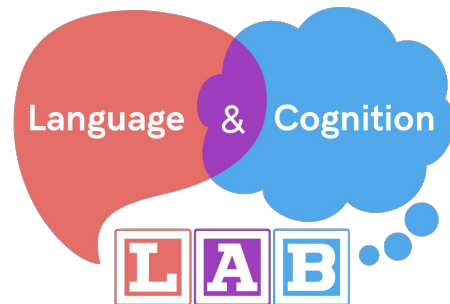
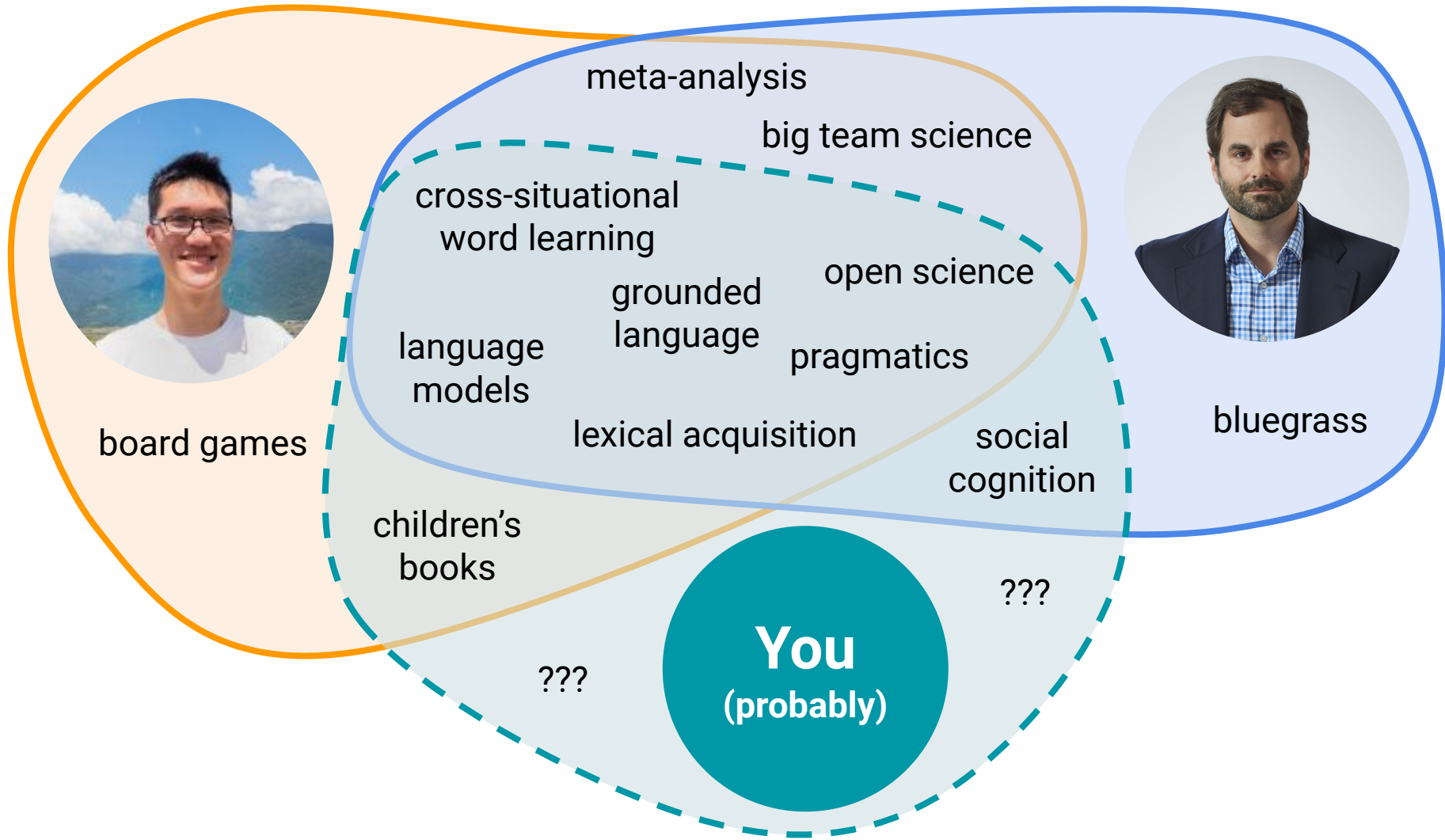


Using Wordbank for Language Development Research

Alvin Tan & Michael C. Frank





Roadmap

- Introduction
 - CDIs
 - Wordbank
 - What can we learn?
- Practical session
 - `tidyverse` refresher
 - `wordbankr`
 - Case study: Vocabulary and grammar
- What have we learnt? (Research talk – Mike)

Pre-workshop survey



Measuring language development



Direct production measurement
([LENA Team, 2020](#))

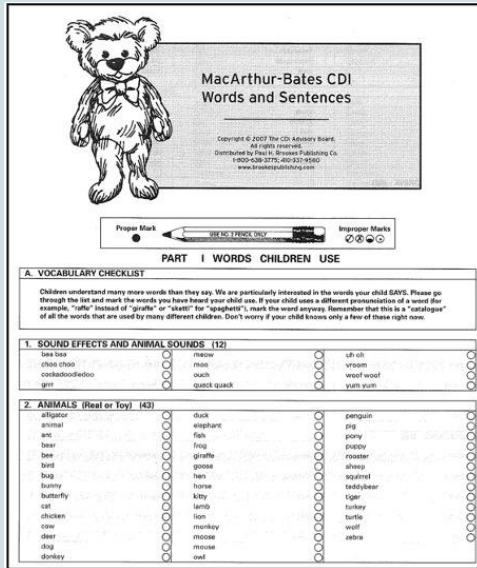


Picture-based vocabulary tests
([Dunn, 2018](#); [Brownell, 2010](#))



Looking while listening
([Fernald et al., 2008](#); [Byers-Heinlein, 2018](#))






Measuring language development



The image shows the MacArthur-Bates Communication Development Inventory (CDI) form. It features a teddy bear illustration and a section for marking words. The form is titled 'MacArthur-Bates CDI Words and Sentences' and includes copyright information. Below the title, there is a section for 'PART I WORDS CHILDREN USE' and a 'VOCABULARY CHECKLIST' with two main categories: '1. SOUND EFFECTS AND ANIMAL SOUNDS (12)' and '2. ANIMALS (Real or Toy) (43)'. Each category lists various words and sounds with corresponding checkboxes for marking.

MacArthur-Bates CDI Words and Sentences

Copyright © 2007 The CDI Inventory Board. All rights reserved.
Distributed by The In. Brown Publishing Co.
1-800-438-2777, 402-333-9580
www.mcdiinventory.com

Proper Mark:  **USE PENCIL ONLY**  Improper Marks:   

PART I WORDS CHILDREN USE

A. VOCABULARY CHECKLIST

Children understand many more words than they say. We are particularly interested in the words your child **SAYS**. Please go through the list and mark the words you have heard your child use. If your child uses a different pronunciation of a word (for example, "saffe" instead of "graffe" or "skate" for "skateboard"), mark the word anyway. Remember that this is a "catalogue" of all the words that are used by many different children. Don't worry if your child knows only a few of these right now.

1. SOUND EFFECTS AND ANIMAL SOUNDS (12)

<input type="checkbox"/> bee bee	<input type="checkbox"/> meow	<input type="checkbox"/> uh oh
<input type="checkbox"/> chick chick	<input type="checkbox"/> moo	<input type="checkbox"/> vroom
<input type="checkbox"/> coo-coo-doo-doo	<input type="checkbox"/> oach	<input type="checkbox"/> woof woof
<input type="checkbox"/> grrr	<input type="checkbox"/> quack quack	<input type="checkbox"/> yuh yuh

2. ANIMALS (Real or Toy) (43)

<input type="checkbox"/> alligator	<input type="checkbox"/> duck	<input type="checkbox"/> penguin
<input type="checkbox"/> animal	<input type="checkbox"/> elephant	<input type="checkbox"/> pig
<input type="checkbox"/> ant	<input type="checkbox"/> fish	<input type="checkbox"/> pony
<input type="checkbox"/> bear	<input type="checkbox"/> frog	<input type="checkbox"/> puppy
<input type="checkbox"/> bee	<input type="checkbox"/> giraffe	<input type="checkbox"/> master
<input type="checkbox"/> bird	<input type="checkbox"/> goose	<input type="checkbox"/> sheep
<input type="checkbox"/> bug	<input type="checkbox"/> hen	<input type="checkbox"/> squirrel
<input type="checkbox"/> bunny	<input type="checkbox"/> horse	<input type="checkbox"/> teddy bear
<input type="checkbox"/> butterfly	<input type="checkbox"/> kitty	<input type="checkbox"/> tiger
<input type="checkbox"/> cat	<input type="checkbox"/> lamb	<input type="checkbox"/> turkey
<input type="checkbox"/> chicken	<input type="checkbox"/> lion	<input type="checkbox"/> turtle
<input type="checkbox"/> cow	<input type="checkbox"/> monkey	<input type="checkbox"/> wolf
<input type="checkbox"/> deer	<input type="checkbox"/> mouse	<input type="checkbox"/> zebra
<input type="checkbox"/> dog	<input type="checkbox"/> mouse	
<input type="checkbox"/> donkey	<input type="checkbox"/> owl	

Communicative Development Inventories (CDIs)

Caregiver-report checklists measuring early lexical comprehension and production (and gesture use, emerging morphology and syntax...)

(Dunn, 2018, Brownell, 2018)

CDIs

- Words & Gestures (8–18mo; receptive and expressive vocabulary)
- Words & Sentences (16–30mo; expressive vocabulary)

1. <i>Sound Effects and Animal Sounds</i> (12)				<i>English (American) WS</i>			
baa baa	<input type="radio"/>	grrr	<input type="radio"/>	ouch	<input type="radio"/>	vroom	<input type="radio"/>
choo choo	<input type="radio"/>	meow	<input type="radio"/>	quack quack	<input type="radio"/>	woof woof	<input type="radio"/>
cockadoodledoo	<input type="radio"/>	moo	<input type="radio"/>	uh oh	<input type="radio"/>	yum yum	<input type="radio"/>
2. <i>Animals (Real or Toy)</i> (43)							
alligator	<input type="radio"/>	cow	<input type="radio"/>	horse	<input type="radio"/>	puppy	<input type="radio"/>
animal	<input type="radio"/>	deer	<input type="radio"/>	kitty	<input type="radio"/>	rooster	<input type="radio"/>
ant	<input type="radio"/>	dog	<input type="radio"/>	lamb	<input type="radio"/>	sheep	<input type="radio"/>
bear	<input type="radio"/>	donkey	<input type="radio"/>	lion	<input type="radio"/>	squirrel	<input type="radio"/>
bee	<input type="radio"/>	duck	<input type="radio"/>	monkey	<input type="radio"/>	teddybear	<input type="radio"/>
bird	<input type="radio"/>	elephant	<input type="radio"/>	moose	<input type="radio"/>	tiger	<input type="radio"/>

Why CDIs?

- Comprehensive view of child's abilities
- Less susceptible to child temperament/mood
- Reliable and valid ([Jarůšková et al., 2023](#))
- Inexpensive and easy to administer
- Adaptations in >120 languages/dialects
- Online implementation in Web-CDI ([DeMayo et al., 2021](#))




2022




Wordbank

An open database of children's vocabulary development

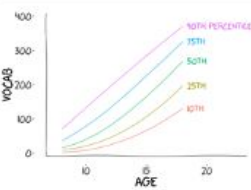


Wordbank

An open database of children's vocabulary development

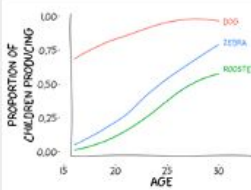


Wordbank contains data from 79,027 children and 93,258 CDI administrations, across 40 languages and 76 instruments:




Vocabulary Norms

Explore vocabulary size growth curves for various languages and demographic groups.




Item Trajectories

Explore trajectories of individual words, word categories, and grammar items.




Wordbank

An open database of children's vocabulary development



Wordbank contains data from 75,144 children CDI administrations, across 40 languages and 56 instruments:



2015

2019

Wordbank team



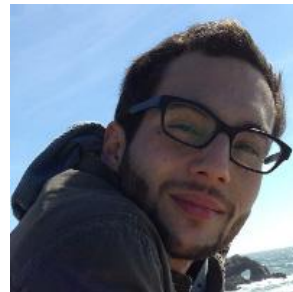
Mike Frank



Virginia Marchman



Mika Braginsky



Dan Yurovsky



Danielle Kellier



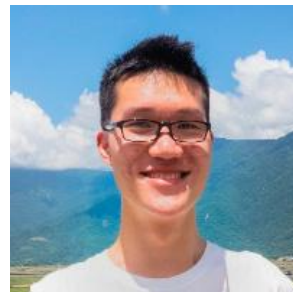
Benny deMayo



George Kachergis



Jess Mankewitz

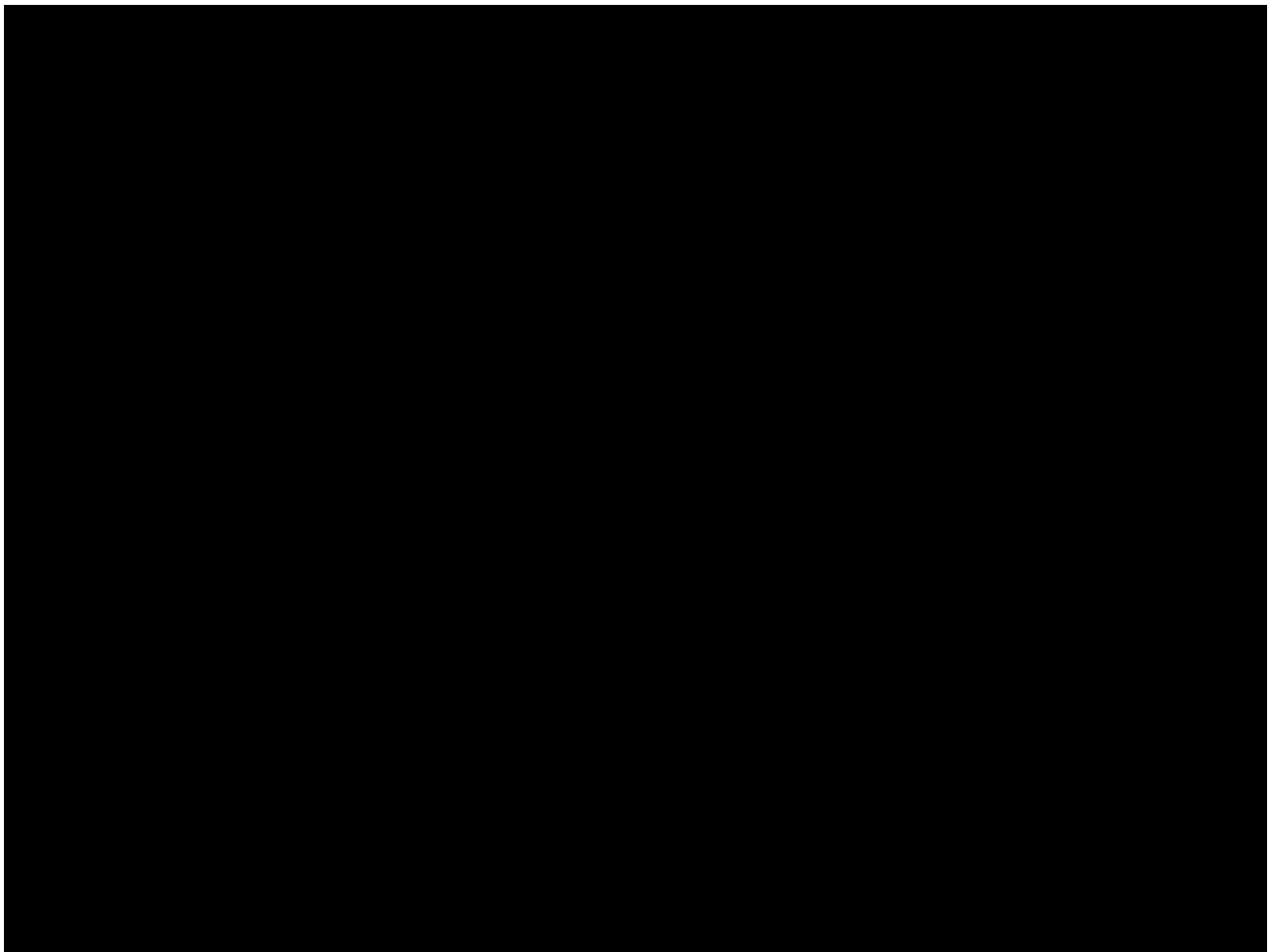


Alvin Tan

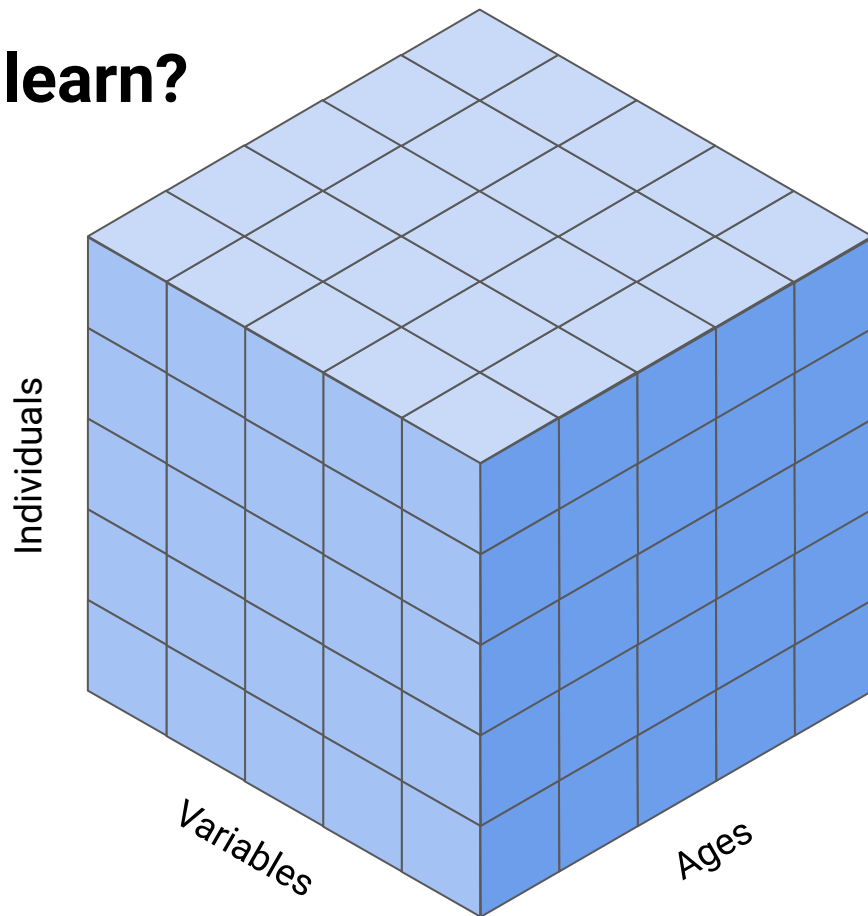


Henry Mehta

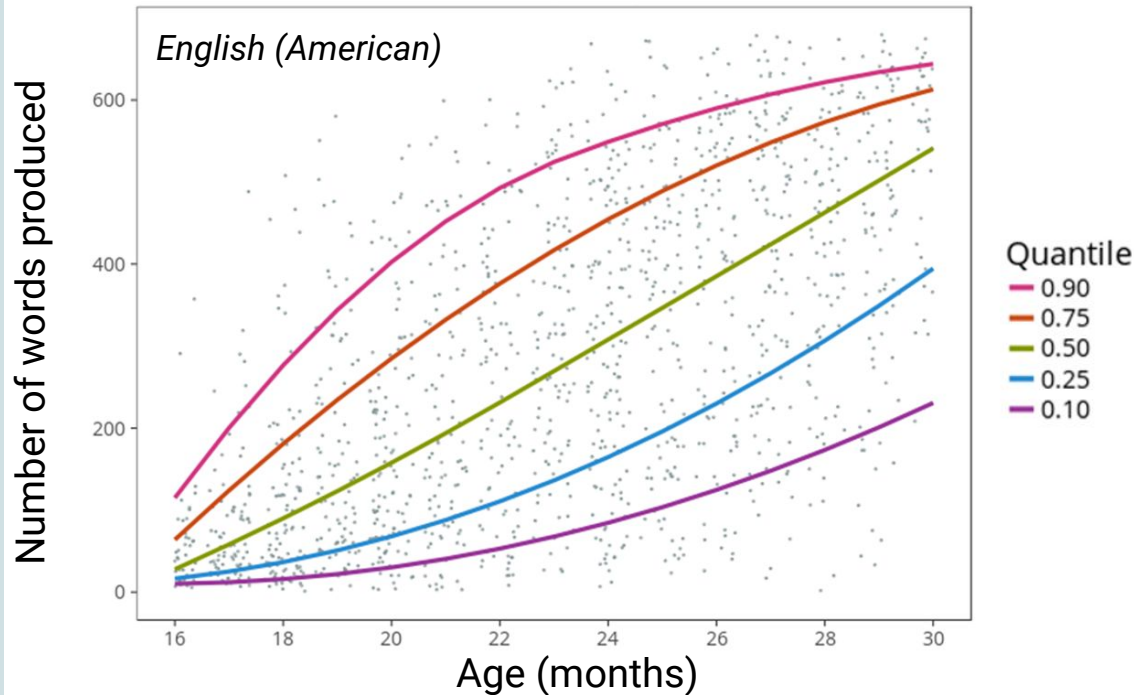
+ dozens of contributors!



What can we learn?



Large variability in children's vocabularies over development



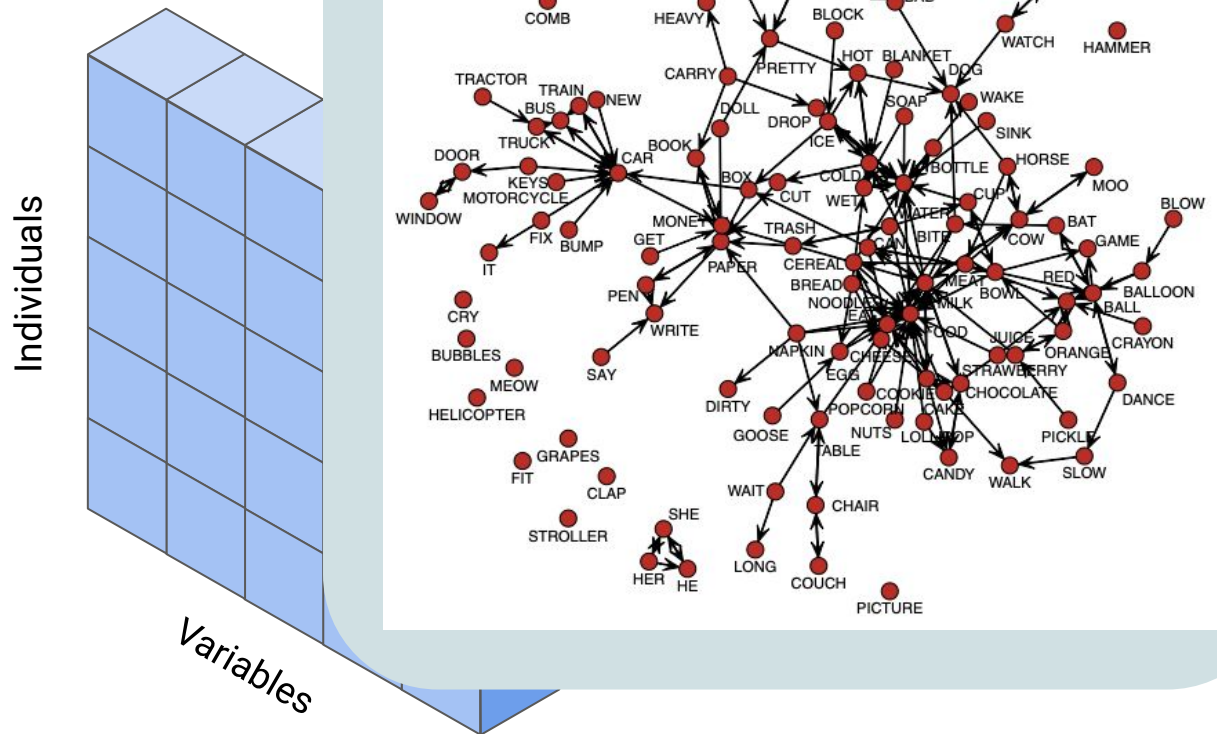
Individuals

Ages

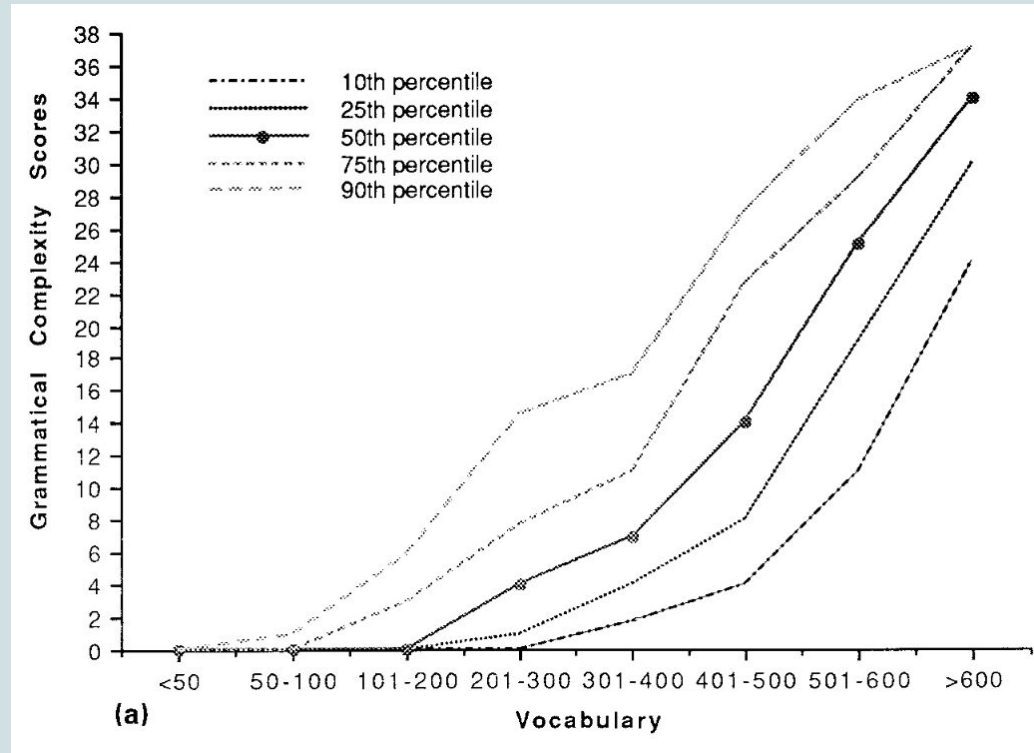
	Croatian	Danish	English (American)	Italian	Korean	Norwegian	Russian	Spanish (Mexican)	Swedish	Turkish
1	grandma	peekaboo	sister	peekaboo	child's name	daddy	grandma	milk	child's name	mommy
2	mommy	pattycake	brother	daddy	again	peekaboo	daddy	bye	mommy	daddy
3	daddy	child's name	daddy	mommy	money	mommy	grandpa	yum yum	daddy	yum yum
4	bye	daddy	mommy	baby food	aunt	no	yum yum	daddy	peekaboo	water
5	child's name	mommy	child's name	child's name	cracker	child's name	need	mommy	bye	outing
6	grandpa	no	bottle	hi	play	hi	can	water	bath	child's name
7	woof woof	yum yum	peekaboo	grandma	peekaboo	bye	weee	no	hi	baby food
8	cat	hi	no	water	grandma	pattycake	thump	ouch	no	up
9	vroom	bye	bye	no	with	yum yum	hide and seek	ball	lamp	clap
10	yum yum	ouch	bath	woof woof	food	good night	meow	tickle	look	ball

Consistency in early vocabulary across languages

Preference for acquiring words with more potential links to their lexicons



Grammar and lexicon co-develop



Accessing Wordbank data

- Downloading...



- ...or using wordbankr

- Reproducible workflow
- Version control
- Flexible updating

wordbankr

R package for accessing the Wordbank database.

Installation

To install the released version on [CRAN](#):

```
install.packages("wordbankr")
```



To install the latest development version:

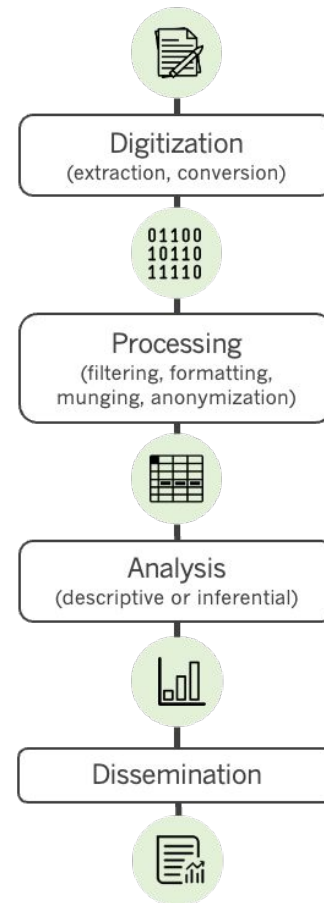
```
# install.packages("devtools")  
devtools::install_github("langcog/wordbankr")
```



Reproducibility: What

		Data	
		Same	Different
Code	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Analysis pipeline



Reproducibility: Why

		Data	
		Same	Different
Code	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Provenance,
transparency,
& verification

Error detection
& mitigation

Efficiency

Reproducibility: How

Reproducible reports: R Markdown (.Rmd) /
Quarto (.qmd) documents

- Embed all your processing and analysis code in your report
- Avoid updating / copy-paste errors
- Save pain when changing formats
- Interpretable and transparent
- Quick and ✨pretty✨



Practical session

- Go to <https://posit.cloud> and register (if you haven't)
- Create a “New RStudio project”
- Download the Rmd file for the practical at <https://github.com/alvinwmtan/2023-mpal-wordbank> and upload it into your project (the “Upload” button is in the bottom right panel)

