Navigating a quasiregular system: Efficiency and generalization in the early stages of learning to read

Matt Cooper Borkenhagen
University of Wisconsin – Madison

An enduring challenge

- Children learn within constraints (Chomsky, 1965; Griffiths, 2020; Hart & Risley, 1995; Seidenberg & MacDonald, 1999)
  - Memory (MacDonald, Just, & Carpenter, 1992)
  - Environment (Hart & Risley, 1995; Hoff, 2013)
  - Instruction (Darling-Hammond, 2000; Seidenberg, 2017)
  - Educational opportunity (Reardon, 2013)
  - Time (Fisher et al., 1981)
- Those that teach children are constrained too
- Constraints permeate how we understand kids, learning, education
- Especially in learning about print

The special case of learning about print

- Children must learn how print and speech relate
- Necessary but not sufficient for learning to read
- Short term: as much knowledge about print as possible
  - Crack the code
  - Achieve escape velocity
- Long term: comprehension
- Therefore: the goal is efficiency

Get the child as much knowledge as possible to enable them to read words (they have not yet seen)

Existing solutions?

- Enhance the child's skills with phonemes
- Teach the child about stable relations between print and speech
- Increase practice with print
- Rinse, repeat until the goal is accomplished *<u>or we run out of time</u>*

The problem?

- Which patterns matter most?
- Resources are limited
- Time

Seidenberg, Cooper Borkenhagen, & Kearns (2020)
Reading Research Quarterly

A different approach

- Learning is computational (Seidenberg & McClelland, 1989; Seidenberg et al., 2020)
- Embody learning in neurobiologically plausible learner
- Maps printed word to a spoken word
  - Seidenberg & McClelland (1989); Plaut et al. (1996)
- Expose the learner to many different learning environments
  - Sets of different single-syllable words
- Select sets of words that have good properties
  - Learning
  - Generalization

Study 1

REMEMBER:
Computer simulations,
not humans!

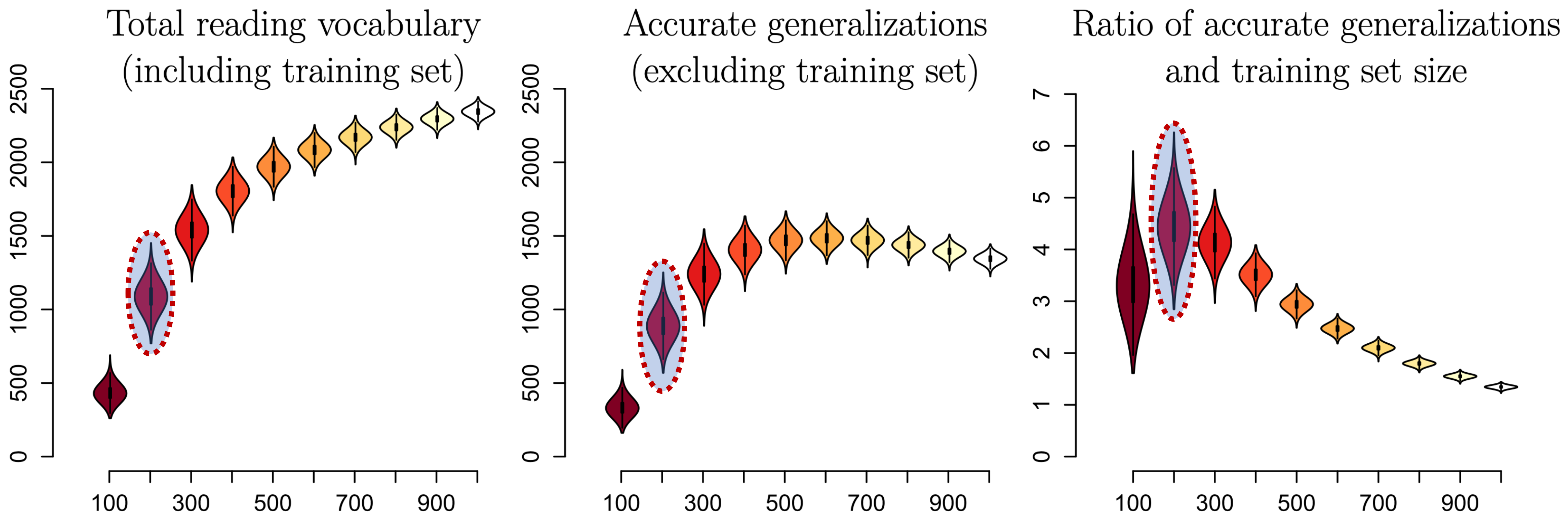
- How much generalization can you get from a subset of large pool of words
  - Selection pool: corpus of 2881 monosyllabic words
  - Subsets: range from 100 to 1000, increments of 100 (*k*)
- Randomly select *k* words from the selection pool
- Train the learner until it gets all those words right
- Test the model on the rest of the words
- Repeat for many, many samples of words

Cox, Cooper Borkenhagen, & Seidenberg (2019)
Proceedings of the Cognitive Science Society

- A large pool of words
- Take random samples (100, 200…1000)
- Train the computational learner on random sample
- See how well it does on the set of words it hasn't been taught

What are we looking for?

- For a set of *k* words, how much generalization can you get?
- How much more generalization do you get for more words (as *k* increases)


Total reading vocabulary
(including training set)
Accurate generalizations
(excluding training set)
Ratio of accurate generalizations
and training set size
0
500
1000
1500
2000
2500
0
1
2
3
4
5
6
7
100
300
500
700
900

Model Level	η_p^2	$\Delta \mathbf{R^2}$
Word length	0.002	0.001
Orth. Neighbors	0.006	0.005
Phon. Neighbors	0.000	0.000
Consistency	$\boxed{0.137}$	0.136

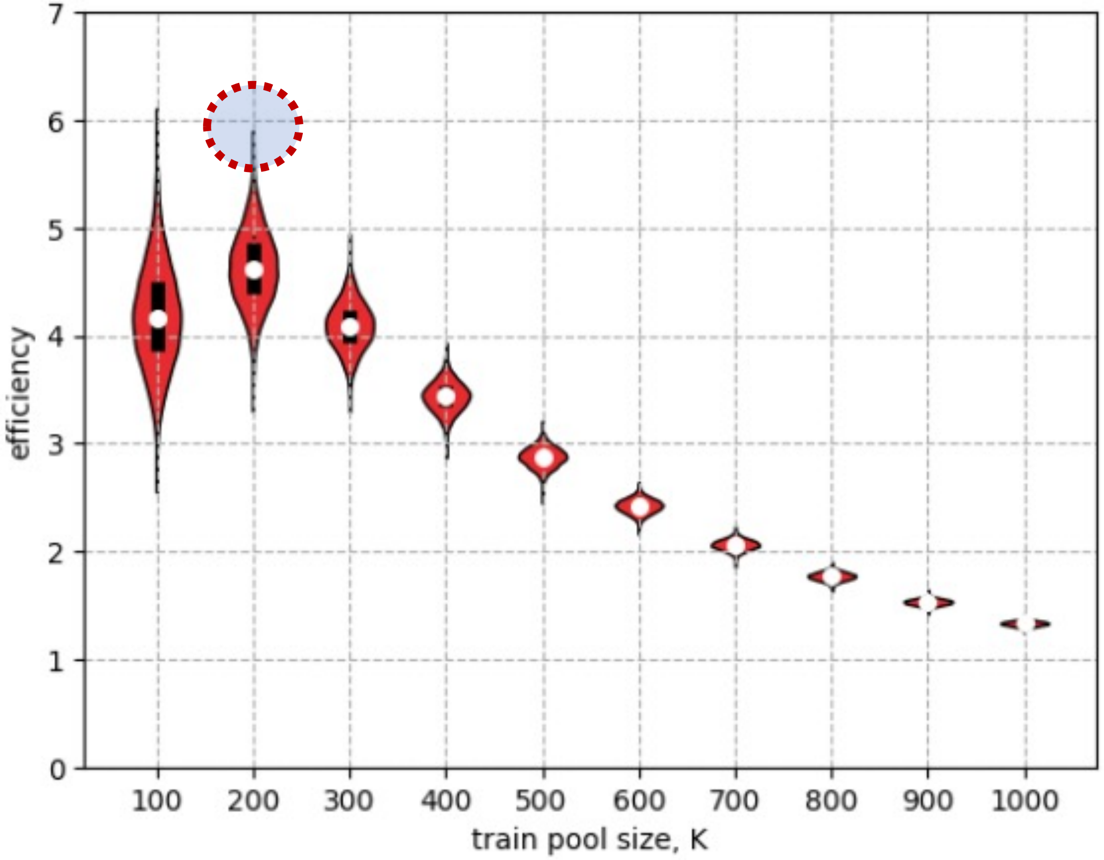
Study 2

- Sets of words vary in how good their structure is
- Good = generalizable
- But can sets be arranged sequentially to promote generalization?

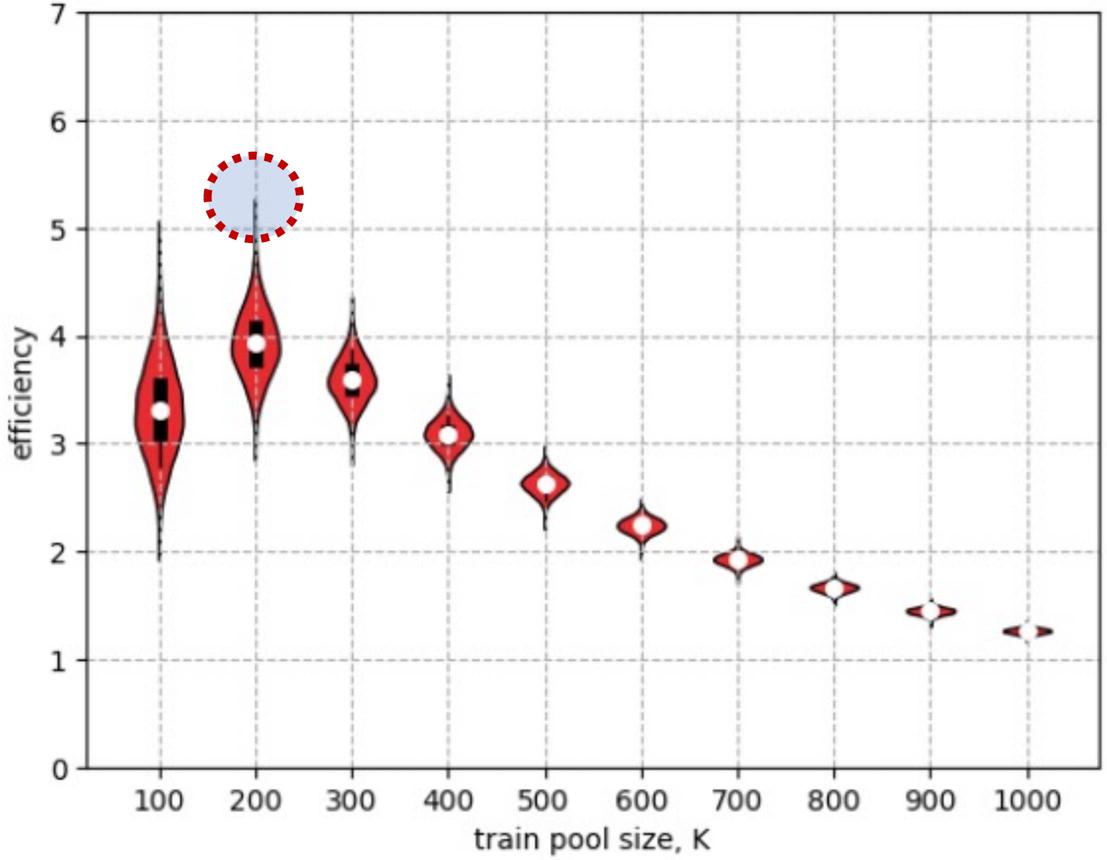
Procedure

- Optimize the sequence in which they are learned
- A two step approach
- Step 1: select a good set from a large corpus (Study 1)
- Step 2: manipulate the sequence to enhance generalization

Step 1: Find a good set (same as study 1)



Kid words

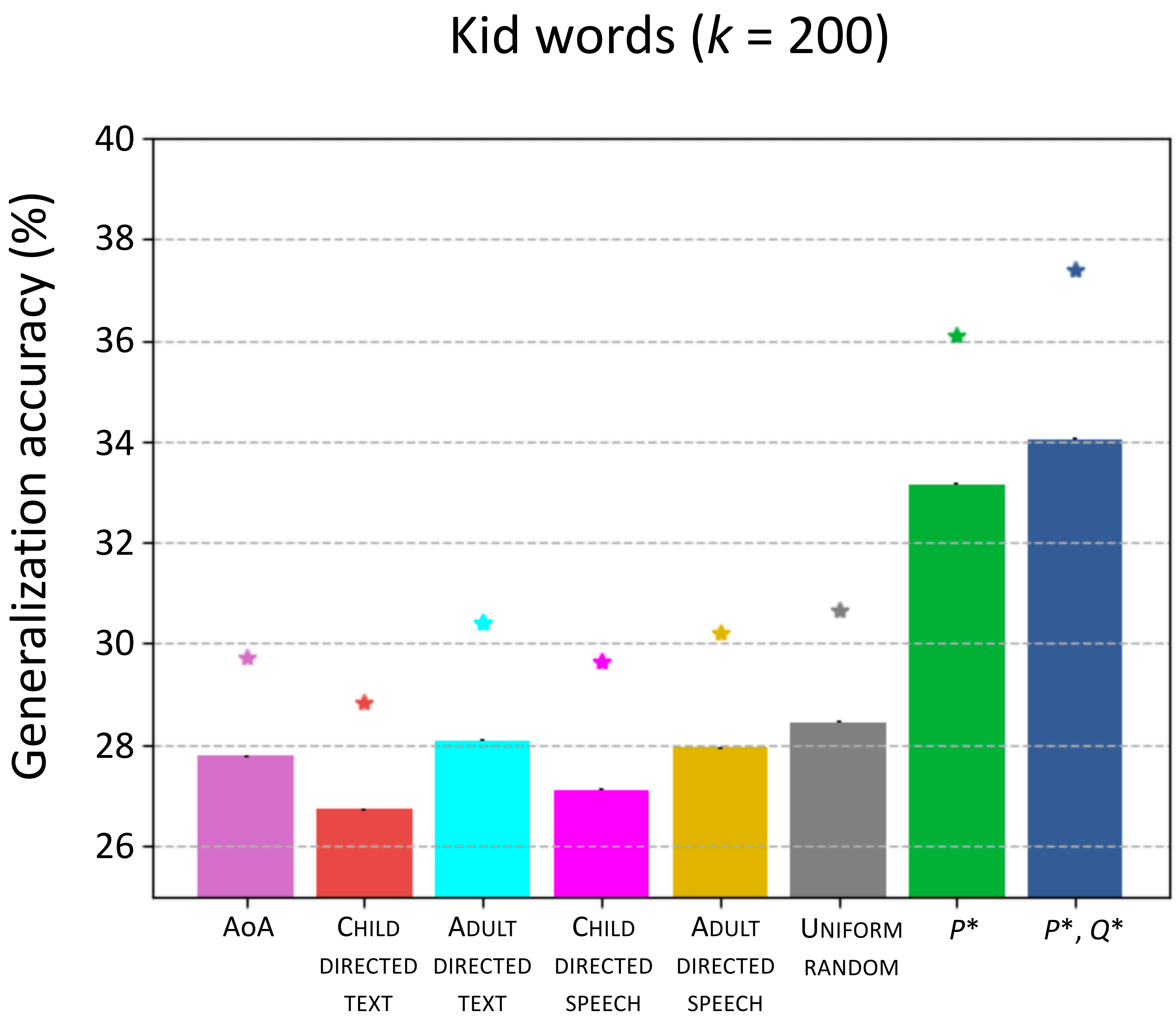


Adult words

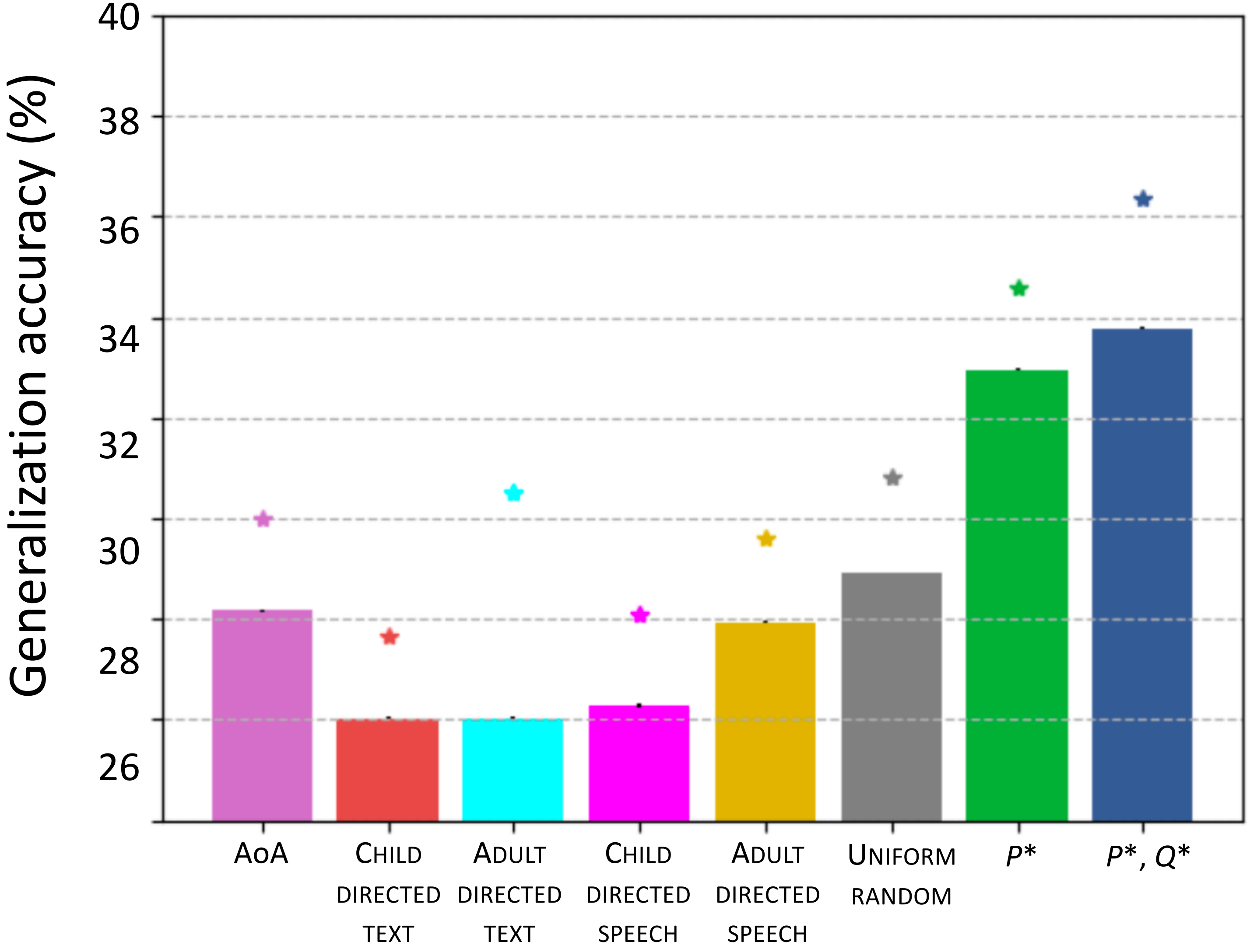
Step 2: Sequential optimization

- Start with a good set to build a good sequence
  - Kid words (set of 200 words)
  - Adult words (set of 200 words)
- Assign a probability to each word at the start of training (called *P*)
- And to each word at the end of training (called *Q*)
- Titrate the starting value with the ending value
- Train over a finite number of learning trials (10k)
- Adjust P, Q so they point towards better and better generalization
- Continue to adjust, until generalization plateau

Sen, et al. (2020) Preprint: arxiv.org/pdf/2006.16470.pdf


Kid words (k = 200)
Generalization accuracy (%)
40
38
36
34
32
30
28
26
AoA
CHILD DIRECTED TEXT
ADULT DIRECTED TEXT
CHILD DIRECTED SPEECH
ADULT DIRECTED SPEECH
UNIFORM RANDOM
P*
P*, Q*

Adult words (k = 200)



Summary

- Sets of words vary in their capacity to generalize
- Num of words selected for teaching can be adjusted intelligently
- Good sets = “bang for your buck” (efficiency)
- Sequencing can enhance learning, generalization
- The properties of good sets of words are multifaceted (quasiregular)
- Good words (in terms of structure) aren’t those that are in naturalistic environments (AoA, child directed text)

Takeaways

- Embrace efficient instruction
- Use corpora of words that are relevant for learning (Compton et al., 2014; Seidenberg et al., 2020)
- Constraints *should be* built into constructing learning environments
- Estimate what is realistic given the amount of time allocated
- Isolate smaller number of valuable words, increase efficiency

Thank you

Interdisciplinary Training Program in Education Sciences

UNIVERSITY OF WISCONSIN–MADISON

This work has been supported by the Vilas Trust and Deinlein Language and Literacy Fund at UW-Madison and by the Institute of Education Sciences, US Department of Education, through Award #R305B150003 to UW-Madison. The opinions expressed are those of the authors and do not represent views of the US Department of Education.

cooperborken@wisc.edu

References

Compton, D. L., Miller, A. C., Elleman, A. M., & Steacy, L. M. (2014). Have we forsaken reading theory in the name of "quick fix" interventions for children with reading disability?. *Scientific Studies of Reading*, *18*(1), 55-73.

Cox C.C., Cooper Borkenhagen, M., & Seidenberg, M.S. (2019). Efficiency of learning in experience-limited domains: Generalization beyond the

WUG Test. Proceedings of the 41st Annual Meeting of the Cognitive Science Society, Montreal, Quebec, Canada.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1981). Teaching behaviors, academic learning time, and student achievement: An overview. *The Journal of classroom interaction*, *17*(1), 2-15.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive psychology*, *24*(1), 56-98.

Chomsky, N. (1965). Aspects ofthe theory ofsyntax. *Cambridge, MA: MITPress*

Griffiths, T. L. (2020). Understanding Human Intelligence through Human Limitations. *Trends in Cognitive Sciences*.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive science*, *23*(4), 569-588.

Hoff, E. (2013). Interpreting the early language trajectories of children from low SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, *49*(1), 4–14.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education policy analysis archives*, *8*, 1-44.

Seidenberg, M.S. (2017). *Language at the speed of sight: How we read, why so many can't, and what can be done about it*. New York: Basic Books.

Seidenberg, M. S., Cooper Borkenhagen, M., & Kearns, D. M. (2020). Lost in translation? Challenges in connecting reading science and educational practice. *Reading Research Quarterly*, *55*, S119-S130.

Sen, A., Cox, C. R., Borkenhagen, M. C., Seidenberg, M. S., & Zhu, X. (2020). Learning to Read through Machine Teaching. *arXiv preprint arXiv:2006.16470*.

Reardon, S. F. (2013). The widening income achievement gap. *Educational leadership*, *70*(8), 10-16.