UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Determining the core vocabulary of Setswana-speaking Grade R learners as used during school activities

by

Morwesi Gaopalelwe Mogatusi

A mini-dissertation submitted in partial fulfilment of the requirements for the degree

Master in Augmentative and Alternative Communication

in the Centre for Augmentative and Alternative Communication

UNIVERSITY OF PRETORIA

FACULTY OF HUMANITIES

SUPERVISOR: Professor Kerstin Tönsing

CO-SUPERVISOR: Dr Refilwe Morwane

November 2022

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITY OF PRETORIA

DECLARATION OF ORIGINALITY

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

Full names of student: Morwesi Gaopalelwe Mogatusi

Student number: [redacted]

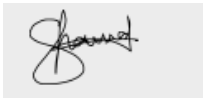
Declaration

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

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

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

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

SIGNATURE OF STUDENT: 

SIGNATURE OF SUPERVISOR: 

ETHICS STATEMENT

The author, whose name appears on the title page of this dissertation, has obtained, for the research described in this work, the applicable research ethics approval.

The author declares that he/she has observed the ethical standards required in terms of the University of Pretoria's Code of ethics for researchers and the Policy guidelines for responsible research.

ACKNOWLEDGEMENTS

To myself,
Gaopi, you are already someone. “You are already someone. You are already something. You are who you are, you are who you are meant to be. You are exactly where you need to be. You are meant to question and wrestle with yourself and the world until you find your own answer. You are answerable to your source and your soul. You will keep running your own race, a good race and you are already a winner in your race.”
Excerpt from *These Things Really Do Happen To Me*
by Khaya Dlanga

The Lord is good and merciful! I bow before the Lord as I complete this chapter for He has carried my work to completion. May I never forget how good God has been to me, moreover the strength and divinity He has bestowed upon me. I owe it all to Him.

To my supervisor, Prof Kerstin Tönsing. This endeavour would have not been possible if it wasn’t for your mentorship. What a privilege it has been to have worked with such an insightful woman. Your never-ending patience and support throughout this process have been truly invaluable. I have learned, transformed and became a whole new person because of your influence. I stand on the shoulders of giants and you are one of them. May your light continue to shine for others. A big thank you to also my co-supervisor, Dr Refilwe Morwane.

To my parents, Edwin and Catherine Mogatusi. Words can never express my gratitude to both of you. The support received from you during this period has been astounding. Mom, thank you for always supporting me in chasing my dreams and helping me achieve them in the best way possible. Thank you Dad for forever believing in my abilities and never failing to remind me of that.

To my brother, Katlego Mogatusi, boy you have been a true right-hand man and have placed your hearts best interests in the success of my studies. Thank you for all the assistance you provided when technicalities failed me. What you have done for me can never be undone in my heart. I hope this serves as an exemplary record that anyone can achieve anything they put their mind to. I wish the same trajectory and even better for you.

To my very supportive partner, your constant belief in me has kept my spirit alive. Thank you for your unwavering support. May we continue soaring.

I would like to also extend my words of gratitude to my Moses Kotane Hospital colleagues for their never-ending support and words of encouragement during the rainy days. I know that I am loved and supported because of you ladies.

To my research assistants (Didintle & Pini), thank you so much for your support. I truly appreciate all the contributions towards the creation of this mini-dissertation.

Finally, a special thanks to my study participants, the caregivers of children involved, school principals and teachers, and the Director of the North West Department of Education for their cooperation and willingness in allowing me to conduct my research study. It meant a lot.

ABSTRACT

Background: When designing AAC systems for children who have not yet developed conventional literacy skills, a process of vocabulary selection is employed. Core vocabulary is a list of words used frequently by particular age group in spontaneous conversation. Core vocabulary lists are one of the useful lists consulted when selecting vocabulary for AAC systems. South Africa is a multilingual country having 11 official languages and only three studies have explored the core vocabularies of the following South African languages: Afrikaans, isiZulu and Sepedi. Due to the uniqueness of languages, the core vocabulary list of a language cannot simply be translated to another language for AAC use because linguistic meaning and will be lost. This study was therefore conducted to determine the core vocabulary of Setswana speaking Grade R learners, which then can be added to the pool of multilingual vocabularies available for use by individuals in need of AAC.

Methods: Language samples of six preschool children recruited from three different schools were collected during regular school activities through recordings. Covid-19 regulations were adhered to at all times during collection of data. Body worn audio recording devices were used to record the spontaneous speech of preschool children and was then transcribed, coded and analyzed. The data was analyzed using Microsoft Excel™.

Results: From a composite script of 18,099 intelligible words, a total of 1,112 different words were identified. The type token ratio of this sample was 0.06. An analysis was conducted on these number of different words used, how frequently used the words were and how commonly used the words were among the participants. A total of 249 core words and 863 fringe words were identified.

Conclusions: The findings of this study were comparable and consistent with those found in other core vocabulary studies of other languages, in that the Setswana core vocabulary consisted of a smaller proportion of words which were used more frequently and represented a large portion of the composite sample. The Setswana core vocabulary list used in this study can be consulted and used as a resource during vocabulary selection of designing an AAC system for children with Setswana language backgrounds.

Keywords: Augmentative and Alternative Communication; Preschool Children; Setswana; Core Vocabulary; Vocabulary Selection

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. PROBLEM STATEMENT, DEFINITIONS AND ABBREVIATIONS

1.1 Problem statement

Aided augmentative and alternative communication (AAC) is typically chosen as the main form of communication for individuals with severe physical disabilities that hinder the use of speech and manual signs (von Tezchner & Martisen, 2000). Early AAC implementation for children with complex communication needs (CCN) is essential to support the development of children's language and communication skills (Romski et al., 2010).

Often, such AAC systems contain words that are preselected and these words are often represented by symbols that a child can use to express themselves (Mngomezulu et al., 2019). These preselected words should be meaningful and functional, effective in establishing social interaction and also be age appropriate as well as applicable to the background, gender and environment of the child who uses AAC (Trembath et al., 2007). Selecting vocabulary for a child who will use AAC typically requires the practitioner to choose a small set of words from hundreds of thousands of possibilities and may thus be an intricate and lengthy process which requires a significant amount of time (Trembath et al., 2007). While the vocabulary of speaking children grows to be increasingly diverse, vocabulary in an AAC system should be limited to limit working memory and navigation demands on the user (Beukelman & Mirenda, 2013). At the same time, the vocabulary should not unnecessarily compromise the user's expressive abilities and opportunities. There are limited guidelines documented in the literature for the vocabulary selection process (Fallon et al., 2001).

Using core vocabulary is one technique that can be employed when selecting vocabulary for AAC (Boenisch & Soto, 2015). Core vocabulary refers to words and/or messages that occur very frequently in conversation and are used commonly by a variety of individuals (Beukelman & Light, 2020). These words make up majority (up to 80%) of an individual's spoken language (Bean et al., 2019). Including core vocabulary words in an AAC system can provide individuals with CCN the expressive power to produce many different phrases and sentences (Mothapo et al., 2021).

With 11 official languages, South Africa carries great linguistic diversity with it (Dada et al., 2017). In the design and customisation of aided AAC, it is necessary to continuously consider the cultural and linguistic diversity to broaden the availability of appropriate linguistically and culturally relevant AAC systems for people with CCN (Soto &

Yu, 2014). Since core vocabulary consists of many structured words, that is, words that relate to the grammar and syntax of a language, it would be amiss to translate the core vocabulary found in one language into another language mainly because languages differ in their grammatical structure. Among the 11 official languages of South Africa, only three (3) studies thus far have explored and determined core vocabulary in three South African languages, that is, Afrikaans (Hattingh & Tönsing, 2020), Sepedi (Mothapo et al., 2021) and isiZulu (Mngomezulu et al., 2019). No studies have addressed the core vocabulary of the other eight official South African languages. Amongst these neglected languages is Setswana. No study has established a core vocabulary for Setswana-speaking children. Moreover, the number of children who may require AAC is growing each year.

According to the 2011 Census, there were 14 418 114 children aged between five and 19 in South Africa. As 8% of the population was Setswana-speaking at the time, this meant that about 1.15 million children would have been estimated to be Setswana-speaking (Statistics South Africa, 2011). Given the population growth of about 18% since the 2011 statistics, this would lead us to 1.36 million children who are Setswana-speaking now. With an estimated 1.3% global incidence of CCN (Dada et al., 2017), children needing AAC services who are also from a Setswana language background are estimated to be about 18 000 in number. These children currently have limited access to AAC services in their home language. A Setswana core vocabulary list could be a useful resource to equip interventionists with a word list for vocabulary selection of AAC for children with CCN from Setswana language backgrounds. In addition, a comparison of such a list to the Sepedi list (Mothapo et al., 2019) could highlight some interesting similarities and differences, because Setswana and Sepedi both belong to the Sotho group of African languages. This means the languages share significant similarities in grammatical structure, and similarities in the core word lists are expected.

1.2 Definition of terms

This dissertation makes frequent use of the following terms and are therefore defined here for clarity.

1.2.1 Augmentative and alternative communication (AAC)

AAC describes a field of clinical and educational practice, including research practices, that supplement and aim to compensate for temporary or permanent disabilities

which limit activity and/or participation in comprehending or communicating through speech, language or writting (ASHA, 2022; Beukelman & Light, 2020).

1.2.2 Code-switching

The process of using two languages or more in a single conversation (Mothapo et al., 2019). This occurs when one inserts words from another language into a sentence or phrase (Zirker, 2007).

1.2.3 Commonality score

A score assigned to a particular word to indicate the total number of participants who used it. For example, in this study where six participants took part, if a particular word occurred in all the participant language samples, the commonality score would then be 6. Therefore, any word with a commonality score of 6 meant that it was used by all participants at least once.

1.2.4 Content words

Content words are words that are used in a language and carry meaning with them. These are words such as nouns, verbs, adverbs and adjectives (Trembath et al., 2007). These words can appear in isolation for labelling purposes. However, these words are typically inadequate in conveying more complex messages on their own, they require the use of structured words to form syntactical constructions (Sutton, Soto et al., 2002).

1.2.5 Core vocabulary

Core vocabulary has been defined by the field of AAC as words that appear most frequent and are commonly used in natural conversations (Witkowski & Baker, 2012). These words are constant across different settings (Beukelman & Mirenda, 2013; Boenisch & Soto, 2015). In this study, core vocabulary is determined by the frequency of word occurrence and commonality across speakers. Words that occur with a minimum frequency of 0.5% and a commonality score of at least 3 (i.e., at least three speakers used the word) in the composite sample are considered core words.

1.2.6 Fringe vocabulary

Fringe vocabulary are words and messages that are unique to a particular individual and context, they often occur with a lower frequency and a lesser commonality in speech

samples (Banajee et al., 2003; Beukelman & Mirenda, 2013; Trembath et al., 2007). In this study, all words that occurred with a frequency of less than 0.5 per 1 000 words (0.5‰) and/or had a commonality score of less than 3 were classified as fringe words.

1.2.7 Grade R

This grade is the first class in the South African foundation phase of basic education. It is also referred to as 'reception'. This is the class children enrol in before they commence their first year of formal school. There is a national curriculum specified for South African schools in the Curriculum and Assessment Policy Statement (CAPS, 2012) which guides the teaching of this grade as well its learning outcomes. In this mini-dissertation, Grade R learners will be referred to as pre-schoolers.

1.2.8 **Grammatical variation/inflexion form**

Words typically contain root word parts and additional morphemes are often added to indicate various aspects such as, for example, tense. In Setswana, *tla* is a root word which can be written as *tlaya* in the present tense meaning 'come' and can be written as *tlile* in the past tense meaning 'came'. The addition of morphemes to the root part is what makes it inflected (Quirk et al., 1985). The additional morphemes in root words can be replaced to form different other words with the same root. The inflected form in a word does not change its part of speech classification (Payne, 1997).

1.2.9 Graphic symbols

Graphic symbols are static in nature and are often represented in the form of line drawing pictures and/or picture representations (Smith, 2006). These are various commercially and freely obtainable collections of symbols which are used to encode messages in AAC (Beukelman & Light, 2020). Most graphic symbol collections may have similarities with other symbolic systems, however, most of these symbols lack other aspects of language, such as arbitrariness of letters and the duality of patterning (Smith, 2006).

1.2.10 Heteronyms

Words that are written or spelled the same but are different in pronounciation and in meaning (mean different things). Setswana has heteronyms such as the word *tlhaga* which can be used as a noun and a verb. *Tlhaga* can refer to 'grass' (noun) or can refer to 'appear' (verb).

1.2.11 Lemma

The lemma is the form of the word that is typicvally found as the headword in a dictionary. For example, the lemma of the English words *sit, sits, sat*, and *sitting* is *sit*, because this form would be found as the head word in the dictionary. The lemma is typically the uninflected form of the word.

1.2.12 Parts of speech

A set of word groups that are used to classify words according to their syntactical functioning or purpose in the language (Croft, 2000). Mojapelo (2007) refers to this process of classification as grammatical classification. Verbs, nouns, concords and conjunctions are a few examples of these word groups. Another term used to refer to these word groups is 'word classes'.

1.2.13 Root word

The central morphemes of a word that carries lexical meaning are called root words. Any word contains a root which has only one morpheme and cannot be further divided into other smaller meaningful units (Howard, 2003).

1.2.14 Structure words

Structure words are the inverse of content words. These words have less meaning in isolation but have a high grammatical function because they assist content words (nouns, verbs, pronouns, etc.). The grammatical correctness of sentences can be attributed to the use of structure (Banajee et al., 2003). Pronouns, prepositions, conjunctions and auxiliary verbs (amongst others) are classified as structure words. In Setswana, since the language is written disjunctively, structure words include morphemes to indicate concords, participle tenses etc.

1.2.15 Type token ratio

The type token ratio is an indicator of lexical diversity, this ratio has been used extensively in child language research to show the variation of words in a language sample (Richards, 1987). The ratio is calculated by dividing the number of different words (NDW) by the total number of words (TNW) (Kettunen, 2014).

1.3 List of abbreviations

AAC: Augmentative and Alternative Communication

ASHA: American Speech-Language-Hearing Association

CAPS: Curriculum and Assessment Policy Statement

CCN: Complex Communication Needs

CIA: Central Intelligence Agency

CS: Code Switch

DoE: Department of Education

LoLT: Language of Learning and Teaching

MS: Microsoft

NDW: Number of Different Words

POPIA: Protection Of Personal Information Act

SLT: Speech-Language Therapist

TNW: Total Number of Words

TNDW: Total Number of Different Words

TTR: Type Token Ratio

2. LITERATURE REVIEW

To frame this study, some background will be provided on AAC systems, particularly graphic symbol-based systems. Vocabulary selection methods will be reviewed. Core vocabulary lists from other studies and their significance will be discussed. The need for core vocabulary studies in South Africa is looked into thereafter. Lastly, the Setswana language will be briefly discussed.

2.1 Augmentative and alternative communication (AAC)

One of the significant developmental achievements of young children is the ability to use speech and language to communicate. Within the first years of life, children transition from using only vocalisations (sounds) and body movements to ultimately communicating through speaking in complex sentences (producing words and sentences). Children progress through these stages of language development as they learn to express themselves and convey their emotions, wishes, observations and experiences. As we grow, the ability to produce speech then becomes largely automatic (Beukelman & Mirenda, 2013). For some individuals, however, effortless oral speech communication is not an option due to various disabilities, and, as a result, they struggle to meet their communication needs (Beukelman & Ray, 2010). These individuals often require AAC to supplement their limited expressive speech (American Speech-Language-Hearing Association (ASHA), 2022). AAC is the cornerstone of communication habilitation as well as rehabilitation for this group; its effects are documented in the increasing research which reports positive outcomes such as communication enhancement, language development, increased participation, supported comprehension and a decline in frustration and problem behaviours (Beukelman & Light, 2020).

AAC comprises unaided and aided modes. *Unaided* modes involve non-spoken methods of communication that are produced using the body only, such as gestural cues, manual signs or facial expressions (ASHA, 2022). *Aided* modes involve external support such as communication boards with graphic symbols or objects, printed words, and traditional orthography as well as computers, tablets, and other handheld mobile devices with software or applications that have speech-generation capabilities (ASHA, 2022).

When a child is unable to speak or sign manually due to severe physical or motor impairment, they often require other means of expression. Due to this child's limitations and

inability to use speech or manual signs, aided forms of language are then considered to enable this individual to communicate (Von Tetzchner & Stadskleiv, 2016). Graphic symbols are often what is used to represent concepts or referents, that is, language. By selecting certain graphic symbols, the child can then communicate a message. The graphic symbols act as a type of language representation (Von Tetzchner & Stadskleiv, 2016). Since graphic symbols exist outside of the communicator and are selected instead of produced like in natural speech, these symbols (and the vocabulary they represent) need to be preselected to be made available to the child.

2.2 Vocabulary selection

Once the need for aided AAC has been recognised, the typical course of action is to select, design or customise a graphic symbol-based AAC system for the child (Von Tetzchner & Stadskleiv, 2016). This typically includes selecting and/or customising the vocabulary that is to be included in the system. For individuals who are still in the early stages of language development and are not yet able to spell words, vocabulary selection can be categorised into two categories, namely, vocabulary that is needed to communicate essential messages and vocabulary that is needed to develop language skills (Beukelman & Light, 2020). The two vocabulary categories are coverage vocabulary and developmental vocabulary, respectively. When selecting initial vocabulary for a child with CCN, Musselwhite and St. Louis (1998) recommend that the initial vocabulary selected should be vocabulary that is highly needed or desired by the individual, have the ability to be regularly used, be able to be used in multiple semantics and pragmatic ways and be able to be used currently but not limited to future use. Similarly, Soto and Cooper (2021) concurred that the first words that are taught to a child who has difficulty communicating should be i) age-appropriate and well aligned with their development (i.e., selected among first words used by children of the same age with typical development); ii) selected to support the development of grammar through the inclusion of different word classes; and iii) be applicable to the child's communicative needs in different settings e.g., outside the classroom – these are still the followed guiding principles that inform AAC practice for over 40 years (Soto & Cooper, 2021; Beukelman & Light, 2020; Holland, 1975; Lahey & Bloom, 1977).

Different types of sources and different team members can be consulted in the quest of selecting the best vocabulary for the individual in need (Crestani et al., 2010). Collaboration on vocabulary selection is an important part of shared decision-making in the

development of AAC for the person with CCN. Researchers suggest that when selecting vocabulary, collaboration with family and professionals should be conducted to customise the vocabulary and ensure that it is personally meaningful (Soto & Cooper, 2021). To select and customise vocabulary according to a child's needs, AAC practitioners need to consult a combination of sources to identify potential vocabulary and then discuss this with the child's family and other professionals, thus ensuring a thorough consultation process (Soto & Cooper, 2021).

As mentioned previously, various methods and sources can be used to facilitate vocabulary selection (Bean et al., 2019; Beukelman & Light, 2020). These include environmental or ecological inventories, informant lists, communication diaries and the use of existing vocabulary lists (published resources) to guide selection (Beukelman & Light, 2020). Each of these methods and sources has its advantages and disadvantages. Furthermore, each of these methods can work differently for the individual with CCN and should be carefully considered when selecting vocabulary. Some of the factors to consider during this selection process includes the ease of use by the system recommended, the setting where the system will be used and the child's cognitive abilities (Beukelman & Light, 2020).

Environmental or ecological inventories have an advantage in that they can be used to document how the individual with CCN participates in and interacts in various activities or environments, and also which words and messages are communicated by peers with and without disabilities in those situations. Such inventories result in word lists that are environment-specific and functional in high-priority situations. The disadvantage, however, is that there is still a degree of inference in assuming that the individual with CCN will want to or need to communicate about the same topics as peers. Additionally, the words obtained in this manner may be very context-bound, whereas much of communication for more advanced communicators is decontextualised (e.g., talking about a past event).

Utilising informant lists for vocabulary selection also has its advantages. These lists are often generated by individuals that spend the most time with the person with CCN, and the most common informants are caregivers, peers, siblings, friends, teachers, and other caregivers (Beukelman & Light, 2020). Informants may, for example, keep a communication diary and record words that they feel are needed in different situations throughout the day. These informants can offer valuable vocabulary suggestions that can be used in the process of

vocabulary selection. Results from previous studies suggest that each informant related to the child with CCN has the potential to contribute an important number of fringe words to the vocabulary of children (Beukelman & Light, 2020). The disadvantage of this method is the fact that the selection of vocabulary may be biased and influenced by the informant's expectation of which words the child should - as a result, some words may be less useful to the child. Also, informants tend to think of content words that have semantic meaning. Structured words that are important for building sentences are often omitted, and informant-generated vocabularies tend to have more nouns and verbs (Banajee et al., 2003).

2.3 Core vocabulary lists

Core vocabulary lists are another known resource to SLPs which can be utilised as a guiding tool for vocabulary selection. Core vocabulary refers to high-frequency words that make up about 80% of an individual's spoken language and are frequently used across different activities and environments by different individuals of the same age group (Soto & Cooper, 2021). For core vocabulary lists to be established, researchers often record and collect samples of spoken language from a group of typically developing children at a preschool or daycare during routine activities such as playtime, snack time or others (Laubscher & Light, 2020). The most frequent and common words are then identified from these speech samples. Trembath et al. (2007) indicated that for words to be considered 'core', they must have a frequency occurrence of at least 0.5 in 1 000 and must be used by at least 50% of the participants.

The advantage of core vocabulary is that by virtue of its small size, this vocabulary can be exhaustively displayed on a communication device (Banajee et al., 2003; Soto & Cooper, 2021). Utilizing core vocabulary in AAC allows individuals to use novel utterances that can serve various syntactic, semantic and pragmatic functions (Mngomezulu et al., 2019; Soto & Cooper, 2021). It has been suggested that including core vocabulary words in an AAC system can grant individuals with CCN access to the production speech through various different phrases and sentences, therefore improving communication expression (Mothapo et al., 2021). Core vocabulary consists of words from different word classes. This includes many structured words and closed class words as these are words that appear more frequently in the spontaneous speech of peers of the same age without disability (Witkowski & Baker, 2012; Yorkston et al., 1998). These structured words allow for the building up of novel sentences because they include parts of speech such as prepositions, conjunctions, and other

parts of speech (Mngomezulu et al., 2019). Structured words obtained from a core list provide a framework for functional language use by the individual with CCN by contributing to the grammatical correctness of sentences and thus syntax of a language (Hatting et al., 2020, Banajee et al.,2003). So, although structured words have little semantic meaning they perform an important grammatical function (Hatting et al., 2020). To communicate, a combination of content and structured words is needed.

Core vocabulary lists have been determined for several different languages, based on spoken language samples from various populations collected in various contexts.Van Tilborg and Deckers (2016) reviewed the different language samples of various core vocabulary studies from different populations, specifically, those with typical and atypical development. This review revealed that core vocabularies found in various language samples were comparable for: typical and atypical populations, individuals with primary language impairments, second language learners, monolinguals, bilinguals, AAC users, and individuals with physical and/or intellectual disabilities (Van Tilborg & Deckers, 2016). Specifically, the authors noted that core vocabulary of the most studied groups contained a high number of structure words - especially for the first 50 or 100 words. The number of content words (words with semantic meaning, e.g., nouns) increased after 100+ words (Mothapo, 2019).

This study made a comparison of the proportion of content and structure words found in the top 100, top 200, and total core word lists that were established for three South African core vocabulary studies: the Afrikaans core vocabulary list (Hattingh et al., 2020), Sepedi core vocabulary (Mothapo et al., 2021) and Zulu core vocabulary study (Mngomezulu, et al., 2019) See Table 1. This comparison was done to determine whether Van Tilborg and Deckers's (2016) postulation on the structure and content word proportion found in core vocabulary is also true for South African languages. Table 1 indicates the percentages accounted for by structured versus content words in core word lists.

Table 1

Comparison of the Proportions of Content versus Structure Words in the Top 100, 200 and Complete Core Word Lists as Determined in Three South African Core Vocabulary Studies for isiZulu, Afrikaans and Sepedi

	isiZulu		Afrikaans		Sepedi	
	Structure	Content	Structure	Content	Structure	Content
Top 100 words	60.8%	35.6%	44%	56%	56%	44%
Top 200 words	29%	68%	34%	66%	40%	60%
Above 200 words	32%	56,4%	32%	68%	38%	62%

In all three core vocabulary studies (Hattingh, et al., 2020; Mngomezulu et al., 2019; Mothapo et al., 2021), it was noted that the proportion of structure words was highest among the top 100 words. However, structure words proportionally decrease when the top 200 and all core words are taken into consideration (Mothapo, 2019). Therefore, the proportion of content words (nouns, verbs, adjectives and adverbs) increases beyond the top 100 (Mngomezulu, 2019). This is true for all three lists compared. A similar pattern was observed in an English core vocabulary study by Boenish and Soto (2015).

Although there are clear similarities in the characteristics of core vocabularies determined for different languages, the actual core word lists cannot be assumed to be identical or translation equivalents of each other across languages. Words are used to encode the grammar, morphology and semantic meanings of a language. As the grammatical and morphological structure of languages differs, and because the meaning boundaries of words that represent concepts differ between languages, words that most frequently occur in one language cannot be assumed to be the most frequently used words of another language (Soto & Cooper, 2021).

To confirm that languages cannot have the same core vocabulary lists, Mothapo (2019) undertook a comparison between two South African core lists: Sepedi- (Mothapo, 2019) and isiZulu (Mngomezulu, 2017), which indicated an overall overlap of 41.9%. This percentage value undoubtedly reflects that there can never be complete equivalence between core lists in different languages, meaning that most words found in an isiZulu list will more

than likely not appear or overlap with the Sepedi list. Researchers need to bear in mind that although there are similarities in the trends of words occurring in core lists across languages, there are also significant differences in that core lists can never reflect the exact core word list found in a different language.

Though the benefits of core vocabulary have been well documented, Laubscher and Light (2020) provided some critiques on the appropriateness of utilizing the core word approach during selection or teaching of expressive language for beginning communicators (children who have started producing their first words but have not yet began to put those into two or more word utterances) after comparing five English core vocabulary lists whih were established based on the speech samples of two- to six-year-old children. The core vocabulary lists were developed by Banajee et al., (2003), Beukelman et al., (1989), Fried-Oken et al., (1992), Marvin et al., (1994) and Trembath et al., (2007).

It was discovered that 80% of the words found in the English version of MBCDI (MacArthur Bates Communicative Development Inventories- a standardized assessment tool that assesses the early vocabulary skills for typically and atypically developing children) were not reflected in all the five core vocabulary lists (Laubscher & Light, 2020). It is for this reason that Soto and Cooper (2021) advocate that practitioners should combine various methods when selecting vocabulary for beginning communicators. These methods may include using communication diaries, categorical inventories, environmental inventories, and/or published and validated wordlists produced by children aged similarly to the age of those in the MBCDI. They confirmed that although published word lists seem to offer a solution towards selecting vocabulary, the solution is not complete and requires additional guides to ensure that the relevance of the vocabulary selected and customisation to the needs of each child and their language developmental stage (Trembath et al., 2007).

Finally, the vocabulary selected should undergo an ongoing process of vocabulary maintenance. If words that were previously chosen for the AAC system are now used less frequently due to reduced usefulness or have expired their time, these should be eliminated to make space for newly selected words (Beukelman & Light, 2020).

2.4 The need for core vocabulary studies in South African languages

South Africa is a country with a population of approximately 60 million people whom represent a diverse range of cultural, linguistic, and religious practices; as well as different ethnicities and nationalities spread across the nine South African provinces (Moonsamy et al., 2017). Among its 11 official languages, it is reported that the most commonly spoken languages are isiZulu (22.7%), isiXhosa (16%), Afrikaans (13.5%), English (9.6%), Setswana (8%), and Sesotho (7.6%) (Statistics South Africa, 2011). These languages are mostly spoken by the population residing in urban and peri-urban areas of Johannesburg, Cape Town, Durban, Pretoria and some rural areas in other parts of the country. Other languages are spoken by a minority in several different rural areas (Statistics South Africa, 2011).

Speech-language therapists are clinical practitioners who work in the prevention, as well as the assessment, diagnosing and treatment of speech, language and swallowing disorders of children and adults (ASHA, 2022; Pillay et al. 2020). AAC is one of the branches in the field and falls within the scope of practice of SLTs. It is often imperative to understand the history of this field and how it has contributed to the lack of SLT resources and the current disparities in the field.

The development of the speech-language pathology field in South Africa needs to be understood in the context of its pre- and post-democratic history. The SLT profession was previously embedded in the service delivery models of the Western and/or colonisation (Moonsamy et al., 2017). During the apartheid era, ethical standards of practice were a priority and were always strived for, however, services were only available and accessible to the minority white population which would now be categorised as unethical according to the South African constitution. The voices and needs of the non-white people, that is, majority of South African population, who were in need of SLT services appeared silent and were not catered for (Moonsamy et al., 2017). As a result of these apartheid policies, SLT services were predominantly accessed and available to the white population only thus marginalising accessibility to rehabilitation services for special needs children from non-white backgrounds. In addition, access to quality basic education was also limited to the white population and thus opportunities for tertiary qualifications or academic degrees were restricted for the majority of non-white persons, therefore limiting non-white people who could qualify in the SLT profession.

After this, the field of speech-language pathology now comprises a limited and insufficient number of speech-language pathologists (SLTs) to provide services to the population at large. Furthermore, the number of qualified SLTs do not represent and parallel the linguistic and cultural needs of the country's population. The distribution of SLTs across the private and public sectors is unequal, therefore, causing a great disparity in the supply and need for culturally and linguistically appropriate resources or services in the SLT field (Pascoe & Norman, 2011; Pillay et al., 2020). AAC in a diverse country like South Africa needs to cater to the language needs of all individuals with CCN. Most citizens in the country speak an African language as a home language, whereas English is often an additional language. Therefore, the need for the provision and development of linguistically appropriate and culturally relevant tools for individuals with CCN has been a repeated call in the field. Without linguistically and culturally appropriate resources, the provision of AAC services in South Africa will continue to remain limited (Dada et al., 2017; Maguvhe, 2014). Moonsamy et al. (2017) also added that South African SLPs are responsible for cultivating clinical practices that are relevant to the needs of the South African population, furthermore should ensure that the resources used in practice are customised to the language, culture, choice of content and familiarity. Service providers need to develop and practice cultural competence in a country of diversity and difference.

It is thus established that the ability to access appropriate and relevant forms of AAC plays a significant role in reducing communication challenges of individuals with CCN and promoting participation and inclusivity in society (McNaughton & Babb, 2021). The lack of appropriate AAC systems in African languages is a recognised barrier (Dada et al., 2017). Designing and customising such systems requires knowledge of the grammar and the vocabulary of the intended language (Soto & Yu, 2014). A start has been made by establishing core vocabulary lists for three South African languages, namely Afrikaans (Hattingh & Tönsing, 2020), Sepedi (Mothapo et al., 2021) and isiZulu (Mngomezulu et al., 2019). Since core vocabulary cannot be translated from one language to another, the need for a Setswana core vocabulary list is observed and this is what this study aims to address.

2.5 The Setswana language

Setswana is a spoken language in Southern Africa and belongs to the Sotho group, alongside Sepedi (Northern Sotho) and Sesotho (Southern Sotho) (Mahura & Pascoe, 2016; Cole, 1955). Setswana is also a cross-border language as it is also spoken in other countries such as Botswana, Namibia and Zimbabwe (Mahura & Pascoe, 2016; Cole, 1955). Out of an estimated 60 million citizens in South Africa (Worldometer 2020), Setswana is spoken by approximately 8% of the population which is more than 4 million people (Statistics South Africa, 2011). Considering the population growth of about 18% and a 1.3% incident rate of CCN, an estimated 18 000 children with limited speech in South Africa could benefit from a Setswana AAC system. This number increases when children in Botswana and Zimbabwe are also considered.

In South Africa, Setswana is the fifth-largest language group (Statistics South Africa, 2011). The Setswana language is divided into four subgroups of different dialects: Central Setswana (Sehurutshe & Sengwaketse), Southern Setswana (Setlhaping, Setlhware, Serolong), Northern Setswana (Sekwena, Sengwato, Setswana) and Eastern Setswana (Transvaal Sekgatla, West-Transvaal Sekwena) (Mokgoko, 2019). These dialects were birthed as a result of migration to different areas in search of employment where Setswana-speaking people interacted with other language speakers thus influencing the use of standard Setswana (Mokgoko, 2019). Although these are the existing different dialects, they do not deviate substantially from the standard language. Standard Setswana is reportedly based on the Sehurutshe dialect and is nearly identical to it (University of Wisconsin, 2022). This dialect also provides the basis for formal written language. The speech samples collected for the current study were obtained in the Rustenburg region of the North-West Province. This population predominantly speaks the Transvaal Sekgatla and West-Transvaal Sekwena dialects. Although regional specificity may be present in the sample, it is expected to only minimally influence the resulting core word list.

Regarding its morphological typology and orthographic conventions, Setswana is a predominantly agglutinating language, which is written disjunctively (Malema et al., 2020). This means that the many morphemes that exist in the language are usually sequenced without the morphemes changing form or pronunciation and that many morphemes are written as one orthographic word, separated from others by the orthographic space. For example, morphemes such as *fa* and *ka* are separate orthographic words, forming a short

phrase such as *ka fa* ('in here'). As Mothapo (2019) discussed in the case of Sepedi, the implications for core vocabulary are that the orthographic space can be productively used to separate the units that are counted when establishing word frequency counts that underlie the establishment of a core vocabulary list.

Morphological variations of nouns (plurals, locatives, diminutives), for example, *batho (motho), ntlung (ntlu), nnyenyane (nnye)* and verbs (moods and tenses), for example, *batle (batla), bone (bona)* as well as adjectives (agreement with noun), such as *dintsinyana (ntsi), montle (ntle)* exist in the Setswana language similar to the Sepedi language. These words were grouped under the root word in the Sepedi study (Mothapo et al., 2021) and this study will undertake the same technique. The Setswana language has seven noun classes (*Mo, Mo, Le, Se, N, Lo, Bo, Go*) wherein the noun consists of a pre-morpheme (class prefix) and a root (Harman, 1980). Other noun words may consist of a premorpheme, a root and a postmorpheme e.g., *motsomi* → *mo* (premorpheme) *tsom* (root) i (post-morpheme). The pre-morpheme which can be singular or plural gives the noun a characteristic which distinguishes it from other parts of speech and makes the noun congruous with the rest of the sentence (Harman, 1980). The Setswana nouns are thus divided into various classes according to the various pre-morphemes (Harman, 1980).

2.6 Summary

This literature review established the merit of core vocabulary lists as one source of vocabulary that can be included in graphic symbol-based AAC systems for preliterate children. The need to establish language-specific core word lists was highlighted, and the lack of such lists for South African languages was discussed. The structure of the Setswana language and some implications for a core vocabulary study were also discussed.

3. METHODOLOGY

This part of the mini-dissertation will focus on the employed research methodology for this study. The main aim of the study is highlighted, followed by the sub-aims as well as the research design employed. A summary of the phases conducted is illustrated. The setting of the study, the description of participants and the materials and equipment used are also elaborated upon. The pilot study is then described, including the aims, followed by the procedures, results and recommendations. The main study is then described, including the procedures for data collection, and data analysis, and factors influencing validity and reliability are thereafter reported. Finally, ethical considerations and principles upheld in the study are discussed.

3.1. Aims

3.1.1 Main aim

The study's main aim was to determine and describe the words that are most commonly and most frequently used by Setswana speaking Grade R learners without disabilities during regular school-based activities.

3.1.2 Sub-aims

The sub-aims of the study were:

(i) To establish the words, the total number of different words (TNDW), and the frequency of use of each word used by Setswana-speaking Grade R learners without disabilities during regular school activities from transcribed language samples;

(ii) To differentiate core and fringe vocabulary based on frequency and commonality criteria and to describe NDW and coverage of both;

(iii) To further describe core words by commonality scores, differentiating content versus structure words and classifying them into parts of speech; and

(iv) To compare the coverage of parts of speech found in the Setswana core vocabulary to the core vocabulary established in two other linguistically related languages (Sepedi and isiZulu).

3.2 Research design and stages

A descriptive observational study design was adopted for this study (McMillan & Schumacher, 2010). This design is best suited to answer the research question as it asks about a naturally occurring phenomenon and allows the researcher to study the participants in their natural environment without manipulating any variables, thus increasing the external validity of the study. Six (6) Grade R learners aged 5;0 to 6;11 who speak Setswana as a home language were recorded using body-worn recorders and small microphones during their regular school activities. Due to the time-consuming nature of observations and transcriptions, only a small group of participants were involved. This may reduce the representativeness of results (Cresswell, 2009). The external validity of the results is highly dependent on whether the population studied is representative of the population proposed to be studied (Aggarwal & Ranganathan, 2019).

3.2.1 Stages of the study

The stages undertaken by the researcher to conduct the study are shown in Figure 1. They include (1) Material development, (2) Participant recruitment, (3) Pilot study, and (4) Data collection and analysis.

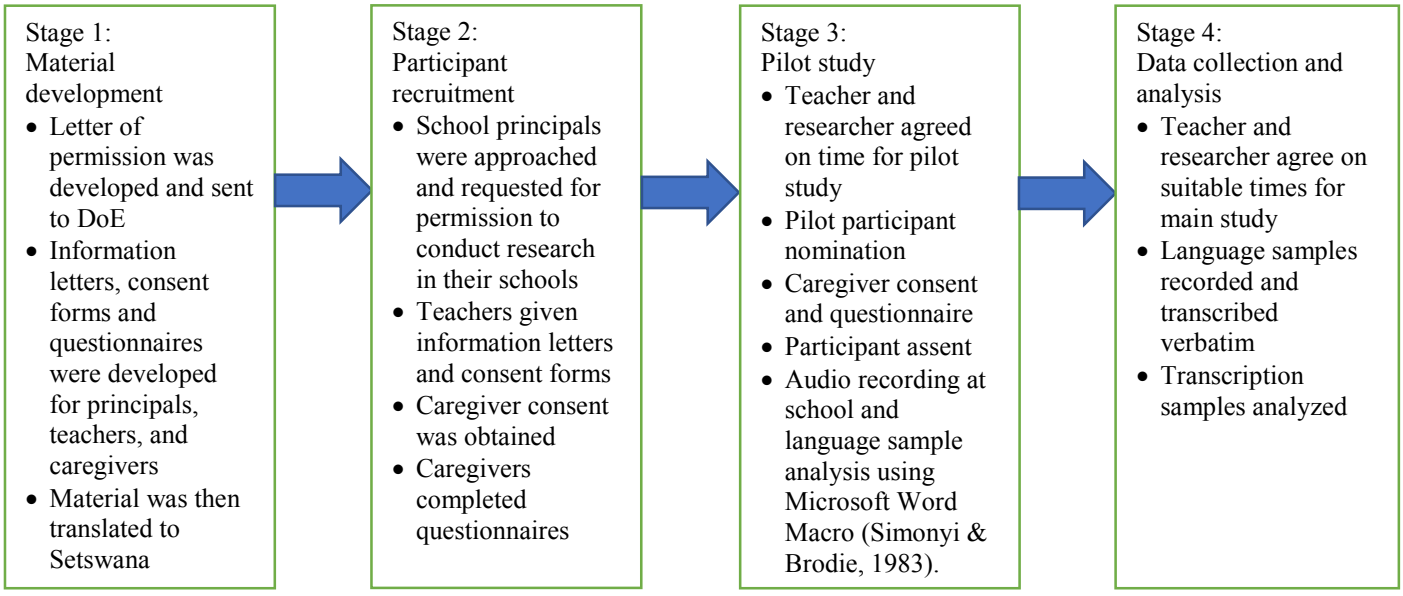


Figure 1. Stages of the study

3.3 Setting

The chosen schools wherein the participants were obtained were all situated in a semi-rural area towards the North of Rustenburg City Centre, in the Mogwase and Ledig location. Both locations are situated in the Moses Kotane Local Municipality under the Bojanala district region. The schools are scattered in the area and are situated about 25 to 30 km from one another. The three selected preschools used Setswana as the primary language of learning and teaching (LoLT). The preschools had access to water and sanitation as well as electricity. However, most preschools did not have access to the internet but used a telephone for communication with other stakeholders. All three preschools followed the CAPS curriculum (CAPS, 2012) in their teaching activities.

The map provided in Figure 2 shows the Moses Kotane region in the Northwest Province.

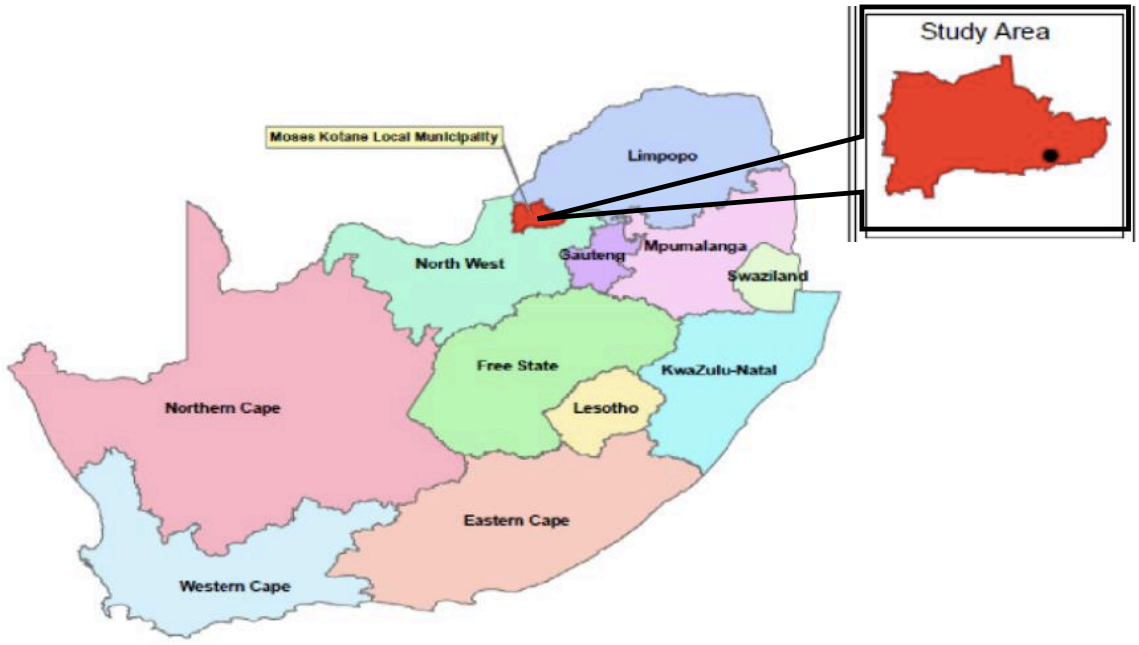


Figure 2. Area of Moses Kotane Municipality where the study was conducted (Source: Mosime, 2014) (https://repository.nwu.ac.za/bitstream/handle/10394/15444/Mosime_DK.pdf?sequence=1&isAllowed=y)

3.4 Participants

3.4.1 Participant recruitment and sampling

Research approval was obtained from the Research Ethics Committee of the Faculty of Humanities of the University of Pretoria (Appendix A). The researcher then obtained approval for participant recruitment in the schools from the North West Department of Education (DoE) (Appendix B). The convenience sampling method was utilised to select three public schools with Grade R in Mogwase and Ledig, Rustenburg, near the researcher's residence. Convenience sampling is a non-probability sampling method that involves the sample being drawn from a population which is close to the researcher (McMillan & Schumacher, 2010). This method of sampling was used due to the language used in the area of the researcher as well as the researcher's interests in the Setswana language. Setswana was the language of instruction in all the Grade R classrooms at all three schools.

The principals of the schools were provided with information letters detailing all information regarding the study (Appendix C) and permission forms where they had the opportunity to grant or decline permission to conduct research in the Grade R classrooms of their schools. All the principals that were approached granted permission for the study. Thereafter, Grade R classroom teachers were approached and provided with information letters and consent forms (Appendix D). They were requested to grant or decline written consent to conduct the study in their classrooms.

Once teachers provided consent, purposive sampling was then used to select talkative learners from the class to be included in the study. Teachers were asked to nominate a girl and boy from each of their classes whom they perceived as talkative. The teachers were thereafter required to provide the nominated learners' caregivers with information letters (see Appendix E) that describe all aspects of the study in English and Setswana. Caregivers needed to first grant or decline consent on the form (Appendix E). Those that granted consent were further asked to complete a questionnaire (Appendix G).

Participant assent was sought before the study and visual aids were used to enhance comprehension. The researcher arranged to meet with each potential participant and explained the study to him/her in Setswana using child-friendly language according to a

script (Appendix F). The script was followed to keep instructions standard, succinct and easily understood. Potential participants were shown pictures to improve understanding. Potential participants were then allowed to give or decline assent using verbal communication as well as marking their answers on a picture-based assent form using a marker pen or crayon.

3.4.2 Participant selection criteria

The participant selection criteria are presented in Table 2.

Table 2

Participant Selection Criteria

Criterion	Exclusion	Justification	Measure used
Participants should be between the ages of 5;0 and 6;11.	Learners younger and older than the prescribed age	Learners of this age have relatively mature speech and language skills (Owen & Leonard, 2002)	Caregiver Questionnaire (Appendix G)
Participants should have no speech and language impairments or developmental concerns	Learners with speech and language impairments	Impairments and other developmental concerns may have effects on speech and language (Wallace et al., 2015)	Caregiver Questionnaire (Appendix G)
Learners should have been enrolled for at least 1 month at the school	Learners enrolled for less than 30 days in the school	The participant should be comfortable and familiar with the environment to reduce novelty effects that may affect talkativeness (Trembath et al., 2007)	Caregiver questionnaire (Appendix G) Teacher nomination
The child should have Setswana as their home language and as the language of learning and teaching (LoLT)	Learners who only speak Setswana at school but not in the home context	Chances of obtaining speech samples that include code-switching and mixing should be minimised (Bosma & Blom, 2019)	Preschool Questionnaire (Appendix H) Caregiver Questionnaire (Appendix G)

3.4.3 Descriptive criteria

The following table will describe the participants of this study according to their descriptive data to detail all factors related to them, their school and home environment as well as the languages they were exposed to.

Table 3

Description of Participants

Participant number	**School/ Site number**	**Age**	**Gender**	**Number of months enrolled in preschool**	**Frequency of school attendance**	**Home language**	**Other Languages exposed to at home**	**Languages exposed to via media (TV/Radio)**	**Monthly income of household**
P1	1	5;1	F	2 months	Daily	Setswana	English, SeSotho	English, Setswana	<R7,275
P2	1	5;1	M	3 months	Daily	Setswana	English, Shona	English	<R7,275
P3	2	5;2	M	3 months	Daily	Setswana	isiZulu, English	English, Setswana	<R7,275
P4	2	5;6	F	3 months	Daily	Setswana	isiXhosa	English	<R7,275
P5	3	6;2	M	3 months	Daily	Setswana	English	English	<R7,275
P6	3	6;1	F	3 months	Daily	Setswana	English, SeSotho, isiZulu	English	<R7,275

3.5 Materials and equipment

The equipment and material used in the study were carefully chosen for the suitability of the design employed and ethical considerations. All Covid-19 regulations were upheld during data collection.

3.5.1 Equipment

Speech samples were collected through digital voice recorders (Olympus, Model DM 650) and (Philips, Model DVT 6010) placed inside padded pouches worn around the waist. A small lapel microphone (Audio Technica Lavalier Microphone, ATR 3350) was used to obtain a speech sample from each of the participants. The microphone was attached to the top part of the participant's shirt or collar using a clip.

The audio files retrieved from the recorders were saved onto a laptop (MacBook Air 14"). Earphones were used to play back the audio recordings during the transcription phase. Transcriptions were done in Microsoft Word™ documents. A Microsoft macro function found in Microsoft Word™ was used to run frequency counts of words obtained.

Figure 3. Digital voice recorders. Olympus, model DM 650 and Philips, model DVT 6010

Figure 4. Small lapel microphone (Audio Technica Lavalier ATR 3350)

Figure 5. Participants wearing the recording equipment

3.5.2 Materials

3.5.2.1 Information letters and permission/ consent forms

Information letters and concomitant permission/consent forms were drafted for the various persons that needed to be informed and provide permission or consent before the commencement of the study. These were: the school principals (Appendix C), the teachers (Appendix D) and the caregivers (Appendix E). All letters (except the DoE permission letter and teacher letter) were translated to Setswana by the researcher and then checked and corrected by a linguist, and the co-supervisor who are both fluent in Setswana to ensure accurate language translation.

Both English and Setswana information letters and forms were given to principals and caregivers to allow them to choose their preferred language. Teachers were also given their information letter and consent form in English. Most individuals in South Africa conduct their higher education in English, therefore literacy skills in English are often better than literacy skills in their home language.

Each letter set out the purpose of the study, the intended procedures (including Covid-19 safety protocols), the intended use of the data, the risk and benefits, and the ethical principles that would be adhered to. The expectations of the party addressed should they provide permission/consent were also clearly stipulated.

3.5.2.2 Assent script and form

A child-friendly script was drafted by the researcher in Setswana explaining all aspects of the study (Appendix F). She included a picture illustrating each aspect described including Covid-19 protocols to be adhered to during the study. The researcher compiled an assent form with pictures to allow the learners to give or decline assent (see Appendix F). The assent script and form are given in both English and Setswana for the sake of the reader's convenience, although only the Setswana version was used with the learners. A similar procedure of translation only to English (not back translation) was followed.

3.5.2.3 Caregiver questionnaire

The caregiver questionnaire was devised to gather information relating to specific selection criteria, such as the child's age, developmental milestones, the length and frequency of the child's attendance at the school as well as the use of Setswana as the main language in the home. Furthermore, information about the child's language exposure via family members and the media was also included, as were questions about the household income, that is, above or below the taxable income. Information about the availability of water, electricity, and toilet facilities at home was also gathered. This information helped attain a rich picture of the range of participants (Shenton, 2004). The researcher based the questionnaire on those previously constructed by Mothapo (2019) and Hatting (2019) for similar studies. The questionnaires mentioned were also drafted in English and translated into Setswana. These were all checked by two fluent Setswana-speaking linguists. They were sent to the caregivers in both English and Setswana (see Appendix G).

3.5.2.4 Preschool questionnaires

A preschool questionnaire (Appendix H) was devised to gather information and establish whether the school's language of learning and teaching (LoLT) met the selection criteria of the study i.e Setswana. Aditionally, the questionnaire gathered descriptive information about the exposure and use of other languages (e.g., in child-to-child interactions), the number of learners in the preschool classes as well as the daily programme and the curriculum used. The available facilities/services at each school were also noted for descriptive purposes. The questionnaire was based on Mothapo (2019) who conducted a similar study. The researcher drafted the questionnaire in English and then translated it into

Setswana. A fluent Setswana speaker then checked the translated version. The last item on the questionnaire listed the participant selection criteria and then requested that a boy and a girl is nominated from a class by their teacher for inclusion in the study.

3.5.2.5 Transcription and coding rules

A set of transcription rules were developed to guide the transcription process of the voice recordings. The transcription rules were developed based on Du Bois (1991) and Trembath (2007). The rules ensured that transcriptions are conducted consistently. The transcription rules used are shown in Appendix J.

Coding rules were developed. The accuracy and correctness of adding codes to the words transcribed was guided by the rules. The coding rules were created to accommodate the Microsoft Excel™ program such that inflected forms of words can be counted together with their root words. The coding rules used are shown in Appendix K.

3.6 Pilot study

A pilot study was conducted to ensure that the materials and procedures proposed are appropriate for the main study (McMillan & Schumacher, 2010). This assisted the researcher to test and review the recruitment strategy, selection criteria, procedures for data collection and the analysis where necessary.

The pilot study participant was a female of five years who attended Grade R at Site 1. The classroom she was selected from was made up of 43 learners. The school had access to the internet and a telephone and had adequate sanitation. As with most learners coming from the areas around, their household income was below the taxable income cut-off. The participant spoke Setswana as a home language and was exposed to English, isiXhosa and isiZulu languages at home. The recruitment, consent and assent processes were followed as described in Section 3.4.1. The pilot study participant met the selection criteria described in Table 2. Procedures outlined in Section 3.7 were followed. The equipment was fitted on the participant as per Section 3.5.1. The recording was taken over one day only. The researcher requested the teacher to check the child's comfort on the equipment over two-hourly intervals as well as to check the recorder for continuous functioning. The participant was closely

observed to ensure no interruption to her daily routine. The pilot study findings were used to give the researcher an idea of the effectiveness of the recording equipment.

Table 4 gives an overview of the aims of the pilot study, the materials and procedures used, the results and the subsequent recommendations.

The pilot study provided the researcher with useful information on the procedures of the study; this provided the researcher with better ideas for ensuring a thorough understanding of the procedures and cleared out all concerns raised by caregivers regarding the participants' safety and confidentiality. All recommended findings were implemented in the main study.

Table 4

Pilot Study Findings

Aim	Materials	Procedures	Results	Recommendations
To determine whether the selected method of recruiting participants through teacher assistance is successful	Teacher information letter and nomination	Information letters and details about selection criteria were delivered and discussed with the teacher, who then nominated a child to participate.	The teacher nominated an appropriate participant	None
To determine whether caregiver consent could effectively be obtained through the school.	Caregiver consent form and questionnaire (Appendix E and G, respectively)	The class teacher was requested to send a caregiver info sheet and consent letter to the participants' caregivers on behalf of the researcher.	The teacher self-handed the letters to the caregivers of the child due to fear of paper damage/loss. When received, the caregivers requested to meet the researcher and have an opportunity to ask questions. A meeting with caregivers was thus conducted.	Where an in-person meeting is required by the caregivers, the teacher will inform the researcher and the researcher will arrange to meet the caregivers to explain the study in person and any address questions.
To ensure that child assent procedures are effective to promote participants' comprehension of the study and gain informed assent.	Assent script (Appendix F)	The assent script was presented verbally to the child and supported with pictures. Their comprehension of procedures was estimated by their responses.	The participant showed understanding and gave assent by putting a circle on their response	None
To ensure that the proposed Covid-19 infection control measures to be adhered to during data collection were appropriate and effective.	Covid-19 protocol (Appendix F)	Covid-19 protocol outlined in the child assent script was highlighted to the participants. Teacher feedback was used to assess their usefulness or any associated observed risks.	The researcher sanitised equipment before and after use by the participant, social distance was also maintained when assent was received. The researcher also checked with the teacher if the equipment posed any health and safety risks. The teacher reported no risks and affirmed to make sure all learners wear masks at all times	Continued compliance with Covid-19 protocol. The researcher informed the teacher when recordings revealed no wearing of masks and teacher-reinforced adherence to safe covid protocols.

Aim	Materials	Procedures	Results	Recommendations
To ensure that the recording equipment is safe and poses no risk of injuries to participants.	Waist pouches, digital voice recorders and lapel microphones	Microphone lapels were attached to the child's collar and recording equipment was placed inside pouches and fastened around the child's waist. The teacher was asked to report on child comfort.	The participant was comfortable during all times and always reported when the pouch became loose to be tightened or when needing to go to the bathroom	None
To ensure that the chosen equipment was effective in recording intelligible samples that can be reliably transcribed	Lapel microphones (Audio Technica Lavalier Microphone, ATR3350), digital voice recorders (Olympus, Model DM 650) and earphones.	The researcher transferred audio files from the recorders to the laptop. These were listened to through earphones. All words uttered by the participant were transcribed. An independent person cross-checked transcriptions against the audio and transcription reliability were calculated. An agreement of at least 90% was considered reliable.	Typical earphones did not allow enough noise cancellation which affected the transcription reliability. All words were transcribed verbatim from audio to Word document.	Noise-cancelling headphones are to be used to listen to the recording when transcribing. This improved accuracy as some words were muffled by wearing masks. Cross-checkers listened to audio via electronic files.
To determine whether transcription rules were comprehensive to guide the transcription process	Transcription rules (Appendix J)	The researcher applied transcription rules during the transcription process to obtain good-quality transcription.	The transcription rules used were sufficient and helped with avoiding repetitions or song vocabulary affecting the core list.	None
To determine if the coding system is effective to generate word counts that are appropriate for the study	Coding rules (Appendix K) to ensure morphological variations of nouns and verbs are counted together, that code switches to other languages are identifiable, and that all heteronyms and polysemous words are counted separately.	The coding rules were applied after the transcription process to obtain a quality-coded transcription. Codes were refined in the process of application.	The coding rules used had symbols used to code parts of speech. Microsoft Macro counts symbols as separate words therefore symbols had to be replaced with numbers	More coding rules were needed. Coding rules were amended.

Aim	Materials	Procedures	Results	Recommendations
To determine whether the analysis process will yield the desired outcomes as outlined in the aims of the study.	Microsoft Word™ Macro (Simonyi & Brodie, 1983)	The researcher downloaded a macro system on MS Word™ then used the MS macro system to generate word frequency report	The Word Macro report provided only word occurrences. The researcher was able to get the total number of words on MS Word™ but had to calculate the number of different words.	The researcher can tabulate all word occurrences on MS Excel™ to determine the total number of words and calculate the total number of different words by eliminating heteronyms and adding these together.

3.7 Procedures

3.7.1 Data collection

Appropriate days and times for data collection were agreed upon with the teachers at each school. The researcher arrived at the respective schools every morning on the days agreed upon to fit the participants with recording equipment and remove them every afternoon. Verbal orientation to the study was provided to the teacher and participants before data collection and teachers also received an instruction sheet (Appendix I) to refer to during data collection. During meetings with the participants, the researcher and learners maintained a safe social distance of 1.5-2 metres from each other (except for times when equipment is fitted) and they wore face masks at all times.

Each participant was fitted with a small recording device in a padded pouch worn around the waist, with a small lapel microphone clipped to the collar of their shirt or T-shirt. Participants and other learners in class were requested not to fiddle or play with the recording equipment fitted. This is to reduce transmission of Covid-19 by touching the equipment as well as to eliminate the risk of switching off the recorders or damaging the equipment. All study procedures including the wearing of pouches, recorders and microphones were explained to the learners thoroughly. The researcher granted each participant an opportunity to give assent to take part or to decline taking part in the study. Participants were reminded that they are allowed to withdraw at any point of the study with no negative outcomes. The participants were shown how not to interfere with the microphones by touching or blocking them as this might jeopardise the audibility of the data collected. Participants were encouraged to report to their teacher if they have any difficulty with the equipment or would like it to be taken off. The participants were also asked to behave as they typically would on any other day. Teachers were requested to monitor the participants and ensure that the child is safe and comfortable at all times of the study. Each participant was recorded until 3 000 orthographic words had been reached.

According to the Standard Operating Procedure for the Prevention, Containment and Management of Covid-19 in schools and school communities (2020) by the South African Department of Basic Education, the following procedures were to always be observed in schools: (1) Every learner, staff member and visitor should wear a cloth mask at all times; (2) all persons should avoid contact with others through shaking hands or hugging; (3) everyone

should wash their hands frequently or use an alcohol-based hand sanitiser to practice uninterrupted hygiene; (4) everyone should practice social distancing of 1.5-2 metres. These procedures were all adhered to when the researcher met with participants. Teachers furthermore monitored that the wearing of recorders did not in any way jeopardise participants' adherence to these procedures.

In addition, recording equipment was not shared by the participants. Each learner was allocated one recording device which was used only by them throughout this study phase until the data collection process was complete. The recording equipment was disinfected daily and was stored in a sealable plastic bag with a colour and number code for each participant. This assisted with preventing cross-contamination from surfaces of equipment used by different participants.

3.7.2 Transcription, coding and data analysis

The audio files obtained from the participants were transferred from the recorders to a laptop daily. The researcher and research assistant listened to the audio recording and manually transcribed everything that the target participant said following the transcription rules (Appendix J). To check reliability, a second person cross-checked all transcriptions with the audio recording, similar to the procedures used by Romski et al. (2010) and Barton-Hulsey et al. (2017). An agreement of 90% or more was deemed acceptable (Ayres & Ledford, 2014). The following agreement formula was used:

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

Table 5

Percentage Agreement of Transcription per Participant Transcript

Particiapnts	1	2	3	4	5	6	Average
Percentage agreement (%)	98.2	98.5	92.3	93.2	96.3	98	96.5

The percentage agreement of transcriptions ranged from 92.3% to 98% per participant, with a mean or average of 96.5%. As an agreement of 90% and above was satisfactory, the total average represents a good transcription agreement of all transcripts and indicates reliability.

Coding was then added to the transcription according to the coding system developed (see Appendix K). This allowed the researcher to perform frequency counts in a manner that appropriately deals with code switches to other languages, morphological variations of nouns and verbs, as well as heteronyms and polysemous words. The reliability of the coding was also checked. A second independent coder coded 20% of all transcriptions. Codes were compared and percentage agreement was determined, where an agreement of 90% or more was deemed acceptable (Ayres & Ledford, 2014). The following formula was used:

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

Table 6

Percentage Agreement of Coding 20% of Each Transcript

Participants	1	2	3	4	5	6	Average
Percentage agreement (%)	98	97	97	97.4	97.7	98.5	97.6

The percentage agreement of coding ranged from 97% to 98.5% per participant, with an average or mean of 97.6%. Percentage agreement of 90% or more was regarded acceptable (Ayres & Ledford,2014), this thus represents a good agreement between the coders and further indicates that coding was conducted consistently and reliably.

The MS Word™ document generated during transcription has a function to report the TNW in the document which the researcher needed for analysis. Thereafter, a MS Word™ macro system (Simonyi & Brodie, 1983) was run on the coded composite transcript in an MS Word™ document to determine the frequency counts of all words. Once the MS macro system had run frequency counts, the TNDW of the composite sample was displayed in a pop-up report. The same method on the MS macro system was used to determine the TNDW for each transcript and was noted by the researcher for result reporting and tabulation

Results from the frequency counts were used to determine the total number of words or morphemes, the total number of different words (or in some cases lemmas) and the frequency of each word or lemma across all transcripts (Miller & Iglesias, 2012). The results from the frequency count of the composite script were then transferred to an MS Excel™

Sheet where the total number of words or morphemes was determined. The MS Excel™ Sheet displayed the frequency of each word or lemma separately.

Various columns of information were created in the MS Excel™ sheet to present information in a readable format. The commonality of each word was then checked manually by the researcher in each of the six scripts to determine how many participants used the word. Each word was then classified by parts of speech and the coverage of each word was calculated.

3.7.3 Reliability and validity

Reliability and validity are two important research concepts that define the quality of the study. The reliability of a study evaluates the consistency of what is being measured and the validity of a study measures the accuracy of how the study is conducted. These concepts were considered for this study to evaluate whether the methods employed increase or reduce the validity of the study, similarly, to also evaluate whether the results obtained would be replicable if obtained from the same context at a different time. The pilot study served to ensure that procedures for data collection can be reliably executed and that recordings of sufficient quality are obtained to ensure reliable transcription and, therefore, reliable frequency counts.

Participants may often react to the novelty of wearing recording equipment, which could influence their language samples because participants may talk more or talk less and/or talk about the equipment. Therefore, the first 20 minutes of all recordings were omitted from transcription and analysis. Also, all utterances where the participants talk about the recording equipment or recording process were omitted from the analysis. Moreover, all recordings took place during the entire school day and not only at a specific time of the day and thus these factors strengthened the study's internal validity.

Another measure used to strengthen the study's internal validity involved having the same researcher giving the same instructions, using the same tools, materials, and protocols and following the same procedures with all participants (McMillan & Schumacher, 2010). Transcription reliability was ensured by using the same transcription rules throughout transcription. Additionally, transcriptions were cross-checked with voice recordings to ensure

the accuracy of transcriptions, as described in Section 3.7.2 (Barton-Hulsey et al., 2017; Romski et al., 2010). The reliability of coding was assessed by measuring the extent to which two or more independent coders code data in the same way (Freelon, 2013). The reliability of the transcription was high, and the reliability of the coding was also similarly high.

The sample size used was limited, thus affecting the generalisability (external validity). However, external validity was improved by including learners from three schools, rather than from only one classroom in one school. Vocabulary frequency results that could be attributed to a specific focus or theme in one class were therefore somewhat ameliorated. Additionally, both boys and girls were included to avoid gender bias in the results.

3.8 Ethical issues

The ethical principles for researching as outlined in the Belmont Report (1978) were upheld in this study. The study offered no harm or risk to participants and thus adheres to the principle of non-maleficence. Participants were constantly reminded to always approach the teacher if the recording equipment causes discomfort or if they wish for it to be removed. Teachers were instructed to adjust or remove the equipment according to the participants' requests and also at the teacher's discretion thus minimising the risk of injury.

Although the study did not benefit participants directly, it did, however, offer beneficence to the AAC field through research and offered a resource for clinicians and practitioners who will need a Setswana vocabulary list when selecting vocabulary for AAC systems.

Justice was maintained throughout, that is, all participants were treated equally. Participants involved in the study were not significantly advantaged nor disadvantaged compared to those who were not selected to participate, therefore there was no risk of justice being violated.

Consent or permission was obtained from school principals, the teachers and most importantly the caregivers of participants as children are a vulnerable research group. Child assent was also sought to ensure that an informed decision was made on whether or not the child wanted to participate. Details of the study were explained to participants using child-friendly language and ensuring comprehension of the process before data collection. The

Setswana assent sheet supplemented by pictures was used to circumvent any comprehension or literacy issues by the learners. With the above-mentioned, autonomy was ensured for all participants. In addition, participants also had the right to withdraw at any given point of the study without any negative effects. The caregivers also had the right to choose whether to make the recordings available to other researchers through the SADiLAR (South African Digital Language Resource Centre) repository where only the written form of recordings will be made available for reuse or further analysis.

Confidentiality of participant details was and will continue to be upheld by not disclosing the participants' identities; this was done through the exclusion of mentioned names from the recordings when transcribing. Research assistants signed a confidentiality agreement form to adhere to and protect participants' privacy, thereby upholding the POPI Act of 2013 (POPIA, 2013). Names of teachers and classmates, as well as place names (proper nouns) mentioned in recordings, were substituted by a code. Only the researcher, research assistant and supervisor had access to the recordings. The data will be kept safely at the Centre for Augmentative and Alternative Communication (CAAC) at the University of Pretoria for 15 years.

3.9 Summary

This chapter aimed to describe the research methodology followed. It outlined the main aims and sub-aims and also described the research design followed. The research design section elaborates on the stages of the study as well as the phases it underwent. The following section described the pilot study where the aims, procedures and measures were explained which influence the results and the recommendations. This then introduced a section that describes the participants, their recruitment and the selection criteria carried out. Lastly, the chapter described the procedures of the study and ethical considerations of the study as well as the reliability and validity of the study.

4. RESULTS

4.1 Description of the sample

The transcribed language samples contained between 2 941 and 3 126 words (TNW) per participant. The TNW, NDW and TTR per participant (before the removal of unintelligible words) are given in Table 5.

Table 7

TNW, NDW and TTR per Participant

Participant	TNW	NDW	TTR
Participant 1	3126	449	0.14
Participant 2	3065	523	0.17
Participant 3	2941	634	0.22
Participant 4	3029	583	0.19
Participant 5	3037	583	0.19
Participant 6	3048	677	0.22

These samples were then combined to form a composite sample that contained 18,246 words in total. From this composite sample, all unintelligible words were removed, leaving the composite script with 18,099 intelligible words. The NDW of the composite script totalled 1,112. The type-token ratio of the composite sample was 0.06.

As prescribed by the coding system used in this study, the inflected forms of words (nouns, verbs or pronouns) were reduced to their root word (lemma) and counted together with their respective root words to calculate NDW. Thus, words with similar roots were counted together. Furthermore, words that had the same spelling but with different meanings (heteronyms) and words that have different meanings (although derived from the same origin, i.e., polysemous words) were counted separately.

Following the POPI Act (2013) and the study's aim to preserve the participants' privacy, learners's names mentioned by the participants were replaced with CN, teachers' names with TN and the names of places with PN. Each of these codes was thus counted separately when calculating NDW.

4.2 Core and fringe vocabulary

To establish the core vocabulary, words had to meet two criteria. The two criteria used in this study were also employed in previous studies (Hattingh et al., 2020; Mngomezulu et al., 2019; Mothapo et al., 2020). The first criterion required that the word had to occur in the sample with a frequency of 0.5 % per mille, that is, occurring once or more per 2,000 words.

The frequency of occurrence was calculated by dividing the total number of occurrences of each unique word in the sample by the TNW in the composite sample and multiplying by 1 000 (Mngomezulu, 2017; Mothapo, 2019). The formula used is given below:

$$\frac{Total\ number\ of\ occurrence}{Total\ number\ of\ words} X\ 1{,}000 = frequency\ per\ mille$$

The second criterion required that the word had to have a commonality score of at least 3, indicating that the word has been used by three out of the six participants (i.e., the word had to be used by at least 50% of the participants).

A Microsoft Excel™ spreadsheet was used to arrange all words according to frequency per mille. The commonality score of each word that had a frequency count of 0.5‰ or more was then determined. Using these two criteria, a core list of 249 Setswana words was established. The core word list with frequency and commonality scores for each word can be found in Appendix L.

The number of occurrences of all 249 core words was summed up and divided by the TNW (intelligible words) to obtain the coverage percentage which amounted to 86.27%. This thus indicates the coverage of the core vocabulary, indicating that 86.27% of the words used by the participants during conversations were core words. This also means that the core words covered >80% of the conversations.

The remaining 863 words were categorised as fringe words. Though fringe words have a high NDW, their coverage was only 13.73%. Figure 6 shows the number of different fringe words and the number of different core words in the composite sample, and Figure 7 shows the coverage of core words versus fringe words.

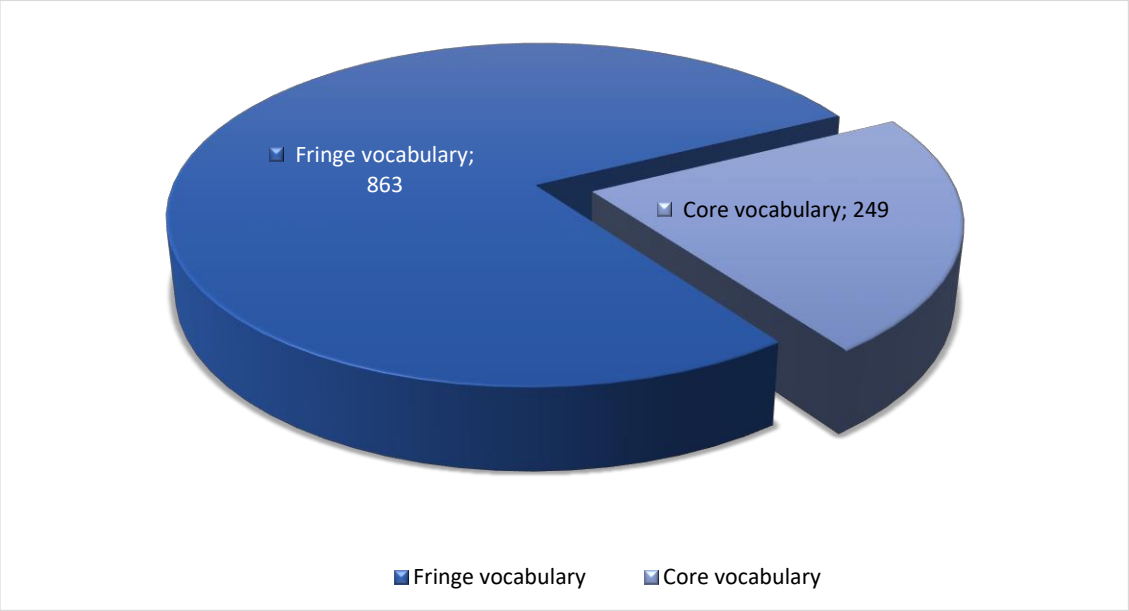


Figure 6. The number of different words in the composite sample that were designated as core versus fringe vocabulary

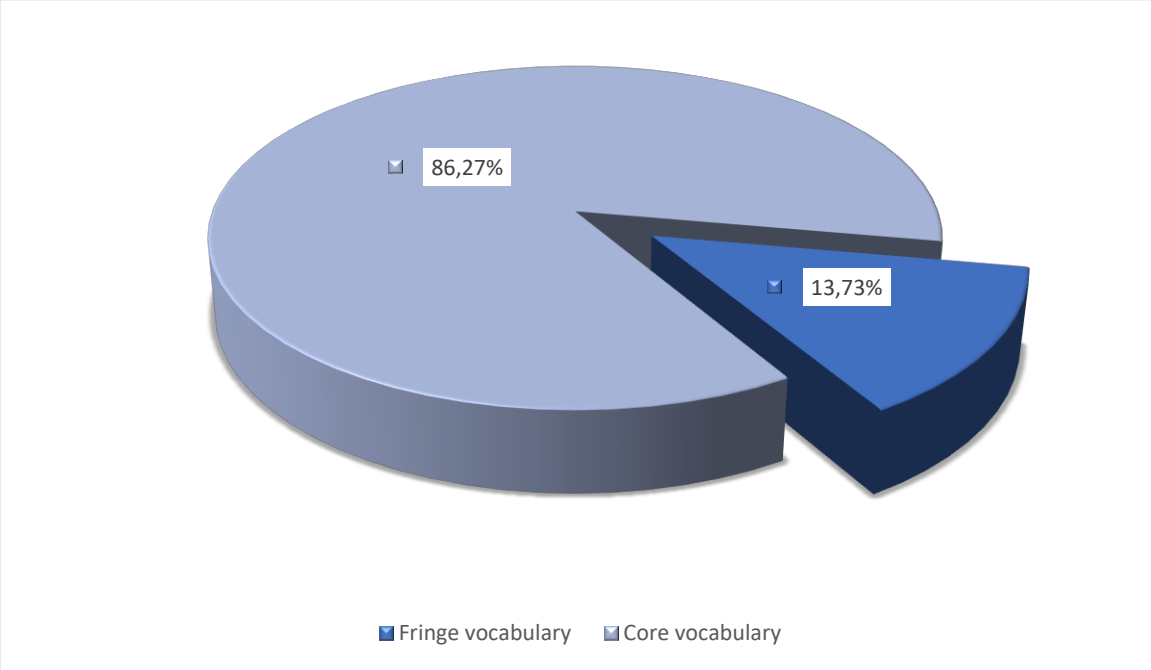


Figure 7. Coverage of core and fringe vocabulary

4.3 Further description of the core vocabulary

4.3.1 Commonality scores of core vocabulary by frequency

The average commonality score decreases with the reduced frequency of the words. Within the top 25, all the words were used by all participants. Subsequently, within the top 50 words, all words were used by all participants except for one word which was used by only three participants. The top 226 to 249 words were used by about four participants on average.

Figure 8 shows the range of commonality scores as the frequency of the core vocabulary reduces. The average commonality score among all core words amounted to 4.85, thereby showing that the core words established were used by most participants.

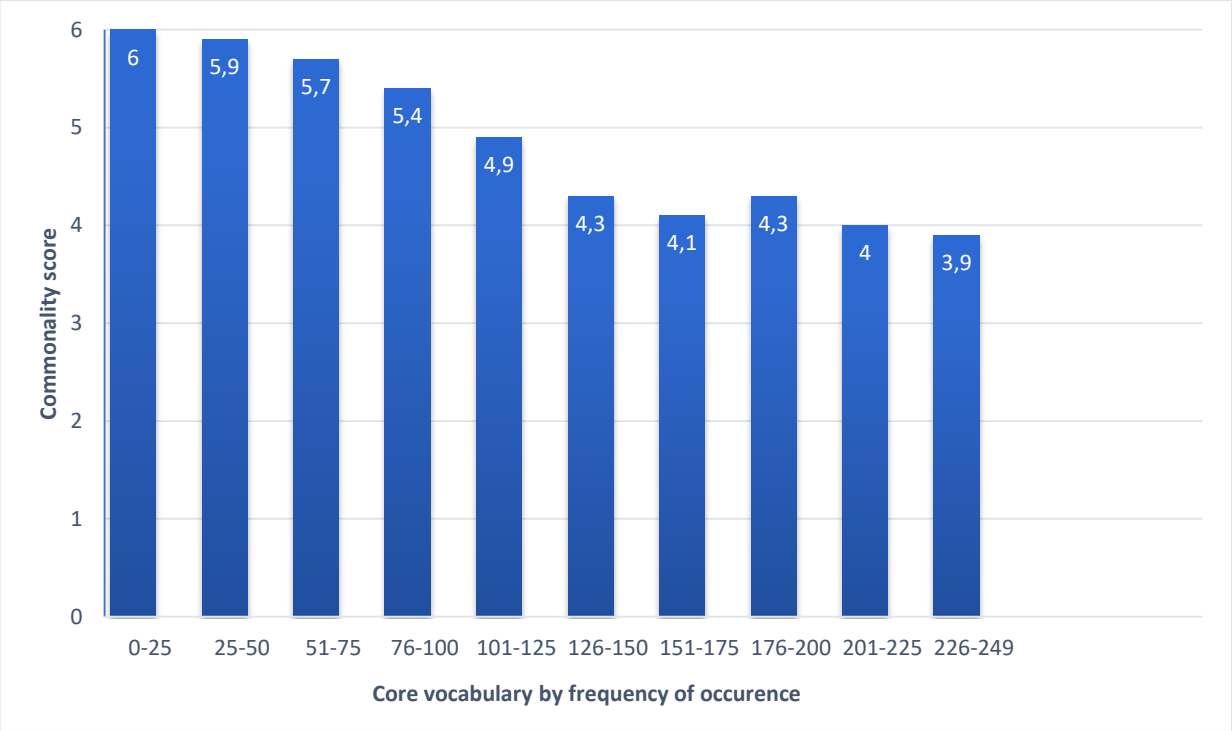


Figure 8. The average commonality scores of core vocabulary grouped by frequency of occurrence

4.3.2 Classification into content and structure words

The core vocabulary list was further analysed and was classified into content and structure words. Content words are words that have meaning on their own. These include nouns, verbs, adjectives, and adverbs (Shi et al., 2006). Structure words, also known as function words, are syntactical words that give meaning to sentences wherein content words appear. Setswana structure words include prepositions, concords, conjunctions and pronouns.

Figure 9 gives a visual representation of the number of different content and structure words found in the core vocabulary.

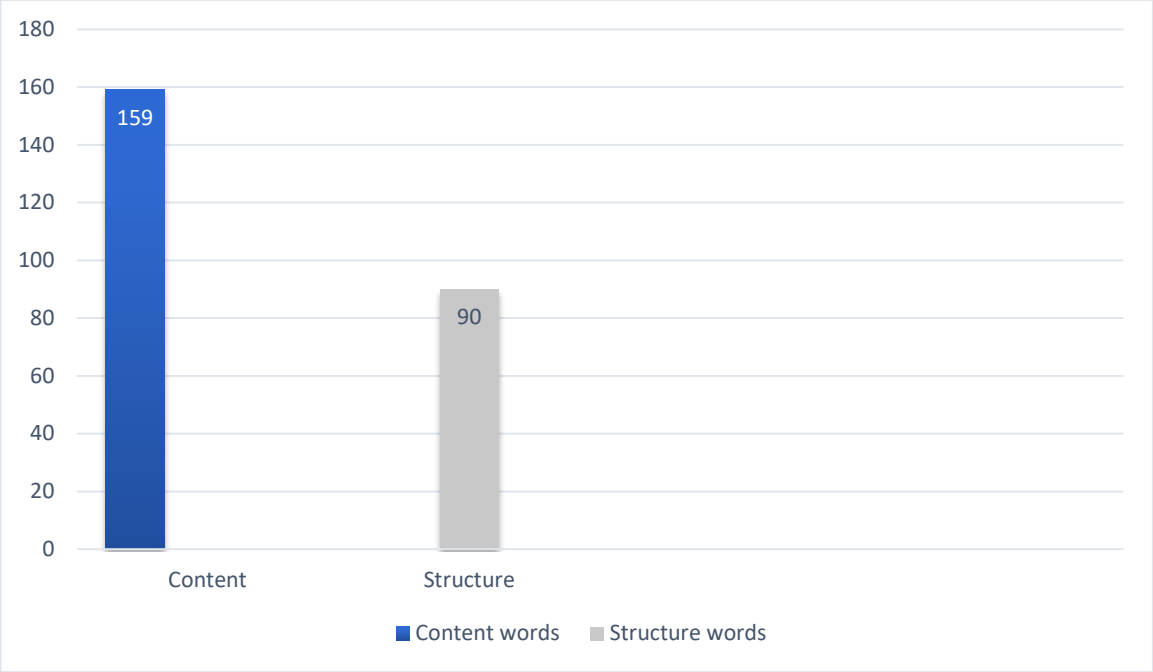


Figure 9. Number of content and structure words found in the core vocabulary

One hundred and fifty-nine content words constituted 64% of the core vocabulary, meanwhile, ninety words constituted 36% of the core vocabulary.

4.3.3 Classification by parts of speech

All core words established were further classified by parts of speech. This was done by locating each of the core words in the *Oxford (Setswana-Seesimane) Thanodi ya Sekolo / School Dictionary* (Phuti, 2019). The classification stipulated by the dictionary was used to tag the words by word class. When the dictionary was not comprehensive enough, a grammar book by Cole (1975) and a morphology book by Krüger and Pretorius (2006) were consulted for further reference. Code switches to English were classified according to the English section of the dictionary. The classification of core words was conducted by the student and verified by the supervisor. The number of different core words falling into different word classes, the number of occurrences in the core sample as well the frequency of occurrence of each part of speech that appeared in the core vocabulary were calculated. Table 8 displays the results obtained. The parts of speech are arranged from most to least frequently occurring.

Table 8

Word Classes/Parts of Speech Occurring in the Core Vocabulary Sample

Parts of speech	NDW	Proportional percentage in core (by NDW)	Number of occurrences in sample	Frequency of occurrence (‰) in total sample
Verbs	75	30.1%	3,029	167.4‰
Nouns	48	19.3%	1,651	91.2‰
Adverbs	23	9.2%	948	52.4‰
Pronouns	22	8.8%	2,636	145.6‰
Interjections	22	8.8%	1,169	64.6‰
Demonstratives	14	5.6%	657	36.3‰
Concords	13	5.2%	3,479	192.2‰
Adjectives	8	3.2%	164	9.1‰
Possessive concords	7	2.8%	783	43.26‰
Auxiliary verbs	6	2.4%	528	29.2‰
Conjunctions	5	2%	281	15.5‰
Enumerative	4	1.6%	67	3.7‰
Quantitative	1	0.4%	12	0.7‰
Prefix	1	0.4%	210	11.6‰
Total	249	99.8%[a]	15,614	862.8‰

[a] Due to rounding the percentage does not add up to exactly 100%.

Figures 10 and 11 illustrate the NDW by parts of speech and the frequency of occurrence of each part of speech in the total sample.

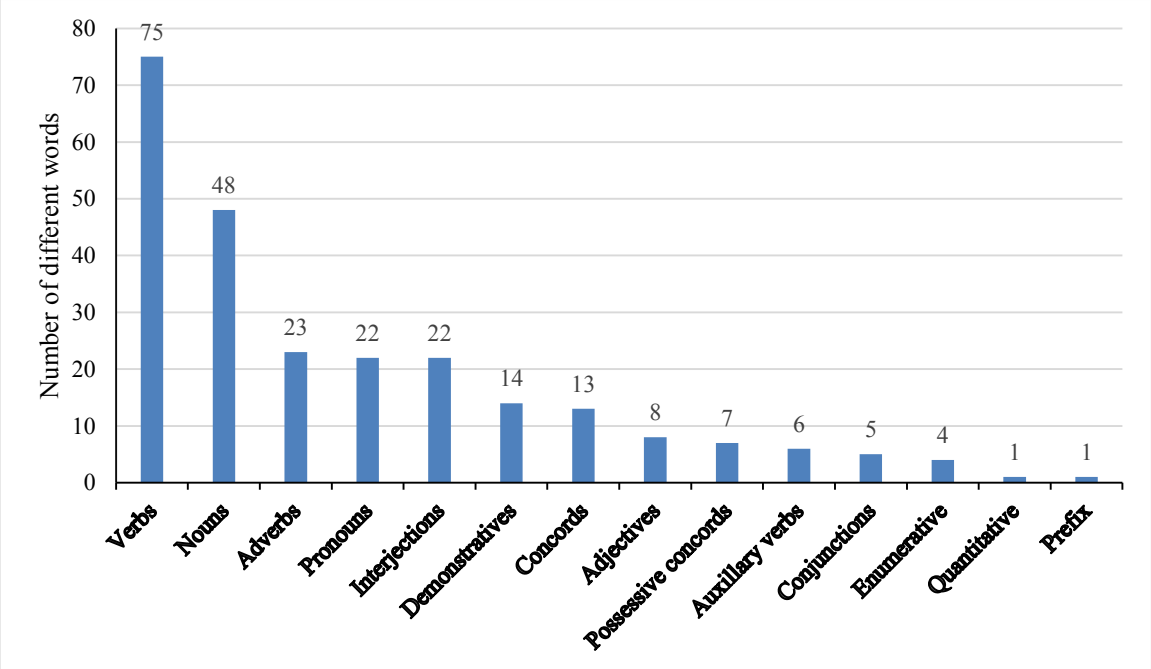


Figure 10. Number of different words of core vocabulary according to parts of speech

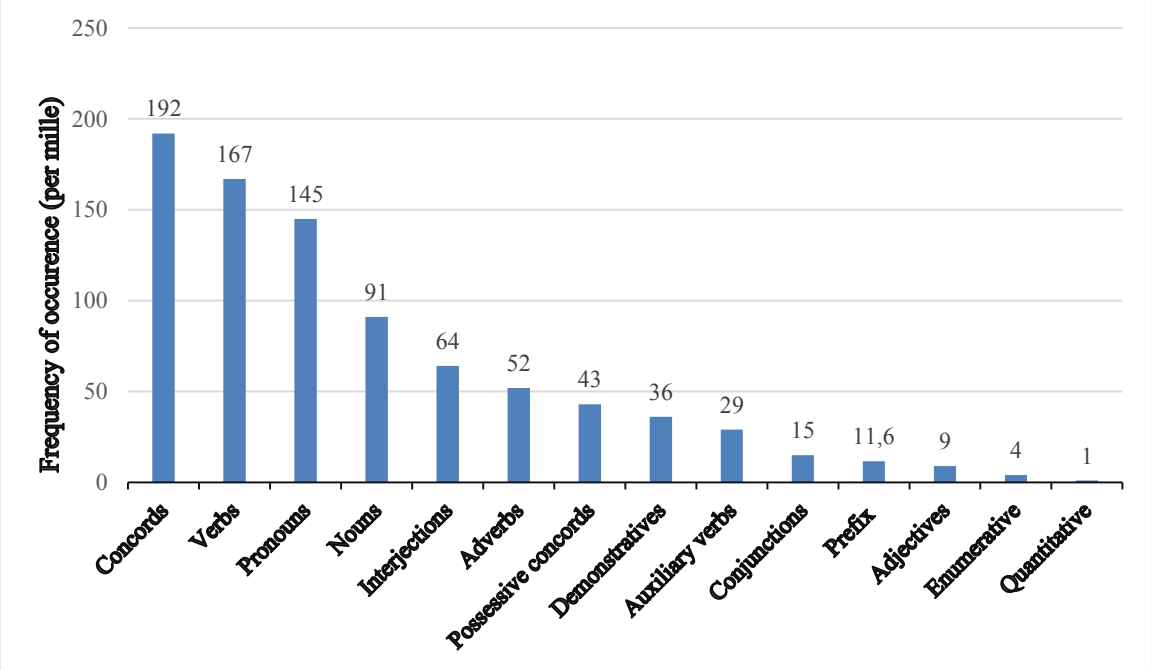


Figure 11. Frequency of occurrence of the different parts of speech in core vocabulary sample

Figure 10 illustrates that the core sample comprised a high number of different verbs (75) and nouns (48) and as such a low number of different enumeratives.

Figure 11, in turn, shows that concords were the most frequently used part of speech in the core sample with a high occurrence of 192‰. Other parts of speech occurring frequently were verbs, pronouns, nouns and interjections. Altogether these top five parts of speech accounted for a frequency of 659‰ during the participants' conversations. In comparing the two figures, it becomes clear that the NDW in a part-of-speech category does not directly predict the frequency with which this category is used in the sample. For example, although only 13 different concords were found, they accounted for 192‰ (i.e., nearly 20%) of the words used.

4.3.4 Comparison of core vocabulary parts of speech across three African languages

Due to the scope of the mini-dissertation is limited, only the coverage (i.e., frequency of occurrence) of the different parts of speech found in the Setswana core vocabulary was compared to that found in two core vocabulary lists for other African languages, namely Sepedi (Mothapo, 2019) and isiZulu (Mngomezulu, 2017). All three languages belong to the Niger-Congo language group (Eberhard et al., 2022) whereas Sepedi and Setswana also belong to the same subgroup, namely the Sotho language group. Similarities between the proportions of different parts of speech may, therefore, be expected, due to the similarities in language structure.

However, the comparison needs to be interpreted in light of some differences between the studies. Mngomezulu (2017) analysed her language sample by frequency of formatives (morphological analysis) rather than by frequency of orthographic words, and, therefore, differences may be expected. Also, dictionaries and grammar books are not always in complete agreement as to the classification of a word into a specific part of speech, and this may also lead to some differences. In the current study, for example, concords and possessive concords were classified into different parts of speech, while Mngomezulu (2017) and Mothapo (2019) classified all concords together.

Figure 12 shows the frequency with which core words from different parts of speech occurred in the samples.

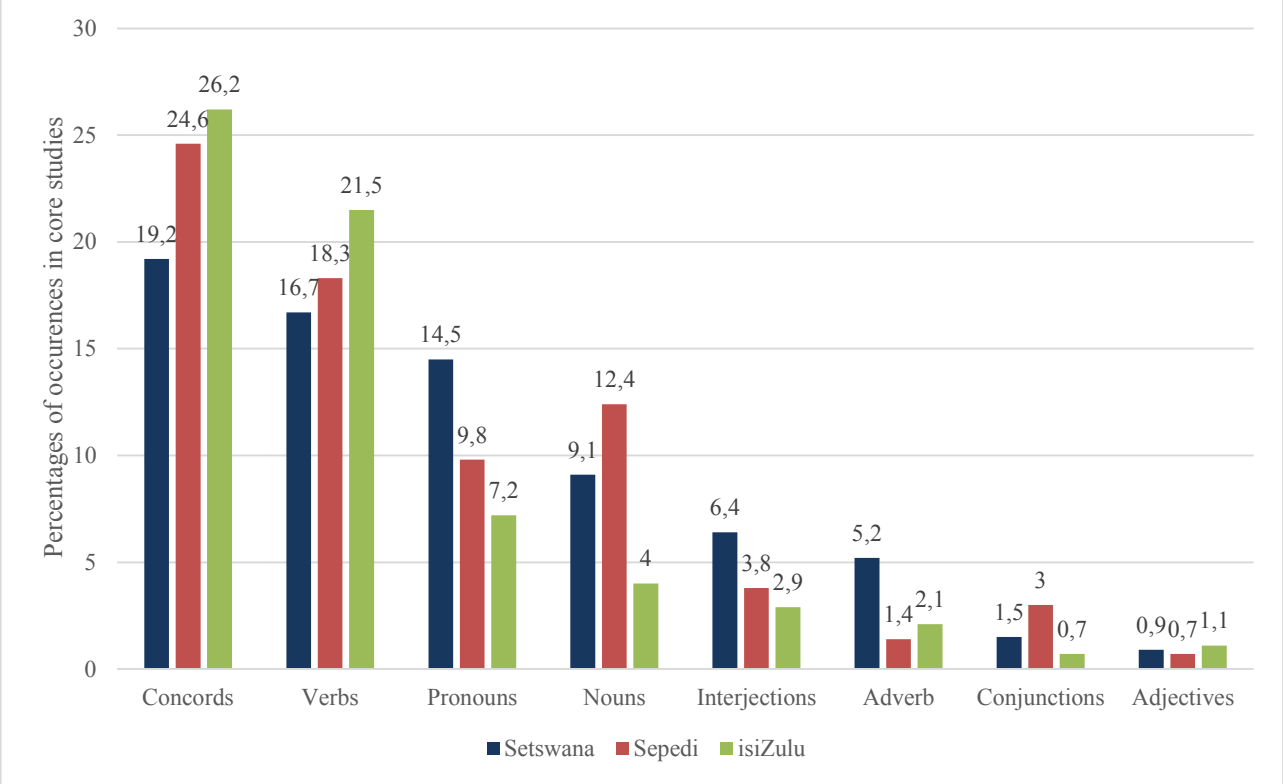


Figure 12. Percentage of occurrence of Setswana, isiZulu and Sepedi core words by various parts of speech

Concords and verbs were the core words that, as a group, occurred most frequently in the samples. Pronouns occurred more frequently in the Setswana study compared to the other two languages. The category of nouns showed a more prominent occurrence in the Sepedi study compared to the other two languages. However, in all three studies, concords, verbs, pronouns and nouns were the four most frequently used parts of speech. In contrast, conjunctions and adjectives showed a similarity in their lower frequency of occurrence.

4.4 Summary

The results of this study were presented according to the sub-aims of the study. The TNW and NDW were identified in the language samples of six Setswana-speaking learners during regular preschool activities. The frequency and commonality criterions were applied in order to determine the Setswana core vocabulary list. The Setswana core vocabulary list consists of 249 words which covered 86,27% of the learners's discourse. Futhermore, the Setswana core vocabulary list was described by commonality scores; content versus structure words and lastly classified by parts of speech.

5. DISCUSSION

5.1 Characteristics of the sample, core, and fringe vocabulary

The TNW in the composite sample of this study comprised 18,099 words. The NDW of the sample totalled 1,112. The type-token ratio was determined to be 0.06. The type-token ratio is used to evaluate the linguistic diversity of the TNW in the sample. According to Richard (1989), a smaller ratio is indicative of less unique or different words used. The participants in this study reused the same vocabulary when in conversation. The total number of words obtained from the study was high, however, the number of unique words was rather less. This finding indirectly expresses the concept of core vocabulary. The learners in this study reused words often in their everyday conversations.

The same trend is identifiable in other core vocabulary studies wherein similar size samples were obtained and when compared, the TTRs are almost similar. In the Sepedi study, Mothapo et al. (2021) collected a composite sample of 17,569 words and found a TTR of ~0.06. In the English study conducted by Boenisch and Soto (2015), a composite sample of 19,885 words was reported with a similar TTR of ~0.07. Furthermore, the isiZulu study conducted by Mngomezulu et al.(2017) which analysed the most frequently used formatives (morphemes) rather than orthographic words also yielded a TTR of ~0.06 on a composite sample of 20,137 formatives.

Two other core vocabulary studies have shown a similarity between calculated TTRs. The study conducted by Trembath et al. (2007) with English-speaking Australian children yielded a composite sample of 18,000 with a TTR of ~0.08. Similarly, the Afrikaans study conducted by Hatting et al. (2020) yielded a TTR of ~0.08 on a composite of 39,645 words.

It is notable that the TTRs in all these mentioned studies are similar and mostly comparable, which suggests a similar trend in the way different languages are used by children across different contexts. Speakers of languages typically use the same words over and over to convey different types of meanings (Baker & Change, 2006). It is very unlikely that a speaker will use a completely different set of words from one day to the next. The lower TTR found in these studies additionally suggests that including the most frequently occurring words on an AAC system may be a plausible and usable method for equipping individuals with CCN with some of the needed vocabulary for their everyday conversations.

The core word list established in this study amounted to 249 words. This was comparable to core lists previously established by Mothapo et al. (2021), Hattingh and Tönsing (2020) and Mngomezulu et al. (2019) who established core lists of 226 Sepedi words, 239 Afrikaans words, and 238 isiZulu formatives respectively. Similar findings have also been reported by Trembath et al. (2007) who established a list of 263 English core words and Robillard et al. (2014), who established a list of 216 French core words. All these studies used the same criteria to define a core word/formative, namely a frequency of occurrence of at least 0.5‰ and a commonality criterion of at least 50% (i.e., at least half of the participants sampled had to use the word). These findings, therefore, strengthen the premise that a small pool of words or other semantic units is frequently reused by speakers across different languages.

Furthermore, similar characteristics were found regarding the coverage that the core vocabulary provided across each language sample collected in the different studies. The coverage of the Setswana core words accounted for 86% of the sample – very similar to the coverage of the Sepedi core which accounted for 88% of the sample (Mothapo et al., 2021). The coverage of the isiZulu core formatives was 88.9% (Mngomezulu et al., 2019). The coverage of the English, French and Afrikaans core vocabularies (Hattingh & Tönsing, 2020; Robillard et al., 2014; Trembath et al., 2007) varied between 79.4% and 80.2% – interestingly somewhat lower than the coverage found in the African languages. One reason may be that concord (the agreement of one word with another to ensure grammaticality) is linguistically expressed by orthographically separate words in Setswana and Sepedi (e.g., short words like *o, le,* and *ba*). This gives rise to a part of speech called ‘concord’ which does not exist in Indo-European languages like French and English. While concords are not orthographically separate in isiZulu, the analysis on the formative level conducted by Mngomezulu et al. (2019) did result in them being separated as unique units. As a result, these frequently reused units (concords) may have given rise to higher coverage of the Setswana, Sepedi and isiZulu core vocabulary.

In all studies, the fringe vocabulary consisted of many more different words than the core vocabulary, but these words had a much smaller coverage of the collected language sample. The Setswana fringe vocabulary consisted of 17,850 words, covering only about 14% of the sample. The coverage of the fringe vocabularies established for French (Robillard et

al., 2014), English (Trembath et al., 2007), Afrikaans (Hattingh & Tönsing, 2020), isiZulu (Mngomezulu et al., 2019) and Sepedi (Mothapo et al., 2021) also ranged between 11 and 20%, with the number of different fringe words ranging from 14,432 to 39,415. This attests to the infrequent use of a great many different words that make up the fringe vocabulary in different languages. It also illustrates the challenge that fringe vocabulary poses in its inclusion on AAC systems – the vast number of fringe words requires that their inclusion on systems is considered thoughtfully and that a logical organisation is adopted to ensure that they are easy to find (Fallon et al., 2003; Thistle & Wilkinson, 2015). Personalisation is more important when considering fringe vocabulary than core vocabulary – while core vocabulary is common across speakers, fringe vocabulary is often unique to the interests, preferences and contexts of an individual (Trembath, 2007).

5.2 Further description of the Setswana core vocabulary

The Setswana core vocabulary showed a high degree of commonality across participants, with the most frequently used words (i.e., top 25) being used by all six participants. Similar findings were reported by Beukelman et al. (1991) and Trembath et al. (2007). This underscores the notion that core vocabulary is used across speakers, and suggests that it may be generally useful on AAC systems that cater for children who are beginning to learn to combine words and produce sentences (Laubscher & Light, 2020).

The Setswana core vocabulary list was also categorised into content versus structure words, for example, *bona* (verb) falls within content words and the word *ke* (concord) would be categorised as a structured word. The core vocabulary comprised about two-thirds of content words and about one-third of structure words. These findings are comparable to those of Mothapo et al. (2021), Boenisch and Soto (2015), Mngomezulu et al. (2019) and Hattingh and Tönsing (2020), whereby similar ratios of content to structure words were found. The most frequently occurring structure word in this study was found to be *ke* (‘I’) and the most frequently occurring content word was *bona* (‘look’*).* The word *ke* can refer to different people, depending on the speaker, and hence lacks specificity and is not complete on its own. On the other hand, the word *bona* (the verb form in this case) describes a specific action and can be used on its own, as an imperative (‘Look!’). However, it is clear that in both cases, combining these words with other words can enable the user to express a variety of different meanings.

Although the Setswana core vocabulary contained fewer structure core words than content core words, structure core words covered more than 50% of the total sample. On the other hand, a high number of different content core words were found, but these only covered about one-third of the composite sample. Fringe words accounted for the remaining 13.7% of the composite sample. A similar trend has been observed in other studies (Mothapo et al., 2019, Hattingh & Tönsing, 2019, Mngomezulu et al., 2017), where fewer structure words accounted for high coverage of the total sample.

Structure words have shown a significant occupancy in the core vocabulary, given their high frequency of use. Structure words are not always prioritised in AAC systems (McFadd & Wilkinson, 2010; Adamson et al., 1992). One reason for this could be that content words are much easier to represent by symbols than it is to represent structure words (Mngomezulu et al., 2017; Hattingh & Tönsing, 2019; Smith & Witten, 1993). Adamson and colleagues (1992) added that when selecting vocabulary for AAC, food items and object names are usually the first symbols added to an AAC system seeing that beginning communicators usually need vocabulary for more tangible concepts. Most structure words, in turn, are abstract and fulfil a grammatical function rather than a lexical one. Laubscher and Light (2020) argue that structure words on an AAC system become important when morphosyntactic skills emerge but question their usefulness in AAC systems for beginning communicators. According to these authors, the first 50 words produced by (English-speaking) young children are mostly nouns. However, it should also be noted that structure words appear frequently once children start to combine words. Banajee et al. (2003), for example, found that the 23 words most frequently used by two- to three-year-old English-speaking toddlers were almost all structure words.

The inclusion of structure words in an AAC system requires thoughtful consideration. One disadvantage of omitting them is the risk to stunt the user's expressive abilities. AAC devices that are too limiting tend to be abandoned (Moorcroft et al., 2019). Also, since it has been established that aided input or modelling is an effective method of improving the communication skills of children who require AAC (Allen et al., 2017; Dada et al., 2020; Sennott et al., 2016), communication partners may need access to structure words to provide models that are more complex than those produced by the child using the system, to scaffold language development (Von Tetzchner & Stadskleiv, 2016). However, structure words

always need to be combined with content words to provide a system that enables the generation of novel utterances (Mngomezulu et al., 2019).

The Setswana core vocabulary established in this study could be classified into 14 different parts of speech (see Table 8 in Section 4). Many of these parts of speech are also noted in other studies which used similar classifications (Boenisch & Soto, 2015; Hattingh & Tönsing, 2020; Mngomezulu et al., 2019; Mothapo et al., 2021; Robillard et al., 2014; Trembath et al., 2007). Studies that analysed units on a morphological level included other different parts of speech such as prefixal formatives, suffixes and vocative formatives (Shin & Hill, 2016; Mngomezulu et al., 2019). Some of the parts of speech identified in this study are unique to the specific group of African languages to which Setswana belongs. These languages have an elaborate noun class system, and each noun used in a sentence requires that agreement be made to the specific class through the use of concords. Concords are, therefore, a part of speech that is unique to these languages, and the Setswana core vocabulary contained several concords (Harman, 1980).

When analysing the NDW per part of speech category, it was discovered that the Setswana core vocabulary contained a large number of different nouns and verbs. These two categories accounted for about half of the NDW in the core vocabulary. Mothapo et al. (2021) and Mngomezulu et al. (2019) found similar patterns in the Sepedi and isiZulu core vocabulary, with verbs/verb roots and nouns/noun roots containing the highest NDW. Verbs were similarly prominent in number in the English and Afrikaans core vocabularies established by Boenisch and Soto (2015) and Hattingh and Tönsing (2020) respectively. Nouns were a little fewer in number in the latter two studies, with other parts of speech containing a higher NDW (e.g., adverbs). Robillard et al. (2014) found only one noun in the core vocabulary of French children. Linguistic similarities between the African versus the Indo-European languages[1] are once again evident. As in other studies, the Setswana core vocabulary also contained a fair number of different adverbs, pronouns, and interjections. A lower number of demonstratives, adjectives, concords and other parts of speech was found.

[1] Although Afrikaans is a language that originated in South Africa, its linguistic classification is Indo-European, as it has its roots in the Dutch language (Eberhard et al., 2022).

It is clear that the Setswana core vocabulary is characterised by the presence of many different parts of speech – in this regard, it also resembles core vocabularies in other languages. The importance of including different parts of speech on AAC systems that intend to give access to novel utterance generation is thereby highlighted – sentences contain different parts of speech. A predominance of nouns to the exclusion of other parts of speech on AAC systems will not allow the flexible combination of words to generate a range of different meanings (Mngomezulu et al., 2019).

When assessing the frequency with which different parts of speech in the Setswana core vocabulary were used (i.e., coverage), the similarities to isiZulu and Sepedi were once again noted. For example, core concords, verbs, pronouns and nouns all appear with high frequency in the sample. Setswana concords surpass all other parts of speech when comparing the frequency of use. As previously explained, concords ensure agreement between nouns and other parts of speech in the sentence. They can also fulfil a pronominal function, where they stand in the place of a noun (e.g., *Ba a bala* – 'They read', where the concord *ba* takes the pronominal function expressed by the pronoun 'they' in English). Concords are part of almost all Setswana sentences. This is also true for the Sepedi and isiZulu languages, as can be seen by the fact that concords were the most frequently used part of speech in the core vocabularies established by Mngomezulu et al. (2019) and Mothapo et al (2021). Concords are not found in many of the other languages in which core vocabulary studies were conducted (e.g., French, English, and Afrikaans). However, the very high occurrence of pronouns in the Afrikaans study (Hattingh & Tönsing, 2020) suggests that words with pronominal functions are also used with a high frequency in other languages. Interestingly, the frequency of use of pronouns was also high in this study (third highest when comparing parts of speech) and in the Sepedi (Mothapo et al., 2021) and isiZulu (Mngomezulu et al., 2019) core vocabularies. The inclusion of concords on a Setswana AAC system seems imperative if grammaticality and sentence production are envisaged. Yet their inclusion also poses challenges, as concords are not easily represented by graphic symbols – their meaning depends on the noun they are referring to. One may also argue that concords can be inferred by the communication partner in many instances, and their omission may not always impede understanding. As Sutton et al. (2002) argued, familiar communication partners can often discern the meaning of ungrammatical utterances, and the omission of grammar elements may increase the rate and decrease physical efforts of production for the person using AAC.

Verbs occurred with the second highest frequency when comparing parts of speech. The 75 different core verbs found in this study occurred 3,029 times in the sample, covering about 17% of the total sample. Verbs have also been found to be frequently used in other core studies, for example, Mothapo et al., (2021) found that 83 different verbs were used with a frequency of almost 20%, while Mngomezulu et al., (2019) found that 81 different verbs (verb roots and verbal auxiliaries) were used with a frequency of just over 20% in the total sample. This suggests that verbs are among the topmost used parts of speech in most languages and their significance is seen in the meaning they bring to sentences. Mairal and Gil (2006) affirm that verbs are frequently used across all human languages and are used in almost every sentence. It, therefore, seems imperative to include specifically frequently used verbs in AAC systems. Many core verbs found in this study seem to be usable in various contexts, such as *tla* ('come'), *fa* ('give'), *baa* ('put') and *fetsa* ('finish'). A few may have been specific to the age group and the school context, such as *tshameka* ('play') and *khalara* ('colour in'). The importance of including both more generic verbs and also customising verbs to the age and context where the AAC system will be used is, therefore, illustrated.

Besides core concords, verbs and pronouns, core nouns were also used relatively frequently, appearing 1 651 times in the sample and covering about 9% of it. The coverage provided by core nouns/noun roots was slightly higher in Sepedi (about 12%) but lower in Afrikaans and isiZulu respectively (about 3-4%). Some influence of the context in which the study was conducted can be seen in the nouns identified as core words. The school context may have led to the inclusion of nouns like *moswinki* ('swing'), *kherayone ('crayon')* and *sekolo ('school').* The fact that data collection took place during the Covid-19 pandemic led to the inclusion of the codeswitch noun '*mask*'. Nouns carry lexical meaning and establish a clear frame of reference in communication. They are especially important when communicating about objects and persons not present, as these need to be specified through pertinently referring to them using symbols. In contrast, objects and persons who are present can be indicated, for example, by eye or finger pointing. Although nouns can be replaced by pronouns or concords in subsequent sentences, these pronouns and concords will remain unspecific if not linked to the noun they replace. Many nouns are easy to represent by graphic symbols due to their concrete nature which lends itself to highly iconic representation. AAC systems should not only prioritise nouns for this reason but should also include other parts of

speech that occur in high frequencies (Mothapo et al., 2021). However, nouns should likewise not be omitted due to the important lexical meaning they convey.

Code-switching is not a part of speech but was, however, observed in the Setswana core list. This is when more than one language is used in a single conversation (Reyes, 2004). The Setswana language speakers were influenced by colonisation and language contact with Western language speakers (Harman, 1980). Due to this reason, the Afrikaans language, spoken by the colonisers of the Batswana settlements (the Dutch) was adopted by the Batswana population. Since then, Afrikaans borrowed words and code-switches were then observed in the Setswana language. The words seen in the list are *moes* ('had to/ must've'), *man* ('man') and *maar* ('but'). Code-switching is a common phenomenon among multilingual populations and has been observed in the conversations of South Africans (du Plessis, 2006; Slabbert & Finlayson, 2000). English code switch words were also found in the core list, namely: '*sharp', 'why', 'mask*', and '*sorry*'. Mothapo et al. (2021) as well as Hattingh and Tönsing (2020) also found English code switches in the Afrikaans and Sepedi core vocabularies. In South Africa, children are likely to be exposed to English, for example, through television and radio. English has been described as the de facto lingua franca in South Africa (Khokhlova, 2015) and is, therefore, more prone to appear in various languages spoken by children exposed to it.

5.3 Summary

The established Setswana core vocabulary list showed similarity in TNDW and coverage to other core vocabulary lists established. Similarities were also observed in the proportion of content and structure words as well as the coverage of each of these categories. Once again, the importance of structure words for sentence building was highlighted by the frequency with which the structure words occurred in the composite sample. The percentage of occurrence of parts of speech also showed similarities with related languages, namely Sepedi and isiZulu. The content of the core vocabulary list showed some influence of context, but many generic words were also included. The core word list represents the first effort to establish a resource for the selection of vocabulary for children in need of AAC from a Setswana language background, in an attempt to relate more equity in the availability of appropriate resources for children with CCN from diverse backgrounds.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary of main findings

Establishing core vocabulary lists is one of the widely used methods to obtain vocabulary to include in graphic symbol-based AAC systems (Bean et al., 2019; Beukelman & Light, 2020). Preselecting vocabulary for preliterate individuals may not be a straightforward process as these individuals often do not influence the vocabulary selected for their interactions. Thus, selecting appropriate and relevant vocabulary for an individual with CCN may be a daunting process for any AAC team and needs to be conducted thoroughly and meticulously. Core vocabulary lists are one resource that AAC teams can consider in selecting vocabulary for preliterate individuals.

The main aim of this study was to identify the core vocabulary of Setswana-speaking Grade R learners without disabilities, as a resource to guide AAC system vocabulary selection for pre-schoolers from a Setswana language background who need AAC. Six learners were recruited from three different preschools. Each participant was fitted with a voice recorder and lapel microphone during their preschool daily activities. A range of 2,900 to 3,100 words was collected from each participant. After excluding all unintelligible units from the transcribed composite sample, a total of 18,099 orthographic words were obtained.

Frequency counts and commonality scores were calculated to determine a core vocabulary list of 249 different words. This core list accounted for 86% of the composite sample. These findings were consistent with results from core vocabulary studies in various other languages such as Sepedi (Mothapo et al., 2021), English (Boenisch & Soto, 2015; Trembath et al., 2007), isiZulu (Mngomezulu et al., 2019) and Afrikaans (Hatting & Tönsing, 2020). As in other languages, Setswana speakers, therefore, reuse a small set of words frequently, and these words cover a significant proportion of their conversations. The characteristics of the Setswana core vocabulary in terms of commonalty scores, content versus structure words and parts of speech were found to be largely similar to other core vocabulary studies (Boenisch & Soto, 2015; Mngomezulu et al., 2019; Mothapo et al., 2021; Robillard et al., 2014; Trembath et al., 2007). More similarities were found with the core vocabularies of other African languages of the same language family, specifically, isiZulu and Sepedi as compared to core vocabularies in Indo-European languages, such as English, French and Afrikaans.

The established Setswana core vocabulary list can be utilised as a resource for vocabulary selection for Setswana AAC systems intended for preliterate children. The use of this specific vocabulary in AAC can enhance the language production by pre-schoolers in need of AAC through access to a range of content and structure words from various parts of speech to enable the ability to formulate grammatically correct sentences, expression of novel meaning and language learning.

6.2 Implications for practice

The established Setswana core vocabulary of 249 words can be utilised as a vocabulary source when selecting vocabulary for inclusion in a Setswana graphic symbol-based AAC system for preschool-aged children. The Setswana language is spoken by approximately 8% of the South African population (Statistics South Africa, 2011), 77% of the Botswana population and 5% of the Namibian population (CIA, 2022). Van Tilborg and Deckers (2016) suggested that vocabulary can be selected from core lists for individuals of other ages or used across other settings. This list is, therefore, expected to have a significant clinical application in South Africa and across neighbouring countries to serve a large population of children with complex communication needs who have Setswana language backgrounds. This action, however, should be exercised with caution, keeping in mind this specific vocabulary list may have been influenced by the specific context, dialect and population.

This core list is not intended to be used in isolation but rather in combination with other context-specific vocabularies such as fringe vocabulary which is specific to the individual and culturally appropriate (Balandin & Iacono, 1998; Beukelman et al., 1991; Robillard et al., 2014). Other reliable sources of vocabulary such as informants or environmental inventories (Fallon et al., 2001; Sturm & Clendon, 2004) should be consulted to ensure informed decisions and a designed system that speaks to the needs of the individual with CCN. The core list established together with others can be beneficial to individuals with CCN.

Though this vocabulary list was obtained from speech samples of Grade R learners without disabilities, studies have shown that children who have disabilities as well as those without disabilities use the same words frequently in their daily conversations, therefore, showing great similarities in core vocabularies between both groups (Boenisch, 2014;

Deckers et al., 2017; Robillard et al., 2014). This proves that the communication needs and opportunities of these heterogenous groups are comparable, and the list can help meet the demands of their communication and interaction environments (Hattingh et al., 2020).

Clinicians and AAC team members can use this list for other intervention goals outside of AAC. For example, it can be consulted in the development of formal and informal language and vocabulary assessments. It can also be used as a guide for vocabulary used by Setswana children of the studied age and can thus be used to guide language therapy goals. It is important, however, to note the small participant group used in the study, thus the list should be used with caution and over-generalisation of the results should be avoided.

Given that South Africa is a multilingual country with most communication therapists being monolingual (Crago et al., 1997), it is important to make multilingual and contextually relevant resources available to the population in need (Pascoe & Norman, 2011). The availability of such resources will help to preserve language and richness. That said, AAC system developers should explore how the Setswana list can co-occur with other languages on high-tech AAC system devices and how multilingual children can use this. Multilingual disabled children should also be allowed to learn their native languages and not be restricted to English language use as a common language.

6.3 Critical evaluation of the study

6.3.1 Strengths

This is the first study that aimed at identifying a Setswana core vocabulary list based on the language samples of Setswana Grade R learners. This list provides a novel resource for AAC practitioners in selecting the vocabulary items for pre-schoolers needing aided AAC systems.

In addition, the adopted data collection method and analytic processes strengthened the internal validity of this study, thereby validating the results of the study. Learners were recorded in natural settings where naturally occurring conversations took place, thus minimising any influence by the researcher. The 20-minute initial warm-up period omitted from the transcription and analysis process was useful in reducing participant reactiveness to the fitted equipment. During the recording review and transcription, it was found that most

learners spoke predominantly about the recording equipment in the first 10 to 15 minutes. This was also true in the pilot study. Therefore, the exclusion of this vocabulary in the first 20 minutes from the analysis strengthened the internal validity of the results. Learners were recorded throughout the preschool day, thus minimising the possibility of vocabulary reflecting only one activity, such as playtime or circle time only.

Comparing each transcription against the original recording by an independent person increased the transcription reliability. The inter-rater reliability varied from 92.3% to 98.5% when assessing 20% of each transcript; this shows that the coding was executed reliably. These two aspects strengthened the internal validity of the results.

Root words of all words (verbs, nouns, adverbs, adjectives, etc.) were counted together. This was done to avoid greatly diffusing the core list by separating inflected and uninflected forms of words. Inflected and uninflected forms of words are often represented by the same graphic symbol on an AAC system. Therefore, this method of analysis allowed for a better reflection of the frequencies of words as represented on AAC systems.

6.3.2 Limitations

Participants recruited in this study were from three different school sites, however, the sample size comprised only six participants, thus making up a relatively small sample size. This could affect the representativeness of the data to a larger population. In addition, having participants from similar ages (between 5 and 6 years of age) as well as collecting data within a short period (two to three days per child) could also influence the generalisability of the data obtained. It is not known whether the same core vocabulary would be established if the data were to be collected from a different participant group using dialectically unique characteristics specific to their region (Hatting et al., 2020).

Although a cool-off period of 20 minutes was exercised at the beginning of each transcription, participants were still prone to the Hawthorne effect. As known with observational designs, participant reactivity is a common and often unavoidable phenomenon. This was also true for this study. The recordings showed that participants had an ongoing cognisance of the fitted equipment which influenced how they behaved around their peers and the conversations they had. This may have affected the study's internal validity.

The background noise noted in the recordings could have affected the transcription process (transcription accuracy). Perhaps the audio recordings could have been supplemented with video recordings to improve accuracy, however, this would have disrupted the classroom routines and would have led to other ethical considerations such as confidentiality and privacy concerns.

In the study by Mothapo et al. (2021), it was noted that the frequency and commonality score criterion is somewhat arbitrary. Shin and Hill (2016) report that there has not been a scientific justification for using 50% (≥3) and a frequency count of 0.5 per cent per mille as criteria for the inclusion of words in the core list. These authors also mention that there are other methods of objective analysis, for example, grouped frequency counts, which could be used to define core versus fringe vocabulary (Shin & Hill, 2016).

Lastly, due to the limitation of available Setswana grammar books, the words established in the Setswana core list were classified according to their part-of-speech label described in the *Oxford Setswana-Seesimane School Dictionary* (Phuti, 2019). Though dictionaries provide a wide range of grammatical terms, it was noted that the dictionary used had a limited and superficial classification of parts of speech. Wachal (1994) attests that grammatical information presented in the prefatory material of most dictionaries is typically limited. In most cases, grammarians would typically classify words differently and add more grammatical refinement to each entry of part of speech. This factor should be kept in mind when doing a comparison with other core lists of parts of speech.

6.4 Recommendations for further studies

Due to the limited sample used in this study, it would be worthwhile to replicate this study with a greater sample size. This will allow for generalisation. The replicated study can be conducted in a different population that speaks a different dialect of the Setswana language. Other studies can explore samples taken in different contexts, for example, at home and comparisons of the core lists can be made to examine similarities. Samples from children of a younger and older age may also be useful.

The data obtained from this study (18,099 words) can be further analysed to investigate conversational topics among children of the same age. This data can also be further explored to examine different communication functions and fringe vocabulary usage of pre-schoolers.

Future studies can investigate and determine the most appropriate graphic symbol representation of the structure words with specific attention on uninflected forms of words, as these words are particularly difficult to represent graphically on AAC systems for preliterate children (Hattingh et al., 2020). In addition, researchers can explore the best possible organisation of these words on AAC grids to allow for easy access, and efficient communication and to promote AAC acceptance (Moorcroft et al., 2019).

Lastly, intervention studies are required to explore the language development of children with CCN when such vocabulary is implemented into their therapy goals, this can broaden the horizon on intervention methods and help communication partners understand how best to teach and implement the vocabulary in natural settings. These studies are important to establish the effects of core vocabulary on AAC systems and provide empirical evidence rather than just theoretical support (Hattingh et al., 2020).

6.5 Summary

This section provided the summary of the study as well as the summary of findings obtained, it also discussed the overall implications for practice with the results obtained. A critical evaluation of the study was provided by identifying the strengths and limitations associated. Lastly, recommendations for further studies were also discussed.

REFERENCES

Adamson, L. B., Romski, M. A., Deffebach, K., & Sevcik, R. A. (1992). Symbol vocabulary and the focus of conversations: Augmenting language development for youth with mental retardation. *Journal of Speech, Language, and Hearing Research*, *35*(6), 1333-1343.

Aggarwal, R., & Ranganathan, P. (2019). Study designs: part 2–descriptive studies. *Perspectives in clinical research*, *10*(1), 34. https://doi.org/10.4103/picr.PICR_154_18

Allen, A. A., Schlosser, R. W., Brock, K. L., & Shane, H. C. (2017). The effectiveness of aided augmented input techniques for persons with developmental disabilities: A systematic review. *Augmentative and Alternative Communication*, *33*(3), 149-159. https://doi.org/10.1080/07434618.2017.1338752

American Speech-Language-Hearing Association. (2022). Augmentative and alternative communication. 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*. National Development Agency.

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

Baker, B.R., & Chang, S.-K. (2006). A Mandarin language system in augmentative and alternative communication (AAC). *International Journal of Computer Processing of Languages, 19*(04), 225–237. https://doi.org/10.1142/S0219427906001438

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

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

Bean, A., Cargill, L. P., & Lyle, S. (2019). Framework for selecting vocabulary for preliterate children who use augmentative and alternative communication. *American Journal of Speech-Language Pathology*, *28*(3), 1000–1009. https://doi.org/10.1044/2019_AJSLP-18-0041

Beukelman, D., Jones, R., & Rowan, M. (1989). Frequency of word usage by nondisabled peers in integrated preschool classrooms. *Augmentative and Alternative Communication, 5*(4), 243–248. https://doi.org/10. 1080/0743461891233127529

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

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

Beukelman, D.R., & Light, J.C. (2020). *Augmentative and alternative communication: Supporting children and adults with complex communication needs*. Paul H. Brookes.

Beukelman, D.R., & Ray, P. (2010). Communication supports in pediatric rehabilitation. *Journal of Pediatric Rehabilitation Medicine,* 3, 279-288. https://doi.org/10.3233/PRM-2010-0139

Boenisch, J. (2014). Kernvokabular im Kindes- und Jugendalter: Vergleichsstudie zum Sprachgebrauch von Schülerinnen und Schülern mit und ohne geistige Behinderung und Konsequenzen für die UK (Core vocabulary in childhood and adolescence: Comparative study in language use of learners with and without intellectual disability and consequences for AAC). UK & Forschung: Special Supplement of Unterstützte Kommunikation, 3(1), 4–23.

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. https://doi.org/10.3109/07434618.2014.1001521

Bosma, E., & Blom, E. (2019). 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*, *23*(6), 1431–1447. https://doi.org/10.1177/1367006918798972

Curriculum and Assessment Policy Statement. (2012). CAPS for Foundation Phase. Retrieved from: https://www.education.gov.za/Curriculum/CurriculumAssessmentPolicyStatements(CAPS)/CAPSFoundation.aspx

Cole, D. T. (1955). *An introduction to Tswana grammar*. Longman.

Cole, D. T. (1975). *An introduction to Tswana grammar*. Longman.

Central Intelligence Agency (Ed.). (2011). *The World Factbook 2011*. Central Intelligence Agency.

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

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

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

Dada, S., Murphy, Y., & Tönsing, K. (2017). Augmentative and alternative communication practices: A descriptive study of the perceptions of South African speech-language therapists. *Augmentative and Alternative Communication*, *33*(4), 189–200. https://doi.org/10.1080/07434618.2017.1375979

Deckers, S. R. J. M., Van Zaalen, Y., Van Balkom, H., & Verhoeven, L. (2017). Core vocabulary of young children with Down syndrome. *Augmentative and Alternative Communication, 33*, 77–86. https://doi.org/10.1080/07434618.2017.1293730

Du Bois, J. W. (1991). Transcription design principles for spoken discourse research. *Pragmatics*, *1*(1), 71-106. https://doi.org/ 10.1075/prag.1.1.04boi

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

Eberhard, D. M., & Mangulamas, M. (2022). Who texts what to whom and when? Patterning of texting in four multilingual minoritised language communities and a preliminary proposal for the language repertoire matrix. *International Journal of the Sociology of Language*, *2022*(276), 169-205.

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. https://doi.org/10.1044/1058-0360(2001/010)

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 children based on developmental and environmental language sources. *Augmentative and Alternative Communication, 8*(1), 41–56. https://doi.org/10. 1080/07434619212331276033

Harman, F. J. N. (1980). *Pula I: Tswana Grammar.* F.J.N Harman Publishers.

Hattingh, D. (2019). *The core vocabulary of South African Afrikaans-speaking preschoolers without disabilities*. (Mater's dissertation, University of Pretoria).

Hattingh, D., & Tönsing, K. M. (2020). The core vocabulary of South African Afrikaans-speaking Grade R learners without disabilities. *South African Journal of Communication Disorders*, *67*(1), 1–8. https://doi.org/10.4102/sajcd.v67i1.701

Holland, A. L. (1975). Language therapy for children: Some thoughts on context and content. *The Journal of Speech and Hearing Disorders, 40*(4), 514–523. https://doi:10.1044/jshd.4004.514

Howard, K. (2003). *Language socialization in a Northern Thai bilingual community*. University of California.

Kettunen, K. (2014). Can type-token ratio be used to show morphological complexity of languages? *Journal of Quantitative Linguistics*, *21*(3), 223–245. doi:10.1080/09296174.2014.911506

Khokhlova, I. N. (2015). Global English: Africanisation of the English Language. *Review of European Studies*, (*7*,) p 201.

Krüger, C. J. H., & Pretorius, R. S. (2006). *Introduction to the Morphology of Setswana* (69). Lincom Europa.

Lahey, M., & Bloom, L. (1977). Planning a first lexicon: Which words to teach first. The *Journal of Speech and Hearing Disorders, 42(*3), 340–350. https://doi:10.1044/jshd.4203.340

Laubscher, E., & Light, J. (2020). Core vocabulary lists for young children and considerations for early language development: a narrative review. *Augmentative and Alternative Communication*, *36*(1), 43–53. https://doi.org/10.1080/07434618.2020.1737964

Mairal, R., & Gil, J. (2006). *Linguistic universals* (1st ed.). Cambridge University Press.

Maguvhe, M. O. (2014). Augmentative and alternative communication: Requirements for inclusive educational interventions. *International Journal of Educational Sciences*, *7*(2), 253–260. https://doi.org/10.1080/09751122.2014.11890187

Mahura, O. O., & Pascoe, M. (2016). The acquisition of Setswana segmental phonology in children aged 3.0–6.0 years: A cross-sectional study. *International Journal of Speech-Language Pathology*, *18*(6), 533–549. https://doi.org/10.3109/17549507.2015.1126639

Malema, G., Okgetheng, B., Tebalo, B., Motlhanka, M., & Rammidi, G. (2020). Complex Setswana Parts of Speech Tagging. *Proceedings of the First Workshop on Resources for African Indigenous Languages*, 21–24. https://aclanthology.org/2020.rail-1.4

Marvin, C. A., Beukelman, D. R., Brockhaus, J., & Kast, L. (1994). What are you talking about?'': Semantic analysis of preschool children's conversational topics in home and preschool settings. *Augmentative and Alternative Communication, 10*(2), 75–86. https://doi.org/10.1080/ 07434619412331276780

McFadd, E., & Wilkinson, K. (2010). Qualitative analysis of decision making by speech-language pathologists in the design of aided visual displays. *Augmentative and Alternative Communication*, *26*(2), 136-147.

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

McNaughton, D., & Babb, S. (2021). Supporting participation and communication in employment and volunteer activities for adolescents with complex communication needs. In B. T. Ogletree (Ed.), *Augmentative and Alternative Communication: Challenges and Solutions* (p. 117). Plural Publishing.

Miller, J., & Iglesias, A. (2012). Systematic analysis of language transcripts (SALT), research version 2012 [computer software]. *SALT Software, LLC*.

Mngomezulu, J. R. (2017). *Determining an AAC core vocabulary for Zulu-speaking preschool children*. (Master's dissertation, University of Pretoria).

Mngomezulu, J., Tönsing, K. M., Dada, S., & Bokaba, N. B. (2019). Determining a Zulu core vocabulary for children who use augmentative and alternative communication. *AAC: Augmentative and Alternative Communication*, *35*(4), 274–284. https://doi.org/10.1080/07434618.2019.1692902

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

Mokgoko, B. D. (2019). *Teaching Setswana as a home language subject in secondary schools in a linguistically diverse community in North West* [Master's thesis, University of Pretoria]. UPSpace. http://hdl.handle.net/2263/70424

Moonsamy, S., Mupawose, A., Seedat, J., Mophosho, M., & Pillay, D. (2017). Speech-language pathology and audiology in South Africa: Reflections on transformation in professional training and practice since the end of apartheid. *Perspectives of the ASHA Special Interest Groups*, *2*(17), 30–41.https//doi.org/10.1044/persp2.SIG17.30

Moorcroft, A., Scarinci, N., & Meyer, C. (2019). Speech pathologist perspectives on the acceptance versus rejection or abandonment of AAC systems for children with complex communication needs. *Augmentative and Alternative Communication*, *35*(3), 193-204. https://doi.otg/ 10.1080/07434618.2019.1609577

Mosime, D. K. (2014). *The management of potable water supply in Mogwase Township, Moses Kotane Local Municipality* [Masters dissertation, Potchefstroom Campus of the North-West University]. Retrieved from: https://repository.nwu.ac.za

Mothapo, N. R. B. (2019). *Determining the core vocabulary used by Sepedi- speaking preschool children during regular preschool-based activities* [Master's thesis, University of Pretoria]. UPSpace. http://hdl.handle.net/2263/71481

Mothapo, N. R. B., Tönsing, K. M., & Morwane, R. E. (2021). Determining the core vocabulary used by Sepedi-speaking children during regular preschool activities. *International Journal of Speech-Language Pathology*, 1–10. https://doi.org/10.1080/17549507.2020.1821774

Musselwhite, C.R & St. Louis. (1998). *Communication programming for persons with severe handicaps: Vocal and augmentative strategies*. Little, Brown.

Owen, A. J., & Leonard, L. B. (2002). Lexical diversity in the spontaneous speech of children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *45*, 927–937. https://doi.org/10.1044/1092-4388(2002/075)

Pascoe, M., & Norman, V. (2011). *Contextually relevant resources in speech-language therapy and audiology in South Africa-are there any? 58*(1), 2–5. https://doi.org/hdl.handle.net/11427/19940

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

Phuti, G., B. (Ed.). (2019). *Oxford Setswana-Seesimane Thanodi ya Sekolo/ School Dictionary* (1st ed.). Oxford University Press.

Pillay, M., Tiwari, R., Kathard, H., & Chikte, U. (2020). Sustainable workforce: South African audiologists and speech therapists. *Human Resources for Health*, *18*(1), 1-13. https://doi.org/10.1186/s12960-020-00488-6

Protection of Personal Information Act (POPIA). (2013). *Section 115 short title and commencement*. https://popia.co.za/section-115-short-title-and-commencement/

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

Richards, B. (1987). Type/token ratios: What do they really tell us?. *Journal of Child Language*, *14*(2), 201-209. https://doi.org/10.1017/S0305000900012885

Robillard, M., Mayer-Crittenden, C., Minor-Corriveau, M., & Bélanger, R. (2014). Monolingual and bilingual children with and without primary language impairment:

core vocabulary comparison. *Augmentative and alternative communication*, *30*(3), 267-278. https://doi.org/10.3109/07434618.2014.921240

Romski, M., Sevcik, R. A., Adamson, L. B., Cheslock, M., Smith, A., Barker, R. M., & Bakeman, R. (2010). Randomised comparison of augmented and nonaugmented language interventions for toddlers with developmental delays and their parents. *Journal of Speech, Language, and Hearing Research*, *53*(2), 350–364. https://doi.org/10.1044/1092-4388(2009/08-0156).

Sennott, S. C., Light, J. C., & McNaughton, D. (2016). AAC modeling intervention research review. *Research and Practice for Persons with Severe Disabilities*, *41*(2), 101-115. https://doi.org/10.1177/1540796916638822

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*, *22*(2), 63-75. https://doi.org/10.3233/EFI-2004-22201

Simonyi, C., & Brodie R. (1983). Microsoft Word. Word-processor software. Microsoft Corporation.

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

Slabbert, S., & Finlayson, R. (2000). " l'm a cleva!": the linguistic makeup of identity in a South African urban environment. *International Journal of the Sociology of Language*, *2000*(144), 119-136. https://doi.org/10.1515/ijsl.2000.144.119

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

Smith, T. C., & Witten, I. H. (1993). *Language inference from function words* (Computer Science Working Papers No. 93/3). Hamilton

Soto, G., & Cooper, B. (2021). An early Spanish vocabulary for children who use AAC: developmental and linguistic considerations. *Augmentative and Alternative Communication, 37*(1), 64-74. https://doi.org/10.1080/07434618.2021.1881822

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. https://doi.org/10.3109/07434618.2013.878751

Standard Operating Procedures for teachers, non-teaching staff and learners on the coronavirus outbreak in South Africa (2020). *Standard operating procedure for the prevention,*

containment and management of Covid-19 in schools and school communities. http://www.education.gov.za

Statistics South Africa (2011). *Census 2011: Census in brief.* http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_ in_brief.pdf

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

Sturm, J. M., & Clendon, S. A. (2004). Augmentative and alternative communication, language, and literacy: Fostering the relationship. *Topics in language disorders*, *24*(1), 76-91.

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

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). Whurr.

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

University of Wisconsin (2022). *About Setswana.* https://wisc.pb.unizin.org/lctlresources/chapter/about-setswana/

van Tilborg, A., & Deckers, S. R. J. (2016). Vocabulary selection in AAC: Application of core vocabulary in atypical populations. *Perspectives on Augmentative and Alternative Communication, 1,* 125-138. https://doi.org/10.1044/persp1.SIG12.125

Von Tetzchner, S., Martinsen, H. (2000). *Introduction to augmentative and alternative communication: Sign teaching and the use of communication aids for children, adolescents and adults with developmental disorders*. Whurr.

von Tetzchner, S., & Stadskleiv, K. (2016). *Constructing a language in alternative forms. In MM Smith & J. Murray (Eds.), The silent partner? Language, interaction and aided communication* (pp. 17–34). J&R Press.

Wachal, R. S. (1994). The dictionary as grammarian: Part-of-speech definitions and labels. *Dictionaries: Journal of the Dictionary Society of North America*, *15*(1), 159-170. https://doi.org/ 10.1353/dic.1994.0002

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), e448–e462. https://doi.org/10.1542/peds.2014-3889

Witkowski, D., & Baker, B. (2012). Addressing the content vocabulary with core: Theory and practice for nonliterate or emerging literate students. *Perspectives on Augmentative and Alternative Communication*, *21*(3), 74–81. https://doi.org/10.1044/aac21.3.74

Worldometer. (2020). *South African population (live).* https://www.worldometers.info/world-population/southafrica-population/

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

Zirker, K. A. H. (2007). *Intrasentential vs. intersentential code switching in early and late bilinguals* (Unpublished doctoral dissertation). Brigham Young University. Department of Linguistics and English Language.

APPENDIX A: APPROVAL BY RESEARCH ETHICS COMMITTEE OF THE FACULTY OF HUMANITIES

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

2 February 2022

Dear Miss MG Mogatusi

Project Title:	Determining the core vocabulary of Setswana-speaking Grade R learners as used during school activities
Researcher:	Miss MG Mogatusi
Supervisor(s):	Prof KM Tönsing
Department:	Centre for Augmentative and Alternative Communication
Reference number:	[redacted]
Degree:	Masters

I have pleasure in informing you that the above application was **approved** by the Research Ethics Committee on 27 January 2022. Data collection may therefore commence.

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 the 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 Karen Harris
Chair: Research Ethics Committee
Faculty of Humanities
UNIVERSITY OF PRETORIA
e-mail: tracey.andrew@up.ac.za

Room 7-27, Humanities Building, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa
Tel +27 (0)12 420 4853| Fax +27 (0)12 420 4501| Email pghumanities@up.ac.za | www.up.ac.za/faculty-of-humanities

APPENDIX B: DEPARTMENT OF EDUCAITON PERMISSION

education
Lefapha la Thuto la Bokone Bophirima
Noordwes Departement van Onderwys
North West Department of Education
NORTH WEST PROVINCE

Garona Building, Mmabatho
1st Floor, East Wing,
Private Bag X2044,
Mmabatho 2735
Tel.: (018) 388-3433
Fax.: 086-514-0126
e-mail: sgedu@nwpg.gov.za

OFFICE OF THE SUPERINTENDENT-GENERAL

Enq. : Dr T Phorabatho
Tel. : 018 388 3071/3433

To: Prof Kerstin Tonsing
University of Pretoria
Faculty of Humanities

From: Ms S M Semaswe
Superintendent-General

Date : 21 October 2021

PERMISSION TO CONDUCT RESEARCH: MS GAOPALELWE MOGATUSI

Permission is hereby granted to you to conduct research in the department as requested, subject to the following conditions:

- You contact the relevant School Principals for your target schools about your request with this letter of permission;
- Considering that your research will involve both Educators and Learners, the general functionality of the school should not be compromised by the research process.
- The participation in your project will be voluntary.
- The principles of informed consent and confidentiality will be observed in strictest terms, and
- The findings of your research should be made available to the North West Department of Education upon request.

Best wishes

Mrs S M Semaswe
Superintendent-General

APPENDIX C: PRINCIPAL INFORMATION LETTER AND PERMISSION FORM

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho
Centre for Augmentative and Alternative Communication

Humanities 100. Since 1919

Date:_____

Dear Principal

Re: Permission to conduct a research study at your school premises

My name is Gaopalelwe Mogatusi. I am a Master's student at the Center for Augmentative and Alternative Communication (AAC) at the University of Pretoria. As part of my studies, I am completing a small-scale research project. My research study intends to obtain and analyze the vocabulary of Setswana-speaking children in preschool. The title of my study is: "Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities".

I would like to request your permission to recruit participants for this study from your school and conduct the study at your school premises. I have obtained approval for the study from the North West Department of Education (DOE) (pleaes see attached).

Rationale
Children with speech impairments often struggle to speak due to congenital or acquired disabilities. When a child with a disability is unable to use natural forms of speech, they often benefit from using other forms of communication systems. For example, we may give them a computer or a board with pictures that represent words or messages to help them communicate. Choosing the appropriate words and messages for such systems is important to ensure that children can use them to communicate effectively in various situations. My research intends to study the words Setswana-speaking children use on a day-to-day basis. These words can then be used as a guideline when choosing words for a communication system for children with Setswana language backgrounds who cannot speak.

What will be expected of your school?
I will require teachers to assist me by nominating two (2) learners, preferably a boy and a girl, aged between 5 years to 6 years 11 months who could possibly participate in the study. The teachers will also be asked to complete a background questionnaire about the preschool as well as to send information letters and consent forms to the parents of the nominated learners and return to me after completion.

When parents/guardians consent for their children to participate in the study, and the children also agree, the children 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 will communicate with the teacher regarding fitting times of the audio recorder every day before class starts. The audio recorder and microphone will be fitted and removed every day.

Centre for Augmentative and Alternative Communication
Communication Pathology Building
Lynwood Road, Hatfield
University of Pretoria, Private Bag X20
Hatfield 0028, South Africa
Tel +27 (0)12 420 2001 | Fax +27 086 510 0841
Email: saak@up.ac.za | Web address: www.caac.up.ac.za

Data collection should take not more than five days. This data collection will not disturb the normal school day routine.

If learners experience any problems with the recorder, the teacher or researcher may help the learner to adjust the recorder or remove it should the learner wish to remove it. Teachers may also decide to switch off and/or remove the audio recorders for any activities during which they deem recording inappropriate or during which the device would hinder the learner's activities.

What will be expected of the participants during the study?
The participating learners will be expected to meet me with the teacher present. Picture-based explanations will be used to explain the process in an age-appropriate manner to the learner. I will then ask the learner if they are willing to participate in the study, and the learner will have an opportunity to 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.

The learners will be instructed not to play with or adjust the recording devices. They may ask for assistance from their teacher at any time should they feel uncomfortable or want to stop participating. The teacher will assist the learner with adjusting or removing the device. Learners may stop to participate 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 or legal guardians and assent from the participants themselves will be obtained prior to conducting the study. All participants and their parents/legal guardians will be made aware of their right to withdraw from the study at any point in time without any negative consequences to themselves.
- The speech sample recordings collected during the study will be assessed only by the researcher and her supervisor.
- All identifying information pertaining to individuals and the schools will be kept confidential from those external to the study. Any identifying information will be removed from the transcription and replaced by numerical codes (e.g. names of people and places will not be transcribed). No individual or school names will be mentioned in any published data.

Covid-19 protocols:
Covid-19 protocols in the prevention of the transmission of the SARS-COV-19 virus will be adhered to and abided by as per the World Health Organization (WHO) guidelines (2019) and the National Department of Basic Education Standard Operating Procedures for teachers and learners on the coronavirus outbreak in South Africa (2020). The following will be observed:

- Each child should continue wearing their face cloth masks at all times during data collection
- Hand hygiene and etiquette will be observed by the researcher at all times (i.e hand sanitizing and no shaking hands or hugging)
- Social distancing will be maintained with the children when orienting them to the study, requesting assent and on any contact days were the researcher is onsite
- Each child will be allocated one recording device which will be used solely on them throughout data collection to minimize potential for contamination

- Each recording device will be disinfected daily and will be kept in a sealable ziplock bag to prevent it from contact with other contaminated surfaces

Who will have access to the results of the study?
The research study data will be secured and safely stored for 15 years in both hard copy and electronic format at the Center for AAC in University of Pretoria. Should parents give permission, the voice recordings may be made available to other researchers who request them. However, these researchers will also keep the recordings confidential and only use them for research purposes. The transcription of the audio recording with all personal data removed will be made available on the website of the South African Digital Language Resource Centre (https://repo.sadilar.org/). The data obtained from the research will be used for writing a research report (a Master's mini- dissertation), writing scientific papers as well as for educational and research purposes. The results will not be linked to specific participants. The school and participants' identities, as well as personal information, will not be disclosed. A summary of the results will be made available for any interested staff members or parents.

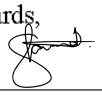
What are the risks and the benefits?
The learners and teachers at your school will not be put at risk during the study. The learners will not miss out on any of their school activities through participating in this research. The audio-recorders will be fitted to participants within padded body-word pouches to minimize discomfort. Participation is voluntary at all times and participants may withdraw at any time without negative consequences. The participating learners may request to have the equipment removed at any time. Teachers may also stop audio recordings and/or remove the audio recording equipment from the child at any time at their own discretion.

Potential benefits of this study 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 use Setswana as their language of communication.

I would appreciate it if you could complete the attached form to indicate whether you give permission to include participants at your school in the study. For any further information, please contact me or my supervisor using the contact details below.

I look forward to your response.

Regards,



Gaopalelwe Mogatusi
Cell: [redacted]
Email: gmogatusi@gmail.com

Date

Professor Kerstin Tonsing
Centre for Augmentative and Alternative Communication (CAAC)
Cell: [redacted]

Date

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Centre for Augmentative and
Alternative Communication

Humanities 100. Since 1919

Principal permission form

Principals name: _____

School's name: _____

Title of study: **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**

Researcher: Gaopalelwe Mogatusi
Master's Student
Centre for AAC
University of Pretoria
Cell: [redacted]
gmogatusi@gmail.com

Supervisor: Kerstin Tönsing
Associate Professor
Centre for AAC
University of Pretoria
Cell: [redacted]
kerstin.tonsing@up.ac.za

I, _____ (Name and surname)

(please tick box that applies)

☐ give permission to the researcher to recruit learners from the school named above for possible participation in the study entitled: **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**, conducted by Gaopalelwe Mogatusi, 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 at the school. I understand that the data will be stored for 15 years at the CAAC and that all data will be treated confidentially. I understand that the data may be used for a scientific article, research reports or presentations. I understand that the transcription of the audio recording with all personal data removed will be made available on the website of the South African Digital Language Resource Centre (https://repo.sadilar.org/) and may be used for further analysis. I understand that, should parents give permission, the audio recordings will be made available to other researchers for research purposes.

OR

☐ do not give permission to Gaopalelwe Mogatusi to recruit learners from the preschool named above for possible participation in the study entitled **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities.**

_____
Principal Signature

_____
Date

School stamp

Centre for Augmentative and Alternative Communication
Communication Pathology Building
Lynwood Road, Hatfield
University of Pretoria, Private Bag X20
Hatfield 0028, South Africa
Tel +27 (0)12 420 2001 | Fax +27 086 510 0841
Email: saak@up.ac.za | Web address: www.caac.up.ac.za

Setswana version:
LEKWALO LA MOGOKGO LE KOPO YA TETLELELO

Go Mogokgo **Letlha:_____**

Re: Kopo ya go letlelelwa go dira patlisiso mo sekolong

Leina la me ke Gaopalelwe Mogatusi. Ke moithuti wa Master's kwa lefapheng la Center for Augmentative and Alternative Communication (AAC) kwa Yunibesithi ya Pretoria. Moano mongwe wa dithuto tsa me ke go dira porojeke e nyenyane ya patlisiso. Maikaelelo a patlisiso ya me ke go utlwelela puo ya Setswana mo baneng ba sekolo fa ba dira ditiro tsa sekolo. Leina la patlisiso ya me ke: "Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities".

Ka jalo ke kopa go letlelelwa go dirisa sekolo seno go dira patlisio ya me le go thusiwa go tlhopha bana mo sekolong ba ba ka tsayang karolo mo patlisisong ya me. Ke dumeletswe go dira patlisiso ena ke ba lefapha la thuto – North West Department of Education (DoE) (bona matlhare a a kgomagantsweng).

Botlhokwa ba patlisiso

Bana ba ba sa kgoneng go bua gantsi ya be ele ka ntlha ya bogole ba matsalo kgotsa bogole bo bo itlhagetseng. Fa ngwana a na le bogole ba puo, gantsi o sokola go buisana le batho ba bangwe mme ebile o tlhoka mekgwa e farologaneng ya go bua gore a kgone go buisana le batho. Sekai, ngwana yo o sa kgoneng go bua a ka fiwa khomputara kgotsa boto e e nang le ditshwantsho tse di emelelang mafoko le dipolelo gore a tle a e dirise go bua le batho. Go tlhopha mafoko kgotsa dipolelo tse di tla beiwang mo botong kgotsa khomputara ke tiro e e botlhokwa tota gonne re tshwanetse go netefatsa gore a dirisiwa ka tsela e e ba tswelang mosola. Patlisiso ya me ke go batla mafoko a Setswana a a dirisiwang ke bana ba mo sekolong letsatsi le letsatsi. Mafoko a, a ka kgona go dirisiwa fa ngwana wa segole sa puo a direlelwa boto ya puo ya Setswana.

Ke eng se sekolo se tlhokang go se dira?
Ke tlile go kopa gore barutabana ba nthuse go tlhopha bana ba le babedi mo sekolong, mosimane le mosetsana ba ka nna dingwaga di le 5 to 6 years 11 months ba ba ka tsayang karolo mo patlisisong ena. Barutabana ba tlile go kopiwa go tlatsa letlhare la dipotso tsa sekolo le go romela makwalo a kopo go batsadi ba bana ba ba tlhopilweng mme morago ba busetse matlhare ao kwa go nna.

Ga batsadi/ batlhokomedi ba bana ba dumela gore bana ba tseye karolo mo patlisisong, le bana bao ba dumela, ba tla neelwa rekhota e e tla thusang mmatlisisi go rekhota mafoko a bana ba dirisang fa sekolong.

Centre for Augmentative and Alternative Communication
Communication Pathology Building
Lynwood Road, Hatfield
University of Pretoria, Private Bag X20
Hatfield 0028, South Africa
Tel +27 (0)12 420 2001 | Fax +27 086 510 0841
Email: saak@up.ac.za | Web address: www.caac.up.ac.za

Mmatlisisi o tla buisana le morutabana ka ga nako ya go apesa bana mechini ya audio rekhota letsatsi le letsatsi pele sekolo se tsena. Mmatlisisi o tla apesa bana direkhota mme gape a ba apole mechini ya go rekhota morago ga letsatsi la sekolo. Patlisiso e e tlile go tsaya lobaka lo ka nna malatsi a le matlhano mme ga nkitla e kgoreletsa bana mo tsamaisong ya bona ya dithuto jaaka ka metlha.

Fa bana ba kgorelediwa ke rekhota, morutabana kgotsa mmatlisisi ba tla thusa ngwana go e baakanya kgotsa go e ntsha.
Barutabana ba letlelelwa go tima le go tshuba rekhota ga ba dira tiro e ba sa batleng go utlwiwa mo go yona kgotsa ga rekhota e kgoreletsa ngwana go dira tiro ya sekolo.

Ke eng se se solofetsweng mo bananeng ba ba tsayang karolo mo patlisisong?
Bana ba ba tsayang karolo ba tlile go kopana le mmatlisisi le morutabana wa bona. Ba tlile go tlhalosetswa tsamaiso ya patlisiso sentle ka ditshwantsho le ka puo e e ba tshwanetseng. Morago ba tla kopiwa go tlhopa gore a tota ba dumela go tsaya karolo, ngwana mongwe le mongwe o tla neelwa chono go araba ka disthwantsho le ka puo.

Ga ngwana a dumetse, o tla kopiwa go apara sekgwama sa letheka se se nang le rekhota e e tsamaisanang le maekrofunu. Maekrofunu oo o tla ngaparetswa mo molaleng wa sekhipa gore o kgone go rekhota mafoko a ngwana a a buang.

Bana ba tla laelwa gore ba seke ba tshameka ka mechini ya go rekhota kgotsa ba kgotlha-kgotlha di rekhoto. Bana ba tla dumelelwa go kopa thuso mo go morutabana nako engwe le engwe ga ba kgoreletsiwa ke mochini wa go rekhota. Morutabana o tla thusa ngwana go baakanya mochini wa go rekhota kgotsa go e ntsha gotlhelele. Bana ba letlelelwa go emisa go tsaya karolo nako nngwe le nngwe go sena ditlamorago.

Melao e e latelang e obametswe mo patlisisong ena:

- Tumelano e e kwadilweng go tswa go batsaya-karolo kgotsa batsadi le batlhokomedi ba bana e tla amogelwa pele ga tshimologo ya patlisiso. Batsaya karolo le batsadi/batlhokomedi ba tla sedimosetswa ka ditshwanelo tsa bona mo patlisisong le gore ba ka ikgogela morago nako ngwe le ngwe go sena ditlamorago dipe.
- Direkhoto tsa puo tse di tla tseiwang mo patlisisong di tlile go atlha-atlhiwa ke mmatlisisi le mmatlisisi-mogolo fela.
- Tshedimosetso le maina a batsaya karolo di tla sireletswa, mme tsa tsholwa jaaka khupamarama. Ka jalo, maina a batsaya karolo a tla emelelwa ke dinomoro eseng maina a bona (sekao, maina a batho le maina a mafelo a ka seke a dirisiwa). Ga gona leina la motho kgotsa leina la sekolo le le tla phasalatsiwang mo mafaratlhatlheng.

Melawana ya Covid-19
Melao ya go fokotsa tshwaetsego ya mogare wa SARS-COV-19 e tla obamelwa jaaka e kaetswe ke World Health Organization (WHO) (2019) le ba National Department of Basic Education (DBE) (2019) go fokotsa mogare wa covid-19 mo baneng kwa dikolong. Melao e e tla obameliwang mo patlisisong ke:

- Ngwana mongwe le mongwe o tla tsweletsa go rwala maseke wa bona ka dinako tsotlhe tsa patlisiso.
- Tlhokomelo ya diatla le go fokotsa tshwarano ka diatla (jaaka, go tlhapa matsogo le go sa atalne kampo go tshwarana ka matsogo).
- Go ema sekgele se le sengwe magareng fa bana ba rutiwa ka patlisiso ena le fa ba kopiwa tetla ya go tsaya karolo kampo fa mmatlisisi a le teng fa sekolong.
- Ngwana mongwe le mongwe o tla fiwa mochini wa go rekhota mme a o dirisa a le mongwe mo tsamaisong ya patlisiso go fokotsa tshwaetsego.
- Mochini mongwe le mongwe wa go rekhota o tla phepafadiwa letsatsi lengwe le lengwe mme ebile o tla tsengwa mo sekgwameng gore o seke wa kopana le dilo tse di na leng mogare wa Covid-19.

Ke mang a tla nnang le tetla ya patlisiso?
Direkhoto tsa patlisiso di tla bolokiwa ka pabalesego dingwaga di 15 kwa lefapheng la Center for AAC kwa Yunibesithi ya Pretoria. Fa batsadi ba neela tetla, direkhoto tsa puo di ka dirisiwa ke babatlisisi ba bangwe ba ba kopileng. Ka jalo, babatlisisi bao ba tla boloka direkhoto sentle mme ba di dirisetsa fela dithuto kgotsa dipatlisiso tse dingwe. Direkhoto fela eseng maina a batho, ke tsone di tla tlhagisiwang mo mafaratlhatlheng a South African Digital Language Resource Centre (https://repo.sadilar.org/). Direkhoto tse di tla tseiwang mo patlisisong e, di tlile go dirisetswa go kwala lekwalo la Master's (Master's mini-dissertation), go kwala makwalo a saense, a tlileng go phatlhalatswa gape le go dirisetswa thuto le dipatlisiso tse dingwe. Maina a sekolo le mogokgo a tla nna sephiri ebile ga nkitla a phatlhalatswa.

Tshobokanyo ya patlisiso e tla nna teng go badiwa ke barutabana kgotsa babereki ba sekolo ba ba nang le kgatlhegelo ya se se bonweng mo patlisisong.

Ke eng dikotsi le dipoelo?
Bana ba sekolo le barutabana ga nkitla ba beiwa mo botshosetsing bope. Baithuti ga ba kitla ba kgoreletswa go tsaya karolo mo tsa dithutong. Barutabana ga ba kitla ba kgoreletswa go dira tiro ya bone jaaka ka gale.
Mechini ya go rekhota e tla apesiwa batsaya karolo ka tsela e bolokegileng. Go tsaya karolo go tla ka boithaopi ka dinako tsotlhe ebile batsaya karolo ba ka ikgogela morago nako nngwe le nngwe go se na ditlamorago. Batsaya karolo ba ka kopa go ntshiwa mochini wa go rekhota nako nngwe le nngwe. Barutabana ba ka emisa go rekhota/ go ntsha mochini wa go rekhota mo ngwaneng ga ngwana a kopile jalo.

Dipoelo tsa patlisiso e, ke go godisa patlisiso ka bophara mo lefapheng la AAC le go neela tshedimosetso gore ke mafoko a a feng a a tlhokegang go direla ngwana wa segole sa puo thekenoloji ya Setswana ya AAC gonne Setswana e le puo ya gagwe ya gae.

Nka itumelela gore o tlatse foromo e e romeletsweng (bona ka fa morago) go kaela fa o mpha tetla ya go tsaya baithuti ba sekolo jaaka batsaya karolo ya patlisiso ye. Fa o na le dipotso dingwe, ka kopo leletsa nna kgotsa mmatlisisi mogolo wa me ka go dirisa dinomoro tse di tlhagisistsweng fa.

Ke solofetse go utlwa tsibogo ya gago.

Wa lona,

Gaopalelwe Mogatusi
Mmatlisisi
Mogala: [redacted]
Email: gmogatusi@gmail.com

Letlha

Professor Kerstin Tonsing
Centre for Augmentative and Alternative Communication (CAAC)
Mmatlisisi-mogolo
Mogala: [redacted]
Email: kerstin.tonsing@up.ac.za

Letlha

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Centre for Augmentative and
Alternative Communication

Humanities 100. Since 1919

Tetla ya Mogokgo

Leina la mogokgo: _____

Leina la sekolo: _____

Leina la patlisiso: **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**

Mmatlisisi: Gaopalelwe Mogatusi
Moithuti wa Master's
Lefapha la AAC
Yunibesithi ya Pretoria
Mogala: [redacted]
gmogatusi@gmail.com

Mmatlisisi mogolo: Kerstin Tönsing
Associate Professor
Lefapha la AAC
Yunibesithi ya Pretoria
Mogala: [redacted]
kerstin.tonsing@up.ac.za

Nna, _____ (Leina le sefane)

(tshwaya mo lebokoswaneng)

☐ ke fa tetla go mmatlisisi go tlhopha baithuthi mo sekolong gore ba tseye karolo mo patlisisong e e bidiwang: **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**, e e tsamaisiwang ke Gaopalelwe Mogatusi, ka fa tlase ga pono le botsamaisi bogolo ba ga Kerstin Tonsing. Tetla e, e diriwa jaana ka boithaopo ebile ke tlhaloganya gore re ka gogela morago ka nako engwe le engwe. Ke tlhaloganya gore bana ba ba tla tsayang karolo ba tlile go rekhotiwa ka mechini fa sekolong. Ke tlhaloganya gore direkoto tsa patlisiso ye di tla bolokiwa ka pabalesego dingwaga di le 15 kwa lefapheng la CAAC le gore direkoto di tla tshwariwa ka khupamarama. Ke tlhaloganya gore direkhoto tse di ka dirisetswa makwalo a saense, makwalo magolo le ditlhalosetso. Ke tlhaloganya gore maina a bana le maina a sekolo a tla phomoliwa gotlhelele fa go phatlhalatswa patlisiso ye mo mafaratlhatlheng a South African Digital Language Resource Centre (https://repo.sadilar.org/) le gore patlisiso ye e ka dirisetswa dipatlisiso tse dingwe. Ke tlhaloganya gore fa batsadi ba neela tetla, direkhoto tsa puo di ka kgona go dirisiwa ke babatlisisi ba bangwe go etleletsa dipatlisiso kgotsa thuto.

KGOTSA

☐ ga ke fe Gaopalelwe Mogatusi tetla ya go tlhopha baithuthi mo sekolong go tsaya karolo mo patlisisong e e kaetsweng e e bidiwang: **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**

_____
Tshaeno ya mogokgo

_____
Letlha

Setempe sa sekolo

APPENDIX D: TEACHER INFORMATION LETTER AND PERMISSION FORM

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho
Centre for Augmentative and Alternative Communication

Humanities 100. Since 1919

Dear Class Teacher Date: _____

Re: Consent to nominate children and for recordings to take place in classroom

My name is Gaopalelwe Mogatusi. I am a Master's student at the Center for Augmentative and Alternative Communication (AAC) at the University of Pretoria. As part of my studies, I am completing a small-scale research project. My research study intends to obtain and analyze the vocabulary of Setswana-speaking children in preschool. The title of my study is: "Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities".

I would like to request your permission and assistance in selecting learners from your class for possible participation in my study and to allow for recording to take place in your classroom. I have obtained approval from the school principal and the Department of Education (DOE) (please see attached).

Rationale/ Reason for study
Children with speech impairments often struggle to speak due to congenital or acquired disabilities. When a child with a disability is unable to use natural forms of speech, they often benefit from using other forms of communication systems. For example, we may give them a computer or a board with pictures that represent words or messages to help them communicate. Choosing the appropriate words and messages for such systems is important to ensure that children can use them to communicate effectively in various situations. My research intends to study the words Setswana-speaking children use on a day-to-day basis. These words can then be used as a guideline when choosing words for a communication system for children with Setswana language backgrounds who cannot speak.

What will be expected of you?
I will require you to assist me by selecting two (2) learners, preferably a boy and a girl, aged between 5 years to 6 years 11 months who could possibly participate in the study. You will be requested to send information letters and consent forms to the parents of the selected learners and return to me after completion. I will also ask you to kindly complete a background questionnaire about the preschool.

When parents/guardians consent/agree for their children to participate in the study, and the children also agree, each child 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 will communicate with you regarding fitting times of the audio recorder every day before class starts. The audio recorder and microphone will be fitted and removed every day. Data collection should take not more than five days. This data collection will not disturb the normal school day routine.

Centre for Augmentative and Alternative Communication
Communication Pathology Building
Lynwood Road, Hatfield
University of Pretoria, Private Bag X20
Hatfield 0028, South Africa
Tel +27 (0)12 420 2001 | Fax +27 086 510 0841
Email: saak@up.ac.za | Web address: www.caac.up.ac.za

If learners experience any problems with the recorder, you may help the learner to adjust the recorder or remove it should the learner wish to remove it. You may also decide to switch off and/or remove the audio recorders for any activities during which you deem recording inappropriate or during which the device would hinder the learner's activities.

What will be expected of the participants during the study?
The participating learners will be expected to meet me with the teacher present. Picture-based explanations will be used to explain the process in an age-appropriate manner to the learner. I will then ask the learner if they are willing to participate in the study, and the learner will have an opportunity to 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.

The learners will be instructed not to play with or adjust the recording devices. They may ask for assistance from their teacher at any time should they feel uncomfortable or want to stop participating. The teacher will assist the learner with adjusting or removing the device. Learners may stop to participate 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 or legal guardians and assent from the participants themselves will be obtained prior to conducting the study. All participants and their parents/legal guardians will be made aware of their right to withdraw from the study at any point in time without any negative consequences to themselves.
- The speech sample recordings collected during the study will be assessed only by the researcher and her supervisor.
- All identifying information pertaining to individuals and the schools will be kept confidential from those external to the study. Any identifying information will be removed from the transcription and replaced by numerical codes (e.g. names of people and places will not be transcribed). No individual or school names will be mentioned in any published data.

COVID-19 PROTOCOLS:
Covid-19 protocols in the prevention of the transmission of the SARS-COV-19 virus will be adhered to and abided by as per the World Health Organization (WHO) guidelines (2019) and the National Department of Basic Education Standard Operating Procedures for teachers and learners on the coronavirus outbreak in South Africa (2020). The following will be observed:

- Each child should continue wearing their face cloth masks at all times during data collection
- Hand hygiene and etiquette will be observed by the researcher at all times (i.e., hand sanitizing and no shaking hands or hugging)
- Social distancing will be maintained with the children when orienting them to the study, requesting assent and on any contact days were the researcher is onsite
- Each child will be allocated one recording device which will be used solely on them throughout data collection to minimize potential for contamination
- Each recording device will be disinfected daily and will be kept in a sealable zip lock bag to prevent it from contact with other contaminated surfaces.

Who will have access to the results of the study?
The research study data will be secured and safely stored for 15 years in both hard copy and electronic format at the Center for AAC in University of Pretoria. Should parents give permission, the voice recordings may be made available to other researchers who request them. However, these researchers will also keep the recordings confidential and only use them for research purposes. The transcription of the audio recording with all personal data removed will be made available on the website of the South African Digital Language Resource Centre (https://repo.sadilar.org/). The data obtained from the research will be used for writing a

research report (a Master's mini- dissertation), writing scientific papers as well as for educational and research purposes. The results will not be linked to specific participants. The school and participants' identities, as well as personal information, will not be disclosed. A summary of the results will be made available for any interested staff members or parents.

What are the risks and the benefits?
Neither the learners nor you will be put at risk during the study. The learners will not miss out on any of their school activities through participating in this research. The audio-recorders will be fitted to participants within padded body-word pouches to minimize discomfort. Participation is voluntary at all times and participants may withdraw at any time without negative consequences. The participating learners may request to have the equipment removed at any time. You may also stop audio recordings and/or remove the audio recording equipment from the child at any time at your own discretion.

Potential benefits of this study 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 use Setswana as their language of communication.

I would appreciate it if you could complete the attached form to indicate whether you give permission to include participants in your class in the study. For any further information, please contact me or my supervisor using the contact details below.

I look forward to your response.

Regards,

Gaopalelwe Mogatusi
Cell: [redacted]
Email: gmogatusi@gmail.com

Date

Professor Kerstin Tonsing
Centre for Augmentative and Alternative Communication (CAAC)
Cell: [redacted]
Email: kerstin.tonsing@up.ac.za

Date

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Centre for Augmentative and
Alternative Communication

Humanities 100. Since 1919

Teacher consent form

Teacher's name: _____

School's name: _____

Title of study: **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**

Researcher: Gaopalelwe Mogatusi
Master's Student
Centre for AAC
University of Pretoria
Cell: [redacted]
gmogatusi@gmail.com

Supervisor: Kerstin Tönsing
Associate Professor
Centre for AAC
University of Pretoria
Cell: [redacted]
kerstin.tonsing@up.ac.za

I, _____ (Name and surname)

(please tick box that applies)

☐ give consent to participate and assist in the study entitled: **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**, conducted by Gaopalelwe Mogatusi, 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 in the classroom. I understand that the data will be stored for 15 years at the CAAC and that all data will be treated confidentially. I understand that the data may be used for a scientific article, research reports or presentations. I understand that the transcription of the audio recording with all personal data removed will be made available on the website of the South African Digital Language Resource Centre (https://repo.sadilar.org/) and may be used for further analysis. I understand that, should parents give permission, the audio recordings will be made available to other researchers for research purposes.

OR

☐ do not give consent to participate in the study entitled **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities.**

_____
Teachers Signature

_____
Date

APPENDIX E: CAREGIVER INFORMATION LETTER AND CONSENT FORM

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Centre for Augmentative and Alternative Communication

Humanities 100. Since 1919

Date: _____

Dear Sir/Madam

Re: Request to allow for your child's participation in a research study

My name is Gaopalelwe Mogatusi. I am a Master's student at the Center for Augmentative and Alternative Communication (AAC) at the University of Pretoria. As part of my studies, I am completing a small-scale research project. My research study intends to obtain and analyze the vocabulary of Setswana-speaking children in preschool. The title of my study is: "**Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**".

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

Why is this study important?
Children with speech impairments often struggle to speak due to congenital or acquired disabilities. When a child with a disability is unable to use natural forms of speech, they often benefit from using other forms of communication systems. For example, we may give them a computer or a board with pictures that represent words or messages to help them communicate.

Choosing the appropriate words for such systems is important to ensure that children can use them to communicate effectively in various situations. My research intends to study the words Setswana-speaking children use on day to day basis. These words can be used as a guideline when choosing words for a communication aid (a system with words that one can point at to express themselves) for children with Setswana language backgrounds who cannot speak.

What will be expected of your child?
Should you give consent for your child to participate in the study, the following will be expected from him or her:

- To meet with me and the teacher present where picture-based explanations will be used to explain the process in an age-appropriate manner to the child. I will then ask the child if they are willing to participate in the study, and the child will have an opportunity to respond using pictures as well as verbally.
- 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 microphone clipped to his/her shirt/top during normal school activities for about 3-5 days. The recorder will be switched on to record your child's speech.
- Your child will be asked not to play with or to adjust any of the equipment. He/she may ask for assistance at any time from his/her teacher if he/she experiences any discomfort or if he/she does not want to participate any longer. The teacher will assist your child with removing or adjusting the device. The teacher may also switch off or remove the recording equipment at any time when she feels that recording is not appropriate or when the equipment could interfere with your child's activities.
- Your child may choose to withdraw from the study at any time without any negative consequences.

Centre for Augmentative and Alternative Communication
Communication Pathology Building
Lynwood Road, Hatfield
University of Pretoria, Private Bag X20
Hatfield 0028, South Africa
Tel +27 (0)12 420 2001 | Fax +27 086 510 0841
Email: saak@up.ac.za | Web address: www.caac.up.ac.za

What are my child's rights?
You and your child can choose to participate or not to participate in this study. You or your child may stop participating in the study at any given moment if you wish to. Nothing bad will happen to you or your child if you choose not to participate or to stop participating. It will be ensured that your child understands that he/she can ask the teacher to take off the recording equipment at any time. If you or your child decide to stop taking part, all recordings of your child will be destroyed.

Your child's name, as well as any other personal information, will only be available to the researcher for administrative purposes. All the personal 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 and my supervisors. Any personal information such as names of people and places will be removed when I write down the words that your child used. The written form of what your child said (with all personal information removed) will be made available on the website of the South African Digital Language Resource Centre (https://repo.sadilar.org/). The reason for this is that other researchers may be able to use this information to understand children's language development. However, no personal information will be shared in the process. When I speak, or write about the study, no personal information about your child, yourself or the school will be shared.

What will happen after I collect the information?
The personal 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 voice recordings will only be listened to by me and my supervisors. Any personal information such as names of people and places will be removed when I write down the words that your child used. If you give additional permission, the voice recordings may be made available to other researchers who request them. However, they will also keep the recordings confidential and only use them for research purposes. However, it is your choice whether you want to grant this additional permission or not.

The written form of what your child said (with all personal information removed) will be made available on the website of the South African Digital Language Resource Centre (https://repo.sadilar.org/). The reason for this is that other researchers may be able to use this information to understand children's language development. However, no personal information will be shared in the process.

The information collected will be used for writing a Master's mini-dissertation, writing scientific papers and for presentation at professional conferences and seminars. However, 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.

Covid-19 protocols:
Covid-19 protocols in the prevention of the transmission of the SARS-COV-19 virus will be adhered to and abided by as per the World Health Organization (WHO) guidelines (2019) and the National Department of Basic Education Standard Operating Procedures for teachers and learners on the coronavirus outbreak in South Africa (2020). The following will be observed:

- Each child should continue wearing their face cloth masks at all times during data collection.
- Hand hygiene and etiquette will be observed by the researcher at all times (i.e., hand sanitizing and no shaking hands or hugging).
- Social distancing will be maintained with the children when orienting them to the study, requesting assent and on any contact days were the researcher is onsite.
- Each child will be allocated one recording device which will be used solely on them throughout data collection to minimize potential for contamination.
- Each recording device will be disinfected daily and will be kept in a sealable ziplock bag to prevent it from contact with other contaminated surfaces.

What are the risks and the benefits of participating in this study?
Your child will not be disadvantaged or harmed in any way during this study. He/she will also not be absent from any school activities. 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, or if he/she wants to stop participating

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

I would appreciate if you would complete the attached consent 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.

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

Gaopalelwe Mogatusi
Cell: [redacted]
Email: gmogatusi@gmail.com

Date

Professor Kerstin Tonsing
Centre for Augmentative and Alternative Communication
(CAAC)
Cell: [redacted]
kerstin.tonsing@up.ac.za

Date

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Centre for Augmentative and Alternative Communication

Humanities 100. Since 1919

Parent Informed Consent

Name of Child: _____

Name of Parent/Caregiver: _____

Project title: **Determining the core vocabulary of Setswana-speaking Grade R as used during preschool activities**

Researcher: Gaopalelwe Mogatusi
Master's Student
Centre for AAC
University of Pretoria
Cell: [redacted]
gmogatusi@gmail.com

Supervisor: Kerstin Tönsing
Associate Professor
Centre for AAC
University of Pretoria
Cell: [redacted]
kerstin.tonsing@up.ac.za

I, _____ (Name and surname)

(Please tick box that applies)

☐ give consent for my child to participate in the study entitled ***Determining the core vocabulary of Setswana-speaking Grade R as used during preschool activities***, conducted by Gaopalelwe Mogatusi, 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 my child will be audio-taped for data collection purposes. I understand that the data may be used for a scientific article, research reports or presentations. I understand that the written form of what my child said (with all personal information removed) will be made available on the website of the South African Digital Language Resource Centre (https://repo.sadilar.org/), in order to be reused for further analysis. I understand that all personal 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 of Setswana-speaking Grade R as used during preschool**.

Additionally, I

☐ *give permission for the recordings of my child's speech to be made available to other researchers. I understand that recordings will be kept confidential by these researchers and only used for research purposes.*

OR

☐ *do not give permission for the recordings of my child's speech to be made available to other researchers.*

_____ _____
Parent's Signature Date

Setswana version:

LEKWALO LA TSHEDIMOSETSO LE TETLELELO YA MOTSADI/ MOTLHOKOMEDI WA NGWANA

Go Rre/Mme Letlha: _____

Re: Kopo ya go letlelela ngwana wa gago go tsaya karolo mo patlisisong

Leina la me ke Gaopalelwe Mogatusi. Ke moithuti wa Master's kwa lefapheng la Center for Augmentative and Alternative Communication (AAC) kwa Yunibesithi ya Pretoria. Moano mongwe wa dithuto tsa me ke go dira porojeke e nyenyane ya patlisiso. Maikaelelo a patlisiso ya me ke go utlwelela puo ya Setswana mo baneng ba sekolo fa ba dira ditiro tsa sekolo. Patlisiso ya dithuto tsa me ke: "Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities".

Ke kopa tetla ya gore ngwana wa gago a tseye karolo mo patlisisong ena

Goreng patlisiso e e le botlhokwa?

Bana ba e leng digole mme ebile ba sa kgone go bua, gantsi ba sokola go buisana le batho ba bangwe. Bana ba, ba tlhoka mekgwa e farologaneng ya go buisana le batho, sekai, ngwana yo o sa kgoneng go bua a ka fiwa khomputara kgotsa boto e e nang le ditshwantsho tse di emelelang mafoko le dipolelo gore a tle a e dirise go bua le batho. Go tlhopha mafoko kgotsa dipolelo tse di tla beiwang mo botong kgotsa khomputara ke tiro e e botlhokwa tota gonne re tshwanetse go netefatsa gore a dirisiwa ka tsela e e ba tswelang mosola. Patlisiso ya me ke go batla mafoko a Setswana a a dirisiwang ke bana ba mo sekolong letsatsi le letsatsi. Mafoko a, a ka kgona go dirisiwa fa ngwana wa segole sa puo a direlelwa boto ya puo ya Setswana

Go solofetswe eng mo ngwaneng wa gago?

Fa o ka letla ngwana wa gago go tsaya karolo mo patlisisong eno, se se solofelwang mo go ena ke se se latelang:

- Ngwana o tlile go kopana le nna le morutabana wa gage. Ke tlile go tlhalosetsa ngwana ka ga tsamaiso ya patlisiso eno ke dirisa ditshwantsho le puo e e tlhofofaditsweng e e tshwanetseng bana. Morago ke tla kopa ngwana go kaela fa a eletsa go ka tsaya karolo mo patlisisong. Ngwana o tla nna le moetlo wa kaela keletso ya gagwe a dirisa ditshwantsho le puo.
- Fa ngwana a dumetse go tsaya karolo, o tla kopiwa go apara kgetsana ya letheka e e nang le rekhota e e tsamaisanang le maekrofunu. Maekrofunu o tla ngaparetswa mo molaleng wa sekhipa sa ngwana. Se se tla diriwa phakela mongwe le mongwe go ka nna malatsi a ka nna 3-5 gore mochini o rekhote mafoko a ngwana a a buang kwa sekolong.
- Ngwana o tla laelwa gore a seke a tshameka ka mochini wa go rekhota kgotsa ao kgotlha-kgotlha. Ngwana o dumelelwa go kopa thuso mo go morutabana nako nngwe le nngwe fa a kgoreletsiwa ke mochini wa go rekhota. Morutabana o tla thusa ngwana go baakanya rekhota kgotsa go o ntsha gotlhelele.
- Ngwana o letlelelwa go emisa go tsaya karolo ka nako engwe le engwe mme go sena ditlamorago.

Centre for Augmentative and Alternative Communication
Communication Pathology Building
Lynwood Road, Hatfield
University of Pretoria, Private Bag X20
Hatfield 0028, South Africa
Tel +27 (0)12 420 2001 | Fax +27 086 510 0841
Email: saak@up.ac.za | Web address: www.caac.up.ac.za

Ke eng ditshwanelo tsa ngwanake?

Motsadi le ngwana ba na le tshwanelo ya go tlhopha go tsaya karolo kgotsa go tlhopha go sa tseye karolo mo patlisisong. Ba ka ikgogela morago nako engwe le engwe fa ba rata. Ga gona ditlamorago fa motsadi le ngwana ba ikgogela morago. Go tlile go netefatsiwa gore ngwana o tlhaloganya gore a ka tsibosa morutabana fa mochini wa go rekhota o mo kgoreletsa. Fa motsadi kgotsa ngwana ba ikgogela morago, direkhoto tsa ngwana di tla latlhiwa.

Leina la ngwana wa gago le direkhoto tse dingwe tsa gagwe di tla dirisetswa fela mabaka a patlisiso eseng sepe gape. Direkhoto tsotlhe di tlile go bolokiwa ka pabalesego mme tsa seke tsa fiwa ope gape. Rekhoto ya puo ya ngwana e tlile go utlwelelwa ke mmatlisisi le mmatlisisi mogolo fela. Maina a batho ba bangwe a ngwana a ka a buang mo puong ya gagwe mo rekhoteng a tla phomoliwa fa re kwala puo yotlhe ya ngwana. Rekhoto ya puo ya ngwana e ka fitlhelelwa mo mafaratlhatlheng a South African Digital Language Resource Centre (https://repo.sadilar.org/). Mosola wa se ke gore babatlisisi ba bangwe le bone ba kgone go bona le go ithuta ka puo ya bana. Ga gona maina a bana bape ba ba tsayang karolo mo patlisisong e, a tlileng go tlhagisiwang gope fa re kwala kgotsa re bua ka patlisisio ena.

Go tla diregang morago ga dipatlisisio?

Dikgang tse o tla re bolelelang tsona kaga ngwana mmogo le direkhoto tsa ngwana di tla bolokega ka pabalesego dingwaga di ka nna 15 kwa lefapheng la Center for Augmentative and Alternative Communication.

Puo ya di rekhoto e tlile go utlwelelwa ke mmatlisisi le mmatlisisi mogolo fela eseng gope gape. Dikgang tsa maina a ngwana le tse dingwe tse di supisang ga nkitla di sedimosetswa fa go kwadiwa puo eo ngwana a e buileng. Fa o neela tetla tlaleletso, direkhoto tsa puo di tla dirisiwa ke babatlisisi ba bangwe b aba kopileng. Ka jalo, babatlisisi bao ba tla boloka direkhoto sentle mme ebile ba di dirisetsa dipatlisiso fela. Mme fela, ke ka thato ya gago go dumela kgotsa go ganela seno.

Se se kwadilweng ka puo (ntle le maina) se tla phasalatswa mo mafaratlhatlheng a South African Digital Language Resource Centre (https://repo.sadilar.org/). Lebaka ke gore babatlisisi ba bangwe le bone ba kgone go ithuta ka puo ya bana. Ga go maina ape a batsaya karolo a a tlhagisiwang mo dipatlisisong.

Direkhoto tse di tla tseiwang mo patlisisong e, di tlile go dirisetswa go kwala lekwalo la Master's (Master's mini-dissertation), go kwala makwalo a saense, a a tlileng go phatlhalatswa gape le go dirisetswa thuto le dipatlisiso tse dingwe. Ga gona maina ape a batsaya karolo a a tla tlhagisiwang mo dipatlisisong.

Fa o rata go ka bona tshobokanyo ya patlisiso, o ka leletsa nomoro ya me go nkitsise. Fa mmatlisisi mongwe a batla go dirisa dikgang tsa patlisiso e, re tlile go go itsise le go go kopa pele.

Melawana ya Covid-19
Melao ya go fokotsa tshwaetsego ya mogare wa SARS-COV-19 e tla obamelwa jaaka e kaetswe ke World Health Organization (WHO) (2019) le ba National Department of Basic Education (DBE) (2019) go fokotsa mogare wa covid-19 mo baneng kwa dikolong. Melao e e tla obameliwang mo patlisisong ke:

- Ngwana mongwe le mongwe o tla tsweletsa go rwala maseke wa bona ka dinako tsotlhe tsa patlisiso.
- Tlhokomelo ya diatla le go fokotsa tshwarano ka diatla (jaaka, go tlhapa matsogo le go sa atalne kampo go tshwarana ka matsogo).
- Go ema sekgele se le sengwe magareng fa bana ba rutiwa ka patlisiso ena le fa ba kopiwa tetla ya go tsaya karolo kampo fa mmatlisisi a le teng fa sekolong.
- Ngwana mongwe le mongwe o tla fiwa mochini wa go rekhota mme a o dirisa a le mongwe mo tsamaisong ya patlisiso go fokotsa tshwaetsego.

- Mochini mongwe le mongwe wa go rekhota o tla phepafadiwa letsatsi lengwe le lengwe mme ebile o tla tsengwa mo sekgwameng gore o seke wa kopana le dilo tse di na leng mogare wa Covid-19.

Ke eng dikotsi le dipoelo mo patlisisong e?

Ngwana wa gago ga kitla a nna mo kotsing epe ka ntlha ya patlisiso e. Ngwana wa gago ga nkitla lofisiwa sekolo. Ngwana wa gago o tla tswelela ka tsa sekolo jaaka metlha gonne mochini o tla morekhota a ntse a le fa sekolong le bana ba bangwe. Morutabana wa ngwana wa gago o tla netefatsa gore ngwana o apara mochini sentle ka tsela e bolokegileng. Morutabana o tla thusa go ntsha mochini kampo go o baakanya fa o sa nna sentle kgotsa fa ngwana a batla go o rola gotlhelele.

Dipoelo tsa patlisiso e ke go re naya tshedimosetso ka ga mafoko a Setswana a a buiwang thata ke bana a a ka dirisetswang thekenoloji ya Setswana ya AAC go thusa bana ba ba sa kgoneng go bua.

Nka itumela fa o ka tlatsa foromo e e romeletsweng (bona ka fa morago) go kaela fa o letla ngwana wa gago go tsaya karolo mo patlisisong ye kgotsa nyaya.

Fa o na le dipotso dingwe, ka kopo leletsa nna kgotsa mmatlisisi mogolo wa me ka go dirisa dinomoro tse di tlhagisistsweng fa. Ka kopo busetsa foromo e tladitsweng kwa sekolong.

Wa lona,

Gaopalelwe Mogatusi
Mmatlisisi
Mogala: [redacted]
Email: gmogatusi@gmail.com

Letlha

Professor Kerstin Tonsing
Centre for Augmentative and Alternative Communication (CAAC)
Mmatlisisi-mogolo
Mogala: [redacted]
kerstin.tonsing@up.ac.za

Letlha

- Mochini mongwe le mongwe wa go rekhota o tla phepafadiwa letsatsi lengwe le lengwe mme ebile o tla tsengwa mo sekgwameng gore o seke wa kopana le dilo tse di na leng mogare wa Covid-19.

Ke eng dikotsi le dipoelo mo patlisisong e?

Ngwana wa gago ga kitla a nna mo kotsing epe ka ntlha ya patlisiso e. Ngwana wa gago ga nkitla lofisiwa sekolo. Ngwana wa gago o tla tswelela ka tsa sekolo jaaka metlha gonne mochini o tla morekhota a ntse a le fa sekolong le bana ba bangwe. Morutabana wa ngwana wa gago o tla netefatsa gore ngwana o apara mochini sentle ka tsela e bolokegileng. Morutabana o tla thusa go ntsha mochini kampo go o baakanya fa o sa nna sentle kgotsa fa ngwana a batla go o rola gotlhelele.

Dipoelo tsa patlisiso e ke go re naya tshedimosetso ka ga mafoko a Setswana a a buiwang thata ke bana a a ka dirisetswang thekenoloji ya Setswana ya AAC go thusa bana ba ba sa kgoneng go bua.

Nka itumela fa o ka tlatsa foromo e e romeletsweng (bona ka fa morago) go kaela fa o letla ngwana wa gago go tsaya karolo mo patlisisong ye kgotsa nyaya.

Fa o na le dipotso dingwe, ka kopo leletsa nna kgotsa mmatlisisi mogolo wa me ka go dirisa dinomoro tse di tlhagisistsweng fa. Ka kopo busetsa foromo e tladitsweng kwa sekolong.

Wa lona,

Gaopalelwe Mogatusi
Mmatlisisi
Mogala: [redacted]
Email: gmogatusi@gmail.com

Letlha

Professor Kerstin Tonsing
Centre for Augmentative and Alternative Communication (CAAC)
Mmatlisisi-mogolo
Mogala: [redacted]
kerstin.tonsing@up.ac.za

Letlha

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Centre for Augmentative and Alternative Communication

Humanities 100. Since 1919

Tetlelo ya motsadi

Leina la ngwana: _____

Leina la motsadi/motlhokomedi: _____

Leina la patlisiso: **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**

Mmatlisisi: Gaopalelwe Mogatusi
Moithuti wa Master's
Lefapha la AAC
Yunibesithi va Pretoria
Mogala: [redacted]
gmogatusi@gmail.com

Mmatlisisi mogolo: Kerstin Tönsing
Associate Professor
Lefapha la AAC
Yunibesithi va Pretoria
Mogala: [redacted]
kerstin.tonsing@up.ac.za

Nna, _____ (Leina le sefane)

(tshwaya mo lebokoswaneng)

☐ ke neela ngwanake tetla ya go tsaya karolo mo patlisisong e e bidiwang: ***Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities***, e e tsamaisiwang ke Gaopalelwe Mogatusi, ka fa tlase ga pono le botsamaisi bogolo ba ga Mmatlisisi mogolo Kerstin Tonsing. Tetla e, Ke e dira jaana ka boithaopo ebile ke tlhaloganya gore re ka gogela morago ka nako engwe le engwe. Ke tlhaloganya gore ngwanake o tlile go rekhotiwa ka mechini fa sekolong. Ke tlhaloganya gore direkhoto tsa patlisiso di tla bolokiwa ka pabalesego dingwaga di le 15 kwa lefapheng la CAAC le gore direkhoto di tla tshwariwa ka khupamarama. Ke tlhaloganya gore direkhoto tse, di ka dirisetswa makwalo a saense, makwalo magolo le ditlhalosetso. Ke tlhaloganya gore maina a ngwana a tla phomoliwa gotlhelele fa go phatlhalatswa patlisiso ye mo mafaratlhatlheng a South African Digital Language Resource Centre (https://repo.sadilar.org/) le gore patlisiso ye e ka dirisetswa dipatlisiso tse dingwe. Ke tlhaloganya gore dikgang tsotlhe tse di tla tsewang mo patlisisong e di tla tshwarwa ka khupamarama.

KGOTSA

☐ Ga ke neele ngwanake tetla ya go tsaya karolo mo patlisisong e e bidiwang: ***Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities.***

TETLA TLALELETSO,

☐ *Ke neela tetla gore direkhoto tsa puo tsa ngwanake di ka dirisiwa ke babatlisisi ba bangwe. Ke tlhaloganya gore direkhoto di tla bolokiwa ka nako tsotlhe mme di tla dirisetswa fela dipatlisiso.*

KGOTSA

☐ *Ga ke neele tetla gore direkhoto tsa ngwanake di dirisiwe ke batlisisi ba bangwe.*

_____ _____

Tshaeno ya motsadi Letlha

APPENDIX F: ASSENT SCRIPT AND RESPONSE FORM
ENGLISH ASSENT SCRIPT

	Hello, my name is Gaopi. I would like to find out what words children like you use when they are at school, speaking to their teachers and friends. I want to ask you if you would like 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.
	You will keep on wearing your mask even when you are wearing the recorder. I will also keep mine on so that we don't spread the corona virus. Do not fiddle around or touch the microphone during recordings.
	I will keep the microphone I give you in a sealable plastic bag to protect it from germs like the corona virus. Only you will use this microphone alone until I stop recording you. I will also clean your recorder every day to make sure it does not get germs that make you sick.

	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

SETSWANA ASSENT SCRIPT

	Dumela, Leina la me ke Gaopi. Ke tsena sekolo sa go ithuta ka mafoko a bana ba a dirisang ga ba bua fa sekolong. Ke tlile kwano go le kopa gore le nthuse ka patlisiso yame. Fa o dumela,,se ke se re tlileng go se dira:
	Ke tlile go kopa gore o apare mochini wa go rekhota–medumo mo lethekeng (bontsha), mme mochini oo re tla o tsenya ka fa gare ga kgetsana e nyenyane. Ke tlile go ngaparetsa maekhrofounu mo sekipeng sa gago. Ke tlo reetsa mafoko a gago fa o bua le ditsala le morutabana/mistresse wa gago.
	Puo le medumo ya gago e tlile go reetswa ke nna fela le tichere wa me o motona wa ko yunibesithi. Re tlile go reetsa fela medumo le se o se buang.
	Fa mochini kgotsa maeke o sa nna sentle kgotsa o go ngapa, o seke wa leka go o ipakanyetsa. O ise letsogo la gago godimo, o kope tichere go go thusa.

	O tla rwala maseke wa gago ka dinako tsotlhe le fa re rekhota. Le nna ke tla nna ke rwele wa me gore re seke ra tshwaetsana ka corona virus. O seke wa tshwara mochini wa rekhota ka menwana fa o rekhota.
	Ke tla tsenya mochini wa go rekhota mo plastikeng gore ke o sireletse o seke wa nna le corona virus. Ebile gape ke tla phepafatsa rekhota ya gago letsatsi le letsatsi gore e seke ya nna le megare e e ka go lwatsang.
	Fa o sa dumele go apara mochini le maeke, o bolelele tichere gore a go apole one. Ga gona motho yo tlileng go ngalang gore oo apotse kampo sepe se se maswe se se tla go diragallang. A wa tlhaloganya?
	O kgona go tlhopa gore o batla go apara mochini kgotsa ga o batle. Ga gona sepe se se maswe se se tla go diragallang.

CHILD-FRIENDLY RESPONSE FORM (ENGLISH)

Name: _____
Date of birth: _____
Date: _____
Name of the study: **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**
Researcher: Gaopalelwe Mogatusi

Did you understand everything I explained to you? 	
Do you understand that you can choose to participate or not to?  	
Do you understand that you can stop when you want to? 	
Do you understand that I will record the words that you say? 	

Do you have any questions? 	
Are you happy with how I answered your question? 	
Do you want to be part of the study?	

CHILD-FRIENDLY RESPONSE FORM (SETSWANA)

TUMELANO YA GO TSAYA KAROLO

Leina: _____
Matsalo (Letsatsi, kgwedi le ngwaga): _____
Letlha: _____
Leina la patlisisio: **Determining the core vocabulary of Setswana-speaking Grade R learners as used during preschool activities**
Mmatlisi: Gaopalelwe Mogatusi

A o tlhalogantse tsotlhe tse ke di buileng? 	
O tlhaloganya gore o kgona go dumela kgotsa go gana go tsaya karolo?  	
O tlhaloganya gore o ka nne wa emisa ga o batla? 	
O tlhaloganya gore mochini oo o tlile go tsaya mafoko a gago a otlhe a o a buang, ke be ke kgona go a utlwelela? 	

O nale potso kgotsa go na le selo se o sa se tlhaloganyeng? 	
A o itumetseletse dikarabo tse ke go fileng tsona? 	
O batla go tsaya karolo mo patlisisong ena ya me?	

APPENDIX G: CAREGIVER QUESTIONNAIRE

English version
(Based on Mothapo et al.,2019 & 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

1. Does your child speak Setswana as a home language?

☐ Yes

☐ No

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

☐ Yes

☐ No

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

_____

3. 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: _____

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

☐ Yes

☐ No

If not, please describe your concerns: _____

_____

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

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

6. 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

7. 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

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

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

_____

10. Does the child enjoy watching the television (TV) OR listening to the radio?
    Yes ☐ No ☐

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

    _____

11. Does the child enjoy watching videos or listening to music on the cellphone?
    Yes ☐ No ☐

    If yes, what languages is your child exposed to on the cellphone?

    _____

Information about the facilities in the home surroundings

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

    ☐ Yes

    ☐ No

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

    ☐ Yes

    ☐ No

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

    ☐ Yes

    ☐ No

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

    ☐ less than R 7275

    ☐ more than R 7275

Thank you so much for your time and effort in assisting me with my study!

Setswana version
FOROMO YA MOTSADI

(Foromo e e kaetswe ke patlisiso ya ga Mothapo et al.,2019 le Mngomezulu, 2017)

Letlha: _____

Leina la ngwana: _____

Matsalo: _____

Ngwana ke mong: _____

Leina la motho yo tlatsang foromo: _____

Kamano le ngwana: _____

Mogala: _____

Taelo: Ka kopo araba dipotso tse di latelang ka go tshwaya Karabo e e maleba.

Tlhagiso ka ga ngwana

16. Ngwana wa gago o bua Setswana jaaka puo ya gae?

☐ Eya

☐ Nyaa

17. A ngwana wa gago o bua dipuo tse dingwe?

☐ Eya

☐ Nyaa

Fa karabo ya gago e le eya, ke dipuo dingwe tse di feng tseo ngwana a di buang? Tlhalosa:

_____

18. A o tshwenyegile ka tse di latelang mo ngwaneng?:

Pono: Eya ☐ Nyaa ☐

Fa Karabo ya gago ele eya, tlhalosa: _____

Go utlwa: ☐ Eya ☐ Nyaa

Fa Karabo ya gago ele eya, tlhalosa:_____

Go tsamaya: ☐ Eya ☐ Nyaa

Fa Karabo ya gago ele eya, tlhalosa:_____

Go bua: ☐ Eya ☐ Nyaa

Fa Karabo ya gago ele eya, tlhalosa:_____

Go akanya/go tlhaloganya: ☐ Eya ☐ Nyaa

Fa Karabo ya gago ele eya, tlhalosa:_____

19. A o nagana gore go gola ga ngwana wa gago go tsamaisana le dingwaga tsa gage?

☐ Eya

☐ Nyaa

Fa Karabo ya gago ele nyaa, tlhalosa ditshwenyego tsa gago:_____

_____

20. Ngwana wa gago o simolotse go bua lefoko le le lengwe leng (mama, papa, dijo)? Tshwaya ka gangwe fa tlase

Dikgwedi 0-6	Dikgwedi 7-12	Dikgwedi 13-18	Dikgwedi 19-24	Go feta dingwaga tse pedi

21. Bo kgatsadie kgotsa bana ba bangwe ba ba nnang le ngwana mo ntlung

Mong (Mosimane/Mosetsana)	Dingwaga	Kamano le ngwana	Puo e e dirisiwang thata ke ngwana yo	Dipuo tse dingwe tse di dirisiwang ke ene ngwana yo

22. Bagolo ba ba nnang mo ntlung le ngwana

Mo eng (Mosimane/ Mosetsana)	Dingwaga	Kamano le ngwana	Puo e e dirisiwang thata ke mogolo yo	Dipuo tse dingwe tse di dirisiwang ke mogolo yo

23. Ke puo e feng e e dirisiwang **thata** mo metlotlong ya mo gae? _____

24. Ke dipuo tse dife tse dingwe tse go buisaniwang ka tsona mo gae? Tlhalosa:

_____

25. A ngwana o rata go lebelela TV kgotsa go reetsa radio? ☐ Eya ☐ Nyaa

Fa Karabo ya gago ele eya, ke dipuo tse dife tse ngwana a di utlwang mo TV kgotsa mo radiong?

_____

26. A ngwana o rata go lebelela di video mo founung kgotsa go utlwelela dipina mo founung?
☐ Eya ☐ Nyaa

Fa o re eya, ke dipuo dife tse ngwana a di reetsang mo radiong?

_____

Ditlhagiso tsa legae

27. A go na le motlakase mo ntlung?

☐ Eya

☐ Nyaa

28. A go na le metsi a pompo mo ntlung?

☐ Eya

☐ Nyaa

29. A go na le ntlwana ya boithuso mo ntlung?

- ☐ Eya
- ☐ Nyaa

30. Ka kopo tshwaya fa tlase madi a le a kokoanyang kgotsa le a dirisang mo gae kgwedi le kgwedi

- ☐ botlase ba R 7275
- ☐ go feta R 7275

Ke lebogela nako ya gago le matsapa a gago mo go nthuseng ka patlisiso ya me!

APPENDIX H: PRESCHOOL QUESTIONNAIRE
(Based on Mothapo et al., 2019 & Mngomezulu, 2017)

English version

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 Setswana the primary language of instruction used in the preschool?

   ☐ YES

   ☐ NO

2. Do you use Setswana 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?
   _____

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

☐ NO

If yes, please specify: _____

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 or daily program?

☐ YES

☐ NO

If yes, please describe the daily program: _____

_____

Information about the facilities at the preschool

17. How many classrooms does the school facility have? _____

18. Do you have running water at your preschool? YES ☐ NO ☐

19. Do you have electricity at your preschool? YES ☐ NO ☐

20. Do the children have a playground at the preschool? ☐ YES ☐ NO

21. Do the children have an indoor water facility in the preschool (such as for a basin and washing dishes)? YES ☐ NO ☐

22. Do the children have an indoor toilet facility in the preschool? ☐ YES ☐ NO

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

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

☐ YES

☐ NO

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

26. Is the preschool fenced? YES ☐ NO ☐

27. 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 Setswana speaking preschool children between the ages of 5 years, 0 months and 6 years 11 months old. Please nominate two (2) children (including children from both gender) who

- are aged 5-6 years,
- speak Setswana as a first language,
- in your view have adequate speech and language skills for their age,
- tend to be talkative,
- have been in Grade R for at least one month,
- attend Grade R for at least three days a week.

Then kindly send the packages provided to you (containing caregiver information letters and consent forms and caregiver questionnaires) to the parents/legal guardians of the nominated children. The researcher will then collect any consent forms of parents consenting for their child to take part from you.

Thank you so much for your time and effort in assisting me with my research study!

DIPOTSO KA SEKOLO
(Foromo e e kaetswe ke patlisiso ya ga Mothapo et al., 2019 le Mngomezulu, 2017)

Setswana version

Mosola wa dipotso tse ke go itse ka ga sekolo

Letlha: _____
Motho yo o tlatsang foromo: _____
Maemo a gago mo sekolong: _____
Leina la sekolo: _____

Taelo: Ka kopo araba dipotso ka go tshwaya
Tshedimosetso ka dipuo tse di dirisiwang mo sekolong

1. A Setswana ke puo ya ntlha e e dirisiwang mo sekolong?
   ☐ Eya
   ☐ Nyaa
2. A le dirisa Setswana jaaka puo ya thuto mo phaphusing ya boithutelo?
   ☐ Eya
   ☐ Nyaa
3. Fa Karabo ya gago ele nyaa, ke dipuo tse di feng tse dingwe tse di dirisetswang go ruta?
   Tlhalosa:_____
   _____
4. Ke puo e e feng e bana ba e dirisang gantsi mo phaphusi boithutelong?
   _____
5. Ke dipuo tse di feng tse dingwe tse bana ba di dirisang go buisana le bana ba bangwe?
   _____
6. Go na le bathusi ba barutabana ba ba kae mo phaphusing ya boithutelo ya lona? (Ga ba seyo, kwala 0)
   _____
7. Ke dipuo tse dife tse bathusi ba barutabana ba di dirisang mo phaphusing?
   _____
8. Ke dipuo dingwe tse di feng gape tse bathusi ba barutabana ba di dirisang le bana?
   _____

Tshedimosetso ka ga bana ba sekolo le lenaneo la sekolo

9. Go na le bana ba ba kae mo phaphusing ya boithutelo?_____

10. Dingwaga tsa bana ba ba leng mo **phaphusing** di dikae? Go simolla ka dingwaga _____ (bonnyenyane) go dingwaga di le_____ (botona).

11. Go na le bana ba ba kae mo sekolong? _____

12. Go na le diphaphusi boithutelo di le dikae mo sekolong? _____

13. Sekolo sa lona se na le lenaneo la thuto?
    - ☐ EYA
    - ☐ NYAA

Fa o re eya, tlhalosa: _____

14. Dingwaga tsa bana ba **sekolo** di dikae? Go simolla ka dingwaga _____ (bonnyenyane) go dingwaga di le _____ (botona).

15. A bana ba ba leng mo phaphusing ba kgona go amana le bana ba bangwe mo sekolong? Tlhalosa: _____

_____

16. A sekolo se se latela tiro ya letsatsi ka metlha?
    - ☐ EYA
    - ☐ NYAA

Fa o re eya, ka kopo tlhalosa tiro ye ya letsatsi le letsatsi: _____

_____

Tshedimosetso ka di dirisiwa tsa sekolo

17. Go na le di phaphusi boithutelo tse kae mo sekolong? _____

18. Le na le mesti a pompo mo sekolong? Eya ☐ Nyaa ☐

19. Le na le motlakase mo sekolong? Eya ☐ Nyaa ☐

20. A bana ba na le lebala la metshameko? Eya ☐ Nyaa ☐

21. A bana ba na le sekotlolo sa gotlhpa mabogo/beisini mo gare ga phaphusi?
Eya ☐ Nyaa ☐

22. A bana ba na le ntlwana boithuselo mo sekolong? Eya ☐ Nyaa ☐

23. Go na le dintlwana boithuselo tse kae mo sekolong? (Kwa ntle kgotsa ka fo gare)
_____

24. A badiri ba sekolo ba na le dintlwana boithuselo tsa bona mo sekolong?

☐ Eya

☐ Nyaa

25. Badiri ba sekolo ba na le dintlwana tse kae mo sekolong? (Kwa ntle kgotsa ka fo gare)
_____

26. A lebala la sekolo le sekeleditswe ka fense? Eya ☐ Nyaa ☐

27. A sekolo se na le di dirisiwa tse? Tshwaya tse di leng teng.

Mogala	Fekese	Mafaratlhatlha

<u>Go tlhopha batsaya karolo</u>

Moano wa patlisiso e ke go amogela mafoko a Setswana a a dirisiwang ke bana ba sekolo ba dingwaga di le 5 go fitlha go dingwaga di le 6 ka dikgwedi di le 11. Ka kopo tlhopha bana ba le babedi (2) (basimane le basetsana) ba ba:

- na leng dingwaga di le 5-6,
- buang Setswana jaaka puo ya gae,
- go ya ka wena, ba na le puo e e tlhwatlhwa e e tshwanetseng dingwaga tsa bona,
- Ba bua go fetisa ba bangwe,
- Ba tsene mophato wa Grade R go feta kgwedi e le nngwe,
- Ba tsena sekolo malatsi a mararo a beke kgotsa go feta.

Romela foromo ya tetla ya batsadi (Appendix D) gore batsadi ba neye tetla pele ga mmatlisisi a neelwa maina a bana. Foromo ya dipotso tsa batsadi (Appendix E) e tsamaye le foromo ya tetla go ya go batsadi. Mmatlisisi o tla tsaya di foromo tse di signilweng ke batsadi mo go wena.

Ke lebogela nako ya gago le matsapa a go nthusa ka patlisiso ya me!

APPENDIX I: TEACHER INSTRUCTIONS

TEACHER INSTRUCTIONS FOR RECORDINGS AND EQUIPMENT

Thank you for allowing me to work with learners in your class for this study. Please take note of the following during the course of the study.

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 irritating 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.

Note:
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 [redacted] .

Yours sincerely
Mogatusi MG

APPENDIX J: TRANSCRIPTION RULES

Transcription rules based on Du Bois (1991) and Trembath (2007) were used to compile transcription rules. Additional rules were added to accommodate the Setswana language.

Rule	**Example** (where applicable)
1. The first 20 minutes of each recording will be omitted in transcription to eliminate novelty effects.	
2. One document per child will be used. The learner's initials will be used to identify the document. Note: If two learners have the same initials then numbering will differentiate them. Utterances should be transcribed individually and each new sentence on a new line-An utterance is a word production defined by intonation or a pause of greater than 2 seconds.	'Remoratile Matu' and 'Remoratile Morwane' being the name of two learners. Numbers 1 and 2 will then differentiate them: Codes RM1 & RM2 will be used respectively, as the document name. Transcript one: RM1 Transcript two: RM2: *Nna ke ja dijo.* *Tshaba foo!.*
3. Transcription will commence 20 minutes into a recording and stop when 3 000 orthographic words have been obtained.	(See bottom left word count on word document to keep track of words)
4. Recordings may be presented in short clips and not one long clip. The symbol = will be used to indicate the end of a recording in transcription. This will mark transcription of a new recording clip on the transcript.	Re a tsamaya. = (End of clip & Beginning of new clip) Mphe dijo = Mme, ke kopa TissueCS.
5. Environmental noises, sounds or words uttered by other people **will not** be included in transcription	A teacher or other peers speaking in the background of the learner's conversation will not be transcribed. Only the learners' productions will be considered.
6. No proper nouns should be transcribed. The following codes will be used: CN (child name), TN (teacher name), PN (other person's name) and PPN (place name). This is to protect the safety and confidentiality of participants.	*Malome PN ke ole!* *Mabane ke ne ke ile ko PPN*

Rule	**Example** (where applicable)
7. Production of interjections such as "heeh!", fillers such as "err" / "mmm" or 'uhm' will be transcribed with phonetical consistency (similarly through-out)	Example 'Uhm' consistently written as is, not changed to 'ahm' or 'ehm'
8. Interjections will be written as one word and will be counted in the analysis	Interjections (words expressing reactions or emotions) "He eh" – usually written disjunctively (separately words) should be written as one word = "heeh"
9. Vulgar words will be transcribed and will not be excluded.	
10. Conversations regarding the recording equipment or the research study will **not** be transcribed and analyzed.	Conversations like: "Founu ya ka e a wa" "Ba tsentse mochini ka fa" Will **NOT** be transcribed
11. Every utterance will end with a punctuation mark, either a full stop to indicate end of production, question mark for questions, exclamation marks for emphasis and interrupted words will be indicated by – .	Question • *….kae?* Exclamation • *….Ena!* Interrupted words • *Tliii-*
12. Numbers uttered will be transcribed as words. Numbers consisting of more than one word will be written as one word	Numbers *Three* for 3 *Five* for 5 Number > one word =written as one word *Twentyfour* for 24 *Fiftyfive* for *55*
13. Syllable, sound repetitions and prolongations will be transcribed as one word.	Brrr brrr brrrrr = br Mmm mmm mmmm = mm
14. Names of nursery rhymes, cartoons, and books will be **transcribed as one word** (i.e., no space between words), and placed within parenthesis. Words spoken by participants that are parts of rote recitations, such as prayers, nursery rhymes, reciting the alphabet and songs will also be transcribed as one word and placed within parenthesis.	Cartoons *Cool cats= (Coolcats)* Songs Twinkle twinkle = (*Twinklestar)* Happy Birthday= (*HappyBirthday)* Prayer Se re yang go se ja= (Serayanggoseja)

Rule	**Example** (where applicable)
15. If a part of the word is unintelligible, the whole word will be transcribed as unintelligible, using the code ?? Similar applies to an unintelligible part of an utterance or a whole utterance that is unintelligible.	S: Re ya ko ??
16. Setswana spelling rules will be adhered to during the transcription process. In case of uncertainty, A bilingual Setswana-Seesimane Thanodi ya Sekolo (Bilingual Setswana-English school dictionary) should be consulted. Mispronunciations by children will be transcribed as if they did not occur, and the correct spelling of the intended word will be used.	Children's phonological processes should not affect the correct spelling of the word. X **Ny**o**ny**ane √ Nonyane Most children may assimilate sounds in words due to developmental articulation processes
17. Words spoken in other languages (code switches) will be transcribed using the orthography of the target language (e.g., English). The word will be identified as belonging to another language by adding CS at the end of the word (no space). Loan words that are pronounced with Setswana pronunciation do not count as code switches. They will be transcribed according to Setswana spelling rules.	Code switch: WhyCS ToiletCS Not a code switch but a loan word: Phathi (English – party)

APPENDIX K: CODING RULES

Coding rules were created to accommodate the Microsoft Word Macro (Simonyi,1983) function and to accommodate the Setswana language.

Coding rules were particularly created to:
1) Enable the Word Macro function to easily enable the identification of verbs, nouns and adjectives that share the same root, so that all morphological variations can be counted under one root word.
2) Avoid counting heteronyms and polysemous words that have obvious different lexical meanings as one vocabulary item.

The coding rules pertaining to these two objectives are set out in more detail here following.

1. ***Coding of inflected forms of nouns, verbs, and adjectives***

Morphological variations of certain words will be transcribed in such a way that Microsoft Macro is able to identify the root word (lemma) and the word variations. The root word will be transcribed first followed by a number code (refer to under heteronyms below) for the morpheme inflection. Thereafter, either the part of the word that represents a morphological variation or the whole inflected word will be transcribed. Specific examples pertaining to nouns, verbs and adjectives are presented in the table following.

Coding Rules:

Part of speech	**Word form**	**Root word Example**	**Grammatical variations**	**Manner of coding**	**Example**	**Example in sentence and in coded sentence**
Nouns	Singular form	Lebati	Plural Mabati	(lemma)1(plural prefix)	Lebati1ma	Mama o reka mabati Coded: Mama o reka lebati1ma
	Locative	Lebati	Lebating	(lemma)1(locative suffix)	Lebati1ng	Ema mo lebating Coded: Ema mo lebati1ing

Part of speech	Word form	Root word Example	Grammatical variations	Manner of coding	Example	Example in sentence and in coded sentence
Verbs	Imperfect indicative mood	Bula	Negative form-1 Bule	(lemma)3(negative)	Bula3e	Lebati ga le bule Coded: Lebati ga le bula3e
			Negative form-2 Bulege	(lemma)3(negative)	Bula3ege	Lebati ga le bulege. Coded: Lebati ga le bula3ege
			Object concord Mpulele	(lemma)3(object concord)	Bula3Mpulele	Mpulele lebati Coded: Bula3Mpulele lebati
	Perfect indicative mood	Kwala	Kwadile	(lemma)3(past tense inflection)	Kwala3dile	Rre o kwadile lekwalo Coded: Rre o kwala3dile lekwalo
	Imperative mood	Itaya	Se itaye	(lemma)3(negative imperative suffix)	Itaya3e	Se itaye motho o Coded: Se itaya3e motho o
	Subjunctive	Bina	Bine	(lemma)3(negative subjunctive suffux)	Bina3e	Gore o se bine Coded: Gore o se bina3e

Part of speech	Word form	Root word Example	Grammatical variations	Manner of coding	Example	Example in sentence and in coded sentence
Adjectives	Adjectives modified by class prefixes	Ntle	(Dintle/ sentle/ mantle)	(lemma)5(class prefix)	Ntle5se	Ditlhako tsa gago di dintle Coded: Ditlhako tsa gago di ntle5di
	Complex adjective/ Modified adjective	Ntle	Sentlenyana	(lemma)5 (modified adjective + class prefix)	Ntle5nyanase	Selo se, se sentlenyana Coded: Selo se, se ntle5nyanase

The examples given are not exhaustive.

2. *Coding of heteronyms*

The Setswana language has heteronyms (e.g., bona-see & bona-them) and polysemous words (e.g., letlhare-leaf, letlhare-page).
These are words written and spelled the same but have different meanings and are pronounced differently to give meaning to a sentence. It is important to differentiate between such words as each holds significance to the semantics of the language. The purpose of coding heteronyms is to separate words that are spelled exactly the same in the transcript (taking coding into consideration) but have different meanings.

The first method of separating heteronyms that belong to different parts of speech is to code them with a number designating different parts of speech. The following numbers were used:

Number codes
Nouns-1
Pronoun-2
Verb- 3
Adverb-4
Adjectives-5
Interjection-6
Concord- 7
Enumerative-8
Demonstrative-9
Quantitative-10
Possessive-11
Ideophone-12
Conjunction-13
Preposition-14

When a heteronym had different meanings but belonged to the same part of speech, the letters 'a' and 'b' were placed after the number to separate the two words. A complete list of all coded heteronyms and homonyms is provided below.

Heteronym & Homonym Codes:

Heteronym	**Variations**	**Codes assigned**	**Example**
Tlhaga Tlhaga-appear Tlhaga-grass	Tlhaga (Verb)	Tlhaga3	O tlile go tlhaga3 mo telebishineng gompieno.
	Tlhaga (Noun)	Tlhaga1	Tlhaga1 ele e tala.
Ya	Ya (Verb)	Ya3a	
	Ya (Possessive)	Ya11	

Heteronym	Variations	Codes assigned	Example
Fitlha (Verb)	Fitlha-*Bury*	Fitlha3a	Fitlha selo seo
	Fitlha- *Arrive*	Fitlha3b	Bona ya ka e fitlha fo kae
Nna Nna-sit Nna-me	Nna (Verb)	Nna3	Nna3 fa fatshe.
	Nna (Pronoun)	Nna2	Le nna2 ke moithuti.
<u>Polysemous verbs:</u> The word "Paka" can have two different verbial meanings	1.Paka=Pack 2.Paka=Testify (Both verbs)	Verb codes: 3a & 3b 1.Paka3a 2.Paka3b	 Coded: Rakgadi ga a paka3a beke ya me gompieno. Coded: Ke boleletse Mme, o tlile go paka3b gore o ba bone kae!

In coding heteronyms and homonyms, transcribers will use their knowledge of Setswana to determine which part of speech does the word in a sentence represent if it is a heteronym.

APPENDIX L: SETSWANA CORE LIST

CORE VOCABULARY LIST WITH FREQUENCY OF OCCURRENCE, COMMONALITY SCORE AND COVERAGE

Words	No. of occurrences	Frequency permille	Commonality	Content/Structure	Part of speech	Coverage
Ke (I)	1278	70,611636	6	structure	Concord	7,0611636
o2 (He/ She)	935	51,66031272	6	structure	pronoun	5,166031272
go (to)	592	32,70898945	6	structure	concord	3,270898945
a7 (then, is, are)	490	27,07331897	6	structure	Concord	2,707331897
pn (pesronsname)	338	18,67506492	6	content	Noun	1,867506492
wêna (you)	283	15,636223	6	structure	pronoun	1,5636223
wa7 (his/hers/their)	274	15,13895795	6	structure	possessive concord	1,513895795
ka7	269	14,8626996	6	structure	Concord	1,48626996
tlaa (will)	241	13,3156528	6	structure	auxillary verb	1,33156528
re2 (we)	234	12,9288911	4	structure	pronoun	1,29288911
nna2 (me)	217	11,98961269	6	structure	Pronoun	1,198961269
ba2 (they)	222	12,26587104	6	structure	pronoun	1,226587104
bona3 (look)	217	11,98961269	6	content	Verb	1,198961269
mo7	207	11,43709597	6	structure	Concord	1,143709597
ga15	210	11,60285099	6	structure	prefix	1,160285099
e7 (it)	192	10,6083209	4	structure	concord	1,06083209
mme/mma (mom)	148	8,177247362	6	content	noun	0,817724736
ee (yes)	177	9,779545831	6	structure	interjection	0,977954583
fa4 (here)	165	9,116525775	6	content	adverb	0,911652577
heeh/e-e (No)	179	9,890049174	6	structure	interjection	0,989004917
e9 (this one)	176	9,72429416	6	structure	demonstrative	0,972429416
ya11 (of)	152	8,398254047	6	structure	possessive concord	0,839825405

Words	No. of occurrences	Frequency permille	Commonality	Content/Structure	Part of speech	Coverage
le13 (and)	147	8,12199569	6	structure	Conjunction	0,812199569
tlhe!	140	7,735233991	6	structure	interjection	0,773523399
le2	138	7,624730648	6	structure	pronoun	0,762473065
di7 (they are)	129	7,127465606	6	structure	concord	0,712746561
ile (will)	127	7,016962263	6	structure	auxillary verb	0,701696226
ka11 (mine)	124	6,851207249	6	structure	possessive concord	0,685120725
ngwana (child)	118	6,519697221	6	content	noun	0,651969722
eng? (what)	115	6,353942207	6	structure	pronoun	0,635394221
itse (know)	114	6,298690535	6	content	verb	0,629869054
o9 (that)	110	6,07768385	6	structure	demonstrative	0,607768385
ne (was)	104	5,746173822	6	content	verb	0,574617382
ga7	103	5,69092215	6	structure	concord	0,569092215
tsamaya (go)	101	5,580418808	6	content	verb	0,558041881
batla (look for)	99	5,469915465	6	content	verb	0,546991546
ya3 (go)	98	5,414663794	6	content	verb	0,541466379
tse (these)	98	5,414663794	6	structure	demonstrative	0,541466379
etsa (make)	97	5,359412122	6	content	verb	0,535941212
tsaya3 (take)	95	5,248908779	6	content	verb	0,524890878
ja3e (eat)	95	5,248908779	6	content	verb	0,524890878
kae (where)	94	5,193657108	6	content	adverb	0,519365711
foo (there)	88	4,86214708	6	content	adverb	0,486214708
o7	90	4,972650423	6	structure	concord	0,497265042
se2 (it)	88	4,86214708	6	structure	pronoun	0,486214708
nna3 (sit)	88	4,86214708	6	content	verb	0,486214708
wa7 (their)	87	4,806895409	6	structure	possessive concord	0,480689541
kwa (at)	86	4,751643737	6	content	adverb	0,475164374
yoh!	83	4,585888723	3	structure	interjection	0,458588872

Words	No. of occurrences	Frequency permille	Commonality	Content/Structure	Part of speech	Coverage
ema (stop)	80	4,420133709	6	content	verb	0,442013371
le9	80	4,420133709	6	structure	demonstrative	0,442013371
tlogela3 (leave)	75	4,143875352	6	content	verb	0,414387535
jaana (like this)	73	4,03337201	6	content	adverb	0,403337201
sa7 (of)	73	4,03337201	6	structure	possessive concord	0,403337201
akere (right?)	72	3,978120338	6	structure	interjection	0,397812034
re3 (say)	72	3,978120338	6	content	verb	0,397812034
neh (okay?)	70	3,867616995	5	structure	interjection	0,3867617
motho (person)	69	3,812365324	6	content	noun	0,381236532
gore (so that)	68	3,757113653	6	structure	Conjunction	0,375711365
mang (who)	68	3,757113653	6	structure	pronoun	0,375711365
se9 (this)	67	3,701861981	6	structure	demonstrative	0,370186198
kopa (ask)	66	3,64661031	6	content	Verb	0,364661031
seka (should not)	66	3,64661031	6	structure	auxillary verb	0,364661031
tana3 (climb)	62	3,425603625	5	content	verb	0,342560362
selo (thing)	61	3,370351953	6	content	Noun	0,337035195
heh? (huh?)	61	3,370351953	5	structure	interjection	0,337035195
gago (your)	60	3,315100282	6	structure	pronoun	0,331510028
wa3 (fall)	60	3,315100282	6	content	verb	0,331510028
tlaya (come)	66	3,64661031	6	content	verb	0,364661031
tla (will)	59	3,25984861	6	structure	auxillary verb	0,325984861
kwala (write)	56	3,094093596	5	content	Verb	0,30940936
utlwa (hear)	54	2,983590254	5	content	Verb	0,298359025
bona2 (them)	54	2,983590254	6	structure	pronoun	0,298359025
monna (man)	53	2,928338582	4	content	noun	0,292833858
lona2 (it)	53	2,928338582	6	structure	pronoun	0,292833858
se4 (not)	53	2,928338582	6	content	adverb	0,292833858
teng (present)	53	2,928338582	6	content	adjective	0,292833858

Words	No. of occurrences	Frequency permille	Commonality	Content/Structure	Part of speech	Coverage
na (have)	53	2,928338582	6	content	verb	0,292833858
gatwe(apparently)	51	2,81783524	6	content	adverb	0,281783524
tla3e (come)	51	2,81783524	6	content	verb	0,281783524
fetsa (finish)	50	2,762583568	6	content	Verb	0,276258357
tsa (their)	50	2,762583568	6	structure	possessive concord	0,276258357
hee!	49	2,707331897	3	structure	interjection	0,27073319
mm (yes)	49	2,707331897	5	structure	interjection	0,27073319
fa3e (give)	46	2,541576883	3	content	verb	0,254157688
se7 (is)	45	2,486325211	6	structure	concord	0,248632521
tlisa (bring)	43	2,375821869	6	content	verb	0,237582187
tsena3 (enter)	42	2,320570197	5	content	Verb	0,23205702
pele (in front)	42	2,320570197	6	content	adverb	0,23205702
dilo (things)	41	2,265318526	5	content	noun	0,226531853
sharpcs (sharp)	40	2,210066855	5	structure	interjection	0,221006685
rona2 (us)	40	2,210066855	6	structure	pronoun	0,221006685
mancs (man)	40	2,210066855	6	content	noun	0,221006685
ba7 (are)	40	2,210066855	6	structure	concord	0,221006685
baa (put)	39	2,154815183	6	content	verb	0,215481518
apole1 (apple)	38	2,099563512	4	content	Noun	0,209956351
fela (only)	38	2,099563512	6	content	adverb	0,209956351
mama (mom)	38	2,099563512	4	content	Noun	0,209956351
eishcs	38	2,099563512	5	structure	interjection	0,209956351
mongwe (certain thing/person)	37	2,04431184	6	content	noun	0,204431184
eh!	37	2,04431184	4	structure	interjection	0,204431184
saancs (son)	35	1,933808498	3	content	noun	0,19338085
robala (sleep)	35	1,933808498	5	content	verb	0,19338085
ah!	33	1,823305155	4	structure	interjection	0,182330515

Words	No. of occurrences	Frequency permille	Commonality	Content/Structure	Part of speech	Coverage
jaanong (and now)	34	1,878556826	6	structure	conjunction	0,187855683
khalara (colour in)	34	1,878556826	3	content	verb	0,187855683
tshaba (run away)	34	1,878556826	4	content	Verb	0,187855683
tsenya3 (put in)	34	1,878556826	6	content	verb	0,187855683
betsa (beat)	33	1,823305155	6	content	verb	0,182330515
tswa (come out)	33	1,823305155	6	content	verb	0,182330515
tshameka (play)	33	1,823305155	6	content	Verb	0,182330515
ntsha3 (take out)	32	1,768053484	6	content	Verb	0,176805348
papa (dad)	32	1,768053484	6	content	Noun	0,176805348
tshwara3 (touch)	31	1,712801812	6	content	Verb	0,171280181
bolaisa (tell on)	31	1,712801812	5	content	Verb	0,171280181
metsi (water)	30	1,657550141	5	content	Noun	0,165755014
onecs (one)	30	1,657550141	5	content	noun	0,165755014
tlhapa3 (bath)	29	1,60229847	6	content	verb	0,160229847
tshwara3(arrest)	29	1,60229847	6	content	verb	0,160229847
dijo (food)	28	1,547046798	4	content	noun	0,15470468
gape (again)	28	1,547046798	6	content	adverb	0,15470468
siana (run)	28	1,547046798	4	content	verb	0,15470468
fale (over there)	27	1,491795127	4	content	adverb	0,149179513
pushacs (push)	27	1,491795127	4	content	verb	0,149179513
mhmm	27	1,491795127	3	structure	interjection	0,149179513
geh (then)	27	1,491795127	3	content	adverb	0,149179513
letsogo (arm)	26	1,436543455	5	content	Noun	0,143654346
dlalacs (play)	26	1,436543455	3	content	verb	0,143654346
gobala (get hurt)	26	1,436543455	3	content	verb	0,143654346
mosimane (boy)	26	1,436543455	5	content	noun	0,143654346
tlile (shall)	25	1,381291784	5	structure	auxillary verb	0,138129178
monate (nice)	25	1,381291784	5	content	adjective	0,138129178
bontsha3e (show)	24	1,326040113	5	content	verb	0,132604011

Words	No. of occurrences	Frequency permille	Commonality	Content/Structure	Part of speech	Coverage
maaka (lies)	24	1,326040113	5	content	Noun	0,132604011
mmu (soil)	24	1,326040113	4	content	Noun	0,132604011
tshela3a (pour)	24	1,326040113	4	content	verb	0,132604011
tswala (close)	24	1,326040113	3	content	verb	0,132604011
a11 (of the)	23	1,270788441	6	structure	Possessive concord	0,127078844
le7	23	1,270788441	6	structure	concord	0,127078844
cn (childname)	23	1,270788441	3	content	noun	0,127078844
jang (how)	23	1,270788441	5	content	adverb	0,127078844
panana (banana)	23	1,270788441	3	content	noun	0,127078844
waka (mine)	23	1,270788441	4	structure	pronoun	0,127078844
ausi (sister)	22	1,21553677	3	content	noun	0,121553677
nngwe (one)	22	1,21553677	5	content	enumerative	0,121553677
twocs (two)	22	1,21553677	5	content	enumerative	0,121553677
bo7	21	1,160285099	6	structure	concord	0,11602851
feta (pass)	21	1,160285099	4	content	verb	0,11602851
moswinki (swing)	21	1,160285099	3	content	noun	0,11602851
nx!	21	1,160285099	4	structure	interjection	0,11602851
rata (love)	21	1,160285099	5	content	verb	0,11602851
tshomi (friend)	21	1,160285099	3	content	noun	0,11602851
adima (borrow)	20	1,105033427	3	content	verb	0,110503343
bitsa (call)	20	1,105033427	5	content	verb	0,110503343
buka (book)	20	1,105033427	5	content	noun	0,110503343
ene (him/her)	20	1,105033427	5	structure	pronoun	0,110503343
itaya3e (beat)	20	1,105033427	3	content	verb	0,110503343
kherayone1di (crayon)	20	1,105033427	4	content	noun	0,110503343
lela (cry)	20	1,105033427	6	content	Verb	0,110503343
mare(maar)-(but)	20	1,105033427	4	structure	conjunction	0,110503343

Words	No. of occurrences	Frequency permille	Commonality	Content/Structure	Part of speech	Coverage
bolelela (tell)	19	1,049781756	4	content	verb	0,104978176
ole (that)	19	1,049781756	3	structure	demonstrative	0,104978176
sekolo1 (school)	19	1,049781756	5	content	noun	0,104978176
whycs (why?)	19	1,049781756	4	content	adverb	0,104978176
isa (take to)	18	0,994530085	3	content	verb	0,099453008
maskcs (mask)	18	0,994530085	5	content	noun	0,099453008
morago (back)	18	0,994530085	4	content	adverb	0,099453008
nnye5 (small)	18	0,994530085	3	content	adjective	0,099453008
ba9	17	0,939278413	4	structure	demonstrative	0,093927841
be4 (then)	17	0,939278413	3	content	adverb	0,093927841
botsa (ask)	17	0,939278413	4	content	verb	0,093927841
ntle5 (beautiful)	17	0,939278413	4	content	adjective	0,093927841
thata5 (difficult/hard)	17	0,939278413	6	content	adjective	0,093927841
dira (do)	16	0,884026742	3	content	verb	0,088402674
la (of)	16	0,884026742	5	structure	demonstrative	0,088402674
lebelela3etse (look)	16	0,884026742	4	content	verb	0,088402674
naare (what exactly)	16	0,884026742	6	content	adverb	0,088402674
tsona (them)	16	0,884026742	4	structure	pronoun	0,088402674
kwano (here)	16	0,884026742	4	content	adverb	0,088402674
fatshe (down)	15	0,82877507	6	content	adverb	0,082877507
hae!	15	0,82877507	5	structure	interjection	0,082877507
ithaya (think)	15	0,82877507	6	content	verb	0,082877507
bana (kids)	15	0,82877507	5	content	noun	0,082877507
mmele (body)	15	0,82877507	3	content	noun	0,082877507
namela (climb)	15	0,82877507	4	content	verb	0,082877507
ohh!	15	0,82877507	6	structure	interjection	0,082877507

Words	No. of occurrences	Frequency permille	Commonality	Content/Structure	Part of speech	Coverage
ppn (proper noun/place names)	15	0,82877507	4	content	Noun	0,082877507
sorrycs (sorry)	15	0,82877507	3	structure	interjection	0,082877507
abuti (brother)	14	0,773523399	4	content	noun	0,07735234
apola3 (undress)	14	0,773523399	5	content	verb	0,07735234
ebile (even)	14	0,773523399	3	content	adverb	0,07735234
ise (not yet)	14	0,773523399	5	content	adverb	0,07735234
tsoga (wakeup)	14	0,773523399	3	content	verb	0,07735234
yoo (that one)	14	0,773523399	5	structure	demonstrative	0,07735234
ga4	14	0,773523399	3	content	adverb	0,07735234
baakanya (fix)	13	0,718271728	3	content	verb	0,071827173
bula(open)	13	0,718271728	4	content	verb	0,071827173
ene (him/her)	13	0,718271728	5	structure	Pronoun	0,071827173
fe? (which)	13	0,718271728	5	structure	pronoun	0,071827173
koo (there)	13	0,718271728	3	structure	demonstrative	0,071827173
moscs (moes-Afrikaans)	13	0,718271728	5	structure	interjection	0,071827173
mosetsana (girl)	13	0,718271728	4	content	noun	0,071827173
pencilcs (pencil)	13	0,718271728	4	content	noun	0,071827173
seo (that)	13	0,718271728	5	structure	demonstrative	0,071827173
threecs (three)	13	0,718271728	3	content	enumerative	0,071827173
toiletcs (toilet)	13	0,718271728	5	content	noun	0,071827173
botlhoko (painful)	13	0,718271728	4	content	adjective	0,071827173
botlhe (all of)	12	0,663020056	4	content	quantitative	0,066302006
bua (talk)	12	0,663020056	4	content	verb	0,066302006
hehe	12	0,663020056	3	structure	interjection	0,066302006
iyoh!	12	0,663020056	4	structure	interjection	0,066302006
kaka (pooping)	12	0,663020056	3	content	verb	0,066302006

Words	No. of occurrences	Frequency permille	Commonality	Content/Structure	Part of speech	Coverage
kuka (carry)	12	0,663020056	3	content	verb	0,066302006
kana (by the way)	12	0,663020056	5	structure	conjunction	0,066302006
nwa (drink)	12	0,663020056	4	content	verb	0,066302006
sekeletsa (round)	12	0,663020056	3	content	noun	0,066302006
senya3 (ruin)	12	0,663020056	5	content	verb	0,066302006
tlhaba3 (stab)	12	0,663020056	4	content	verb	0,066302006
tlhapi (fish)	12	0,663020056	3	content	noun	0,066302006
tsaka (mine)	12	0,663020056	3	structure	pronoun	0,066302006
tseo (those)	12	0,663020056	4	structure	demonstrative	0,066302006
tshwana (same/ be alike)	12	0,663020056	5	content	verb	0,066302006
bola3dile (rot)	11	0,607768385	4	content	verb	0,060776838
eo (that)	11	0,607768385	5	structure	demonstrative	0,060776838
gradecs (grade)	11	0,607768385	3	content	noun	0,060776838
haha	11	0,607768385	6	structure	interjection	0,060776838
koloi (car)	11	0,607768385	3	content	noun	0,060776838
leina (name)	11	0,607768385	4	content	noun	0,060776838
leswana (spoon)	11	0,607768385	3	content	noun	0,060776838
me (myself)	11	0,607768385	5	structure	Pronoun	0,060776838
ntlu1 (house)	11	0,607768385	3	content	noun	0,060776838
ntsi5 (many)	11	0,607768385	5	content	adjective	0,060776838
pedi (two)	11	0,607768385	4	content	noun	0,060776838
phala3 (surpass)	11	0,607768385	5	content	verb	0,060776838
sele9 (that)	11	0,607768385	5	structure	demonstrative	0,060776838
sona2 (it)	11	0,607768385	6	structure	Pronoun	0,060776838
thusa3 (help)	11	0,607768385	3	content	verb	0,060776838
apara (wear)	10	0,552516714	3	content	verb	0,055251671
bolelo (hot)	10	0,552516714	3	content	adjective	0,055251671
eseng (not)	10	0,552516714	4	content	adverb	0,055251671

Words	No. of occurrences	Frequency permille	Commonality	Content/Structure	Part of speech	Coverage
jumpacs (jump)	10	0,552516714	3	content	verb	0,055251671
kgona (can)	10	0,552516714	5	structure	auxillary verb	0,055251671
mmala (colour)	10	0,552516714	4	content	noun	0,055251671
nne (four)	10	0,552516714	4	content	enumerative	0,055251671
ntate (father)	10	0,552516714	4	content	noun	0,055251671
omanya (shout)	10	0,552516714	3	content	verb	0,055251671
sekhafitini (lunchbox)	10	0,552516714	6	content	noun	0,055251671
tlhogo (head)	10	0,552516714	3	content	noun	0,055251671

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

APPENDIX M: DECLARATION OF LANGUAGE EDITING

VSE

Van Schalkwyk Editorial Services

Email: arayofhope1@gmail.com

LinkedIn profile: https://www.linkedin.com/in/ar%C3%A9-van-schalkwyk-0214202a/

04/11/2022

DECLARATION OF PROFESSIONAL EDIT

DETERMINING THE CORE VOCABULARY OF SETSWANA-SPEAKING GRADE R LEARNERS AS USED DURING SCHOOL ACTIVITIES

by

Morwesi Gaopalelwe Mogatusi

I declare that I have edited this mini-dissertation. My involvement was restricted to language usage and spelling, completeness and consistency, reference style, and formatting of headings, captions and tables of contents. I did no structural rewriting of the content and did not influence the academic content in any way.

Mr Aré van Schalkwyk
BA (Languages)
Accredited service provider of the University of Pretoria, Stellenbosch University, the University of Johannesburg, and other institutions