UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Determining the core vocabulary of Sepedi-speaking Grade R learners from the Sekhukhune district during regular school activities

by

Mmoto Charmaine Moswathupa
Student no: 23871165

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

Master's in Augmentative and Alternative Communication

in the Centre for Augmentative and Alternative Communication

UNIVERSITY OF PRETORIA

FACULTY OF HUMANITIES

SUPERVISOR: Professor Kerstin Tönsing
CO-SUPERVISOR: Ms Rahab Mothapo

October 2024
UNIVERSITY OF PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA 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: Mmoto Charmaine Moswathupa

Student number: 23871165

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

"Perhaps this is the moment for which you have been created."

This mini-dissertation was only possible with the support and guidance of the following individuals and institutions.

Firstly, my kind supervisor, Professor K. Tönsing, and co-supervisor, Ms. Mothapo, I greatly appreciate your invaluable support and guidance throughout my journey. Your expertise and knowledge have brought this project to life.

I extend my deepest gratitude to my funders, the learners, principals, teachers, and caregivers who contributed in this study. Thank you for your cooperation and patience throughout.

To my parents, Maube and Ngwanaditle Moswathupa and my siblings Natasha and Priviledge, thank you for always being my source of peace and resilience in this world. Your prayers and words of encouragement have kept me going.

My incredible friends, Amukelani, Kgaogelo, Pretty, and Shantel, your constant support, prayers, and love have carried me through this journey. Thank you for inspiring me, being my research assistant, being accommodative, and being patient throughout. To my partner, thank you for walking this journey with me, believing in me, and creating a supportive environment when all was falling apart.

Lastly, my amazing team at Daveyton Main Clinic, thank you for your support and your words of encouragement.

This project received financial support from the South African Centre for Digital Language Resources (SADiLaR). SADiLaR is a research infrastructure established the Department of Science and Technology of the South African government as part of the South African Research Infrastructure Roadmap (SARIR).

ABSTRACT

Background: Augmentative and alternative communication (AAC) systems has been used widely to improve the communication of individuals with reduced or no functional speech. Customising AAC systems for children with limited literacy skills requires others to pre-select a restricted number of words as their vocabulary, which can be challenging. Various resources, including core vocabulary lists, have guided the vocabulary selection process. Core vocabulary are words that are frequently used in conversations amongst a group of people. These words are language-specific and are influenced by geographical features. So far, there is only one Sepedi core vocabulary list based on speech samples collected in the Capricorn district in Limpopo province, South Africa. Thus, the current study aims to determine the core vocabulary of Sepedi-speaking Grade R learners in the Sekhukhune district in Limpopo province. This core vocabulary list can provide dialect-specific words and supplement the existing Sepedi core list when designing AAC systems.

Methods: Six preschool children from two schools without speech/language difficulties were audio-recorded during regular school activities. Over the period of a week, their language samples were collected using small body-worn audio recording devices. The samples were then transcribed and tagged using Microsoft Word™. The data was analysed using Microsoft Word™ and Microsoft Excel™.

Results: The combined sample comprises 19,316 intelligible words and 1,068 different words were identified. The core vocabulary list consists of words with a minimum frequency of 0.05‰ and used by at least half the participants. This resulted in a core vocabulary list of 255 words. The 255 core words accounted for 88.7% of the composite sample.

Conclusions: The results are consistent with findings in the existing literature on core vocabulary in other languages, as it comprises a small set of words that are commonly and frequently used. The Sepedi core vocabulary list compiled in this study can serve as a vocabulary resource when designing AAC systems for Sepedi-speaking children from the Sekhukhune district and may be used in conjunction with the existing Sepedi core words list to provide a core vocabulary resource for the larger Sepedi speaking population.

Keywords: Augmentative and alternative communication, core vocabulary, preschool children, Sepedi, vocabulary selection

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Meaning
AAC	Augmentative and Alternative Communication
CCN	Complex Communication Needs
LoLT	Language of Learning and Teaching
TTR	Type-Token Ratio

LIST OF APPENDICES

1. PROBLEM STATEMENT AND TERMINOLOGY

1.1 Problem statement

In children without disabilities, the acquisition of expressive and receptive language development occurs mostly in a seamless and natural manner. Children are exposed to spoken communication within their environment (Beukelman & Light, 2020; Laubscher & Light, 2020) as they interact with and observe others. Over time, children come to understand and use the words they hear and build up a large vocabulary that they can use to communicate about topics of interest to them. When it comes to children with complex communication needs (CNN) who require augmentative and alternative communication (AAC), however, expressive language development proceeds rather differently. Until such children are literate, their expressive vocabulary is typically preselected by others, most often AAC specialists and/or speech-language therapists. When it comes to graphic symbol-based aided AAC systems, vocabulary selection can be especially challenging. A balance must be struck between providing enough words to enable a child to communicate about desired topics and to expand their expressive language, and providing a reasonable number of words to keep learning and navigation demands within reasonable limits. A variety of vocabulary selection strategies have been noted in the literature, for example, relying on core vocabulary lists compiled from spoken language corpora in specific contexts, using lists from other AAC experts and informants, as a guide for selecting the appropriate vocabulary to aid the child's interaction and improve language development (Bean et al., 2019).

Previous studies by Murray et al. (2019) and Thistle and Wilkinson (2015) show that clinicians often include core vocabulary when designing aided AAC systems for children who are not yet literate. This is supported by the findings from (Dada et al., 2017), who reported that over 50% of speech therapists in South Africa introduce core vocabulary in their initial vocabulary selection when customising an AAC system. Core vocabulary comprises of a variety of words which are frequently used by speakers of a specific language across environments. These words have been described as the most central to the language (Beukelman & Light, 2020). Typically, core vocabulary consists of 200–300 words that represent about 80% of the words used within typical interactions. Core vocabulary is typically made up of a variety of parts of speech, including words from closed and open word classes (van Tilborg & Deckers, 2016). Core words are typically not topic specific. Core vocabulary words are presumed to be useful on an AAC system as they represent a relatively small pool

of words that are nevertheless useable across settings (Deckers et al., 2017). These words may also support individuals with CNN to combine words to produce phrases and sentences (Mothapo et al., 2021), thereby expanding their ability to produce novel meanings. Additionally, core vocabulary may assist communication partners in facilitating modelling and creating communication opportunities in different settings (Soto & Cooper, 2021).

There are a variety of core vocabulary lists in various languages such as English, Isizulu, Afrikaans, and French, however, due to the complexity and diversity of the different language structures, one cannot assume that the commonality and frequency of a specific word will be the same across different languages (Soto & Cooper, 2021). Similarly, a core vocabulary list based on a corpus collected from one specific sample of participants from one specific geographical area is unlikely to be representative of the core vocabulary of all speakers of the language. Hattingh (2018), for example, compared five English core vocabulary lists collected from different geographical areas, where 183 different words were compared and only 26% of these words were found in all five lists. Although some of the differences could be attributed to different transcription methods, others could be due to samples being collected in different areas, from different age groups and during different activities. Therefore, although overlap is expected, it is likely that differences can also be expected.

Currently, one Sepedi core vocabulary list exists, established by Mothapo (2019) and Mothapo et al. (2021) and is based on language samples collected from six participants residing in the Capricorn district in Limpopo province, South Africa where the most prevalent dialect is Kopa (Mojela, 2013). While this is a laudable start, the extent to which the list is representative may be limited due to the small sample size. Mothapo et al. (2021) collected a sample of 17 569 words which resulted in a total of 226 core vocabulary words. As the existing core vocabulary list is based on speech samples of children from the Capricorn district, it might be difficult to apply it to AAC users in the Sekhukhune district. Therefore, supplementing this list with additional word frequency data collected from additional participants from a different geographical area, may increase the external validity of the Sepedi core vocabulary list. This work would contribute towards addressing the need to develop AAC and speech- and language resources for so-called low resource languages (Nekoto et al., 2020) and for underrepresented and marginalised populations (Henrich et al., 2010; Kathard et al., 2011; Pascoe et al., 2013).

1.2 Terminology

This section defines the important and frequently used terminologies in the study in alphabetic order to improve the reader's understanding of the specific concepts used within the study. Each terminology is discussed below:

1.2.1 Augmentative and alternative communication

AAC focuses on using other forms of communication to either replace or enhance natural speech (Beukelman & Mirenda, 2013). These methods provide individuals with communication difficulties with skills to express themselves and engage in conversations within their society. AAC methods include the use of aided systems (e.g., graphic symbol-based electronic or non-electronic systems) and the use of unaided systems such as gestures, manual sign language, and facial expressions (Hall et al., 2022).

1.2.2 Content words

Content words refer to open-class words with specific meanings such as nouns, main verbs, adverbs, and adjectives (Trembath et al., 2007). They provide semantic values without linking them to other words to be meaningful (Tsai, 2023).

1.2.3 Core vocabulary

Core vocabulary is a set of words that are small in number, are not influenced by changes in the environment, and are frequently and commonly used among peers of the same age group (Banajee et al., 2003). These words cover a larger portion of everyday conversation.

1.2.4 Dialect

According to Crystal (2011), a dialect develops when there are regional and social influences on a standard language. These influences come about because of geographic barriers that separate a group of people or if there is any social division in the group. These influences alter the pronunciation and accent of that language.

1.2.5 Fringe vocabulary

This refers to words that are specific to an individual or an activity (Beukelman & Light, 2020). These words reflect an individual's interests, beliefs, and personalities (Wofford et al., 2022), and the words are topic-specific.

1.2.6 Function words

These are also called structure words. They do not have semantic meaning; however, they fulfil a grammatical function (Hattingh & Tönsing, 2020), thus, they link words in a sentence (Shin & Hill, 2016).

1.2.7 Grade R learner

This term refers to children enrolled in Grade R (i.e., 'Reception')—the first compulsory year of schooling in South Africa (Department-of-Basic-Education, 2022). In the current study, this term is used to refer to the participants at times to highlight that they were all enrolled in Grade R. However, the terms 'participant' and 'child' will be used for ease of reading when the purpose is to highlight that these were children participating in the study.

1.2.8 Graphic symbols

These refers to graphic elements such as icons, line drawings, and images used in the AAC field to convey a message. However, they share little similarities with the morphological aspects of spoken languages (Smith, 2006).

1.2.9 Individuals with complex communication needs

The term 'complex communication needs' (CCN) refers to individuals with speech production difficulties or limited communication skills who experience limited participation in their everyday activities (Light & McNaughton, 2012). This includes children with developmental delays and neurodevelopmental disorders who have restricted participation.

1.2.10 L1 speaker

This term refers to individuals who are the first speakers of a language.

1.2.11 Number of different words

'Number of different words' refers to the total number of unique words found in the sample.

1.2.12 Root word

The most basic unit of a word that carries the lexical meaning on its own. These base forms may change their meaning by adding prefixes and suffixes to form new words (Yurtbaşı, 2015).

1.2.13 Total number of words

The total number of words is the overall number of words used by participants in the study. Two counts were done in this regard, a count including unintelligible words (total number of words including unintelligible words) and a count excluding unintelligible words (total number of words excluding unintelligible words). The final analysis was based on the total number of words excluding unintelligible words.

1.2.14 Vocabulary selection

This is a process of carefully selecting vocabulary that will assist children with communication difficulties to meet their daily communication needs across different context. This process is crucial as it impacts the child's ability to develop language skills and overall communication (Bean et al., 2019).

2. LITERATURE REVIEW

This section aims to present background literature that is relevant to the current study. The literature on graphic symbols used in the field of AAC, on vocabulary selection in the field of AAC, on core vocabulary and its relation to language development in children with CCN, on the Sepedi language, and on the development of core vocabulary lists within the South African region, will be discussed.

2.1 Augmentative and alternative communication and graphic symbols

Spoken language offers children without speech and language difficulties a medium to engage and interact within their society. AAC offers children with CCN an alternative to speech/spoken language to express themselves and communicate with their partners. AAC is a division of clinical practice that aims to address the needs of individuals with CCN (American Speech-Language-Hearing Association, 2023). AAC systems provide a way to support communication among different individuals through aided and unaided methods (Hall et al., 2022). AAC is used by a wide variety of individuals including children with neurodevelopmental disorders, adults with acquired speech and language disorders, children who are at risk for speech and language difficulties, as well as those with articulation or fluency challenges (Beukelman & Light, 2020). These systems are used to aid both comprehension and expressive communication, and they afford individuals a means to express their thoughts and feelings and form interpersonal relations. AAC positively impacts children and their families' interactions (Muttiah et al., 2022). Early implementation of AAC in children with CCN is important for the development of their receptive and expressive language skills as well as their literacy skills at a later stage.

AAC systems can be either aided or unaided (Sevcik et al., 2004). Aided AAC systems consist of the use of external support or devices to exchange messages (Hall et al., 2022). Examples of aided AAC systems include paper-based systems that provide visual referents to express oneself, like communication boards, picture exchange systems, and alphabet boards (Hall et al., 2022), as well as electronic systems such as speech-generating devices that are battery/electricity powered. Unaided AAC systems are body-based and do not require extrinsic support, for example, using signs from official sign languages, facial expressions, eye movements, gestures, and all-natural nonverbal forms of communication (Moorcroft et al., 2019). For all these aided and unaided AAC systems to be effective, they need to be appropriate

to the individual's environment, relevant according to their needs and capabilities, and evolve and adjust over time (Hall et al., 2023).

Graphic symbols are commonly used to represent meanings on aided communication aids for individuals who are not (yet) literate (Loncke, 2020). Many times, one graphic symbol represents one concept, word, or message. Graphic symbol-based systems require vocabulary to be preselected and organised across the pages of the system/device. Children with CNN depend on others to preselect the relevant words (represented by graphic symbols). To use the system, the child needs to remember where the relevant graphic symbol is located on the communication board or within the electronic system, and needs to locate and somehow select or indicate the symbol to communicate the word or message. This process places high demands on the child's working memory, as they navigate the pre-selected vocabulary and select the appropriate graphic symbol, while maintaining the conversational interest of their partners. As a result, a graphic symbol-based vocabulary is typically smaller than the spoken language vocabulary of typically developing peers (Stadskleiv et al., 2022). It is, therefore, necessary to prudently consider which words to select.

2.2 Vocabulary selection

The selection of vocabulary is an important task in the intervention process, where the AAC team (person in need of AAC, professionals, and communication partners) pre-select words, symbols, and phrases that will assist an individual in engaging in meaningful conversations across contexts. The vocabulary should also enable the individual to fulfil a variety of communication needs, including developing interpersonal relationships, expressing needs and wants, and expressing interests (Trembath et al., 2007). Pre-selecting vocabulary for preschool children is challenging as they may not be able to participate in the selection process (Trembath et al., 2007), thus making it more time-consuming and complicated. Pre-selected vocabulary should be age-appropriate, culturally, and linguistically appropriate, and support the language development of individuals who need AAC (Johnson et al., 2017; Laubscher & Light, 2020). The intended context, the intended time span for usability, and its ability to facilitate interaction and appropriate grammatical structures should be considered when selecting vocabulary for pre-literate individuals who use AAC (Bean et al., 2019). When the pre-selected vocabulary is inappropriate or insufficient, children with CNN may experience discrepancies between the available vocabulary on their aided system and the vocabulary they would like to express, thus impacting their language development (van Tilborg & Deckers, 2016). Vocabulary selection for children who are not yet literate should involve their direct

communication partners to ensure the vocabulary supports conversational interactions within their environment.

As mentioned, the vocabulary introduced should afford the child the support to engage in conversations and facilitate language skills. The vocabulary must ensure that the individual with CCN is able to convey messages and respond to conversational demands throughout their daily activities, and the vocabulary should introduce words that are new to the user to target semantics and morphological skills (Beukelman & Light, 2020). The vocabulary selected should include functional words, content words core vocabulary, and personalised vocabulary to afford the user the ability to attain communication competence (Beukelman & Light, 2020). There are various techniques suggested to facilitate vocabulary selection. According to Mngomezulu et al. (2019) and Trembath et al. (2007), these techniques include asking informants such as communication partners, referring to published core vocabulary lists, and using environmental inventories to determine communication demands. Each of these techniques has advantages and disadvantages, however, all the suggested techniques should culminate in the introduction of vocabulary that facilitates social interaction and supports language development (Beukelman & Light, 2020).

Consulting with informants such as caregivers, speech therapists, and teachers to compile word lists, offers valuable daily vocabulary including personalised words that are familiar to the individual needing an AAC system (Beukelman & Light, 2020). Other techniques include using vocabulary checklists based on existing inventories such as language-specific core vocabulary lists or the MacArthur-Bates communicative developmental inventories (Fenson, 2007) that are based on language skills of children without disabilities. These lists have been found helpful by different team members to facilitate effective communication (Beukelman & Light, 2020). The AAC team must ensure that the vocabulary checklists selected are relevant to the individual based on age, culture, and environmental demands. An environmental inventory entails the AAC team to identify critical words that peers without communication difficulties use in specific activities (Beukelman & Light, 2020). However, these lists consist of activity-based vocabulary that may be difficult to use or are irrelevant in other communication settings (Johnson et al., 2017). Introducing a vocabulary that is only relevant in one context but not in others where the child needs to communicate, may result in the abandonment of the AAC system and exclusion from societal interactions.

Using more than one source of vocabulary has been recommended to mitigate the limitations that each individual technique entails (Bean et al., 2019). This will allow communication partners to validate the proposed vocabulary socially (Johnson et al., 2017) and

expose the child to a variety of vocabulary that is relevant in the home and school environment. As mentioned, many service providers report on including core vocabulary in the systems they provide to individuals who use AAC (Beukelman & Light, 2020; Dada et al., 2017).

2.3 Core vocabulary

(Beukelman & Light, 2020) refer to core vocabulary as words that are commonly and frequently used by individuals to communicate with each other in different settings. While exact operational definitions vary, most core vocabulary lists contain approximately 200–300 of the most commonly and frequently used words as extracted from the corpora of spoken language (Van Tilborg & Deckers, 2016). In addition, Van Tilborg and Deckers (2016) allude that core vocabulary refers to words used frequently by a specific demographic group and that they have widespread usage across situations. Researchers from different contexts have used conversational samples of individuals without communication difficulties to develop core vocabulary lists that will best suit the specific CCN population in terms of cultural and linguistic domains (Mothapo et al., 2021). For example; there are core vocabulary lists based on the various languages such as French (Robillard et al., 2014), Mandarin (Tsai, 2023), and Korean (Shin & Hill, 2016). Moreover, the sample size, age, location, and sampling context may all influence the resulting core vocabulary list. This is evident in the study by Hattingh (2018), where the author compared the first 100 words of the four existing English core vocabulary lists, namely: (Beukelman et al., 1989; Boenisch & Soto, 2015; Stuart et al., 1993; Trembath et al., 2007). All the aforementioned studies are based on different geographic locations and populations. Beukelman et al. (1989) examined the core vocabulary of six English preschoolers from Nebraska, while Boenisch and Soto (2015) focused their study on 22 school-aged native English speakers and eight English second language speakers in San Francisco Bay; the two abovementioned studies are based on populations in two different areas in the United States. Furthermore, the study by Trembath et al. (2007) is based on English-Australian preschoolers. Stuart et al. (1993) compiled a core vocabulary list based on adult native English population from Nebraska. The comparison indicated only 26% of words were common across all the lists. This indicates the influence of geographic region, the age of participants, and the environment on core vocabulary.

Core vocabulary lists in various languages consist of a mixture of content and function vocabulary (Tsai, 2023). Typically, many function words (including conjunctions, prepositions, and pronouns) are found in the top 50–100 words (Van Tilborg & Decker, 2016). Function words (also called structure words) have little semantic meaning; however, they fulfil

a grammatical role. Examples of such words are conjunctions, prepositions, concords, and particles. Examples in English include words like '*after'* and '*we*', in Sepedi '*go*' and '*ka*', and in Isizulu '*kodwa*' and '*ngoba*'. Such words are typically overlooked when informants suggest words for inclusion in the system, as they are abstract. Consulting core vocabulary lists when selecting vocabulary can help to counter this trend and to move away from over-representing content words in AAC systems, as they are mostly topic-specific (Bean et al., 2019). According to Hattingh and Tönsing (2020) and Mngomezulu et al. (2019), core vocabulary lists are effective for supporting the development of early syntactic skills (Mothapo et al., 2021). Various researchers have suggested the use of core vocabulary when selecting vocabulary for AAC systems, as they can be taught across settings, are smaller in number, and reduce cognitive and motor demands (Van Tilborg & Deckers, 2016). Beukelman and Light (2020) and van Tilborg and Deckers (2016) argue that an AAC system should not only be compiled using core vocabulary, but also include the individual's personal fringe vocabulary to help them meet their daily communication needs across context.

2.4 Core vocabulary and language development

Banajee et al. (2003) conducted a core vocabulary study with toddlers without disabilities and found that the most frequently and commonly used words were function words and verbs rather than nouns. (Laubscher & Light, 2020), however, stated that a prominence of function words does not support early language development according to the Communication Development Inventories (Fenson, 2007). Furthermore, clinicians preferred to introduce nouns and verbs during AAC intervention as they are easier to represent on aided systems (McFadd & Wilkinson, 2010; Mngomezulu et al., 2019). Mothapo et al. (2021) argued that a preponderance of nouns limits the individual who uses AAC to single symbol utterances which focus only on content words and neglect morphological and syntactical development. The inclusion of core vocabulary with its various parts of speech (e.g., verbs, pronouns, prepositions) can facilitate the transition from single-word utterances to forming phrases and clauses (Bean et al., 2019; Mothapo et al., 2021). Vocabulary selection has direct implications for the expressive language development of children using AAC. When the AAC team selects vocabulary that is relevant for a longer period and is usable across multiple settings, it will increase exposure to the words and consequently, the likelihood that children with CCN will learn these words (Bean et al., 2019). Exposure to this relevant vocabulary over time, may result in an expanded semantic representation of these words, enabling the child to use them to fulfil a variety of pragmatic functions during communication interactions (Bean et al., 2019).

Clinicians may use communication developmental inventories as resource tools to guide vocabulary selection that supports language development for young children. Frick Semmler et al. (2023) further suggest that speech therapists must select core vocabulary with caution and consider the language developmental norms when selecting the vocabulary. Discretion may, therefore, be needed when introducing core vocabulary as part of an aided AAC system. Although core vocabulary is essential, children also need access to vocabulary that facilitates language development. Thus, the need to use both the core vocabulary list and Communication Development Inventories (Fenson, 2007), as well as the individual's fringe words are essential in guiding vocabulary selection. Furthermore, access to both core and fringe vocabulary supports ongoing language development in terms of lexicon and grammatical diversity (Binger et al., 2024).

2.5 Core vocabulary lists in South Africa

South Africa has an estimated population of 62.0 million (Statistic S.A., 2022) with 12 official languages attesting to high linguistic and cultural diversity. As per the latest Census 2022, the six most widely spoken languages of the 12 official languages are Isizulu 24.4%, IsiXhosa 16.3%, Afrikaans 10.6%, Sepedi 10%, English 8.7%, and Setswana 8.3% (Statistics, 2022). Indigenous languages are mostly preferred as home language for the South African population (Mesthrie, 2002). Given the linguistically and culturally diverse population, speech-language therapists are often required to provide AAC services to individuals whose home language and culture differs from theirs, thus, making the selection and customisation of an aided AAC system challenging in terms of the symbols and vocabulary to include (Tönsing et al., 2018). There are limited resources to guide vocabulary selection for graphic symbol-based AAC systems in the South African context (Pascoe & Norman, 2011).

Core vocabulary lists that are relevant to the South African context have been established for only four of the 11 official spoken languages. Currently, lists exist only for isiZulu (Mngomezulu et al., 2019), Afrikaans (Hattingh & Tönsing, 2020), Sepedi (Mothapo et al., 2021), and Setswana (Mogatusi, 2022) populations. While this is a laudable start, the external validity of these lists is limited due to the limited number of participants from whom speech samples were collected. The developers of the core vocabulary list in isiZulu (Mngomezulu et al., 2019), Sepedi (Mothapo et al., 2021), and Setswana (Mogatusi, 2022) based their lists on speech samples from six participants (Grade R learners) each, while the Afrikaans core vocabulary list is based on 12 participants in Grade R (Hattingh & Tönsing, 2020). In addition, the participants in each study resided in a specific geographical location,

speaking a specific regional dialect of the respective languages. For example, the participants in the study by Mothapo et al. (2021) came from the Capricorn district where the most prevalent Sepedi dialect is Kopa (Mojela, 2013). This might limit the generalisability of the results to a larger population. Collecting speech samples for additional participants from different regions may strengthen the external validity of the core vocabulary lists.

2.6 Sepedi language

Sepedi is one of the Southern African languages from the Sotho-Tswana language group (Prinsloo, 2014). The Sotho-Tswana cluster shares some level of linguistic features. Existing studies have shown some discrepancies in the official name of the language, where it is either referred to as Northern Sotho, Sesotho, Leboa, or Sepedi (Rakgogo & van Huyssteen, 2018), however, for the purpose of this study the term Sepedi will be used as the official name. Sepedi is spoken across all nine of South Africa's provinces, however, it is frequently spoken in three of the nine South African provinces, namely Limpopo, Mpumalanga, and Gauteng (Statistics S.A., 2022). Sepedi is the most prevalent first language (55.5%) in the Limpopo province (Statistics S.A., 2022). Sepedi is predominantly spoken by at least 80.9% of the population in the Sekhukhune district in the Limpopo province (SekhukhuneDistrictMunicipality, 2017).

Mojela (2013) highlighted that there are approximately 30 dialects in the Sepedi language. The dialects include Pedi, Matlala, Lobedu, Hananwa, Mphahlele, Kwena, Birwa, Kone, and others (Rakgogo & Zungu, 2022). Dialects are standard language variations that are expressed as a different pronunciation of certain words as well as accent (Crystal, 2011). Due to inequalities and lack of representation on the Language Board of 1963, various Sepedi dialects were excluded during the standardisation of the language (Mojela, 1999). This resulted in speakers of other dialects being forced to learn the standard Sepedi, more so for educational purposes (Mokgokong, 1966) and that resulted in an inferiority complex in the first language speakers of the excluded dialects (Mojela, 1999). These dialects are specific to geographical regions within the provinces. Some show influences of nearby languages, for example, Lobedu is influenced by Tshivenda, and Pulana is influenced by Xitsonga (Mokgokong, 1966).

According to Mokgokong (1966), the North-Western Sotho dialects, such as Kopa, Mphahlele, Matlala, and Mamabolo are prominently spoken within the Capricorn district. In contrast, Central Sotho dialects, such as Pedi, Kone, and Tau are mostly spoken in the Sekhukhuneland. These two groups of dialects share a large amount of common linguistic features such as vowel phonemes. However, there are some words and phrases that are used in

the Kopa dialects that will be unintelligible to the Pedi speakers (Mokgokong, 1966). These variations include mostly the phonological and syntactical variations, and fewer on the lexical level, which was documented in length by (Mokgokong, 1966), where the author explained that words may acquire a different meaning or connotation in a certain dialect as compared to others.

The Sepedi language orthographic representation was brought about by missionaries from European countries in an attempt to translate the Bible (Prinsloo, 2014). Their orthography has been criticised as being based largely on Pedi and Kopa, or a mixture of the two (Mojela, 2008; Rakgogo & Zungu, 2022). Mokgokong (1966) emphasised that since the Pedi and Kopa mixture dialect was the first abstract system of the language, it was adopted by various writers. However, some of those words do not represent the typical Pedi dialect. This resulted in discrepancies between the spoken language and written texts within the societies. Due to the poor representation of speakers of various dialects on the Sepedi Language Board (Mojela, 1999), the standardised language is based on only a few dialects. These are the dialects spoken within the Sekhukhune district, the southern part of the Capricorn district, and the south-eastern part of the Waterberg district (Mojela, 2013). Those dialects have influenced the orthographic components of the Sepedi language (Mojela, 2013; Mokgokong, 1966). Furthermore, Prinsloo (2014) identified that some Sepedi linguistic phenomena have been excluded from the standard language and are not classified, which creates challenges when learning and teaching those words which are frequently used by L1 speakers in specific areas. For example, the Kone tribe use the word '*ka moswane'* which translates to '*tomorrow*' in English, which is not represented in the standard language.

2.7 Summary

This section highlighted the importance of core vocabulary lists as a resource when customising aided AAC systems. Core vocabulary is language and region-specific. The existing core vocabulary lists of four of the 11 official spoken South African languages were also reviewed. Lastly, the Sepedi language structure and the need to establish regional-specific core word lists to strengthen the generalisability of the existing Sepedi list were also discussed.

3. METHODOLOGY

In this chapter the main aim and sub-aims of the study are outlined, the study design applied is explained, and the participants involved are described. The materials and equipment used are highlighted as well as the details regarding the procedures followed during the pilot study, data collection, and data analysis. Lastly, the ethical considerations and principles upheld during this study are described.

3.1 Research aims

3.1.1 Main aim

The main aim of the study is to establish a core vocabulary list based on language samples from Sepedi-speaking Grade R learners in the Sekhukhune district collected during school activities and to compare it to the core vocabulary list established by Mothapo et al. (2021).

3.1.2 Sub-aims

The sub-aims of the study were:

i. To describe the speech sample by total number of words, number of different words and type-token ratio (TTR);
ii. To determine a core vocabulary list for spoken Sepedi from the Sekhukhune district by identifying the most frequent and common words used by Sepedi-speaking Grade R learners during school activities;
iii. To differentiate content and function words contained in this list;
iv. To describe the word list according to parts of speech; and
v. To compare this list to the existing Sepedi core vocabulary list established by Mothapo et al. (2021).

3.2 Research design

A quantitative, non-experimental descriptive observational design was implemented in the current study. This design was appropriate for the descriptive nature of the research question, which required the observation of a phenomenon that is taking place naturally, without manipulation of variables by the researcher (Thompson & Panacek, 2007). In this study, spoken language samples were obtained of the participants through audio recordings of

their speech in their natural environment (Omair, 2015). A typical threat to the internal validity of observational data is reactivity of the participants due to being observed. The researcher met with the participants prior to collecting the data to build a rapport and walk them through the equipment. In addition, the researcher limited her visibility around the schools to ensure that participants are less reactive, and the first 20 minutes of recordings were not transcribed.

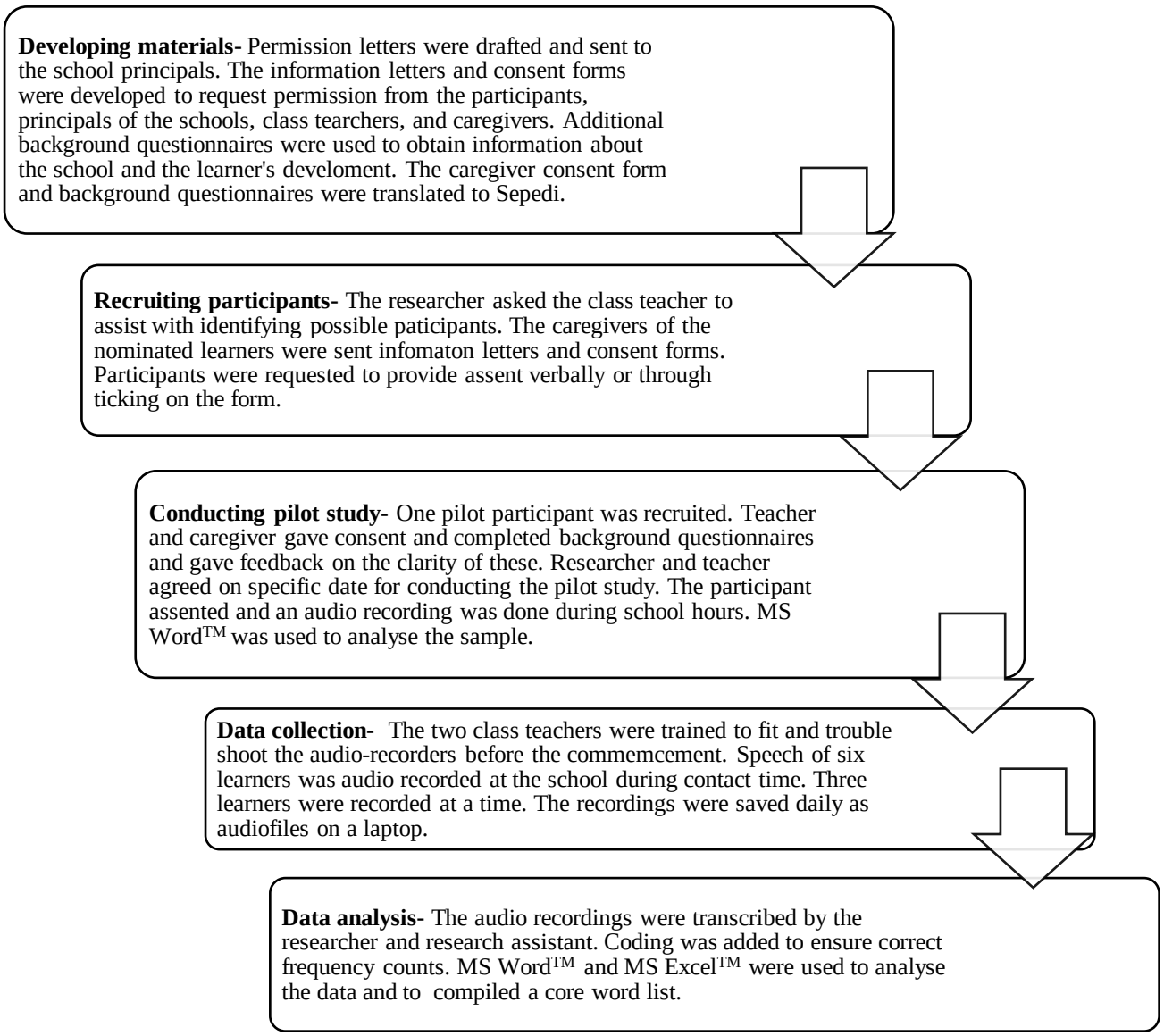
In this study, as in many other observational designs, the data collection and analysis were time-consuming. For this reason, only a limited number of participants were included, which may have limited the degree to which the sample was representative of the study population, affecting external validity of the study (Thompson & Panacek, 2007). However, the strict selection criteria applied afforded the researcher participants that resembled the proposed population. Further, a relatively large corpus was collected from each child, increasing the representativeness of the corpus.

3.2.1 Stages of the study

Figure 1 below details the process that was undertaken to conduct the study.

Figure 1

Stages of the study



3.3 Study setting

The study took place in two rural public schools in the Sekhukhune district in the Limpopo province, the northernmost province of South Africa. The Sekhukhune district consists of four local municipalities. Both schools were situated in the Makhuduthamaga local Municipality, approximately 30 km from each other. The Makhuduthamaga Municipality has the second largest population in the district and numbered 340 328 people in 2022 (StatsSA, 2022). The two schools were classified as Quintile one as the learners did not pay any school fees and received stationery and textbooks from the schools (Ogbonnaya & Awuah, 2019). Learners in public schools are expected to pay a fee to assist with the day-to-day expenses of the school, however, due to the poor economic status in most communities, some of these

learners were unable to afford schooling (Department of Basic Education, 1996). Nevertheless, to improve access to education for all, the government classified the schools based on the economic levels of the surrounding communities. These schools are classified from the poorest (Quintile one) to the least poor (Quintile five) schools. Learners from Quintile one to three are exempt from paying school fees. The teachers at the schools follow the national Curriculum Assessment Policy Statement introduced by the Department-of-Basic-Education (2011) and used Sepedi as the language of learning and teaching (LoLT); most of the learners' home language was also Sepedi. Both schools had access to electricity and sanitation facilities, however, they had interrupted water supply, no access to the internet and landline, and learners had no access to a playground within the institution. During breaktime the learners spent their time on the school premises, playing under the trees interacting with other learners, while others preferred to sit in the classroom.

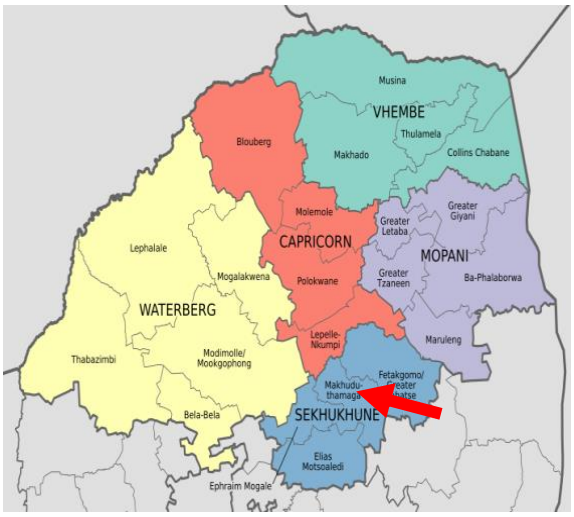
Figures 2 and 3 below, show the Sekhukhune district and the Makhuduthamaga local municipality where the study took place.

Figure 2

The Sekhukhune district (highlighted in red)

Figure 3

The Makhuduthamaga local municipality (see red arrow)



Source: Maps copied from Wikipedia (**https://commons.wikimedia.org/wiki/File:Map_of_Limpopo_with_municipalities_named_and_districts_shaded_(2016).svg**)

The schools' setting comprised of both schools having one Grade R classroom, and both schools followed the Curriculum Assessment Policy Statement for teaching and learning. At Site 1, the classroom accommodated 35 learners where the age of the youngest learner was

5;0 (years; months) and the age of the oldest learner in class was 6;10. The teacher and learners predominantly communicated in Sepedi, even during lesson time, however, due to the exposure to English, learners will code switch some words during their conversations. Site 2 has 34 learners in the classroom, the youngest learner's age was 4;11 and the age of oldest learner was 5;11. The teacher used Sepedi and English interchangeably when communicating with the learners (this is a school initiative to improve literacy skills in English). The learners predominantly spoke Sepedi amongst each other with a few instances of code-switching to English.

3.4 Participants

3.4.1 Participant sampling and recruitment

For the purposes of the study, approval was obtained from the Faculty of Humanities at the University of Pretoria (Appendix A) and the Limpopo Department of Basic Education (Appendix B). Two public primary schools with Grade R learners in the Makhuduthamaga local municipality within the Sekhukhune district were identified based on a combination of purposive and convenience sampling (Etikan et al., 2015). The schools selected were closer to the researcher's place of residence and were approximately 30 km away from each other. Both the schools used Sepedi (and specifically the Sekhukhune dialect) as the LoLT, and this sampling may reduce the likelihood of dialect variations amongst the participants.

The principals of the schools were contacted, and permission was requested to conduct the study involving the Grade R learners on the school premises during school time. The researcher also requested that classroom teachers assist with completing a questionnaire and identifying and recruiting suitable participants. The principals were provided with information letters pertaining to the study processes and intent. They were afforded the opportunity to grant or decline permission for the study in writing by means of a consent form appended to the information letter (Appendix C).

The Grade R class teachers were approached and provided with information letters (Appendix D), detailing all study aspects. They were invited to be a part of the study, and their consent was solicited via a consent form appended to the letter. Upon providing written consent, teachers were requested to purposefully select three learners from their class who met the selection criteria, and whom they deemed to be talkative children. The inclusion of talkative has no necessary influence on their vocabulary, these children were expected to be using similar language structures with their peers. The reason for asking teachers to select talkative children was that speech samples of 3 000 words would be gathered in the shortest possible period,

thereby minimising the data collection time and any inconvenience that may have been caused. This approach also follows the precedent set by other core vocabulary studies based on the recordings of preschoolers' speech (Trembath et al., 2007). Furthermore, the teachers were requested to fill in a questionnaire that provides background information about the preschool (Appendix E). The teachers sent information letters and consent forms to the legal guardians/parents of the selected children. Etikan et al. (2016) highlighted that purposive sampling allows the researcher to select participants based on the qualities they possess. One teacher was requested to select two boys and one girl, and the other teacher was requested to select two girls and one boy, to achieve a balance in gender.

The caregivers of the nominated learners were provided with information letters explaining the study and consent forms (Appendix F) requesting proxy consent to allow their children to take part in the study. Furthermore, the consent forms and information letters were made available for caregivers in English and Sepedi. The caregivers who consented to their child's participation were requested to return the signed consent forms granting permission to the class teacher and were further requested to fill in a questionnaire to help determine the child's developmental background (Appendix G).

To request assent from the children, the nominated learners whose caregivers consented to their participation met the researcher individually, with their teacher present, on a scheduled day at the school. The researcher introduced herself and spent some time building a rapport with the child. She then explained the study verbally in Sepedi according to a script and pointed to the pictures contained on the script (Appendix H) to scaffold the child's comprehension. The researcher then read a set of questions off a response form (Appendix I), scaffolding the questions with pictures contained on the form to support comprehension. The questions were intended to ensure that children had understood all parts of the study as well as their rights, and to ask whether they agreed to participate. Children were encouraged to answer each question verbally and by marking their option of choice ('yes' or 'no' represented with pictures) with a pencil.

3.4.2 Selection criteria

In this study, the selection criteria set out in Table 1 below, were applied.

Table 1

Participant selection criteria

Criterion	Justification	Measure used
Participants should be aged 5;0 (years: months) to 6;11 at the start of data collection.	Speech and language skills are fairly mature at this age, and grammar skills are well-developed (Owen & Leonard, 2002).	Caregiver questionnaire (Appendix G)
Participants should not have any language delays or impairments	The aim is to obtain speech and language sample from children without disabilities	Caregiver questionnaire (Appendix G) preschool background questionnaire (Appendix E)
Participants should be attending Grade R at a public school.	Participants in Grade R were included as the Grade R curriculum still allows for more semi-structured and unstructured activities that may lead to more child-initiated and child-to-child interactions, rather than teacher-led interaction (Department of Basic Department-of-Basic-Education, 2011). The speech samples collected were, therefore, less likely to be overly influenced by the teacher or the learning material and themes.	Preschool background questionnaire (Appendix E)
Participants should be enrolled at the school for at least one month.	The participants should be familiar and comfortable in the environment to ensure that their conversations are not inhibited.	Preschool background questionnaire (Appendix E)
Participants should be based in the Sekhukhune district	The Sekhukhune district was targeted to provide some geographical variation to the previously conducted core vocabulary study (Mothapo et al., 2021).	Recruitment will take place here.
Participants should have Sepedi as their first language and LoLT	This will limit the chances of code switching and mixing, as the Sepedi language (and specifically the Sekhukhune dialect) was the target of study.	Caregiver questionnaire (Appendix G), preschool background questionnaire (Appendix E)

3.4.3 Descriptive criteria

Table 2 below, provides a description of the participants that were involved in this study, including the description of personal factors, as well as their home and school environment. These were not selection criteria, but are provided so that readers can contextualise results in the light of who the participants were.

Table 2

Description of participants

Site	Participant Number	Age	Gender	Frequency of School Attendance	Home Language	Other Languages Exposed to at Home Through Conversations	Languages Exposed to Through Media (TV/radio/audio-visual media on smartphone)	Monthly Household Income
1	1	5;7	Female	Daily	Sepedi	English	English	<7979
2	2	5;2	Female	Daily	Sepedi	English	English and Sepedi	<7979
1	3	6;1	Female	Daily	Sepedi	English	English	<7979
2	4	5;0	Male	Daily	Sepedi	English	English	<7979
2	5	5;6	Male	Daily	Sepedi	English	English	<7979
1	6	5;10	Male	Daily	Sepedi	English	English	<7979

The table above shows that the learners' ages ranged from 5;0 to 6;1. The mean age was 5;5. While all participants use Sepedi as their primary language of communication at home, they were additionally exposed to English through conversations and media. Only one caregiver reported that their child is exposed to Sepedi via media channels. All caregivers reported that their monthly household income is less than the minimum taxable income for the year, which is less than R7 979.00 per month. Regarding socio-economic status, caregivers were further asked to indicate their accessibility to utilities such as water and electricity. All households had access to electricity in their houses, however, of the six households only two had access to running water in the houses.

3.5 Materials and equipment

Materials utilized for the study are described in this section, as is the equipment used. Materials include consent letters, assent script, as well as questionnaires to obtain background information.

3.5.1 3.5.1 Materials

3.5.1.1 Information letters and consent forms

Information letters for school principals, Grade R teachers, and parents were drafted detailing the information about the study and the process that will take place during the study (Appendices C–F). Letters with permission forms were sent to the principals (Appendix C), while letters with consent forms were sent to teachers (Appendix D) and caregivers (Appendix F) prior to the commencement of the study, to give all stakeholders an opportunity to give or decline permission/consent to participate. All the letters were drafted in English. The caregivers were given an additional Sepedi version of the letters and consent forms to ensure that they could comprehend the contents fully. Since teachers and principals are professionals who would have completed tertiary training in English, a Sepedi version of letters and forms was not deemed necessary for them.

All the letters highlighted the nature of the study, what was expected of participants, participants' rights, ethical principles that would be observed, potential benefits of the study, and how data collected was going to be used and stored. The consent forms and information letters sent to class teachers also requested that they assist the researcher with the recruitment process and provide background information of the school.

3.5.1.2 Assent script and response form

A child-friendly script was developed in Sepedi based on and guided by the study conducted by Mogatusi (2022) which focused on the Setswana language. This was used for the purpose of explaining all aspects of the study to potential child participants. The script was supplemented with pictures to enhance understanding (see Sepedi version and English translation in Appendix H—translation provided for the sake of readers unfamiliar with Sepedi). A Sepedi response form was developed (Appendix I—includes English translation) with a set of questions to ensure the children's understanding of all aspects of the study as well as their rights and to request whether a child agreed to participate. Pictures were included to represent each question and the response options ('yes' and 'no').

3.5.1.3 Caregiver questionnaire

Participants' background information was obtained through a caregiver/parental questionnaire, which was developed based on the study by Mogatusi (2022). The questionnaire includes questions pertaining to the selection criteria, such as the participant's age, main language used in the home, and speech-language development (Appendix G), as well as additional descriptive criteria, such as overall development, family composition, and exposure to additional languages. This information was deemed important to understand factors that might be influencing the participants' spoken language. The questionnaire provided was made available in both Sepedi and English to ensure that caregivers did not experience any language barriers.

3.5.1.4 Preschool background questionnaire

A preschool background questionnaire was developed based on the study conducted by Mogatusi (2022) to gather information pertaining to describing the daily preschool routine as well as language practices followed in the classroom (Appendix E). This information was obtained to highlight the LoLT at the schools, the exposure to different languages, and the nature of activities in the classroom (e.g., where they are mostly teacher-led, or learner-led). It also helped readers understand the study setting regarding the teacher-learner ratio in rural public schools.

3.5.1.5 Transcription rules

The set of transcription rules (Appendix J) were compiled based on the rules developed for previous studies by Mogatusi (2022); Mothapo et al. (2021); and Trembath et al. (2007). These rules were used to maintain consistency in the way that audio recordings were transcribed, thereby ensuring the reliability of the results. These rules were applied equally by both the first transcriber and the second transcriber.

3.5.1.6 Tagging rules

A set of tagging rules (Appendix K) was compiled as guidance to ensure that the samples are tagged correctly and consistently to improve the reliability of the findings. The rules were informed by a previous study conducted by Mothapo et al. (2021) and the Oxford *Pukuntšu ya Sekolo* (school dictionary) (De Schryver, 2007). These codes are intended to differentiate homonyms and to ensure that they are counted separately. Additionally, they ensure that different variations of words (past tenses, present continuous tenses, etc.) are related to their root forms and that they are all counted together under the root words.

3.5.2 Equipment

Small digital voice recorders (two Phillips voice tracers DVT 6010, one Olympus voice recorder DS-50, two Phillips voice tracers DVT6110, and one Olympus voice recorder DM 650) contained in body worn pouches were used together with lapel microphones (Audio technical Lavalier Microphones ATR3350iS) to collect the participants' speech samples during their school day. The equipment had two parts: the voice recorders fitted in pouches around the participants' waists; and the lapel microphones attached to the participants' clothing (collar of the shirt; see Figure 4). The recorded audio samples were saved on a laptop for later transcription. The laptop and headphones were used during the transcription of the samples for better audibility and playbacks. The transcriptions were done using Microsoft Word documents (MS Word™). Word frequency analyses were conducted using the word frequency macro on Microsoft Word developer as well as MS Excel™.

Figure 4

Participants on site fitted with the pouches and microphones

3.6 Pilot study

A pilot study was conducted once approval was given for the study by the Ethics Committee of the Faculty of Humanities at the University of Pretoria and the Limpopo Department of Basic Education. The pilot study took place in one of the recruited schools with one participant to ensure that the recruitment processes, selection criteria, equipment, and materials used were safe, appropriate, and effective in obtaining the relevant data. The pilot study aimed to evaluate the planned study design, recruitment processes, and data collection procedures (Lowe, 2019).

A female child aged 6;4 (years; months) was recruited to participate in the pilot study. The participant attended Grade R at the school designated as Site 1. The participant met the selection criteria as stated in Table 1. The teacher facilitated the procedure of sending the consent forms and caregiver background questionnaire home and ensured that they were returned to the school. The participant spoke Sepedi as her home language, and she was exposed to English at home through television and radio. The family has no access to running water and their household income is below the taxable income cut-off in South Africa.

Table 3 below, gives an overview of the aims of the pilot study, the materials and procedures used, the results, and the subsequent recommendations for the main study. The findings of the pilot study provided the researcher with guidance on how to structure the data collection and analysis procedures in the main study. Feedback was provided to the researcher on the clarity and appropriateness of the consent forms and background questionnaires by the participant's teacher and caregiver.

Table 3

Pilot study aims, materials, procedures, results, and recommendations

Aim	Materials	Procedures	Results	Recommendations
To determine the effectiveness of partnering with the class teacher during the selection of participants.	Teacher information letters and nomination sheet (Appendix E)	The researcher provided the teacher with the letter and consent form requesting the teacher to nominate the participant from their classroom.	The teacher selected an appropriate participant.	The same process should be followed for the main study. Teachers will be encouraged to communicate with the researcher when they have any challenges and uncertainty relating to the selection criterion.
To determine if obtaining caregiver consent through the school was effective.	Caregiver consent forms (Appendix F)	The teacher sent the information letter and consent form to the caregiver and requested them to return them to the school.	The researcher sent the letters through the teacher, this was done in person. The researcher had a telephonic meeting with the caregiver to ensure that all the information was understood.	When providing the letters to parents, the teacher will encourage caregivers to contact the researcher telephonically should they have questions about the study.
To determine the comprehensiveness and clarity of the school background questionnaire and caregiver questionnaire.	School background questionnaire and caregiver questionnaire (Appendix E and Appendix F)	The caregiver and teacher were requested to complete the questionnaires.	The teacher was able to complete the questionnaires independently. The caregiver had challenges concerning questions relating to access to water due to the wording used.	The researcher changed the wording on Question 13 of the caregiver questionnaire (from “access to running water in the house” to “access to running water from the taps”).

Aim	Materials	Procedures	Results	Recommendations
To ensure that the child assent procedures were effective.	Child script and assent form (Appendix H and Appendix I)	The child was given information according to a script, and then asked questions to ensure comprehension. Lastly, the child was asked whether she granted assent. The researcher supplemented all verbal explanations with pictures to increase comprehension.	The participant was able to understand all aspects of the study and the provided visual aids seemed to improve understanding. the participant was able to answer all the questions by pointing to the desired response option and the researcher assisted with circling it.	None
To determine the most effective way of fitting participants with the equipment to ensure that it does not pose any harm and that it was effective in recording intelligible speech samples.	Voice recorders, lapel microphones, and pouches	The researcher fitted the equipment on the participant and recorded her speech during a school day. The recording was checked for intelligibility.	The pouch containing the voice recorder, and the lapel microphone were well-fitted around the participant's waist and collar of her shirt, however, the intelligibility of the recording was slightly compromised. This may have been affected by the placement of the lapel microphone and overall background noise in the classroom.	The researcher aimed to ensure that the lapel microphones are fitted in a way that they are free from obstruction by the participant's shirt or hair movements, to ensure good sound quality of the recording.
To determine the comprehensiveness and usefulness of the	Transcription rules (Appendix J)	The audio files were transferred from the recorders to a laptop. The	The transcription rules compiled assisted with reducing the over-	Specific additions were made to the transcription rules to

Aim	Materials	Procedures	Results	Recommendations
transcription rules in ensuring consistency of the samples.		files were listened to through headphones and transcribed into a word document based on the transcription rules.	representation of words that were repeated by the participant and to transcribe songs fairly to ensure that the word count is not skewed.	ensure that unintelligible utterances are easily identified, these utterances will be transcribed as xxx and all the words in the transcription will be in small letters.
To ensure that the tagging rules and the website can be applied successfully to conduct the desired analyses.	Tagging rules (Appendix k)	The transcribed Word file was coded based on the rules. The frequency word website was used to analyse the sample to generate word frequency.	The analyses were successful. TTR, number of different words, as well as frequency counts could be established per word and per root word. Additional tagging rules were required to separate homonyms according to various part of speech.	The word document will be tagged using the tagging rules. The document will be analysed on the Word frequency website. A tagging key is required to keep track of the tagging rules to count root words of morphological variations of verbs and nouns, as well as separating homonyms.
To determine if words could be classified by parts of speech	The Sesotho sa Leboa *Pukuntšu ya Sekolo* (school dictionary) (De Schryver, 2007)	The Ms Excel sheet was used to classify the top 30 words in the pilot study findings based on their root form.	The researcher was able to classify the number of different words based on their root form and calculate the total number of different words.	None

3.7 Procedure

3.7.1 Data collection

As this was an observational study, data was collected in the form of audio-recordings of children's speech. The researcher and teachers agreed on specific dates and set times for data collection at the selected schools. The researcher informed the teacher about the procedures to follow in terms of time of arrival, equipment to be used, and how to troubleshoot any problems with the audio recorders should the need arise. The selected participants were requested to provide or decline assent to be a part of the study, and they were also informed that they could withdraw from the study at any time as they wish. The researcher met with the participants to build a rapport with them, discussed the procedures to follow, and explained the reasons for the procedures. The researcher informed the participants and their classmates not to play with the equipment (audio recorder, microphone, and pouches). Each participant provided assent by either pointing to the pictures or circling the picture with a pencil; they all indicated that they understood the procedures. The researcher reminded the participants of the data collection process and procedures every morning before fitting the participants with the recording equipment.

The researcher arrived at the selected school every morning to fit the pouches with the recorders around each of the participant's waist and clip the microphone on their shirt collar and return in the afternoon to remove the recording equipment. The participants were encouraged to inform their teacher should they wish to have the equipment temporarily removed or withdraw from the study during the school day. The class teachers were requested to monitor the participants to ensure they were comfortable. Teachers were requested to remove or adjust the equipment if they could see that children were uncomfortable. The teachers were able to remove and fit the pouches and microphones when participants wanted to use the bathroom, this was done upon the participant's request. The participants were recorded on consecutive weekdays until the language sample from each participant comprised at least 3 000 words. The data collection process took 4 recording days to attain at least 3 000 words with Participant 6 from Site 1 having a shortfall of 153 words. Due to the shortfall amounting to less than 1% of the composite sample, it was decided that another day of recording was not needed. Table 4 below, provides an overview of the total number of words recorded and the time taken per participant.

Table 4

Total number of words (including unintelligible words) per participant, and number of days taken to record the sample

Participants	Total number of words with Unintelligible Words	Total Number of recording Days
1	3 102	2
2	3 182	3
3	3 801	4
4	3 450	3
5	3 085	4
6	2 847	4

3.7.2 Research assistants

The study involved six research assistants who helped during the transcriptions and tagging of the speech samples for reliability checks. Two assistants were qualified high school teachers who majored in Sepedi, one occupational therapist, one speech therapist, one software engineer, and one professional language translator. All research assistants are Sepedi home language speakers, and five of them used Sepedi as their first language during their high school studies.

3.7.3 Transcription

Each participant's recordings were initially transcribed into a Microsoft Word™ document by the researcher. It was done based on the transcription rules (Appendix J) set by the researcher, based on previous studies by Mogatusi (2022), Mothapo (2019), and Trembath et al. (2007) to ensure that there was consistency in the transcription of all the samples. Transcriptions were checked against the audio recording by a research assistant to ensure reliable transcription (O'Connor & Joffe, 2020). The research assistants marked any changes to the first transcriber's version using the track change's function. These additions, deletions or changes constituted disagreements. The percentage of agreement between the first and second transcriber was calculated as a measure of inter-rater reliability of the transcription. The following agreement formula was applied:

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

Agreement of 80% or more was deemed acceptable, this was to ensure that only a maximum of 20% of the data might be deemed incorrectly transcribed (McHugh, 2012).

Table 5 below, gives the percentage agreement obtained for each of the six transcripts. The overall percentage agreement was 88.2%, which constitutes good agreement and suggests that the speech samples were reliably transcribed. In cases of disagreement, the researcher and research assistants informally reviewed the speech sample to reach an agreement.

Table 5

Percentage agreement of transcriptions

Participants	**1**	**2**	**3**	**4**	**5**	**6**	**Overall**
Percentage agreement obtained (%)	90%	80.6%	80%	95.9%	91.8%	90.8%	88.2%

3.7.4 Tagging and analysis

The data were quantitively analysed, which entailed frequency counts and commonality counts of the words. Firstly, tags were added by the researcher to each transcript according to a set of tagging rules compiled (Appendix K), adapted from Mothapo (2019). Tags ensured that: (a) homonyms were counted as different words and different forms of a noun (e.g., plural form, diminutive form, locative form); and (b) different forms of a verb (e.g., tenses, moods) were all counted under one root word (dictionary form of the noun/verb). A research assistant independently checked the tagged transcripts against the tagging rules (Appendix K) and the Oxford *Pukuntšu ya Sekolo* (school dictionary) (De Schryver (2007) to ensure reliability. The research assistant marked any instances in which they disagreed with the tag assigned (a different or no tag should have been assigned), or where they felt a tag should have been omitted. These instances were indicated as disagreements. An agreement consisted of any tagging applied by the researcher that the research assistants confirmed as correct. Agreement between the tagging performed by the researcher and research assistant of 80% or more was considered acceptable (McHugh, 2012). Agreement of the tagging was calculated according to the following formula:

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

The percentage agreement of the tagging per transcript and overall is provided in Table 6 below.

Table 6

Percentage agreement of tagged transcripts

Participants	1	2	3	4	5	6	Overall
Percentage agreement obtained (%)	95.3%	96.8%	89.9%	91.5%	93.8%	90.6%	92.9%

The overall agreement for the tagging of transcripts was 92.9%, indicating good agreement. The researcher checked the disagreements and made a final decision regarding the tagging. Thereafter, one composite transcript file was created by copying all transcripts into one Word document. Microsoft Word™ loaded with an additional word frequency count macro (Simonyi and Brodie (1983) was used to analyse the tagged transcripts per child and the composite tagged transcript, as well as to determine the word frequency count per participant and on the combined sample. The 'Word Count' function on Microsoft Word™ was used to identify the number of words in the transcript to ensure that each participant achieved the 3 000-word target. This analysis also provided the total number of words per transcript and the total number of words of the composite sample (including unintelligible words). A word frequency macro was run on each participant file and on the composite file to identify the number of different words used per child and for the composite sample. The frequency of unintelligible words was determined per child sample and for the composite sample, and subtracted from the total number of words, to arrive at a total number of words (intelligible words).

The TTR per participant and for the composite sample were then calculated by dividing the number of different words by the total number of words (intelligible words). The word frequency macro's output from the composite file was then used to create a table to classify the different words alphabetically, together with their frequency counts. This table was then copied into an Excel sheet. The researcher manually combined morphological variations of nouns and verbs under the lemma (root word). The frequency per mille was then calculated for each (root) word by dividing the total number of occurrences of a word by the total number of words within the composite transcript and multiplying by 1 000. The formula applied was as follows:

$$\frac{Total\ number\ of\ occurrance}{Total\ number\ of\ intelligeble\ words} \times 1\ 000 = frequency\ per\ mille$$

The words were arranged by frequency (from highest to lowest) on an Excel sheet. All words that had a per mille frequency count of at least 0.5‰ (meaning they appeared at least once in every 2 000 words) were identified.

Commonality counts were then determined for each word appearing in the composite sample with a frequency of 0.5‰ or more, by identifying the number of participants that used each of the words identified. All words appearing in the composite sample with a frequency of 0.5‰ or more and used by at least 50% of the participants was designated as core words. The core words obtained from the composite sample were classified into different parts of speech according to Oxford *Pukuntšu ya Sekolo* (school dictionary) (De Schryver, 2007) and the coverage of each word was calculated. Additionally, the core word list compiled was compared to the previous list compiled by Mothapo et al. (2021).

3.8 Reliability and validity

In quantitative research, reliability describes the consistency with which measurements and/or procedures are conducted. Internal validity describes the extent to which findings are warranted by the study procedures, while external validity describes the generalisability of findings (Heale & Twycross, 2015). In the current study, the reliability of transcription and of tagging was ensured by determining the percentage of agreement between the researcher and research assistant. The research assistant cross-checked the transcribed speech samples against the audio recordings to increase reliability, this was done based on the transcription rules (Appendix J). Overall reliability of the transcription was 88.2%, representing good reliability. To ensure the reliability of the tagging, the research assistant checked the tagged samples against the tagging rules and noted any disagreements. The overall percentage agreement on the tagging was 92.9%, which is a representation of good reliability.

To decrease the possibility of novelty effects, the first 20 minutes of the recorded speech samples obtained on the first day of data collection were not included in the transcription and all the references made about the recording equipment were omitted (Trembath et al., 2007). The researcher built a rapport with each participant and made them aware of the equipment before the data collection period to ensure they were comfortable with it. The researcher applied the same process to collect the speech samples and was consistent across participants in fitting the equipment, giving the same instructions to the participants and class teachers, and consistently obtaining 3 000 words from each participant.

To increase external validity, the data was collected from two schools that are situated in neighbouring villages approximately 30 km apart. The data was collected throughout the

school day, including in the classroom and on the playground. The study sample includes the same number of males and females.

3.9 Ethical issues

The ethical principles and guidance for conducting research as suggested in the Belmont Report (Biomedical & Research, 1978) were upheld in this study. The following principles were applied: justice; beneficence; confidentiality; autonomy; and non-maleficence.

The researcher requested ethical clearance from the Faculty of Humanities Research Committee, University of Pretoria. Approval was obtained from the Humanities Ethics Committee before the study commenced. Permission from the Limpopo Department of Education was granted to conduct the study at two selected public primary schools (where the LoLT in the classrooms is Sepedi) in the Sekhukhune district (Makhuduthamaga municipality). The principals of the two schools were asked for permission to allow the study to take place on the school premises and during school hours. Caregivers were provided with detailed information letters in English and Sepedi concerning the nature of the study and were asked for their consent to allow their children to participate in the study. The class teachers were provided with English consent forms to participate in the study. The researcher engaged in a rapport-building activity in the classroom to ensure that the participants were comfortable before the commencement of their participation. The participants were provided with an assent script and forms in Sepedi and English, the script included child-appropriate pictures and friendly language. Through the letters and the assent script, the researcher fully disclosed the purpose of the study and what it entails to the participants, caregivers, and teachers. All participants were made aware that participation is voluntary and that non-participation will not pose negative consequences for them.

Autonomy was maintained by informing participants of all their rights, including the right to discontinue their participation at any time. Teachers were requested to indicate their choice to participate or not via a signed consent form. Parents/legal guardians were requested to indicate their choice to allow their child to participate or not via a signed consent form. Children were requested to indicate their choice to participate or not via an assent form. All participants were treated with fairness and equality to maintain **justice**.

The researcher ensured that the recorders in the pouches and microphones were fitted securely. To ensure that **non-maleficence** was upheld, the researcher asked the participants to indicate any physical discomfort resulting from the pouch on their waist or the microphone on their collar to their teacher immediately. Teachers were asked to assist or remove equipment

as they see fit. This was done to ensure that the study did not pose any risk to participants. The pouches were placed on the side of the participant's waist to avoid interference with their activities throughout the day. The participants were further encouraged to inform their class teachers should they experience any discomfort.

The researcher safeguarded that participants' information remained private and **confidential** by de-identifying transcriptions. Personal identifiers such as names of participants and names of the schools were removed from the transcriptions and replaced with codes to ensure that participants and schools remain anonymous. The recordings collected and scanned completed questionnaires will be securely stored on an external drive format at the Centre for AAC (University of Pretoria) once the study is completed. Only the researcher and supervisor have access to identifiable data, such as the responses to questionnaires and audio recordings. The participants' recordings may be shared with other researchers with the caregivers' written permission, and these researchers are obliged to 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/), and on the University of Pretoria research data repository and platform (University of Pretoria, n.d.).

The research findings do not benefit the participants directly; however, the research findings will **benefit** the AAC field in both the Department of Health and Education by providing a more extensive Sepedi core vocabulary list when selecting aided AAC systems for individuals with CNN.

3.10 Summary

This section focused on outlining the research methodology that was implemented in the study. The chapter reports on both the main aim and sub-aims of the research and the type of research design that was followed. It further reported on the pilot study and main data collection procedures.

4. RESULTS

4.1 Description of the sample

The transcribed language samples consisted of a total number of words (intelligible words) ranging between 2 817 and 3 783 words per participant and a combined sample of 19 316 words. Table 7 below illustrates the total number of words (intelligible words), number of different words (morphological variations counted separately), and TTR per participant. The combined sample's number of different words came to 1 917 intelligible words, with the morphological variations of different words counted separately. The composite sample resulted in a total TTR of 0.10.

Table 7

The total number of words (intelligible words), number of different words (morphological variations counted separately), and TTR per participant

Participant	Total Number of Unintelligible Words	Total number of words (Intelligible Words)	Number of different words	TTR
Participant 1	37	3 065	614	0.20
Participant 2	25	3 157	704	0.22
Participant 3	18	3 783	696	0.18
Participant 4	20	3 430	659	0.19
Participant 5	21	3 064	588	0.19
Participant 6	30	2 817	689	0.24
Total	151	19 316	1 917	0.10

In the next step of the analysis, morphological variations of nouns, verbs, and adjectives were combined under their lemma (root word), as guided by the tagging rules. For example, the singular, plural, and locative form of the same nouns (e.g., *kereke*, *dikereke*, *kerekeng*) were all combined under the singular form. The number of different words (morphological variations combined) was then recalculated and was found to be 1 068 across the whole sample. It should be noted that the number of different words of 1 068 is an approximation, as the proper nouns

referring to teachers' names, children's names, and place names were collectively replaced by the abbreviations 'CN', 'TN' and 'PN'. This artificially lowered the number of different words somewhat. The TTR based on the number of different words (morphological combinations combined) was 0.06.

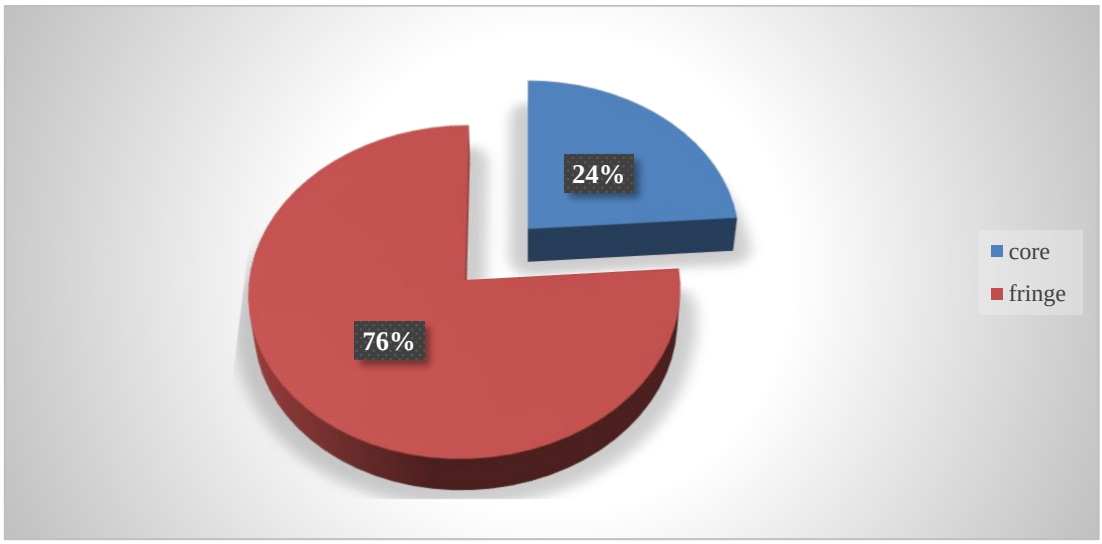
4.2 Core and fringe vocabulary

The total sample output was analysed using two criteria to identify core words. These criteria were used in previous studies by (Mngomezulu et al., 2019); Mogatusi, (2022); and Mothapo (2019) to determine the core words. Firstly, the words that had a frequency count of at least 0,5 per mille were determined. Secondly, the words that met this criterion were inspected for commonality, and all words that had a commonality score of at least 3 or more (meaning that the word was used by three participants or more) were identified. This ensured that at least 50% of the participants used the words designated as core. A total number of 262 words were found to meet the frequency criterion. Seven of these did not meet the commonality criterion, and hence were excluded. After applying both criteria, a total core list of 255 core words was established. The Sepedi core vocabulary word list based on the Sekhukhune district with the frequency count, commonality count, and their part of speech are documented (see Appendix L).

Of the 1 068 words, 813 words did not meet the criteria to be included as core vocabulary. These words were classified as fringe vocabulary, which makes up a larger proportion of the number of different words. Figure 5 below, is an illustration of the percentage of core and fringe words making up the total list of number of different words from the overall sample.

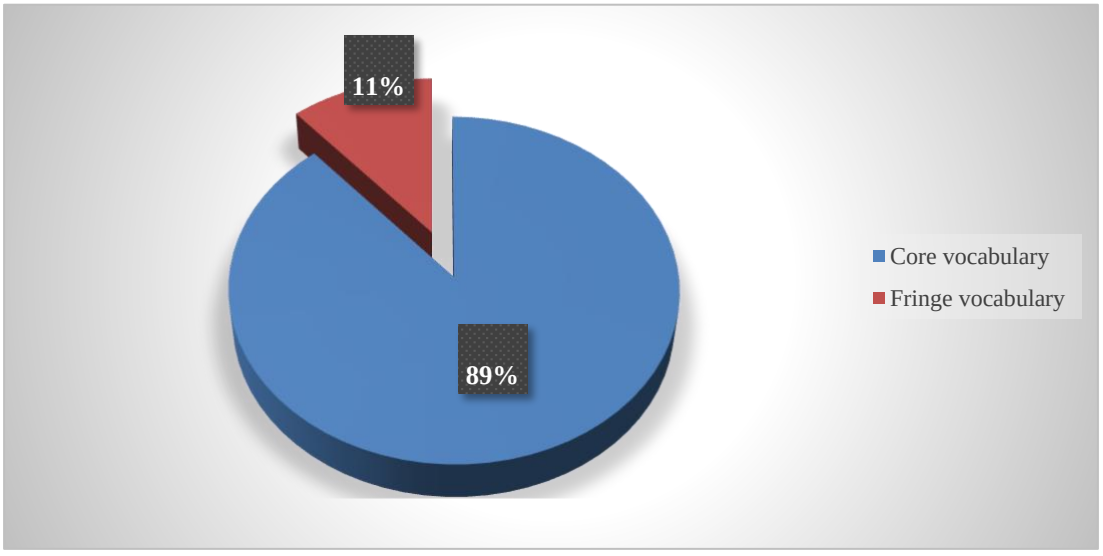
Figure 5

The proportion of core and fringe words in the total number of different words of the overall sample



To determine the coverage of the core words in the overall sample, the summed frequency of 886.8‰ was divided by 1000 and multiplied by 100 to give a score of 88.7%. This is an indication that 88.7% of words that were used by participants during conversations were core words, while fringe vocabulary had a coverage of only 11.3%. This affirms that fringe words, although much larger in number (number of different words), were not commonly and frequently used amongst the participants. Figure 6 below, demonstrates the proportional coverage of core and fringe vocabulary on overall sample.

Figure 6 *The proportional coverage of core and fringe vocabulary*

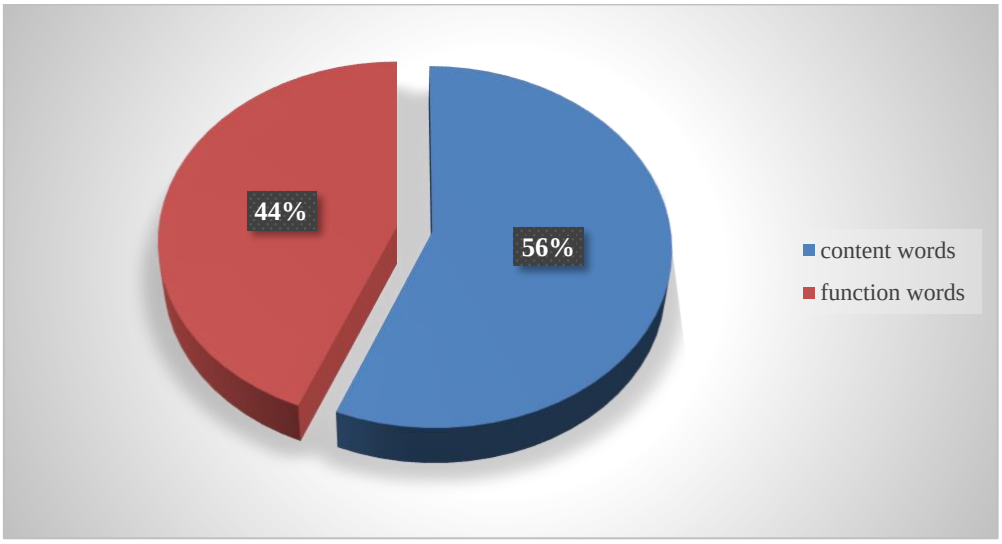


4.3 Core vocabulary: Content versus function words

The core words were differentiated as being either content or function words. From the core word list, 143 of the words were found to be content words, and the remaining 112 were function words. This is demonstrated in Figure 7 below, showing that 56% of the core word list consists of content words and 44% consists of function words. Furthermore, the top 25 core words were analysed to classify the top number of different words as either being content words vs function words. The findings indicates that 19 words are classified as function words. These words have a commonality score of 6, which indicated that they were used by all the participants within the sample.

Figure 7

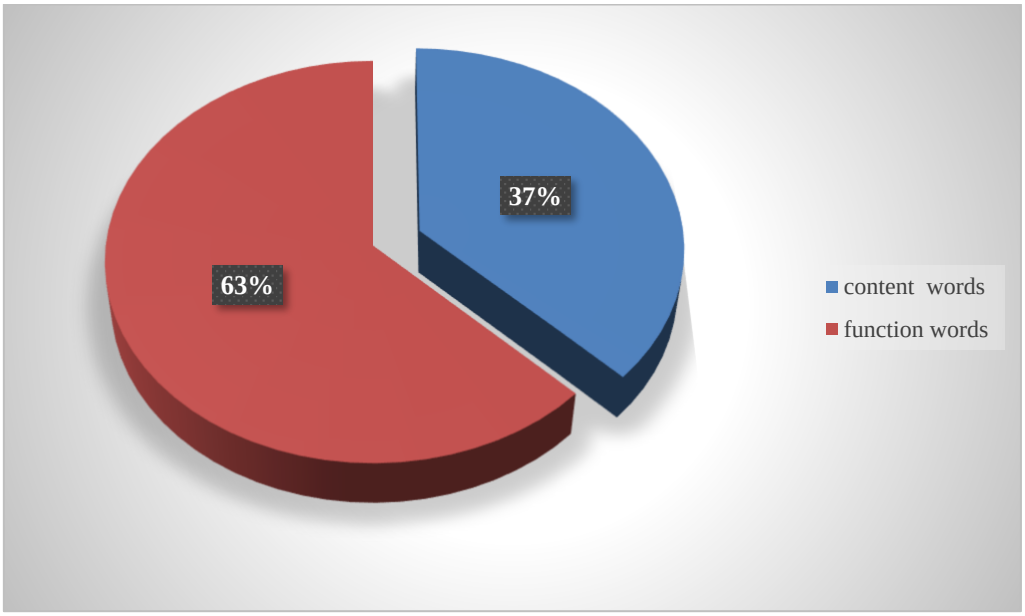
Proportion of content and function words in the core vocabulary



Furthermore, the coverage of the content and function words appearing in the core vocabulary list was calculated by summing up the frequency per mille of all the content words and all the function words and converting it to a percentage. Figure 8 below, shows the coverage of content versus function words appearing in the core vocabulary list. The figure shows that 63% of the core words used by participants were function words. The overall coverage of the top 19 function words accounted for 17% of the overall 88.7% core vocabulary.

Figure 8

Coverage of content vs function words on the core vocabulary list



4.4 Core vocabulary classification according to part of speech

The core words were classified according to various parts of speech (nouns, verbs, adverbs, adjectives, etc). The Sepedi *Pukuntšu ya Sekolo* (school dictionary) (De Schryver, 2007) was used to classify the Sepedi words, while the code-switched English words were classified according to the English part of the same dictionary. The number of different words per part of speech was established, as was the frequency with which words belonging to that part of speech was used in the sample (coverage per part of speech). The results are displayed in Table 8. The parts of speech are arranged by decreasing frequency of use.

From Tables 8 and 9 below, it is apparent that the core word list contained 85 different verbs, 47 different nouns, 29 different concords, 21 different interjections, 15 different adjectives, and 11 different pronouns. The remaining words came from various parts of speech, all numbering less than 10 words per part of speech. This is also displayed in Figure 10. The 29 concords were used with a frequency of 279.1‰. In contrast, the 58 verbs were used with a frequency of 186.7‰, and the 47 nouns with a frequency of 107.5‰. All other parts of speech categories were used with a frequency of less than 100‰ (i.e., with a frequency of less than 10%). This is displayed in Figure 11.

Table 8

Part of speech represented on the core vocabulary list

Part of Speech	**Number of different words**	**Frequency of Occurrence in the Core List**
Concords (total)	29	279.1
Subject concords	*16*	*225.9*
Object concords	*7*	*31.5*
Possessive concords	*6*	*21.7*
Verbs (total)	85	186.7
Main verbs	*74*	*169.5*
Auxiliary verbs	*7*	*11.7*
Modal verb	*1*	*1.2*
Copulative verbs	*3*	4.3
Nouns (total)	47	107.5
Common nouns	*43*	*57.1*
Proper nouns	*4*	*50.4*
Interjections	21	54
Demonstrative particles/pos	7	36.7
Pronouns (total)	11	52.2
Reflexive pronoun	*1*	*1.3*
Communal possessive pronoun	*1*	*0.7*
Possessive pronoun	*4*	*15.4*
Absolute pronoun	*5*	*34.8*
Future morphemes	3	22.9
Locative particles	5	21.5
Adjectives	15	19.5
Copulative particle	1	17.2
Conjunctions	8	16.1
Negative morphemes	3	12.7
Connective particle	1	12.3
Adverbs	6	11.9
Infinitive prefix	1	7.7
Instrumental particle	1	6.5
Aspectual prefix	3	5.8
Demonstrative copulative pos	2	5.5
Preposition	1	4.1
Present tense morpheme	1	3.4
Temporal particle	1	1.2
Potential morpheme	1	0.98
Agentive particle	1	0.7
Hortative particle	1	0.6
Total	**255**	**886.8**

Figure 9

Number of different words per part of speech found in the core vocabulary list

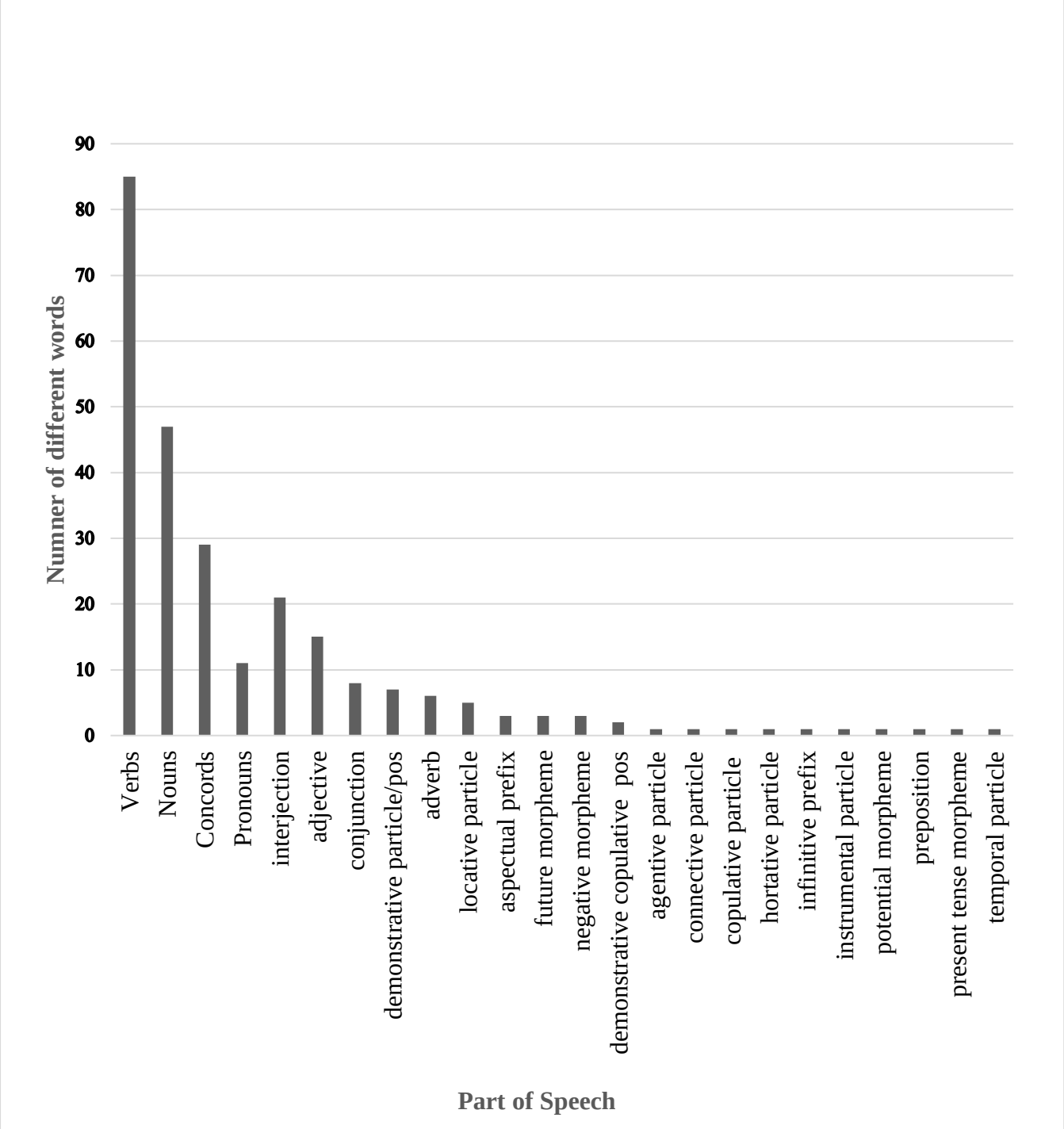
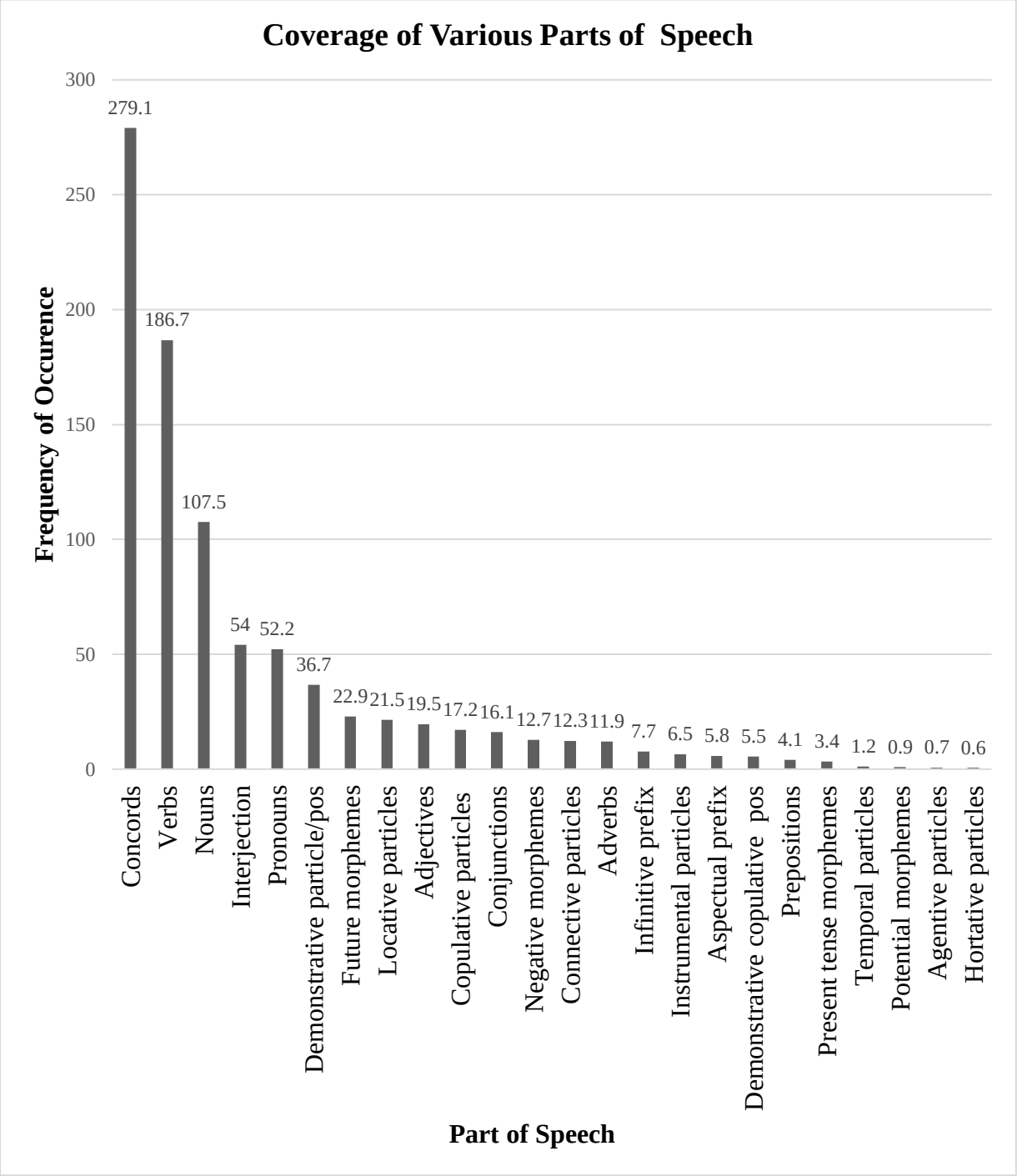


Figure 10

Part of speech coverage on the core word list



4.5 Comparison of current word list to previously compiled core vocabulary list

The core word list established in the current study was compared to the core vocabulary list compiled by Mothapo et al. (2021) based on speech samples from six preschoolers in the Capricorn district. Different dialects of Sepedi, namely Tau and Kone are spoken in the

Sekhukhune district versus the dialects that are mostly prominent in the Capricorn district (i.e., Kopa, Matlala, Mphahlele) (Mothapo (2019), with minor variations of the orthographic representation. The core vocabulary list compiled by (Mothapo et al., 2021) consists of 226 core root words compared to the 255 core root words compiled in the current study.

The most frequently used 100 core words in the current study were used with a frequency of 721.89‰, compared to 759.46‰ obtained by (Mothapo, 2019). The top 100 words in both studies have an average commonality score of 5.8 and contain more function words than content words. The current study's top 100 words consist of 59% function words, and Mothapo (2019) top 100 core words contain 53% function words.

Further comparisons were made regarding the parts of speech found in the lists. These comparisons are likely influenced by several factors, including some differences in transcription rules, regional dialect, and different part-of-speech classifications based on the grammar books and dictionaries consulted in each study. For example, in this study, the concords were further classified as either subject, object, or possessive concords, and the same phonetic form may be classified as three different words based on the grammar function it fulfilled. In contrast (Mothapo et al., 2021) did not further specify the type of concord, and therefore a phonetic form (for example, 'ba') was only classified as one word (a concord) rather than as three different words (subject, object, and possessive concords). In the current study, words were classified according to parts of speech in line with De Schryver (2007). Mothapo (2019) used De Schryver (2007) and additional grammar books (Poulos & Louwrens, 1994; Van Wyk et al., 1992) to identify the part of speech category for each word, resulting in a slightly different classification of the part of speech.

The top frequently used 100 words per list were compared based on the number of different words falling into different part-of-speech categories. The results are displayed in Table 9 below.

Table 9

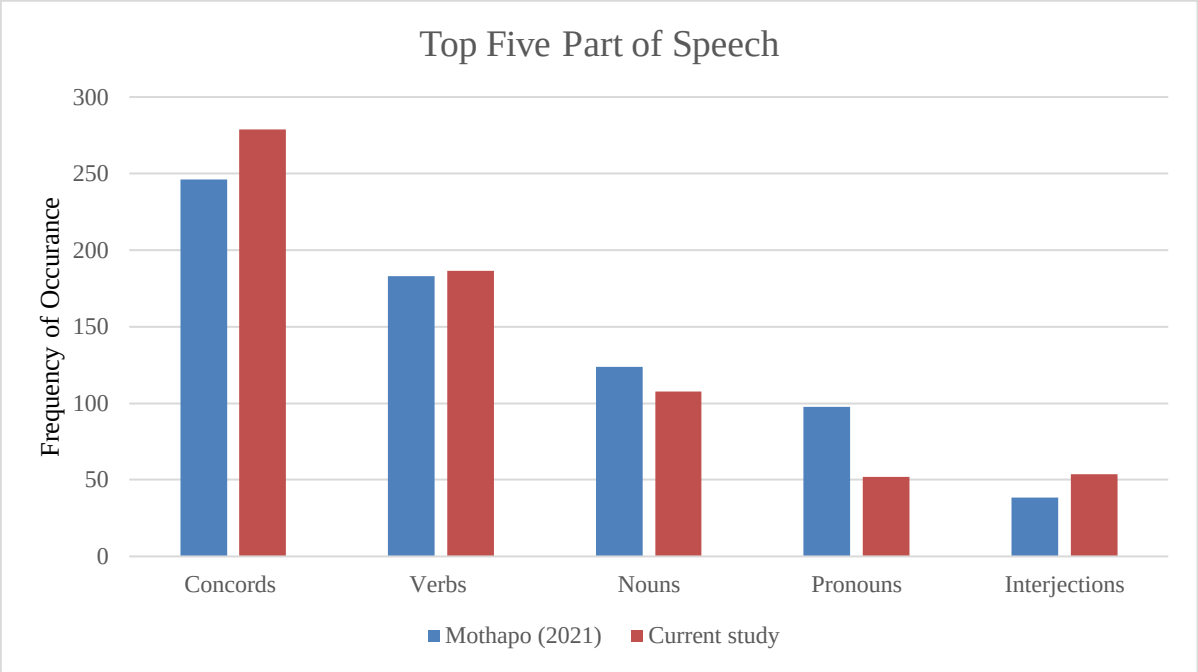
Comparison of the top 100 part of speech

Parts of Speech in Top 100 number of different words	**Current Study: Sekhukhune District**	**Mothapo et.al (2021): Capricorn District**
Adjectives	2	1
Adverbs	2	3
Aspectual prefix	1	1
Conjunctions	2	5
Connective particle	1	0
Concords	25	16
Copulative particle	1	1
Demonstrative particles/pos	3	1
Demonstrative copulative pos	2	0
Future morphemes	3	2
Hortative particle	0	1
Interjections	8	7
Infinitive prefix	1	0
Instrumental particle	1	0
Locative particles	2	2
Negative morpheme	1	2
Nouns	11	13
Potential morpheme	0	1
Preposition	1	1
Pronouns	5	14
Verbs	27	28
Present tense morpheme	1	1

The core vocabulary lists were further compared based on the frequency of occurrence of various part of speech. Figure 11 below, compares the frequency of occurrence of the five most frequently used part of speech found in the top 100 words. To make the comparison more balanced, all the concords, nouns, pronouns, and verbs in this current core vocabulary list were counted together.

Figure 11

Comparison of the five most frequently occurring parts of speech based on the top 100 words



From Figure 11, it is apparent that the frequency with which core words belonging to different parts of speech occurred in the two samples were quite similar, especially regarding verbs. Concords and interjections appeared to be more frequently used in this study as compared to that of (Mothapo et al., 2021), while the participants in the study by (Mothapo et al., 2021) were found to be using more nouns and pronouns in their conversations. This may, however, have been a matter of classification by slightly different rules.

A further comparison was made by determining how many content words and how many function words that appeared in the current list also appeared in Mothapo's (2019) list. Words were considered the same if they had the same identical meaning, had the same part of speech classification, and when they had a similar orthographic form (e.g., *bedi* and *pedi* were considered the same word, as they both refer to 'two' despite the slight orthographic phonetic variation. A total of 167 words were found to be common among the two lists. This number represents 65% of the words appearing in the current list, and 74% of the words appearing in Mothapo's list. These differences in percentage are due to the larger number of words appearing in the current list (255 words as compared to 226 words in Mothapo's list). A composite list is provided in (Appendix M). The list is divided into three sections, namely the overlapping words, unique words found in this study, and unique words found in Mothapo (2019). Of the 142 content words in the current list, 83 were also found in Mothapo's (2019)

list. This amounts to 58% of the content words in the current list overlapping with Mothapo's (2019) list. Of the 113 function words, 84 (i.e., 74%) were also found in Mothapo's (2019) list. When interjections are disregarded from this list, the overlap is even higher, with 78 of 92 words (i.e., 85%) having an equivalent word in Mothapo's (2019) list. The higher percentage of overlap among function words is expected, as function words typically provide the grammar structure of the language and are therefore, less dependent on the activity and context of use. Interjections, however, even though they are classified as function words (since they have no specific inherent lexical meaning) may differ much more from region to region.

To understand the influence of dialectical variations, the two-word lists were perused for words that have the same meaning and the same part of speech classification, yet have different phonological and orthographic forms. Thirteen words were found to represent such dialectical variations. Table 10 below, provides the dialectical variations and their part-of-speech classification.

Table 10

Words from the two vocabulary lists with dialectical variations

Words from the current study (based on the Sekhukhune district)	**Words from the study by (Mothapo, 2019) (based on the Capricorn district)**	**Part of Speech Classification**	**English Translation**
a	*ga*	negative morpheme	does not
yey	*haay*	interjection	hey
bapala	*raloka*	verb	play
gešu	*gešo*	pronoun	of our home
itia	*betha*	verb	hit
mfanaka	*mfana*	noun	friend
šo	*šule*	pronoun	there she/he is
ria	*dira*	verb	do
tjo	*yoh*	interjection	oh my gosh
sepela	*tsamaya*	verb	go

5. DISCUSSION

5.1 Characteristics of the speech sample

The study involved six participants who produced a combined total of 19 316 intelligible words (total number of words) and as mentioned above, 1 068 number of different words with an overall TTR of 0.06. The TTR reflects the proportion of unique words as compared to the total number of words within the sample (Richards, 1987). Richards (1987) further stated that a low TTR indicates that participants repeated most words during conversations. The author also outlined that the overall sample size (total number of words) directly influences the TTR. The larger the sample size, the higher the chances of re-used words in the sample. This was echoed by (Covington & McFall, 2010).

The low TTR found in the current study was, therefore, expected as a rather large sample was collected. Two previous core vocabulary studies conducted with Sepedi and Setswana Grade R learners respectively, found the same TTR (0.06) based on a speech samples of a similar size (Mogatusi, 2022; Mothapo, 2019)(Mothapo, 2019). Even in other studies targeting non-African languages such as English (Trembath et al. (2007) and Afrikaans (Hattingh & Tönsing, 2020), TTRs between 0.06 and 0.08 were found. This indicates that preschoolers and school-age children across languages use limited unique words in their conversational interactions. It suggests that children reuse the same words in conversations, and hence, it may be possible to determine a limited set of frequently used words (core words) for inclusion on an AAC system for individuals with CCN.

5.2 Characteristics of the Sepedi core vocabulary list (Sekhukhune district)

From the total sample of 19 316 words, a core vocabulary list of 255 words could be determined by applying the frequency and commonality criteria. The number of unique words on this list compared well to other core vocabulary lists that used the same criteria to establish core vocabularies (i.e., a frequency count of 0.5 per mille and a commonality score of at least 50% of the population). For example, four other core vocabulary lists based on South African Grade R learners' speech samples were established by (Mothapo et al., 2021) in Sepedi (based in the Capricorn district), (Mogatusi, 2022) in Setswana, (Mngomezulu et al., 2019) in isiZulu, and (Hattingh & Tönsing, 2020) in Afrikaans. These lists contained between 226 and 249 words. Similar findings were reported from other geographical regions outside of South Africa. Trembath et al. (2007) established a total of 263 core vocabulary words from an English-

Australian population,(Robillard et al., 2014) identified a French core word list comprising 216 words based on samples from Canadian children.

It is interesting that core vocabularies across languages from similar and different linguistic families seem to number in the region of 200–260 words. It suggests that the specific language has little influence on the core vocabulary size, and that all languages contain a small pool of frequently used words. It is plausible to suggest that such words are valuable and relevant for AAC users whose vocabulary may need to be limited in order to avoid the cognitive overload that comes with attempting to memorise the location of a large number of words on an aided AAC system (Yorkston et al., 1988).

The 255 core word list established in the current study accounted for 89% of the composite sample, which is consistent with the findings of the same language by Mothapo et al. (2021), who found that the core vocabulary established in their study covered 88% of the sample collected. Similar findings were observed by Mngomezulu et al. (2019) who reported a coverage of 88% of the core formatives that were found in their study; and Mogatusi (2022), who reported 86% coverage on the Setswana sample. Coverage of the Afrikaans and some of the English core vocabularies are less than that established in the African language studies. For example, Hattingh and Tönsing (2020) established an Afrikaans core word list that covered 79.4% of the sample and Trembath et al. (2007) reported 79.8% core coverage based on an English language sample. This suggests that core vocabulary accounts for 79–90% coverage of language samples in different languages. This finding is like that of Van Tilborg and Deckers (2016), who report it to be 80%.

Mogatusi (2022) argued that the presence of concords in most African languages tends to increase the core coverage. Since concords do not appear in Indo-European languages, the coverage of core may be smaller. Concords are a set of closed-class morphemes, limited in number, that are used as bounding words to show a relation between a noun and another part of speech in a sentence, and can be used as substitutions when referring to a previously mentioned noun (Poulos & Louwrens, 1994). In Sepedi, these words are used to connect a subject to a verb, substitute nouns, and link a possessive word with its possessor (Faab, 2010). This word class features prominently in most African languages, such as Setswana (Mogatusi, 2022), Sepedi (Mothapo, 2019), and isiZulu (Mngomezulu et al., 2019). These individual words have their own orthography and are used in most of the sentences, which resulted in large coverage in the speech samples.

The remaining 11% of the sample was covered by fringe vocabulary. These words constitute a larger total number of words, however, have a small coverage of the composite

sample. They are more personalised words that each participant used during their conversations. These words reflect the participant's interests and specific environment (Trembath et al., 2007). This smaller coverage was consistent with other studies, including those based on African languages (Mngomezulu et al., 2019; Mogatusi, 2022) where coverage ranged from 12% to 14% and smaller coverage were also found in studies based on the Indo-European languages (Hattingh & Tönsing, 2020; Robillard et al., 2014; Trembath et al., 2007), where coverage ranged from 20.2% to 20.6% . Including fringe vocabulary on AAC systems is essential to connect the persons with CCN to their environment, allow them to communicate their family members' names, and engage in school and work activities. Determining an AAC user's fringe vocabulary can be quite complex since it cannot be predicted or even generalised to different persons who use AAC, and the selection of fringe words requires regular updating to ensure that the words and symbols are suitable for the activity at hand. AAC experts are required to explore the user's environment and daily activities to conceptualise the required fringe vocabulary.

5.3 Content versus function words in the Sepedi core vocabulary list

In the current study, a higher number of unique content words was found in the core word list as compared to the number of unique function words, however, function words had a higher coverage. Function words are essential in sentence construction as they convey grammatical relations. They are necessary in understanding the relationship between content words, for example, connecting subject to verbs (Corver & van Riemsdijk, 2013). The findings indicate that function words were primarily used during participant's interactions. These findings are consistent with (Mothapo et al., 2021), who also identified a higher coverage of core function words, compared to core content words. Additionally, 76% of the top 25 core words in the current study were classified as function words, which was found to be the same in the study by Mothapo (2019). This reiterates that function words are commonly used and are crucial during communication. It was found that the most frequently used word in this sample is the function word *ke* translated as 'I', which is consistent in other core vocabulary lists by Mogatusi (2022), Mothapo (2019), and Trembath et al. (2007).

The study has a fair number of content words, which covered over a quarter of the sample. In a study conducted by McFadd and Wilkinson (2010), most participants included more content words on the AAC system, this is based on the findings that children who rely of AAC to communicate use mostly single-word utterances to interact within their environment (Solomon-Rice et al., 2017), which may be a reflection of limited access to vocabulary on their

AAC systems (McFadd & Wilkinson, 2010). This is outlined by (Mngomezulu et al., 2019; Mothapo, 2019; Trembath et al., 2007), who argued that including content words is essential for AAC users to be effective communicators and respond to communication demands and may limit working memory demands (Thistle & Wilkinson, 2013). While this is true, the inclusion of function words is as important in conveying a multi-word message, which facilitates language and literacy development, especially for beginner AAC users who are on the single-word level and have the potential to progress to multi-word utterances guided by their linguistic and cognitive abilities. The development of grammar is useful for individuals to effectively communicate their thoughts and feelings to familiar and unfamiliar communication partners. Furthermore, the inclusion of function words on the core vocabulary list may facilitate literacy skills during school activities (Riccelli-Sherman, 2017).

The Sepedi core word list consists of a proportion of content and function words that will efficiently provide a baseline of vocabulary to be included in an AAC system. However, to avoid system abandonment, this list should be used in conjunction with various fringe words that are of personal interest to the user in order to support efficient communication across contexts (Moorcroft et al., 2020). Fringe words increase motivation to use AAC systems, since they allow the expression of personal identity, experiences, and preferences (Wofford et al., 2022). Furthermore, these words personify the AAC systems to be more responsive to the individual's social context.

5.4 Parts of speech found in the current core vocabulary list

The Sepedi core vocabulary list identified in this study included 38 different parts of speech, including three types of concords, four types of verbs, interjections, two types of nouns, four types of pronouns, various particles, various morphemes, prepositions, adverbs, and adjectives. In the current study, different types of concords, verbs, nouns, pronouns, morphemes, and particles were counted separately. However, for the purpose of comparison to other studies, all concords, nouns, verbs, and pronouns were combined and counted together.

Coverage analyses showed that concords, verbs, nouns, interjections, and pronouns were the parts of speech used most frequently by the participants. Similar results were found in the Sepedi and Setswana studies by Mothapo (2019) and Mogatusi (2022) respectively. Across all three studies, participants used concords, verbs, nouns, pronouns, and interjections more frequently during their conversations than other parts of speech, such as adjectives, morphemes, etcetera. This reflects the similarity of the Sepedi and Setswana languages—they both belong to the Sotho-Tswana languages, which share overlapping vocabulary and a very

similar language structure (Prinsloo, 2014). These similarities are expected, since both languages are derived from the same language cluster (Mokgokong, 1966).

Concords were also found to be frequently used by isiZulu-speaking preschoolers (Mngomezulu et al., 2019; Mogatusi, 2022). As previously explained, these words provide links between nouns and other parts of speech to explain the relationships between them and enable grammatical expression. Concords are a prominent structural feature of many African languages (Mngomezulu et al., 2019; Sengani, 2013).

Verbs, nouns, pronouns, and interjections are prominently represented on this core vocabulary list, as in other core vocabulary lists in African languages (Mngomezulu et al., 2019; Mogatusi, 2022) as well as Indo-European languages (Hattingh & Tönsing, 2020). Verbs and nouns are content words that are required in many languages to convey meaning. Robillard et al. (2014) reported that monolingual French children's core vocabulary contained only one noun, and Tsai (2023), conducting a similar study looking into Mandarin core vocabulary, only found one verb on the list. This reflects the influence of languages on core vocabulary, thus the need to compile core words in different languages to ensure that the core vocabulary is responsive to communicational demands. Furthermore, core vocabulary must reflect different parts of speech aligned to the spoken language that is represented on the system for it to be meaningful and useful in different communication settings.

Although interjections typically do not provide a specific lexical meaning, they were prominent in this study and other similar studies (Hattingh & Tönsing, 2020; Mothapo et al., 2021). These utterances assist speakers in expressing their current mental state during conversations (Ameka, 2006). However, they do not have a consistent orthographic representation contained in a dictionary, which may influence their use in different languages and regions. These words can be expected to differ from one area to the next based on linguistic and cultural differences (Wierzbicka, 1992). The comparison with the findings by (Mothapo et al., 2021), showed significant differences in the type of interjections used by preschoolers. The variations are to be expected, since the interjection word class is open-ended in nature, allowing the creation of new interjections. In addition, interjections are typically not dictionarised, meaning that they are not standardised in spelling. Interjections are therefore, open to change and open to influence by geographic space, society, and age group (generational differences) (Norrick, 2009).

Code-switching was observed in the sample. Eleven English words form part of the core vocabulary list. This phenomenon is prevalent in other studies based on the South African context, for example, (Hattingh & Tönsing, 2020; Mogatusi, 2022; Mothapo et al., 2021)

reported the presence of English words in their samples. Mothapo (2019) reported that code-switching occurs due to exposure to different languages in society. Preschoolers may be exposed to English due to exposure to language via the media (including social media) and web content. In 2023 about a quarter of the South African population owned a smartphone with Internet capabilities, hence, exposure to such content is ubiquitous (Statista, 2023). In addition, English is prominently used in educational settings, where it is often an additional or the only language of teaching and learning (Malebese & Tlali, 2020). Furthermore, the absence of certain words in African languages compel speakers to use the English words (Mothapo, 2019). Code-switched words may well be relevant for AAC users to participate in various activities without communication breakdowns caused by limiting vocabulary. In South Africa, there is an influence of cross-linguistic contact (Tönsing et al., 2018), which increases the exposure of various languages in societies. As a result, most South Africans are bilingual or multilingual. Moreover, speakers code switch to improve comprehensibility and accommodate communication partners (De Kadt, 2005).

Coverage of various parts of speech on an AAC system is crucial for effective communication. The current core vocabulary list is relevant to the language and includes relevant parts of speech within the language (based on similarities with Mothapo (2019)). This provides the user with various words and symbols to formulate multi-word utterances. This inclusion mimics the language development of typical developing children by offering children who uses an AAC the opportunity to communicate beyond verbs and nouns.

5.5 Comparison of the current list with the Sepedi list established by Mothapo (2019)

The core words in this study were compared to the list by Mothapo et al. (2021). Both lists are based on the Sepedi language from different geographic regions. The list by Mothapo et al. (2021) was based on speech samples collected from children from the North-Western Sotho cluster where the most prevalent dialects are Tokwa and Kopa (Mokgokong, 1966). The participants in the current study spoke the most prevalent dialects in the Sekhukhuneland, that is Pedi and Tau. A total of 165 words appeared on both lists, indicating that approximately two thirds of the current list overlapped with that of Mothapo (2019), while nearly three quarters of Mothapo's list overlapped with the current list. In many ways this represents quite a substantial overlap, considering that words were collected in different contexts and at different times (Mothapo's sample was collected in 2018, whereas the sample for this study was collected in 2024). When comparing five English lists, for example, Hattingh (2018) found that

only 26% of the total number of different words across these lists appeared in all five lists. However, it should be noted that these lists were compiled across different age groups and very different geographies (different states in the USA and in Australia). Further, the more lists are added to a comparison, the less likely the statistical chance that a word will be found on all lists.

In other ways, the differences found between the two Sepedi lists should still alert one to the fact that core lists established for a language are not immutable absolutes, but always just one version of the truth—frequency lists based on samples collected at a particular time and in a particular context (Laubscher & Light, 2020). Therefore, no core vocabulary list should be seen as static or finite, and this must be considered when they are incorporated into an AAC system.

A greater overlap was found in the function words as compared to content words. This was expected, since these words provide grammatical support and are unlikely to be influenced by the context, activity, or personal preference (Soto & Tönsing, 2024). In addition, it suggests that dialectical variations did not express themselves primarily in grammatical structure, but that the grammar structures used in both contexts were similar. This is important to consider in AAC vocabulary selection for children who require access to words that enable sentence-building—the function words contained in the composite list (Appendix M) may provide a useful grammatical framework for children in need of AAC.

Most of the function words that were unique to either of the lists were classified as interjections. This was not surprising since this word class is open to modification based on environmental and cultural influence. These words are expected to vary from one region to the next (Soto & Tönsing, 2024). The current study had different types of interjections, which were not found in the Mothapo (2019) study. This indicates that participants from Sekhukhuneland used more interjections during their conversations. The lack of orthographic form for these utterances may have influenced how they were transcribed in the two studies, thus resulting in variation in the orthography of the words.

There was considerably less overlap between the content words of the two lists. This was to be expected, as content words are typically more context- and activity-specific and may also represent the preferences of the speaker. While there was an overlap of words, such as *tlaleya* (report), *sekolo* (school), *pedi* (two), this may have resulted from the fact that both samples were collected in preschool settings, where children engage in educational activities and spend time playing and socialising with each other. This suggests that AAC experts may pre-select context-specific content words (Marvin et al., 1994), for example, for the school

setting. Other content words such as *nwa* (drink) and *abuti* (brother) were unique to one of the lists. Verbs and nouns specifically are open-class words, meaning that there are many of them and new words can be added at any time to these categories. The popular topics in the media, the themes and activities presented by the teacher in class, and the play activities that children choose for themselves as well as the toys or other artefacts available in the environment, can all influence the content words they use. Furthermore, since the use of content words is influenced by the environment (Marvin et al., 1994), children use words based on events that they are exposed to. For example, participants in the study by Trembath et al. (2007) frequently used words such as 'spiderman' and 'swing' which were influenced by the events and artefacts in that environment.

Dialectical variation may have had a limited influence on the differences between the core vocabulary lists. This is expected, since different dialects share linguistic features as they are based on the same standard language form (Crystal, 2011). Ten words with dialectical variations from the two lists were noticed, including verbs, nouns, pronouns, interjections, and negative morphemes. Some dialectical variations are not represented in the Sepedi Oxford dictionary *pukuntšu ya Sekolo* (school dictionary) (De Schryver, 2007). Earlier studies (Mojela, 2007; Mokgokong, 1966) outlined that other Sepedi dialects are not defined in the official standard language, thus, resulting in some of the vocabulary being excluded. The current core vocabulary list caters to AAC users in Sekhukhuneland, ensuring that they are not forced to communicate using the standard language in a society that speaks another dialect (Mokgokong, 1966). The list may also provide an additional set of words that various Sepedi AAC users require from different areas, and it offers region-specific core words that afford individuals a core vocabulary that is relevant and accepted by their society. Since both lists are based on the Sepedi language, the common words indicate that the two core word lists may be generalised with caution across the Bapedi regions.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary of main findings

Core vocabulary lists have been used in the field of AAC to guide vocabulary selection for individuals with CCN, while reducing the time and complications involved in preselecting vocabulary for individuals who are not yet literate and require aided AAC systems to communicate (Boenisch & Soto, 2015; Johnson et al., 2017). The lists provide a pool of relevant words to various settings and can be used with different communication partners. These words are helpful for providing a wide vocabulary that supports language and literacy development (Riccelli-Sherman, 2017). Core word lists are developed to be used in conjunction with other vocabulary sources, for example, words suggested by informants (Beukelman & Light, 2020) in order to include an individualised fringe vocabulary for full expression.

This study aimed to determine the Sepedi core vocabulary from six Grade-R learners in the Sekhukhune district, which resulted in a composite of 19 316 intelligible words and a core vocabulary list of 255 words. The core words accounted for nearly 90% of the overall conversations. This coverage is consistent with existing literature on core vocabulary lists such as lists compiled for isiZulu (Mngomezulu et al., 2019) and Setswana (Mogatusi, 2022) that were also compiled based on speech samples collected from preschool children. The greater coverage suggests that these words can be useful on an aided AAC system as they are relevant to different situations. Of the 255 words, function words had a greater coverage than content words. This higher function word coverage is observed in other core vocabulary lists based on African languages such as isiZulu and Setswana. Function words provide grammatical support for the content words, where a child is likely to have a more comprehensive participation and aid literacy development. Furthermore, the core word list has a great representation of various parts of speech that are frequently used in the Sepedi language including, concords, nouns, verbs, and interjections.

The current core words list serves as a foundation of core word resources for the Sekhukhune population. Due to the dialectical variation and environmental influences, the core words are somewhat different from the existing Sepedi core word list by (Mothapo, 2019). However, this list can also provide an additional set of Sepedi core words that may be used in conjunction with the existing list to guide the vocabulary selection of Sepedi-speaking children with CCN. These two lists offer a vast number of words that were used by at least 50% of the participants and the words are relevant and applicable for the Sepedi-speaking population.

6.2 Implications for the study

Vocabulary selection in AAC requires the use of various language resources to select words that support language development as well as reflect the individual's personality and interests. This has been documented as being a challenging task for therapists (Boenisch & Soto, 2015). The adoption of core vocabulary lists as a resource for vocabulary selection on aided AAC systems for children has been observed in various languages such as English (Banajee et al., 2003) to provide vocabulary for children who are pre-literate. Children with CCN are expected to participate within their society with limited communication breakdowns. The inclusion of core words that are based on the specific region's language, dialect, and pronunciation may lead to increased participation in various activities such as family interactions, church services, and schooling activities. Although core words are important during participation, access to fringe vocabulary on an AAC system is also necessary to ensure that individuals with CCN can participate fully in different activities using topic-specific language.

The 255 Sepedi core words provide a vocabulary resource for speech therapists and other experts when providing AAC services to Sepedi-speaking school-aged children from the Sekhukhune district. This is especially true when deciding on which words to include on an aided system of a Sepedi speaker from Sekhukhuneland, to ensure that their vocabulary is the same as that which persons without disabilities use in their everyday environment and that they can interact in their daily and scholastic activities. Sepedi is the most prominently spoken language in Limpopo, at approximately 55.5% (Statistics, 2022), and in the Sekhukhune district, it is the prominent home language of approximately 81% of the population (SekhukhuneDistrictMunicipality, 2017). These core words may be used in conjunction with the existing Sepedi core vocabulary list to cater to a greater population in the Sekhukhune and Capricorn districts. Since the Sepedi language has approximately 23 dialects, careful consideration should be applied before generalising this list to the larger population. Caution when including these words is important; thus, family consultation should be applied to ensure that the words are relevant in their society.

Publishing these core words may serve as a language resource for AAC developers when designing high-tech AAC systems for the Sepedi population (Mogatusi, 2022). However, a graphic symbol system based on the core vocabulary list alone is insufficient for optimal communication and participation. AAC experts must consult various informants, such as their interests and developmental inventories, to ensure that the words are appropriate and relevant to the individual and their different communication contexts.

6.3 Critical evaluation of the study

6.3.1 Strengths

The core word list generated in this study serves as a complement and extension of the existing Sepedi core vocabulary list established by Mothapo et al. (2021). The two core word lists are based on different geographical regions within the Limpopo province, South Africa. The word similarities between the two strengthen the lists' generalisability to the larger Sepedi population. Furthermore, the current list provides the core words based on the Pedi dialect, which is important when providing AAC services to this region. Offering an AAC system based on the regional language may increase the chances of acceptability of the system within the environment, while supporting participation (Beukelman & Light, 2020).

The analysis that focused on root words can be regarded as a strength. Using root words instead of inflected forms on the aided systems reduces the vocabulary size and memory demands related to system navigation (Lund et al., 2017). The inclusion of root words ensures that individuals with CCN can express themselves using a system that provides an efficient vocabulary layout that is easily navigated.

The study implemented measures to increase internal validity by ensuring that the recordings occurred in a similar context and with participants of the same age. Measures were applied to reduce the novelty effect on the overall samples, as reported (Trembath et al., 2007). This was done by excluding the first 20 minutes of the participants' first samples to ensure that they were comfortable with the equipment and by exclusion of words related to the recording process to ensure that the language sample was a true representation of everyday speech and that it was not influenced by the recordings and equipment.

The transcription and tagging process was found to be reliable. When checked by a research assistant, an inter-rater reliability average of 88% was established regarding the transcription. The inter-rater reliability of the tagging process was found to be 92%, representing good reliability.

Additional measures such as selecting participants based on two geographical areas that are at least 10 km apart, were put in place to increase external validity. This ensures that the language used is more general in the region.

6.3.2 Limitations

The study includes six participants from two sites in the Sekhukhune district to increase external validity. However, the sample size is too small to strengthen the core word list's generalisability to the greater population of Sepedi speakers. Furthermore, the words used by

participants are influenced by the school setting, as they are all in preschool and with a mean age of 5;5 (years; months). Since the recording took place over a limited duration of time, these words may not be frequently used by the participants throughout their school year. In addition, the core vocabulary established may not be appropriate to contexts other than the school setting,

As discussed earlier, Sepedi has approximately 23 dialects. Some dialects are more closely related with limited variation, whereas others are influenced by neighbouring languages such as Tshivenda and Xitsonga, thus resulting in a greater variation from other Sepedi dialects. Generalising this core word list to populations who use and understand other dialects such as Lobedu and Pulana, should be done with caution (Mokgokong, 1966).

Although the first 20 minutes of the first recordings were excluded, the samples indicated that children used some words related to the data collection procedures. While these were omitted from the analysis, it still indicates that participants were somewhat conscious of the recording equipment, which could have influenced the type of language they used and the conversations they engaged in. Increasing the duration of recording may desensitise the participants to the process.

6.4 Recommendations for further studies

Since the current study focused on the Sepedi dialect that is spoken in a district close to the Capricorn district and may thus, be relatively similar to the dialect spoken there (Mothapo, 2019), the variation in findings may be minimised. Conducting a similar study with children speaking other dialects that have a wider language variation such as Pulana, Tokwa, Lobedu, etcetera (Mokgokong, 1966), will provide a rich language source for the Sepedi population. The dialects-specific core vocabulary is essential for AAC service delivery in terms of vocabulary selection for individuals with CCN in those regions. These language sources provide speech therapists with an understanding of the language and grammatical knowledge of the different dialects.

Speech samples collected in a different environment, such as a home or community setting, may provide an additional Sepedi corpus to strengthen the generalisability of the existing core word lists across contexts. Only by collecting samples across contexts can it be clarified which words are truly used frequently across different communication partners and settings.

Future studies are needed to determine how these core words can best be represented and organised within a Sepedi graphic symbol-based AAC system. Studies would then also be needed to determine what intervention and teaching strategies may be successful in helping

children learn to use the system to communicate effectively in their everyday lives. Studies could test, for example, whether the inclusion of core words and specifically function words in the aided system will enhance the development of multiple-word utterances. Moreover, further analyses of the current transcriptions to identify co-occurring word sequences may provide guidance when deciding on the set of words to introduce on an AAC system at a time and how to arrange the vocabulary for easy access. Such studies could also help in developing word prediction software for Sepedi electronic AAC systems.

REFERENCES

Ameka, F. K. (2006). Interjections. In *Encyclopedia of language & linguistics* (pp. 743-746). Elsevier.

Banajee, M., Dicarlo, C., & BURAS STRICKLIN, S. (2003). Core vocabulary determination for toddlers. *Augmentative and Alternative Communication*, *19*(2), 67-73.

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/https://doi.org/10.1044/2019_AJSLP-18-0041

Beukelman, & Light. (2020). Augmentative and alternative communication: Supporting children and adults with complex communication needs. https://doi.org/https://stars.library.ucf.edu/etextbooks/588

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/https://doi.org/10.1080/07434618912331275296

Beukelman, D. R., & Mirenda, P. (2013). *Augmentative & alternative communication : supporting children and adults with complex communication needs* (4th ed.). Paul H. Brookes Pub. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=571668

Binger, C., Magallanes, P., Miguel, V. S., Harrington, N., & Hahs-Vaughn, D. (2024). How Toddlers Use Core and Fringe Vocabulary: What's in an Utterance? *American Journal of Speech-Language Pathology*, 1-30.

Biomedical, U. S. N. C. f. t. P. o. H. S. o., & Research, B. (1978). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. The Commission. https://books.google.bi/books?id=lw9fWOREf08C

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

Corver, N., & van Riemsdijk, H. (2013). *Semi-lexical categories: The function of content words and the content of function words* (Vol. 59). Walter de Gruyter.

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of quantitative linguistics*, *17*(2), 94-100. https://doi.org/https://doi.org/10.1080/09296171003643098

Crystal, D. (2011). *A dictionary of linguistics and phonetics*. John Wiley & Sons.

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/https://doi.org/10.1080/07434618.2017.1375979

De Kadt, E. (2005). English, language shift and identities: a comparison between'Zulu-dominant'and'multicultural'students on a South African university campus. *Southern African Linguistics and Applied Language Studies*, *23*(1), 19-37. https://doi.org/https://doi.org/10.2989/16073610509486372

De Schryver, G.-M. (2007). *Pukuntšu ya polelopedi ya sekolo : Sesotho sa Leboa le Seisimane : e gatišitšwe ke Oxford = Oxford bilingual school dictionary : Northern Sotho and English*. Oxford University Press Southern Africa.

Deckers, S. R., Van Zaalen, Y., Van Balkom, H., & Verhoeven, L. (2017). Core vocabulary of young children with Down syndrome. *Augmentative and Alternative Communication*, *33*(2), 77-86. https://doi.org/https://doi.org/10.1080/07434618.2017.1293730

Department-of-Basic-Education. (2011). *Curriculum and Assessment Policy Statement (CAPS) –Foundation Phase Home Language Grades R-3*. Government press.

Department-of-Basic-Education. (2022). *Basic Education Laws Amendment (BELA) Bill* Pretoria: Government Gazette No. 45601

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, *5*(1), 1-4. https://doi.org/10.11648/j.ajtas.20160501.11

Faab, G. (2010). *A morphosyntacic description of Northern Sotho as a basis for an automated translation from Northern Sotho into English* University of Pretoria]. http://hdl.handle.net/2263/28569

Fenson, L. (2007). MacArthur-Bates communicative development inventories.

Frick Semmler, B. J., Bean, A., & Wagner, L. (2023). Examining core vocabulary with language development for early symbolic communicators. *International Journal of Speech-Language Pathology*, 1-10. https://doi.org/https://doi.org/10.1080/17549507.2022.2162126

Hall, N., Juengling-Sudkamp, J., Gutmann, M. L., & Cohn, E. R. (2022). *Fundamentals of AAC: a case-based approach to enhancing communication*. Plural Publishing.

Hall, N., Juengling-Sudkamp, J., Gutmann, M. L., & Cohn, E. R. (2023). *Fundamentals of AAC : a case-based approach to enhancing communication*. Plural Publishing, Inc. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=3254594

https://public.ebookcentral.proquest.com/choice/PublicFullRecord.aspx?p=6949907

Hattingh, D. (2018). *The core vocabulary of South African Afrikaans-speaking preschoolers without disabilities* University of Pretoria (South Africa)]. https://www.proquest.com/dissertations-theses/core-vocabulary-south-african-afrikaans-speaking/docview/2901496395/se-2?accountid=14717

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/http://dx.doi.org/10.4102/sajcd.v67i1.701

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-based nursing*, *18*(3), 66-67. https://doi.org/https://doi.org/10.1136/eb-2015-102129

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences*, *33*(2-3), 61-83. https://doi.org/https://doi.org/10.1017/S0140525X0999152X

Johnson, E., Bornman, J., & Tönsing, K. (2017). Model for vocabulary selection of sensitive topics: An example from pain-related vocabulary. Seminars in Speech and Language,

Kathard, H., Ramma, L., Pascoe, M., Jordaan, H. L., Moonsamy, S., Wium, A.-M., Du Plessis, S., Pottas, L., & Khan, N. B. (2011). How can speech-language therapists and audiologists enhance language and literacy outcomes in South Africa?(And why we urgently need to). https://doi.org/http://hdl.handle.net/2263/18805

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/https://doi.org/10.1080/07434618.2020.1737964

Light, J., & McNaughton, D. (2012). Supporting the communication, language, and literacy development of children with complex communication needs: State of the science and future research priorities. *Assistive technology*, *24*(1), 34-44. https://doi.org/https://doi.org/10.1080/10400435.2011.648717

Loncke, F. (2020). *Augmentative and alternative communication: Models and applications* (Vol. 1). Plural publishing.

Lowe, N. K. (2019). What is a pilot study? *Journal of Obstetric, Gynecologic & Neonatal Nursing*, *48*(2), 117-118. https://doi.org/https://doi.org/10.1016/j.jogn.2019.01.005

Lund, S. K., Quach, W., Weissling, K., McKelvey, M., & Dietz, A. (2017). Assessment with children who need augmentative and alternative communication (AAC): Clinical decisions of AAC specialists. *Language, speech, and hearing services in schools*, *48*(1), 56-68. https://doi.org/https://doi.org/10.1044/2016_LSHSS-15-0086

Malebese, M. e. L., & Tlali, M. F. (2020). Teaching of English first additional language in rural learning environments: a case for problem-based learning. *International Journal of Inclusive Education*, *24*(14), 1540-1551. https://doi.org/https://doi.org/10.1080/13603116.2018.1544300

Marvin, C., Beukelman, D., & Bilyeu, D. (1994). Vocabulary-use patterns in preschool children: Effects of context and time sampling. *Augmentative and Alternative Communication*, *10*(4), 224-236. https://doi.org/https://doi.org/10.1080/07434619412331276930

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. https://doi.org/https://doi.org/10.3109/07434618.2010.481089

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, *22*(3), 276-282. https://hrcak.srce.hr/89395

Mesthrie, R. (2002). *Language in South Africa*. Cambridge University Press.

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

Mogatusi, M. G. (2022). *Determining the core vocabulary of Setswana-speaking Grade R learners as used during school activities* University of Pretoria (South Africa)]. https://www.proquest.com/dissertations-theses/determining-core-vocabulary-setswana-speaking/docview/2901813233/se-2?accountid=14717

Mojela, V. (2007). Polysemy and homonymy: Challenges relating to lexical entries in the Sesotho sa Leboa–English Bilingual Dictionary. *Lexikos, 17*. https://doi.org/https://doi.org/10.5788/17-0-537

Mojela, V. (2008). Standardization or stigmatization? Challenges confronting lexicography and terminography in Sesotho sa Leboa. *Lexikos*, *18*(1), 119-130. https://hdl.handle.net/10520/EJC60633

Mojela, V. (2013). A balanced and representative corpus: the effects of strict corpus-based dictionary compilation in Sesotho sa Leboa. *Lexikos*, *23*(1), 286-296. https://doi.org/https://hdl.handle.net/10520/EJC146571

Mojela, V. M. (1999). *Prestige terminology and its consequences in the development of Northern Sotho vocabulary* University of South Africa Pretoria, South Africa].

Mokgokong, P. C. (1966). *A dialect-geographical survey of the phonology of the Northern Sotho area* University of South Africa].

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.org/https://doi.org/10.1080/07434618.2019.1609577

Moorcroft, A., Scarinci, N., & Meyer, C. (2020). 'We were just kind of handed it and then it was smoke bombed by everyone': How do external stakeholders contribute to parent rejection and the abandonment of AAC systems? *International Journal of Language & Communication Disorders*, *55*(1), 59-69. https://doi.org/https://doi.org/10.1111/1460-6984.12502

Mothapo, N. R., kerstinTönsing, Refilwe Morwane. (2019). *Determining the core vocabulary used by Sepedi-speaking preschool children during regular preschool-based activities* [masters's thesis, university of pretoria]. http://hdl.handle.net/2263/71481

Mothapo, N. R., 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*, *23*(3), 295-304. https://doi.org/https://doi.org/10.1080/17549507.2020.1821774

Muttiah, N., Seneviratne, A., Drager, K. D., & Panterliyon, N. A. (2022). Parent perspectives on augmentative and alternative communication in Sri Lanka. *Augmentative and Alternative Communication*, *38*(3), 173-183. https://doi.org/https://doi.org/10.1080/07434618.2022.2046854

Nekoto, W., Marivate, V., Matsila, T., Fasubaa, T., Kolawole, T., Fagbohungbe, T., Akinola, S. O., Muhammad, S. H., Kabongo, S., & Osei, S. (2020). Participatory research for

low-resourced machine translation: A case study in african languages. *arXiv preprint arXiv:2010.02353*. https://doi.org/https://doi.org/10.48550/arXiv.2010.02353

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

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: debates and practical guidelines. *International journal of qualitative methods*, *19*, 1609406919899220. https://doi.org/https://doi.org/10.1177/1609406919899220

Ogbonnaya, U. I., & Awuah, F. K. (2019). QUINTILE RANKING OF SCHOOLS IN SOUTH AFRICA AND LEARNERS'ACHIEVEMENT IN PROBABILITY. *Statistics Education Research Journal*, *18*(1), 106-119. https://doi.org/https://doi.org/10.52041/serj.v18i1.153

Omair, A. (2015). Selecting the appropriate study design for your research: Descriptive study designs. *Journal of health specialties*, *3*(3), 153. https://doi.org/10.4103/1658-600X.159892

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, 45*, 927-937. https://doi.org/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? https://doi.org/http://hdl.handle.net/11427/19940

Pascoe, M., Rogers, C., & Norman, V. (2013). Are we there yet? On a journey towards more contextually relevant resources in speech-language therapy and audiology. https://doi.org/http://hdl.handle.net/11427/19941

Poulos, G., & Louwrens, L. J. (1994). *A linguistic analysis of Northern Sotho*. Via Afrika.

Prinsloo, D. J. (2014). Lexicographic treatment of kinship terms in an English/Sepedi–Setswana–Sesotho dictionary with an amalgamated lemmalist. *Lexikos*, *24*, 272-290. https://doi.org/https://doi.org/10.5788/24-1-1263

Rakgogo, T. J., & van Huyssteen, L. (2018). Exploring the Northern Sotho language name discrepancies in informative documentation and among first language speakers. *South African Journal of African Languages*, *38*(1), 79-86. https://doi.org/ https://doi.org/10.1080/02572117.2018.1429871

Rakgogo, T. J., & Zungu, E. B. (2022). The elevation of Sepedi from a dialect to an official standard language: Cultural and economic power and political influence matter.

Literator-Journal of Literary Criticism, Comparative Linguistics and Literary Studies, *43*(1), 1827.

Riccelli-Sherman, A. (2017). *Using a core vocabulary intervention to improve communication of students who use Augmentative and Alternative Communication (AAC)* University of St. Francis].

Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of child language*, *14*(2), 201-209. https://doi.org/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/https://doi.org/10.3109/07434618.2014.921240

SekhukhuneDistrictMunicipality. (2017). *Sekhukhune District Language Policy*.

Sengani, T. M. (2013). Controversies around the so-called alliterative concord in African languages: A Critical Language Awareness on communicative competence with specific reference to Tshivenda 1. *South African Journal of African Languages*, *33*(2), 189-201. https://doi.org/10.1080/02572117.2013.871461

Sevcik, R., Romski, M., & Adamson, L. (2004). Research directions in augmentative and alternative communication for preschool children. *Disability and rehabilitation*, *26*(21-22), 1323-1329. https://doi.org/https://doi.org/10.1080/09638280412331280352

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

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

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/https://doi.org/10.1080/09638280500077747

Solomon-Rice, P. L., Soto, G., & Heidenreich, W. (2017). The impact of presupposition on the syntax and morphology of a child who uses AAC. *Perspectives of the ASHA Special Interest Groups*, *2*(12), 13-22. https://doi.org/https://doi.org/10.1044/persp2.SIG12.13

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/https://doi.org/10.1080/07434618.2021.1881822

Soto, G., & Tönsing, K. (2024). Is there a 'universal'core? Using semantic primes to select vocabulary across languages in AAC. *Augmentative and Alternative Communication*, *40*(1), 1-11. https://doi.org/https://doi.org/10.1080/07434618.2023.2243322

Stadskleiv, K., Batorowicz, B., Sandberg, A. D., Launonen, K., Murray, J., Neuvonen, K., Oxley, J., Renner, G., Smith, M. M., & Soto, G. (2022). Aided communication, mind understanding and co-construction of meaning. *Developmental neurorehabilitation*, *25*(8), 518-530. https://doi.org/https://doi.org/10.1080/17518423.2022.2099030

Statista. (2023). *Smartphone users in South Africa from 2014 to 2023* Retrieved 09 August from

Statistics, S. A. (2022). *Census 2022*. Retrieved 30 July from https://census.statssa.gov.za/#/

StatsSA. (2022). *Census*. Pretoria

Stuart, S., Vanderhoof, D., & Beukelman, D. (1993). Topic and vocabulary use patterns of elderly women. *Augmentative and Alternative Communication*, *9*(2), 95-110. https://doi.org/https://doi.org/10.1080/07434619312331276481

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

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

Thompson, C. B., & Panacek, E. A. (2007). Research study designs: Non-experimental. *Air medical journal*, *26*(1), 18-22. https://doi.org/https://doi.org/10.1016/j.amj.2006.10.003

Tönsing, K. M., Van Niekerk, K., Schlünz, G. I., & Wilken, I. (2018). AAC services for multilingual populations: South African service provider perspectives. *Journal of communication disorders*, *73*, 62-76. https://doi.org/https://doi.org/10.1016/j.jcomdis.2018.04.002

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/https://doi.org/10.1080/13668250701689298

Tsai, M.-J. (2023). Core vocabulary for AAC practice from Mandarin Chinese-speaking Taiwanese without disabilities. *Augmentative and Alternative Communication*, *39*(2), 73-83. https://doi.org/https://doi.org/10.1080/07434618.2023.2199855

van Tilborg, A., & Deckers, S. R. (2016). Vocabulary selection in AAC: Application of core vocabulary in atypical populations. *Perspectives of the ASHA Special Interest Groups*, *1*(12), 125-138. https://doi.org/https://doi.org/10.1044/persp1.SIG12.125

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

Wierzbicka, A. (1992). The semantics of interjection. *Journal of pragmatics*, *18*(2-3), 159-192. https://doi.org/https://doi.org/10.1016/0378-2166(92)90050-L

Wofford, M. C., Ogletree, B. T., & De Nardo, T. (2022). Identity-focused practice in augmentative and alternative communication services: A framework to support the intersecting identities of individuals with severe disabilities. *American journal of speech-language pathology*, *31*(5), 1933-1948. https://doi.org/https://doi.org/10.1044/2022_AJSLP-21-00397

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/https://doi.org/10.1080/07434618812331274807

Yurtbaşı, M. (2015). Building English vocabulary through roots, prefixes and suffixes. *Global Journal of Foreign Language Teaching*, *5*(1), 44-51. https://doi.org/http://dx.doi.org/10.18844/gjflt.v5i0.39

Appendix A
Approval by the Research Ethics Committee of the Faculty of Humanities

APPENDIX A:
Approval by the research ethics committee of the Faculty of Humanities

Faculty of Humanities

14 November 2023

Dear Miss CM Moswathupa

Project Title:	Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune District during regular school activities
Researcher:	Miss CM Moswathupa
Supervisor(s):	Prof KM Tönsing
Department:	Centre for Augmentative and Alternative Communication
Reference number:	23871165 (HUM024/0723)
Degree:	Masters

I have pleasure in informing you that the above application was **approved** by the Research Ethics Committee on 14 November 2023. Please note that before research can commence all other approvals must have been received.

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

Appendix B
Limpopo Department of Basic Education permission letter

LIMPOPO
PROVINCIAL GOVERNMENT
REPUBLIC OF SOUTH AFRICA

DEPARTMENT OF
EDUCATION

CONFIDENTIAL

Ref: 2/2/2 Enq: Makola MC Tel No: 015 290 9448 E-mail:MakolaMC@edu.limpopo.gov.za
Moswathupa M
Private Bag X20
Hatfield
0028

RE: REQUEST FOR PERMISSION TO CONDUCT RESEARCH

1. The above bears reference.
2. The Department wishes to inform you that your request to conduct research has been approved. Topic of the research proposal: **"Determining the core vocabulary of Sepedi-speaking Grade R learners from the Sekhukhune district during regular school activities."**

3.The following conditions should be considered:

3.1 The research should not have any financial implications for Limpopo Department of Education.

3.2 Arrangements should be made with the Circuit Office and the School concerned.

3.3 The conduct of research should not in anyhow disrupt the academic programs at the schools.

3.4 The research should not be conducted during the time of Examinations especially the fourth term.

3.5 During the study, applicable research ethics should be adhered to; in particular the principle of voluntary participation (the people involved should be respected).

3.6 Upon completion of research study, the researcher shall share the final product of the research with the Department.

Cnr 113 Biccard & 24 Excelsior Street, POLOKWANE, 0700, Private Bag X 9489, Polokwane, 0700
Tel:015 290 7600/ 7702 Fax 086 218 0560

The heartland of Southern Africa-development is about people

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

APPENDIX C

Principle information letter and permission form

Date: _____

Dear Principal

PERMISSION TO CONDUCT RESEARCH AT YOUR SCHOOL PREMISES

My name is Charmaine Moswathupa. I am a Master's student at the Centre for Alternative and Augmentative Communication at the University of Pretoria. During my studies, I will be conducting a small-scale research project. The title of my study is "Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities". This study is an extension of a study that was conducted by Mothapo et al,. (2021) in the Capricorn region.

I would like to request permission to recruit participants from your school, and to conduct the research on the school premises during school time. I have obtained approval to conduct the research from the Limpopo Department of Education (attached approval).

The rationale of the study

Children who cannot speak adequately to meet all their communication needs require other ways to express themselves. One alternative form includes communication boards or electronic communication software with picture symbols. The selection of relevant words to be included on such boards or software is an important process as it will allow the individual to effectively communicate across different situations in their society (school, home, and playgrounds). My research aims to obtain the Sepedi words used daily by children who can speak. I want to compile a list of these frequently used words so that speech therapists and AAC experts can use this list to select words for communication boards or electronic systems for children who cannot speak.

What will I expect from your school?

I will request the Grade R teacher to complete a short preschool background questionnaire regarding the daily routine and language(s) used in class. I will also ask the teacher to assist me in nominating three learners aged 5 to 6 years from their class who could participate in this study. The teacher will be asked to send the information letters and consent forms to the caregivers of the nominated learners and return them to me after completion.

When the caregivers and learners agree to be a part of the study, I will discuss suitable times for recording the children's speech with the teacher. On the days agreed upon, the teacher and I will fit the participants with a microphone and small voice recorder (in a pouch fastened around the waist) in the morning. The recorder will be switched on to record the words that the child is using throughout the school day. The recorders will be removed before the child goes home.

The teacher will be asked to monitor the child during the day to ensure that the recording equipment does not interfere with their activities. Teachers may remove the equipment at any time if they feel it is interfering with the child's activities or if it is inappropriate to record the

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

activity. Children may also request the teacher to assist them if they experience discomfort and they may also request the teacher to remove the equipment. The period of collecting data will be no more than 5 school days.

What will be expected of the participants during the study?
The nominated learners whose caregivers have consented to the study and the teacher of these learners will meet with me. I will use a child-friendly picture-based information sheet to explain the processes that we will follow throughout the data collection period. I will then use pictures to ask the learners if they want to participate in the study, and they will be allowed to respond using the pictures or verbally.

If the learners agree, they will be fitted with the recorder in the pouch on their waist and the microphone on their collar to record their language throughout the selected days. The participants and their classmates will be instructed not to play with the equipment. They may request help from their teacher if the recorders cause them discomfort or if they want to stop taking part in the study. The teacher will adjust or remove the recording equipment. Learners may stop participating at any time without any negative consequences to themselves.

The following ethical principles will be upheld in the study:

- Conditional approval for the study has been obtained from the Ethics Board of the Faculty of Humanities, University of Pretoria, and from the Limpopo Department of Education.
- Written consent from the participants' caregivers and verbal assent from the participant will be obtained prior to the commencement of the research.
- Participants and their caregivers will be informed of all their rights and that they may withdraw from the study at any point without any negative consequences.
- Confidentiality: The speech samples obtained from the learners will be accessed by the researcher, a research assistant and the supervisor only, unless caregivers give written permission to share these with other researchers. The identifying information of the participants and the school will be removed from the transcriptions and be replaced with numbers. No individual or school names will be mentioned in any published data.

Who will have access to the study?
The recordings collected will be securely stored on an external drive format at the Centre for AAC (University of Pretoria). The participants' recordings may be shared with other researchers with the caregivers' written permission, and these researchers are obliged to 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/) and on the University of Pretoria research data repository and platform (https://researchdata.up.ac.za/). The data collected will be used to write a research report as a Master's mini dissertation and write scientific papers. It may also be presented at professional conferences. A summary (with the de-identified data with no personal information) of the study findings will be made available for caregivers and staff members on request.

What are the risks and benefits of the study?
The study poses no risk to the participants, their classmates, and teachers. Care will be taken not to interfere with any school activities. The equipment used does not pose any risk of harm, as the padded pouches will be comfortably fitted on the participants' waists. The equipment may be removed at any point when the participants experience any discomfort. The teacher will be

encouraged to use his/her discretion to identify any discomfort or unsettledness and remove the equipment immediately.

The study does not benefit the participants or school directly, however the findings have a potential of providing speech language therapists and other professions in both Department of Basic Education and Department of Health with guidelines for Sepedi vocabulary selection when customizing a communication system for children who use Sepedi.

I would appreciate it if you could complete the attached form to indicate whether you give permission or not to include participants at your school in the study. For any further information, please contact me or my supervisor using the contact details below.
Kind regards

02/11/2023

Charmaine Moswathupa
Master's student in AAC
Cell No: 0626059781
Email address: mmotomoswathupa@gmail.com

Date

02/11/2023

Professor Kerstin Tönsing
Centre for Alternative and Augmentative Communication
Tel: 012 420 4729
Email address: 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

APPENDIX C: Principal permission form

Name of the School: _____

Name of the Principal: _____

Title of study: **Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities**

Researcher: Charmaine Moswathupa
Master's student at the Centre for AAC
University of Pretoria
Cell: 0626059781
Email: mmotomoswathupa@gmail.com

Supervisor: Professor Kerstin Tönsing
Professor at the Centre for AAC
University of Pretoria
Tel: 012 420 4729
Email: Kerstin.tonsing@up.ac.za

I _____ **(Name and surname)**

(please tick relevant option:)

☐ give permission to conduct the study titled: **Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities** at the above-mentioned school. I understand that the study will be conducted by Charmaine Moswathupa under the supervision of professor Kerstin Tönsing. My permission is voluntary, and I understand that I may withdraw it at any time. I understand that Grade-R teachers will be requested to recruit learners, that learners will be audio-recorded throughout the school days, and that the teachers may assist with fitting and removing the recorders. I understand that the data obtained will be stored at the Centre for AAC for 15 years and it will be treated with strict confidentiality. I understand that the data may also be used for research presentations, reports, and scientific articles. 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 on the University of Pretoria research data repository and platform (https://researchdata.up.ac.za/) 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 Charmaine Moswathupa to recruit learners and assistance from the teachers from the school mentioned above for possible participation in the study titled: **Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities.**

School stamp

_____
Principal Signature

_____
Date

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

Appendix D
Teacher 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

APPENDIX D
Teacher consent letter

Dear class teacher **Date:_____**

Consent to nominate learners for participation and for recordings to take place in the classroom.

My name is Charmaine Moswathupa. I am a Master's student at the Centre for Alternative and Augmentative Communication at the University of Pretoria. During my studies, I will be conducting research project on a personal preferred topic. My topic of interest focuses on obtaining and analyzing the core vocabulary of Sepedi-speaking Grade R learners during everyday school activities within the Sekhukhune region. The topic of my study is "Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities". This study is an extension of an exciting study that was conducted by Mothapo et al. (2021) in the Capricorn region.

I would like to request permission and assistance in nominating possible study participants from your classroom, to conduct the research within the classroom and during conduct time. I have obtained approval to conduct the research from the Limpopo Department of Education (attached approval).

The rationale of the study
Individuals with complex communication needs require alternative forms of expressive communication other than speech. This alternative includes aided systems such as picture communication or computer-based communication boards. The selection of vocabulary to be included on the aided system is an important process as it will allow the individual to effectively communicate across different situations in their society (school, home, and playgrounds). My research aims to obtain the Sepedi words used daily by children to compile a core vocabulary list that may assist in knowing the relevant words to be included. This list may be a guideline for speech therapists and AAC experts when selecting vocabulary for children with speech and language challenges.

What will I expect from you?
I will ask you to complete a background questionnaire about the languages used in the Grade R classroom, about the daily routine in class as well as the curriculum, about class size and also about some of the facilities at the school. This should take no more than 10 minutes of your time. I will then ask for your assistance in nominating three learners aged 5;0 to 6 years and 11 months old from your class who could be my participants in this study. You will be asked to send the information letters and consent forms to the caregivers of the nominated learners and return them to me after completion.

When the caregivers and learners agree to be a part of the study, I will ask you to assist with fitting the participants with the microphone and pouch carrying the audio-recorder on their waist in the morning and they will be removed in the afternoons. The recorders will record the words

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

that the specific learner wearing the recorder is using throughout the school day. This data collection procedure will not interfere with the learners' school activities.

During the date collection period, you will be asked to assist the participants with any technical issues they may experience and remove/switch off the recorders upon their wish. You may also remove the equipment should you feel that the equipment is inappropriate to wear or that recording is inappropriate during specific activities. I will communicate with you daily before fitting the equipment to provide any clarity and questions arising from the previous day. The period of collecting data will be no more than 5 school days.

What will be expected of the participants during the study?
I will have a meeting with you and the nominated learners, once their parents/caregivers have given consent for them to participate. During the meeting, I will then use a child-appropriate picture-based information sheet to explain the processes that will follow throughout the data collection period. I will then use pictures to ask the learners to be a part of the study, and they will be allowed to respond using the pictures or verbally. If the learners agree, they will be fitted with the pouch on their waist and the microphone on their collar to record their language throughout the day. The participants and their classmates will be required not to play with the equipment and should inform you if they experience any issues or discomfort.

The following ethical principles will be upheld in the study:

- Confidentiality- The speech samples (recordings) obtained from the learners will be accessed by the researcher, a research assistant and the supervisor only, unless caregivers give written permission to share these with other researchers. The identifying information of the participants and the school will be removed from the transcriptions and be replaced with numbers.
- Written consent from the participant's caregivers and verbal assent from the participant will be obtained prior to the commencement of the research.
- Participants and their caregivers will be informed of all their rights and that they may withdraw from the study at any point.

Who will have access to the study?
The data collected will be securely stored on an external drive format at the Centre for AAC (University of Pretoria). The participant's voice recordings may be shared with other researchers with the caregivers' written permission, and these researchers are obliged to keep the recordings confidential. 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 on the University of Pretoria research data repository and platform (https://researchdata.up.ac.za/). The data collected will be used to write a research report as a Master's Mini dissertation and write scientific papers. A summary (with the de-identified data with no personal information) of the study findings will be made available for caregivers and staff members on request.

What are the risks and benefits of the study?
The study poses no risk to the participants, their classmates, and teachers. The equipment used does not exert any harm, the pouches will be comfortably fitted on the participant's waist. The equipment may be removed at any point when the participants experience any discomfort. The teacher will be encouraged to use his/her discretion to identify any discomfort or unsettle and stop the voice recordings to ensure voluntary participation.

The study does not benefit the participants or school, however the findings have a potential of influencing the AAC field in both Department of Basic Education and Department of health by proving them with guidelines for Sepedi vocabulary selection when customizing aided AAC systems for children who uses Sepedi.
I would appreciate it if you could complete the attached form to indicate whether you give permission or not to include participants from your class in the study. For any further information, please contact me or my supervisor using the contact details below.
Kind regards

Charmaine Moswathupa
Master's student in AAC
Cell No: 0626059781
Email address: mmotomoswathupa@gmail.com

Date

Pı rstin Tönsing
Centre for Alternative and Augmentative Communication
Tel: 012 420 4729
Email address: 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

Name of the School: _____

Name of the Teacher: _____

Title of study: **Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities**

Researcher: Charmaine Moswathupa
Master's student at the Centre for AAC
University of Pretoria
Cell: 0626059781
Email: mmotomoswathupa@gmail.com

Supervisor: Professor Kerstin Tönsing
Professor at the Centre for AAC
University of Pretoria
Tell: 012 420 4729
Email: Kerstin.tonsing@up.ac.za

I _____ **(Name and surname)**
(Please tick relevant option:)

☐ give consent to assist the researcher in recruiting learners from my class for possible participation in the study titled: **Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities ,** which will be conducted by Charmaine Moswathupa under the supervision of professor Kerstin Tönsing. I understand the information provided by the researcher, that Grade-R learners will be audio-recorded throughout the school days, that the teachers may assist with putting and removing the recorders and that participation is voluntary. I understand that I may withdraw from the study at any time.
I understand that the data obtained will be stored at the Centre for AAC for 15 years and it will be treated with strict confidentiality. I understand that the data may also be used for research presentations, reports, and scientific articles. I understand that the transcription of the audio recording with all learners personal information removed will be made available on the website of the South African Digital Language Resource Centre (https://repo.sadilar.org/) and on the University of Pretoria research data repository and platform (https://researchdata.up.ac.za/)and may be used for further analysis I understand that data may be used for further analysis and with the permission of the caregiver, the data will be made available to other researchers.

OR

☐ I do not give consent to assist Charmaine Moswathupa in recruiting learners from my class for possible participation in the study titled: **Determining the core vocabulary of Sepedi-speaking Grade R learners from the Sekhukhune district during regular school activities** .

_____ _____
Teacher Signature Date

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

Appendix E
Preschool background and teachers' questionnaire

APPENDIX E: PRESCHOOL BACKGROUND QUESTIONNAIRE

(Based on Mogatusi, 2023)

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

Date: _____

Name of class teacher: _____

Name of School: _____

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

Information about the language(s) used at the school.

1. Is Sepedi the primary language of instruction used within the school premises?

   ☐ YES

   ☐ NO

2. Do you use Sepedi during teaching and learning (contact time)?

   ☐ YES

   ☐ NO

3. What other languages do you use during contact time? 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 to communicate 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

17. If yes, please describe the daily program:_____

_____

Information about the facilities at the preschool

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

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

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

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

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

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

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

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

☐ YES

☐ NO

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

27. Is the preschool fenced? ☐ YES ☐ NO

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

A landline	A telephone	A fax machine	Internet

Nomination of participants

The study is aimed at obtaining speech and language sample from Sepedi-Speaking Grade R learners who are of ages between 5years and 6years and 11months old. Please nominate 3 learners from your class. The nominated learners should be of both genders (boys and girls) and meet the following criteria:

- Speak Sepedi as their First/home language,
- Tend to be talkative,
- Have been enrolled in Grade R for at least one month,
- Live in the Sekhukhune region,
- Have adequate speech and language skills for their age.

Please send the received packages containing the caregiver information letter, consent form and questionnaires to each of the nominees' caregivers/parents. The researcher will make arrangements to collect any completed consent forms from the school.

Thank you for your assistance.

Appendix F
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

APPENDIX F1
Caregiver information letter and consent form (English version)

Date: _____

Dear Sir/Madam

Request to allow your child to participate in a study

My name is Charmaine Moswathupa. I am a Master's student at the Centre for Alternative and Augmentative Communication at the University of Pretoria. During my studies, I will be conducting a research project. I am interested in finding out what words Sepedi-speaking Grade R learners use every day during school activities. The title of my study is "Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities".

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

Why is this study important?
Some children with disabilities struggle to speak. These children need other ways of communicating with those around them. For example, we may give them a computer or a board with pictures that represent words or messages to help them communicate.

It is important that we choose appropriate words to include in the communication systems we give to the children. Without the right words, they cannot communicate with those around them. In this study, I want to find out which words children without disabilities use every day. I want to compile a list of these words that children without disability are using so that it can be used to guide speech therapists when making communication systems for children who cannot speak. Including such words on the communication systems we give to children who cannot speak will ensure that they can use the systems to communicate in the same way as other children use speech.

What will be expected of me and my child should I give permission to participate?
I will ask you to complete a background questionnaire to tell me about your child's development and also about the languages he//she uses in daily life. This should take no more than 10 minutes of your time. I will have a meeting with your child and the class teacher. I will introduce myself and explain the study to your child in Sepedi using pictures to help him/her understand. I will then use pictures to ask your child if he/she wants to be a part of the study, and your child can respond using the pictures or by saying yes or no. Your child is free to choose if he/she wants to take part or not.

If your child agrees, he/she will be expected to carry a pouch/bag on their waist (which has a small audio-recorder) and a microphone on the collar of their shirt to record what they say throughout the day. The period of recording will be no more than five days. This is to determine the words that your child is using during his/her typical school 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

Your child and his/her classmates will be asked not to play with the equipment and should inform their teacher if they experience any issues or discomfort. Should your child feel any discomfort, his/her class teacher will be able to remove the audio-recorder and microphone. His/her teacher may also remove the equipment if they interfere with the learning activities.

What are my child's rights?
Your child is not forced to participate in this study. I will make sure your child understands everything about the study. You or your child can choose to not be a part of this study and you can later choose to stop participating in the study. Your child can ask the teacher at any time to remove the equipment. Nothing bad will happen to you or your child if you decide to stop.

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, a research assistant and my supervisors. Voice recordings will only be shared with other researchers if you give permission (see the form on the last page). 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 of your child 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 information received from you and your child will be securely stored on an external drive format at the Centre for AAC (University of Pretoria). Voice recordings will only be shared with other researchers if you give me permission (see the form on the last page).

The written form of what your child said (with all personal data like names, places etc. removed) will be shared on the website of the South African Digital Language Resource Centre (https://repo.sadilar.org/) and on the University of Pretoria research data repository and platform (https://researchdata.up.ac.za/) to help other researchers in understanding the Sepedi language development in children. Should you wish to obtain a copy of the written form of what your child said, you may contact me and I will provide it.

The information I receive from you and your child will be used to write a Master's Mini dissertation and scientific articles. I may also present the information at conferences. However, I will never mention your child's name or any other personal information when I write or speak about the study.

If you would like to receive a copy of your child's recording or the written form of what your child said, kindly let me know, and I will supply it to you. I will also give you a summary of the study if you so wish.

What are the risks and benefits of the study?
The study carries no harm to your child, their classmates, and teachers. No one will be harmed during the study. The equipment used is not harmful, the pouches will be comfortably fitted on the participant's waist. The study will take place during daily school activities (in the classroom, playground, etc.) and will not interfere with your child's learning. The equipment may be

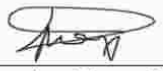
removed at any point when your child experiences any discomfort. His or her teacher will be monitoring your child to check for any discomfort and stop the voice recordings to make sure that your child participates freely in the classroom.

The study does not directly benefit you, your child, or the school; however, the findings have the potential for assisting us to know which words we should include when making communication systems for children who uses Sepedi.

I would appreciate if you could complete the attached consent form to let me know if you give permission for your child to take part in the study or not. Please return the form to your child's class teacher.

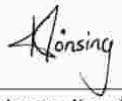
For any further information, please contact me or my supervisor using the contact details below.

Kind regards



Charmaine Moswathupa
Master's student in AAC
Cell No: 0626059781
Email address: mmotomoswathupa@gmail.com

Date



Professor Kerstin Tönsing
Centre for Alternative and Augmentative Communication
Tel: 012 420 4729
Email address: 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

Parental Consent form

Name of child: _____

Name of caregiver: _____

Title of study: Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities

Researcher: Charmaine Moswathupa
Master's student at the Centre for AAC
University of Pretoria
Cell: 0626059781
Email: mmotomoswathupa@gmail.com

Supervisor: Professor Kerstin Tönsing
Professor at the Centre for AAC
University of Pretoria
Tell: 012 420 4729
Email: Kerstin.tonsing@up.ac.za

I _____ (Name and surname)
(please tick applicable box)

☐ give consent for me and my child to participation in the study titled: Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities , which will be conducted by Charmaine Moswathupa under the supervision of Professor 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 be stored for 15 years at the CAAC and that all data will be treated confidentially. I understand that my child will be audio-recorded 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 to allow me and my child to participation in the study titled: Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities.

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.*

_____ _____
Caregiver Signature Date

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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Centre for Augmentative and
Alternative Communication

APPENDIX F2
Caregiver information letter and consent form (Sepedi version)

letsatsikgwedi: _____

Mohlomphegi/ Mohumagadi

Tumelelo ya go dumelela ngwana wagago go tsea karolo thutong.

Leina laka ke Charmaine Moswathupa. Ke moithuti wa Master's kua Center for Alternative and Augmentative Communication ka Yunibesithing ya Pretoria. Nakong ya dithuto tša ka, ke tla be ke swara projeke ya nyakišišo. Ke na le kgahlego ya go hwetša gore ke mantšu afe ao barutwana ba Mphato wa R bao ba bolelago Sepedi ba a šomišago letšatši le letšatši nakong ya mediro ya sekolo. Thaetlele ya thuto ya ka ke "Go laetša tlotlontšu ya motheo ya baithuti ba Mphato wa R bao ba bolelago Sepedi go tšwa seleteng sa Sekhukhune nakong ya mediro ya ka mehla ya sekolo ".

Ke rata go kgopela tumelelo ya go akaretša ngwana wa gago thutong ye.

Ke ka baka la'ng thuto ye e le bohlokwa?
Bana ba bangwe bao ba golofetšego ba palelwa ke go bolela. Bana ba ba hloka ditsela tše dingwe tša go boledišana le bao ba ba phelago le bona. Mohlala, re ka ba nea khomphutha goba puku eo e nago le diswantšho tšeo di emelago mantšu goba melaetša bakeng sa go ba thuša go boledišana.

Go bohlokwa gore re kgethe mantšu a maleba ao re swanetšego go a akaretša ditshepedišong tša poledišano tšeo re di fago bana. Ka ntle le mantšu a swanetšego, ba ka se kgone go boledišana le bao ba ba lego kgauswi le bona. Thutong ye, ke nyaka go hwetša gore ke mantšu afe ao bana bao ba se nago bogole ba a šomišago letšatši le letšatši. Ke nyaka go kgoboketša lenaneo la mantšu a ao bana bao ba se nago bogole ba a šomišago gore le kgone go šomišwa go hlahla baalafi ba polelo ge ba direla bana bao ba sa kgonego go bolela ditshepedišo tša kgokagano. Go akaretša mantšu a bjalo ditshepedišong tša poledišano tšeo re di fago bana bao ba sa kgonego go bolela go tla netefatša gore ba kgona go šomiša ditshepedišo tšeo go boledišana ka tsela yeo bana ba bangwe ba šomišago polelo.

Ke eng seo se tlago letelwa go nna le ngwana wa ka ge nka fa tumelelo ya go tšea karolo?
Ke tla go kgopela gore o tlatše lenaneopotšišo la morago go mpotša ka ga kgolo ya ngwana wa gago le gape ka maleme ao a a šomišago bophelong bja letšatši le letšatši. Se ga se sa swanela go tšea metsotso e fetago e 10 ya nako ya gago. Ke tla ba le kopano le ngwana wa gago le morutiši wa ka phapošing. Ke tla itsebiša le go hlalosetša ngwana wa gago thuto ka Sepedi ke šomiša diswantšho go mo thuša go kwešiša. Ke moka ke tla šomiša diswantšho go botšiša ngwana wa gago ge eba a nyaka go ba karolo ya thuto, gomme ngwana wa gago a ka araba ka go šomiša diswantšho goba ka go re ee goba aowa. Ngwana wa gago o lokologile go kgetha ge e ba a nyaka go tšea karolo goba aowa.

Ge ngwana wa gago a dumela, go tla letelwa gore a rwale mokotla lethekeng la gagwe (wo o nago

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

le audio-recorder ye nnyane) le maekrofouno molala wa hempe ya gagwe go rekota seo a se bolelago letšatši ka moka. Nako ya go rekota e ka se fete matšatši a mahlano. Se ke go hwetša mantšu ao ngwana wa gago a a šomišago nakong ya letšatši la gagwe la sekolo le le lwaelegilego.

Ngwana wa gago le bao a tsenago le bona ka klaseng ba tla kgopelwa gore ba se bapale ka didirišwa gomme o swanetše go tsebiša morutiši wa bona ge a sa iketle. Ge ngwana wa gago a ka ikwa a sa iketla, morutiši wa gagwe wa ka phapošing o tla kgona go tloša sedirišwa sa go rekota modumo le maekrofouno. Morutiši wa gagwe le yena a ka tloša didirišwa ge di šitišana le mediro ya go ithuta.

Ditokelo tša ngwana wa ka ke dife?
Ngwana wa gago ga a gapeletšege go tšea karolo thutong ye. Ke tla kgonthiša gore ngwana wa gago o kwešiša se sengwe le se sengwe ka thuto. Wena goba ngwana wa gago le ka kgetha go se be karolo ya nyakišišo ye gomme ka morago o ka kgetha go kgaotša go tšea karolo thutong. Ngwana wa gago a ka kgopela morutiši nako efe goba efe gore a ntšhe didirišwa. Ga go selo se sebe seo se tlago go go diragalela goba go ngwana wa gago ge e ba o dira phetho ya go kgaotša.

Leina la ngwana wa gago, gammogo le tshedimošo efe goba efe ye nngwe ya motho, e tla hwetšagala fela go monyakišiši ka mabaka a taolo. Tshedimošo ka moka ya motho ka noši le direkhoto tša ngwana wa gago di tla bolokwa di bolokegile gomme di ka se abelanwe le motho le ge e le ofe. Dikgatišo tša lentšu di tla theeletšwa feela ke nna, mothuši wa nyakišišo, le baokamedi ba ka. Direkoto tša lentšu di tla abelanwa fela le banyakišiši ba bangwe ge o fa tumelelo (bona foromo yeo e lego letlakaleng la mafelelo). Tshedimošo efe goba efe ya motho ka noši go swana le maina a batho le mafelo e tla tlošwa ge ke ngwala mantšu ao ngwana wa gago a a dirišitšego.

Foromo ye e ngwadilwego ya seo ngwana wa gago a se boletšego (ka tshedimošo ka moka ya motho ka noši e tlošitšwe) e tla hwetšagala mo wepsaeteng ya Senthara ya Methopo ya Polelo ya Ditšitale ya Afrika Borwa (https://repo.sadilar.org/) le ka polokelong ya datha ya nyakišišo ya Yunibesithi ya Pretoria le sefala (https://researchdata.up.ac.za/), lebaka la se ke gore banyakišiši ba bangwe ba ka kgona go šomiša tshedimošo ye go kwešiša kgolo ya polelo ya bana. Ge o nyaka go hwetša khopi ya foromo ye e ngwadilwego ya seo ngwana wa gago a se boletšego, o ka ikgokaganya le nna gomme ke tla go fa.

Lega go le bjalo, ga go na tsebišo ya motho ka noši ya ngwana wa gago yeo e tlago go abelanwa tshepedišong yeo. Ge ke bolela, goba ke ngwala ka thuto, ga go na tshedimošo ya motho ka noši ka ngwana wa gago, wena goba sekolo yeo e tlago abelanwa.

Go tla direga eng ka morago ga ge ke kgobokeditše tshedimošo?
Tshedimošo yeo e amogetšwego go tšwa go wena le ngwana wa gago e tla bolokwa ka polokego ka sebopegong sa drive ya ka ntle kua Center for AAC (University of Pretoria). Dikgatišo tša lentšu di tla abelanwa fela le banyakišiši ba bangwe ge o mpha tumelelo (bona foromo yeo e lego letlakaleng la mafelelo).

Foromo ye e ngwadilwego ya seo ngwana wa gago a se boletšego (ka data ka moka ya motho go swana le maina, mafelo bjo bjo. e tlošitšwe) e tla abelanwa mo wepsaeteng ya Senthara ya Methopo ya Polelo ya Ditšitale ya Afrika Borwa (https://repo.sadilar.org/) go thuša ba bangwe.

Tshedimošo yeo ke e hwetšago go tšwa go wena le ngwana wa gago e tla šomišwa go ngwala

lengwalo la Master's Mini le dihlogo tša mahlale. Nka tšweletša gape tshedimošo dikhonferenseng. Lega go le bjalo, nka se tsoge ke boletše leina la ngwana wa gago goba tsebišo le ge e le efe e nngwe ya motho ka noši ge ke ngwala goba ke bolela ka thuto yeo.

Ge e ba o rata go amogela khopi ya kgatišo ya ngwana wa gago goba foromo ye e ngwadilwego ya seo ngwana wa gago a se boletšego, ka botho ntsebiše, gomme ke tla go aba yona. Gape ke tla go fa kakaretšo ya thuto ge o rata.

Ke dikotsi le mehola efe ya nyakišišo ye?
Thuto ye ga e rwale kotsi go ngwana wa gago, bao ba tsenago le bona ka klaseng le barutiši. Ga go na motho yo a tlago go gobala nakong ya thuto, didirišwa tšeo di šomišitšwego ga di kotsi, mekotla e tla tsenywa gabotse lethekeng la motšwasehlabelo. Thuto e tla direga nakong ya mediro ya sekolo ya letšatši le letšatši (ka phapošing ya borutelo, lepatlelong, bj.bj.) gomme e ka se šitiše thuto ya ngwana wa gago. Didirišwa di ka tlošwa nakong le ge e le efe ge ngwana wa gago a itemogela go se iketle le ge e le gofe.

Morutiši wa gagwe o tla be a beile ngwana wa gago leihlo go lekola go se iketle le ge e le gofe le go emiša direkhoto tša lentšu go kgonthiša gore ngwana wa gago o tšea karolo ka bolokologi ka phapošing ya borutelo. Thuto yeo ga e hole wena, ngwana wa gago goba sekolo; le ge go le bjalo, dikhwetšo di na le bokgoni bja go re thuša go tseba gore ke mantšu afe ao re swanetšego go a akaretša ge re dira ditshepedišo tša poledišano tša bana bao ba šomišago Sepedi.

Nka thabela ge o ka tlatša foromo ya tumelelo yeo e kgomareditšwego go ntsebiša ge e ba o nea ngwana wa gago tumelelo ya go tšea karolo thutong goba go se bjalo. Hle bušetša foromo yeo go morutiši wa ka phapošing ya ngwana wa gago.
Bakeng sa tshedimošo efe goba efe ye nngwe, hle ikopanye le nna goba mookamedi wa ka o šomiša dintlha tša kgokagano tše di lego ka mo tlase.

Wa lena,

Charmaine Moswathupa
Moithuti wa Master's ka AAC
Nomoro ya sele: 0626059781
Aterese ya imeile : mmotomoswathupa@gmail.com

Letšatšikgwedi

Professor Kerstin Tönsing
Centre for Alternative and Augmentative Communication
Nomoro ya mogala: 012 420 4729
Aterese ya imeile : Kerstin.tonsing@up.ac.za

Letšatšikgwedi

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Fakulteit Geesteswetenskappe
Lefapha la Bomotho

Centre for Augmentative and Alternative Communication

Foromo ya Tumelelo ya Batswadi

Leina la ngwana: _____

Leina la motswadi: _____

Thaetlele ya thuto : Go laetša tlotlontšu ya motheo ya baithuti ba Mphato wa R bao ba bolelago Sepedi go tšwa seleteng sa Sekhukhune nakong ya mediro ya ka mehla ya sekolo

Monyakišiši : Charmaine Moswathupa
Moithuti wa Master's kua Center for AAC
Yunibesithi ya Pretoria
Mogala: 0626059781
Imeile: mmotomoswathupa@gmail.com

Mookamedi : Professor Kerstin Tönsing
Moprofesara kua Centre for AAC
Yunibesithi ya Pretoria
Mogala: 012 420 4729
Imeile : Kerstin.tonsing@up.ac.za

Nna _____ (Leina le sefane)
(Hle swaya lepokisi leo le šomago)

☐ Ke fa tumelelo yaka le ngwana wa ka go tšea karolo thutong yeo e nago le sehlogo se se rego: Go laetša tlotlontšu ya motheo ya baithuti ba Mphato wa R bao ba bolelago Sepedi go tšwa seleteng sa Sekhukhune nakong ya mediro ya ka mehla ya , yeo e tlago swarwa ke Charmaine Moswathupa ka fase ga tlhokomelo ya Moprofesara Kerstin Tönsing. Tumelelo ya ka ke ya boithaopo gomme ke kwešiša gore nka gogela morago go tšea karolo ga ngwana wa ka thutong nako efe goba efe. Ke kwešiša gore datha e tla bolokwa mengwaga ye 15 go CAAC le gore datha ka moka e tla swarwa ka sephira. Ke kwešiša gore ngwana wa ka o tla rekotwa ka modumo bakeng sa merero ya go kgoboketša data. Ke kwešiša gore datha e ka šomišwa go sehlogo sa mahlale, dipego tša nyakišišo goba ditlhagišo. Ke kwešiša gore foromo ye e ngwadilwego ya seo ngwana wa ka a se boletšego (ka tshedimošo ka moka ya motho ka noši e tlošitšwe) e tla hwetšagala mo wepsaeteng ya Senthara ya Methopo ya Polelo ya Ditšitale ya Afrika Borwa (https://repo.sadilar.org/), e le gore e šomišwe gape bakeng sa tshekatsheko ye nngwe. Ke kwešiša gore tshedimošo ka moka ya motho ka noši yeo e šomišitšwego le yeo e hweditšwego nyakišišong ye e tla swarwa bjalo ka sephiri.

Goba

☐ Ga ke nee tumelelo ya gore nna le ngwana wa ka re tšee karolo thutong yeo e nago le sehlogo se se rego: Go laetša tlotlontšu ya motheo ya baithuti ba Mphato wa R bao ba bolelago Sepedi go tšwa seleteng sa Sekhukhune nakong ya mediro ya ka mehla ya sekolo

Go oketša moo, ke

☐ *fa tumelelo ya gore direkhoto tša polelo ya ngwana wa ka di hwetšagale go banyakišiši ba bangwe.Ke kwešiša gore direkhoto di tla bolokwa e le sephiri ke banyakišiši ba gomme di šomišetšwa fela merero ya nyakišišo.*
Goba

☐ *GA KE nee tumelelo ya gore direkhoto tša polelo ya ngwana wa ka di hwetšagale go banyakišiši ba bangwe.*

_____ _____
Mosaeno wa Mohlokomedi Letšatšikgwedi

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

Appendix G
Caregivers' questionnaire

APPENDIX G1 : CAREGIVER QUESTIONNAIRE (English version)

(Based on Mogatusi, 2023)

Name of child: _____

Date of birth: _____

Gender: _____

Name of caregiver: _____

Relationship with the child: _____

Cell phone numbers: _____

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

Information about the child

1. Does your child speak Sepedi 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?: _____

_____

3. Are you concerned about your child's:

Vision: ☐ Yes ☐ No If yes, please explain: _____

Hearing: ☐ Yes ☐ No If yes, please explain: _____

Walking: ☐ Yes ☐ No If yes, please explain: _____

Talking: ☐ Yes ☐ No If yes, please explain: _____

Thinking: ☐ Yes ☐ No If yes, please explain: _____

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

☐ Yes

☐ No

If not, please explain your concerns: _____

_____

5. At what age did your child started 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. Does your child have siblings and /or other children living in the same household? If yes, please list them in the table below.

Gender (Male/Female)	Age	Relationship to your child	Language used mostly by this child	Other languages used by this child

7. Who are the adults living with the child at home? Please list them in the table below.

Gender (Male/ Female)	Age	Relationship to your child	Language used mostly by this adult	Other languages used by this adult

8. Which language is mostly used during conversations at home? _____

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

_____

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

less ☐ than R 7979

☐ more than R 7979

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

APPENDIX G2: CAREGIVER QUESTIONNAIRE (Sepedi V ersion)

(Based on Mogatusi, 2023)

Leina la ngwana: _____

Letsatsi la matswalo: _____

Bong bja ngwana: _____

Leina la Motswadi: _____

Kamano le ngwana: _____

Nomoro ya sellathekeng: _____

Taetso: ka kgopelo, hle araba dipotsiso tse dilatelago kago swaya kgetho ye oe kgethilego.

Tshedimoso ka ga ngwana

1. Ana ngwana wa gago o bolela Sepedi bjalo ka polelo ya gae?

   ☐ Ee

   ☐ Aowa

2. Ana ngwana wa gago o bolela maleme a mangwe?

   ☐ Ee

   ☐ Aowa

Ge e ba Karabo e le ee, ke Maleme afeng ao ngwana wa gago a a bolelang ? _____

_____

3. Na o tshwenyegile ka tseo dilatelago mo go ngwana wa gago:

Mahlo goba pono: ☐ Ee ☐ Aowa

Ge Karabo e le Ee, Hlalosa: _____

Go kwa ☐ Ee ☐ Aowa

Ge Karabo e le Ee, Hlalosa: _____

Go sepela ☐ Ee ☐ Aowa

Ge Karabo e le Ee, Hlalosa: _____

Go Bolela ☐ Ee ☐ Aowa

Ge Karabo e le Ee, Hlalosa: _____

Go nagana ☐ Ee ☐ Aowa

Ge Karabo e le Ee, Hlalosa: _____

4. O nagana gore ngwana wa gago o gola ka mo go tlwaelegilego go ya ka mengwaga ya gagwe?

☐ Ee

☐ Aowa

Ge Karabo e le Aowa, Hlalosa di pelaelo tsa gago:

_____

_____

5. Ngwana wa gago o thomile neng go bolela lentsu la mathomo (e.g. mama, papa, dijo)? Hle swaya kgetho e tee.

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

6. Na ngwana wa gago o na le bana babo / goba bana ba bangwe bao ba dulago le bona ka lapeng. Hle ba ngwale mo tafoleng ye e lego ka mo tlase.

Bong (Mosemane/mosetsana)	Mengwaga	Kamano le ngwana wa gago	Polelo yeo e somisiwago kudu ke ngwana	Maleme a mangwe ao ngwana a a sumisago

7. Batho ba bagolo bao ba dulago le ngwana wa gago ka gae. Hle ba ngwale mo tafoleng ye e lego ka mo tlase.

Bon (Monna/ Mosadi)	Mengwaga	Kamano le ngwana wa gago	Polele yeo e sumisiwago kudu	Maleme a mangwe ao motho yo a a sumisago

8. Ke polele efe yeo e somisiwago kudu dipoledisanong tsa ka gae? _____

9. Ke maleme a feng amangwe ao a somisiwago dipoledisanong tsa ka gae? Hlalosa.

_____

10. Na ngwana wa gago o thabela go bogela TV goba go theeletsa seyalemoya?
Ee ☐ Aowa ☐

Ge Karabo e le Ee, ke maleme afe ao ngwana wa gago a a theeletsang mo TV gobe seyalemoya?

_____

11. Na ngwana wa gago o thabela go bogela dibidio goba go theeletsa mmino ka sellathekeng?
Ee ☐ Aowa ☐

Ge Karabo e le Ee, ke maleme afe ao ngwana wa gago a a sumisago mo sellathekeng?

_____

Tshedimoso ka ga dinolofatsi tikologong ya ka gae

12. Na o na le phihlelelo ya mohlagase ka ngwakong?

☐ Ee

☐ Aowa

13. Na o nale phihlelelo ya meets ka ngwakong?

- ☐ Ee
- ☐ Aowa

14. Hle laetsa maseleng ao lapa la gago le kgonago go a boloka goba go a dirisa kgwedi ka kgwedi .

Ka ☐ tlase ga R 7979

Go ☐ feta R 7979

Ke leboga kudu nako ya lena le maiteko a lena a go nthusa thutong le dipotsisong tsaka!

Appendix H
Participants' assent script

APPENDIX H

Child assent script (English version)

	Hi, my name is Charmaine, I am a student at university. I would like to find out what kind of words children like you use daily when speaking to their friends and teacher. Will you be interested in helping me with that? If you say yes, we will do the following:
	I will ask you to carry this bag on your waist. This bag has a voice recorder. It will record every word you say to your teacher and friend throughout the day. I will also clip this microphone on the collar of your shirt. (The researcher will demonstrate the placement of the equipment to the child)
	Only I and another person helping me can listen to your recording. I will not give it to anyone else.
	Please do not play with the equipment. If the microphone or the pouch on your waist is uncomfortable, tell your teacher, and she/he will help you or even take it off.

	Please tell your teacher if you want to take off the bag or the microphone. She/he will take it off, and nobody will be angry at you (not your teacher, me, caregiver, or principal).
	You can choose to carry this bag on your waist and the microphone or not. If you say no, nothing bad will happen to you.

APPENDIX H2

Sengwalwa sa tumelelo ya ngwana (Sepedi version)

	Thobela, leina laka ke Charmaine, ke moithuti wa yunibesithing. Ke rata go hwetša gore ke mantšu a mohuta mang ao bana ba go swana le wena ba a šomišago letšatši le letšatši ge ba bolela le bagwera ba bona le morutiši. Na o tla ba le kgahlego ya go nthuša ka seo? Ge o ka re ee, re tla dira tše di latelago:
	Ke tla go kgopela gore o rwale mokotla wo lethekeng. Mokotla wo o na le sedirišwa sa go rekota mantšu. E tla rekota lentšu le lengwe le le lengwe leo o le bolelago go morutiši le bagwera wa gago letšatši ka moka. Ke tla kgaola le maekrofouno yo mo molala wa hempe ya gago. (Monyakišiši o tla bontšha go bewa ga didirišwa go ngwana)
	Ke nna le motho yo mongwe yo a nthušago feela bao ba ka theetšago kgatišo ya gago. Nka se e nee motho yo mongwe.
	Hle o se ke wa bapala ka didirišwa. Ge maekrofouno goba mokotla wa lethekeng la gago o dira gore oseke wa phuthologa, botša morutiši wa gago, gomme o tla go thuša goba gaešita le go o apola.

	Hle botša morutiši wa gago ge e ba o nyaka go apola mokotla goba maekrofouno. O tla e apola, gomme ga go na motho yo a tlago go go galefela (e sego morutiši wa gago, nna, mohlokomedi, goba hlogo ya sekolo).
	O ka kgetha go rwala mokotla wo lethekeng la gago le maekrofouno goba go se rwale. Ge o ka re aowa, ga go selo se sebe seo se tlago go go diragalela.

Appendix I
Participants' response form

APPENDIX I : Child - friendly response form

Name and Surname: _____

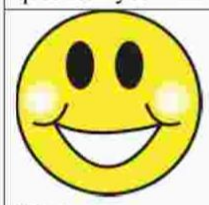
Date of birth: _____

Date: _____

Name of the study: **Determining the core vocabulary of Sepedi- speaking Grade R learners from the Sekhukhune district during regular school activities.**

Researcher: Charmaine Moswathupa

 Did you understand everything I explained to you? (What will happen during the week)	
 Do you understand that you can choose to be a part of this study or not	
 Do you understand that you may choose to stop participating in this study at any time?	

 Do you understand that you I will record everything you say to your friends and teachers throughout the day?	
  Please think about it. Do you have any questions you want to ask me?	
 Did I answer your question? Are you happy with my answer?	
Do you want to take part in this study?	

APPENDIX I2: Child - friendly response form (Sepedi version)

Leina le Sefane: _____

Matswalo: _____

Letsatsikgwedi: _____

Leina la nyakisisi: **Go laetša tlotlontšu ya motheo ya baithuti ba Mphato wa R bao ba bolelago Sepedi go tšwa seleteng sa Sekhukhune nakong ya mediro ya ka mehla ya sekolo**

Monyakisisi: Charmaine Moswathupa

 Na o kwešišitše tšohle tšeo ke go hlalositšego tšona? (Tseo ditlo diragalago mo bekeng)	
 Na o a kwešiša gore o ka kgetha go ba karolo ya thuto ye goba aowa?	
 Na o a kwešiša gore o ka kgetha go kgaotša go tšea karolo thutong ye nako le ge e le efe?	

 Na o kwešiša gore ke tlo rekota tšohle tšeo o di bolelago go bagwera ba gago le barutiši letšatši ka moka?	
  Ke kgopela o nagana ka yona. Naa o na le dipotšišo tšeo o nyakago go mpotšiša tšona?	
 Na ke arabile potšišo ya gago? O thabile ka karabo yaka?	
Na o nyaka go tšea karolo thutong ye?	

Appendix J
Transcription rules

APPENDIX: J: Transcription rules

These rules are adopted based on Trembath et al. 2007, Mothapo et al. (2021), and Mogatusi (2022).

Transcription rules	Example/explanation
1. Deleting the first 20 minutes of each sample to ensure accurate spontaneity of each sample and reduce novelty effects.	This will ensure that novelty effects will not skew the results.
2. Any comment that participants may make about the equipment or the research process will not be transcribed.	This will ensure that participant reactivity will not unduly influence results.
3. The researcher will transcribe all recordings into Word documents. A research assistant with knowledge of the Sepedi language will check for errors in the transcripts to ensure the reliability of the data.	
4. One document will be used to transcribe the recordings for each participant.	This will facilitate the calculation of a commonality score. However, for the final analyses, all transcripts will be merged into one composite transcript.
5. The de-identification of each transcript will be done by abbreviating the participant's first and last name and attaching the numerical number according to the order of the samples. This is to ensure confidentiality.	For example, the transcript of Moswathupa Charmaine who was the 1st participant to be recorded, will be named 'MC1'
6. The transcription of each recording will stop when 3 000 intelligible and orthographic words are obtained.	The total word count will be indicated on the bottom section of the Word document.
7. Additional speech, including environmental sounds and sounds made by other speakers, will not be included.	The study aims to collect all the words uttered only by target participants to arrive at valid word frequency counts.

Transcription rules	Example/explanation
8. Utterance will be transcribed individually. Every sentence will be transcribed on a new line.	Any word production followed by a pause of greater than 2 seconds or intonation will be transcribed as an utterance.
9. Every statement will end with a punctuation mark. Full stop indicates a statement, a question mark indicates questions, and * indicates interrupted utterances.	"Etla mo**.**" "ke mang yo a Mpitšago**?**' "ke ya*."
10. Numbers uttered will be written in word form.	5 will be written as 'five'
11. Codes in capital letters will be used to represent all people's names and other proper nouns such as the name of a town or city. The following codes will be used: For children's names, CN, for teacher's name, TN, and PN for the name of the place.	"Ke Puku ya CN."
12. Vulgar utterances will be transcribed and counted in the analysis.	The aim is not to censor the participants' speech samples.
13. Theme-specific utterances such as songs and repetitive games will be transcribed as one word.	Singing of 'Happy birthday', prayer songs, and repetitive games played during break time. For example: happybirthdatoyouhappybirthdaytoyou-happybithdaydearCNhappybirthdaytoyou. This will ensure that rote recitals will not skew the word count.
14. Sound, syllable, and word repetition and prolongation of words will be transcribed as individual words. The repeated part will be ignored.	"sh sh shut" will be transcribed as "shut".
15. Words that are code-switched will be transcribed using the orthography of the language of origin. This will be indicated by "CS" on the transcribed document. However, words that are loaned from other languages will be transcribed using the Sepedi spelling rules.	"Toilet" will be transcribed in English as follows: toiletCS "lefesetere" which is loaned from Afrikaans, will be transcribed according to the Sepedi rules.

Transcription rules	Example/explanation
16. Sepedi orthographic rules will be applied in accordance with the Oxford *pukuntšu ya sekolo* dictionary. The researcher and research assistant will apply their knowledge and experience when agreeing on the spelling of words that might not have orthographic representation.	Consistent spelling is important to ensure that word frequency counts are accurate.
17. Interjections and word fillers will be transcribed as one-word.	These words are used to express feelings and reactions. These words will be transcribed in a consistent form phonetically to ensure consistency. Example: “EE”, “Hee”, “HE-Eh”, “AOWA”

Appendix K
Tagging rules

APPENDIX: K

Tagging rules. Adapted from Mothapo et al. (2021)

The tagging rules were created to accommodate the Sepedi Language within the Microsoft word program.

1. These rules were applied to enable the easy identification of different parts of speech and that the words that share the same morphological variation are counted under one root form.
2. The base form of the words will be identified first and followed by the different variations.
3. For code-switching, the code "@cs" is used to identify the use of another language within the sample.
4. These rules ensure that the heteronymous words with different meanings are not counted under the same vocabulary unit.

Tagging rules: **Inflected forms verbs and nouns**

Part of speech variation	**Code**
Negative form 1	**@a**
Negative form 2	**@b**
Object concord relating to the 1st person	**@c**
Negative form relating to the 1st person	**@d**
Applied verbal extension	**@e**
Object concord	**@f**
Plural	**@g**
Locative	**@h**

Part of Speech	Lemma (base form)	Lemma Example	Grammatical Variations	Examples	The Code Used in the Sample	Examples of Sentences and Coded Sentences
Verbs	Indicative mood	Bona	Negative form 1	Bone	Bona@a/bone	Mahlo a gagwe ga a bone. **Mahlo a gagwe ga a bona@a/bone.**
			Negative form 2	Bonale	Bona@b/bonale	Mahlo a gagwe ga a bonale. **Mahlo a gagwe ga a bona@b/bonale.**
			Object concord relating to the first person.	Mpona	Bona@c/Mpona	O ile ge a mpona a tšaba. **O ile ge a bona@c/mpona a tshaba@s.**
			Negative form relating to the first person.	Mpone	Bona@d/Mpon e	Obe a sa mpone. **O be a sa bona@d/mpone.**
			Applied verbal extension.	Bonela/bon etše/bonang	Bona@e/Mpone la	O nyaka go mponela. **O nyaka go bona@e/mponela.**
	Imperative mood	Dula	Applied verbal extension.	Dudiša	Dula@e/dudiša	Ke tla go dudiša. **Ke tlago dula@e/dudusa@s.**

	Interrogative mood	Etla	Negative form Object concord	Ga a tle Ntlela.	Etla@a/etle Etla@f/ntlela	Ga a tle? **cn etla@a/etle?** O tla ntlela? **O tla etla@f/ntlela?**
Nouns	Singular form (Leina)	Mohlare	1. Plurals	Mehlare	Mohlare@g/me hlare	Bana ba bapala ka mehlare. **Bana ba bapala ka mohlare@g/mehlare.**
			2. Locative	Mohlareng	Mohlare@g/mo hlareng	Go dutše banna kua mohlareng. **Go dutse@s banna kua mohlare@g/mohlareng.**
		Ngaka	1. Plural	Dingaka	Ngaka@g/dinga ka	Ke bone dingaka ka labobedi. **Ke bone ngaka@g/dingaka ka labobedi.**
			2. Locative	Ngakeng	Ngaka@h/ngak eng	O ile ngakeng lehono. **O ile ngaka@h/ngakeng lehono.**

Additional tagging of heteronyms and polysemous words was done with specific codes to avoid over-counting of words or under-representation of other words. These words are spelled the same but differ in their lexical meaning. Some of them are pronounced differently. These rules will be used with the knowledge of Sepedi language to clearly indicate how the words should be transcribed on the samples.

Tagging rules: **Heteronyms**

Part of speech variation	**Key Code**
Noun	**@i**
Verb and axillary verb	**@j**
Demonstrative particle	**@k**
Infinite prefix	**@1**
Object concord	**@m**
negative morpheme	**@n**
copulative verb	**@o**
Connective particle	**@p**
Locative particle	**@q**
Object concord 2nd person singular	**@r**
past tense morpheme	**@s**
Infinitive Prefix	**@t**
Future morpheme	**@u**
Possessive concord	**@w**
Subject concord	**@x**
Agentive particle	**@y**
Possessive pronoun	**@z**
Present tense morpheme	**@xx**
Instrumental particle	**@qq**
Copulative particle	**@oo**
Temporal particle	**@pp**
Hortative particle	**@yy**
Aspectual prefix	**@tt**

Word	Part of Speech variation	Coding sample	Examples in
Thaba	1. Noun	Thaba@i	Ke thaba@i ya gesu@s ela.
	2. Verb	Thaba@j	Ke a thaba@j ge kele bona.
Noka	1. Noun	Noka@i	Noka@i e tletse@s meetse.
	2. Noun	Noka@i	Noka@i ya ka e bohloko.
	3. Verb	Noka@j	Malome o noka@j nama ka letswai.
Le	1. Plural	Le@g	Le wena ga oye kerekeng?
	2. Connective particle	Le@p	O boile le bo mang?
Mo	1. Locative particle	Mo@q	O e bee mo@q godimo ga koloi.
	2. Concord	Mo@m	Ke mosadi o mo@m telele.
	3. Demonstrative particle	Mo@k	Ke mmone malabo a tlile mo@k.
Go	1. Infinite prefix	go@l	Ke dieta tsa go@l ya sekolong.
	2. Object concord 2nd person singular	go@r	Ke go@r bone maloba.
A	1. Concord	a@m	Gase a@m mpona.
	2. demonstrative particle	a@k	Meetse a@k ke a mabose.
	3. past tense morpheme	a@s	Cn ga a@s sa mpolediša.

Word	**Part of Speech variation**	**Coding sample**	**Examples in**
Ga	1. negative morpheme	ga@n	Bana ba ga@n ba loka.
	2. Locative particle	ga@q	Nna ke ile ga@q bo Tumi.
Ba	1. concord	ba@m	Bana ba@m ba bolela kudu.
	2. copulative verb	ba@o	Bare ba@o etla.
Sa	1. concord	sa@m	Tliša senepe sa@m Lebo.
	2. negative morpheme	sa@n	Ga a sa@n mpolediša.
	3. prefix	sa@t	Ke sa@t ile go bolela le yena.
Tla	1. Verb	tla@j	E tla@j le yona.
	2. Future morpheme	tla@u	Barile e tla@u tla@j lehono.
Ya	1. Verb	ya@j	Ke ya@j go botsa@s.
	2. Possessive concord	ya@w	Ke koloi ya@w papa.

Appendix L
Sepedi core vocabulary list

APPENDIX L: Sepedi Core Vocabulary List

Core word list with frequency of occurrence, commonality score, and part of speech classification

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
ke@x (i)	985	50.9465191	6	Subject concord	function word
cn (child's name)	901	46.6018413	6	Proper noun	content word
o (you)	777	40.1882694	6	Subject concord	function word
a@x (she, he, they)	442	22.8612806	6	Subject concord	function word
e (it)	440	22.7578359	6	Subject concord	function word
nna (i)	432	22.3440571	6	Absolute pronoun	content word
mo@k (here)	337	17.4304334	6	Demonstrative particle/pos	function word
ke@oo (is, are)	332	17.1718217	6	Copulative particle	function word
ba@x (they)	300	15.5167063	6	Subject concord	function word
wa@x (you)	294	15.2063722	6	Subject concord	function word
bona (see)	291	15.0512051	6	Verb	content word
se (she, he, it)	288	14.8960381	6	Subject concord	function word
ee (yes)	285	14.740871	6	Interjection	function word
tlo (will, shall)	285	14.740871	6	Future morpheme	function word
ka@q (inside, into, in)	252	13.0340333	6	Locative particle	function word
le@p (and, with)	237	12.258198	6	Connective particle	function word
tla@j (come)	229	11.8444192	6	Verb	content word
re@x (we)	214	11.0685838	6	Subject concord	function word
a@n	197	10.1893038	6	Negative morpheme	function word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
re@j (say)	188	9.72380263	6	Verb	content word
ye (this)	181	9.36174615	6	Demonstrative particle/pos	function word
ngwala (write)	165	8.53418848	6	Verb	content word
wena (you)	155	8.01696493	6	Absolute pronoun	function word
ya@w (of)	150	7.75835316	6	Possessive concord	function word
go@m (you)	149	7.70663081	6	Object concord	function word
go@t (to)	148	7.65490845	5	Infinitive prefix	function word
ka@z (my, mine)	140	7.24112962	6	Possessive pronoun	function word
di (they)	139	7.18940726	6	Object concord	function word
go@x (it)	138	7.13768491	6	Subject concord	function word
nyaka (want, search)	128	6.62046136	6	Verb	content word
ka@qq (with, by means of)	126	6.51701665	6	Instrumental particle	function word
ka@x (i)	118	6.10323782	6	Subject concord	function word
akere (isn't it)	111	5.74118134	6	Interjection	function word
le@x (you, she, he, it)	111	5.74118134	6	Subject concord	function word
mo@m (her, him)	110	5.68945898	6	Object concord	function word
gore (so)	108	5.58601428	6	Conjunction	function word
tše (these ones)	108	5.58601428	6	Demonstrative particle/pos	function word
ria (do, make)	102	5.27568015	6	Verb	content word
fa (give)	101	5.22395779	6	Verb	content word
ko (i am going to)	99	5.12051309	6	Future morpheme	function word
re@m (us)	91	4.70673425	6	Object concord	function word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
swara (hold)	87	4.49984483	6	Verb	content word
gago (yours)	85	4.39640012	6	Possessive pronoun	function word
nto (a thing)	84	4.34467777	6	Noun	content word
mang (who)	83	4.29295542	6	Noun	content word
botša (tell)	82	4.24123306	6	Verb	content word
fela (finish)	81	4.18951071	6	Verb	content word
kae (where)	80	4.13778835	6	Adverb	content word
kua (over there)	80	4.13778835	6	Locative particle	function word
bea (put, place)	79	4.086066	6	Verb	content word
so (like this)	79	4.086066	5	Preposition	function word
tšea (take)	79	4.086066	6	Verb	content word
ja (eat)	74	3.82745423	6	Verb	content word
tša (of)	73	3.77573187	5	Possessive concord	function word
wa@w (of)	73	3.77573187	6	Possessive concord	function word
motho (person)	72	3.72400952	6	Noun	content word
aowa (no)	71	3.67228716	4	Interjection	function word
ahh	68	3.5171201	6	Interjection	function word
tseba (know)	68	3.5171201	6	Verb	content word
ya@x (she, he, it, they)	68	3.5171201	6	Subject concord	function word
še (here it is)	67	3.46539774	6	Demonstrative copulative pos	function word
tloga (leave)	67	3.46539774	4	Verb	function word
gape (again)	66	3.41367539	6	Adverb	content word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
haeh	66	3.41367539	5	Interjection	function word
rubba (erase)	66	3.41367539	6	Verb	content word
tšhela (pour)	66	3.41367539	5	Verb	content word
a@xx	65	3.36195304	6	Present tense morpheme	function word
bo (it)	65	3.36195304	6	Subject concord	function word
ema (stop)	65	3.36195304	6	Verb	content word
la (you)	64	3.31023068	6	Subject concord	function word
tlalea (report)	64	3.31023068	5	Verb	content word
mmemo (teacher)	63	3.25850833	6	Noun	content word
ge (when, while)	59	3.05161891	6	Conjunction	function word
tla@u (shall, will)	59	3.05161891	6	Future morpheme	function word
no (only, just)	56	2.89645185	6	Aspectual prefix	function word
we	56	2.89645185	6	Interjection	function word
be	55	2.84472949	6	Auxiliary Verb	function word
eng (what)	54	2.79300714	6	Noun	content word
kwa (hear, feel, taste, smell)	54	2.79300714	6	Verb	content word
tn (teacher's name)	54	2.79300714	6	Proper noun	content word
le@m (you, it)	50	2.58611772	6	Object concord	function word
dira (do, make)	48	2.48267301	5	Verb	content word
ae	47	2.43095066	6	Interjection	function word
gešu	47	2.43095066	6	Possessive pronoun	function word
gona (there)	47	2.43095066	6	Absolute pronoun	function word
bedi (two)	46	2.3792283	6	Adjective	content word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
ga@w (of)	46	2.3792283	5	Possessive concord	function word
mmm	45	2.32750595	5	Interjection	function word
ne (have)	45	2.32750595	6	Copulative verb	function word
ntša (take out)	43	2.22406124	5	Verb	content word
dula (sit)	42	2.17233888	6	Verb	content word
ngwe	42	2.17233888	6	Adjective	content word
itia (hit)	40	2.06889418	5	Verb	content word
mama (mother)	40	2.06889418	4	Noun	content word
meetse (water)	40	2.06889418	5	Noun	content word
šo (here she/he is)	40	2.06889418	4	Demonstrative copulative pos	function word
ngwana (child)	39	2.01717182	5	Noun	content word
pela (haste)	38	1.96544947	4	Noun	content word
sepela (walk)	38	1.96544947	5	Verb	content word
tee (one)	38	1.96544947	6	Adjective	content word
tjo	38	1.96544947	6	Interjection	function word
rena (we)	37	1.91372711	6	Absolute pronoun	function word
kgopela (ask)	35	1.8102824	6	Verb	content word
ya@j (go)	35	1.8102824	6	Verb	content word
owo (okay)	34	1.75856005	6	Adjective	content word
ra (we)	34	1.75856005	5	Subject concord	function word
hmm	33	1.7068377	5	Interjection	function word
mara (but)	33	1.7068377	6	Conjunction	function word
nke (as if)	33	1.7068377	6	Conjunction	function word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
puku (book)	33	1.7068377	6	Noun	content word
fase (down)	32	1.65511534	6	Noun	content word
yey	32	1.65511534	5	Interjection	function word
adima (borrow)	31	1.60339299	5	Verb	content word
gae (home)	31	1.60339299	5	Noun	content word
le@k (this)	31	1.60339299	6	Demonstrative particle/pos	function word
maan	31	1.60339299	5	Interjection	function word
nnyane (small, few)	31	1.60339299	6	Adjective	content word
tsena (enter, get in)	31	1.60339299	6	Verb	content word
yo (go and, go to)	31	1.60339299	5	Aspectual prefix	function word
ga@q (at the place of)	30	1.55167063	6	Locative particle	function word
mo@q (on)	30	1.55167063	6	Locative particle	function word
one@cs	30	1.55167063	3	Adjective	content word
sa@w (of)	30	1.55167063	5	Possessive concord	function word
sa@x (she, he, it)	30	1.55167063	6	Subject concord	function word
ebile (then)	29	1.49994828	6	Conjunction	function word
fora (decieve)	29	1.49994828	6	Verb	content word
namela (climb)	29	1.49994828	5	Verb	content word
bapala (play)	28	1.44822592	5	Verb	content word
kgona (can)	28	1.44822592	6	Verb	content word
le@o (is, are, am)	28	1.44822592	5	Copulative verb	function word
why@cs	28	1.44822592	4	Conjunction	function word
yena (her/him)	28	1.44822592	6	Absolute pronoun	function word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
gabotse (well)	27	1.39650357	5	Adverb	content word
khalara (colour)	27	1.39650357	4	Noun	content word
nnang (refuse)	27	1.39650357	6	Interjection	function word
red@cs	27	1.39650357	6	Adjective	content word
topa (pick up)	27	1.39650357	5	Verb	content word
tšwa (come out/from)	27	1.39650357	6	Verb	content word
ile (went)	26	1.34478121	6	Verb	content word
lena (yours)	26	1.34478121	5	Possessive pronoun	function word
tshela (jump over)	26	1.34478121	5	Verb	content word
yah	26	1.34478121	5	Interjection	function word
ga@n	25	1.29305886	5	Negative morpheme	function word
kgale (long ago)	25	1.29305886	6	Noun	content word
sa@tt (still)	25	1.29305886	5	Aspectual prefix	function word
school-peke (schoolbag)	25	1.29305886	6	Noun	content word
sekolo (school)	25	1.29305886	6	Noun	content word
te	25	1.29305886	6	Reflexive pronoun	function word
apara (wear)	24	1.24133651	5	Verb	content word
go@q (at, from, to)	24	1.24133651	6	Locative particle	function word
ke@j	24	1.24133651	6	Auxiliary verb	function word
lebelela (look for/at)	24	1.24133651	4	Verb	content word
monna (man, male person)	24	1.24133651	3	Noun	content word
phaka	24	1.24133651	4	Verb	content word
sa@n	24	1.24133651	5	Negative morpheme	function word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
baitše (they said)	23	1.18961415	4	Verb	content word
ganijwale	23	1.18961415	4	Adverb	content word
ka@pp (at, on)	23	1.18961415	5	Temporal particle	function word
nka (i can/may)	23	1.18961415	6	Modal verb	function word
blue@cs	22	1.1378918	4	Adjective	content word
bula (open)	22	1.1378918	6	Verb	content word
duma (wish)	22	1.1378918	6	Verb	content word
gana (refuse)	22	1.1378918	6	Verb	content word
ka@j (never)	22	1.1378918	4	Auxiliary verb	function word
napile (then)	22	1.1378918	3	Auxiliary Verb	function word
two@cs	22	1.1378918	5	Adjective	content word
aga	21	1.08616944	3	Interjection	function word
iša (take to)	21	1.08616944	6	Verb	content word
jwang (how)	21	1.08616944	6	Adverb	content word
ma	21	1.08616944	5	Interjection	function word
pencil@cs	21	1.08616944	6	Noun	content word
swana (the same)	21	1.08616944	6	Verb	content word
tafola (table)	21	1.08616944	6	Noun	content word
bolela (talk/speak)	20	1.03444709	6	Verb	content word
gafa (mad)	20	1.03444709	4	Verb	content word
kgwatha (touch)	20	1.03444709	6	Verb	content word
oe (that one)	20	1.03444709	5	Demonstrative particle/pos	function word
sesi (sister)	20	1.03444709	6	Noun	content word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
simba (chips)	20	1.03444709	5	Noun	content word
tšhaba (run away, escape)	20	1.03444709	5	Verb	content word
yellow@cs	20	1.03444709	5	Adjective	content word
yona (it)	20	1.03444709	6	Absolute pronoun	function word
ba@k (these ones)	19	0.98272473	6	Demonstrative particle/pos	function word
bitša (call)	19	0.98272473	6	Verb	content word
ka@u (can, could)	19	0.98272473	5	Potential morpheme	function word
kgetha (choose)	19	0.98272473	3	Verb	content word
reng (say what)	19	0.98272473	5	Verb	content word
tsamo (go and)	19	0.98272473	5	Auxiliary verb	function word
ba@j (furthermore, and)	18	0.93100238	4	Auxiliary verb	function word
bowa (come)	18	0.93100238	5	Verb	content word
letsogo (arm)	18	0.93100238	5	Noun	content word
reka (buy)	18	0.93100238	5	Verb	content word
šia (here they are)	18	0.93100238	4	Verb	content word
tjebanna	18	0.93100238	6	Interjection	function word
tšhelete (money)	17	0.87928002	5	Noun	content word
ba@m (them)	16	0.82755767	6	Object concord	function word
lahla (throw away)	16	0.82755767	5	Verb	content word
moka (all, every)	16	0.82755767	5	Noun	content word
rata (love, like)	16	0.82755767	5	Verb	content word
šala (stay, remain)	16	0.82755767	4	Verb	content word
sega (laugh)	16	0.82755767	6	Verb	content word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
sekotlelo (plate)	16	0.82755767	5	Noun	content word
senya (destroy)	16	0.82755767	6	Verb	content word
apola (undress)	15	0.77583532	6	Noun	content word
dilo (things)	15	0.77583532	6	Noun	content word
neh	15	0.77583532	5	Interjection	function word
ngwatha (break off a piece)	15	0.77583532	4	Verb	content word
nwa (drink)	15	0.77583532	4	Verb	content word
tala (green/blue)	15	0.77583532	5	Adjective	content word
gabo (of their family)	14	0.72411296	4	Communal possessive pronoun	function word
ke@y (by, with)	14	0.72411296	5	Agentive particle	function word
lehono (today)	14	0.72411296	4	Noun	content word
maaka (lies)	14	0.72411296	6	Noun	content word
ntšhi (many, a lot)	14	0.72411296	4	Adjective	content word
torowa (draw)	14	0.72411296	5	Verb	content word
wo (this one)	14	0.72411296	4	Demonstrative particle/pos	function word
abuti (brother)	13	0.67239061	3	Noun	content word
bjale (now)	13	0.67239061	5	Adverb	content word
faka (put in)	13	0.67239061	3	Verb	content word
koko (grandmother)	13	0.67239061	5	Noun	content word
maabane (yesterday)	13	0.67239061	4	Noun	content word
phaphuši (classroom)	13	0.67239061	4	Noun	content word
rela (say)	13	0.67239061	5	Verb	content word
šapo (okay)	13	0.67239061	5	Adjective	content word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
seeta (shoe)	13	0.67239061	4	Noun	content word
tharo (three)	13	0.67239061	5	Adjective	content word
wa@j (fall)	13	0.67239061	4	Verb	content word
ai	12	0.62066825	5	Interjection	function word
chaya	12	0.62066825	4	Verb	content word
ga@yy	12	0.62066825	4	Hortative particle	function word
grade@cs	12	0.62066825	3	Noun	content word
hlatswa (wash)	12	0.62066825	4	Verb	content word
righte (right)	12	0.62066825	4	Noun	content word
baba (spicy, pain)	11	0.5689459	4	Verb	content word
ene (and)	11	0.5689459	4	Conjunction	function word
fihla (arrive)	11	0.5689459	6	Verb	content word
goba (or)	11	0.5689459	6	Conjunction	function word
hwetša (find)	11	0.5689459	6	Verb	content word
ntle (out, outside)	11	0.5689459	4	Noun	content word
number@cs	11	0.5689459	5	Noun	content word
robala (sleep)	11	0.5689459	5	Verb	content word
seng (which/who is not)	11	0.5689459	5	Copulative verb	function word
thoma (start, begin)	11	0.5689459	5	Verb	content word
botse (beauty)	10	0.51722354	4	Noun	content word
ditšhila (dirt)	10	0.51722354	4	Noun	content word
eban	10	0.51722354	4	Interjection	function word
haah	10	0.51722354	3	Interjection	function word

root word	Number of occurrences	Frequency per mille	Commonality	Part of speech	Content/function words
hlogo (head)	10	0.51722354	4	Noun	content word
mfanaka (my boy/friend)	10	0.51722354	4	Proper noun	content word
pn (name of place)	10	0.51722354	4	Proper noun	content word
šupa (point)	10	0.51722354	5	Verb	content word
three@cs	10	0.51722354	5	Adjective	content word
tllaseng (classroom)	10	0.51722354	3	Noun	content word

Appendix M
Core vocabulary comparison with Mothapo (2019) findings

APPENDIX M

Core words overlapping between the two core vocabulary lists

Root word	Part of speech	Root word	Part of speech
A@M	Object concord	BOWA	Verb
A@X	Subject concord	BULA	Verb
A@XX	Present tense morpheme	CN	Proper noun
ADIMA	Verb	DI	Object concord
AHH	Interjection	DILO	Noun
AKERE	Interjection	DIRA	Verb
AOWA	Interjection	DITŠHILA	Noun
APARA	Verb	DULA	Verb
BA@J	Auxiliary verb	E	Subject concord
BA@K	Demonstrative particle/pos	EBILE	Conjunction
BA@M	Object concord	EE	Interjection
BA@X	Subject concord	EMA	Verb
BAPALA	Verb	ENG	Noun
BE	Auxiliary verb	FA	Verb
BEA	Verb	FASE	Noun
BEDI	Adjective	FELA	Verb
BITŠA	Verb	FIHLA	Verb
BO	Subject concord	FORA	Verb
BOLELA	Verb	GA@N	Negative morpheme
BONA	Verb	GA@Q	Locative particle
BOTŠA	Verb	GA@W	Possessive concord
GA@YY	Hortative particle	JWANG	Adverb
GABO	Communal possessive pronoun	KA@J	Auxiliary verb
GABOTSE	Adverb	KA@PP	Temporal particle
GAE	Noun	KA@Q	Locative particle
GAFA	Verb	KA@QQ	Instrumental particle
GAGO	Possessive pronoun	KA@U	Potential morpheme
GANA	Verb	KA@X	Subject concord
GAPE	Adverb	KA@Z	Possessive pronoun
GE	Conjunction	KAE	Adverb
GEŠU	Possessive pronoun	KE@J	Auxiliary verb
GO@M	Object concord	KE@OO	Copulative particle

Root word	Part of speech	Root word	Part of speech
GO@Q	Locative particle	KE@X	Subject concord
GO@T	Infinitive prefix	KE@Y	Agentive particle
GO@X	Subject concord	KGALE	Noun
GONA	Possessive concord	KGETHA	Verb
GORE	Conjunction	KGONA	Verb
HAEH	Interjection	KGOPELA	Verb
HLOGO	Noun	KUA	Locative particle
HMM	Interjection	KWA	Verb
ILE	Verb	LA	Subject concord
IŠA	Verb	LE@K	Demonstrative pos
JA	Verb	LE@M	Object concord
LE@O	Copulative verb	NGWE	Adjective
LE@P	Connective particle	NKA	Modal verb
LE@X	Subject concord	NNA	Absolute pronoun
LEBELELA	Verb	NNYANE	Adjective
LENA	Possessive pronoun	NO	Aspectual prefix
LETSOGO	Noun	NTO	Noun
MAAKA	Noun	NTŠHA	Verb
MABANE	Noun	NTŠHI	Adjective
MAMA	Noun	NYAKA	Verb
MANG	Noun	O	Subject concord
MARA	Conjunction	PN	Proper noun
MEETSE	Noun	RA	Subject concord
MO@K	Demonstrative particle/pos	RATA	Verb
MO@M	Object concord	RE@J	Verb
MO@Q	Locative particle	RE@M	Object concord
MOKA	Noun	RE@X	Subject concord
MONNA	Noun	REKA	Verb
MOTHO	Noun	RENA	Absolute pronoun
NAMELA	Verb	ROBALA	Verb
NGWALA	Verb	SA@N	Negative morpheme
NGWANA	Noun	SA@TT	Aspectual prefix
NGWATHA	Verb	SA@W	Possessive concord
SA@X	Subject concord	TŠHABA	Verb
SE	Subject concord/demonstrative pos	TŠHELA	Verb
ŠE	Demonstrative copulative pos	TŠWA	Verb

Root word	Part of speech	Root word	Part of speech
SEETA	Noun	WA@J	Verb
SEGA	Verb	WA@W	Possessive concord
SEKOLO	Noun	WA@X	Subject concord
SEPELA	Verb	WENA	Absolute pronoun
SESI	Noun	WHY@CS	Conjunction
SO	Preposition	WO	Demonstrative particle/pos
SWANA	Verb	YA@J	Verb
SWARA	Verb	YA@W	Possessive concord
THOMA	Verb	YA@X	Subject concord
TLA@J	Verb	YE	Demonstrative particle/pos
TLA@U	Future morpheme	YENA	Absolute pronoun
TLALEA	Verb	YONA	Absolute pronoun
TLO	Future morpheme	TŠEA	Verb
TN	Proper noun	TSEBA	Verb
TŠA	Possessive concord	TSENA	Verb
TŠE	Demonstrative particle/pos		

Unique core words found in the current study

Root word	Part of Speech	Root word	Part of Speech
ABUTI	Noun	KO	Future morpheme
AE	Interjection	KOKO	Noun
AGA	Interjection	LEHONO	Noun
APOLA	Noun	MA	Interjection
BABA	Verb	MAAN	Interjection
BAITŠE	Verb	MMEMO	Noun
BJALE	Adverb	MMM	Interjection
BLUE@CS	Adjective	NAPILE	Auxiliary verb/perfect
BOTSE	Noun	NE	Copulative verb
CHAYA	Verb	NEH	Interjection
DUMA	Verb	NKE	Conjunction
EBAN	Interjection	NTLE	Noun
ENE	Conjunction	NUMBERS@CS	Noun
FAKA	Verb	NWA	Verb
GANIJWALE	Adverb	OE	Demonstrative particle/pos
GOBA	Conjunction	ONE@CS	Adjective
GRADE@CS	Noun	OWO	Adjective
HAAH	Interjection	PENCIL@CS	Noun
HLATSWA	Verb	PHAKA	Verb
HWETŠA	Verb	PHAPHUŠI	Noun
KGWATHA	Verb	PUKU	Noun
KHALARA	Noun	RED@CS	Adjective
RELA	Verb	TLOGA	Auxiliary verb
RENG	Verb	TOPA	Verb
RIGHTE	Noun	TOROWA	Verb
RUBBA	Verb	TSHELA	Verb
ŠALA	Verb	TŠHELETE	Noun
ŠAPO	Adjective	TWO@CS	Adjective
SCHOOL-PEKE	Noun	WE	Interjection
SEKOTLELO	Noun	YAH	Interjection
SENG	Copulative verb	YELLOW@CS	Adjective
SENYA	Verb	YO	Aspectual prefix
ŠIA	Verb	SKHAFTIN	Noun
SIMBA	Noun	ŠUPA	Verb

Root word	Part of Speech
TAFOLA	Noun
TALA	Adjective
TE	Reflexive pronoun
TEE	Adjective
THARO	Adjective
THREE@CS	Adjective
TJEBANNA	Interjection
TLLASENG	Noun

Unique core words found in Mothapo (2019)

Root words	Part of speech	Root word	Part of Speech
APEŠA	Verb	KOLOI	Noun
APEYA	Verb	KOTO	Adjective
AYEYE	Interjection	KUDU	Adverb
BOGOBE	Noun	KUKA	Verb
BOKAKA	Noun	LEINA	Noun
BOLO	Noun	LELEKERE	Noun
BONTŠHA	Verb	LEOTO	Noun
EH	Interjection	LERAGO	Noun
FETA	Verb	LLA	Verb
FIŠA	Verb	MAMAGO	Noun
GAGWE	Pronoun	MAMAKA	Noun
GARE	Noun	MMATA	Noun
GATA	Verb	MOLA	Pronoun
GENO	Pronoun	MORAGO	Noun
GOBATŠA	Verb	MOSADI	Noun
GOLO	Adjective	MOŠIMANE	Noun
GOPOLA	Verb	MPYA	Noun
HEH	Interjection	NA	Verb
HEY	Interjection	NAA/NA	Question particle
HLAPA	Verb	NGWANENYANA	Noun
HLOBOLA	Verb	NKGA	Verb
JERSEY[CS]	Noun	OHO	Interjection
KAKA	Verb	PARTY	Noun
KGANTHE	Conjunction	RAGA	Verb
ROGA	Verb	TSENELELA	Verb
ROTA	Verb	TŠONA	Pronoun
SEETA	Noun	WOLA	Pronoun
SEJO/DI	Noun	YELA	Pronoun
SEKHIPHA	Noun	YOH	Interjection
SELO	Noun	SORI	Interjection
SETULO	Noun	ŠULE	Pronoun
TLOGA	Verb	TLAPA	Verb
TLOGELA	Verb	TLIŠA	Verb
TOILET[CS]	Noun		

Appendix N
Declaration of Language Editing

JANINE ELLIS

LANGUAGE EDITING / TRANSCRIPTION / TYPING

janine.ellis4@gmail.com

Cell: 083-6563660

Client

Mmoto Charmaine Moswathupa (s/no. 23871165)
Mini-dissertation
University of Pretoria

P O Box 28164
Sunridge Park
6008

09 October 2024

DECLARATION

To whom it may concern,

I hereby declare that I language edited and proofread the mini-dissertation authored by **Mmoto Charmaine Moswathupa**, titled: ***Determining the core vocabulary of Sepedi-speaking Grade R learners from the Sekhukhune district during regular school activities***

All aspects of this mini-dissertation were carefully looked at, corrections made, and suggestions given with regards to certain wording, sentence structure, grammar, spelling, and punctuation, however, the academic content was not influenced in any way. The layout and presentation as well as the referencing of this mini-dissertation were edited as per the referencing and technical/style template/guide provided by the client. Final acceptance of all proposed corrections/changes/comments is at the discretion of the author.

Kind regards

Janine Ellis

Janine Ellis

Appendix O
Turnitin report

turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	CM (Charmaine) Moswathupa
Assignment title:	AAK 895 Year 2 Full mini-dissertation
Submission title:	Moswathupa Charmaine mini-dissertation 131024.docx
File name:	Moswathupa_Charmaine_mini-dissertation_131024.docx
File size:	7.54M
Page count:	162
Word count:	27,254
Character count:	156,337
Submission date:	13-Oct-2024 09:01AM (UTC+0200)
Submission ID:	2483514102

Determining the core vocabulary of Sepedi-speaking Grade R learners from the Sekhukhune district during regular school activities