary study conducted by Mngomezulu (2017) analysed the language samples by the frequency of formatives (on a morphological level), and the current study analysed the vocabulary by orthographic words, direct comparisons are complicated, since the unit of analysis is not comparable. However, it was possible to

compare the percentage of occurrences of specific parts of speech that were defined similarly in the two studies. The results are given in Figure 4.11.

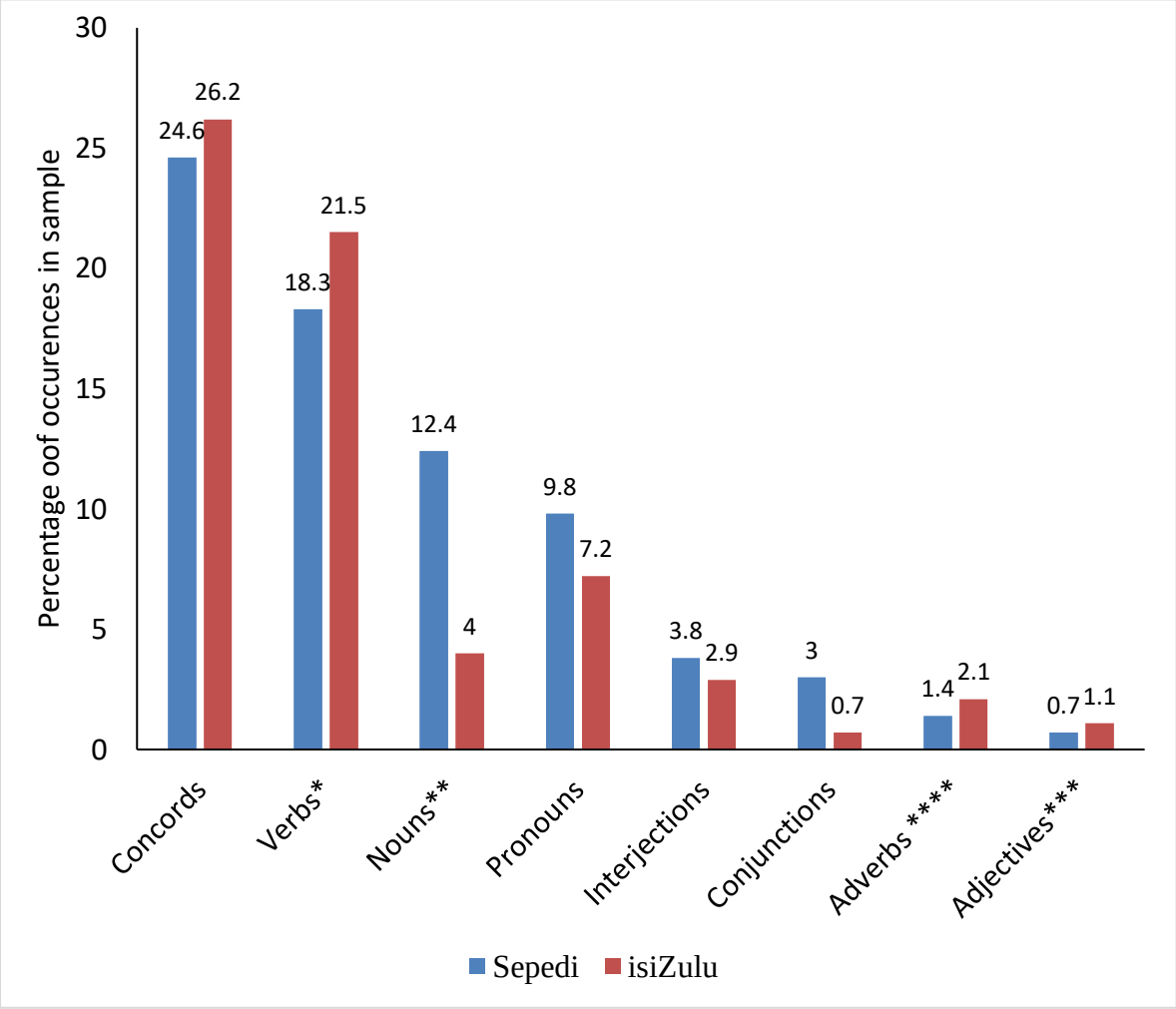


Figure 4.11. Percentages of occurrences of parts of speech between the two studies

Note. *Verb roots and verbal auxiliaries in the isiZulu study; ** Noun roots in the isiZulu study; *** Adjective roots in the isiZulu study; ****Adverbial roots in the isiZulu study.

It is interesting to note that concords and verbs were found to be the most and second-most frequently occurring parts of speech in both studies. Pronouns also occurred frequently, although nouns occurred more frequently in the Sepedi study. Interjections, conjunctions adverbs and adjectives occurred with a lower frequency.

As noted, because of different units of analysis, word-for-word comparisons were difficult. At present, the authors are not aware of a Sepedi-isiZulu dictionary that could facilitate comparisons. However, Mngomezulu (2017) translated certain formatives in her

core vocabulary (particularly those she designated as content formatives) into English. These included noun roots, verb roots, pronouns, adverbs and adverb roots, adjective roots and conjunctions, amounting to a total of 131 different formatives. These formatives tended to have equivalents in Sepedi on the level of orthographic words (see also Chapter 2, Section 2.7.2.2, for an explanation of linguistic versus orthographic words) and were therefore more easily comparable than the other formatives found in Mngomezulu's (2017) study. Therefore, the equivalent parts of speech in the Sepedi core list (nouns, verbs, pronouns, adverbs, adjectives and conjunctions) were also translated into English. The two translated lists were compared per part of speech to determine the number of overlapping words. The results are presented in Table 4.4. Further details are provided in Appendix N.

Table 4.4

Summary of the Findings of the Comparison of Sepedi Core Words and Core Words Found in the isiZulu List

Part of speech	Number of isiZulu formatives	Number of Sepedi words	Number of units that overlapped in meaning	Number of units that did not overlap in meaning	Percentage overlap
Verbs	62	83	84	62	57.9
Nouns	32	49	27	54	33.3
Pronouns	13	24	18	17	48.6
Adjectives	11	6	8	9	47.1
Adverbs	8	6	2	12	14.3
Conjunctions	5	7	6	6	50.0
Totals	131	175	145	160	41.9

It is clear from the table that there was some overlap in the content of the isiZulu and the Sepedi core vocabulary lists, with more overlap among verbs and conjunctions than among other parts of speech.

4.5.2 Comparison to English core vocabulary lists

This comparison was done to understand the semantic similarity of the current core vocabulary list with that of some English lists. This was done in a two-step process. Firstly, the top 100 words of the Sepedi core vocabulary obtained were translated into English

according to the translations given by the Oxford Pukuntšu ya Sekolo dictionary (de Schryver, 2007). For those Sepedi words in the list that had more than one translation, all translations were given. The translations of the top 100 Sepedi words according to the dictionary were then matched to the top 100 words of three English core vocabulary lists – particularly lists compiled from speech samples of preschoolers and published in 1998 or later (i.e., not more than 20 years ago). These lists were identified from the studies consulted by Hattingh (2018) as described in Table 2.1 to compile the English composite list (as described in Section 2.6). The sources are described in Appendix O. A Sepedi word was considered to have an equivalent word in the English lists if at least one of its translations was identical to a word in the English list. Furthermore, if it was found that the Sepedi word was translated by an English phrase, then all the words of the phrase had to be present in the English list in order for it to be considered an equivalent. Appendix O gives the complete results of the comparison, while Table 4.5 indicates the number of words in the top 100 Sepedi core words that had equivalents in the top 100 most frequently used words of the three English core vocabulary lists. It was found that 55 of the top 100 Sepedi words had an equivalent in all three English comparison lists.

Table 4.5

Summary of the Findings of the Comparison of Top 100 Sepedi Core Words and Top 100 Core Words from Three English Lists

Number of words in the Sepedi List	**Number of English lists in which equivalent words were found**
55	3
9	2
8	1
28	0

4.6 Summary

The chapter presented the results of the study according to the sub-aims of the study. First, the TNW and NDW that were identified in the samples collected from Sepedi-speaking children during regular preschool activities were determined. Second, a Sepedi core vocabulary consisting of 226 words was established by applying both a frequency and a commonality criterion. It was found that these 226 words covered 88.1% of the children's

conversations. Third, the core vocabulary was further described by classifying core words as either content or structure words. It was found that 144 different content words constituted 64% of the core vocabulary, while 82 structure words accounted for the remaining 36% of the core vocabulary.

Core words were also classified into different parts of speech. It was found that the core vocabulary contained a high number of different verbs (83 different verbs), nouns (49 different nouns), and pronouns (24 different pronouns). In contrast, the core vocabulary only contained one of each of the following parts of speech: copulative particles, demonstrative particles, hortative particles, present tense morphemes, prepositions, question particles, potential morphemes and infinite prefixes. Regarding the frequency with which different parts of speech were used, verbs, nouns and pronouns were used relatively frequently; however, the 18 copula represented the most frequently used part of speech overall.

Lastly, the Sepedi core vocabulary was compared to an isiZulu core vocabulary list and also to three English core vocabulary lists. Similarities and differences were found.

Chapter 5

DISCUSSION

5.1 Introduction

In this chapter, the results of the study are discussed. The composite Sepedi speech sample was compared to that of other studies. The parameters (number and coverage) of the identified Sepedi core vocabulary were compared to those found in other studies. The proportions of core versus content vocabulary, as well as the composition of the core vocabulary by parts of speech, are discussed with reference to the implications for vocabulary selection for AAC systems. Finally, the results of comparisons of the lexical meanings of the Sepedi core vocabulary to an isiZulu list and three English lists are discussed with reference to the implications for translatability of the core vocabulary.

5.2 The composite speech sample and its parameters

The combined speech sample contained a total number of 17 569 words (TNW). The NDW of the combined speech sample amounted to 1 023. The TTR was also determined. This ratio is used to indicate the number of unique words in the TNW in the combined speech sample. The TTR was ~0.06. Since the ratio is related to the number of unique words (i.e. the higher the number of unique words, the higher the ratio), it follows that the participants in the study re-used words often, since the total number of words used was substantially higher than the unique number of words.

This phenomenon is also observable in other studies where similar size samples were collected, and where comparable TTRs were found. Boenisch and Soto's (2015) study, for example, yielded a TTR of ~0.06 on a composite sample of 19 885 words from eight English second language speakers, while Trembath et al.'s (2007) composite sample of 18 000 words (spoken by six English-speaking Australian children) contained 1 411 unique words (TTR = ~0.08). Although Mngomezulu's (2017) study on six isiZulu-speaking children identified the most frequently used morphemes or formatives rather than the most frequently used orthographic words, the TTR of ~0.06 on a composite sample comprising 20 137 formatives is still similar to that found in the current study.

It is interesting that the TTRs in the studies are so similar. One contributing factor may have been the age of the participants. This study, along with that of Mngomezulu (2017), included participants of age five to six years. Trembath et al. (2007) included children aged three to five years. The similar ages may have prompted a similar TTR, since it has been suggested that TTR is linked to age (Richards, 1987). Although Boenisch and Soto's (2015) participants were older (7-14 years old), they were second-language speakers, and this may have led to a similar TTR to that found in the studies involving younger children.

For AAC system design, the TTR suggests that it is possible to find words that are re-used often in conversations, and to include these on the AAC system.

5.3 Sepedi core vocabulary and its parameters

By applying the frequency and commonality criteria, a core vocabulary of 226 words covering nearly 90% of the speech sample was identified. The remaining 797 words (fringe vocabulary) were used with a frequency of just over 10%. This finding illustrates that while speakers do re-use words, they reuse a small set of words continually (across speakers and different preschool activities), confirming the existence of a core vocabulary as defined in the field of AAC (Trembath et al., 2007; Robillard et al., 2014). This core vocabulary accounts for a large proportion of spoken conversations, while the fringe vocabulary is used much less frequently. Inclusion of these very frequently used words in an AAC system may be beneficial to assist the person using the system to access relevant words to construct sentences for spoken conversations.

The proportional coverage of core to fringe vocabulary found in this study is similar to that found in other studies in other languages. For instance, Mngomezulu (2017) found that 221 isiZulu formatives in the sample also covered nearly 90% of conversations. English studies have typically found core-fringe ratios closer to 80%:20%, giving rise to the so-called 80%-20% rule. This rule deduces that approximately 80% of speech samples comprise core vocabulary, while the remaining 20% comprises fringe vocabulary (Baker, Hill, Devylder, 2000). Trembath et al. (2007), for example, 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. It is interesting to note that

slightly higher ratios were found in Sepedi and isiZulu (nearly 90% to 100%). This may in part be accounted for by a different linguistic structure of the language – African languages tend to be agglutinating with frequent usage of morphemes, whereas English is primarily analytic, with fewer morphemes.

5.4 Sepedi core vocabulary and its classification

5.4.1 Content versus structure grouping

The core vocabulary list was further classified into content versus structure words. It was found that the core vocabulary was composed of 144 different content words (64% of NDW of core) and 82 structure words (36% of NDW of core). These findings are comparable to those of Boenisch and Soto (2015) who determined a core vocabulary list of 300 words. About 68% of these words were content words and 32% were structure words. The isiZulu study by Mngomezulu (2017) also found that content formatives (amounting to 151) were more in number than structure formatives (amounting to 66). The classification of words in terms of content or structure was done similarly in the studies highlighted above and the current study. The three studies deemed content words to be verbs, nouns, adjectives and adverbs.

When comparing the findings of this study to those of other studies, there seems to be a variation. Trembath et al. (2007) and Robillard et al. (2014) indicated that the core vocabularies identified in their respective studies consisted mostly of structure words rather than content words. However, these authors did not discuss how the classification was done, and the exact proportions of content and structure vocabulary were not outlined.

Although there were fewer structure words in the Sepedi core vocabulary, they covered more than half (55.4%) of the total sample, while the content core vocabulary covered 32.7% (with the remaining 11.9% covered by fringe words). A similar pattern was found by Mngomezulu (2017), who found that structure formatives covered 54.2% of her sample, while content formatives covered 34.7% (with the remaining 11.1% being covered by the fringe formatives). None of the other studies documented this aspect of the results.

The comparison of content and structure coverage was also explored in the top 20, 50, 100, 200 and the total sample (Figure 4.4 in Chapter 4). It was found that the occurrence of

structure words is highest among the top 20 core words, followed by the top 50 words and the top 100 words, but decreases systematically as the top 200 and all core words are considered. This highlights the importance and prominence of structure words in the core vocabulary, given their high coverage, as well as their prominence in the top most frequently used words.

AAC systems often contain primarily content words, with an absence of structure words (Trembath et al., 2007). This could be related to the way vocabulary is selected for AAC systems and the issue of how these words are represented. For example, content words (particularly verbs and nouns) have more concrete referents than structure words, which makes it easier to represent them graphically compared to structure words, which are usually more abstract in nature and are less referential than content words (Mngomezulu, 2017). Also, when informants (e.g., parents or other AAC team members) select vocabulary for a particular AAC system, content words may be easier to think of because of their concrete nature. Since structure words often do not carry lexical meaning in themselves (e.g. the word *tlo* meaning 'shall' or 'will'), they may not come to mind when thinking of words needed for expression. However, structure words have been found to be important, as these are words that make sentences grammatically correct. In English, words such as 'of' and 'the' allow for sentence building and permit the expression of more sophisticated and nuanced meanings.

However, structure vocabulary alone would not provide access to a generic AAC system that enables the person using it to create novel messages. Content vocabulary is needed, as content words add the specific details of subjects, verbs (actions) and objects, as well as descriptive details in the form of adjectives and adverbs. The core vocabulary included 144 content words. While most content words seem quite generic and not context-specific (e.g., *selo* - 'thing', *bona* – 'see', *nyaka* – 'want), a few do seem to reflect the preschool context (e.g., *raloka* – 'play, *sekolo* – 'school') and specific preschool routines (e.g. the code switch *toilet*). Similar findings were made in previous studies, where context did seem to have a small influence on specifically nouns and verbs in the core vocabulary. For example, the word *ngwenya*, which means 'crocodile', appeared in the isiZulu core list by Mngomezulu (2017), as the children read a story about a crocodile at school. The words *Spiderman, swing* and *plane* appeared in the core vocabulary list compiled by Trembath et al. (2007). The appearance of *Spiderman* in the core vocabulary was attributed to the study coinciding with the release of the feature film, Spiderman. The children in the preschools at which Trembath et al. (2007) conducted the study showed great interest in and enthusiasm for

the film. The appearance of the words *swing* and *plane* could be explained by the presence of swings on the playground and the fact that one of the preschools at which the study was conducted was near an airport. This shows the impact of time and context on core vocabulary.

5.4.2 Sepedi core vocabulary and parts of speech

The Sepedi core vocabulary identified in this dissertation found 20 different parts of speech. These parts of speech included those familiar to those from an English language background (i.e. verbs, nouns, conjunctions, interjections, adverbs, adjectives) and a few different groups (such as concords, hortative particles, demonstrative morphemes, aspectual prefixes and past tense morphemes) that were added as they are specific to various African languages belonging to the same language family as Sepedi (e.g., Sesotho and Nguni languages).

The most frequently use part of speech was concords, although only 18 different concords were found in the core. This part of speech forms part of structure vocabulary, and a limited number of concords exist in the Sepedi language (closed class words). These include subject, object, and possessive concords corresponding to each of the 18 noun classes. Subject concords form part of nearly all sentences, with some exceptions such as imperatives (commands) or elliptical sentences. Concords have an important grammatical function. Subject concords, for example, link the subject to the verb. Concords are a feature of other Sesotho and Nguni languages, and Mngomezulu (2017) also found a limited number of concords being used with a very high frequency (highest frequency among all parts of speech). It is clear that concords play a prominent role in the language, and their inclusion in an AAC system would allow the person using the system to build grammatically complete sentences. The high frequency of a part of speech such as concords (a category that does not exist in many other languages, e.g., English) once again underlines the importance of language-specific studies to identify core vocabulary.

Verbs, nouns and pronouns were also frequently used parts of speech, with a high number of different words falling into these categories. Similar findings were made in other studies. Boenisch and Soto (2015), for example, found that these three parts of speech accounted for over half (nearly 60%) of the 300 most frequently used words of native and ESL speakers, whereas eight other parts of speech made up the remaining proportion. In the

current study, these three parts of speech accounted for nearly 70% of the core words (see Table 4.3).

In the current study, 83 different verbs were found in the core vocabulary, accounting for nearly 40% of the core words. Prominence of verbs was also found in other core vocabulary studies, where verbs were found to be the part of speech that dominated the core vocabulary. For example, Fallon and colleagues (2001) found that verbs accounted for 29% of their core vocabulary, while Boenisch and Soto (2014) found that 26 to 28% of their core vocabularies consisted of verbs. Shin and Hill (2016) found that 22% of their Korean core vocabulary was made up of verbs. Mngomezulu (2017) also found that verb roots and verbal auxiliaries dominated the core vocabulary, accounting for around 37% of the core vocabulary. According to linguists, verbs play a prominent role in almost all human languages across the world (Mairal & Gil, 2006), and almost all sentences contain verbs.

Forty-nine nouns were part of the core vocabulary in this study, being used with a frequency of 12.4% in the sample. Most other studies also identified nouns (or formatives related to nouns) as part of the core vocabulary, with some finding nouns to be used relatively frequently (Mngomezulu, 2017; Shin & Hill, 2016). Boenisch and Soto (2015) found that nouns were used less frequently in the top 100 most frequently used English words in comparison to the top 300 most frequently used words. In contrast, Robillard et al. (2014) found very limited use of nouns among the core vocabulary identified from French children. Although the use of codes to de-identify children's and teachers' names led to a degree of over-counting, it is still worth noting that children's names (replaced by the code CN) were the most frequent noun and second most frequent 'word' used in the sample. This emphasises the importance of including these in an AAC system, as also noted by Trembath et al. (2007).

Based on these findings, nouns seem to form a pivotal part of the core vocabulary in various languages. Nouns are usually easier to represent in graphic symbol format. However, the core vocabulary should not be misunderstood to be noun-based only. Other parts of speech should also be recognised, as they also appear in high frequencies. An AAC system should not be dominated by nouns, to prevent limitations in expressive capacity (Snodgrass, Stoner, & Angell, 2013). Such as system will limit the expression of more varied and nuanced meanings and will inhibit the development of more complex syntax (Robillard et al., 2014).

Regarding the other parts of speech, verbal affixes together were used with a frequency of about 12%. These grammatical markers would enable the expression of more specific meanings using the correct grammar. Ungrammatical telegraphic messages are common among persons using graphic symbol-based AAC systems (Trudeau et al., 2007; Von Tetzchner & Grove, 2003), and partners are often heavily involved in co-constructing messages in order to make meaning (Brekke & Von Tetzchner, 2003; Solomon-Rice & Soto, 2011). Correct grammar may not always be a priority in AAC intervention (Binger & Light, 2008; Sutton et al., 2002), since it may be time-consuming and cumbersome to construct grammatical messages, and it may also be regarded as 'too difficult' for the person using AAC. As a result, AAC systems may not even always include grammatical markers (Binger & Light, 2008; Sutton et al., 2002). However, it is clear that five- to six-year-olds use a variety of Sepedi grammatical markers (as also evidenced in their use of inflections of verbs, and different grammatical forms of verbs and nouns). Being able to construct grammatically correct sentences enables greater communicative autonomy, with less reliance on partners for interpretation and less risk of misinterpretation (Solomon-Rice, Soto, & Heidenreich, 2017). Inclusion of verbal affixes may therefore be considered on Sepedi AAC systems, and appropriate grammar interventions, such as those found to be effective in various studies (Binger & Light, 2008; Soto & Clarke, 2017), can be considered in order to teach their use.

It is interesting to note that interjections (also termed exclamations in some studies) were included in the core vocabulary. Interjections have consistently been found in the core vocabulary of various other languages (Boenisch & Soto, 2015; Mngomezulu, 2017; Shin & Hill, 2016). In this study, 13 interjections were found in the core, used with a frequency of 3.2%. Interjections typically have important pragmatic functions in discourse, such as initiating a turn, holding the floor, or gaining attention, while also signalling emotion or internal states to the listeners (Norrick, 2009). Children who require AAC have the same communicative needs as their peers and should therefore have access to the same pragmatic functions in order to participate effectively in communicative interactions.

The use of code switching is common and even inevitable in multilingual societies such as South Africa (du Plessis, 2006). Three English words regarded as code switches were found in the core vocabulary, namely the English words *why*, *jersey*, and *toilet.* Code switching describes the use of more than one language in the context of a single conversation

(Reyes, 2004). For example, the children infused the words *toilet* or *why* into a Sepedi sentence. However, the phonological and morphological character of the word was maintained as in the original language; the word was not inflected in the way other Sepedi words are inflected. Code switching often occurs because certain communities have more than one language to communicate with. It may also occur because certain words do not exist in a language. An example of a word that does not exist in Sepedi, which was found in the composite sample, is 'bathtub'. The children used this word to provide the partner with a particular understanding of the idea during the conversation. This is an example where code switching is also used to extend communicative competence for achieving conversational goals during peer interaction (Reyes, 2004).

In the sample, a number of loan words could also be observed. These words are taken from another language, but used within the morphological context of one's own language. For example *mara*, which means 'but' (conjunction) and comes from the Afrikaans 'maar', is a word that was found in the core vocabulary. This is considered a loan word, because the original pronunciation has been changed to the typical Sepedi phonology. A word with the same meaning does exist in Sepedi (*eupša*). These words may be used even if words that represent the concept exist in the Sepedi language (Kosch, 2006). Loan words exist in all languages of the world. Even English words such as 'realise' and 'language' originally came from Latin, but were adopted into the English language. It remains important to note that the nature of language by definition is changeable (Besio & Chinato, 1996; Smith, 2006), and that cross-linguistic influence of languages on each other is observable around the world.

In summary, the core vocabulary determined by this study consisted of words belonging to various parts of speech. When selecting vocabulary, one therefore needs to ensure that a variety of words from different parts of speech are included in order to provide a vocabulary that can cover the generation of a variety of novel messages and enable communicators to engage with a variety of partners in a variety of communication contexts. The presence of parts of speech in the core vocabulary that are specific to the linguistic typology of the language (e.g., concords and verbal affixes) supports the argument that language-specific studies are the most reliable way of identifying the most frequently used words in a language. Furthermore, in multilingual societies, code switches should also be considered for inclusion in a system, as these may appear frequently in the conversations of speaking individuals.

5.4.3 Grammatical variations of words

For all the core vocabulary studies discussed thus far, the authors chose to count inflected forms/grammatical variations of words either separately or as one. Many articles reporting on core vocabulary studies do not report on the number of inflected versus uninflected forms of words. In the current study, inflected forms of nouns, verbs and adjectives were counted together for the purpose of calculating frequency and commonality scores. However, note was also taken of the number and types of inflected forms to obtain an indication of the number of inflected words in the core vocabulary. All verbs appeared in the imperfect form. More than three quarters of the verbs found in the core vocabulary also appeared in the negative imperfect form, while just over half also appeared in the perfect and perfect negative form. The object concord indicating action to the first person was used to modify just over one quarter of the verbs appearing in the core vocabulary. Of the nouns, just under half were used in both plural and singular form, while a small proportion was also used with a locative extension.

It is clear from the data that Sepedi-speaking children aged five to six use a variety of grammatical inflections, specifically pertaining to verbs. As with verbal affixes, these grammatical variations add nuance and specificity to the meanings expressed, and various authors have argued for access to grammar on AAC systems as a method to increase communicative autonomy (Binger & Light, 2008; Binger, Maguire-Marshall, & Kent-Walsh, 2011; Sutton et al., 2002). Features such as automatic grammar support and inflection popups have been added to various AAC systems (e.g., Proloquo2Go[8]). Decisions about when to introduce such grammatical features on AAC systems used with children should be made taking into account language development and learning demands. Some scholars argue that grammar features should be available on the AAC system from the start, in order for the system not to limit the language development of the child using the system, and also argue for targeted grammar intervention for children using graphic symbol-based AAC systems (Binger, 2008; Fey, 2008).

[8] Proloquo2Go is a product of Assistive Ware, Laurierstraat 193, and 1016 PL Amsterdam, Netherlands. https://www.assistiveware.com

5.5 Comparing the Sepedi core vocabulary list to English and isiZulu core lists

The lexical meanings of words in the current Sepedi list were compared to those found in the isiZulu list by Mngomezulu (2017). Because of a different unit of analysis used in the isiZulu list, only certain parts of speech (verbs, nouns, pronouns, adjectives, adverbs and conjunctions) could be compared. Fewer than half of the words overlapped in meaning. Although the use of a different unit of analysis complicated comparisons somewhat, this finding still suggests that core vocabulary varies from one language to another. It also shows that there is minimal translatability even within languages from a similar language family. This supports the idea that core vocabulary is not translatable from one language to another.

The lexical content of the top 100 entries in the Sepedi list were further compared to the top 100 entries from three English core lists from two similar studies (Trembath et al., 2007; Boenisch & Soto, 2015). It was found that just over half of the words overlapped between all the lists. The translation criteria were very lenient. This supports the idea that core vocabulary is not translatable from one language to another (Liu & Sloane, 2006; Mngomezulu, 2017; Shin & Hill, 2016) . Although the unit of analysis was the same (i.e. through orthographic spaces), the core vocabularies determined still varied substantially. This may be because the languages are structurally different. Even the meanings of words that did have some semantic overlap may not map completely onto each other in English and Sepedi. For example, the word *nyaka* can be translated as ‘want’, ‘look for’ or ‘search’ in English.

5.6 Summary

The results of the study were discussed in this chapter. The composite Sepedi speech sample displayed similar TTRs to those found in other studies. The Sepedi core vocabulary identified was found to be similar in number of different words and coverage to those found in previous studies. The findings that the core vocabulary consisted of both structure and content words, as well as different parts of speech, were discussed with reference to the implications this has for vocabulary selection for AAC systems. Finally, the lexical meanings of the core vocabulary words were compared where possible to those found in an isiZulu list and to those found in three English lists. Limited semantic overlap was found, supporting the notion that language-specific studies are needed to determine core vocabularies.

Chapter 6

CONCLUSION

6.1 Introduction

The aim of the chapter is to provide a summary and a critical evaluation outlining the strengths and limitations of the study. The implications for clinical practice are also discussed. Recommendations for future research are provided.

6.2 Summary of the study

AAC systems that make use of graphic symbols require the preselection of relevant vocabulary suitable to meet the communication needs of the person using it. Including core vocabulary (the most commonly and frequently used words) in such a system has been suggested as an integral step in providing access to a measure of generative language.

The main aim of this study was to identify the core vocabulary of Sepedi-speaking preschool children for the purpose of informing AAC communication system design for preschoolers from Sepedi language backgrounds in need of AAC.

Six children from three different preschools participated in this study. Each of the children wore a voice recorder while continuing to participate in their typical preschool routine. Recording continued until 3 000 words (including unintelligible words) per child had been obtained. After combining transcripts and removing unintelligible words, a composite sample of 17 569 words remained.

Frequency and commonality scores were used to determine a core vocabulary of 226 different words. These 226 words accounted for 88.1% of the composite sample. These findings were comparable to those of English studies and studies in other languages (Boenisch & Soto, 2015; Trembath et al., 2007). The Sepedi core vocabulary consisted of a larger number of content words than structure words. Structure words, however, had greater coverage, accounting for 55,4% of all words occurring in the sample. Regarding parts of speech, verbs and concords had the highest frequency of occurrence. When the semantic

meaning of the core words identified were compared to those found in an isiZulu list and different English lists, limited overlap was found.

The identified Sepedi core vocabulary list represents a resource that can be used for vocabulary selection for Sepedi AAC systems intended for children. The use of such a list can ensure that words that are typically used with a high frequency by preschool children are included in order to give access to a system with a measure of generativity in expression.

6.3. Critical evaluation of the study

6.3.1 Strengths

This is the first study that aimed at identifying a Sepedi core vocabulary from preschoolers' spoken language samples. This list can now be used as an additional resource by AAC team members when selecting vocabulary for Sepedi AAC systems to be used by preliterate children in need of AAC.

The use of the observational design permitted the recording of speech produced by preschool children during naturally occurring activities. The 20-minute warm-up period (which was omitted from the transcription and analysis) that was applied after children had been fitted with the recording equipment reduced participant reactivity to the equipment; this strengthened the internal validity of the results. The risk of the vocabulary only reflecting one type of activity, for example circle time or break time only, was minimised by recording during typical preschool activities continuously throughout the day. This increased the external validity of the data.

The checking of each transcription against the original recording by an independent person increased the transcription reliability. The inter-rater reliability of the coding varied from 92% to 96% on assessing 20% of each transcript, showing that the coding was executed reliably. These aspects strengthened the internal validity of the results.

Various measures were implemented during the data analysis process in order to maximise rigour and the internal validity of the study. Words with the same spelling yet different meanings (heteronyms and polysemes) were separated to avoid over-counting and to

enable accurate classification into the various parts of speech. Other studies have not necessarily been rigorous in this regard (e.g. Trembath et al., 2007; Boenisch & Soto, 2015).

Certain grammatical variations (inflected forms) of nouns, verbs and adjectives were counted together, so as not to lose the common elements within these words and dilute the core vocabulary. Since it is likely that the different grammatical forms of these words have the same graphic representation, it seems most appropriate to count them under a common lemma or root. However, note was still taken of the different inflected forms. This method of analysis allowed for a more nuanced analysis and may assist practitioners to make judgements about the necessity to include inflectional morphology on an AAC system.

6.2.2 Limitations

Although three different sites were used for the study, the sample comprised only six participants. This introduces a limitation concerning the extent to which the core vocabulary can be regarded as completely representative and to what extent it can be generalised to the larger population. Also, the sites were relatively homogenous (preschools from the same area), the time span of collecting data was relatively short (two to three days per child) and children were similar in age (five to six years). This also affects the generalisability of the results.

Participant reactivity remained an unavoidable factor, as with all observational designs, and this could have affected the internal validity. Although it seemed that the children conversed freely about various topics, they may still have changed their behaviour in response to the presence of the recorders.

The noise in the classroom may have affected the accuracy of the transcriptions, directly affecting the accuracy of the results. One solution would have been to collect supplementary visual data, e.g. by means of video recordings. The author acknowledges that there would be pros and cons about adding this medium.

The frequency and commonality score criteria for words to be regarded as core are somewhat arbitrary. There is no scientific justification for using a commonality score of ≥3 (50%) and a frequency count of ≥0.5‰ as criteria for the inclusion of words in core vocabulary (Shin & Hill, 2016). There may be other methods of analysis, such as grouped

frequency counts (Shin & Hill, 2016), which represent a more objective way of defining core versus fringe vocabulary.

6.4. Clinical implications

The determined core vocabulary list of 226 words can be used as a source to guide vocabulary selection for children from a Sepedi language background in the preschool phases who require an AAC system. Sepedi home language speakers represent the fifth largest group in South Africa, and the list is therefore expected to have clinical application to a sizeable population of children with complex communication needs.

Individuals (i.e. clinicians, teachers, caregivers, AAC interventionists) who intend to introduce AAC to children who require it may use the list along with other sources — vocabulary selection requires numerous reliable sources in order to provide the person using the system with a comprehensive set of words to express different communication needs in multiple contexts. The list is not developed with the intent to be used in isolation, but should be used with child- and context-specific fringe vocabulary (Robillard et al., 2014) that reflects culturally and individually appropriate ways of communicating.

Specifically, the core word list can alert individuals (i.e. clinicians, teachers, caregivers, AAC interventionists) and other team members to the words that are most frequent and common in conversations of preschool children, and to the structure words that are most likely necessary to build grammatical sentences. Such structure words are often omitted from AAC systems (especially when sources other than core lists are used to preselect vocabulary) based on their abstract nature and difficulty posed when representing them graphically. The importance of these words in an AAC system is that they provide the grammatical aspect of the language and they ensure that sentences can be formulated that are syntactically correct.

Research has stipulated that core vocabulary is known to be useful across settings and individuals. As a result, this list may be used across varying ages and settings. However, one would need to ensure that there are no dialectical differences and that the list is still applicable to the particular individual at a different age. Therefore, this application should have to be done with caution and along with other lists, as some of the words are age- and

context-specific. Examples of such words are *bapala/raloka* (play) and *mpya* (dog). Such words would have to be used in cases where they are person-, age- and context-appropriate.

The selection of vocabulary for an AAC system is only one challenge when designing a system. Some of the other challenges include the representation of abstract (structure) core vocabulary items graphically. One would also need to be vigilant of the cognitive and physical demands of the lay-out and organisation of the system. The effort of generating a message should not exceed the benefit of sharing the message (Von Tetzchner, Grove, Loncke, Barnett, Woll & Clibbens, 1996). One should always ensure that the AAC user remains motivated to use the system to prevent demotivation and negative feelings towards the system. In addition, the design and implementation of a system should ensue in a way that is not only linguistically, but also culturally congruent for the child and family (Khoza-Shangase & Mophosho, 2018), as the system is likely to be under-used or abandoned if this is not the case (McCord & Soto, 2004; Soto & Yu, 2014).

The transcription and coding conventions developed may also be useful to clinicians (and possibly researchers in language development) for transcribing and analysing language samples from Sepedi-speaking children. Language sample analysis is a tool used by speech language pathologists in various languages to identify language delay/disorders and to monitor development and progress (Miller & Iglesias, 2012).

6.5 Recommendations for future research

A recommendation for future studies is to replicate the study with a greater sample size. This will enhance the extent to which the results may be applied to the larger population (generalisation). A more detailed description of participants' exposure to multiple languages and inclusion of participants with varied exposure could also allow a more rigorous analysis of the impact of multilingualism and code switching on the Sepedi core vocabulary. The study could also be replicated among different populations to examine possible influences of age, gender, and geographical areas on Sepedi core vocabulary, especially considering that the Sepedi language has numerous dialects in different areas. Samples recorded in different communication contexts would enable the determination of similarities between the core vocabulary used in varying contexts (for example at home versus at school).

The fringe vocabulary obtained from this study could be further analysed to determine typical semantic categories of fringe vocabulary that may be appropriate for preschool children. The sample may furthermore be analysed for overall conversational topics that could guide vocabulary selection and even lay-out of vocabulary on the AAC systems of preschoolers.

Intervention studies are of great importance to determine how applicable vocabulary items are for AAC users in different settings and contexts. It is also imperative to determine if the vocabulary selected allows for generation of new individualised utterances. The lack of such studies represents a gap in terms of evidence-based practices.

To conduct appropriate and effective AAC interventions for persons from a Sepedi language background, language assessment tools that are appropriate for this population need to be developed. There is as yet a critical lack of appropriate language assessment (Kathard et al., 2011; Bornman, Sevcik, Romski, & Pae, 2010). Therefore more studies should be conducted to devise such tools in the South African languages.

The transcription conventions and coding rules developed for this study may be useful to core vocabulary studies in other Sesotho languages (e.g., Sesotho and Setswana). These languages have the same linguistic structure and also use disjunctive orthographic conventions. Such studies could build on the analysis methods developed for the current study.

6.6 Summary

This chapter provided a summary of the study and a critical evaluation of it. The implications for clinical practice were discussed. Recommendations for future research were also provided.

REFERENCES

Alant E. (2005). Issues in graphic symbol communication. In E. Alant & L.L. Lloyd (Eds.), *Augmentative and alternative communication and severe disabilities: Beyond poverty* (pp 108-130). London, United Kingdom: Whurr.

American Speech-Language-Hearing Association. (2019). *Augmentative and alternative communication* Retrieved from: https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589942773&section=Overview

Atmore, E., van Niekerk, L., & Ashley-Cooper, M. (2012). *Challenges facing the early childhood development sector in South Africa*. Johannesburg, South Africa: National Development Agency.

Baker, B., Hill, K., & Devylder, R. (2000). Core vocabulary is the same across environments. *Proceedings of the Technology and Persons with Disabilities Conference,* California State University, CA.

Balandin, S., & Iacono, T. (1999). Crews , wusses , and whoppas: Core and fringe vocabularies of Australian meal-break conversations in the workplace. *Augmentative and Alternative Communication, 15,* 95-109. doi:10.1080/07434619912331278605

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

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 nondisabled peers in integrated preschool classrooms. *Augmentative and Alternative Communication, 5*(4), 243–248.

Beukelman, D.R., McGinnis, J., & Morrow, D. (1991). Vocabulary selection in augmentative and alternative communication. *Augmentative and Alternative Communication, 7,* 171–185.

Beukelman, D. R., & Mirenda, P. (2013). *Augmentative & alternative communication. Supporting children & adults with complex communication needs*. Baltimore, MD: Paul H. Brookes.

References

Beukelman, D. R., Yorkstorn, K. M., Poblete, M. & Naranjoo, C. (1984). Frequency of word occurence in communication samples produced by adult communicationn aid users. *Journal of Speech and Hearing*, *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., & Light, J. (2008). The morphology and syntax of individuals who use AAC: research review and implications for effective practice. *Augmentative and Alternative Communication*, *24*(2), 123–138. doi:10.1080/07434610701830587

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

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., & Bryen, D. N. (2013). Social validation of vocabulary selection : ensuring stakeholder relevance. *Augmentative and Alternative Communication*, *29*(2), 174-181. doi:10.3109/07434618.2013.784805

Bornman, J., & Donohue, D. K. (2013). South African teachers' attitudes toward learners with barriers to learning: Attention deficit and hyperactivity disorder and little or no functional speech. *International Journal of Disability, Development and Education*, *60*(2), 85-104. doi:10.3109/07434618.2011.610356

Bornman, J., Sevrick, R.A., Romski, M.A., & Pae, H.K. (2010). Successfully translating language and culture when adapting assessment measures. *Journal of Policy and Practice in Intellectual Disabilities*, *7*(2), 111-118. doi:10.1111/j.1741-1130.2010.00254.x

Bosma, E., & Bloma, E. (2018). A code-switching asymmetry in bilingual children: Code-switching from Dutch to Frisian requires more cognitive control than code-switching from Frisian to Dutch. *International Journal of Bilingualism, 00(0)*, 1–17. doi:10.1177/1367006918798972

Brekke, K. M., & Von Tetzchner, S. (2003). Co-construction in graphic language development. In S. Von Tetzchner & N. Grove (Eds.), *Augmentative and alternative communication: Developmental issues* (pp. 176–210). London, United Kingdom:

References

Whurr.

Burroughs G. E. R. (1957). *A study of the vocabulary of young children.* Birmingham, United Kingdom: Oliver & Boyd.

Chen, M., Chen, Y., & Chen, C. (2013). Do they need specific language training content?: Results of a core vocabulary investigation from children with/without intellectual disabilities in Taiwan. *Proceedings of the 7th International Convention on Rehabilitation Engineering and Assistive Technology*.

Chen, M., Hill, K., & Yao, T. (2009). Preliminary vocabulary frequency findings for Mandarin Chinese AAC treatments. *Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology*,

Chen, M., Ko, C., Hsu, S., Lin, Y. L., Chi, P., & Huang, Q. (2011). Do they narrate in same vocabulary?: the result of core vocabulary investigation from children with/without asperger syndrome in Taiwan. *Proceedings of the 5th International Conference on Rehabilitation Engineering & Assistive Technology*,

Cresswell, J. W. (2009). *Research design: Qualitative, quantitative and mixed methods approaches.* Thousand Oaks, CA: Sage Publications.

Croft, W. (2000). *Explaining language change: an evolutionary approach.* London, United Kingdom: Longman.

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

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

de Groot, A. M. B. (2013). Written language processing. In F. Grosjean & P. Li, *The psycholinguistics of bilingualism* (pp. 171-181). New York, NY: John Wiley & Sons, Inc.

de Klerk, F. W. (2016). The impact of the language of education at schools and universities *Proceedings from Conference the Future of Multiculturalism in South Africa.* Cape Town, South Africa, 1-12.

de Pauw, G, & de Schryver, G-M. (2007). Dictionary writing system (DWS) + corpus query package (CQP): The case of Tshwanelex. *Lexikos*, *17*, 226– 246

du Plessis, S. (2006). *Multilingual preschool learners : a collaborative approach to communication intervention* (Doctoral dissertation). University of Pretoria. Pretoria, South Africa.

de Schryver, G-M. (2007). *Pukuntšu ya polelopedi ya sekolo. Sesotho sa Leboa le Seisimane.*

References

[Oxford Bilingual School Dictionary. Northern Sotho and English]. Cape Town, South Africa: Oxford University Press.

Faaß, G., Heid, U., Taljard, E., & Prinsloo, D. (2009). Part-of-speech tagging of Northern Sotho : Disambiguating polysemous function words. *Proceedings of the EACL 2009 Workshop on Language Technologies for African Languages-AfrLaT 2009*, Greece, 38–45.

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)

Fey, M. (2008). Thoughts on grammar intervention in AAC. *Perspectives on Augmentative and Alternative Communication*, *17*(2), 43–49. doi:10.1044/aac17.2.43

Freelon, D. G. (2013). ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a web service. *International Journal of Internet Science*, *8*(1), 10–16.

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., & Lloyd, L. (1991). Toward a common usage of iconicity terminology. *Augmentative and Alternative Communication*, *7*, 215–220.

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.

Gerber, S., & Kraat, A. (1992). Use of a developmental model of language acquisition : applications to children using AAC systems. *Augmentative and Alternative Communication*, *8*, 19–32. doi:10.1080/07434619212331276013

Hattingh, D. (2018). *The core vocabulary of South African Afrikaans-speaking preschoolers without disabilities* (Master's thesis). University of Pretoria. Pretoria, South Africa.

Hill, K., & Romich, B. (2002). A rate index for augmentative and alternative communication. *International Journal of Speech Technology*, *5*, 57–64.

Howard, K. (2003). *Language socialization in a Northern Thai bilingual community* (Doctoral dissertation). University of California, Los Angeles. USA.

References

Johnson, S. D., Aragon, S. R., Shaik, N., & Palma-Rivas, N. (2000). The influence of learning style preferences on student success in online versus face-to-face environments. *Journal of Interactive Learning Research ,11*(1), 29-49.

Kathard, H., Ramma, L., Pascoe, M., Jordaan, H., Moonsamy, S., Wium, A.M., & Khan, N. B. (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.

Kettunen, K. (2014). Can type-token ratio be used to show morphological complexity of languages? *Journal of Quantitative Linguistics*, *21*(3), 223-245.

Khoza-Shangase, K., & Mophosho, M. (2018). Language and culture in speech-language and hearing professions in South Africa: The dangers of a single story. *South African Journal of Communication Disorders*, *65*(1), 1–7. doi:10.4102/sajcd.v65i1.594

Kilgarriff A., Charalabopoulou F., Gavrilidou M., Bondi Johannessen J., Khalil S., Johansson Kokkinakis S., Lew R., Sharoff S., Vadlapudi R, Volodina E. (2012). Corpus-based vocabulary lists for language learners for nine languages. *Language Resources and Evaluation Journal, special issue, 48,* 121–163.

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.

Kosar, K. R. (2011). *The Quasi Government: Hybrid Organizations with Both Government and Private Sector Legal Characteristics.* Washington, United Stated of America: Congressional Research Service.

Kosch, I. M. (2006). *Topics in morphology in the African lanugage context* (1st ed.). Pretoria, South Africa: Unisa Press.

Kuhl, P.K. (2010). Brain mechanisms in early language acquisition. *Neuron, 67,* 713–727. doi:10.1016/j.neuron.2010.08.038

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.

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

Light, J., Drager, K., McCarthy, J., Mellott, S., Millar, D., Parrish, C., Wellevier, M. (2004). Performance of typically developing four- and five-year-old children with AAC systems using different language organization techniques. *Augmentative and Alternative*

References

Communication, *20*(2), 63–88. doi:10.1080/07434610410001655553

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

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., Quist, R.W., & Windsor, J. (1990). A proposed augmentative and alternative communication model. *Augmentative and Alternative Communication*, *6*(3), 172- 183.

Mairal, R., & Gil, J. (2006). *Linguistic universals* (1st ed.). New York, United States of America: Cambridge University Press.

Mayer-Johnson.(2018). PCS Symbol libraries. Retreived from: https://www.tobiidynavox.com/support-training/downloads/pcs-symbols/libraries/

McCord, M. S., & Soto, G. (2004). Perceptions of AAC: An ethnographic investigation of Mexican-American families. *Augmentative and Alternative Communication*, *20*(4), 209–227. doi:10.1080/07434610400005648

McMillan, J., & Schumacher, S. (2010). *Research in education: Evidence-based inquiry* (7th ed.). New York, NY: Pearson.

Miller, J.F., & Chapman, R. S. (1990). *Systematic analysis of language transcripts.* Madison, WI: University of Wisconsin.

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

Mirenda, P. (2003). Toward functional augmentative and alternative communication for students with autism: manual signs, graphic symbols, and voice output communication aids. *Language, Speech, and Hearing Services in Schools*, *34*(3), 203–216. doi:10.1044/0161-1461(2003/017)

Mngomezulu, J. R. (2017). *Determining an AAC core vocabulary for Zulu-speaking preschool children* (Master's thesis). University of Pretoria. Pretoria, South Africa. Retrieved from https://repository.up.ac.za/handle/2263/64811

Mojapelo, M.L.(2007). *Definiteness in Northern Sotho* (Doctoral dissertation). Stellenbosch: University of Stellenbosch. Stellenbosch, South Africa.

Mojela, V.M. (2010). Borrowing and loan words: the lemmatizing newly acquired lexical items in Sesotho sa Leboa. *Afrilexreeks 20*, 700 – 707

Mokwana, M.L. (2009). *The melting pot in Ga-Matlala Maserumule with special reference to the Bapedi culture, language and dialects* (Master's thesis). University of South Africa.

References

Pretoria, South Africa.

Nong, S., de Schryver, G.-M., & Prinsloo, D. J. (2002). Loan words versus indigenous words in Northern Sotho — A lexicographic perspective. *Afrilexreeks 12*, 1–20.

Norrick, N. R. (2009). Interjections as pragmatic markers. *Journal of Pragmatics*, *41*(5), 866–891. doi:10.1016/j.pragma.2008.08.005

Noruzi, A. (2006). Integrativeness: untenable for world Englishes learners?. Folksonomies: (Un)Controlled Vocabulary? *Knowledge Organization*, 33(4): 199-203.

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

Owen, A. J., & Leonard, L. B. (2002). Lexical diversity in the spontaneous speech of children with specific language impairment: Application of D. *Journal of Speech, Language, and Hearing Research*, *45*, 927-937.

Patel, R., Schooley, K., & Wilner, J. (2007). Visual features that convey meaning in graphic symbols : A comparison of PCS and artists' depictions. *Assistive Technology Outcomes and Benefits*, *4*(1), 62-80.

Payne, T. (1997). Describing morphosyntax; A guide for field linguists. Cambridge: Cambridge University Press.

Perkins, S. C., Finegood E. D., & Swain J. E. (2013). Poverty and language development: roles of parenting and stress. *Innovative Clinical Neurosci*ence, *10*, 10–19.

Poulos, G., & Louwrens, L. J. (1994). *A linguistic analysis of Northen Sotho*. Hatfield, Pretoria, South Africa: Via Afrika.

Prah, K. K. (2007). Challenges to the promotion of indigenous languages in South Africa. Cape Town, South Africa: The Center for Advanced Studies of African Society. 1–35.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). Reviews: A comprehensive grammar of the English language. New York, NY: Longman Inc. 109–111.

Renner, G. (2003). The development of communication with alternative means from Vygotsky's cultural historical perspective. In S. von Tetzchner & N. Grove (Eds.), *Augmentative and alternative communication developmental issues* (pp. 67-82). London, United Kingdom: Whurr.

Reyes, I. (2004). Functions of code switching in schoolchildren's conversations literature review : code switching in children. *Bilingual Research Journal*, *28*, 77–98.

Robillard, M., Mayer-Crittenden, M., Minor-Corriveau, C., & Belanger., R. (2014). Monolingual and bilingual children with and without primary language impairment : Core Vocabulary Comparison. *Augmentative and Alternative Communication*, *30*, 267–

References

278. doi:10.3109/07434618.2014.921240

Romski, M., Sevcik, R. A., Adamson, L. B., Cheslock, M., Smith, A., Barker, R. M. (2010). Randomized comparison of augmented and nonaugmented language interventions for toddlers with developmental delays and their parents. *Journal of Speech, Language and Hearing Research 2010; 53*(2):350-64.

Rooy, S. C. (2006). Integrativeness: |Untenable for world Englishes learners? *World Englishes*, *25*(3), 437–450.

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English- learning infants, *Infancy*, *10*(2), 187–198.

Shin, S., & Hill, K. (2016). Korean word frequency and commonality study for augmentative and alternative communication. *International Journal of Language and Communication Disorders*, *51*(4), 415–429. doi:10.1111/1460-6984.12218

Sigafoos, J., & Drasgow, E. (2001). Conditional use of aided and unaided AAC: A review and clinical case demonstration. *Focus on Autism and Other Developmental Disabilities*, *16*(3), 152-161.

Sigafoos, J., Ganz, J. B., O'Reilly, M. F., Lancioni, G. E., & Schlosser, R. (2007). Assessing correspondence following acquisition of and exchange-based communication system. *Research in Developmental Disabilities*, *71–83.*, 71–83.

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

Smith, M., & Grove, N. (2013). Asymmetry in input and output for individuals who use augmentative and alternative communication. In J. Light, D. Beukelman, & J. Reichle (Eds), *Communicative competence of individuals who use augmentative and alternative communication*. Baltimore, MD: Paul H. Brookes.

Snodgrass, M. R., Stoner, J. B., & Angell, M. E. (2013). Teaching conceptually referenced core vocabulary for initial augmentative and alternative communication. *Augmentative and Alternative Communication, 29*(4):322-333. doi:10.3109/07434618.2013.848932

Solomon-Rice, P., & Soto, G. (2011). Co-construction as a facilitative factor in supporting the personal narratives of children who use augmentative and alternative communication. *Communication Disorders Quarterly*, *32*(2), 70–82. doi:10.1177/1525740109354776

Solomon-Rice, P. L., Soto, G., & Heidenreich, W. (2017). The impact of presupposition on the syntax and morphology of a child who uses AAC. *Perspectives of the ASHA Special*

References

Interest Groups: Augmentative and Alterntive Communication, *2*(12), 13–22. doi:10.1044/persp2.SIG12.13

Soto, G., & Clarke, M. (2017). Effects of a conversation-based intervention on the linguistic skills of children with motor speech disorders who use augmentative and alternative communication. *Journal of Speech Language and Hearing Research*, *60*(7), 1980–1998. doi:10.1044/2016_JSLHR-L-15-0246

Soto, G., & Yu, B. (2014). Considerations for the provision of services to bilingual children who use augmentative and alternative communication. *Augmentative and Alternative Communication*, *30*(1), 83–92. doi:10.3109/07434618.2013.878751

Statistics South Africa (2012). *Census in brief.* Pretoria, South Africa: Author. Retrieved from http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf

Stuart, S., Beukelman, D. R., & King, J. (1997). Vocabulary use during extended conversations by two cohorts of older adults. *AAC: Augmentative and Alternative Communication*, *13*(1), 40–47.

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

The Belmont Report (1978): *Ethical principles and guidelines for the protection of human subjects of research* Retrieved from http://ohsr.od.nih.gov/mpa/belmont.php3

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, United Kingdom: Whurr.

Trembath, D., Balandin, S., & Dark, L. (2006). Why any old words won't do: The importance of vocabulary selection. *Acquiring Knowledge in Speech, Language and Hearing*, *8*(3), 117–119.

Trembath, D., Balandin, S., & Togher, L. (2007). Vocabulary selection for Australian children who use augmentative and alternative communication. *Journal of Intellectual & Developmental Disability*, *32 (4)*, 37–41. doi:10.1080/13668250701689298

References

Trudeau, N., Sutton, A., Dagenais, E., De Broeck, S., & Morford, J. (2007). Construction of graphic sumbol utterances by children,teenagers, and adults: The effect of structure and task demands. *Journal of Speech Language, and Hearing Research, 50*, 1314-1329.

van der Sandt-Koenderman, M., Wiegers, J., & Hardy, P. (2005). Assistive technology a computerised communication aid for people with aphasia. *Disability and Rehabilitation, 27*(9), 529–533. doi:10.1080/09638280400018635

van Tilborg, A., & Deckers, S. R. J. (2016). Vocabulary selection in AAC : Application of core vocabulary in atypical populations vocabulary selection in AAC. *Perspectives of the ASHA Special Interest Groups*, *1*(4), 125–138.

Von Tetzchner, S. & Basil, C . (2011). Terminology and notation in written representations of conversations with augmentative and alternative communication. *Augmentative and Alternative Communication, 27,* 141 – 149. doi:10.3109/07434618.2011.610356

Von Tetzchner, S., & Grove, N. (2003). AAC and language development. In S. Von Tetzchner & N. Grove (Eds.), *AAC: Developmental issues* (pp. 1–23). London, United Kingdom: Whurr.

Von Tetzchner, S., Grove, N., Loncke, F., Barnett, S., Woll, B., & Clibbens, J. (1996). Prelimiminarie to a comprehensive model of augmentative and alternative communication. *Augmentative and Alternative Communication: Europian Perspectives*.

van Wyk, E. B., Groenewald, P. S., Prinsloo, D. J., Kock, J. H. M., & Taljard, E. (1992). *Northern Sotho for first-years.* Pretoria, South Africa: Van Schaik.

Wallace, I. F., Berkman, N. D., Watson, L. R., Coyne-Beasley, T., Wood, C. T., Cullen, K., & Lohr, K. N. (2015). Screening for speech and language delay in children 5 years old and younger : A systematic review. *Pediatrics*, *136*(2),449-458. doi:10.1542/peds.2014-3889

Wilkinson, K. M., & Hennig, S. (2007). The state of research and practice in augmentative and alternative communication for chilldren with developmental/intellectual disabilities. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*, 58–69. doi:10.1002/mrdd.20133

Witkowski, D., & Baker, B. (2012). Addressing the content vocabulary with core: Theory and practice for non-literate or emerging literate students. *Perspectives on Augmentative and Alternative Communication, 21*, 74 – 81.

Zirker, K.A. H. (2007). *Intrasentential vs. intersentential code switching in early and late bilinguals* (Masters thesis). Brigham Young University. Provo, Utah.

APPENDIX A

ALPHABETISED ENGLISH TRANSLATIONS OF THE TOP 100 CORE VOCABULARY ITEMS OF LISTS IN DIFFERENT LANGUAGES AND COMPARISON TO AN ENGLISH COMPOSITE LIST

isiZulu list (Mngomezulu, 2017)		Korean list (Shin & Hill, 2016)		German list (Boenisch, 2014)	
Translated core word	**Equivalent found in English composite list (Hattingh, 2018)**	**Translated core word**	**Equivalent found in English composite list (Hattingh, 2018)**	**Translated core word**	**Equivalent found in English composite list (Hattingh, 2018)**
come	1	a bit	0	a	1
do; make	1	a little	1	again	0
eat; confiscate; eat into; cost	0	a lot	0	alike	0
give	0	again	0	all	1
go to	1	all	1	already	0
he; him; himself; she; her; herself	1	along with	0	also	0
hear; listen; taste; smell; feel; sense; live; be alive	0	and	1	always	0
here	1	at	1	and	1
here	1	be	1	at	1
here they are/here it is/here he is, etc.	1	be	1	be	1
here take	1	be	1	because	1
home	0	because	1	but	1
human being; African; one with human feelings; blunt instrument (as knife)	0	being	0	can	1
I, me, myself	1	by	0	come	1

Appendices

isiZulu list (Mngomezulu, 2017)		Korean list (Shin & Hill, 2016)		German list (Boenisch, 2014)	
Translated core word	**Equivalent found in English composite list (Hattingh, 2018)**	**Translated core word**	**Equivalent found in English composite list (Hattingh, 2018)**	**Translated core word**	**Equivalent found in English composite list (Hattingh, 2018)**
if; when	0	by the way	0	completely	0
indeed; truly	0	can	1	eat	0
it; itself	1	case	0	exactly	0
it; itself	1	continuously	0	for	1
it; itself	1	degree	0	give	0
it; itself	1	different	0	go	1
know	1	do	1	good	1
lady teacher	0	do	1	have	1
lie; untruth	0	do	1	hear	0
look at; watch; admire	1	do	1	hello	0
never	0	each other	0	here	1
of place; here; there	1	English	0	how	1
play; dance; frolic	1	exist	0	I	1
put in; put on; put around; commencement of udder to fill with milk	1	exist	0	if	0
report someone; invent; plot against; be rich	0	friend	0	in	1
request; negotiate for a wife; be almost	0	from now on	0	interjection	0
say; intend; think	1	have	1	interjection	0
search; want; desire	1	here	1	interjection	0
see; understand; give regards	1	home, house	0	Interjection	0
sit; stay; remain	0	how	1	Interjection	0

Appendices

isiZulu list (Mngomezulu, 2017)		Korean list (Shin & Hill, 2016)		German list (Boenisch, 2014)	
Translated core word	**Equivalent found in English composite list (Hattingh, 2018)**	**Translated core word**	**Equivalent found in English composite list (Hattingh, 2018)**	**Translated core word**	**Equivalent found in English composite list (Hattingh, 2018)**
small bag; pocket; purse; fund	0	I	1	it	1
strike; punish; play (as an instrument)	1	I	1	know	1
such as this; like this	1	I	1	let	1
take; marry	1	it	1	like (verb)	1
thing; object	0	it	1	look	1
Translation of the content formative (dictionary)	Equivalent exists	just	1	make	1
walk; go; travel	1	just	1	may	0
we; us	1	Korea	0	me	1
what is it; is it so	1	like that	1	me	1
what; of what sort	1	like this	1	microphone	0
where	1	more	0	more	0
who?	0	not	1	Mrs	0
write; write an examination	0	not exist	0	much	0
yes	1	now	1	must	0
you	1	of	1	my	1
*	0	one	1	no	1
*	0	oneself	0	no	1
*	0	only	0	not	1
*	0	or	1	nothing	0
*	0	person	0	now	1
*	0	quite	0	of	1
*	0	really	0	oh (interjection)	1

Appendices

isiZulu list (Mngomezulu, 2017)		Korean list (Shin & Hill, 2016)		German list (Boenisch, 2014)	
Translated core word	**Equivalent found in English composite list (Hattingh, 2018)**	**Translated core word**	**Equivalent found in English composite list (Hattingh, 2018)**	**Translated core word**	**Equivalent found in English composite list (Hattingh, 2018)**
*	0	reason	0	okay	0
*	0	school	0	on	1
*	0	seem	0	one	1
*	0	similar to	0	one (pronoun - as in one should not do that)	1
*	0	so	1	only	0
*	0	such	0	or	1
*	0	than	0	please	0
*	0	that	1	say	0
*	0	there	1	see	1
*	0	thereby	0	she	1
*	0	thing	0	Should	0
*	0	thing	0	so	1
*	0	this	1	still, more, yet	0
*	0	this	1	stop	0
*	0	this	1	that	1
*	0	thought	1	the	1
*	0	time	1	the	1
*	0	too much	0	the	1
*	0	USA	0	the	1
*	0	way	0	the	1
*	0	we	1	the	1
*	0	we	1	then	1

Appendices

isiZulu list (Mngomezulu, 2017)		Korean list (Shin & Hill, 2016)		German list (Boenisch, 2014)	
Translated core word	**Equivalent found in English composite list (Hattingh, 2018)**	**Translated core word**	**Equivalent found in English composite list (Hattingh, 2018)**	**Translated core word**	**Equivalent found in English composite list (Hattingh, 2018)**
*	0	well	1	this	1
*	0	what	1	three	0
*	0	with	1	times' x	0
*	0	with	1	to	1
*	0	words	0	today	0
*	0	work	0	two	1
*	0	year	0	us	1
*	0	yes	1	wait	0
*	0	young person	0	want	1
*	0	*	0	we	1
*	0	*	0	what	1
*	0	*	0	where	1
*	0	*	0	who	0
*	0	*	0	why	0
*	0	*	0	with	1
*	0	*	0	yes	1
*	0	*	0	yes (contradictory)	1
*	0	*	0	you	1
*	0	*	0	you	1
*	0	*	0	you	1
*	0	*	0	you (plural)	1
*	0	*	0	your	1
Total equivalents	31		45		63

* No equivalent translation into English exists for 51 of the top 100 isiZulu core vocabulary items and 13 of the top 100 Korean core vocabulary items

Appendices

APPENDIX B

ETHICS APPROVAL (UNIVERSITY OF PRETORIA)

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Research Ethics Committee

12 February 2018

Dear Ms Mothapo

Project:	**Determining the core vocabulary used by Sepedi speaking preschool children during regular pre-school-based activities**
Researcher:	**RI Mothapo**
Supervisor:	**Dr K Tőnsing**
Department:	**Centre for Augmentative and Alternative Communication**
Reference number:	**12301656 (GW20171037HS)**

Thank you for your response to the Committee's correspondence of 7 November 2017.

I have pleasure in informing you that the Research Ethics Committee formally **approved** the above study at an *ad hoc* meeting held on 12 February 2018.

Please note that this approval is based on the assumption that the research will be carried out along the lines laid out in the proposal. Should your actual research depart significantly from the proposed research, it will be necessary to apply for a new research approval and ethical clearance.

We wish you success with the project.

Sincerely

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

cc: Dr K Tőnsing (Supervisor)
Prof J Bornman (HoD)

Appendices

APPENDIX C

PRINCIPAL INFORMATION LETTER AND PERMISSION FORM

C.1 English version

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Humanities

_____

The Principal: _____ _____

_____

_____

Date:_____

Dear _____

Re: Permission to conduct research study at your school

My name is Rahab Mothapo. I am currently enrolled for a Master's degree in Augmentative and Alternative Communication (AAC) at the University of Pretoria. The title of my study is: *"Determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities."*

I would be much obliged if you would permit me to include your school, _____ in the study.

Rationale of the study

Children with severe disabilities who cannot speak may be given augmentative and alternative communication (AAC) systems to help them communicate. Choosing the appropriate words for such systems is important to ensure that children can use them to communicate in various situations. For this reason I want to find out what words Sepedi-speaking children use, in order to use this as a guideline in selecting the appropriate words for designing AAC systems for children who understand Sepedi but who cannot speak.

What will be expected of the school?

I will require the help of the teachers to nominate two (2) learners 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 program. I will also require the help of the teacher to send information and consent forms to the parents of the nominated learners 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 learner 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. If children experience any problems with the recorder, the teacher, researcher or research assistant may help the child to adjust the recorder or remove it.

The routines and program of the preschool will not be altered or interfered with during the

Centre for Augmentative and Alternative Communication, Room 2-36, Corn 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

learner's participation. Teachers may decide to remove the recorders from the child for certain activities. Recording will continue on consecutive days until a total of 3000 words have been collected for each participating learner. I expect that this will not take longer than five days.

What will be expected of the learners participating in the study?
The learners 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 learner agrees, he/she will be required to wear a small bag around his/her waist containing a small voice recorder and a microphone clipped to the collar or shirt during their normal routines at preschool, on consecutive days for no longer than an estimated five days.

The learners 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. Learners may stop the procedures at any time without any negative consequences to them.

The following ethical principles will be upheld within this study:

Written consent from all participants' parents and 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 that 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 scientific papers 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 learners has been obtained again.

What are the risks and the benefits?
At no time during the participation in the research will the learners be at risk of any harm. The learners will not miss out on any of their daily programs through participating in this research. Potential benefits of this study may include extending research within the field of AAC by providing guidelines regarding what words to include when designing AAC systems for learners who need AAC and who use Sepedi as their language of communication.

Please complete the attached form to indicate whether or not you give permission for me to conduct this study at your school.

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

Ms N.R.B Mothapo Date
Email: [redacted]
Cell: 0[redacted]

Dr Kerstin Tönsing Date
Centre for Augmentative and Alternative Communication
Email: [redacted]
Office tel: 0[redacted]

Appendices

Principal permission: Reply Slip

Name of principal: _____

Name of School: _____

Project title: Determining the core vocabulary used by Sepedi-speaking preschool children during regular preschool-based activities

Researcher: Rahab Mothapo
Masters student
Cell: [redacted]

Supervisor: Kerstin Tönsing

I, _____(Name and surname)

(Please tick box that applies)

☐ give permission to (student name) to recruit learners from the school named above for possible participation in the study entitled D**etermining the core vocabulary used by Sepedi-speaking preschool children during regular preschool-based activities**, conducted by Rahab Mothapo, under the supervision of Kerstin Tönsing. This permission is voluntary and I understand that I may withdraw at any time. I understand that participating learners 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 an obtained in this study will be treated as confidential.

OR

☐ do not give permission to Rahab Mothapo to recruit learners from the preschool named above for possible participation in the study entitled D**etermining the core vocabulary used by Sepedi-speaking preschool children during regular preschool-based activities**.

School stamp

_____
Principal Signature

_____
Date

Appendices

C.2. Sepedi version

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Humanities

_____

Hlogo ya sekolo: _____

_____ _____

_____

Mohlomphegi _____ Letšatši: _____

Re: Mabapi le tumelelo ya bana ba sekolo sa gago gore ba tšeye karolo go nyakišišo

Lebitso la ka ke Rahab Mothapo. Ke tsenetši go ithutela degree ye kgolo ya Master's ka lefapheng la Augmentative and Alternative Communication (AAC) kua Yunibesithi ya Pitori. Ka dithutong tša ka, ke rata go nyakišiša gore bana bao ba šomišago Sepedi ba šomiša mantšu afe ge ba thoma sekolong. Ka gona, ke kgopela tumelelo ya go šoma le bana ba sekolo sa gago go hwetša seo.

Bohlokwa bja nyakišišo ye ke eng?

Bana ba bantši ba go se itekaneli ba hloka polelo. Augmentative le alternative communication (AAC) e rutwa go bana go ba thuša ka polelo. Kgetho ya mantšu ao a swanetšego e bohlokwa go thuša bana gore ba kgone go ipolelela ka dinako tšohle. Ka lebaka le, ke rata go tseba gore mantšu ao bana ba go bolela Sepedi ba a šomišago ke afe. Se se thuša barutiši gore ba kgone go kgetha mantšu a Sepedi go thuša bana bao ba sa kgonego go ipolelela.

Ke eng seo se lebeletšwego go sekolo sa gago?

Nka thabela thušo ya barutiši gore ba nkgethele barutwana ba babedi bao ba ka tšeago karolo mo nyakišišong ye. Barutiši ba tla kgopelwa go tlatša dipotšišo tša letlakala le tee ka sekolo se le tshipidišo ya sona. Nka thabela le thušo ya bona ka o nthomela mangwalo go batswadi ba bana bao ba ba kgethago gore ba mphe tetla ya go tšwela pele le bana. Ge eba batswadi ba a dumela, le bana ba rata go tšwela pele, ba tla apeša direkhoda tšeo di tla dumelelago gore ke kwe mantšu ao ba a šomišago ge ba le sekolong.
Nna goba bathuši re tla apeša bana direkhoda,ra ba hlobola tšona, ra ba kgauswi ge bana ba apešitšwe direkhoda. Tše ka moka di ka se hlakahlakantšhe thuto tša bona. Ge bana ba hlakana le bothata ka rekhoda ba tla hwetša thušo ya go e lokiša goba go ba hlobola yona. Dithuto tša bana di ka se amiwe ke go rekhodiwa go. Barutiši ba ka ntšha bana direkhoda ka dinako tše ba bonago go swanetše. Bana ba tla rekhodiwa matšatši a go latelana go fihla mantšu a 3000 wo mongwe le wo mongwe. Ke dumela gore se se ka se fete matšatši a 5.

Appendices

Ke eng seo se lebeletšwego go sekolo sa gago?

Barutwana ba tla hlakana le nna le morutiši a le gona. Ke tla ba botšiša gore ba nyaka go tšea karolo go nyakišišo ye goba aowa. Ke tla šomiša ditshwantšho go hlalosa gore re tlo šoma bjang. Le bona ba tla nkaraba ka di tshwantšho le ka polelo.

Ge morutwana a dumela, o tla apara mokotlana wa go ba le rekhoda matšatšing ao a go latelana a mahlano ka nako ya sekolo. Re tla dira se matšatši a go lekana 3 go ya go 5.

Bana ba tla laelwa gore ba se ke ba swara rekhoda tšeo.Ba tla laelwa gore ba botše motho wo mogolo ge ba hloka thušo goba ba nyaka go emiša mošomo wo. Motho wo mogolo o tla thuša ngwana wo, goba a ntšha rekhoda yeo. Barutwana ba ka no emiša mošomo ka nako ye ngwe le ye ngwe ntle le dipoelo tše di mpe.

Melao ye e latelago e tlo swarwa mo nyakišišong ye:

Batswadi ba tla kgopelwa tetla gore bana ba tšee karolo go nyakišišo ye. Le bana bao ba tšeago karolo ba tla kgona go dumela ka bo bona pele ga nyakišišo. Bana ka moka ba tla tšebišwa ka ditokelo tša go emiša ka nako ye ngwe le ye ngwe, ntle le dipoelo tše mpe. Dipoelo tšeo di hwetšwago mo, di tla fihlelwa ke monyakišiši (elego nna), mogolo wa ka le bao ba nthušago.
Ditaba ka moka di tla swarelwa ka sephiring go batho ba bangwe ntle le bao ba badilwego. Ditaba tše dingwe le tše dingwe tšeo di ka ahlolago bana goba barutiši, di tlo tlošwa mo go dipoelo. (go swana le maina a batho, le mafelo di tlo tlošiwa). Ga gona lebitso la motho goba la sekolo leo le tlo tšweletšwago go dipoelo.

Who will have access to the results of the study?

Nyakišišo ye e tlo šwarwa kua Yunibesithi ya Pitori kua Centre ya Augmentative le Alternative Communication mengwaga ye 15. Ditaba tšeo di hwetšwago di tla šomišwa go ngwala letlakala la Master's dissertation,le go ngwala matlakala a dinyakišišo le dipresentašene kua diconferenseng. Letlaka leo le tšweletšago dipoelo ka boripana le tla ba gona gore bao ba ratago go tseba ba be le monyetla.

Ditaba (ntle le tšeo di ka ahlolago bana goba barutiši) di ka šomišwa go tšweletša dinyakišišo tše dingwe. Direkhodo tša mantšu di tla šomišwa morago a go kgopela tetla go batswadi pele.

Dikotsi le dipoelo tše di botse ke dife?

Ga gona nako yeo ngwana wa gago a ka ikhwetšago a le mo kotsing. Ngwana wa gago o tšwetša pele tša sekolo tša tlwaelo le ge a tšeya karolo mo dinyakišišong tše. Morutiši wa ngwana wa gago o tla netefatša gore ngwana wa gago o bolokegile go dira seo se swanetšego. Morutiši o tla thuša ngwana go apola motšhene ge a sa kwane le wona.

Nyakišišo ye e tla thuša gore boramahlale ba kgone hlama metšhene ye AAC yeo e ka thušago bana bago bolela Sepedi eupša ba sa kgone go bolela.

Appendices

Nka thaba ge o ka tlatša foromo yeo e tlago le lengwalo le, go netefatša gore o dumela gore bana ba gago ba tšeye karolo mo. Ke kgopela o buše foromo ye sekolong sa ngwana wa gago.

Ge o sa dumele, ke kgopela o hlatše dipotšiso o di bušetše go nna.

Ge le nyaka ditaba ka thuto tše, ihlakantše le nna goba le mogolo wa ka, ka go šomiša di contact details tšeo di ngwadilego mo fase.

Wa lena,

_____ 19/01/2017

Ms. N.R.B Mothapo Letšatši

Email: [redacted]
Mogala: [redacted]

_____ _____

Dr Kerstin Tönsing Letšatši

Centre for Augmentative and Alternative Communication
Email: [redacted]
Mogala: [redacted]

Appendices

Tumelelo ya hlogo ya sekolo:

Selipi sa go araba

Lebitšo la hlogo ya sekolo: _____

Lebitšo la sekolo: _____

Project title: Determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities

Monyakišiši: Rahab Mothapo
Masters student
Cell: [redacted]

Mogolo wa gagwe: Kerstin Tönsing

Nna, _____ (Lebitso le sefane)

(Ke kgopela o swaye e tee)

☐ Ke fa tumelelo/tetla go Rahab Mothapo gore a bereke le bana ba sekolo sa ka mo nyakišišong ya **determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities** yeo e dirwago ke Rahab Mothapo, ka fase ga tebelelo ya Kerstin Tönsing. Tumelelo ye ga se ka gapeleŠwa, ebile ke a kwišiša gore nka gogela morago nako ye ngwe le ye ngwe. Ke a kwišiša gore bana ba tlo theeletŠwa ka rekhoda. Ke a kwišiša gore ditaba tše di tlo beiwa menwaga ye 15 kua CAAC le gore ditaba tše di ka se fiwe mang le mang. Ke a kwišiša gore ditaba tše di ka šomišwa gape eupša boitsebišo bja ka bo tla dula sephiring.

GOBA

☐ Ga ke fe tumelelo/tetla go Rahab Mothapo gore a bereke le bana ba sekolo sa ka mo nyakišišong ya **determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities.**

.

Stempe sa sekolo

Siknatša ya Hlogo ya sekolo

_____
Letšatši

Appendices

APPENDIX D

PRESCHOOL BACKGROUND QUESTIONNAIRE

D.1. English version

(Based on Mngomezulu, 2017)

The purpose of the questionnaire is to establish knowledge about the preschool environment.

Date: _____

Respondent name: _____

Position held at the preschool: _____

Preschool name: _____

Instruction: Kindly answer each question by ticking the preferred option.

Information about the language(s) used at the preschool

1. Is Sepedi the primary language of instruction used in the preschool?

YES	☐
NO	☐

2. Is Sepedi the only language you use in class for teaching?

YES	☐
NO	☐

3. If not, what other languages do you use for teaching? Please describe: _____

_____

4. Which language do children in your class primarily use to communicate with each other?

_____

Appendices

5. Which other languages do children use among themselves?

_____

6. How many assistants do you have to help in your class? (If none, please indicate 0.)

_____

7. Which language do the assistant(s) use primarily for communicating with the children?

_____

8. Which other languages do the assistants use to communicate with the children?

_____

Information about the children and the preschool program

9. How many children are there in your class? _____
10. How old are the children in your class? From _____years (youngest) to _____years (oldest).
11. How many children are there at the preschool overall?

    _____

12. How many preschool classes are there?

    _____

13. Does your preschool follow a curriculum?

YES ☐ If yes, please specify:_____

NO ☐

14. How old are the children in the preschool overall?

    From _____years (youngest) to _____years (oldest).

15. Do the children in your class get a chance to interact with the other children in the school? Please describe:_____

    _____

16. Does the school follow a daily routine program?

YES ☐

NO ☐

17. Please describe the daily program: _____

_____

Appendices

Information about the facilities at the preschool

18. How many classrooms does the preschool have? _____

19. Do you have running water at your preschool?

YES	☐
NO	☐

20. Do you have electricity at your preschool?

YES	☐
NO	☐

21. Do the children have a playground at the preschool?

YES	☐
NO	☐

22. Do the children have an indoor water facility in the preschool (such as for a basin and washing dishes)?

YES	☐
NO	☐

23. Do the children have an indoor toilet facility in the preschool?

YES	☐
NO	☐

24. How many toilets (indoor or outdoor) are available to the children at the preschool?

_____

25. Do the staff members have their own toilet facility in the preschool?

Appendices

YES	
NO	

26. How many toilets (indoor or outdoor) are available to the staff at the preschool?

_____

27. Is the preschool fenced?

YES	
NO	

28. Does the preschool have these facilities available? Please tick all that apply.

A landline	A telephone	A fax machine	Internet

Nomination of Participants

The goal of the study is to obtain an objective sample of vocabulary used by Sepedi speaking preschool children between the ages of 5 years and 6 years 11 months old. Please nominate two children (a boy and a girl) who speak Sepedi as a first language and in your view have adequate speech and language skills for their age. These children ought to have been at the preschool for at least three months and should attend preschool at least three days a week.

Child's name		
Age		
Gender		

Please send each nominated child's parents a package containing the information letter, consent form and care-giver questionnaire (find attached).

Thank you so much for your time and effort to assist me in the study!

Appendices

D.2. Sepedi version

(Based on Mngomezulu, 2017)

Lebaka la dipotšišo tše ke go hwetša ditaba ka sekolo se le gore se šoma bjang

Letšatši: _____

Lebitso la motho yo a tlatšago: _____

Mošomo wa gago mo sekolong: _____

Lebitso la sekolo se: _____

Taelo: Ke kgopela o arabe potšitšo ye ngwe le yengwe ka go swaya karabo ya nnete.

Ditaba tša go amana le maleme ao a šomišwago mo sekolong

1. Naa Sepedi ke leleme leo le šomišwago go ruta mo sekolong?

EE	
AOWA	

2. Naa Sepedi ke leleme leo le šomišwago go ruta le le tee?

EE	
AOWA	

3. Ge eba ga go jwalo, Maleme a mangwe ao a šomišwago ke afe?

_____

4. Ke leleme lefe leo le šomišwago ke bana ka phaphošing ya gago?

_____

5. Ke maleme afe a mangwe ao a šomišwago ke bana ge ba boledišana?

_____

6. O na le bathuši ba ba kae ka phaposhing? (ge e bag a o nabo, ngwala 0.)

_____

7. Bathuši ba gago ba bolela leleme lefe ge ba bolela le bana?

_____

8. Ke maleme afe a mangwe ao ba a šomišago ge ba bolela le bana?

_____

Appendices

Ditaba tša go amana le bana le tshipidišo ya sekolo

9. Ke bana ba ba kae ka phapošing ya gago? _____
10. Bana ba ka phapošing ya gago ba na le mengwaga e me kae? Mengwaga ya bona e thoma go_____go fihla go _____.
11. Ke bana ba ba kae mo sekolong ge ba feleletše? _____
12. Diphapoši tša mphato wa R ke tše kae?

    _____
13. Mphato wa R wa sekolo sa lena o latela kharikhulamo naa?

EE ☐ Ge eba karabo ke ee, tlhalosa ka botlalo:_____

_____

AOWA ☐

14. Bana ba sekolo sa lena bana le mengwaga e me kae ka moka ga bona?
    Go thoma go _____go fihla go _____.
15. Bana ba ka phapošing ya gago ba hwetša monyetla wa go bolela le go raloka le bana ba ba ngwe mo sekolong? Tlhaloša ka botlalo.

    _____

    _____
16. Sekolo sa lena se latela tshipidišo ye tee tšatši ka tšatši?

EE ☐

AOWA ☐

17. Ke kgopela o tšweletše tshipidišo yeo _____

_____

Appendices

Ditaba tša go amana le sekolo

18. Sekolo se na le phaphoši tše kae tša mphato wa R? _____
19. Le na le meetsi ao a yelelago mo sekolong?

EE ☐
AOWA ☐

20. Le na le mohlagase mo sekolong?

EE ☐
AOWA ☐

21. Go na le lepatlelo la go raloka la bana mo sekolong sa lena?

EE ☐
AOWA ☐

22. Bana ba na le mo go bego le meetsi ka moagong (go swana le bafo mo ba ka hlatswetšago dijwana)?

EE ☐
AOWA ☐

23. Bana ba na le phaphoši ya go botshwelo bja mare ka gare ga moago?

EE ☐
AOWA ☐

24. Bana ba na le di toilete tše kae?

_____

25. Bašomi ba sekolo ba na le toilete ya bona?

Appendices

EE	
AOWA	

26. Gona le ditoilete tše kae tša bašomi? _____

27. Sekolo se na le fense naa?

EE	
AOWA	

28. Ke kgopela o swaye dilo tšeo di bego gona mo sekolong.

Thapo ya mogala wa sekolo	Mogala wa sekolo	Motšhene wa fekse	Internete

Kgetho ya bao ba tlago go šoma le nna

Nyakišišo ye e duma go hwetša sampolo ya mantšu a bana bao ba bolelago Sepedi ba gare ga mengwaga ye 5 le 6 le kgwedi tše 11. Ke kgopela o kgethe bana ba babedi (wa mošimanyana le wa ngwanenyana) bao ba bolelago Sepedi ebile ba na le polelo ya go swanela mengwaga ya bona. Bana ba ba swanetše ba be ba tsene sekolo se kgwedi tše tharo ebile ba etla sekolong matšatši a mararo beke ye ngwe le ye ngwe.

Lebitso la ngwana		
Mengwaga		
Bong		

Appendices

Ke kgopela o romele ngwana wo mongwe le wo mongwe wo a kgethilwego gae le dilo tše di latelago: letlakala la ditaba tša nyakišišo ye, foromo ya tetla ya motswadi le letlakala la dipotšišo ya go ya go motswadi (di hwetše ka morago ga foromo ye).

Ke a leboga gore o mphile nako ya gago le go nthuša!

Appendices

APPENDIX E

INFORMATION LETTERS AND CONSENT FORM FOR PARENTS

E.1 English version

Dear Sir/Madam Date: _____

Re: Permission to conduct research study at your school

My name is Rahab Mothapo. I am currently enrolled for a Master's degree in Augmentative and alternative Communication (AAC) at the University of Pretoria. As part of my studies I want to analyse the vocabulary that Sepedi children typically use in preschool. I therefore want to ask your permission to include your child in such a study.

Why is this study important?

Many children with disabilities have little or no speech. Augmentative and alternative communication (AAC) systems may be introduced to help children communicate. Choosing the appropriate words for such systems is important to ensure that children can use them to communicate in various situations. For this reason I want to find out what words Sepedi-speaking children use, in order to use this as a guideline in selecting the appropriate words for designing AAC systems for children who understand Sepedi but cannot speak.

What will be expected of your child?

Should you give consent for your child to participate in the study, I will meet your child at school and explain the study. I will ask if you child is willing to take part. If he/she agrees to take part, he/she will be expected to wear a pouch around the waist with a voice recorder attached to a small lapel microphone clipped to his/her shirt/top during normal school activities for about 3-5 days.

What are my child's rights?

You and your child can choose to take part in the study or not to take part in the study. You and your child may stop participating in the study at any time. I will make sure that your child understand that he/she can ask the teacher to take off the recording equipment at any time. If you or your child decides to stop taking part, all recordings of your child will be immediately destroyed.

All the information and recordings of your child will be kept safe and will not be shared with anyone. The voice recordings will only be listened to by me, my supervisors, and one independent rater. When I speak or write about the study, no personal information about your child will be shared.

Appendices

What will happen after I collect the information?
The information you share with us about your child and all the recordings made will be securely stored at the University of Pretoria in the Centre for Augmentative and Alternative Communication for 15 years. The information will be used for writing a Master's dissertation, writing scientific papers and for presentation at professional conferences and seminars. No personal information about your child will be shared.

If you want to find out about the results of the study, you can contact me and I will send you a summary.

If another researcher wants to use the recordings, we will first ask you for permission.
What are the risks and the benefits?
At no time during the participation in the research will your child be at risk of any harm. Your child will take part in the normal preschool activities while being recorded so they will not lose out on class time. Your child's class teacher will make sure that your child only wears the recorder when it is safe for him/her to do so. The teacher will also help your child to adjust or remove the recording equipment if it is bothering him/her.

Potential benefits of this study are that it can help us to know what words to program into an AAC system or Sepedi children who cannot speak.

I would appreciate if you would complete the attached form to let me know if give permission for your child to take part or not. Please return the form to your child's preschool.

If you do give permission, would you please also complete the background questionnaire and return it to your child's preschool.

For any further information, please contact me or my supervisor using the contact details supplied below.

Kind regards,

19/01/2018

Ms. N.R.B Mothapo Date
Email: [redacted]
Cell: [redacted]

_____ _____
Dr Kerstin Tönsing Date
Centre for Augmentative and Alternative Communication
Email: [redacted]
Office tel: [redacted]

Appendices

Parent Informed Consent: Reply Slip

Name of Child: _____

Name of Parent/Caregiver: _____

Project title: Determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities

Researcher: Rahab Mothapo
Master's Student
Cell: 079 259 9631

Supervisor: Kerstin Tönsing

I, _____
Name and surname

(Please tick box that applies)

☐ give consent for my child to participate in the study entitled: **Determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities**, conducted by Rahab Mothapo, under the supervision of Kerstin Tönsing. My consent is voluntary and I understand that I may withdraw my child's participation 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 sessions will be video-taped for data collection purposes and may be used for training and conferences. I understand that all information used and obtained in this study will be treated as confidential.

OR

☐ do not give consent for my child to participate in the study entitled: **Determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities**.

_____ _____
Parent's Signature Date

Appendices

E. 2 Sepedi version

Motswadi wa Ngwana Letšatši le Kgwedi: _____

Re: Mabapi le tumelelo ya gore ngwana wa gago a tšee karolo go study goba nyakišišo

Lebitso la ka ke Rahab Mothapo. Ke tsenetši go ithutela degree ye kgolo ya Master's ka lefapheng la Augmentative and Alternative Communication (AAC) kua Unibesithing ya Pretoria. Ka dithutong tša ka, ke rata go nyakišiša gore bana bao ba šomišago Sepedi ba šomiša mantšu a fe ge ba thoma sekolong. Ka gona, ke kgopela tumelelo ya go šoma le ngwana wa gago go hwetša seo.

Bohlokwa bja nyakišišo ye ke eng?

Bana ba bantši ba go se itekaneli ba hloka polelo. Augmentative and alternative communication (AAC) e rutwa go bana go ba thuša ka polelo. Kgetho ya mantšu ao a swanetšego e bohlokwa go thuša bana gore ba kgone go ipolelela ka dinako tšohle. Ka lebaka le, ke rata go tseba gore mantšu ao bana ba go bolela Sepedi ba a šomišago ke afe. Se se thuša barutiši gore ba kgone go kgetha mantšu a Sepedi go thuša bana bao ba sa kgonego go ipolelela.

Ke eng seo se lebeletšwego go ngwana wa gago?

Ge o dumela gore ngwana a tšee karolo go study se, ke tla hlakana le ngwana wa gago kua sekolong kamo hlalosetša gore re tlo bereka bjang. Le yena o tla ba le monyetla wago ikgethela go tšea karolo. Ge a dumetše, o tla apara mokotlana wa goba le voice recorder le microphone ka nako ya sekolo. Re tla dira se matšatši a go lekana 3 go ya go 5.

Ditokelo tša ngwana wa gago ke dife?

Wena le ngwana wa gago le ka ikgethela go tšea karolo goba go se e tšee. Wena le ngwana wa gago le ka emiša go tšea karolo ka nako ye ngwe le ye ngwe. Ke tla netefatša gore ngwana wa gago o a kwišiša gore a ka kgopela Mam wa gagwe gore a hlobole mokotlana wok ka nako ye ngwe le yengwe. Ge le kgethile go emiša gore ngwana a se ke a tšea karolo go ya pele, kgatišo tšeo di bego di tšerwe tša ngwana wa lena di tla lahlwa tša se šomišwe.

Dikgatišo tša ngwana wa gago ka moka di tlo beiwa ka tlhokomelo le šedi ebile di ka se fiwe batho ba bangwe. Dikgatišo di tla theeletšwa ke nna le mogolo wa ka mo dithutong le yo mongwe wa go hlokomela dithuto tše feela. Ge ke tlo ngwala goba ke bolela ka dinyakišišo tše, ga gona taba yeo e hlathago ngwana wa gago yeo ke tla e bolelago.

Appendices

Go tlo direga eng morago ga go hwetša ditaba tše?

Ditaba tšeo re di tšeago go ngwana wa gago le dikgatišo di tla lotwa ka šedi kua Yunibesithi ya Pretoria kua Sentareng ya Augmentative and Alternative Communication mengwaga ye lesome hlano. Ditaba tše di tla šomišwa go ngwala lengwalo la Masters, go ngwala lephephe le go lokišetša dipolelo kua diconferenseng. Ga gona ditaba tša ngwana wa gago tša go mo hlatha tšeo di tla šomišwago.

Ge o nyaka go tseba dipoelo tša dinyakišišo tše, o ka ihlakantšha le nna ka go romela kakaretšo ya tšona. Ge monyakišiši yo mongwe a ka rata go šomiša dikgatišo tše, ke tla kgopela tumelelo go tšwa go wena pele.

Dikotsi le dipoelo tše di botse ke dife?

Ga gona nako yeo ngwana wa gago a ka ikhwetšago a le mo kotsing. Ngwana wa gago o tšwetša pele tša sekolo tša tlwaelo le ge a tšea karolo mo dinyakišišong tše. Morutiši wa ngwana wa gago o tla netefatša gore ngwana wa gago o bolokegile go dira seo se swanetšego. Morutiši o tla thuša ngwana go apola mochene wa go gatiša ge a sa kwane le wona.

Nyakišišo ye e tla thuša gore boramahlale ba kgone hlama mechene ye AAC yeo e ka thušago bana bago bolela Sepedi eupša ba sa kgone go bolela.

Nka thaba ge o ka tlatša foromo yeo e tlago le lengwalo le, go netefatša gore o dumela gore ngwana wa gago a tšee karolo mo. Ke kgopela o buše foromo ye sekolong sa ngwana wa gago.

Ge o sa dumele, ke kgopela o hlatše dipotšišo o di bušetše sekolong sa ngwana wa gago.

Ge le nyaka ditaba ka thuto tše, ihlakantše le nna goba le mogolo wa ka ka go šomiša di contact details tšeo di ngwadilego mo fase.

Wa lena,

_____ 19/01/2017

Ms. N.R.B Mothapo Letšatši
Email: [redacted]
Cell: [redacted]

_____ _____

Dr Kerstin Tönsing Letšatši
Centre for Augmentative and Alternative Communication
Email: [redacted]
Office tel: [redacted]

Appendices

Tumelelo ya motswadi:

Selipi sa go araba

Lebitšo la ngwana: _____

Lebitšo la motswadi: _____

Project title: **Determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities**

Monyakišiši: Rahab Mothapo
Masters student
Cell: [redacted]

Mogolo wa gagwe: Kerstin Tönsing

Nna, _____ (Lebitso le sefane)

(Ke kgopela o swaye e tee)

☐ Ke fa tumelelo/tetla go Rahab Mothapo gore a bereke le bana ba sekolo sa ka mo nyakišišong ya **determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities** yeo e dirwago ke Rahab Mothapo, ka fase ga tebelelo ya Kerstin Tönsing. Tumelelo ye ga se ka gapeleša, ebile ke a kwišiša gore nka gogela morago nako ye ngwe le ye ngwe. Ke a kwišiša gore bana ba tlo theeletšwa ka rekhoda. Ke a kwišiša gore ditaba tše di tlo beiwa menwaga ye 15 kua CAAC le gore ditaba tše di ka se fiwe mang le mang. Ke a kwišiša gore ditaba tše di ka šomišwa gape eupša boitsebišo bja ka bo tla dula sephiring.

GOBA

☐ Ga ke fe tumelelo/tetla go Rahab Mothapo gore a bereke le bana ba sekolo sa ka mo nyakišišong ya **determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities.**

.

_____ _____

Siknatšha ya motswadi Letšatši

Appendices

APPENDIX F

CAREGIVER QUESTIONNAIRE

F.1 English version

QUESTIONNAIRE REGARDING HOME BACKGROUND

(Based on Mngomezulu, 2017)

Date: _____

Child's name: _____

Date of birth: _____

Gender: _____

Name of the person filling in the form: _____

Relationship with the child: _____

Cell phone numbers: _____

Instruction: Kindly answer each question by ticking the preferred option.

Information about the child

3. Does your child speak Sepedi as a home language?

YES	☐
NO	☐

4. Does your child speak other language(s)?

YES	☐
NO	☐

5. If yes, which other languages does the child speak? Please describe: _____

_____

Appendices

6. Are you concerned about your child's

Vision: YES ☐ NO ☐ If yes, please describe: _____
_____

Hearing: YES ☐ NO ☐ If yes, please describe: _____
_____

Walking: YES ☐ NO ☐ If yes, please describe: _____
_____

Talking: YES ☐ NO ☐ If yes, please describe: _____
_____

Thinking: YES ☐ NO ☐ If yes, please describe: _____
_____

7. Are you concerned about anything else about your child's development? ____

8. Do you think your child is currently developing normally for his age?

YES ☐ NO ☐ If not, please describe your concerns: _____
_____

9. At what age did your child begin speaking in single words (e.g. mama, dijo)? Please tick one option.

0-6 months	7-12 months	13-18 months	19-24months	>2 years

Appendices

10. Siblings and /or other children living in the child's household

Gender (male/female)	**Age**	**Relationship to your child**	**Language used mostly by this child**	**Other languages used by this child**

11. Adults living with the child at home

Gender (male/female)	**Age**	**Relationship to your child**	**Language used mostly by this adult**	**Other languages used by this adult**

12. Which language is used **most** in conversations at home?

_____

13. Which other language(s) is/are used in the conversations at home? Please describe

_____

Appendices

12. Does the child enjoy watching the television (TV)?

YES ☐
NO ☐

If yes, to which languages is your child exposed to via TV?

_____

_____

13. Does the child enjoy listening to the radio?

YES ☐
NO ☐

If yes, what languages is your child exposed to via radio? _____

Information about the facilities in the home surroundings

14. Do you have access to electricity in the house?

YES ☐
NO ☐

15. Do you have access to running water in the house?

YES ☐
NO ☐

16. Do you have an indoor toilet facility at home?

YES ☐
NO ☐

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

less than R 6700 ☐
more than R 6700 ☐

Thank you so much for your time and effort to assist me in the study!

Appendices

F. 2 Sepedi version

LETLAKALA LA DITABA TŠA KA GAE

(Go tšwa go Mngomezulu, 2017)

Letšatši: _____

Lebitso la ngwana: _____

Letšatši, kgwedi le,ngwaga ya matswalo: _____

Bong bja ngwana: _____

Lebitso la motho wo a tlatšago foromo ye: _____

Tswalano ya gagwe le ngwana wo: _____

Nomoro ya mogala: _____

Taelo: Tlatša dipotšišo tše ka go kgetha yeo e lego ya nnete

Ditaba ka ngwana

14. Ngwana wa gago o bolela Sepedi bjalo ka leleme la ka gae?

EE	☐
AOWA	☐

15. Ngwana wa gago o bolela dipolelo tše di ngwe?

EE	☐
AOWA	☐

16. Ge e ba karabo ke ee, ke di fe dipolelo tšeo? Di ngwale fa: _____

_____

Appendices

O tshwenyegile ka se sengwe go ngwana wa gago:

Go bona: EE ☐ AOWA ☐ Ge e ba karabo ke ee, tlhaloša fa: _____
_____

Go kwa: EE ☐ AOWA ☐ Ge e ba karabo ke ee, tlhaloša fa_____
_____

Go sepela: EE ☐ AOWA ☐ Ge e ba karabo ke ee, tlhaloša fa:_____
_____

Go bolela: EE ☐ AOWA ☐ Ge e ba karabo ke ee, tlhaloša fa: _____
_____

Go nagana: EE ☐ AOWA ☐ Ge e ba karabo ke ee, tlhaloša fa: _____
_____

17. O tshwenyegile ka se sengwe ka go gola ga ngwana wa gago? ____

18. O nagana gore ngwana wa gago o gola gabotse go swana le bana ba mengwaga ya gagwe?

EE ☐ AOWA ☐ Ge eba karabo ya gago ke ee, tlhaloša fa:_____
_____

19. Ngwana wa gago o boletše lentšu la gagwe la mathomo a na le mengwaga e me kae (e.g. mama, dijo)? Kgetha e tee.

Dikgwedi tše 0-6	Dikgwedi tše 7-12	Dikgwedi tše 13-18	Dikgwedi tše 19-24	Mengwaga ya go feta 2

Appendices

20. Ditaba ka bana ba bangwe ka mo gae

Bong (ngwanenyana goba mošemanyana)	**Mengwaga**	**Tswalano le ngwana wo**	**Polelo yeo ngwana wo a e šomišago kudu**	**Dipolelo tše di ngwe tšeo ngwana wo a di bolelago**

21. Ditaba ka batho ba bagolo ka mo gae

Bong (mosadi goba monna)	**Mengwaga**	**Tswalano le ngwana wo**	**Polelo yeo ngwana wo ae šomišago kudu**	**Dipolelo tše di ngwe tšeo ngwana wo a di bolelago**

22. Ke leleme le fe leo le šomišwago kudu ka gae?

_____

23. Ke maleme afe a mangwe ao a šomišwago ka gae? Tlhaloša fa:

_____

Appendices

12. Ngwana wag ago o rata go lebelela thelebišene na (TV)?

EE ☐
AOWA ☐

Ge e ba karabo ke ee, go šomišwa leleme le fe mo TV? _____

13. Ngwana wag ago o rata go theeletša radio na?

EE ☐
AOWA ☐

Ge e ba karabo ya gago ke ee, ke maleme a fe ao radio e a šomišago?

_____

Ditaba ka tšeo di bego ka gae

14. Le na le mohlagase ka gae?

EE ☐
AOWA ☐

15. Le na le meetsi ago ela ka gae?

EE ☐
AOWA ☐

16. Le na le botshwelo bja mare ka mo gae?

EE ☐
AOWA ☐

17. Laetša gore masheleng ao a šomišwago ka gae kgwedi ka kgwedi a wela kae.

Ka fase ga R 6700 ☐
Ka godimo ga R 6700 ☐

Ke leboga nako le thušo tša gago!

APPENDIX G

CHILD ASSENT SCRIPT

G.1 English version

Child Assent Script

Hello, my name is Rahab.

I want to find out more y about the words children use during school activities. I want to ask you if you want to help me with that. If you say yes, this is what we will do:

I will ask you to carry a small machine (voice recorder) in a bag that you will wear around your waist like this (demonstrate). I will clip a microphone to your shirt. I will record all the words you say to your friends and your teacher so that I can listen to the words you will use throughout the day. Only I and someone helping me will listen to the tape. I will not let anyone else listen to it.

If the recorder or microphone makes you feel uncomfortable, don't try to fix it yourself. Ask your teacher to help you

If you want to stop wearing the recorder and microphone, ask your teacher to take it off. Your teacher will take it off. Nobody will be angry with you if you want to stop.

You can choose to wear the recorder or not.
Nothing bad will happen to you if you don't want to wear it.

Appendices

G.2 Sepedi version

Foromo ya tumelo go tswa go ngwana

Dumela, lebitso la ka ke Rahab.

Ke rata go tseba ka mantsu ao bana ba a somisago ge ba le sekolong. O ka kgona go nthusa? Ge o dumela, re tlo dira so:

Ke tlo go kgopela gore o apare mochene wo ka gare ga pekana mo nokeng ya gago so (bontsha).Ke tlo go apesa sepikara se mo sekipeng sa gago. Ke tlo go theeletsa ge o bolela le bakgotsi ba gago le morutishi wa gago. Seo se tlo kwa ke nna le wo ke berekago le yena feela. Nka se dumeleli yo mongwe a theeletsa akere?

Ge o ekwa o sa kgotsofale, o se ke wa leka go e lokisa kabo wena. Kgopela morutisi gore a go thuse.

Ge o se sa nyaka go apara mochene, botsa morutisi gore a go hlobole wona. Ga gona motho wo a tlo go kwatela ge o nyaka go ema.

O kano ikgethela go apara mochene goba go se o apare. Ga gona se sempe seo se tlilego go go hlagela.

APPENDIX H

CHILD-FRIENDLY RESPONSE FORM

H.1 English version

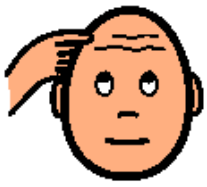
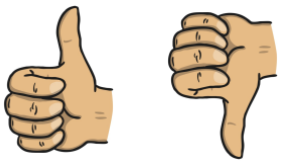
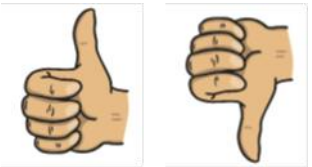
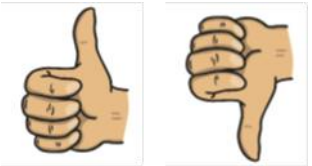
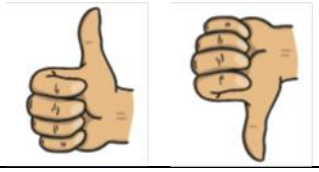
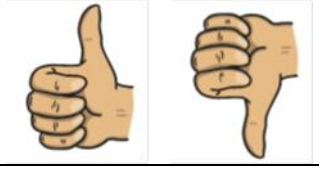
Name: _____

Date of birth: _____

Date: _____

Name of the study: Determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities

Researcher: Rahab Mothapo

	Do you understand everything I explained to you now? Yes No 
	Do you understand that you can choose to do this or not to? Yes No 
	Do you understand that you can stop when you want to? Yes No 
	Do you understand that I will record the words that you say? Yes No 
	Do you want to ask me any questions? Yes No 

Appendices

Are you happy with the way I answered your questions?

Yes   No

Do you want to be part of the study?

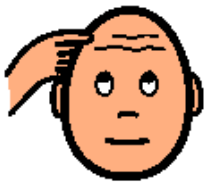
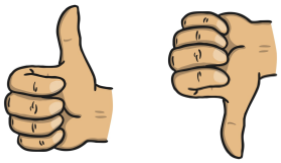
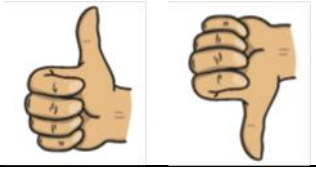
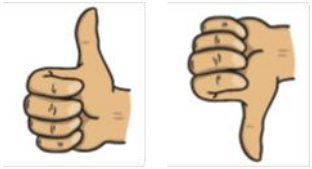
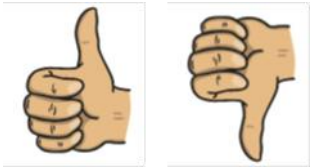
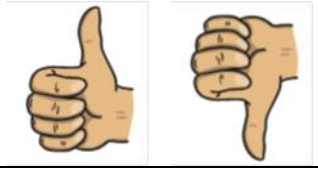
Yes   No

Appendices

H. 2 Sepedi version

Lebitso: _____
Letšatši,kgwedi le ngwaga tša matswalo: _____
Letšatši: _____
Lebitso la nyakišišo: Determining the core vocabulary used by Sepedi speaking preschool children during regular preschool-based activities

Monyakišiši: Rahab Mothapo

	O kwišiša tšohle tšeo ke go hlaloseditšego? **Ee** **Aowa** 
	O a kwišiša gore o ka kgetha go šoma le nna goba go se šome le nna? **Ee** **Aowa** 
	O a kwišiša gore o ka ema ge o nyaka? **Ee** **Aowa** 
	O a kwišiša gore ke tlo theeletša seo o se bolelago? **Ee** **Aowa** 
	O nyaka go mpotšiša se sengwe? **Ee** **Aowa** 

Appendices

O thabile ka mokgwa wo ke arabilego potšišo ya gago?

Ee **Aowa**


O ka rata go bereka le nna?

Ee **Aowa**


Appendices

APPENDIX I

TEACHER INSTRUCTIONS FOR RECORDINGS AND EQUIPMENT

I.1 English version

TEACHER INSTRUCTIONS:

1. Please ensure that the child has the lapel microphone on the chest area and the voice recorder in a pouch around their waist during the recording time.

2. Please ensure that the child does not play with the voice recorder or interfere with the device at any given time.

3. Kindly remove the voice recording device if the child says it causes discomfort or if it is annoying him/her. You can also do so at any time you feel it is unsafe or unsuitable for the child to have the device on.

4. Kindly remove the device if the participant engages in physical activity that may cause damage to the device (i.e. playing on the jungle-gym or on a swing).

5. Please check periodically that the recorder is switched on – we would appreciate if you could do this every two /three hours.

6. I will leave two extra batteries with each teacher, please insert them (as I've shown you) if the device suddenly runs out of power.

Feel free to call/send me a message if you are uncertain about anything during the recording time. I will phone you back/meet you to resolve the issue. My phone number is 079 259 9631.

Yours sincerely

Mothapo N.R.B

Appendices

I. 2 Sepedi version

TAELO TŠA MORUTIŠI:

1. Netefatša gore ngwana o apere microphone mo kgareng ya gagwe le voice recorder e ka gare ga mokotlana mo nokeng ya gagwe ka nako yeo a rekhodiwago ka yona.

2. Netefatša gore ngwana ga a raloke ka motšhene goba ao kgotlakgotla ka dinako tšohle.

3. Ge ngwana a ka re motšhene gao mo sware gabotse, o ka mo tloša wona. O ka mo tloša wona ge wena o bona okare ga se a swanela go o apara ka nako yeo goba o ka hlola kotsi.

4. Tloša ngwana motšhene wo ge a raloka meraloko yeo e ka senyago motšhene (go swana le mo lepatlelong leo ba namelago ba fologa).

5. Netefatša gore recorder yeo e tšhumilwe ka dinako ka moka. O tla netefatša se ka morago ga iri tše pedi goba tše tharo nako le nako.

6. Ke tlo tlogela dibetiri tše pedi le morutiši wo mongwe le wo mongwe. Ke kgopela o di lokele mo motšheneng ge di fedile sebakabakeng.

7. Lokologa go ntshwara ka mogala ge o sa kwišiši se sengwe ke sego. Le molaetša o ka nthomela wona. Ke tla go founela ka lokiša bothata bjo o nago le bjona goba ke tla tla ka sebele. Dinomoro tša ka ke 079 259 9631.

Wa lena

Mothapo N.R.B

Appendices

APPENDIX J

DAILY FEEDBACK FORM FOR TEACHERS

The purpose of this review is to establish whether the school followed its usual program on the day or not and to identify any challenges during daily recordings at the school

Respondent Name: _____

Date: _____

Instruction: Kindly answer each question by ticking the preferred option.

Information on how the recording day went

1. Did you follow the daily routine program today?

YES	
NO	

2. If not, please describe:

_____

_____

3. Did you encounter any problems during recording time today?

YES	
NO	

4. If yes, please describe:

_____

_____

5. How did you resolve the problem?

_____

_____

Thank you so much for your time and effort to assist me in the study!

APPENDIX K
TRANSCRIPTION RULES

Transcription rules based on Trembath et al. (2007),

1. The first 20 minutes of the recording will not be transcribed to rule out the novelty effect.
2. Any comments from participants referring to the recording equipment or the investigation process will be omitted from transcription and final analysis.
3. Utterances will be transcribed individually. Utterance boundaries (words) will be defined by intonation or a pause of greater than 2 seconds.
4. C will be used at the beginning of each line to indicate the child's utterance as required by the SALT program.
5. Every statement will be ended with a punctuation mark. Most will end with a full stop to indicate the end of a statement/utterance, while some will end with a question mark to indicate questions. Interrupted utterances will be indicated by ^.
6. Numbers will be typed as words.
7. Syllable and sound repetitions will be transcribed as an individual word as though the repetition never occurred.
8. Vocalisations that perform the function of filler (e.g. uhmm), a request for clarification (e.g. "heh?") or an indication of agreement/disagreement (e.g. "Hah-eh") will be transcribed in a phonetically consistent form and counted as words. Prolongations of words will be typed as normal words.
9. Imitated noises (e.g., engine sounds, animal sounds) will not be included.
10. Swearwords will be transcribed like other words.
11. Book names and cartoons and production titles consisting of multiple words (e.g., 'Mokgalabje wa sefofu') will be transcribed as one word using an underscore symbol.
12. Words used in songs and repetitive games (e.g., 'Happy birthday to you') will not be included. However, other words spoken by the participants within the songs or rhymes will be included.
13. All children's names spoken will be represented by (CN).
14. All the adult's names spoken will be represented by (TN) and proper nouns (referring to locations) will be identified by (PN) to protect the safety and confidentiality of participants.

15. A single word will be transcribed by a single X; an unintelligible part of an utterance of varying length (more than one word) will be transcribed by XX and a full unintelligible sentence will be represented by XXX.
16. Spelling rules will be adhered to at all times.
17. Code switching will be identified by [CS] for the transcription process and will be included in the word count.

APPENDIX L

CODING RULES

CODING RULES

1. The code [CS] is used for code-switching (which occurs when two languages are used within a sentence or discourse). Therefore this will be applied whenever the participant uses English or any other language in utterances.
2. Morphological variations (inflectional forms) of certain words will be transcribed in such a way that SALT is able to identify the lemma (base form) or root and the different variations. The lemma or root will always be transcribed first, followed by a forward slash and the morpheme or inflected form. Words describing action to the first person will be re-written after a forward slash (e.g. rata/nthate and pusha/mpusha). Verb forms that alter two vowels are also re-written (e.g. swara/swere and apara/apere) Grammatical variations that will be transcribed in this way are the following:

	Lemma	**Example**	**Grammatical variations**	**Example**	**Code in sample**	**Example in sentence**
Nouns	Singular form	Kereke	Plural	Dikereke	Kereke/di	Dikereke tše pedi di tsena ka nako e tee.
			Locative	Kerekeng	Kereke/ngp	Ke be ke le mo kerekeng.
Verbs	Imperfect indicative	Rata	Imperfect tense	Rata	Rata	Ke rata koloi ye.
			Negative form	Rate	Rata/e	Ga ke rate koloi ye.
			Perfect tense	Ratile	Rata/ile	O ratile mpho ya gagwe.
			Negative form	Rata	Rata/e	Ga se a rata mpho ya gagwe.
			Object concord indicating action to the first person (Occlusivation-sound strengthening)	Nthata	Rata/nthata	Rakgadi o a nthata.
			Negative form	Nthate	Rata/nthate	Rakgadi ga a nthate.

	Lemma	Example	Grammatical variations	Example	Code in sample	Example in sentence
			Applied verbal extension(causative +assistive)	Ratela	Rata/ela	Wena ke go ratela borokgo bjo.
			Negative form	Rateli	Rata/eli	Wena ga ke go rateli borokgo bjo.
			Causative verbal extension	Ratiša	Rata/iša	Teye ke e ratiša ke koko wa ka.
			Negative form	Ratiši	Rata/iši	Teye ga ke e ratiši ke koko wa ka.
			Verb + plural marker	Ratang	Rata/ng	Rakgadi wo ke mo ratang.
			Negative form	Rateng	Rata/eng	Rakgadi wo ke sa mo rateng.
Imperative mood			Singular	Eya	Ya/eya	Eya!
			With plural marker	eyang	Ya/eyang	Eyang!

The examples given are the words found in the participants' transcriptions. This does not mean that these are the only examples existing in the Sepedi language.

3. The Sepedi language also has heteronyms and polysemous words. Although these two phenomena differ slightly from each other, they all describe words that are spelled the same (though not necessarily pronounced the same) but that have slightly or completely different meanings, and are often also classified as different parts of speech. It was deemed important to differentiate heteronyms and polysemous words in order to avoid over-counting. An arbitrary code was therefore assigned to allow SALT to differentiate between such words. The table overleaf illustrates the words that have the same spelling but different meanings found in the sample and how they were coded.

Appendices

Word	**Part of speech**	**English translation (approximate)**	**Coding in sample**	**Example in sentence**
A	Concord	He/she/them/of	A[U]	Ga se a tliša dijo.
	Demonstrative particle	These	A[V]	Magapu a ke a rena.
	Imperfect particle	No translation	A[W]	Baithuta ba a lwa.
	Hortative particle	No translation	A[X]	A re boneng!
	Past tense morpheme	No translation	A[Y]	Ga ba a tliša mmereko.
Ba	Concord	They/them/of	Ba[Z]	Baithuti ba itukišetša hlahlobo/ Mma o a ba betha.
	Demonstrative particle	These	Ba[1]	Batho ba ba bolela maaka.
	Copulative verb	Become/be	Ba[2]	Motho e ba leloko la geno ka madi.
Ga	Negative morpheme	No translation	Ga[3]	Ga ke nyake dijo.
	Concord	Of	Ga[4]	O beile setulo godimo ga tafola.
	Locative particle	At	Ga[5]	Kopano e ga Mothapo.
	Hortative particle	No translation	Ga[6]	Ga go tumišwe Morena.
Go	Indefinite subject concord	No translation	Go[R]	O ya sekolong ka fao go ra gore o tlo hloka tshelete.
	Infinite prefix	To	Go[S]	Dikobo tša gago tša go robala.
	Object concord 2nd person singular	No translation	Go[T]	Ga se ka go kwa ge o tsena.

Appendices

Word	**Part of speech**	**English translation (approximate)**	**Coding in sample**	**Example in sentence**
Ka	Prepositions	With/About/Through	Ka[E]	Ntšhi e tšwa ka lefasetere.
	Potential morpheme	No translation	Ka[F]	O ka dira eng?
	Possessive pronoun	Mine	Ka[G]	Nko ye ke ya ka.
Ke	Copulative particle	No translation	Ke[C]	Molato ke eng?
	Subject concord 1st person singular (with a pronomial function)	No translation	Ke[D]	Ke sa sepela.
Le	Connective particle	With/and	Le[H]	Ke sepela le Rakgadi.
	Subject/object concord Class S	No translation	Le[I]	Lelepola le robegile/ Monna o a le aga leboto.
	Subject concord 2nd person plural	You	Le[J]	Le sa ya moketeng?
	Demonstrative copulative	This	Le[K]	Lelepola le.
Mo	Concord	Her/him	Mo[7]	Monna o a mo rata mosadi wa gagwe.
	Demonstrative particle	Here	Mo[8]	Ke mengwaga ke bereka mo.
	Locative particle	On /at	Mo[9]	E bee mo fase.
Re	Verb	Say	Re[P]	Ke be ke re dumela.
	Subject concord) first person plural	We /us	Re[Q]	Re tla kopana gona.

Appendices

Word	**Part of speech**	**English translation (approximate)**	**Coding in sample**	**Example in sentence**
Sa	Concord	He/she/it/of	Sa[10]	Sefatanaga sa ema pele ga ntlo/ selepe sa Tate.
	Aspectual prefix	Still	Sa[11]	Ke sa emetše Karabo ya ka.
	Negative morpheme	No translation	Sa[12]	Ge ba sa nyake, nka se gapeletše.
	Verb	Become light	Sa[13]	Go tla re go sa ra sepela
Tla	Verb	Come	Tla[14]	Ke tla ya gae.
	Future morpheme	will	Tla[15]	Lerato o re a ka tla.
Se	Demonstrative copula	This	Se[N]	Seeta se.
	Negative morpheme	Won't	Se[O]	Nka se go boni.
Ya	Verb	Go	Ya[A]	Ke sa ya kerekeng.
	Possessive concord	Of	Ya[B]	Puku yeo ke ya ka.

APPENDIX M

CORE VOCABULARY LIST WITH ACCOMPANYING FREQUENCY AND COMMONALITY SCORE

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
O	concord	Structure	1046	59,5367	6	
CN	noun	Content	760	43,258	6	
KE[D]	concord	Structure	727	41,3797	6	
GO[T]	concord	Structure	529	30,1099	6	
KE[C]	copulative particle	Structure	444	25,2718	6	
A[W]	present tense morpheme	Structure	416	23,6781	6	
NNA	pronoun	Structure	381	21,6859	6	
LE[H]	conjunction	Structure	335	19,0677	6	
WENA	pronoun	Structure	319	18,157	6	
E	concord	Structure	311	17,7016	6	
BA[Z]	concord	Structure	301	17,1324	6	
GA[3]	negative morpheme	Structure	291	16,5633	6	
RE[Q]	concord	Structure	287	16,3356	6	
KA[E]	preposition	Structure	278	15,8233	6	
YA[A]	verb	Content	276	15,7095	6	Yang(2), eya(16), eyang(1), ile(1) , nkiše(1), yeng(1), ye(62)
BONA	verb	Content	251	14,2865	6	Bone (65), boneng (4), boni(3),

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
						mmone(1), mmoni (3), mmona (4), mmontšha (2), mpona (4), mpontšhe (1), bonang (2), bontšhe (2)
TLO	future morpheme	Structure	233	13,262	6	
YA[B]	concord	Structure	228	12,9774	6	
A[U]	concord	Structure	211	12,0098	6	
SE[O]	Negative morpheme	Content	200	11,3837	6	
WA	concord	Structure	191	10,8714	6	
RE[P]	verb	Content	173	9,8469	6	
MO[8]	demonstrative particle	Structure	170	9,6761	5	
TLA[14]	verb	Content	170	9,6761	6	Tle(7), tlela (1), tlile (17),etla (77), etlang(3)
KA[G]	pronoun	Structure	150	8,5378	6	
YE	pronoun	Structure	138	7,8547	6	
DIRA	verb	Content	129	7,3425	6	Dire (16) , direla (2), direle (2), direng (1), diri (2), dirile (8), dirileng (2), ndira (1), ndire (2), ndirele (1) dirang (3), ntira (1)
AH	interjection	Structure	125	7,1148	6	
NYAKA	verb	Content	121	6,8871	6	Nyake (31), nyakile (3), nyakang (1), nnyaka (2)
A[X]	hortative particle	Structure	120	6,8302	6	
TN	noun	Content	118	6,7164	6	

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
EE	interjection	Structure	117	6,6595	6	
LE[J]	concord	Structure	111	6,3179	6	
AKERE	interjection	Structure	104	5,9195	6	
ENG	noun	Content	104	5,9195	6	
WO	pronoun	Structure	104	5,9195	6	
GAGO	pronoun	Structure	103	5,8626	6	
NGWALA	verb	Content	101	5,7488	6	Ngwadile (17), ngwale (6), ngwalela (2), ngwalelele (1), ngwaleng (1), ngwaletše (1)
MANG	noun	Content	99	5,6349	6	
NTO	noun	Content	89	5,0657	6	
SO	adverb	Content	88	5,0088	6	
DULA	verb	Content	84	4,7811	6	Dule (12), dulang (11), dutše (5) duletšeng (1)
NGWANA	noun	Content	83	4,7242	6	Bana(14)
NA	verb	Content	80	4,5535	6	
TŠEA	verb	Content	79	4,4966	6	Tšee (13) tšeela (1) tšeere (9) and ntšeetše (1)
BOTŠA	verb	Content	78	4,4396	6	Boditše (7), botše (24), botšeng (2), mmotša (3), mmotše (5), mpotše (3), botšang(1)
DI	concord	Structure	74	4,212	6	
AOWA	interjection	Structure	73	4,155	6	

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
MO[7]	concord	Structure	73	4,155	6	
TLALEYA	verb	Content	73	4,155	6	Tlaleye (2), tlaleile (1)
HEH	interjection	Structure	72	4,0981	6	
LA	concord	Structure	71	4,0412	6	
BETHA	verb	Content	70	3,9843	6	Bethe (7), bethela (1), bethi (2), bethile (3), mmetha (9), mmethe (2), mmethile (3), mpetha (6), mpethe (1), mpethela (3), mpethile (2), bethana (1), bethane (1)
KA[F]	potential morpheme	Structure	70	3,9843	6	
MMATA	noun	Content	64	3,6428	6	
JA	verb	Content	62	3,5289	6	Jang(1), je(13), eja(5), jela(1), jele(3), jeng(3), jetše (1), ješa(2), njela(1)
MOTHO	noun	Content	61	3,472	6	Batho(23)
GAPE	Adverb	Content	60	3,4151	6	
KAE	Adverb	Content	60	3,4151	6	
TSEBA	verb	Content	59	3,3582	6	Tsebe (33), tsebeng (1), ntseba (1)
HA-EH	interjection	Structure	56	3,1874	6	
GORE	conjunction	Structure	55	3,1305	6	
TŠA	concord	Structure	55	3,1305	6	
KGONA	verb	Content	51	2,9028	6	Kgone (7), kgoni (11), kgonne (1)

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
SE[N]	pronoun	Structure	49	2,789	6	
YENA	pronoun	Structure	49	2,789	6	
BOLELA	verb	Content	48	2,7321	6	Bolele (2), boleli (4), mpolediša (2), bolelang (1), boletše (3), boletši (1)
LENA	pronoun	Structure	48	2,7321	6	
KUA	locative particle	Structure	47	2,6752	6	
NGWE	adjective	Content	43	2,4475	6	bangwe (8), bongwe (1), dingwe (1), engwe (1), gongwe (2), lengwe (2), mongwe (23), sengwe (2), yengwe (3)
TŠE	pronoun	Structure	43	2,4475	6	
RENA	pronoun	Structure	42	2,3906	6	
SWARA	verb	Content	42	2,3906	5	Sware (3), ntšhwara (5), ntšware (1), ntšhwareng (1), swere (14),
FA	verb	Content	41	2,3337	5	Fe (9), efang (1), file (10), mpha (2), mphe (1)
KGALE	noun	Content	39	2,2198	5	
GE	conjunction	Structure	38	2,1629	5	
SA[10]	concord	Structure	38	2,1629	5	
TSENA	verb	Content	38	2,1629	5	Tsene (5), tseneng (2), tseni (5), tsenang (1), tsenwe (1)
NGWANENYANA	noun	Content	37	2,106	5	Banenyana(17)
LEINA	noun	Content	36	2,0491	5	Maina(4)

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
EH	interjection	Structure	35	1,9921	5	
SA[11]	aspectual prefix	Structure	35	1,9921	6	
BE	verb	Content	34	1,9352	5	
BEA	verb	Content	34	1,9352	6	Beye (4), beyeng (1), beile (6), beileng (1), mpeya (1), mpeye (1).
DLALA	verb	Content	34	1,9352	4	Dlale (9), dlaleli (1), dlaleng(1)
KWA	verb	Content	34	1,9352	4	Kwe(8), ekwang (1), kwele(5)
MAAKA	noun	Content	34	1,9352	5	
GAFA	verb	Content	33	1,8783	5	Ngafiša (2)
GAGWE	pronoun	Structure	33	1,8783	5	
GONA	pronoun	Structure	33	1,8783	5	
YELA	pronoun	Structure	33	1,8783	5	
BO	concord	Structure	32	1,8214	6	
EBILE	conjunction	Structure	32	1,8214	5	
GA[5]	locative particle	Structure	32	1,8214	5	
KGOPELA	verb	Content	32	1,8214	6	Kgopele (3), kgopeli (1), kgopetše (1)
MARA	conjunction	Structure	31	1,7645	6	
SELO	noun	Content	31	1,7645	6	Dilo(15)
TLA[15]	future morpheme	Structure	31	1,7645	6	
THOMA	verb	Content	30	1,7076	5	Thomela (1), thomile (4), thomiše

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
						(2), nthomela (1)
TSAMAYA	verb	Content	30	1,7076	6	tsamaye (1), tsamaile (2), tsamayang(2)
LE[I]	concord	Structure	29	1,6506	5	
TLOGELA	verb	Content	29	1,6506	5	tlogele(2), tlogeleng (1), tlogelang (3), ntlogele (8),ntlogeleng (1)
TŠWA	verb	Content	29	1,6506	6	Tšwe (3), tšwele (7)
BAPALA	verb	Content	28	1,5937	4	Bapadiša (2), bapale (5), bapalang (1)
FASE	noun	Content	28	1,5937	5	
FETŠA	verb	Content	28	1,5937	5	Feditše(10), fetše(3), fetšeng(1), fetši(5)
GO[S]	infinite prefix	Structure	28	1,5937	5	
SA[12]	negative morpheme	Structure	28	1,5937	6	
SEKOLO	noun	Content	28	1,5937	5	Sekolong(20)
MFANA	noun	Content	27	1,5368	4	
MO[9]	locative particle	Structure	27	1,5368	5	
PELE	noun	Content	27	1,5368	4	
TLOGA	verb	Content	27	1,5368	5	Tloge (2), tlogetše (1), tlogile (2), tlogang (1)
YONA	pronoun	Structure	27	1,5368	4	
WOLA	pronoun	Structure	26	1,4799	4	
APARA	verb	Content	25	1,423	4	Apere (5), apare (2), apeše (1), apeša (1)

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
FELA	verb	Content	25	1,423	5	Fedile(2), feditše(1), feleletše(2), fetša(10), fetše(1)
RA	verb	Content	25	1,423	4	Re(1), rela(1), rile(1)
RATA	verb	Content	25	1,423	4	Rate (4), ratile (1)
GEŠO	pronoun	Structure	24	1,366	6	
LE[K]	pronoun	Structure	24	1,366	5	
MAMA	noun	Content	24	1,366	3	
MONNA	noun	Content	24	1,366	3	
SESI	noun	Content	24	1,366	3	
EMA	verb	Content	23	1,3091	3	Eme (7),emeng (1), nkemele (1)
KOTO	adjective	Content	23	1,3091	3	koto (9), bokoto (7), dikoto (1), lekoto (2), makoto (2), mokoto (2)
MAMAKA	noun	Content	23	1,3091	3	
NAA/NA	question particle	Structure	23	1,3091	5	
NTŠHA	verb	Content	23	1,3091	3	Ntšhe (4), ntšhegile (1), ntšhitše (4)
NYANE	adjective	Content	23	1,3091	3	bannyane (3), dinnyane(3), ennyane (1), mannyane (2), mennyane (1), monnyane (4), nnyane (9)
SEETA	noun	Content	23	1,3091	3	Dieta(19)
TŠHABA	verb	Content	23	1,3091	3	Tšhabe (2), tšhabela (1), tšhabang (1)
BITŠA	verb	Content	22	1,2522	4	Bitše (1), mmitše (1), mpitša (3)

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
MORAGO	noun	Content	22	1,2522	4	
MOŠIMANE	noun	Content	22	1,2522	3	Bašimane(16)
ŠE	pronoun	Structure	22	1,2522	4	
LEBELELA	verb	Content	21	1,1953	3	Lebelele (1), lebeleletše (1), lebeleli (1), ntebelela (2), ntebelele (2), ntebeletšeng (1), ntebella (2), lebeletše (3),
MEETSE	noun	Content	21	1,1953	3	
SETULO	noun	Content	21	1,1953	3	Ditulo(1), setulong(2)
LLA	verb	Content	20	1,1384	3	lle (1), llela (5), lliša(1)
MAABANE	noun	Content	20	1,1384	6	
MOKA	noun	Content	20	1,1384	4	
NO	aspectual prefix	Structure	20	1,1384	3	
ROTA	verb	Content	20	1,1384	3	Rote (2), rotetše (2), rotile (3), rotileng (1), ithotela (1), ithotetše (1)
SEPELA	verb	Content	20	1,1384	3	Sepele (3), sepeleng (1), sipidiše (1), sepetše (3)
BA[2]	verb	Content	19	1,0815	6	
BOGOBE	noun	Content	19	1,0815	4	
BULA	verb	Content	19	1,0815	4	Bule (2), butše (1), butšwe (5)
GENO	pronoun	Structure	19	1,0815	4	

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
ROBALA	verb	Content	19	1,0815	4	Robale (3), robaleng (2), robaletše (3), robalang (3)
LELEKERE	noun	Content	18	1,0245	3	Malekere(10)
LETSOGO	noun	Content	18	1,0245	3	Matsogo (11)
SWANA	verb	Content	18	1,0245	4	Swane (2)
TLIŠA	verb	Content	18	1,0245	4	Tliše (3), tlišitše (1)
KGETHA	verb	Content	17	0,9676	4	Kgethe (2), ikgethela (1), kgethile (2)
TLAPA	verb	Content	17	0,9676	4	Tlape (2)
LEOTO	noun	Content	16	0,9107	4	Maoto (14)
LERAGO	noun	Content	16	0,9107	3	Marago (11)
MAMAGO	noun	Content	16	0,9107	4	
RALOKA	verb	Content	16	0,9107	3	Raloke (3), ralokeleng (1)
SEGA	verb	Content	16	0,9107	3	Segiša(6), segiši(1)
TŠHELA	verb	Content	16	0,9107	3	Tšhele (1), tšhelela (1), ntšhela (5), tšhetši (1)
APEŠA	verb	Content	15	0,8538	3	Apeše (3), apešitše (1), nkapeša (1), nkapešang (1), nkapeše (1), nkapešitše (1)
BOELA	verb	Content	15	0,8538	4	Boele (1), boeleng (2)
FETA	verb	Content	15	0,8538	4	Fetile (4)
GANA	verb	Content	15	0,8538	3	Gane (1), nkganeditše (1), nkganetša (1)

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
GATA	verb	Content	15	0,8538	3	Gate (2), gateng (1), gatiwe (3), gatana (1), nkgatha (1)
HAAY	interjection	Structure	15	0,8538	3	
HLOBOLA	verb	Content	15	0,8538	3	Hlobole(2), hlobotše(1)
HLOGO	noun	Content	15	0,8538	3	Hlogong(1)
HMM	interjection	Structure	15	0,8538	3	
KGANTHE	conjunction	Structure	15	0,8538	3	
KOLOI	noun	Content	15	0,8538	4	Dikoloi(1)
OHO	interjection	Structure	15	0,8538	3	
RAGA	verb	Content	15	0,8538	3	Ragela(1), nthaga (1)
REKELA	verb	Content	15	0,8538	3	Rekele (2), ntheketše (1), nthekela (2), reketše (5)
GABO	pronoun	Structure	14	0,7969	3	
HEY	interjection	Structure	14	0,7969	5	
NAMELA	verb	Content	14	0,7969	3	Nameleng (1), nametše (1)
NGWATHA	verb	Content	14	0,7969	3	Ngwathe(1), ngwathela(3), ngwathele(3)
ŠULE	pronoun	Structure	14	0,7969	3	
ILE	verb	Content	13	0,7399	4	
NKGA	verb	Content	13	0,7399	4	Nkgela (1), nkgele (1)
TŠONA	pronoun	Structure	13	0,7399	5	

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
WA	verb	Content	13	0,7399	4	We(2) ,wele(5) ,wiša(1), wišitše(1)
ADIMA	verb	Content	12	0,683	3	Adime (2), nkadime (10)
NKA	future morpheme	Structure	12	0,683	5	
NTŠI	adjective	Content	12	0,683	5	ntši (1), bantši (2), bontši (1), dintši (3), gantši (1), mantši (4)
SEKHIPHA	noun	Content	12	0,683	5	Dikhipha(1), Sekhipheng(2)
TOILET[CS]	noun	Content	12	0,683	5	
WHY[CS]	conjunction	Structure	12	0,683	3	
APEYA	verb	Content	11	0,6261	5	Apeye(1), apeile(1)
AYEYE	interjection	Structure	11	0,6261	4	
BOLO	noun	Content	11	0,6261	4	
FOŠA	verb	Content	11	0,6261	4	Foše(2), fošetša(3), fošetšang(1), fošetše(2)
GABOTSE	adverb	Content	11	0,6261	5	
GOLO	adjective	Content	11	0,6261	5	kgolo (8), legolo (1), magolo (1), mogolo (1)
GOPOLA	verb	Content	11	0,6261	3	
HLAPA	verb	Content	11	0,6261	5	Hlape (2), hlapela (2), hlapile (2)
IŠA	verb	Content	11	0,6261	5	Iše (1), išitše (1)
PEDI	adjective	Content	11	0,6261	5	Babedi (5), mabedi (2)
YOH	interjection	Structure	11	0,6261	3	

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
BONTŠHA	verb	Content	10	0,5692	4	Bontšhe(10)
FIŠA	verb	Content	10	0,5692	4	Fiši (1)
GA[4]	concord	Structure	10	0,5692	5	
GAE	noun	Content	10	0,5692	5	
JERSEY[CS]	noun	Content	10	0,5692	3	
KAKA	verb	Content	10	0,5692	5	Kake(1), kakela(1), kaketše(2), ikaketše(1)
KUDU	adverb	Content	10	0,5692	4	
MPYA	noun	Content	10	0,5692	3	Dimpya(2)
PN	noun	Content	10	0,5692	4	PN(10)
ROGA	verb	Content	10	0,5692	4	Rogana(8), nroge(1)
SEJO/DI	noun	Content	10	0,5692	4	Dijo (10)
SORI	interjection	Structure	10	0,5692	4	
TŠHILA	noun	Content	10	0,5692	3	Matšhila(2)
BOKAKA	noun	Content	9	0,5123	3	Makaka(7)
FIHLA	verb	Content	9	0,5123	4	Fihlile (3)
FORA	verb	Content	9	0,5123	4	
GARE	noun	Content	9	0,5123	4	
GOBATŠA	verb	Content	9	0,5123	4	Gobatše(2), ngobatša(5)
JWANG	adverb	Content	9	0,5123	4	

Appendices

Sepedi word	Part of speech	Content/structure	Number of occurrence	Frequency of occurrence ‰	Commonality score	Inflected grammatical variation (and number of occurrences)
KUKA	verb	Content	9	0,5123	5	Kuke(5), nkukang(1)
MOLA	pronoun	Structure	9	0,5123	3	
MOSADI	noun	Content	9	0,5123	3	
PARTY	noun	Content	9	0,5123	3	diParty(1)
TSENELELA	verb	Content	9	0,5123	4	Tseneleli(1), ntsenelela(1)

APPENDIX N

COMPARISON OF CORE VOCABULARY BETWEEN SEPEDI AND ISIZULU

N. 1 Adjectives

isiZulu		Sepedi			Units that overlapped in meaning
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
Adjective root BI	ugly; bad; evil				0
Adjective root BILI	two	PEDI	Adjective	two	2
Adjective root FIVE	five				0
Adjective root FOUR	four				0
Adjective root HLE	good; beautiful; pretty				0
Adjective root KHULU	large; great	GOLO	Adjective	big	2
Adjective root NCANE	small; few; young	NYANE	Adjective	small	2
Adjective root NE	four				0
Adjective root NYE	some; other	NGWE	Adjective	Another	2
Adjective root ONE	one				0
Adjective root THATHU	three				0
		NTŠI	Adjective	many	0
		KOTO	Adjective	thick/bulky	0

Totals: Number of isiZulu items: 11 Number of Sepedi items: 6 Overlapping items: 8 Non-overlapping items: 9

Appendices

N. 2 Adverbs/adverbial roots

isiZulu		Sepedi			Units that overlapped in meaning
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
Adverb FUTHI	again; once more; perpetually	GAPE	Adverb	again	2
Adverb MANJE	now; at the present time				0
Adverb NAMUHLA	Today				0
Adverb PHELA	indeed; truly				0
Adverb root KHONA	of place; here; there				0
Adverb root LAPHA	Here				0
Adverb root LAPHAYA	yonder				0
Adverb root LAPHO	there				0
		SO	Adverb	like this	0
		KAE	Adverb	where	0
		GABOTSE	Adverb	Well	0
		KUDU	Adverb	very much	0
		JWANG	Adverb	How	0

Totals: Number of isiZulu items: 8 Number of Sepedi items: 6 Overlapping items: 2 Non-overlapping items: 12

Appendices

N. 3 Conjunctions

isiZulu		Sepedi			Units that overlapped in meaning
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
Conjunction ANGITHI	isn't it				0
Conjunction KANTI	just so; in fact; (intj) what of it; what if so; who can stop me	GORE	Conjunction	so that	2
Conjunction KODWA	but	MARA	Conjunction	but	2
Conjunction NGOBA	because				0
Conjunction UMA	if; when	GE	Conjunction	when/while	2
		LE[H]	Conjunction	with/and	0
		WHY[CS]	Conjunction	why	0
		KGANTHE	Conjunction	whereas/while	0
		EBILE	Conjunction	then	0

Totals: Number of isiZulu items: 5 Number of Sepedi items: 7 Overlapping items: 6 Non-overlapping items: 6

Appendices

N. 4 Nouns/noun roots

isiZulu		Sepedi			Commonality scores
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
Noun root AMANGA	lie; untruth	MAAKA	Noun	lies	2
Noun root CHIPS	chips				0
Noun root CLASS	class				0
Noun root CRAYON	crayon				0
Noun root IFONI	phone				0
Noun root IGAMA	name; song; fame	LEINA	Noun	name	2
Noun root IKHAYA	home	GAE	Noun	home	2
Noun root IKHEKHE	cake				0
Noun root IMALI	money				0
Noun root IMOTO	motor-car	KOLOI	Noun	car	2
Noun root INDABA	affair; topic for discussion; story				0
Noun root INDLU	house; hut; room; dwelling place				0
Noun root INGANE	child	NGWANA	Noun	child	2
Noun root INGWENYA	crocodile				0
Noun root INTO	thing; object	SELO	Noun	thing/object	3
		NTO	Noun	thing	
Noun root IPHEPHA	paper				0
Noun root ISANDLA	hand; hand-writing; assistant	LETSOGO	Noun	hand	2
Noun root ISIKHWAMA	small bag; pocket; purse; fund				0
Noun root ISO	eye				0
Noun root ITAFULA	table				0
Noun root TOILET	toilet				0
Noun root UBANI	who?	MANG	Noun	who	2

Appendices

isiZulu		Sepedi			Commonality scores
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
Noun root UGOGO	grand-mother				0
Noun root UKUDLA	food; eating	SEJO/DI	Noun	food	2
Noun root UMAMA	my / our mother				0
Noun root UMISI	lady teacher	TN	Noun	teacher's name	2
Noun root UMLUNGU	white person; European				0
Noun root UMSINDO	noise				0
Noun root UMSIZI	helper; pencil				0
Noun root UMUNTU	human being; African; one with human feelings; blunt instrument (as knife)	MOTHO	Noun	person/human being	2
Child name		CN	Noun	child's name	2
Place name		PN	Noun	place name	2
		LELEKERE	Noun	sweet	0
		SEKHIPHA	Noun	shirt	0
		TOILET[CS]	Noun	toilet	0
		BOLO	Noun	ball	0
		MMATA	Noun	friend	0
		JERSEY[CS]	Noun	jersey	0
		MPYA	Noun	dog	0
		HLOGO	Noun	head	0
		FASE	Noun	down	0
		MAMAGO	Noun	your mother	0
		TŠHILA	Noun	dirt	0
		BOKAKA	Noun	faeces	0
		GARE	Noun	middle/centre	0
		MOSADI	Noun	woman	0
		PARTY	Noun	party	0
		LERAGO	Noun	buttock	0

Appendices

isiZulu		Sepedi			Commonality scores
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
		LEOTO	Noun	foot	0
		ENG	Noun	what	0
		NGWANENYANA	Noun	little girl	0
		KGALE	Noun	long ago	0
		SEKOLO	Noun	school	0
		MFANA	Noun	boy	0
		PELE	Noun	first	0
		SESI	Noun	sister	0
		MAMAKA	Noun	my mother	0
		MAMA	Noun	mother	0
		MONNA	Noun	man	0
		SEETA	Noun	shoe	0
		MORAGO	Noun	behind	0
		MOŠIMANE	Noun	boy	0
		MEETSE	Noun	water	0
		SETULO	Noun	chair	0
		MAABANE	Noun	yesterday	0
		BOGOBE	Noun	porridge	0
		MOKA	Noun	the whole	0

Totals: Number of isiZulu items: 32 Number of Sepedi items: 49 Overlapping items: 27 Non-overlapping items: 54

N. 5 Pronouns/pronoun roots

isiZulu		Sepedi			Commonality scores
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
Pronoun class 1	he; him; himself; she; her; herself	YENA	Pronoun	he/she, he/him	2
Pronoun class 10	it; itself				
Pronoun class 15	it; itself				
Pronoun class 3	it; itself	YONA	Pronoun	she/he/it	8
Pronoun class 5	it; itself				
Pronoun class 6	it; itself				
Pronoun class 7	it; itself				
Pronoun class 9	it; itself				
Pronoun class 2	them; themselves	TŠONA	Pronoun	them/they	
Pronoun first person plural	we; us	RENA	Pronoun	us/we	2
Pronoun first person singular	I, me, myself	NNA	Pronoun	I	2
Pronoun second person plural	You	LENA	Pronoun	you (plural)	2
Pronoun second person singular	You	WENA	Pronoun	You	2
		YELA	Pronoun	that one	0
		GONA	Pronoun	there	0
		WOLA	Pronoun	that one	0
		GEŠO	Pronoun	of our community/ household	0
		LE[K]	Pronoun	this one	0
		ŠE	Pronoun	here he/she is	0
		GENO	Pronoun	of your community /household	0
		GABO	Pronoun	of his/her community/ household	0

Appendices

isiZulu		Sepedi			Commonality scores
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
		ŠULE	Pronoun	there he is	0
		KA[G]	Pronoun	my/of mine	0
		YE[L]	Pronoun	this	0
		WO	Pronoun	This one	0
		SE[N]	Pronoun	This one	0
		GAGO	Pronoun	yours	0
		TŠE	Pronoun	these ones	0
		GAGWE	Pronoun	hers/his	0
		MOLA	pronoun	there	0

Totals: Number of isiZulu items: 13 Number of Sepedi items: 24 Overlapping items: 18 Non-overlapping items: 17

Appendices

N. 6 Verbs/verbal roots

isiZulu		Sepedi			Commonality scores
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
Verb root AZI	Know	TSEBA	Verb	know	2
Verb root BAMBA	catch; grasp; hold; overtake; take on (as seedling); delay; surprise one doing mischief	SWARA	Verb	hold	4
Verb root THATHA	take; marry	TŠEA	Verb	take	
Verb root BEKA	put; place; select pup				0
Verb root BHALA	write; write an examination	NGWALA	Verb	write	2
Verb root BHEKA	look; observe; go towards				
Verb root BONA	see; understand; give regards	BONA	Verb	see	3
Verb root BIZA	call; be expensive; price; be luring	BITŠA	Verb	call	2
Verb root BOLEKA	borrow; lend	ADIMA	Verb	borrow	2
Verb root BUKA	look at; watch; admire				0
Verb root BUYA	return; go back; turn inwards				0
Verb root CEBA	report someone; invent; plot against; be rich	TLALEYA	Verb	tell on	2
Verb root CELA	request; negotiate for a wife; be almost	KGOPELA	Verb	ask/request	2
Verb root CHAMA	Urinate	ROTA	Verb	to urinate	2
Verb root CRAYONA	'the act of crayoning' i.e., colouring				0
Verb root CULA	Sing				0
Verb root DLALA	play; dance; frolic	DLALA	Verb	play	
		BAPALA	Verb	play	4
		RALOKA	Verb	play	

Appendices

isiZulu		Sepedi			Commonality scores
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
Verb root ENZA	do; make	DIRA	Verb	do	2
Verb root FAKA	put in; put on; put around; commencement of udder to fill with milk	BEA	Verb	put	2
Verb root FANA	be like; resemble	SWANA	Verb	be the same	2
Verb root FIKA	come; arrive; reach				0
Verb root FUNA	search; want; desire	NYAKA	Verb	want/search	2
Verb root GIJIMA	run; flow				0
Verb root GQOKA	wear; be clothed	APARA	Verb	put on/wear	2
Verb root HAMBA	walk; go; travel	TŠWA	Verb	go out	4
		TLOGA	Verb	leave	
		TSAMAYA	Verb	walk	
Verb root HLALA	sit; stay; remain	DULA	Verb	sit	2
Verb root IDLA	eat; confiscate; eat into; cost	JA	Verb	eat	2
Verb root IMA	stand; stop				0
Verb root IPHA	give	FA	Verb	give	3
Verb root NIKEZA	give; tell off; pass on; hand over				
Verb root ISHO	say; mean	RE[P]	Verb	say	2
Verb root ITHI	say; intend; think	GOPOLA	Verb	remember/think	2
Verb root IWA	Fall	WE	Verb	fall	2
Verb root KADE	of action just completed				0
Verb root KHIPHA	take out; pull out	NTŠHA	Verb	take out	2
Verb root KHULUMA	talk; speak	BOLELA	Verb	speak/talk	2
Verb root LETHA	Bring	TLIŠA	Verb	bring	2
Verb root NGEKE	Never				0
Verb root NGENA	come in				0
Verb root PHATHA	hold; handle; control; administer				0

Appendices

isiZulu		Sepedi			Commonality scores
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
Verb root PHILA	live; be in good health				0
Verb root PHUMA	come out; go out; lose colour				0
Verb root QALA	begin; commence; annoy	THOMA	Verb	start/to begin	2
Verb root QEDA	finish; complete	FETŠA	Verb	finish	2
		FELA	Verb	finish	2
Verb root SABA	fear; be afraid				0
Verb root SHAYA	strike; punish; play (as an instrument)	BETHA	Verb	hit/beat	2
Verb root SHESHA	make haste; be quick; glide along				0
Verb root SUKA	go off; commence; originate	YA[A]	Verb	go	2
Verb root THANDA	desire; like; love; be inclined; wind; plait	RATA	Verb	love/like	2
Verb root THENGA	Buy	REKELA	Verb	buy for	2
Verb root THOLA	find; get; adopt				0
Verb root TSHELA	Tell	BOTŠA	Verb	tell	2
Verb root UKUBA	to be; if; in order that; because				0
Verb root UKUYA	go to	ILE	Verb	went/to go	2
Verb root VALA	shut; close				0
Verb root VELE	do merely; do originally				0
Verb root VULA	open; commence	BULA	Verb	open	2
Verb root WASHA	do laundry work				0
Verb root WOZA	Come	TLA	verb	come	2
Verb root YEKA	leave off; stop; let go	TLOGELA	Verb	leave behind	2
Verb root YEKELA	leave off; let alone				0
Verb root ZWANI	hear; listen; taste; smell; feel; sense; live; be alive	KWA	Verb	hear	2

Appendices

isiZulu		Sepedi			Commonality scores
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
		NKGA	Verb	smell	0
		NA	Verb	have/had	0
		KGONA	Verb	can/able to	0
		TSENA	Verb	go into/enter	0
		BE	Verb	must be/must become	0
		TSENELELA	Verb	penetrate/go ahead of	0
		RAGA	Verb	kick	0
		KUKA	Verb	carry	0
		GAFA	Verb	go mad/go crazy	0
		APEYA	Verb	to cook	0
		FOŠA	Verb	throw	0
		GOBATŠA	Verb	hurt	0
		FORA	Verb	lie to	0
		NGWATHA	Verb	break off a piece	0
		FIHLA	Verb	arrive	0
		ROGA	Verb	swear	0
		RA	Verb	mean	0
		NAMELA	Verb	climb	0
		EMA	Verb	stand up	0
		TŠHABA	Verb	run away/escape	0
		KAKA	Verb	to defecate	0
		LEBELELA	Verb	look for/at	0
		LLA	Verb	cry/complain	0
		FIŠA	Verb	be hot	0
		SEPELA	Verb	walk towards	0
		BA[2]	Verb	become/be	0
		BONTŠHA	Verb	show	0

Appendices

isiZulu		Sepedi			Commonality scores
Core formative	**English Translation**	**Core word**	**Part of speech**	**English Translation**	
		ROBALA	Verb	sleep	0
		IŠA	Verb	take to	0
		HLAPA	Verb	to wash/bath	0
		KGETHA	Verb	choose	0
		TLAPA	Verb	hit on the face	0
		SEGA	Verb	laugh	0
		TŠHELA	Verb	pour	0
		APEŠA	Verb	help	0
		BOELA	Verb	return to	0
		FETA	Verb	pass	0
		GANA	Verb	refuse	0
		GATA	Verb	step on	0
		HLOBOLA	Verb	take off	0

Totals: Number of isiZulu items: 62 Number of Sepedi items: 83 Overlapping items: 84 Non-overlapping items: 62

APPENDIX O

COMPARISON BETWEEN THE TOP 100 SEPEDI CORE VOCABULARY WORDS AND TOP 100 WORDS FROM THREE ENGLISH CORE VOCABULARY LISTS

Top 100 Sepedi core words	English translation (alphabetised)	Equivalent from top 100 words in English lists			Number of English lists in which equivalent word(s) was/were found
		Trembath Balandin and Togher (2007)	Boenisch and Soto (2015) Native	Boenisch and Soto (2015) ESL	
GAPE	Again				0
AH	Ah	Ah			1
AKERE	Isn't it				0
NGWE	Another				0
KGOPELA	Ask for/request				0
GA[5]	At			At	1
GAFA	Be mad/be crazy				0
THOMA	Begin/start				0
MARA	But	But	But	But	3
KGONA	Can/be able to	Can	Can	Can	3
KA[F]	Can/could	Can	Can	Can	3
NGWANA	Child				0
CN	Child's name				0
TLA[15]	Come	Come	Come	Come	3
DIRA	Do/make	Do	Do	Do	3
GA[3]	Do(es) not	Do not	Do not	Do not	3
JA	Eat			Eat	0
EH	Eh				0
TSENA	Enter/go into				0

Appendices

Top 100 Sepedi core words	English translation (alphabetised)	Equivalent from top 100 words in English lists			Number of English lists in which equivalent word(s) was/were found
		Trembath Balandin and Togher (2007)	Boenisch and Soto (2015) Native	Boenisch and Soto (2015) ESL	
MMATA	Friend				0
FA	Give			Give	1
YA[A]	Go	Go	Go	Go	3
TSAMAYA	Go	Go	Go	Go	3
NA	Had/have	Have	Have	Have	3
YA[B]	He/she/it/of	He/it/of	He/she/it/of	He/she/it/of	3
SA[10]	He/she/it/of	He/it/of	He/she/it/of	He/she/it/of	3
E	He/she/it/they	He/it/they	He/she/it/they	He/she/it/they	3
KWA	Hear/feel				0
YENA	Her/him/she/he	He/him	He/she/him	He/she/her	3
MO[8]	Here	Here	Here	Here	3
GAGWE	Hers/his	His			1
MO[7]	Him/her	Him	Him	Her	3
BETHA	Hit/beat				0
SWARA	Hold				0
KE[D]	I	I	I	I	3
NNA	I/myself/me	I	I	I	3
KE[C]	Is/are	Is			1
BO	It	It	It	It	3
TSEBA	Know	Know	Know	Know	3
MAAKA	Lies				0
SO	Like this	Like this	Like this	Like this	3
NGWANENYANA	Little girl				0

Appendices

Top 100 Sepedi core words	English translation (alphabetised)	Equivalent from top 100 words in English lists			Number of English lists in which equivalent word(s) was/were found
		Trembath Balandin and Togher (2007)	Boenisch and Soto (2015) Native	Boenisch and Soto (2015) ESL	
KGALE	Long ago				0
KA[G]	Mine	Mine		Mine	2
BE	Must be/must become				0
LEINA	Name				0
AOWA	No	No	No	No	3
HA-EH	No	No	No	No	3
WA	Of/you	Of	Of	Of	3
KUA	Over there	Over there			1
MOTHO	Person/human being				0
DLALA	Play	Play	Play	Play	3
BEA	Put	Put	Put	Put	3
RE[P]	Say		Say	Say	2
NYAKA	Search/look for/want	Want	Want	Want	3
BONA	See	See	See	See	3
TLA[15]	Shall/will		Shall/will	Shall/will	2
TLO	Shall/will		Shall/will	Will	2
DULA	Sit down/live/stay				0
GORE	So that		So that	So that	2
BOLELA	Speak/talk/tell	Tell	Tell/talk	Tell	3
SA[11]	Still				0
TŠEA	Take		Take	Take	2
TN	Teacher's name				0
TLALEYA	Tell on	Tell on	Tell on	Tell on	3

Appendices

Top 100 Sepedi core words	English translation (alphabetised)	Equivalent from top 100 words in English lists			Number of English lists in which equivalent word(s) was/were found
		Trembath Balandin and Togher (2007)	Boenisch and Soto (2015) Native	Boenisch and Soto (2015) ESL	
BOTŠA	Tell/inform	Tell	Tell	Tell	3
YELA	That one	That one	That one	That one	3
EBILE	Then		Then	Then	2
GONA	There	There	There	There	3
TŠE	These ones				0
TŠA	They/of	They/of	They/of	They/of	3
DI	They/them	They/them	They/them	They	3
BA[Z]	They/them/of	They	They	They	3
YE	This one	This one	This one	This one	3
SE[N]	This	This	This	This	3
SELO	Thing				0
WO	This	This	This	This	3
RE[Q]	Us/we	We	We/us	We/us	3
RENA	We/ours	We	We	We	3
ENG	What	What	What	What	3
HEH	What	What	What	What	3
GE	When/while		When		1
KAE	Where	Where	Where	Where	3
MANG	Who		Who		1
KA[E]	With/about/through	With	With	With	3
LE[H]	With/and	With/and	With/and	With/and	3

Appendices

Top 100 Sepedi core words	English translation (alphabetised)	Equivalent from top 100 words in English lists			Number of English lists in which equivalent word(s) was/were found
		Trembath Balandin and Togher (2007)	Boenisch and Soto (2015) Native	Boenisch and Soto (2015) ESL	
SE[O]	Won't/will not		Will not	Will not	2
NGWALA	Write				0
EE	Yes	Yes	Yes	Yes	3
WENA	You	You	You	You	3
LE[J]	You	You	You	You	3
LA	You(pl)/of	You/of	You	You/of	3
LENA	You(plural)	You	You	You	3
O	You/her/him/it	You/him/it	You/him/it	You/her/it	3
GAGO	Your(s)	Your	Your	Your	3
GO[T]	No translation				0
A[U]	He/she/them/of	He/she/it/of	He/she/it/of	He/she/it/of	3
A[W]	No translation				0
A[X]	Your(s)	Your	Your	Your	3
NTO	Thing/something		Something	Something	2

APPENDIX P

DECLARATION ON PROFESSIONAL LANGUAGE EDITING

M.B. BRADLEY

P.O. Box 37326
Faerie Glen
Pretoria 0043

072 369 5149

DECLARATION ON EDITING

Student: Ms N.R.B. Mothapo	**Date:** 2019/03/25

Document submitted for editing

Dissertation: Determining the core vocabulary used by Sepedi-speaking preschool children during regular preschool-based activities

The above dissertation (except list of references) was submitted to me for language editing, which was completed on 25 March 2019.

M.B. BRADLEY (MA) - Language editor

Appendices