

## **UC Merced**

### **Proceedings of the Annual Meeting of the Cognitive Science Society**

#### **Title**

Measuring Children's Early Vocabulary in Low-Resource Languages Using a Swadesh-style Word List

#### **Permalink**

<https://escholarship.org/uc/item/45d8890w>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

#### **Authors**

Tan, Alvin Wei Ming  
Kachergis, George  
Marchman, Virginia A  
[et al.](#)

#### **Publication Date**

2023

Peer reviewed

# Measuring Children’s Early Vocabulary in Low-Resource Languages Using a Swadesh-style Word List

**Alvin Tan**

Stanford University, Stanford, California, United States

**George Kachergis**

Stanford University, Stanford, California, United States

**Virginia Marchman**

Stanford University, Stanford, California, United States

**Philip Dale**

University of New Mexico, Albuquerque, New Mexico, United States

**Michael Frank**

Stanford University, Stanford, California, United States

## Abstract

Early language skill is predictive of later life outcomes, and is thus of great interest to developmental psychologists and clinicians. The Communicative Development Inventories (CDIs), parent-reported inventories of early-learned vocabulary items, have proven to be valid and reliable instruments for measuring children’s early language skill. CDIs have been painstakingly adapted to dozens of languages, and cross-linguistic comparisons thus far show both consistency and variability in language acquisition trajectories. However, thousands of languages do not yet have CDIs, posing a significant barrier to increasing the diversity of languages that are studied. Here, we propose a method for selecting candidate words to include on new CDIs, leveraging analysis of psychometric properties of translation-equivalent concepts that are frequently included on existing CDIs. Leveraging 26 datasets from existing CDIs, we propose a list of 229 concepts that have low variability in their cross-linguistic learning difficulty. This pool of common concepts—analogueous to the “Swadesh” lists used in glottochronology—can be used as a starting point for future CDI adaptations. We test how well the proposed list generalizes to data from 8 additional languages.