

ARTICLE

Mean Length of Utterance: A study of early language development in four Southern Bantu languages

Heather BROOKES¹, Patricia MAKAURE¹ , Sefela YALALA¹ , Hannah DANVERS¹, Martin MÖSSMER¹, Francesca LITTLE², Mikateko NDHAMBI⁴, Frenette SOUTHWOOD¹, Babalwa LUDIDI³ and The South African CDI Team¹

¹Department of General Linguistics, Stellenbosch University, South Africa

²Department of Statistical Sciences, University of Cape Town, South Africa

³Department of Communication Sciences & Disorders, University of Cape Town, South Africa

⁴Sefako Makgatho Health Sciences University, South Africa

Corresponding author: Heather Brookes; Email: heatherbrookes@sun.ac.za

(Received 24 May 2022; revised 10 November 2023; accepted 25 November 2023)

Abstract

Mean Length of Utterance (MLU) has been widely used to measure children's early language development in a variety of languages. This study investigates the utility of MLU to measure language development in four agglutinative and morphologically complex Southern Bantu languages. Using a variant of MLU, MLU3, based on the three longest sentences children produced, we analysed the utterances of 448 toddlers (16–32 months) collected using the MacArthur-Bates Communicative Development Inventory, a parent-report tool. MLU3, measured in words (MLU3-w) and morphemes (MLU3-m), significantly correlated with age and other indices of language growth (e.g., grammar and vocabulary). MLU3 measures also accounted for significant variance in language development particular morphosyntactic development. Our results suggest that MLU3-m is a more sensitive measure than MLU3-w. We conclude that MLU measured in morphemes provides a useful addition to other indices of language development in these kinds of morphologically complex languages.

Keywords: Mean length of utterance; Southern Bantu languages; child language development

Introduction

One widely used way of measuring early language development is to determine the mean length of children's utterances (MLU) (R. Brown, 1973; Nice, 1925). MLU is determined

¹SA-CDI team: Monicca Bhuda (University of Mpumalanga), Nina Brink (North-West University), Nomfundo Buthelezi (University of KwaZulu-Natal), William Jiyana (University of Mpumalanga), Portia Khumalo (Stellenbosch University), Muzi Matfunjwa (North-West University), Lufuno Miriri (University of Limpopo), Sibusiso Ndlangamandla (University of South Africa), Helena Oosthuizen (Stellenbosch University), Nomsa Skosana (North-West University), Michelle White (Oslo University), and Katie Alcock (Lancaster University).

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

by counting either words (MLU-w) or morphemes (MLU-m) in a sample of spontaneous recorded utterances (typically 100 per child) and dividing the total words or morphemes by the number of utterances (R. Brown, 1973; Oosthuizen & Southwood, 2009). For example, an utterance such as 'I kicked the ball' can be counted as either 4 words or 5 morphemes where 'kicked' consists of two morphemes.

As children's utterance length in natural interactions varies considerably, calculating an MLU that is representative of a child's language development requires a sufficient number of utterances. Smaller sample sizes may not accurately reflect children's MLU (Oosthuizen & Southwood, 2009). Gavin and Giles (1996) found that a minimum of 175 utterances was necessary for high test-retest reliability, whereas R. Brown (1973) recommended that MLU calculations be based on a sample size of 50-100 utterances. However, he also used the longest utterance produced as an additional measure to determine a child's grammatical development.

Following R. Brown (1973), children's longest utterances have been used as an alternative method for measuring grammatical development to avoid lengthy analyses required with large numbers of utterances in spontaneous samples. This way of measuring children's development, commonly referred to as MSL (Mean Sentence Length) or MLU3 (the term we use in this paper)², has been used in parent report tools on child language development such as the MacArthur-Bates Communicative Development Inventory (MB-CDI; see Fenson et al., 1993 for the original American English MB-CDI). The MB-CDI asks parents to report on their children's lexical production and grammatical development and to provide the three longest sentences produced by their children in the past week (Fenson et al., 1994).

Although MLU has been a widely used proxy measure for child language development in both research and clinical settings (Dethorne et al., 2005), questions have been raised regarding the usefulness and use of MLU (see Ezeizabarrena & Garcia Fernandez, 2018; Jackson-Maldonado & Conboy, 2007; Oosthuizen & Southwood, 2009): Is MLU an accurate measure of language development? What exactly does it measure? Is it a general measure of language development or does it measure a specific ability such as lexical or morphosyntactic development? Does MLU measure the same features across typologically different languages? How should MLU be measured? Are words or morphemes equally suitable units for measuring MLU or is one more accurate than the other for all or only some languages?

MLU seems to be a general measure of language development in languages such as English, but in polysynthetic languages like Canadian Inuktitut, it appears to be primarily a measure of morphosyntactic development (Allen & Dench, 2015). In many languages, using words or morphemes to measure MLU appears to be equally valid, even among languages that are morphologically more complex and typologically distant from English, such as Basque (Ezeizabarrena & Garcia Fernandez, 2018). However, in some synthetic languages like Russian and agglutinative languages like Turkish, morpheme counts are more accurate when calculating MLU (Ege, 2010; Tomas & Dorofeeva, 2019). There are, however, many other typologically distinct and morphologically complex languages, such as Bantu languages, where these questions remain untested.

²The term MLU3 is used to distinguish it from MLU that is based on a large sample of spontaneous utterances (usually between 50 to 175) and MLU2, a term introduced by Johnston (2001), that refers to a spontaneous sample that excludes elliptical responses to questions, imitative utterances and single word yes/no responses (Oosthuizen & Southwood, 2009).

This paper addresses how best to measure language development in Bantu languages which are morphologically complex, agglutinative and typologically distinct (having noun class systems) from languages in previous studies of MLU. In this study, we focus on typically developing monolingual children between 1;4 (years;months) and 2;8 speaking four Southern Bantu languages: isiXhosa, Sesotho, Setswana and Xitsonga. Our aim is to establish whether MLU (specifically MLU3) can be used as a measure of early language development in these languages and how best to calculate it. Is MLU3 a valid and practical measure given that there are few clinical tools to assess language development in Bantu languages? We also want to contribute to broadening knowledge of child language development and the use of MLU by including studies on languages that are under-represented in the child language literature (Kidd & Garcia, 2022).

This study uses MLU3 data comprising children's longest utterances, collected using MB-CDIs adapted for these four languages. These MB-CDIs are part of a project to adapt the MB-CDI for all South Africa's 11 spoken official languages (<https://sa-cdi.org>). MB-CDIs have been developed for over 100 languages worldwide (<https://md-cdi.stanford.edu>). These parent report instruments have proved to be valid and reliable tools for gathering lexical and grammatical norms for language development between 0;8 and 2;6 (Frank *et al.*, 2021).

Previous research

MLU as a measure of child language development

The idea that average sentence length might be a way of assessing children's language development was first proposed by Nice (1925). R. Brown's (1973) use of MLU as a simple index of the development of constructional complexity in children's early language was the impetus for wider uptake. He proposed five stages of children's early morphosyntactic development using MLU and the longest sentence produced at specific ages. Since then, MLU has been extensively applied in child language research and clinical practice (Parker & Brorson, 2005).

Although developed for English, studies using MLU have been done on a variety of different languages, including Afrikaans (Oosthuizen & Southwood, 2009); Basque (Ezeizabarrena & Garcia Fernandez, 2018); Bengali (Gouda *et al.*, 2020); Canadian Inuktitut (Allen & Dench, 2015); Greek (Voniati, 2016); Icelandic (Thordardottir & Weismer, 1998); Irish (Hickey, 1991); Mandarin Chinese (Wu, 2020); Spanish (Jackson-Maldonado & Conboy, 2007); and Turkish (Ege, 2010). MLU has also been used to address questions about bilingual children's development (see Ezeizabarrena & Garcia Fernandez, 2018; Marchman *et al.*, 2004; Meisel, 2011; Thordardottir, 2005).

Overall, MLU has been found to be a reliable measure correlating with age and other measures of general language development (Dethorne *et al.*, 2005; Ezeizabarrena & Garcia Fernandez, 2018; Rice *et al.*, 2006). MLU has also been usefully applied to children with language impairments (Dethorne *et al.*, 2005; Parker & Brorson, 2005; Rice *et al.*, 2006; Wiczorek, 2010; Wu, 2020). However, MLU may not be a valid measure of development at all ages (see Eisenberg *et al.*, 2001, for a discussion). R. Brown (1973) suggests that after Stage 5, at an MLU of four morphemes, the nature of an interaction will shape children's responses more than their underlying linguistic knowledge of a language system (see Southwood & Russell, 2004; Tolentino, 2022).

Several scholars also argue that, in older children, further linguistic complexity may involve internal embedding rather than additional length (Oosthuizen & Southwood,

2009) and a decrease in morphological growth (Parker & Brorson, 2005). Various studies have reported different ages at which MLU may no longer be a sensitive measure (see Allen & Dench, 2015; Cheung, 1998; Ezeizabarrena & Garcia Fernandez, 2018; Oosthuizen & Southwood, 2009; Rice et al., 2006, 2010; Scarborough et al., 1991; Tomas & Dorofeeva, 2019; Wu, 2020).

When using spontaneous language samples, the number of children's utterances needed to accurately measure MLU has also been debated (see Eisenberg et al., 2001, for a discussion). Pragmatic variation means that children may respond with single-word answers in interactions because it is not necessary for them to use additional words. Moreover, a single interactive event may not comprehensively represent a child's language abilities (Oosthuizen & Southwood, 2009; Southwood & Russell, 2004). Although R. Brown (1973) recommended using samples of 50 to 100 utterances, some studies have cautioned against using samples containing fewer than 175 utterances (Gavin & Giles, 1996; Oosthuizen & Southwood, 2009; Tomas & Dorofeeva, 2019).

Another variant of MLU is MLU3, based on the longest sentences a child currently produces. Studies that use MLU3 have found it to be a useful and valid tool for calculating utterance length as a measure of child language development (Allen & Dench, 2015; Ezeizabarrena & Garcia Fernandez, 2018; Fenson et al., 1994; Heilmann et al., 2005; Jackson-Maldonado & Conboy, 2007). Parent report studies find significant correlations between MLU3, age and other indices of language growth such as vocabulary and syntactic development (Ezeizabarrena & Garcia Fernandez, 2018). Jackson-Maldonado and Conboy (2007) suggest that parent report is more efficient as it can "bypass the performance limitations inherent in spontaneous and structured language sampling" (p. 147).

During early childhood, children show considerable variation in language skills when considered by age (Frank et al., 2021). MLU correlates significantly with age, but there is also considerable variability in MLU at specific ages (Dethorne et al., 2005; Tomas & Dorofeeva, 2019; Vasilyeva et al., 2008). R. Brown (1973) has argued that MLU is a more accurate tool for predicting children's expressive language abilities than age because MLU permits the identification of children at similar levels of constructional complexity regardless of chronological age. He holds that "two children matched for MLU are much more likely to have speech that is, on internal grounds, at the same level of constructional complexity than are two children of the same chronological age" (R. Brown, 1973, p. 55).

Some studies based on natural language samples report significantly high correlations between MLU-m and age, supporting the use of MLU-m as a general index of language development (Miller & Chapman, 1981, $r = .88$; Ege, 2010, $r = .81$; Thordardottir & Weismer, 1998, $r = .84$). However, other studies report that MLU-m is significantly but not necessarily always strongly correlated with age, with r scores between .3 and .8 (see Allen & Dench, 2015). Several studies also show that MLU-m has higher correlations with other indices of development than age does (Dromi & Berman, 1982; Thordardottir & Weismer, 1998).

Significant correlations are found between MLU and other indices of language development. For English, Dethorne et al. (2005) found that MLU-m is associated with both lexical and morphosyntactic development, with lexical development – measured as the number of different words – strongly correlating with MLU-m and accounting for 51% variance in MLU-m. Ezeizabarrena and Garcia Fernandez (2018) also report high correlations between MLU and other scales of communicative development, suggesting that MLU is a general measure of early development rather than a measure of one specific component such as semantics or morphosyntax. However, in some studies, MLU correlated better with measures of grammatical development and therefore may be a

more reliable indicator of grammatical ability when compared to other indices (Allen & Dench, 2015; Miller & Chapman, 1981; Thordardottir & Weismer, 1998).

Measurement of MLU in words versus morphemes

A central issue has been whether to measure MLU in words or morphemes. R. Brown (1973) argued that counting morphemes is a more accurate method as it takes inflectional complexity into account. With a language like English that has lower inflection and is morphologically sparse, there may be little difference between calculating MLU in words or morphemes (Ezeizabarrena & Garcia Fernandez, 2018). For instance, Parker and Brorson (2005) found high correlations ($r = .998$) between MLU scales (MLU-m and MLU-w) and age ($r = .69$ for both) in English-speaking children (3;0–3;10), concluding that MLU-w is as effective a measure as MLU-m. Similarly, there appears to be minimal difference between measurements in words versus morphemes in Cantonese and Mandarin Chinese – isolating languages with little inflection (Allen & Dench, 2015; Cheung, 1998; Klee *et al.*, 2004; Wu, 2020). Similar results have also been found for more inflectional languages such as Dutch (Arlman-Rupp *et al.*, 1976), Icelandic (Thordardottir & Weismer, 1998) and Irish (Hickey, 1991), with correlations of between .98 and .99 for MLU-w and MLU-m. Basque, an agglutinative and morphologically rich language, also reports a high correlation between MLU3-m and MLU3-w ($r = .97$) and strong correlations between vocabulary, nominal and verbal morphology, and MLU3-m and MLU3-w with an r range .81–.97 (reduced to .66–.95 when controlling for age (Ezeizabarrena & Garcia Fernandez, 2018)).

Based on significantly high correlations between MLU-m and MLU-w, several studies conclude that MLU-w provides a better alternative to MLU-m for tracking children's language development trajectories because it is easier to count words (Jackson-Maldonado & Conboy, 2007; Parker & Brorson, 2005; Thordardottir & Weismer, 1998). However, whereas measurements using words or morphemes might be similar in some languages, MLU-w is always likely to be equal to, or slightly lower than, MLU-m (Gouda *et al.*, 2020, for Bengali; Oosthuizen & Southwood, 2009). Wiczorek (2010) challenges the conclusion that high correlations between MLU-w and MLU-m mean either can be used, arguing that MLU-w is related to lexical development whereas MLU-m measurements are better indicators of grammatical development.

Some studies claim that a morpheme count is more accurate in morphologically rich languages such as synthetic languages like Russian (Tomas & Dorofeeva, 2019) and Hebrew (Dromi & Berman, 1982); polysynthetic languages like Canadian Inuktitut (Allen & Dench, 2015); or agglutinative languages like Turkish (Ege, 2010). Allen and Dench (2015) found that MLU-w had no significant correlation with age in Inuktitut, a polysynthetic language. They found that MLU-m and Mean Length of Words measured in morphemes and syllables, respectively, correlated significantly with age. However, some studies on morphologically rich languages (such as Basque) report that both MLU-m and MLU-w indicate children's expressive language skills equally well although MLU-w is a consistently lower measure (Ezeizabarrena & Garcia Fernandez, 2018).

Identifying and counting morphemes has its challenges. In synthetic languages like Spanish, Finnish, Dutch, Italian and Icelandic, one morpheme may contain more than one grammatical feature such as tense and person (see Allen & Dench, 2015; Arlman-Rupp *et al.*, 1976; Jackson-Maldonado & Conboy, 2007; Parker & Brorson, 2005, for discussions). Consequently, Allen and Dench (2015) argue that we may underestimate

morpheme counts in synthetic languages compared to agglutinative ones. Researchers need to decide whether to include zero, fused and suppletive morphemes in the morpheme count, and how to handle multimorphemic words such as portmanteaus and compounds (Ezeizabarrena & Garcia Fernandez, 2018). Researchers may find counting morphemes for MLU-m difficult – particularly for cross-linguistic studies – as decisions need to be made morpheme by morpheme and are language dependent (see, e.g., Allen & Dench, 2015, for Inuktitut; Dromi & Berman, 1982, for Hebrew; Eisenberg et al., 2001, general, with reference to American English; Hickey, 1991, for Irish; Tomas & Dorofeeva, 2019, for Russian; Voniati, 2016, for Cypriot Greek).

Whereas studies discuss the operationalization of morphemes and how to ensure consistent computation of an MLU-m index, we have found that most research does not question the definition of a ‘word’ or how to operationalize a word count, especially cross-linguistically. In fact, almost all studies conflate words with orthographic conventions. The only studies that question this issue are Allen and Dench’s (2015) study of Inuktitut, and the Bengali study by Gouda et al. (2020). Allen and Dench (2015) suggest that word counts are less arbitrary than morpheme counts, and that operationalizing a word count is not a challenge for most languages. However, they point out that in morphologically rich polysynthetic languages like Inuktitut – the focus of their study – much of the grammar is realized within the morphology word boundaries, which they then identify orthographically (Allen & Dench, 2015, p. 384). Their solution is to introduce morphemes per word counts as a word can contain ten or more morphemes.

E. K. Brown and Miller (2013, p. 473) conceive of a ‘word’ as an uninterrupted linguistic unit consisting of a stem with or without affixes, where no constituents may be inserted between a stem and its affixes. This definition aligns closely with orthographic word boundaries in most languages where MLU-w has been employed as a measure. Typically, orthographic boundaries (i.e., spaces) between morphemes have been taken to indicate the boundaries of a ‘word’, and each item orthographically separated by a space to be equal to one ‘word’. Given that orthographies are products of convention and consensus, orthography should be treated with caution if used as a basis for linguistic analysis. However, orthographic boundaries may serve as good common-sense dividing lines between ‘words’ (especially for clinical purposes) in languages whose orthographies closely align word boundaries with morpheme boundaries (e.g., English, Dutch). For example, Allen and Dench (2015) view counting words as more practical if, and only if, a language’s writing system “leaves spaces between words,” (p. 384) again equating the concept of ‘word’ with orthographic boundaries. For such languages, researchers and clinicians may perceive MLU-w to be more convenient, faster, easier, more reliable, simpler to implement, more adaptable and less arbitrary in nature than MLU-m (Hickey, 1991; Parker & Brorson, 2005). However, when orthographic word boundaries do not align with linguistic boundaries as an uninterrupted stem + affixes unit, the accuracy of MLU-w as a measure is in doubt.³ In agglutinative Bantu languages, we note that orthographic standards may determine the segmentation of words rather than linguistic criteria. Using orthographic word boundaries for counting words presents challenges in agglutinative languages such as Basque (Ezeizabarrena, personal communication December 22, 2022) and polysynthetic languages such as Inuktitut (Allen & Dench, 2015).

³The orthographic conventions of the Southern Bantu languages present some challenges in this regard and are discussed under *Southern Bantu Languages*.

Southern Bantu Languages

The languages in this study belong to the Bantu language family (Atlantic-Congo) and represent three Southern Bantu language groups: Nguni group (isiXhosa), Sotho-Tswana group (Sesotho and Setswana), and Tswa-Ronga group (Xitsonga). These four languages are agglutinative. They have high inflectional paradigms, strong agreement, high derivation and morpheme-to-word ratio, and a very low number of free roots (Nurse & Philippson, 2006). Both nouns and verbs have abstract roots. Bantu languages categorise nouns using a noun class system – comparable to grammatical genders in some other languages – where each noun class uses a particular morphemic prefix, or ‘noun class prefix’ (Nurse & Philippson, 2006). The noun classes are numbered using an internationally recognized system, and some Bantu languages have up to 19 noun classes (Nurse & Philippson, 2006). Adjectives, verbs, and function words show agreement with the class of the noun they are referring to, overtly expressed with a morphemic prefix (Nurse & Philippson, 2006). Tense, aspect, mood, causativity and negation are marked on the verb with affixes (Nurse & Philippson, 2006). The default word order is Subject-Verb-Object (SVO), but free word order and extraposition for discourse purposes are common (Demuth, 1992; Nurse & Philippson, 2006). The syllable structure is typically Consonant-Vowel (CV). These aspects of Bantu languages contrast with analytic languages like English, that has a weak inflectional paradigm and agreement, high derivation, and a high number of free roots with a low morpheme-to-word ratio (Nurse & Philippson, 2006).

The orthographies of South African Bantu languages have not been standardized to the same degree as those of European languages (Taljard & Bosch, 2006). The Nguni languages use a conjunctive orthography, whereas Sotho-Tswana languages employ a disjunctive orthography. Xitsonga orthography is less clear-cut but is more disjunctive than conjunctive (Lee & Hlungwani, 2017). The different orthographic conventions, decided by missionaries, can be seen more as historical accidents than as indicative of underlying differences in phrase construction in these languages. The disjunctive orthographies of Sesotho and Setswana regularly insert word boundaries between affixal and stem morphemes that do not constitute uninterrupted words (see examples 1a and 1b). In languages with a conjunctive orthography, one orthographic ‘word’ may correspond to two or more linguistic elements (see examples 1c and 2c), whereas in languages with disjunctive orthography, several orthographic ‘words’ may correspond to one linguistic ‘word’ (see examples 3a and 3b).

The following examples demonstrate these orthographic differences across the four languages and the relative consistency of phrase construction between them.⁴

- (1a) Sesotho
Ke a o rata [4 words]
 ke-a-o-rat.a [4 morphemes]
 SM.1SG-PROG-OM.2SG-love.FV
 ‘I love you’

⁴The less common morpheme glosses are as follows: SM = subject marker, OM = object marker, FV = final vowel, HORT = hortative, PVB = preverb, SIM = simultaneous aspect. Roman numerals indicate the number of the relevant noun class (e.g., XV = noun class 15).

- (1b) Setswana
Ke a go rata [4 words]
 ke-a-go-rat.a [4 morphemes]
 SM.1SG-PROG-OM.2SG-love.FV
 'I love you'
- (1c) isiXhosa
Ndiyakuthanda [1 word]
 ndi-ya-ku-thand.a [4 morphemes]
 SM.1SG-PROG-OM.2SG-love.FV
 'I love you'
- (1d) Xitsonga
Ndza ku rhandza [3 words]
 ndza-ku-rhandz.a [3 morphemes]
 SM.1SG.PRES-OM.2SG-love.FV
 'I love you'
- (2a) Sesotho
Ha re yeng ho bapala [5 words]
 ha-re-y-e-ng-ho-bapal.a [7 morphemes]
 HORT.PVB-SM.1PL-go-HORT-AGR.1PL-AGR.XV-play.FV
 'Let's go to play'
- (2b) Setswana
A re ye go tshameka [5 words]
 a-re-y-e-go-tshamek.a [6 morphemes]
 HORT.PVB-SM.1PL-go-HORT-AGR.XV-play.FV
 'Let's go to play'
- (2c) isiXhosa
Masiyokudlala [1 word]
 ma-si-yo-ku-dlal.a [5 morphemes]
 HORT.PVB-SM.1PL-go-XV-play.FV
 'Let's go to play'
- (2d) Xitsonga
Ahi fambe hi ya tlanga [5 words]
 a-hi-famb-e-hi-ya-tlang.a [7 morphemes]
 HORT.PVB-SM.1PL-go-HORT-SM.1PL-go-play.FV
 'Let's go to play'
- (3a) Sesotho
Letsatsi le leng [3 words]
 le-tsatsi-le-le-ng [5 morphemes]
 V-day-SIM-AGR.V-ADJ.one
 'Another day'

- (3b) Setswana
Letsatsi *le lengwe* [3 words]
 le-tsatsi-le-le-ngwe [5 morphemes]
 V-day-SIM-AGR.V-ADJ.one
 ‘Another day’
- (3c) isiXhosa
Olunye *usuku* [2 words]
 o-lu-nye-u-suku [5 morphemes]
 DEM-AGR.XI-ADJ.one-XI-day
 ‘Another day’
- (3d) Xitsonga
Siku *rin’wana* [2 words]
 Ø-siku-ri-n’wana [3 morphemes]
 III-day-SM.III-ADJ.one
 ‘Another day’

Whereas morpheme boundaries are relatively clear, examples 1 to 3 illustrate the challenges that the different orthographic conventions for these morphologically rich, agglutinative languages pose, and that operationalizing word boundaries is not as simple as in analytic or even synthetic languages with relatively disjunctive orthographies. However, following most other studies, we operationalize ‘word’ using the definition of an orthographic word: all linguistic elements that are written separately in the practical orthography and are separated by spaces (Louwrens & Poulos, 2006). Since some of the languages in this study are written conjunctively and some disjunctively, we will see whether the outcome of using morphemes or words in languages with disjunctive and conjunctive orthographies is substantially different. These comparisons will have clinical implications in a multilingual situation when comparing children speaking typologically different languages with different orthographic conventions as is the case in South Africa where different Bantu and Germanic languages are spoken.

The present study

This study explores the use of MLU to measure early language development in morphologically complex Bantu languages, that are also agglutinative and typologically different from languages in previous studies of MLU. As stated above, MLU is generally regarded as a better predictor of language abilities than age in early acquisition, but previous findings differ as to whether MLU measures general language development or morphosyntactic development. Prior research also diverges on what is the most suitable unit in which to measure MLU in typologically different languages. Given these knowledge gaps, especially in relation to morphologically complex languages, our study addresses the following questions:

1. Is MLU a valid measure of language development in (Southern) Bantu languages? What does MLU measure, and is it a more reliable indicator of development than age in these languages?

Given that previous studies found MLU to be a valid measure of general language abilities and often a better predictor than age, we hypothesize that MLU3 will correlate positively with age, vocabulary (i.e., lexical production) and grammar measures, indicating its validity as a measure of general language development and a more reliable indicator than age.

2. Is MLU best measured, from linguistic and clinical perspectives, in words or morphemes in these Bantu languages, given their complex morphology and inconsistent orthographies? Can we claim that words and morphemes are equally valid measures in agglutinative languages?

Given that morphemes have been found to be a more accurate measure in several morphologically complex, agglutinative and synthetic languages, we hypothesize that morphemes will also be a more accurate and appropriate measure of language development than words in these Bantu languages that are morphologically complex and agglutinative: where languages are written conjunctively, the difference between word and morpheme measures will be significant with morpheme measures significantly higher than word measures.

Data collection and methods

The MB-CDI instrument

The MB-CDIs for toddlers (1;4-2;6) were adapted by mother tongue speaking linguists and speech-language therapists of the four languages following the MB-CDI Board's guidelines (<http://mb-cdi.stanford.edu>).⁵ These adaptations were informed by the word-lists in the UK and USA versions of the MB-CDI, and by the grammatical items in the Kenyan Kilifi versions of the MB-CDI⁶ that focus on two Bantu languages as well as by spontaneous speech samples from six children per language collected in their homes. These versions underwent further adaptation in focus groups followed by two pilot studies with participants from communities in which these languages are spoken. At each stage, lexical and grammatical items were evaluated and retained or removed depending on their suitability. A total of 1200 words were tested per language in the first pilot and between 733 and 748 in the second. Forty grammatical items were tested in the first pilot and reduced to 37 items in the second pilot. Where possible, we retained lexicosemantic and grammatical equivalence across the four languages. This study is based on data from the second pilot.

The toddler MB-CDI measures lexical production and early grammatical development. It presents caregivers with a word list grouped into 21 semantic domains and asks them to identify the words their child produces. Then caregivers are asked about their children's early grammatical development. The first two grammar sections focus on word complexity relating to noun class prefixes and verbal affixes. Caregivers are then asked whether their child is combining words and, if they are, to provide their child's three longest utterances of the past week (see Fenson et al., 1994; Frank et al., 2021). Sentence

⁵Permission to conduct these adaptations was obtained from the MB-CDI Advisory Board.

⁶We acknowledge the important contribution made by Katie Alcock who shared her work on the Kenyan Kilifi versions.

length and complexity are then measured by presenting the caregivers with a series of items (e.g., inflected words, phrases, and sentence constructions) with three to four options reflecting different levels of morphosyntactic development. Caregivers are asked to identify which option reflects best what their child currently produces. (Examples of items testing word and sentence complexity and how they are scored can be found in Statistical Analysis below).

Procedure

Whereas CDIs can be answered by caregivers independently in many contexts, in the South African environment, low literacy levels and lack of familiarity with answering questionnaires – especially online – necessitated that we recruit and train fieldworkers from local communities to administer the CDIs either in person or (during COVID) telephonically or online. Fieldworkers recruited participants through local social networks and childcare organizations in their communities. Fieldworkers used the Qualtrics (2019–2020) online survey tool to capture participants' responses.

In addition to completing the CDIs, participants were asked to complete a family background questionnaire that collected information on the mother's pregnancy, the parents' education and employment status, the socio-economic status of the child's family, family composition and housing, the child's medical history (e.g., ear infections), known or suspected areas of communication difficulties (speech-language difficulty, hearing impairments, developmental disability), and the child's level of exposure to other languages.

Participants

Data were collected from caregivers of 472 toddlers (1;4–2;8). The isiXhosa, Sesotho and Setswana toddlers were recruited from urban and rural areas where these languages predominate (See Table 1). Previous studies on language development in African contexts point to differences between rural and urban children (Alcock *et al.*, 2015; Vogt *et al.*, 2015). However, monolingual Xitsonga toddlers could only be recruited from rural and semi-urban areas. (Xitsonga is a minority language, and Xitsonga-speaking toddlers in urban areas acquire other languages from an early age.) We sampled isiXhosa urban speakers in the city of Cape Town in the Western Cape Province and rural speakers in Ilinge in the Eastern Cape Province. Urban Sesotho speakers were drawn from the city of Bloemfontein in Free State Province and rural participants from a rural area 300 km north-east of Bloemfontein in the same province. Urban Setswana speakers were sampled in the town of Rustenburg and rural speakers in the Taung area (394 km from Rustenburg), both in North-West Province. We recruited Xitsonga speakers from semi-rural Giyani and rural Malamulele in Limpopo Province; young Xitsonga-speaking children have little exposure to another language in these areas.

The combined monthly income of households of participants ranged from 0 to 80,000 ZAR (0–4,661 USD). The mean monthly income per household was in the range 2,401 to 5,000 ZAR (140–291 USD), whereas the mean monthly food expenditure per household was in the range 1,201 to 2,000 ZAR (70–117 USD). Regarding parental education, for 29% of children both parents had completed high school or studied beyond, for 21% neither parent had completed high school or studied beyond, and for 38% one parent had completed high school or studied beyond (11% were fathers and 27% mothers). In 12% of

Table 1. Child Demographic Information by Language

Variable	Sesotho	Setswana	IsiXhosa	Xitsonga	Total
Number of Children	113	123	113	99	448
Child Age (Months)					
Range	16–32	16–32	16–32	16–32	16–32
Median	24.0	23.0	22.0	23.0	23.0
Interquartile Range	20–28	19–28	18–27	19–28	19–27
Child Sex					
Female	62	60	54	43	219
Male	51	63	57	56	227
Sex not Specified			2		2
Area					
Rural	58	59	59	99	275
Urban	55	64	54	-	173

the cases, there was incomplete information on parental education. Regarding employment status, for 12% of children, both parents were employed; for 20%, neither parent was employed; and for 41%, one parent was employed (26% were fathers and 15% mothers). Twenty seven percent had incomplete information on employment status. The mean number of adults per household in urban areas was lower ($M = 2.8$, $SD = 1.5$) than in rural areas ($M = 3.5$, $SD = 1.8$). The mean number of children younger than 18 in each household in urban areas was also lower ($M = 1.5$, $SD = 2.3$) than in rural areas ($M = 4.3$, $SD = 5.1$).

All children with suspected language difficulties, as reported by participants, were excluded. We found differences in vocabulary size for children exposed to another language for more than four hours per day (see Southwood et al., 2021); thirteen children were excluded for this reason (isiXhosa = 2, Sesotho = 2, Setswana = 8, Xitsonga = 1). Out of the remaining 459 toddlers, 11 were excluded due to (a) missing vocabulary scores (i.e., only the background questionnaire was completed, not the CDI too), or (b) being outside the target age range. The 448 children included in the study (males: $n = 222$; females: $n = 224$; not specified $n = 2$) were aged 1;4–2;8. We included the children whose gender was not specified as gender comparison was not the focus of the study, and as we do not yet have norms that would allow us to consider gender differences. Table 1 gives the number of children per language, the sample distribution per age in months, gender, and whether rural or urban.

Informed consent was sought from the parents before interviews, after the details of the study had been explained to them. They were informed of their right to withdraw from the study at any point without any negative repercussions. Data were collected anonymously and stored securely. Participating caregivers received a supermarket voucher of approximately 10 USD to thank them for participating. The study was reviewed and approved by the Ethics Committee of the Linguistics Section, and the Faculty of Health Sciences Human Research Ethics Committee, both at the University of Cape Town.

Calculating MLU3-w and MLU3-m

The data for Sesotho, Setswana and Xitsonga were analysed by two linguists and a first language speaker with formal linguistic training at university level. For isiXhosa, a linguist who is a second language speaker of isiXhosa worked with a first language speaker with no formal linguistic training to analyse the data. One of the linguists participated in the analysis of all four languages to ensure consistency across languages.

Uniform inclusion and exclusion criteria were applied across all four languages. As there are morphological differences between the languages – such as the use of pre-prefixes in isiXhosa but not in the other three languages – another set of morpheme-counting rules was also applied, some of which were language-specific.

Inclusion and exclusion criteria for all four languages

1. If the participant indicated that the child was not yet combining words (and therefore did not provide any sentence examples on the CDI), an MLU3 score of 1 was given (isiXhosa = 43, Sesotho = 34, Setswana = 33, Xitsonga = 17).
2. If the participant gave fewer than three sentences as examples, these data sets were included, and an average score was calculated on the number of sentences provided. This did not lead to complications as other criteria (4 and 5) also resulted in only two sentence examples being analysed for some children (Sesotho = 11, isiXhosa = 5, Setswana = 3, Xitsonga = 1).
3. If two sentences were given as one example, they were separated and an average was calculated counting each sentence separately (Sesotho = 4; isiXhosa, Setswana, Xitsonga = 0).
4. Sentence examples were excluded if the meaning of the sentence could not be determined (isiXhosa = 5, Setswana = 3, Xitsonga = 1, Sesotho = 0).
5. If a sentence was entirely in a different language, it was excluded. (There was only one occurrence hereof in the data – namely, ‘I love you mommy’, in English, for one Setswana child).
6. Sentences that contained a loan word from another language were not excluded (Setswana = 10, Sesotho = 5, isiXhosa = 2, Xitsonga = 1).

Morpheme counting and language-specific rules

1. Contractions were counted as separate words (e.g., Setswana *ke a* ‘I am’, written as *ka*; Sesotho *ile go* ‘went to’, written as *ilo*).
2. Noun class prefixes were counted as separate morphemes. This included noun class prefixes applied to foreign loan words (e.g., Sesotho *ma-Simba* ‘Simba chips’). This rule had two exceptions:
 - 2.1. The noun class prefix of mass nouns (where the root is not a free morpheme) was not counted as a separate morpheme (e.g., Setswana *me.tsi* ‘water’, compared to *me-sese* ‘dresses’).
 - 2.2. If the noun class prefix was a null prefix, it was not counted as a separate morpheme (e.g., Xitsonga *Ø-siku* ‘day’, compared to *ri-tiho* ‘finger’).
3. The final vowel in verbs was counted as part of the root.
4. In isiXhosa, nouns have a vowel pre-prefix before the noun class prefix (compare Setswana *le-so* ‘spoon’, to isiXhosa *i-li-so* ‘spoon’). The pre-prefix was counted as a separate morpheme, even if it preceded a bound noun class prefix (see rule 2.1).

- 4.1. In Xitsonga, some data were in a dialect that had both a pre-prefix and a noun class prefix, whereas other data were in a dialect that had no pre-prefix (e.g., *swa-ku-dya* or *swo-dya* 'food'). Rule 4 was applied in instances with a pre-prefix.
5. Morphemes that were reduplicated (onomatopoeically or for emphasis) were counted only once (e.g., isiXhosa *nqo-nqo-nqo* 'knocking', or 'knock-knock'; Setswana *hu-hu-hu* 'barking / woof-woof').

Statistical analysis

To establish whether MLU3 is a valid indicator of language development, and whether MLU3-w or MLU3-m is a better measure, we compared both MLU3-w and MLU3-m scores, based on the three longest sentences caregivers reported for each child, with their vocabulary (i.e., lexical production) and grammar scores, based on the wordlist and grammatical items, respectively, in their completed MB-CDIs. To obtain their MLU3-w and MLU3-m scores, we counted the words and morphemes in each child's reported sentences and calculated simple averages. Their vocabulary scores were obtained by counting the number of words from the CDI wordlist that each caregiver said their child produced. For the four languages, the maximum possible vocabulary score ranged from 733 to 748 lexical items (isiXhosa = 748, Sesotho = 741, Setswana = 734, Xitsonga = 733).

For grammar, caregivers were asked whether their child had started using grammatical features such as plurals and noun class prefixes. These items were scored one for yes and zero for no. Grammatical complexity at the word level was then tested in detail with each option minimally more complex or more grammatically correct than the last. For example, noun prefixes were tested by presenting a word stem with no prefix, then the word with a shadow prefix⁷ followed by the word with its full and correct prefix. For example, in Sesotho, the caregiver could choose between *shemane*, *a-shemane*, or the fully correct form *bashemane* 'boys.' These were scored one, two and three, respectively. For sentence complexity, each item had between two and four options for the participant to choose from, set in order of increasing length and complexity. For example, we asked the caregiver to choose which option best reflected what the child would say for phrases such as 'the ball is on top of the table.' In Xitsonga, the choice was for instance between (a) *bolo tafula* 'ball table', (b) *bolo henhla tafula* 'ball on table', and (c) *bolo yile henhla ka tafula* 'the ball is on top of the table'. These options were scored 1, 2, and 3, respectively, and a fourth option would be scored 4 if the item had one. The maximum possible scores for the entire grammar section ranged from 93 to 103 across the languages (Setswana = 103, Sesotho = 102, isiXhosa = 98, Xitsonga = 93).

Underlying distributions of responses were described through empirical distribution plots, and consideration of skewness and kurtosis parameters. Internal consistency of scales was assessed using Cronbach's alpha. Association between measures was illustrated using pairwise scatterplots and measured using Spearman correlation coefficients, as opposed to Pearson correlation coefficients, to guard against bias due to possible deviations from normality. Regression models were used to further estimate and compare (using adjusted multiple correlation squared [R^2] and Akaike Information Criterion [AIC] statistics) associations between grammar and vocabulary scales with

⁷Shadow prefixes are a feature of development in the acquisition of noun class prefixes (Demuth, 2003)

age, MLU3-m, and MLU3-w measurements. Regression models were validated through an examination of the model residuals.

Results

Cronbach’s Alpha statistics for MLU3-m, MLU3-w, grammar, and vocabulary scales demonstrated high internal consistency, yielding alpha levels above .90 ($\alpha = >.90$). Skewness statistics for MLU3-m (.562), MLU3-w (.399), grammar (.276), and vocabulary (.529) were in the acceptable ranges. Kurtosis ranges for MLU3-w (–.537), grammar (–.753), and vocabulary (–.665) were in the negative but not for the MLU3-m (.249) measure. Overall, data were approximately normally distributed.

Means and standard deviations for MLU3-m, MLU3-w, grammar, and vocabulary are presented per language in Table 2. Scores on all four scales increased with age (in months). However, as expected, there was considerable variability across age within each language and variability between the four languages for MLU3-m, MLU3-w, grammar, and vocabulary.

We hypothesized that MLU3 would correlate with age, vocabulary and grammar measures, indicating its validity as a measure of language development, and that MLU3 would be a more sensitive measure of grammar and vocabulary development than age. Figure 1 presents the pairwise scatterplots for age, MLU3-m, MLU3-w, grammar, and vocabulary variables in each language.

Spearman correlation coefficients are presented, though Spearman and Pearson correlation analyses yielded similar results: all pairwise associations were significant, with all but two of them with $p < 0.001$. However, the relative strength of pairwise correlations differed between scales and within languages. The interpretation guideline of Dancy and Reidy (2007) was used to interpret the strength of correlations. Correlation analysis indicated that age was associated with MLU3-m (isiXhosa $r = .62$, Sesotho $r = .58$, Setswana $r = .40$, Xitsonga $r = .36$). The relationship between age and MLU3-m was weak for Xitsonga, moderate for Sesotho and Setswana, and strong for isiXhosa. There was an association between age and MLU3-w across languages (Sesotho, isiXhosa $r = .59$; Setswana $r = .42$; Xitsonga $r = .38$). This association was moderate for Sesotho, Setswana, and isiXhosa but again slightly weaker in Xitsonga. Age showed significant and moderate

Table 2. Mean and Standard Deviation Scores for MLU3-m, MLU3-w, Grammar, and Vocabulary Scales

Age	Sesotho (n = 113)				Setswana (n = 125)				isiXhosa (n = 113)				Xitsonga (n = 99)																				
	MLU3-m		MLU3-w		Grammar		Vocabulary		MLU3-m		MLU3-w		Grammar		Vocabulary		MLU3-m		MLU3-w		Grammar		Vocabulary										
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD									
16	21	12	1.8	49	37.9	112	23.3	7.8	1.8	35	1.5	41	37.2	9.4	7.7	1.9	1.6	47	1.2	35	27.8	261	28.7	21.2	1.3	39	1.3	30	23.5	160	30.8	24.3	
17	13	19	1.3	39	26.4	64	18.6	4.0	2.2	48	2.0	72	22.1	12.7	6.1	4.8	1.8	1.2	1.5	46	31.8	146	27.3	16.1	2.9	1.6	2.7	1.4	44.2	21.9	35.1	21.7	
18	12	47	1.1	35	30.3	92	18.2	18.1	2.3	1.4	2.0	11	29.6	12.7	12.4	9.2	1.6	1.2	1.3	46	23.6	144	28.1	22.8	3.5	2.0	3.0	1.6	37.9	14.8	28.8	21.5	
19	2.8	1.9	2.5	1.6	44.7	192	32.8	31.8	2.3	39	2.2	40	32.5	14.1	9.5	4.9	1.6	1.6	1.9	1.4	49	28.3	7.7	40.7	18.8	2.8	1.1	2.5	59	43.6	28.1	44.7	30.0
20	2.0	1.6	1.7	1.2	35.1	147	28.4	21.2	2.5	1.5	2.5	1.4	36.6	18.4	17.0	6.1	1.0	0.0	1.0	0.0	20.4	182	26.5	6.8	2.9	1.8	2.4	1.3	38.1	18.9	59.7	26.6	
21	3.1	1.8	2.7	1.2	57.7	31.0	48.8	21.4	2.2	1.2	2.2	1.1	33.4	18.2	16.6	11.9	2.7	2.6	1.6	89	27.3	22.1	25.0	54.4	4.2	1.9	3.8	1.9	50.2	11.4	21.6	21.8	
22	1.0	0.0	1.0	0.0	25.2	2.7	19.6	8.7	3.6	1.1	3.2	1.1	49.5	22.0	28.8	17.5	3.6	2.5	2.1	1.2	29.6	17.4	29.0	13.3	3.3	3.6	2.9	5.5	49.7	14.4	28.2	18.9	
23	2.8	2.2	2.4	1.6	48.5	27.5	53.8	34.6	2.9	1.8	2.6	1.4	40.0	19.7	19.9	11.3	5.3	0.0	2.3	0.0	43.9	0.6	41.2	6.0	2.7	1.3	2.5	1.1	51.0	29.3	50.7	21.1	
24	4.1	3.9	3.5	62.1	14.7	44.5	18.0	3.1	2.1	2.8	1.8	41.7	24.9	19.3	17.9	4.1	2.2	2.4	9.7	43.1	24.4	32.8	27.9	3.9	3.8	3.2	40	86.9	12.2	53.6	27.7		
25	4.0	1.3	3.3	1.0	49.9	14.4	45.2	19.4	3.0	1.4	2.8	1.3	43.3	15.4	22.8	9.7	4.2	1.2	2.2	1.2	50.5	37.9	40.5	20.6	4.2	1.8	3.7	1.4	67.8	18.5	52.4	13.5	
26	3.7	3.8	3.4	51	58.1	22.7	42.9	18.5	3.5	1.7	2.9	1.3	46.4	22.2	34.3	32.0	3.5	1.8	2.3	9.5	39.1	17.4	40.2	14.3	3.6	1.3	3.0	1.2	69.5	24.6	56.5	14.9	
27	4.1	2.2	3.5	1.9	52.3	18.4	41.3	24.9	3.2	1.1	2.7	3.9	36.8	18.4	36.6	18.1	3.6	1.4	2.1	5.1	42.3	20.0	49.0	20.0	4.7	3.8	3.8	1.9	68.5	21.5	59.6	35.2	
28	4.2	3.9	3.7	76	60.5	162	55.1	19.9	3.8	1.9	3.4	1.5	46.6	13.8	30.0	17.7	5.8	4.1	3.0	2.1	41.3	18.7	43.8	26.3	3.3	1.2	3.0	59	63.6	26.9	61.7	30.4	
29	4.5	3.1	4.0	62	63.5	13.9	53.5	31.3	4.0	3.9	3.6	3.8	66.2	18.0	33.6	21.4	5.4	1.1	3.0	3.8	70.6	142	7.8	32.6	3.8	1.2	3.4	81	55.7	11.5	53.8	31.1	
30	6.3	2.4	5.6	1.6	75.7	8.3	56.2	11.2	3.2	1.4	2.8	1.1	43.9	19.2	15.9	7.3	5.3	4.1	2.9	49	60.7	14.1	62.9	14.7	5.0	1.8	4.2	1.6	76.9	17.8	72.9	18.1	
31	4.4	1.7	3.8	1.2	76.3	19.6	58.9	31.1	3.3	1.6	3.3	1.2	52.4	16.6	23.2	9.1	5.3	4.9	2.3	3.3	65.3	4.6	59.1	1.9	4.9	3.5	3.9	1.7	54.5	12.3	63.1	37.9	
32	6.2	2.4	4.2	1.2	81.8	14.6	53.5	7.8	3.8	1.8	3.5	1.7	55.5	25.3	39.7	11.1	7.6	2.5	3.3	1.5	74.5	12.2	68.6	17.9	4.6	2.0	3.8	1.4	71.4	21.2	66.0	25.7	
Total	3.3	1.9	2.9	1.5	51.5	21.7	41.0	24.8	2.9	1.5	2.7	1.8	41.8	18.7	21.2	16.7	3.2	2.3	1.9	99	38.4	21.9	37.8	24.2	3.6	1.7	3.1	1.3	54.4	20.4	48.2	27.9	

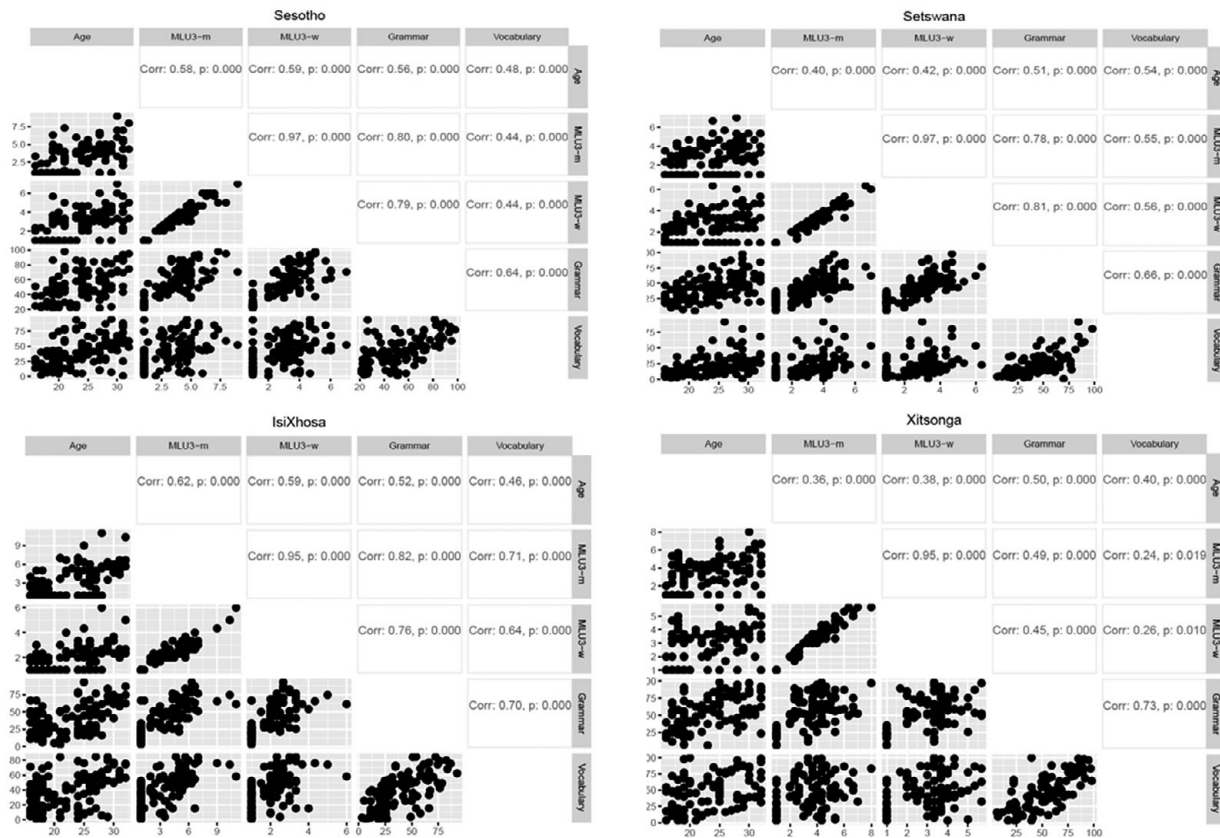


Figure 1. Pairwise Scatterplots for Sesotho, Setswana, isiXhosa, and Xitsonga.

correlations with grammar across languages (Sesotho $r = .56$, isiXhosa $r = .52$, Setswana $r = .51$, and Xitsonga $r = .50$). Moderate correlations were also observed between age and vocabulary scores across languages (Setswana $r = .54$, Sesotho $r = .48$, isiXhosa $r = .46$, and Xitsonga $r = .40$).

MLU3-m and MLU3-w were very strongly correlated in each language (Sesotho, Setswana $r = .97$; isiXhosa, Xitsonga $r = .95$). MLU3-m correlated with grammar across languages (isiXhosa $r = .82$, Sesotho $r = .80$, Setswana $r = .78$, Xitsonga $r = .49$), with strong correlations for Sesotho, Setswana, and isiXhosa, and a moderate correlation for Xitsonga. MLU3-w also correlated with grammar across languages (Setswana $r = .81$, Sesotho $r = .79$, isiXhosa $r = .76$, and Xitsonga $r = .45$), again with strong correlations for Sesotho, Setswana, and isiXhosa but a moderate correlation for Xitsonga. MLU3-m correlated with vocabulary across languages (isiXhosa $r = .71$, Setswana $r = .55$, Sesotho $r = .44$, Xitsonga $r = .24$). The correlations were weak for Xitsonga, moderate for Sesotho and Setswana, and strong for isiXhosa. MLU3-w also correlated with vocabulary development across languages (isiXhosa $r = .64$, Setswana $r = .56$, Sesotho $r = .44$, Xitsonga $r = .26$). A weak correlation was observed for Xitsonga, and moderate correlations for Sesotho, Setswana, and isiXhosa. MLU3-m and MLU3-w had higher correlations with grammar than with vocabulary, for all languages. MLU3-m and MLU3-w also had higher correlations with grammar than age had with grammar, except for Xitsonga. For vocabulary, the correlations between MLU3 variables and vocabulary were similar to the correlations between age and vocabulary for Setswana; however, MLU3 measures in isiXhosa correlated better than age whereas it was the opposite for Sesotho and particularly for Xitsonga.

To compare the reliability of MLU3-m with MLU3-w as measures of language development in these morphologically rich languages, we conducted simple regression analyses (see Tables 3 and 4) and compared the strength of the associations using model fit statistics. Age, MLU3-m, and MLU3-w served as the independent variables whereas grammar and vocabulary were the dependent variables. AIC was used to compare the goodness of fit of the MLU3-m and MLU3-w models, in each language and for all languages combined. Residual distributions did not differ significantly from the normal distribution, neither was significant non-constant variance observed.

Findings based on combined data sets (Tables 3 and 4) indicated that the proportion of variance accounted for by age in grammar (28%) and vocabulary (16%) development was low. MLU3-m accounted for 49% of the variability in grammar and 22% in vocabulary scores across all languages combined. MLU3-w contributed 48% and 16% of the variance in grammar and vocabulary scores, respectively, across all languages combined. Notwithstanding these low R^2 statistics, the AIC showed that of the two MLU3 variables, MLU3-m is the better predictor for grammar and vocabulary.

Within-group statistics for Sesotho (Tables 3 and 4) showed that age accounted for 32% and 20% of the variance in grammar and vocabulary scores, respectively. MLU3-m accounted for 62% of the variance in grammar and 17% in vocabulary measures, whereas MLU3-w explained 59% of the variability in grammar and 17% in vocabulary development. MLU3 variables were thus better predictors of grammar than of vocabulary. The AIC results provide support for MLU3-m being a better model than MLU3-w for grammar. For vocabulary, the AIC also suggests MLU3-m as a better model for Sesotho; however, the R^2 statistics are very low.

Setswana group findings (Tables 3 and 4) indicated that age accounted for 25% and 22% of the variance in grammar and vocabulary measures, respectively. MLU3-m explained significant variance in grammar (55%) and vocabulary (21%) scales, and

Table 3. Simple Linear Regression for Age, MLU3-m, and MLU3-w Predictors of Grammar

Language	Grammar						
	Parameter	Est.	SE	<i>t</i>	<i>p</i>	<i>R</i> ²	AICc
Sesotho	Intercept	-13.2	9.1	-1.5	.148		
	Age	2.7	0.4	7.8	<.001	.32	
	Intercept	22.3	2.5	8.9	<.001		
	MLU3-m	8.9	0.7	13.5	<.001	.62	911.8
	Intercept	20.4	2.8	7.4	<.001		
	MLU3-w	10.9	0.9	12.8	<.001	.59	919.2
Setswana	Intercept	-7.3	7.7	-0.9	.352		
	Age	2.1	0.3	6.4	<.001	.25	
	Intercept	11.1	2.7	4.1	<.001		
	MLU3-m	9.9	0.8	12.2	<.001	.55	989.0
	Intercept	12.9	2.8	4.6	<.001		
	MLU3-w	11.6	0.9	12.6	<.001	.56	984.4
IsiXhosa	Intercept	-18.6	7.8	2.3	.018		
	Age	2.5	0.3	7.5	<.001	.33	
	Intercept	15.1	2.2	6.8	<.001		
	MLU3-m	7.4	0.6	12.9	<.001	.60	918.7
	Intercept	8.9	3.3	2.8	.007		
	MLU3-w	15.3	1.5	10.1	<.001	.48	949.3
Xitsonga	Intercept	-0.6	9.8	-0.1	.949		
	Age	2.4	0.4	5.8	<.001	.25	
	Intercept	27.8	4.8	5.8	<.001		
	MLU3-m	7.5	1.2	6.1	<.001	.27	877.9
	Intercept	2.5	5.4	4.7	<.001		
	MLU3-w	9.4	1.6	5.8	<.001	.25	881.1
Four languages combined	Intercept	-11.7	4.4	-2.6	.009		
	Age	2.5	0.1	13.3	<.001	.28	
	Intercept	19.0	1.5	12.6	<.001		
	MLU3-m	8.4	0.4	20.9	<.001	.49	3761.5
	Intercept	15.9	1.7	9.5	<.001		
	MLU3-w	11.6	0.6	20.6	<.001	.48	3768.0

Table 4. Simple Linear Regression for Age, MLU3-m, and MLU3-w Predictors of Vocabulary

Language	Vocabulary						
	Parameter	Est.	SE	<i>t</i>	<i>p</i>	<i>R</i> ²	AICc
Sesotho	Intercept	−16.1	10.8	−1.5	.138		
	Age	2.4	0.5	5.4	<.001	.20	
	Intercept	23.8	4.1	5.9	<.001		
	MLU3-m	5.2	1.1	4.9	<.001	.17	1020.7
	Intercept	22.3	4.3	5.2	<.001		
	MLU3-w	6.5	1.3	4.9	<.001	.17	1020.8
Setswana	Intercept	−18.0	6.7	−2.7	.009		
	Age	1.7	2.3	5.9	<.001	.22	
	Intercept	5.6	3.0	1.9	.063		
	MLU3-m	5.3	0.9	5.9	<.001	.21	1016.3
	Intercept	5.2	3.1	1.7	.093		
	MLU3-w	6.0	1.0	5.7	<.001	.21	1017.5
IsiXhosa	Intercept	−10.3	8.9	−1.1	.253		
	Age	2.1	0.3	5.5	<.001	.21	
	Intercept	16.1	2.7	5.9	<.001		
	MLU3-m	6.9	0.7	9.8	<.001	.46	966.5
	Intercept	11.1	3.9	2.9	.005		
	MLU3-w	13.9	1.8	7.7	<.001	.35	988.2
Xitsonga	Intercept	−3.6	11.9	−0.3	.763		
	Age	2.2	0.5	4.4	<.001	.16	
	Intercept	32.0	6.3	5.1	<.001		
	MLU3-m	4.6	1.6	2.8	.005	.07	930.7
	Intercept	28.8	6.9	4.2	<.001		
	MLU3-w	6.3	2.1	3.0	.003	.08	929.7
Four languages combined	Intercept	−11.8	5.3	−2.2	.026		
	Age	2.1	0.2	9.3	<.001	.16	
	Intercept	16.0	2.1	7.8	<.001		
	MLU3-m	6.3	0.5	11.4	<.001	.22	4037.7
	Intercept	17.2	2.3	7.3	<.001		
	MLU3-w	7.3	0.8	9.2	<.001	.16	4075.9

MLU3-w explained 56% of the variance in grammar and 21% in the vocabulary scale. As was the case for Sesotho, both MLU3 variables were better predictors of grammar than of vocabulary. AIC showed that MLU3-w is a slightly better model for grammar, with a lower AIC score compared to MLU3-m, though R^2 values for both MLU3-m and MLU3-w are relatively similar. For vocabulary, the AIC indicated that MLU3-m is the better model, although the R^2 values are low.

IsiXhosa group statistics (Tables 3 and 4) showed that age explained 33% and 21% of the variance in grammar and vocabulary development, respectively. MLU3-m explained significant variance in grammar (60%) and vocabulary (46%) measures, whereas MLU3-w accounted for 48% of the variance in grammar and 35% of the variance in vocabulary. MLU3 variables were thus better predictors of grammar than of vocabulary in isiXhosa. AIC findings provide support for MLU3-m being a better model than MLU3-w for grammar and vocabulary development in isiXhosa.

Xitsonga group findings (Tables 3 and 4) indicated that age explained 25% and 16% of the variance in grammar and vocabulary scales, respectively. MLU3-m explained 27% of the variance in grammar and 7% in vocabulary and MLU3-w 25% of the variance in grammar and 8% in vocabulary. Both MLU3 variables were thus better predictors of grammar than of vocabulary. AIC findings indicated that MLU3-m is a better model for grammar than MLU3-w is, although the R^2 values were low. For vocabulary, AIC results showed that MLU3-w is the slightly better model; however, the R^2 values were very low.

To adjust for possible uneven age distributions within language groups, we reran the regression models, including age in all models in addition to either MLU3-m or MLU3-w. The age-adjusted associations of grammar and vocabulary scales with either MLU3-m or MLU3-w were similar to those reported in Tables 3 and 4. For Xitsonga, these adjusted models indicated that age was a stronger predictor of grammar and vocabulary scales than either MLU3-m or MLU3-w.

Discussion

Our first research question was whether MLU3 is a valid measure of early language development in the four Bantu languages concerned. We hypothesized that MLU3 would correlate positively with age, vocabulary and grammar measures, indicating its validity as a measure of development. We also hypothesized that MLU3 could be a more sensitive measure of grammar and vocabulary development than age.

There were moderate correlations between MLU3 measures and age for all languages except Xitsonga, which was on the border of weak to moderate. However, the Xitsonga results for morphemes ($r = .36$) and words ($r = .38$), while lower than the other three languages, are not unusual. Other studies have also reported significant but not strong correlations of .3 between age and MLU measures (Allen & Dench, 2015). There were also moderate to strong correlations between MLU3 variables and other indices of language growth – namely, grammar and vocabulary measures; although correlations with vocabulary were weak for Xitsonga. Overall, our results show that MLU3 scales are developmentally sensitive, correlating with age, vocabulary, and grammar scores, supporting the validity of MLU3 as a measure of language development in these languages. This result is in line with previous studies that demonstrate the validity of MLU3 as a child language development measure (e.g., Allen & Dench, 2015; Ezeizabarrena & Garcia Fernandez, 2018; Fenson et al., 1994; Heilmann et al., 2005; Jackson-Maldonado & Conboy, 2007).

As children's language levels in this age group tend to show considerable variation in relation to age, we considered whether MLU3 might be a more sensitive indicator of

language development than age. We found that MLU3 measures correlated more strongly with grammar than with age in all languages and accounted for a higher proportion of the variance than age in relation to grammar. However, there were mixed results for the correlation between age and vocabulary and between MLU3 measures and vocabulary: overall, there was not much difference except for (a) isiXhosa, where MLU3 accounted for a higher proportion of the variance than age in relation to vocabulary, and (b) Xitsonga, where age was a better predictor of vocabulary size than MLU3 was. Some studies based on natural language samples have shown that MLU-m correlates better than age with both grammatical and lexical indices of development (Dromi & Berman, 1982; Thordardottir & Weismer, 1998). Given our results, age and MLU3 need to be considered in tandem rather than relying on MLU3 as alternative measures, particularly in relation to vocabulary development.

Our second question was whether MLU3-m is a more sensitive measure of language development than MLU3-w. High correlations between MLU3-m and MLU3-w across the four languages suggest that both can be used as measures of language development. Spearman correlations comparing morphemes and words with grammar and vocabulary were also similar across languages. Where correlations are similar between morphemes and words, most studies of other languages – including agglutinative ones like Basque – have concluded that either words or morphemes can be used (Ezeizabarrena & Garcia Fernandez, 2018). However, our model comparison with AIC indicates that morphemes are either an equally or marginally more sensitive measure of grammatical and lexical development in our languages, except for isiXhosa, where morphemes are clearly more sensitive than words. Counting morphemes is therefore more suitable than orthographic word boundaries in agglutinative languages with conjunctive orthographies. These findings are in line with previous research that supports using a morpheme count as a more accurate tool for assessing language complexity in highly inflected languages (Allen & Dench, 2015; Ege, 2010; Meisel, 2011; Tomas & Dorofeeva, 2019).

Previous studies of MLU have almost invariably accepted the notion of a ‘word’ as a linguistic unit (with the exception of Allen & Dench, 2015; Gouda *et al.*, 2020). ‘Word’ boundaries have been seen as consistent with orthographic boundaries – *i.e.*, spaces – although few studies have explicitly stated this assumption. This is not necessarily a problem in languages with standardised and consistent orthographies, and where word boundaries closely align with morpheme boundaries (*e.g.*, English, Dutch). However, this study demonstrates that counting ‘words’ to measure MLU3 may not be as reliable a measure for isiXhosa as it is in Sesotho, Setswana and Xitsonga. We suggest that the word as unit of measurement is theoretically not universally a valid linguistic concept for the analysis of morphosyntactic complexity.

Some studies suggest that MLU is a general measure of early language development based on high correlations between MLU, on the one hand, and grammar and vocabulary scales, on the other (Dethorne *et al.*, 2005; Ezeizabarrena & Garcia Fernandez, 2018). However, other studies have found that MLU correlates better with grammatical development, suggesting it is a measure of specifically morphosyntactic development. This is the case for morphologically rich polysynthetic Inuktitut (Allen & Dench, 2015) and synthetic Icelandic (Thordardottir & Weismer, 1998). Our results support the latter position, with MLU3 measures accounting for a much higher percentage of the variability in grammar than in vocabulary for our agglutinative languages that are also morphologically rich and highly inflected. These differences between studies suggest that MLU may be measuring different kinds of linguistic knowledge in typologically and structurally different languages.

Conclusion

This paper examined the suitability of MLU3-m and MLU3-w as language development measures in four Southern Bantu languages (isiXhosa, Sesotho, Setswana, and Xitsonga) of the Bantu language family (Atlantic-Congo), an understudied group of languages. MLU3 proved to be a sensitive measure of expressive language development in conjunction with age. Although strong correlations were observed between MLU3-m and MLU3-w, MLU3-m was a more reliable measure than MLU3-w across the four languages. MLU3-m appears to be a useful indicator of children's linguistic abilities in this group of languages that are distinct from languages in previous studies of MLU. This study highlights the possible pitfalls of linguistic assumptions about 'words' based on orthographic conventions, particularly for agglutinative languages with conjunctive and inconsistent orthographies.

In terms of clinical implications, MLU counts may be useful for clinicians to obtain a general idea of a child's linguistic development, particularly where there is an absence of standardised assessment instruments. Our findings suggest that MLU3 (which is based on the three longest recent utterances) could provide an alternative measure to time-consuming MLU calculations (based on spontaneous language samples of 50 or more utterances) that clinicians perform as part of language sample analysis to diagnose language disorder or delay. In clinical contexts where a number of morphologically complex languages with different orthographic conventions are spoken, MLU3-m should be used rather than MLU3-w, especially if one wants to establish comparable cross-linguistic norms for children and compare development across languages in countries like South Africa. As our results have shown, MLU3-w is likely to underestimate children's abilities in agglutinative languages with conjunctive orthographies. In the African context, orthographies are inconsistent between related languages and even within one language (see Tucker, 1949, for a discussion). Any clinical application of MLU or MLU3 is necessarily mediated through the language's orthography, as clinicians need to transcribe the utterances before calculating the MLU or MLU3. Inconsistencies in relation to orthography mean that clinicians may need to rely on a more consistent unit of analysis than the word.

Although utterance length measured in morphemes appears to be the better indicator of children's linguistic development, identifying and operationalizing morpheme counts is not always easy in research and clinical settings (Arlman-Rupp et al., 1976). Several studies point out challenges with identifying morphemes and the applicability of rules for the identification of morphemes across different languages. Some argue that syllables are a more reliable and theoretically justifiable measure than morphemes (Arlman-Rupp et al., 1976; Tomas & Dorofeeva, 2019). The alternative of using syllable counts as a measure of language development (MLU-s) still needs to be tested on our data. Testing the sensitivity and specificity of MLU3-m and MLU3-s in differentiating between child speakers of Bantu languages with typical language development and those with delayed or disordered language development will also be necessary. In future studies, syllables should be explored for languages with similar morphological richness, particularly as MLU-m correlates well with MLU-s in Russian (Tomas & Dorofeeva, 2019), and Mean Length of Words – counted in morphemes and syllables – in Inuktitut (Allen & Dench, 2015). Nevertheless, identifying and counting syllables consistently may be as difficult to apply as counting morphemes or words in research and clinical settings (Parker & Brorson, 2005) although Tomas and Dorofeeva (2019) suggest that automated parsing is possible. Both researchers and clinicians may also need to consider what syllables measure in terms of language development and whether they inflate MLU scores as Hickey (1991) has suggested.

As this study is the first attempt to apply MLU as a measurement of development in Southern Bantu languages, further research should include larger samples of these languages and more languages that are typologically similar. A larger sample for Xitsonga, comparisons with similar samples of rural children speaking other Bantu languages as well as careful examination of the development sequence of morphosyntactic features of Xitsonga in comparison to other Southern Bantu languages may shed light on the slightly different trends in relation to Xitsonga in our study. Comparing our findings on isiXhosa, where MLU3-m was a better measurement than MLU3-w, with isiZulu (a closely related agglutinative with conjunctive orthography) may confirm our conclusions. Along with comparisons of more languages, we should also explore alternative measurements such as MLU in syllables, or Mean Length of Words in morphemes as applied by Allen and Dench (2015). In addition, more detailed investigation of the developmental trajectory of semantic and morphosyntactic elements in the acquisition of these languages may guide us on how to interpret MLU counts as measures of developing language complexity. This kind of investigation (a) will help us to understand how MLU relates to linguistic knowledge and the relationship between lexical and morphosyntactic development (Dethorne *et al.*, 2005), given that the languages in our study are typologically and structurally distinct from languages in other studies of MLU; and (b) may aid us in finding other, more sensitive measures such as the average number of grammatical forms in a sample as applied by Tomas and Dorofeeva (2019) to Russian or the number of different words and tense accuracy composite applied by Dethorne *et al.* (2005) to English.

In African contexts, there has been little research on the development of indigenous languages, especially large-scale studies that provide indices of typical development. Consequently, there is a dearth of validated tools for clinicians with which to measure language development. This study is a first step towards establishing the validity of MLU as a useful tool for clinicians to measure early language development in Bantu languages. Research efforts in the pursuit of understanding language-specific acquisition patterns in Bantu languages are of great importance. To date, there are few language assessment tools developed and standardized as accurate diagnostic measures or indicators of either typical or atypical language development in South Africa and most other African countries. The scarcity of such tools places a serious burden on clinicians. The findings of this study, while having theoretical implications (such as encouraging debate on the operationalisation of 'word') are also a step, albeit small, towards addressing the need for linguistically appropriate child language assessment tools in African languages.

Acknowledgements. This study was made possible with the support of the South African Centre for Digital Language Resources (SADiLaR). SADiLaR is a research infrastructure established by the Department of Science and Innovation of the South African government as part of the South African Research Infrastructure Roadmap (SARIR). Additional funding came from The National Research Foundation of South Africa (HSD170602236563). Preliminary work for this research was supported by The British Academy Newton Fund (NG160093) and the National Research Foundation of South Africa/Swedish Foundation for International Cooperation in Research and Higher Education (NRF/STINT160918188417). We thank Dr R Berghoff and Prof. E Bylund at Stellenbosch University, Dr M Ezeizabarrena at University of the Basque Country, and particularly the anonymous reviewers for their helpful comments. We are also grateful to Professor Eric Atmore of the Centre for Early Childhood Development and other early childhood development organisations for their assistance in the field. We thank the MB-CDI advisory board for granting us permission to adapt the MB-CDI for these languages. Any opinion, findings, conclusions or recommendations expressed in this material are those of the authors.

Competing interests. The author(s) declare none.

References

- Alcock, K., Rimba, K., Holding, P., Kitsao-Wekulo, P., Abubakar, A., & Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants: Communicative Development Inventory (CDI) adaptation for two Kenyan languages. *Journal of Child Language*, *42*, 763–85.
- Allen, S. E. M., & Dench, C. (2015). Calculating mean length of utterance for eastern Canadian Inuktitut. *First Language*, *35*(4–5), 377–406. <https://doi.org/10.1177/0142723715596648>
- Arlman-Rupp, A. J. L., van Niekerk de Haan, D., & van de Sandt-Koenderman, M. (1976). Brown's early stages: some evidence from Dutch. *Journal of Child Language*, *3*(2), 267–274. <https://doi.org/10.1017/S0305000900001483>
- Brown, E. K., & Miller, J. E. (2013). *The Cambridge dictionary of Linguistics*. Cambridge University Press.
- Brown, R. (1973). *A first language: the early stages*. Harvard University Press.
- Cheung, H. (1998). The application of MLU in Chinese. *Journal of the Speech-Language-Hearing Association of the Republic of China*, *13*, 36–48. <https://doi.org/10.6143/JSLHAT.1998.06.04>
- Dancey, C., & Reidy, J. (2007). *Statistics without maths for Psychology: using SPSS for Windows* (4th ed.). Pearson.
- Demuth, K. (1992). Acquisition of Sesotho. In D. Slobin (Ed.), *The cross-linguistic study of language acquisition* (Vol. 3, pp. 557–638). Lawrence Erlbaum.
- Demuth, K. (2003). The acquisition of Bantu languages. In D. Nurse & G. Philippson (Eds.), *The Bantu languages* (pp. 209–222). Surrey, England: Curzon Press.
- Dethorne, L. S., Johnson, B. W., & Loeb, J. W. (2005). A closer look at MLU: What does it really measure? *Clinical Linguistics & Phonetics*, *19*(8), 635–648. <https://doi.org/10.1080/02699200410001716165>
- Dromi, E., & Berman, R. A. (1982). A morphemic measure of early language development: data from Modern Hebrew. *Journal of Child Language*, *9*(2), 403–424. <https://doi.org/10.1017/S0305000900004785>
- Ege, P. (2010). Mean length of utterance as a tool for morphological assessment in Turkish children. In S. Topbaş & M. Yavaş (Eds.), *Communication disorders in Turkish* (Vol. 4, pp. 105–118). Multilingual Matters.
- Eisenberg, S. L., Fersko, T. M., & Lundgren, C. (2001). The use of MLU for identifying language impairment in preschool children. *American Journal of Speech-Language Pathology*, *10*(4), 323–342. [https://doi.org/10.1044/1058-0360\(2001\)028](https://doi.org/10.1044/1058-0360(2001)028)
- Ezeizabarrena, M.-J., & Garcia Fernandez, I. (2018). Length of utterance, in morphemes or in words? MLU3-w, a reliable measure of language development in Early Basque. *Frontiers in Psychology*, *8*, 2265. <https://doi.org/10.3389/fpsyg.2017.02265>
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), 1–185. <https://doi.org/10.2307/1166093>
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D. J., Bates, E., & Hartung, J. (1993). *MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore, MD: Brookes Publishing Co.
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: the Wordbank Project*. MIT Press.
- Gavin, W. J., & Giles, L. (1996). Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech, Language, and Hearing Research*, *39*(6), 1258–1262. <https://doi.org/10.1044/jshr.3906.1258>
- Gouda, G., Kumar, S., Sarkar, M., Rashmi, R., Chatterjee, N., & Pani, S. (2020). Mean length of utterance of word and morpheme for 3–5 years Bengali speaking children. *Jadavpur Journal of Languages and Linguistics*, *4*, 1–13.
- Heilmann, J., Weismer, S. E., Evans, J., & Hollar, C. (2005). Utility of the MacArthur-Bates Communicative Development Inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech-Language Pathology*, *14*(1), 40–51. [https://doi.org/10.1044/1058-0360\(2005\)006](https://doi.org/10.1044/1058-0360(2005)006)
- Hickey, T. (1991). Mean length of utterance and the acquisition of Irish. *Journal of Child Language*, *18*(3), 553–569. <https://doi.org/10.1017/S0305000900011247>
- Jackson-Maldonado, D., & Conboy, B. (2007). Utterance length measures for Spanish-speaking toddlers: the morpheme versus word issue revisited. In J. Centeno, R. Anderson, & L. Obler (Eds.), *Communication disorders in Spanish Speakers: Theoretical, research and clinical aspects* (pp. 142–155). Multilingual Matters.

- Johnston, J. R.** (2001). An alternate MLU calculation: magnitude and variability of effects. *Journal of Speech, Language, and Hearing Research*, *44*, 156–164.
- Kidd, E., & Garcia, R.** (2022). How diverse is child language acquisition research? *First Language*, *42*(6), 1–33. <https://doi.org/10.1177/01427237211066405>
- Klee, T., Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Gavin, W. J.** (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1396–1410. [https://doi.org/10.1044/1092-4388\(2004/104\)](https://doi.org/10.1044/1092-4388(2004/104))
- Lee, S. J., & Hlungwani, C.** (2017). Distribution of conjunctive and disjunctive forms in Xitsonga. *Stellenbosch Papers in Linguistics Plus*, *52*, 157–173. <https://doi.org/10.5842/52-0-712>
- Louwrens, L. J., & Poulos, G.** (2006). The status of the word in selected conventional writing systems: the case of disjunctive writing. *Southern African Linguistics and Applied Language Studies*, *24*(3), 389–401. <https://doi.org/10.2989/16073610609486427>
- Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S.** (2004). The language-specific nature of grammatical development: evidence from bilingual language learners. *Developmental Science*, *7*(2), 212–224. <https://doi.org/10.1111/j.1467-7687.2004.00340.x>
- Meisel, J. M.** (2011). Bilingual language acquisition and theories of diachronic change: bilingualism as cause and effect of grammatical change. *Bilingualism: Language and Cognition*, *14*(2), 121–145. <https://doi.org/10.1017/S1366728910000143>
- Miller, J. F., & Chapman, R. S.** (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech, Language, and Hearing Research*, *24*(2), 154–161. <https://doi.org/10.1044/jshr.2402.154>
- Nice, M. M.** (1925). Length of sentences as a criterion of a child's progress in speech. *Journal of Educational Psychology*, *16*(6), 370–379. <https://doi.org/10.1037/h0073259>
- Nurse, D., & Philippson, G.** (Eds.). (2006). *The Bantu languages*. Routledge.
- Oosthuizen, H., & Southwood, F.** (2009). Methodological issues in the calculation of mean length of utterance (MLU). *South African Journal of Communication Disorders*, *56*(1), 76–87. <https://doi.org/10.4102/sajcd.v56i1.194>
- Parker, M. D., & Brorson, K.** (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, *25*(3), 365–376. <https://doi.org/10.1177/0142723705059114>
- Qualtrics.** (2019–2020). Qualtrics [Online software]. <https://www.qualtrics.com>
- Rice, M. L., Redmond, S. M., & Hoffman, L.** (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research*, *49*(4), 793–808. [https://doi.org/10.1044/1092-4388\(2006/056\)](https://doi.org/10.1044/1092-4388(2006/056))
- Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M.** (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research*, *53*(2), 333–349. [https://doi.org/10.1044/1092-4388\(2009/08-0183\)](https://doi.org/10.1044/1092-4388(2009/08-0183))
- Scarborough, H. S., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., & Sudhalter, V.** (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics*, *12*(1), 23–46. <https://doi.org/10.1017/S01427164000936X>
- Southwood, F., & Russell, A. F.** (2004). Comparison of conversation, freeplay, and story generation as methods of language sample elicitation. *Journal of Speech, Language, and Hearing Research*, *47*(2), 366–376. [https://doi.org/10.1044/1092-4388\(2004/030\)](https://doi.org/10.1044/1092-4388(2004/030))
- Southwood, F., White, M. J., Brookes, H., Pascoe, M., Ndhambi, M., Yalala, S., Mahura, O., Mössmer, M., Oosthuizen, H., Brink, N., & Alcock, K.** (2021). Sociocultural factors affecting vocabulary development in young South African children. *Frontiers in Psychology*, *12*, 642315. <https://doi.org/10.3389/fpsyg.2021.642315>
- Taljad, E., & Bosch, S.** (2006). A comparison of approaches to word class tagging: disjunctively vs. conjunctively written Bantu languages. *Nordic Journal of African Studies*, *15*(4), 428–442. <https://doi.org/10.53228/njas.v15i4.37>
- Thordardottir, E.** (2005). Early lexical and syntactic development in Quebec French and English: implications for cross-linguistic and bilingual assessment. *International Journal of Language and Communication Disorders*, *40*(3), 243–278. <https://doi.org/10.1080/13682820410001729655>

- Thordardottir, E., & Weismer, S. E.** (1998). Mean length of utterance and other language sample measures in early Icelandic. *First Language*, **18**(52), 1–32. <https://doi.org/10.1177/014272379801805201>
- Tolentino, T. J.** (2022). *Impact of language sampling context on language productivity and complexity* (Master's thesis). <https://digitalcommons.usf.edu/etd/9822>
- Tomas, E., & Dorofeeva, S.** (2019). Mean length of utterance and other quantitative measures of spontaneous speech in Russian-speaking children. *Journal of Speech, Language, and Hearing Research*, **62**(12), 4483–4496. https://doi.org/10.1044/2019_JSLHR-L-18-0339
- Tucker, A. N.** (1949). Sotho-Nguni orthography and tone-marking. *Bulletin of the School of Oriental and African Studies, University of London*, **13**(1), 200–224. <http://www.jstor.org/stable/609073>
- Vasilyeva, M., Waterfall, H., & Huttenlocher, J.** (2008). Emergence of syntax: commonalities and differences across children. *Developmental Science*, **11**(1), 84–97. <https://doi.org/10.1111/j.1467-7687.2007.00656.x>
- Vogt, P., Mastin, J. D., & Aussems, S.** (2015). Early vocabulary development in rural and urban Mozambique. *Child Development Research*, **2015** (1), 1–15. <https://doi.org/10.1155/2015/189195>
- Voniati, L.** (2016). Mean length of utterance in Cypriot Greek-speaking children. *Journal of Greek Linguistics*, **16** (1), 117–140. <https://doi.org/10.1163/15699846-01601002>
- Wieczorek, R.** (2010). Using MLU to study early language development in English. *Psychology of Language and Communication*, **14** (2), 59–69. <https://doi.org/10.2478/v10057-010-0010-9>
- Wu, S.-Y.** (2020). Mean length of utterance among Mandarin-speaking children with and without DLD. *Child Language Teaching and Therapy*, **36** (6), 165–179. <https://doi.org/10.1177/0265659020945366>

Cite this article: Brookes, H., Makaure, P., Yalala, S., Danvers, H., Mössmer, M., Little, F., Ndhambi, M., Southwood, F., Ludidi, B., & The South African CDI Team (2025). Mean Length of Utterance: A study of early language development in four Southern Bantu languages. *Journal of Child Language* **52**, 244–269, <https://doi.org/10.1017/S0305000923000685>