Handout, Communication Matters, September 2025

MEASURING AAC USER LINGUISTIC COMPETENCE

A Novel Approach

David Niemeijer, PhD, CEO and Founder, AssistiveWare
Erin Sheldon, M Ed, Director of Learning Design, AssistiveWare
Alyssa Hillary Zisk, PhD, AAC Research Team Lead, AssistiveWare

Summary

Introduction

This session explores how to measure linguistic competence for language produced with an AAC system. Light (1989, p. 139) defined linguistic competence as “an adequate level of mastery of the linguistic code, including phonological, morphological, syntactic, and semantic aspects.” We will explore new ways of data collection as well as new ways to measure linguistic competence.

Collecting AAC language use data is important to support AAC user language growth, measure progress towards IEP goals, guide communication partners, and to study AAC language development.

At present, the most common method to collect language use data is using Language Activity Monitoring (LAM) or Automated Data Logging (ADL). This involves the automatic recording of every letter, word, and sentence spoken with the AAC system. Every entry is captured with a timestamp and stored in an exportable log file. While these data are undeniably valuable for the purposes mentioned above, they also represent an invasion of the AAC user’s privacy. Many emergent AAC users are unlikely to be able to truly consent to or reject such data collection.

We will discuss the following questions:

1. Can we collect data in a way that protects AAC user privacy?

2. Can we use existing measures of linguistic competence, such as MLU, or do we need to develop new ones?

Can we collect data in a way that protects AAC user privacy?

The short answer is yes. A number of techniques allow us to guarantee AAC user privacy:

- **Analyze all data on device** rather than exporting a log file so that no one can see what was said when.

- **Aggregate data** on-device over at least 7 days so that any patterns cannot be attributed to a specific day or time, thereby potentially exposing private information

- **Avoid accessing data on fringe word frequency.** Fringe words are, by definition, more personal than core. Instead, focus on a number of on-device computed statistics.

This approach has enabled us to collect and analyze data from thousands of users of symbol-supported AAC and close to a thousand users of text-based AAC. Data was collected and aggregated over a 28-day period between 16 October and 20 November 2024.

Can we use existing measures of linguistic competence, such as MLU, or do we need to develop new ones?

Based on earlier research, we were able to create a number of made-up case studies, ranging from an emergent user of symbol-supported AAC to a fluent user of text-based AAC. By considering typical device configuration (such as vocabulary choice and grid size) and associated communication patterns, we were able to establish a number of challenges when measuring the linguistic competence of AAC users:

- Vocabulary design (including grid size) impacts language production
- Buttons such as “is,” “finished,” “all done,” and “my turn” corrupt language samples by introducing morphology long before AAC users can generate such morphology
- Preprogrammed messages and sentence starters may give a false impression of linguistic competence
- MLU, or Mean Length of Utterance, is a problematic measurement for AAC because it is affected by operational competence and preprogrammed messages.

MLU, or mean length of utterance, where either the number of words or morphemes per utterance is counted, is a measure of the syntactic complexity of language. The underlying assumption is that longer sentences mean more complex sentence structure. In turn, more complex sentence structure reflects greater semantic complexity and greater morphological complexity. Over the last half-century, MLU has become the go-to measure for investigating emerging language abilities for children with and without language impairments. MLU is less suitable for measuring language produced with AAC systems, because it is not just linguistic but interacts with other competences:

- **Operational:** MLU requires clear segmentation of utterances. Building a message, speaking it, and clearing it before the next utterance is operational, not linguistic. Additionally, not every AAC user wants to actively manage the message window when words can be spoken on touch.
- **Strategic:** To speed up communication, AAC users may use incomplete syntax or conversely rely on preprogrammed sentence starters and phrases.

- **Social:** Managing the message window may interfere with social interaction and is often not necessary, as words can be spoken without clearing the window.

As a consequence of the above, MLU is disproportionately high for more emergent users of symbol-supported AAC. Also, because most AAC users speak fewer than two dozen words per day, it is hard to reliably estimate MLU based on data collected in a natural setting.

We concluded that we needed to **develop new methods**. We explored a range of measures to capture the semantic, syntactic, morphological, and phonological dimensions of linguistic competence.

It was found that **density measures are problematic** because they often plateau or decline as language proficiency and communication functions increase. Without exception, the **diversity measures scored better** (higher correlations, more significant, more aligned with the expected trends in relation to reference indicators of language level).

One particular challenge with linguistic diversity measures is that they tend to be **sensitive to sample length**. The likelihood that you use a new word or use a new grammatical form decreases as you write or say more. As natural samples vary greatly in length, we reviewed the literature and decided to use a moving average approach for all our diversity measures, based on the approach Covington and McFall (2010) developed for measuring lexical diversity.

Four measures, one for each dimension of linguistic competence, stood out.

For **semantic competence** we found that lexical diversity measured with a moving average window of 30 words (MATTR-30) is the best measure. The median lexical diversity showed highly significant correlations (R ranges from 0.22 to 0.32) with reference values and followed the expected patterns. It is particularly robust because it can be applied for the most emergent to the most proficient AAC users and because it is based on an open set of words that grows over time.

For **syntactic competence** we found that the combined preposition and conjunction diversity measured with a moving average window of 30 words (MA-UPC-TWR-30) is the best measure. Closed set words kinds such as prepositions and especially conjunctions can be used as indicators of syntactic complexity. Prepositions can be used to track very early language development. Conjunctions are weaker in the early phases but stronger overall. The median preposition and conjunction diversity showed highly significant correlations (R ranges from 0.06 to 0.37) with reference values and followed the expected patterns.

For **morphological competence** we found that the diversity of morphological forms measured with a moving average window of 30 words (MA-UMORPH-TLWR-30) made the most sense on theoretical grounds. However, for analyzing the language of symbol-supported AAC use, we found that all the morphological diversity measures we explored were heavily affected by the presence of morphological forms in standard vocabulary buttons (such as “all done,” “finished,” “is,” “my turn,” etc.). We did find highly significant differences between symbol-supported and text-based AAC.

For **phonological competence** we found that the proportion of unique typed words correctly spelled is the most useful measure on theoretical grounds. By focusing on unique words, the measure is not sensitive to repeated words or repeated misspellings. Even though it appears to be the best measure to assess the phonological competence of AAC users, based on language produced, the challenge is that most AAC users relying on symbol-supported AAC type very little, while proficient text-based AAC users may combine a high language proficiency (in terms of vocabulary and sentence structure) with poor spelling skills, typos related to motor skills, intentional misspellings for pronunciation control, or heavy reliance on word prediction. This means that phonological competence is a poor indicator of overall linguistic competence.

Lexical diversity (calculated with the moving average method) was found to be the strongest and most robust overall measure of linguistic competence. It has the strongest performance in relation to the reference measures. It works for the most emergent to most proficient AAC users and it correlates well (r above 0.55) with other relevant indicators of linguistic competence (syntactic and morphological diversity). It is a very robust measure because it is insensitive to operational competence (unlike MLU) and works with very small language samples (unlike MLU)

AAC design and setup can affect both linguistic competence and the measurement thereof. AAC systems with small grid sizes can slow down the development of linguistic competence and can also lead to an underestimation of linguistic competence. AAC systems with automatic grammaticalization or navigation or many phrases-based buttons can slow down the development of linguistic competence and can also lead to an overestimation of linguistic competence.

It is recommended that clinicians working with MLU exercise caution when analyzing language samples with MLU because of the strategic, social and operational artifacts. It is also recommended that as a field we shift towards measures that are clinically sound yet protect AAC user privacy.

References

Binger, C., Kent-Walsh, J., Harrington, N., & Hollerbach, Q.C. (2020). Tracking Early Sentence-Building Progress in Graphics Symbol Communications. *Language, Speech, and Hearing Services in Schools*, 51(2), 317–328.

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.

Light, Janice (1989). Toward a definition of communicative competence for individuals using augmentative and alternative communication systems. *Augmentative and Alternative Communication*, 5(2), 137-144.

Nippold, M. A., Cramond, P. M., & Hayward-Mayhew, C. (2014). Spoken language production in adults: examining age-related differences in syntactic complexity. *Clinical Linguistics & Phonetics,* 28(3), 195-207.

Potratz, J. R., Gildersleeve-Neumann, C., & Redford, M. A. (2022). Measurement Properties of Mean Length of Utterance in School-Age Children. *Language, Speech, and Hearing Services in Schools*, 53(4), 1088-1100.

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.