Grounding early reading instruction in a theory of learning: Deciding what to teach and why

Matt Cooper Borkenhagen

October 27, 2022

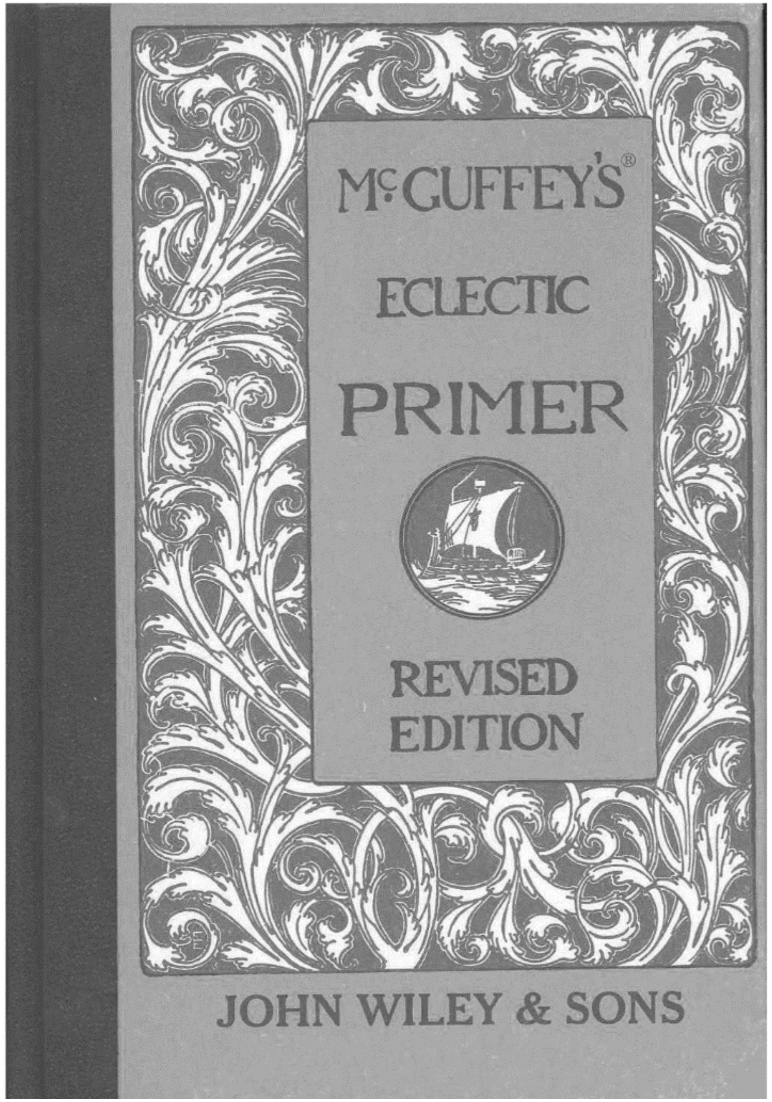
WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Part I: Some old ideas in reading education

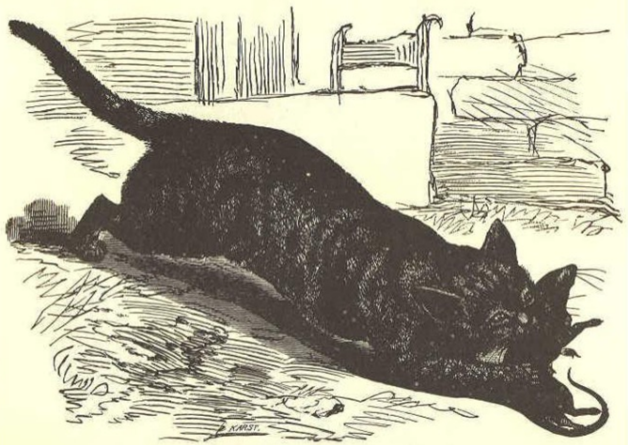
Part 2: Some new ideas in reading education

Orthodoxy: That peculiar condition where the patient can neither eliminate an old idea nor absorb a new one.

Old idea: Special words



LESSON II.



ăt the răn hăș

Ănn

h th ș

The cat the rat

The cat has a rat.

The rat ran at Ann.

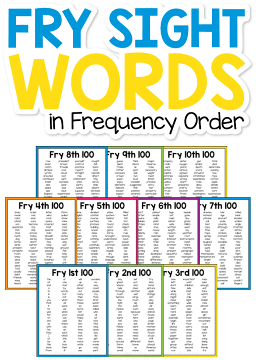
Ann has a cat.

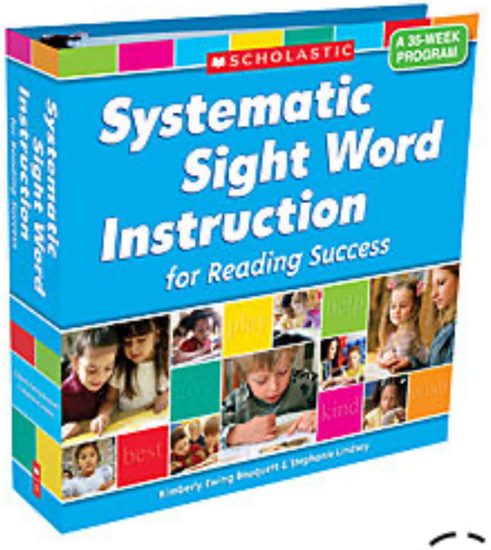
The cat ran at the rat.

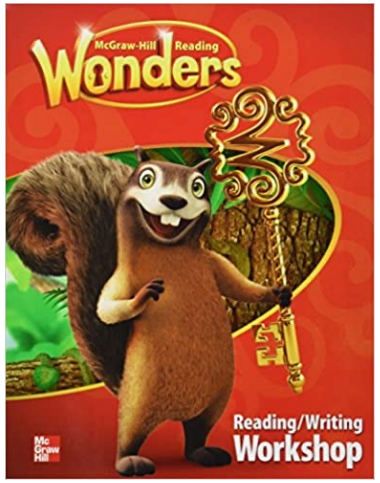
100 Sight Words

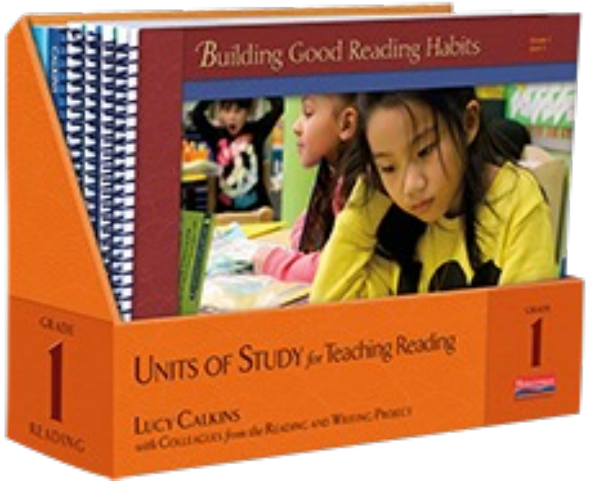
the	he	at	but	there
of	was	be	not	use
and	for	this	what	an
a	on	have	all	each
to	are	from	were	which
in	as	or	we	she
is	with	one	when	do
you	his	had	your	how
that	they	by	can	their
it	I	words	said	if

will	some	two	my	find
up	her	more	than	long
other	would	write	first	down
about	make	go	water	day
out	like	see	been	did
many	him	number	call	get
then	into	no	who	come
them	time	way	am	made
these	has	could	its	may
so	look	people	now	part



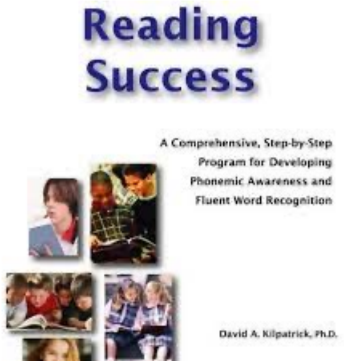






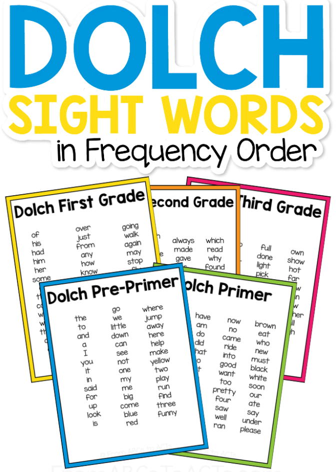
fundations®
Wilson Language Basics

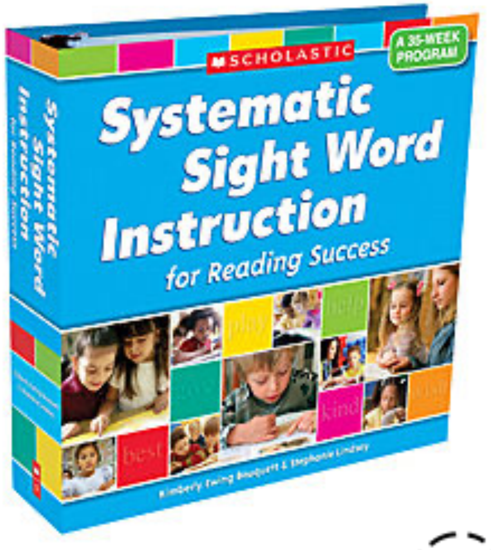
JOURNEYS
COMMON CORE

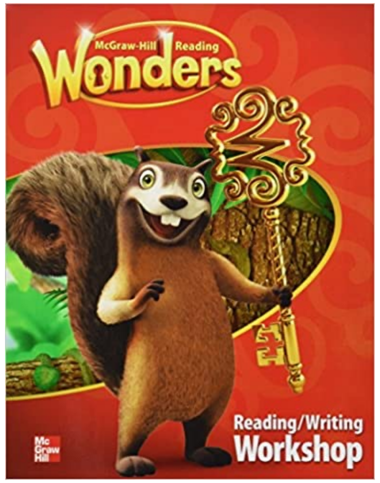


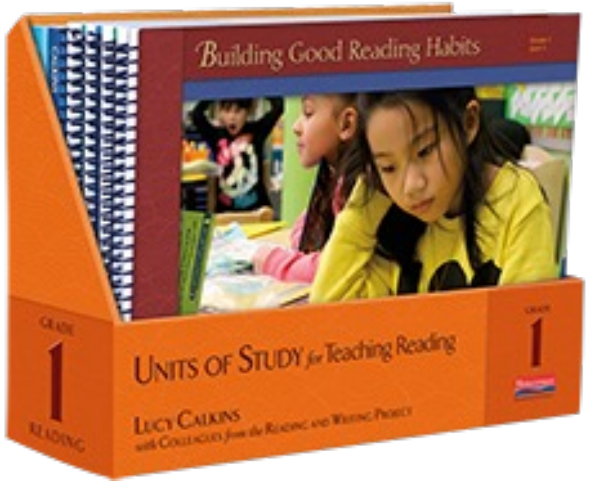
Teachers Pay Teachers

WILSON
WILSON Reading System®



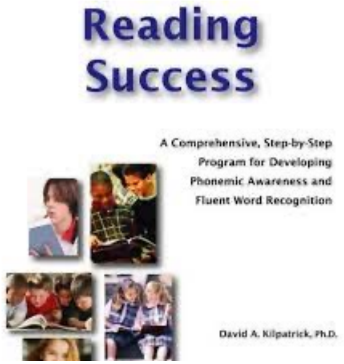




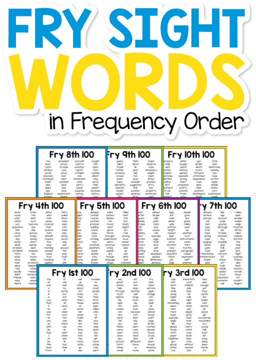


fundations®
Wilson Language Basics

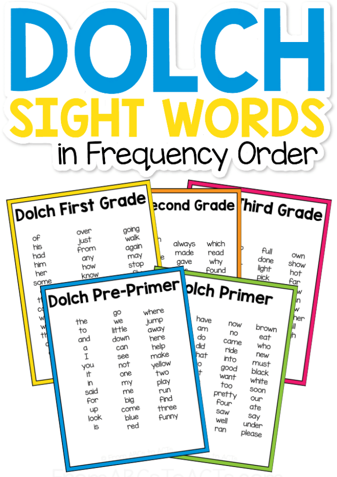
JOURNEYS
COMMON CORE



WILSON
WILSON Reading System®



Teachers Pay Teachers



The New York Times

OPINION

Why Are We Still Teachi[ng …] the Wrong Way[…]

Teacher preparation programs continue t[…] science behind how people becom[…]

Oct. 26, 2018

EDUCATION

Why Millions Of Kids Can't Read And What Better Teaching Can Do About It

January 2, 2019 · 6:00 AM ET

Heard on Morning Edition

EMILY HANFORD

FROM AMERICAN PUBLIC MEDIA

The New York Times

OPINION

There Is a Right Way to Teach [Re]ading, and Mississippi Knows It

[T]he state's reliance on cognitive science explains why.

EdW Popular Literacy Materials Get[…]

READING & LITERACY

Popular Literacy Materials Get 'Science of Reading' Overhaul. But Will Teaching Change?

Lucy Calkins and Jennifer Serravallo Are Among Those Making Shifts

By Sarah Schwartz — October 13, 2021 18 min rea[d]

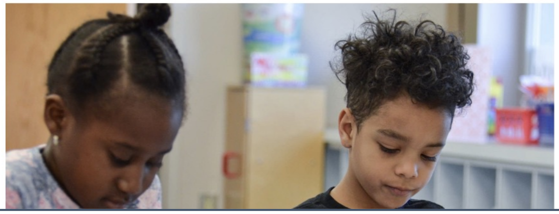
[…]MEDIA

READING

Influential authors Fountas and Pinnell stand behind disproven reading theory

The education professors double down on a flawed approach that encourages pictures and context to read words. Heinemann — their publisher — faces harsh criticism.

November 19, 2021 | by Emily Hanford and Christopher Peak



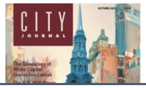
CJ

Many fields suffer from the persistence of received ideas that are unsupportable, but the problem is especially severe in American education. In many cases, systematic research has illustrated that common current practices could be vastly improved on, but such improvements have not been implemented widely.

A 2019 report by Emily Hanford for American Public Media, "At a Loss for Words," described how millions of students

Sign up for Email Alerts

AUTUMN 2021



TEA Texas Education Agency

To the Administrator Addressed

Commissioner Mike Morath

1701 North Congress Avenue • Austin, Texas 78701-1494 • 512 463-9734 • 512 463-9838 FAX • tea.texas.gov

DATE:	September 12, 2019
SUBJECT:	House Bill 3 Reading Practices
CATEGORY:	Early Childhood Education
NEXT STEPS:	Action Required of LEA: Share with appropriate staff

House Bill (HB) 3 was passed by the 86th Texas Legislature, 2019, and signed into law by Governor Abbott on June 11, 2019. The following areas within HB 3 specifically relate to reading practices:

- The Science of Teaching Reading (STR) exam (TEC Sec. 21048 (a-2))
- Reading Standards for Kindergarten Through Third Grade (TEC Sec. 28.0062(a))
    - Literacy Academies
    - Certified Practices
        - Phonics curriculum
        - Placement of highly effective teachers
        - Integrated reading instruments
    - Reading Advisory Board

NORTH CAROLINA GENERAL ASSEMBLY

S386 — **Senate Bill 387 / SL 2021-8** — S388

Excellent Public Schools Act of 2021.
2021-2022 Session

VIEW BILL DIGEST
VIEW AVAILABLE BILL SUMMARIES

EDITION	FISCAL NOTE
Filed	
Edition 1	
Edition 2	
Ratified	
SL 2021-8	

Last Action:	Ch. SL 2021-8 on 4/9/2021
Sponsors:	Berger; Ballard; Lee (Primary) Barnes; Burgin; Corbin; Edwards; Galey; Jarvis; Johnson; Lazzara; Newton; Rabon; Sanderson
Attributes:	Public; Text has changed
Counties:	No counties specifically cited
Statutes:	115C (Chapters); 115C-269.20, 11 270.30, 115C-83.10, 115C-83.11, 83.2, 115C-83.3, 115C-83.4B, 115 83.6, 115C-83.6A, 115C-83.6B, 11 83.7, 115C-83.7A, 115C-83.8, 115 83.9 (Sections)

ALABAMA STATE DEPARTMENT OF EDUCATION

Notes FAQ Contact Us

ALABAMA'S JOURNEY TO READING SUCCESS:

THE *ALABAMA LITERACY ACT* IMPLEMENTATION GUIDE

Every Child.

Every Chance.

Every Day.

Chalkbeat

CURRICULUM AND INSTRUCTION POLITICS & POLICY COLORADO READING

Colorado cracks down on schools using weak reading curriculum. Advocates worry about backpedaling.

By Ann Schimke | Nov 15, 2021, 6:37pm MST

ERIC

Collection Thesaurus
Search education resources | Search | Advanced Search Tips
Peer reviewed only Full text available on ERIC

Mississippi's Literacy-Based Promotion Act: An inside Look

RMC Research Corporation

When students are unable to read by the end of third grade, their risk of falling behind academically grows exponentially. In fact, research shows that nearly nine out of ten high school dropouts were struggling readers in third grade. Students who are not reading proficiently by the end of third grade are four times more likely to drop out of high school, and high school dropouts are not eligible for 90 percent of jobs in the U.S. economy. To address this issue and ensure all students become capable readers by the end of third grade, many states have passed comprehensive K-3 reading policies, including Mississippi. In 2013, the state established the Literacy-Based Promotion Act (LBPA). The law was to ensure a comprehensive approach to teaching all children to read starting as early as kindergarten. The true spirit of the law is to identify K-3 students who need additional help in reading as early as possible and to provide effective instruction and intervention to ensure they read on grade level by the end of third grade. In 2016, the law was amended to include individual reading plans for students identified with a reading deficiency, and a higher cut score was established for third-grade promotion. Additionally, it required teacher candidates to pass a foundational reading test for certification to ensure they have the knowledge and skill to teach all students to read. Since Mississippi enacted the LBPA in 2013 the state's fourth-grade National Assessment of Educational Progress (NAEP) reading scores have substantially improved. In 2013, 21 percent of fourth graders were proficient in reading, and by 2017, 27 percent of fourth graders were reading proficiently: a 6-percentage point increase. Mississippi also decreased the percentage of fourth graders scoring below basic by 7 percentage points, going from 47 percent scoring below basic to 40 percent. Furthermore, Mississippi is second in the nation in learning gains. ExcelinEd contracted with RMC Research Corporation to conduct a study to better understand Mississippi stakeholders' perceptions of and experiences with the LBPA. The questions of interest were as follows: (1) What support strategies and technical assistance do the Mississippi Department of Education (MDE) and school districts provide to improve K-3 literacy? (2) How are teachers' perceptions of the LBPA? (3) What recommendations do stakeholders have for improving the implementation process? and (4) What impact has the LBPA had on districts and schools?

Download full text

ERIC Number: ED613758
Record Type: Non-Journal
Publication Date: 2019-Feb
Pages: 35
Abstractor: ERIC
ISBN: N/A
ISSN: N/A

AZBIGMEDIA

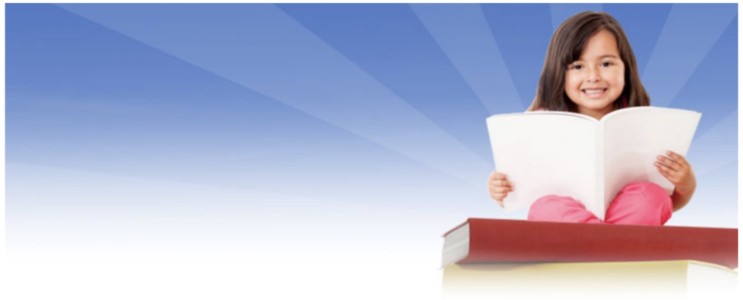
Early childhood literacy bill aims to improve Arizona classrooms

BUSINESS NEWS | 30 Oct, 2021 | ELIZABETH SCHAIBLE

ACT #2019- 523

HB388
200613-3
By Representatives Collins and Baker
RFD: Education Policy
First Read: 09-APR-19

MAY 31 2019 RECEIVED GOVERNOR'S OFFICE



Policy News: Arizona Invests in Accelerating Early Literacy

July 12, 2021

Read On Arizona applauds Governor Ducey and the Arizona legislature for prioritizing early literacy in the FY22 budget and legislation passed this month. SB1572 outlines significant steps for advancing early literacy in our state and was signed into law by Governor Ducey.

Education Next

Can Teaching Be Improved by Law?

At least twenty states have passed or are considering measures related to the science of reading.

Robert Pondiscio

Important in complex print systems

- How to deal with properties of English? (different for other languages)
- At onset of reading instruction, kids know spoken language
- They need to connect it with print
- Important transition; bc of inconsistencies in spelling-sound structure
- A major point of disagreement in education
- Teaching initial vocabulary is key

Castles, Rastle, & Nation (2018) *Psych Science Pub Interest*

Rayner et al. (2002) *Psych Science Pub Interest*

Seidenberg, Cooper Borkenhagen, & Kearns (2021) *RRQ*

One educational solution: special words

- “Sight words” to some (there are many names)
- A common educational idea
- Not clear: the underlying theory that drives it
- Conceptualizations appear to vary
- Teach these special words to accelerate word reading skills
- Often paired with instruction that emphasizes the phonic “rules”

the, one

number, being, Mr., Mrs.

The basic instructional theory

Two dimensions are often considered

Common

Special word

Atypical structure

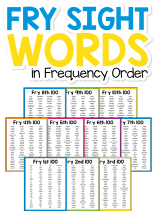
The problem: no principled approach

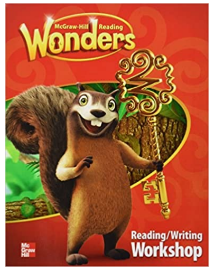
- Do they agree on the dimensions?
- Today’s focus: the “atypical structure” part
- Are the same words, properties included across resources?
- If they're different, how?

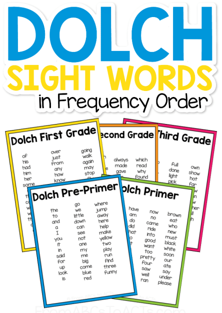
Cooper Borkenhagen, Schilling, & Seidenberg (*in prep*)

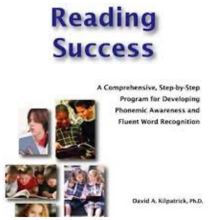
Materials

- Five major (very popular) resources
- Grades K-3
- Dolch, Fry, Fundations, Kilpatrick, Wonders
- 2 word lists (Dolch, Fry), 3 instructional programs (the rest)









- Text-transcribed words, other instructional data
- Constructed a database along with other language data
- Total number of tokens > 250K

Atypical structure: how we measure it

- How many friends v. enemies (of print + speech kind)
- Using body-rime units (vowel and everything following it)
- C<u>AN</u> → F<u>AN</u>, R<u>AN</u>, T<u>AN</u>, B<u>AN</u> (1.0 CONSISTENCY)
- S<u>OWN</u> → GR<u>OWN</u>, FL<u>OWN</u>, FR<u>OWN</u>, D<u>OWN</u> (.5 CONSISTENCY)
- H<u>AVE</u> → G<u>AVE</u>, S<u>AVE</u>, P<u>AVE</u>, CR<u>AVE</u> (0 CONSISTENCY)

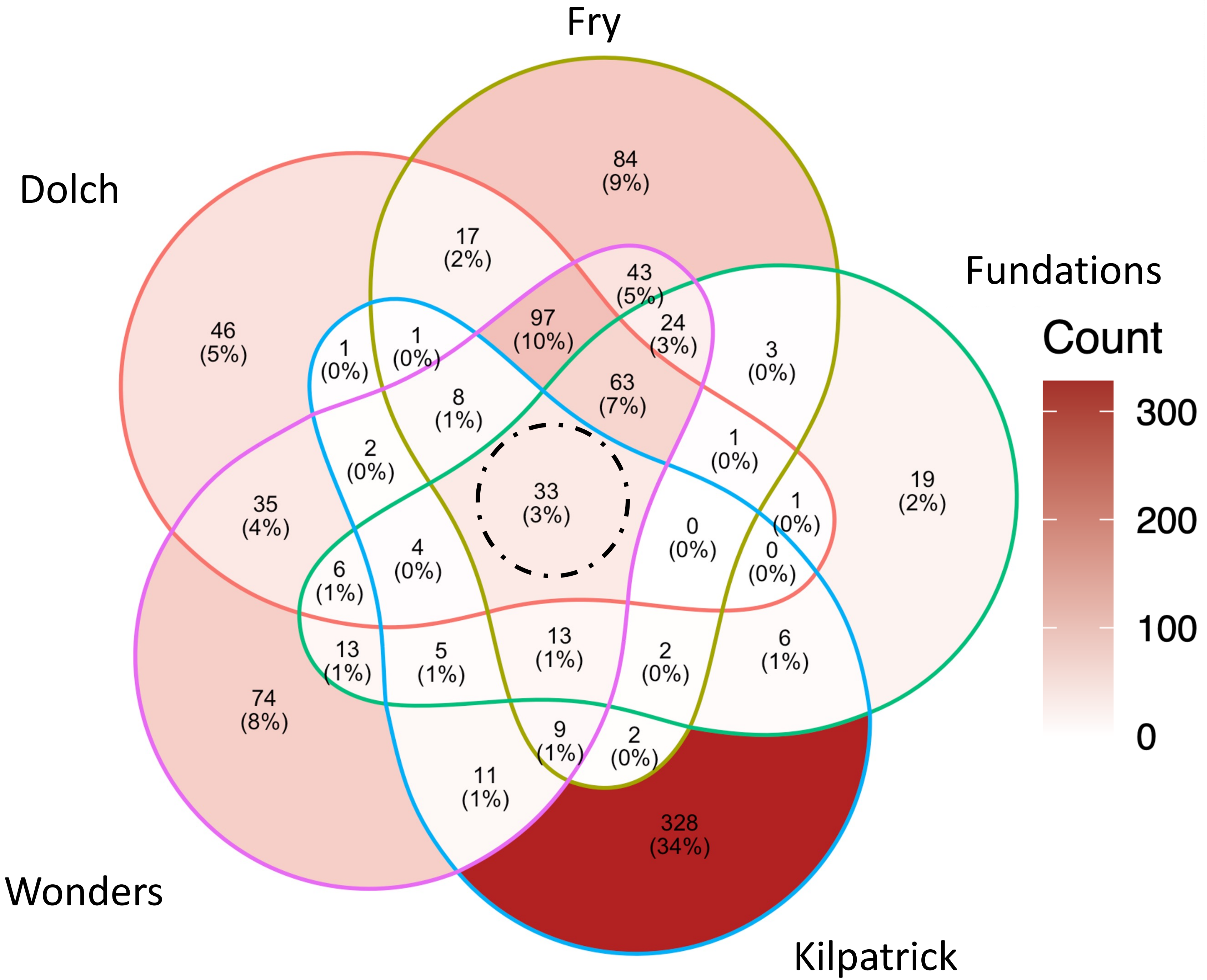
Comparing to a children's book corpus

- To facilitate comparison of resources, we used a reference sample
- Wisconsin Children's Book Corpus (WCBC)
- 250 popular US children's books
- Approximately 10K unique words, 200K total words
- Logic: measure consistency across all words in WCBC corpus
- Standardize the measurement (- mean / SD)
- Take words associated with each resource
- See where it stands against full sample, and other resources

Lewis, Cooper Borkenhagen, Converse, Lupyan, & Seidenberg (2021) *Psych Science*

Cooper Borkenhagen, Lewis, Converse, Borman, & Seidenberg (*in prep*)

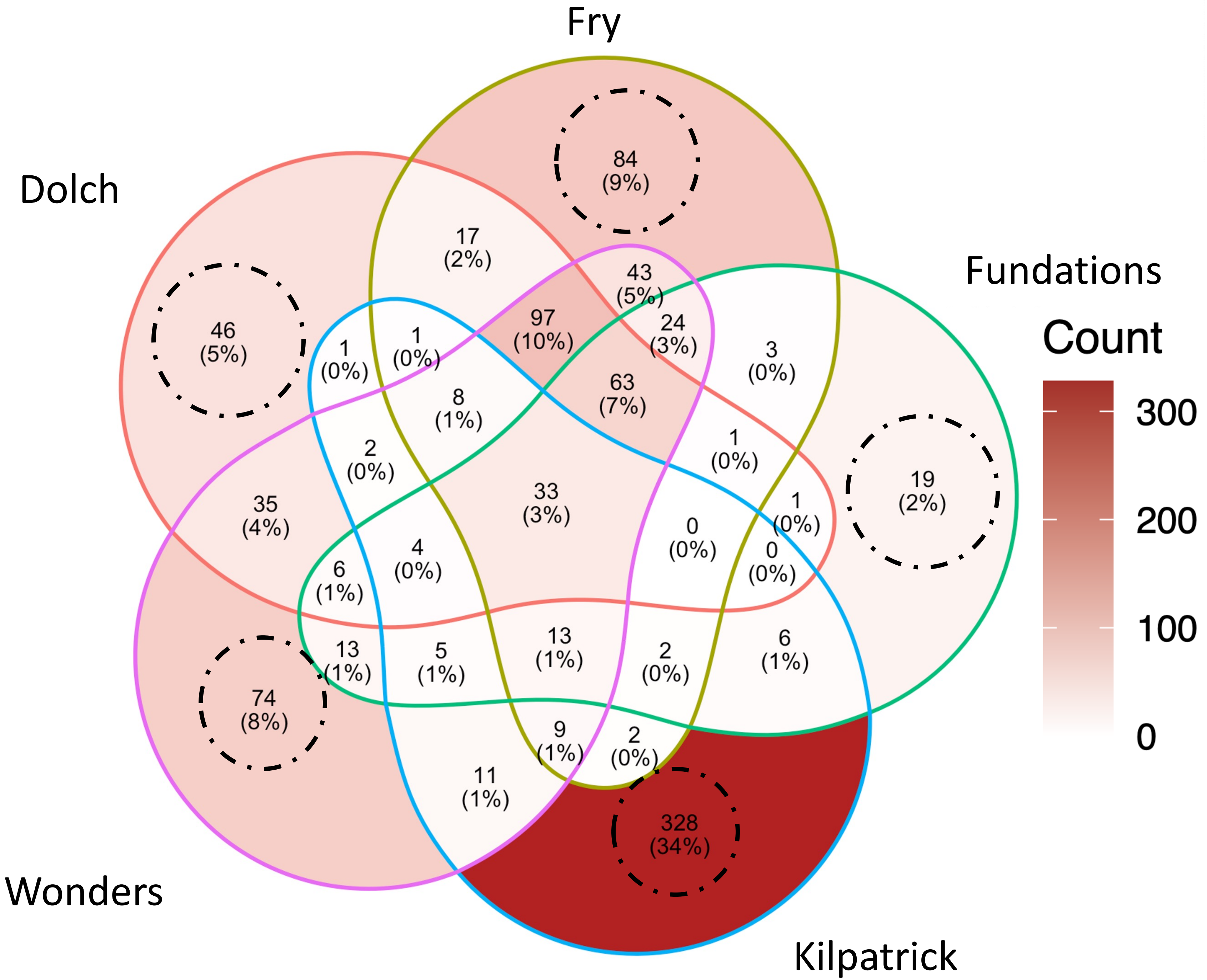

Consistency
Kilpatrick
Dolch
Fry
Fundations
Wonders
Standardized value (within WCBC)
0.0
-0.5
-1.0
Resource

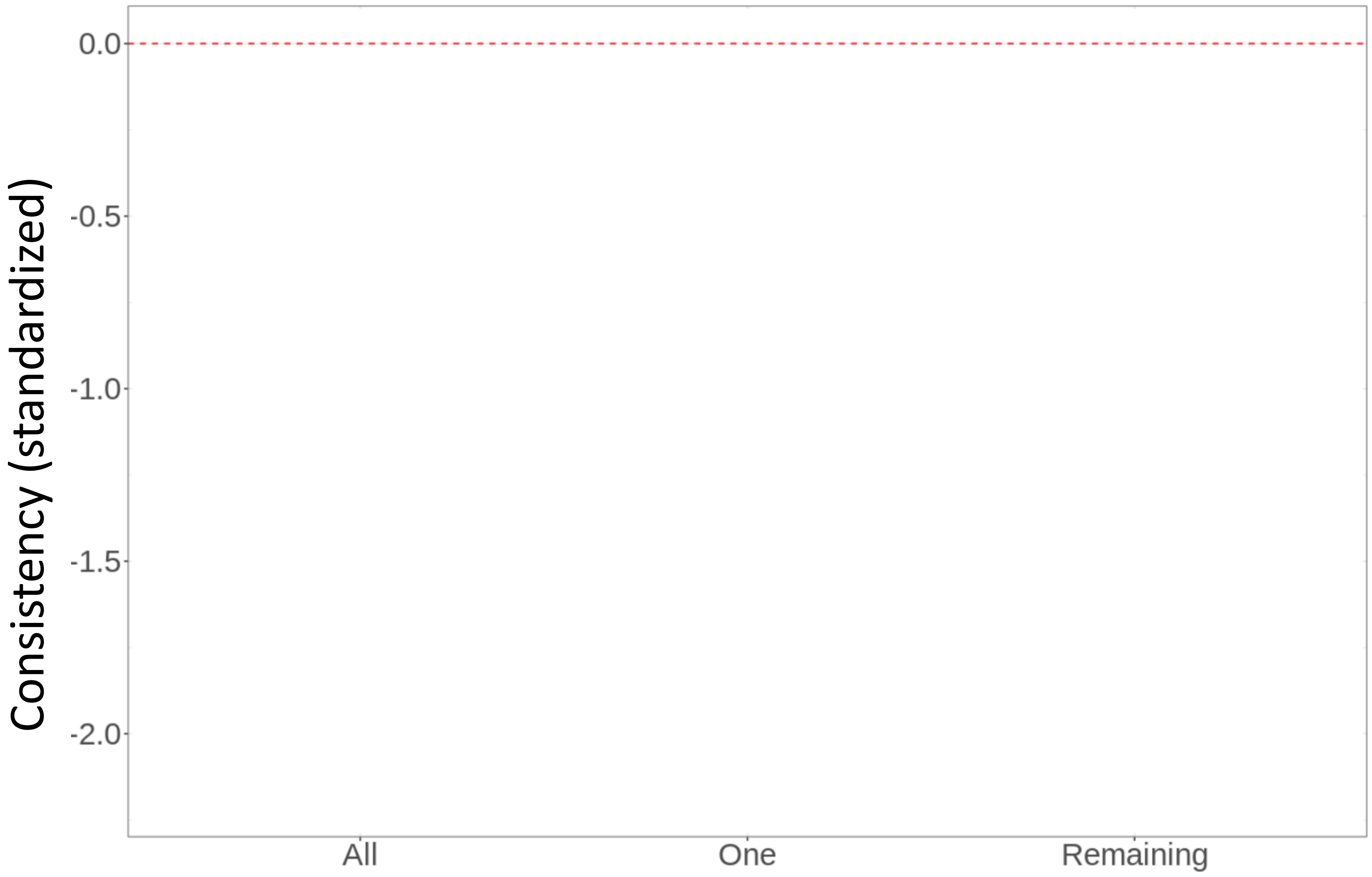


Fry
Dolch
Fundations
Count
300
200
100
0
Wonders
Kilpatrick
84 (9%)
17 (2%)
43 (5%)
46 (5%)
1 (0%)
1 (0%)
97 (10%)
24 (3%)
3 (0%)
63 (7%)
8 (1%)
2 (0%)
1 (0%)
33 (3%)
35 (4%)
19 (2%)
1 (0%)
0 (0%)
0 (0%)
4 (0%)
6 (1%)
13 (1%)
5 (1%)
13 (1%)
2 (0%)
6 (1%)
74 (8%)
9 (1%)
2 (0%)
11 (1%)
328 (34%)

Words Common across All Instructional Sources with Rank Frequency

the (*1*)	one (*32*)	would (*63*)	two (*105*)	been (*212*)
to (*4*)	do (*33*)	from (*66*)	again (*118*)	water (*218*)
of (*8*)	what (*34*)	who (*72*)	put (*126*)	does (*238*)
was (*10*)	are (*40*)	where (*82*)	has (*140*)	many (*253*)
said (*13*)	have (*44*)	come (*89*)	head (*178*)	both (*478*)
is (*15*)	were (*52*)	mother (*90*)	only (*181*)	
his (*17*)	into (*58*)	know (*100*)	once (*197*)	


Fry
Dolch
Fundations
Wonders
Kilpatrick
Count
300
200
100
0
84
(9%)
46
(5%)
17
(2%)
43
(5%)
97
(10%)
24
(3%)
3
(0%)
1
(0%)
1
(0%)
63
(7%)
8
(1%)
19
(2%)
1
(0%)
2
(0%)
35
(4%)
33
(3%)
1
(0%)
0
(0%)
0
(0%)
6
(1%)
4
(0%)
13
(1%)
13
(1%)
5
(1%)
2
(0%)
6
(1%)
74
(8%)
9
(1%)
2
(0%)
11
(1%)
328
(34%)



(the most popular words)
(the least popular words)

Different assumptions about which words

- Very few words (3%) are common to all resources
- The focus is on inconsistent words
- Very little agreement across lists
- This suggests variable experiences for kids
- Differences in conceptualization, implementation

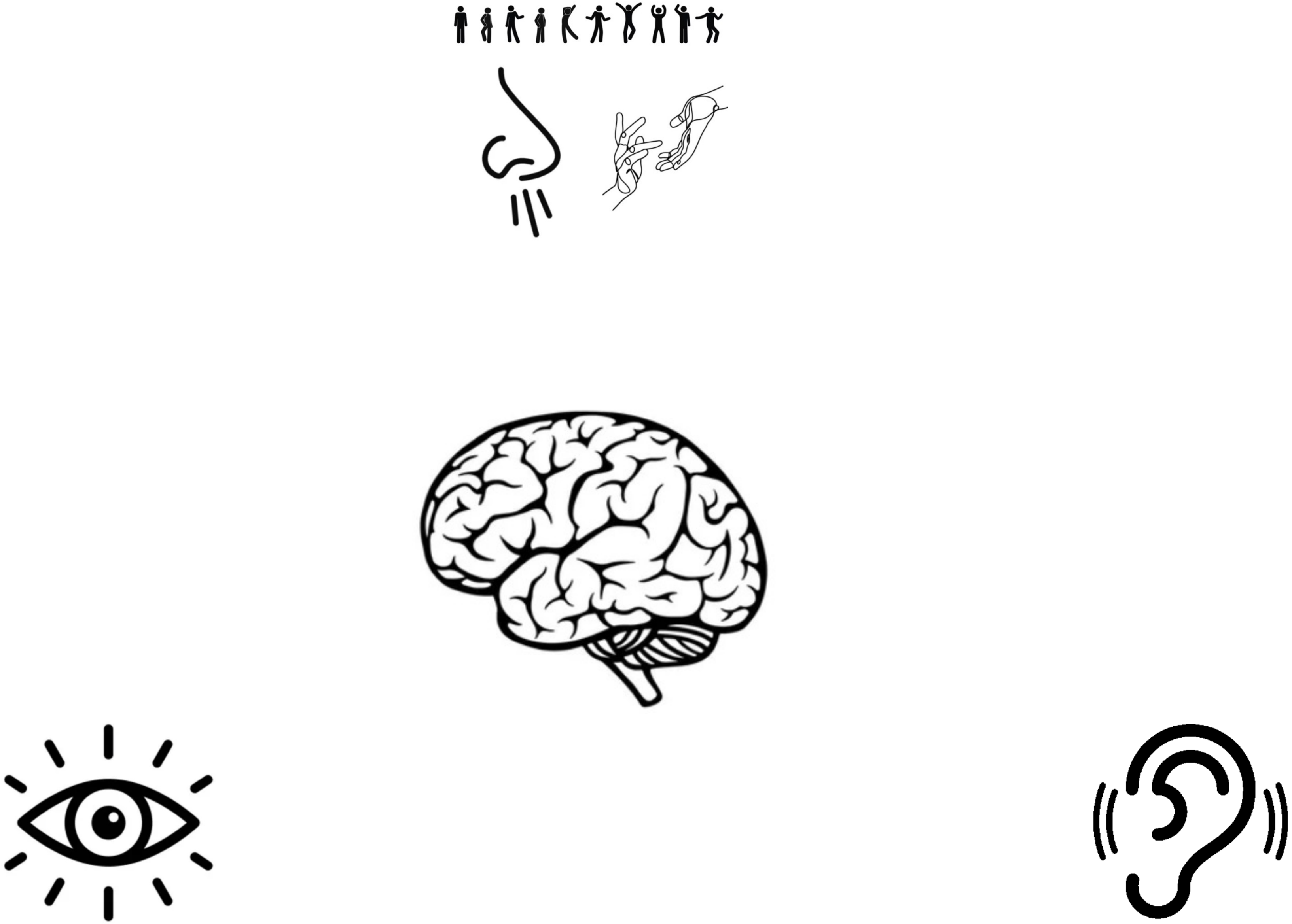
Lack of agreement about what is important!

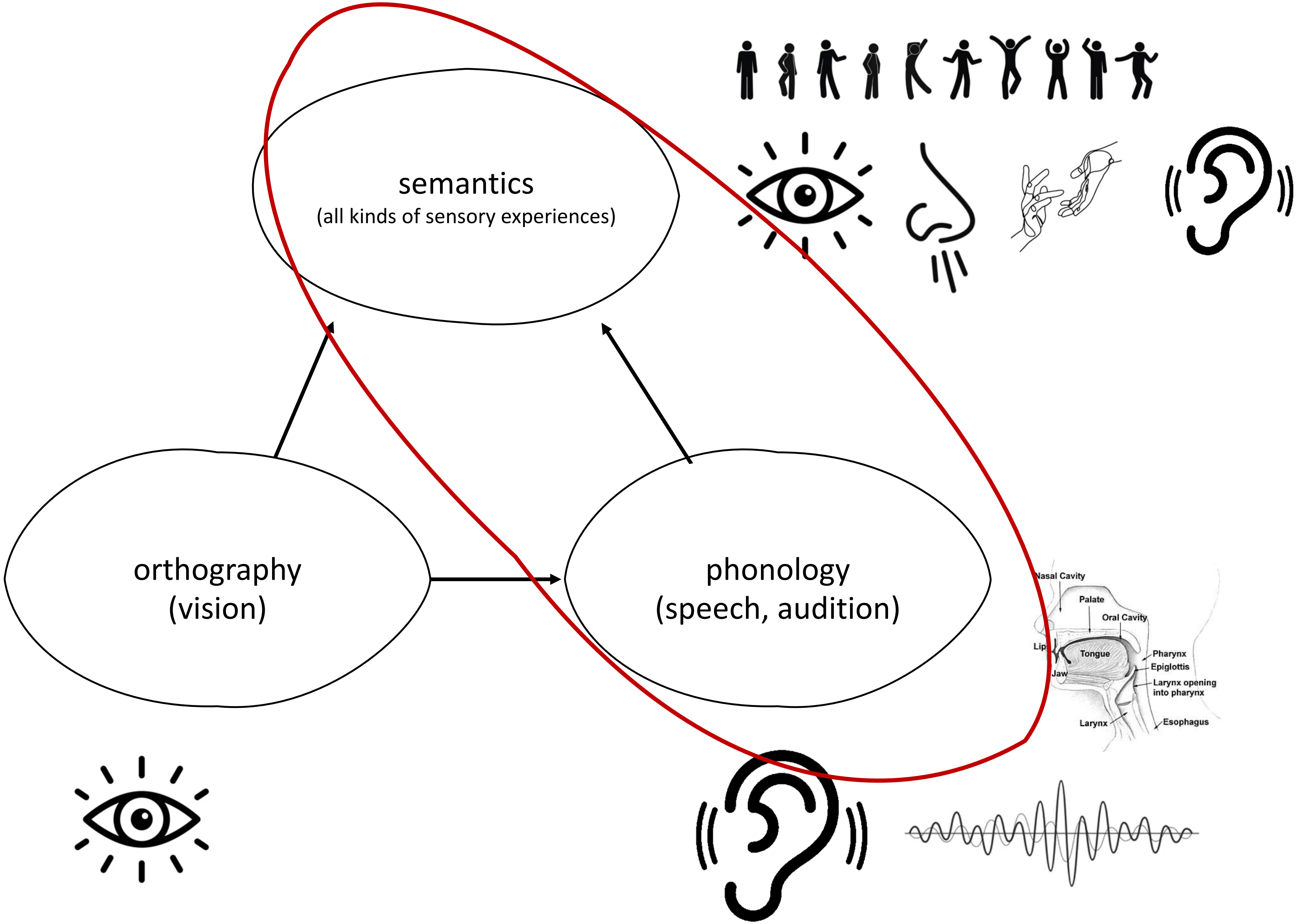
Part 2: Some new ideas

How *could* we find ourselves some special words?

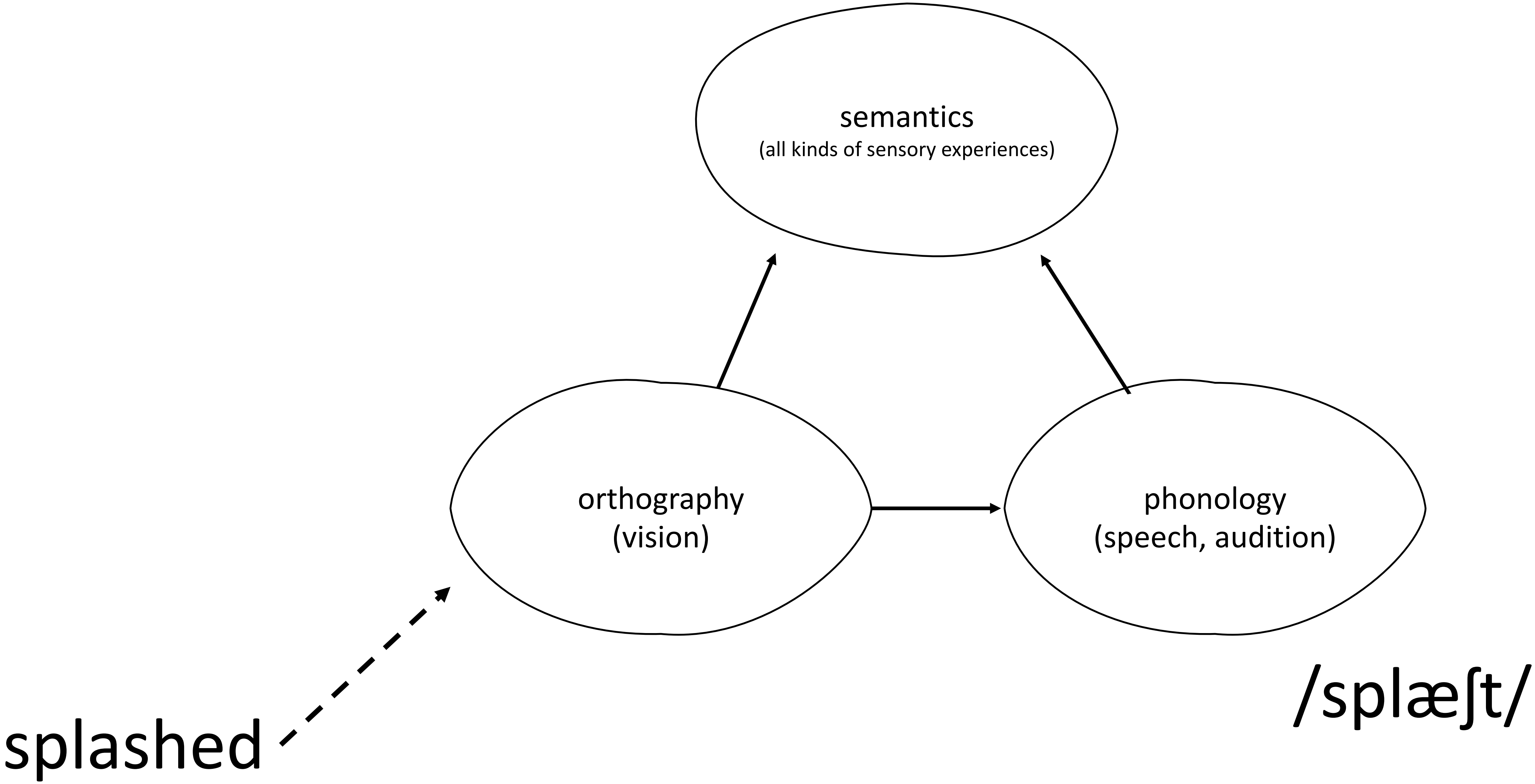
There are other theories, right?

- Cognitive science has some theories of these phenomena
- SOR is having a moment
- I’ll speak to one pocket of the science
- The theory goes by many names




semantics
(all kinds of sensory experiences)
orthography
(vision)
phonology
(speech, audition)
Palate
Oral Cavity
Tongue
Jaw
Pharynx
Epiglottis
Larynx opening into pharynx
Larynx
Esophagus

< water, movement, action, past >

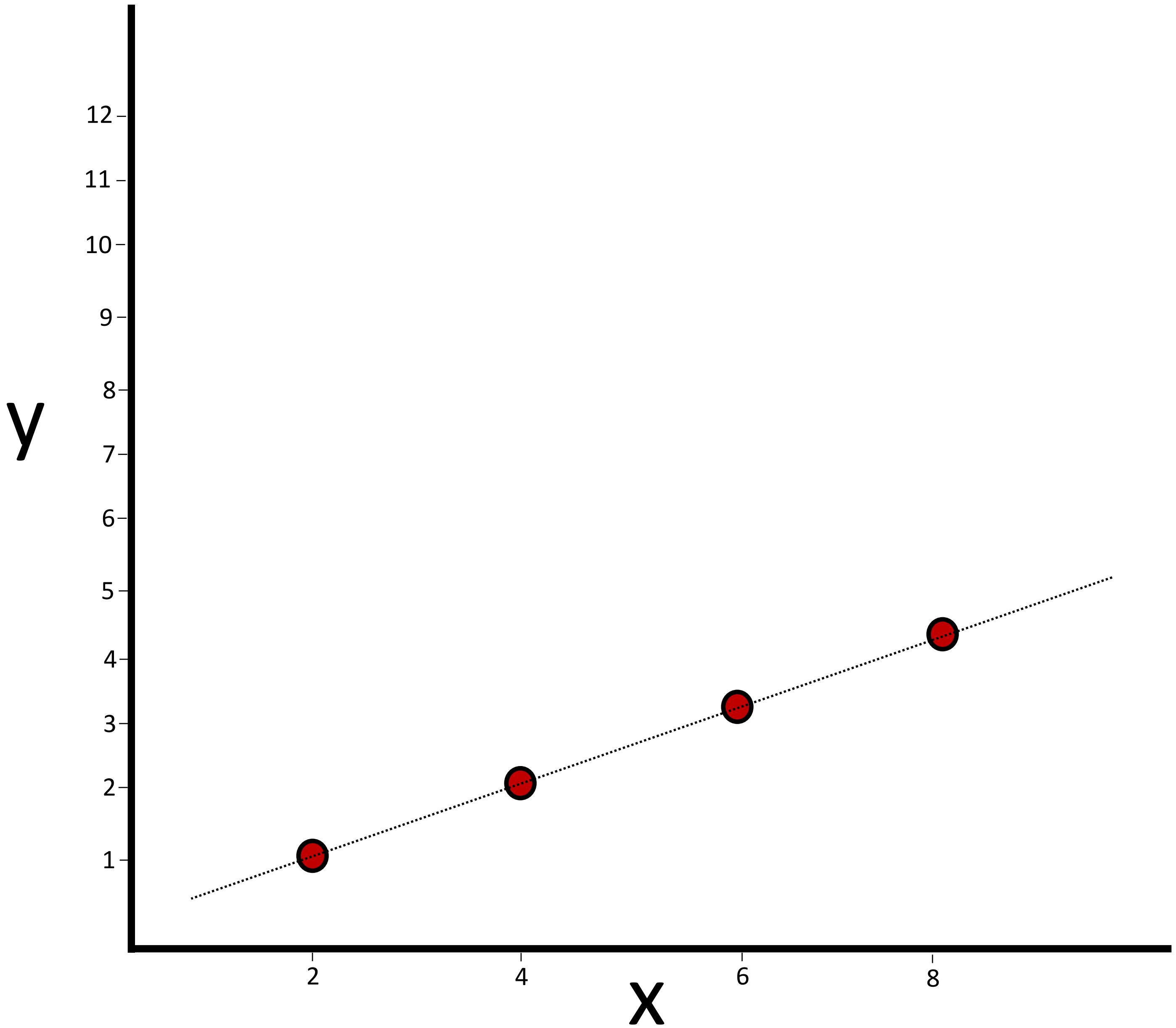


A number of theoretical principles

- Inductive learning processes over language, environment
- Involves encoding of statistical dependencies
- Happens continuously throughout development
- Learning uncovers graded neighborhoods of structure
- Computational mechanisms; neural processes
- So: we use artificial neural networks to simulate learning

$f(x) = y$

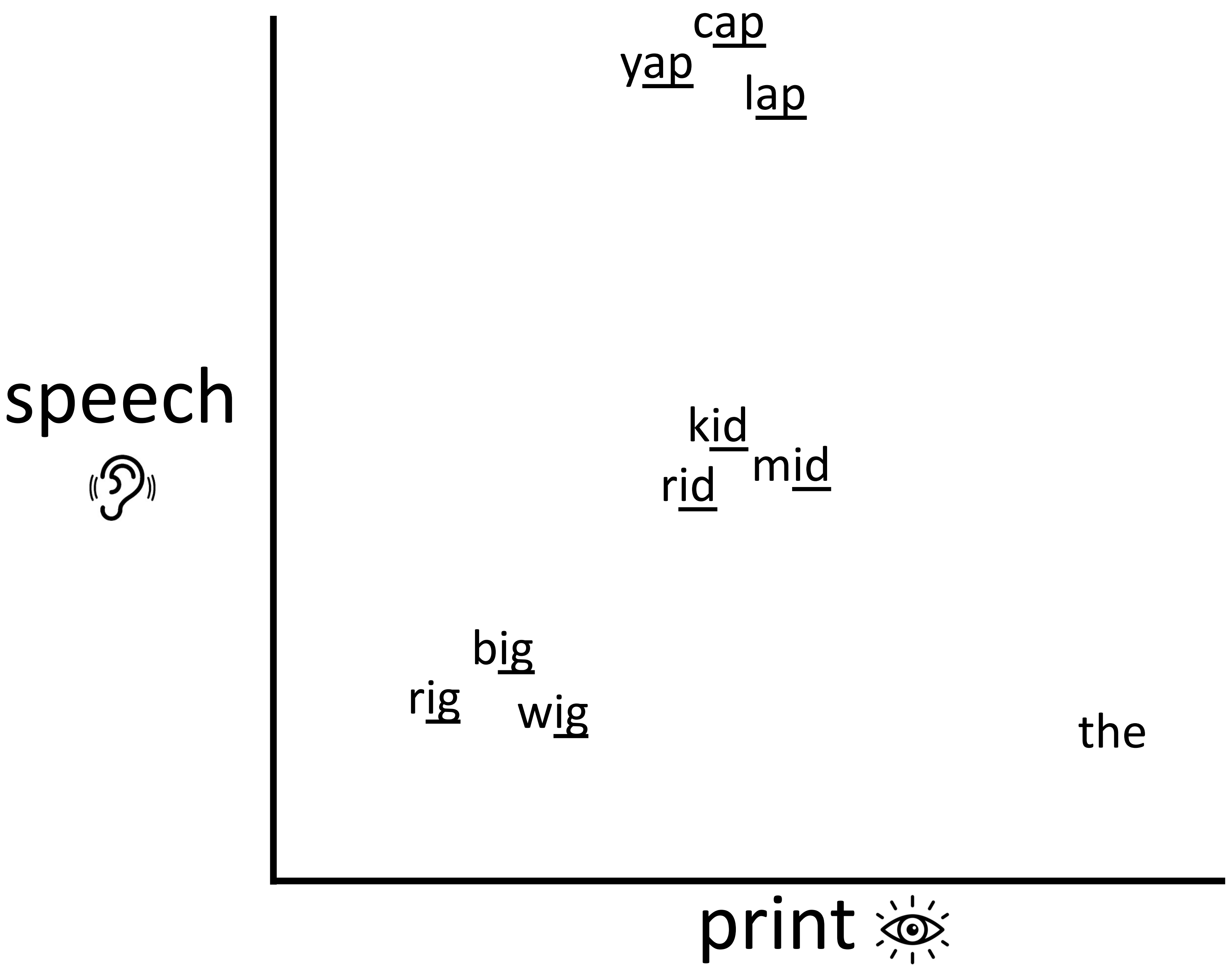
x	y
2	1
4	2
6	3
8	4



yacht

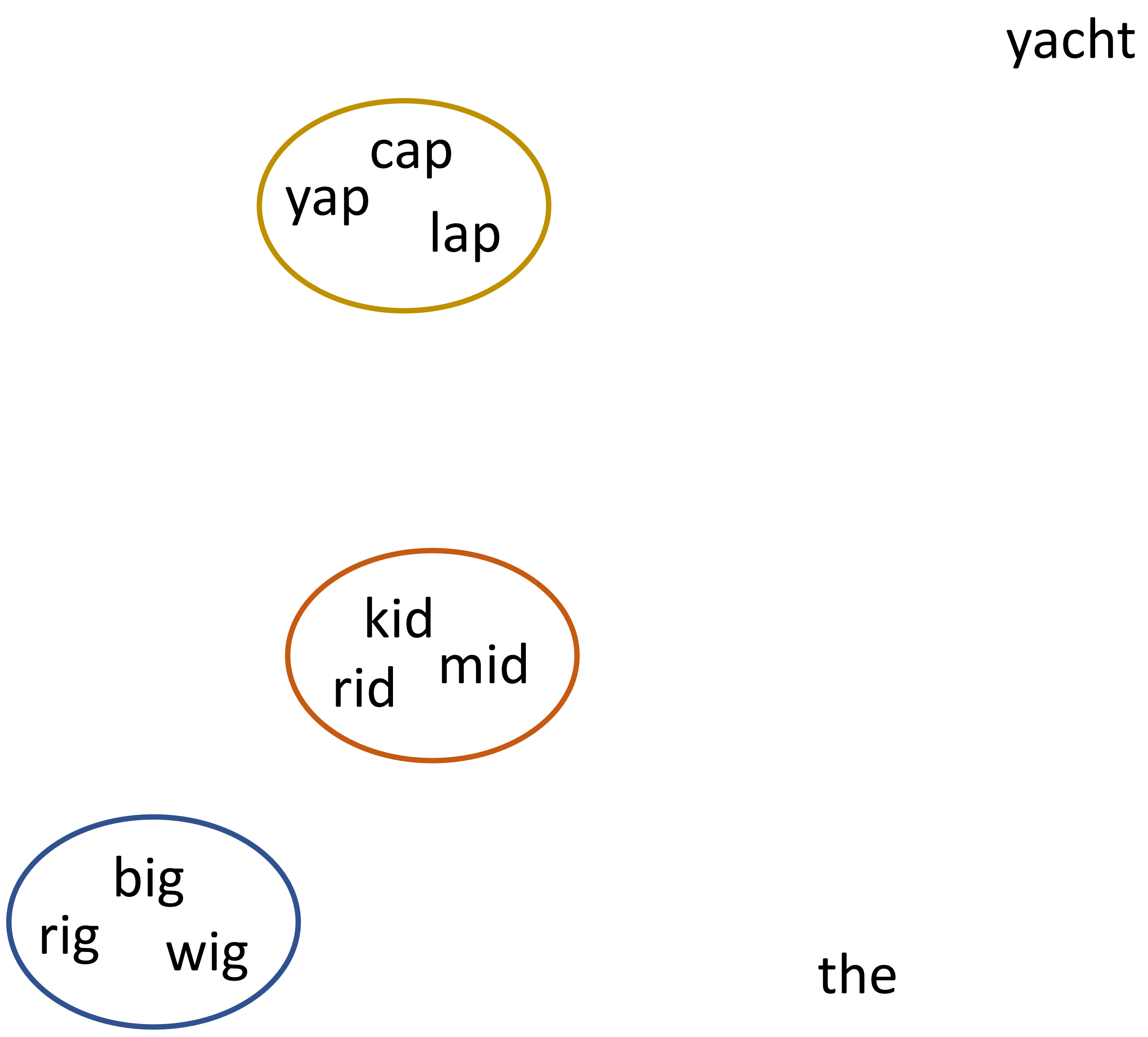
$f(x) = y$

print	speech
big	bɪg
the	ðʌ
kid	kɪd
cap	kæp



$f(x) = y$

print	speech
big	bɪg
the	ðʌ
kid	kɪd
cap	kæp



$f(x) = y$

print	speech
big	bɪg
the	ðʌ
kid	kɪd
cap	kæp

yacht

cap
yap
lap

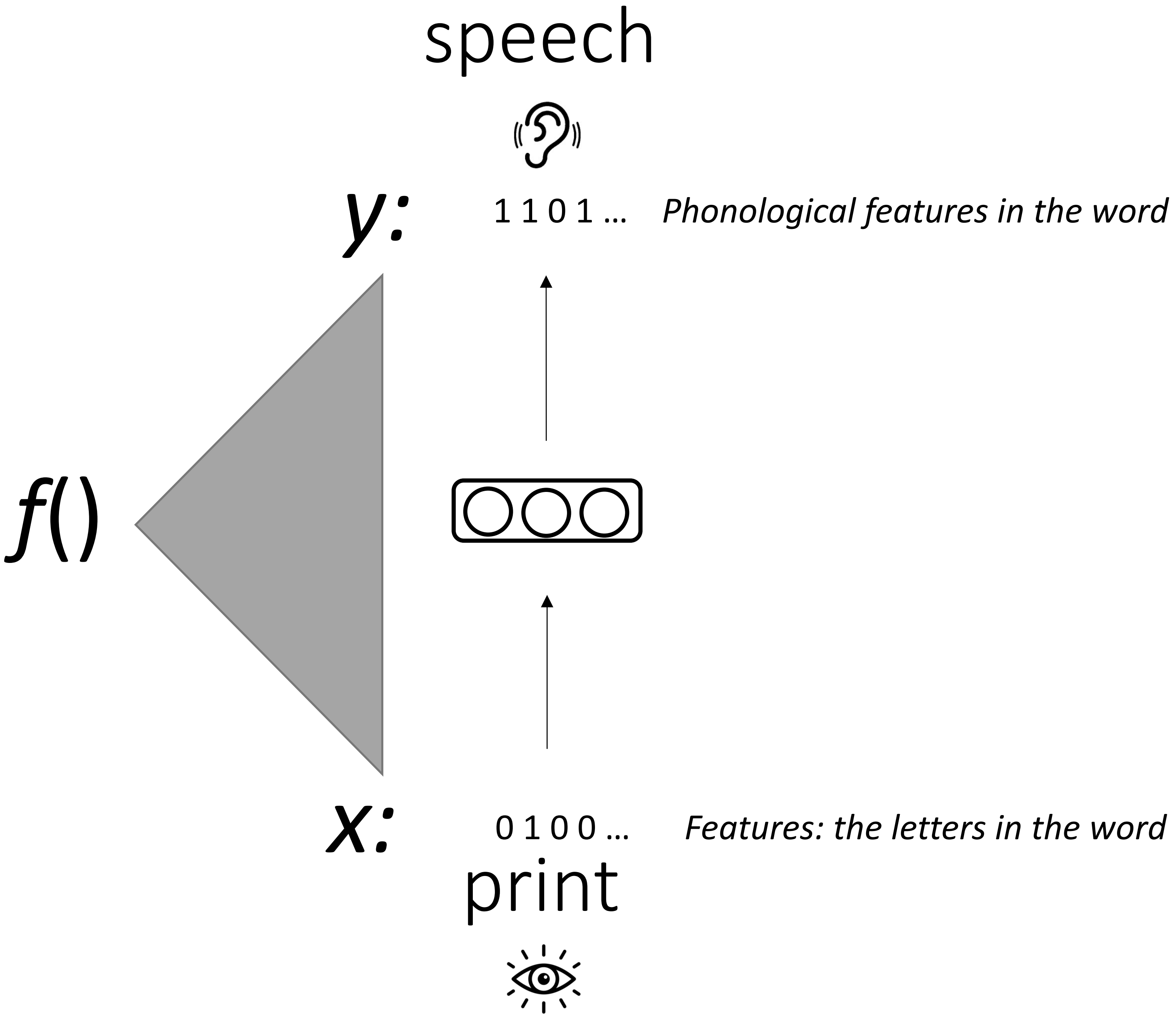
kid
rid
mid

big
rig
wig

the

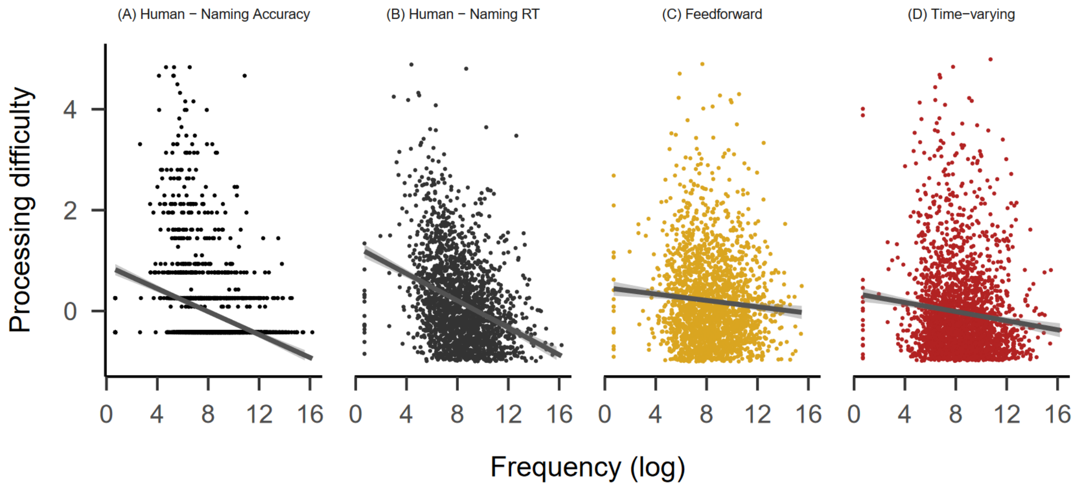
$f(x) = y$

print	speech
big	bɪg
the	ðʌ
kid	kɪd
cap	kæp



These models are established

- The behavior of these models: linked to human behavior
- Range of learning properties akin to those of children, adults
- Decades of research support in cognitive science



McClelland et al. (2010) *Trends in Cognitive Science*
Plaut, McClelland, Seidenberg, & Patterson (1996) *Psych Review*
Seidenberg & McClelland (1989) *Psych Review*
Seidenberg, Cooper Borkenhagen, & Kearns (2020) *RRQ*

Computational models

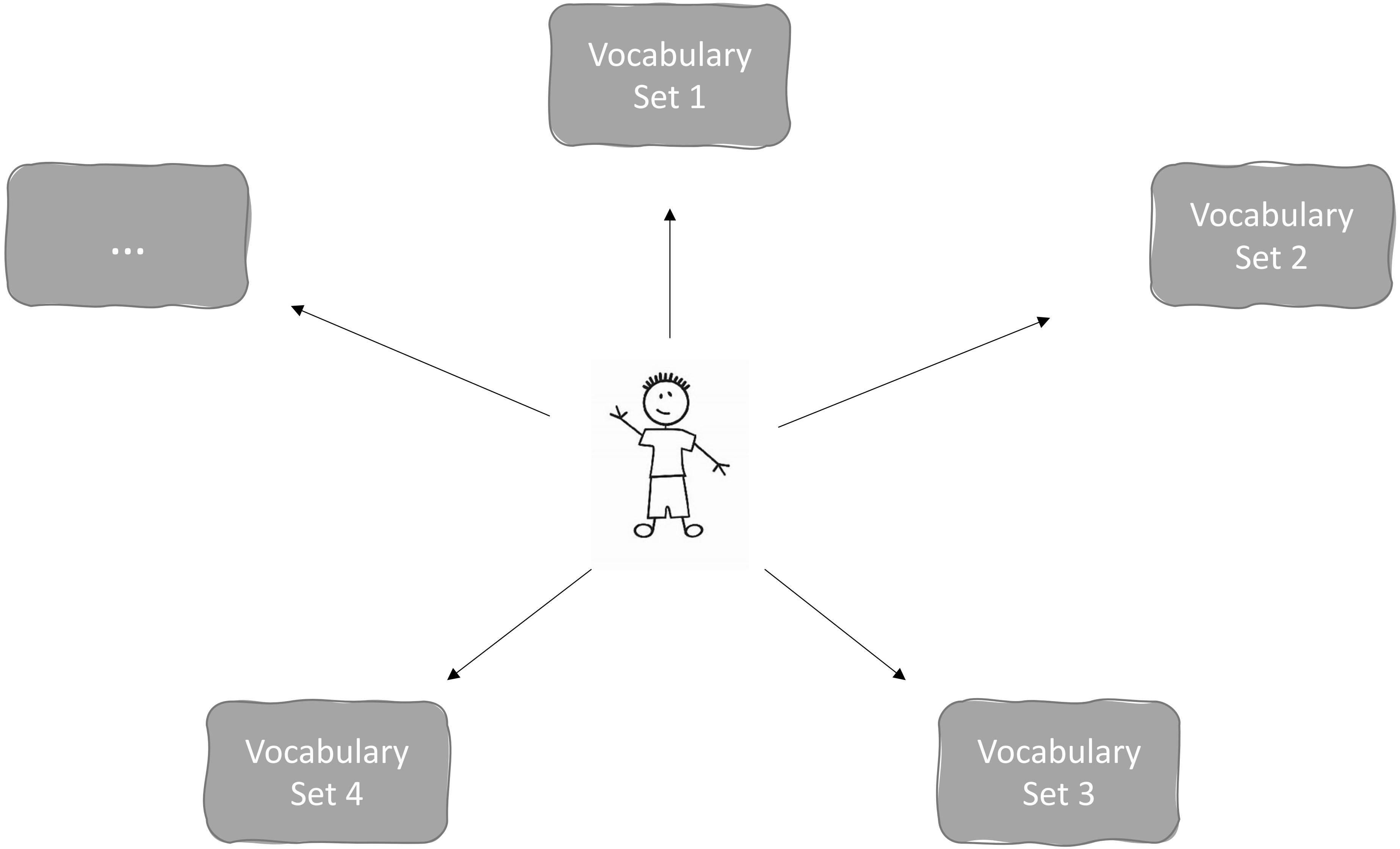
- Quite useful!
- You can poke at it, mess around with environment and other factors
- You can manipulate internal aspects of its structure
- Or its environment
- Experimentation that is not always possible with students
- And at large scale

Models can help us understand what to teach

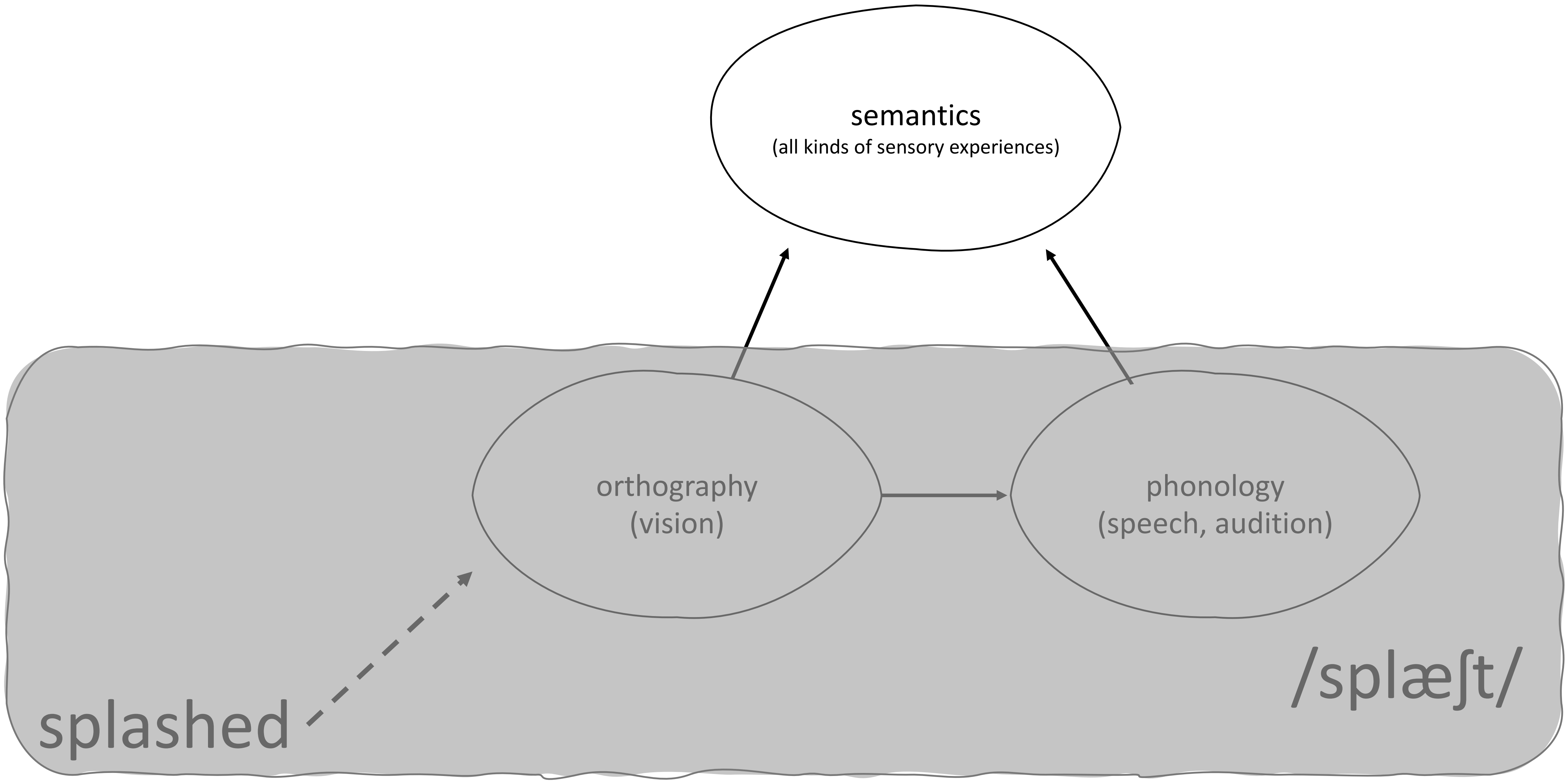
One approach:

- One model architecture, trained many times
- Vary: number of words learned
- Vary: which words are learned

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


Vocabulary Set 1
...
Vocabulary Set 2
Vocabulary Set 4
Vocabulary Set 3

< water, movement, action, past >



Manipulating how many words taught

Training set size:

- Ranged from 100 to 1000 (out of about 3000 total words)
- Increments of 100

Manipulating which words are taught

- Randomly select words for training sets of a given size
- Words used here: monosyllables
- For each of 10 training set sizes, do this 100,000 times
- One million models

Generalization set

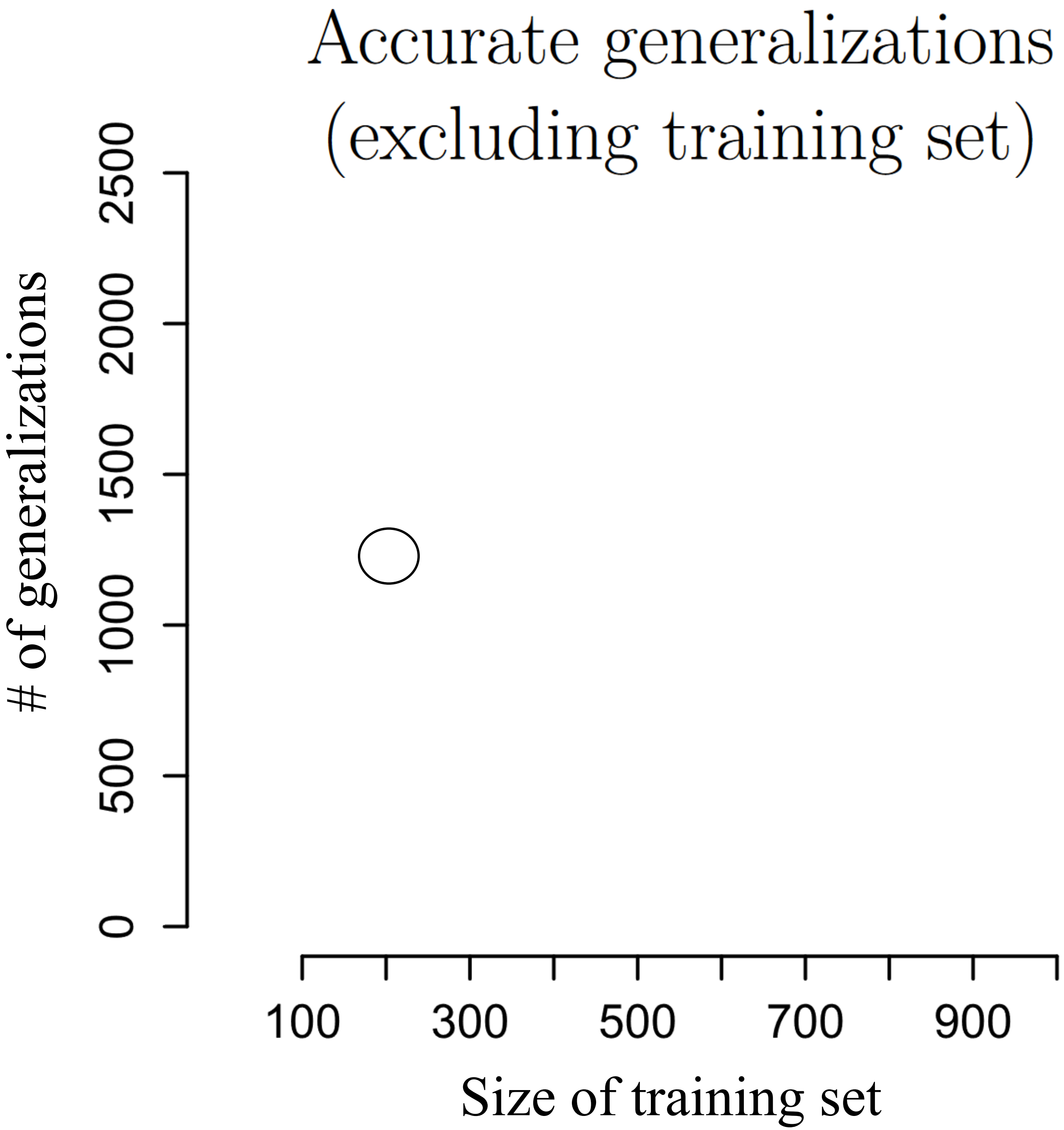
- All the words not selected for the training set
- More like what we ask of kids

Results

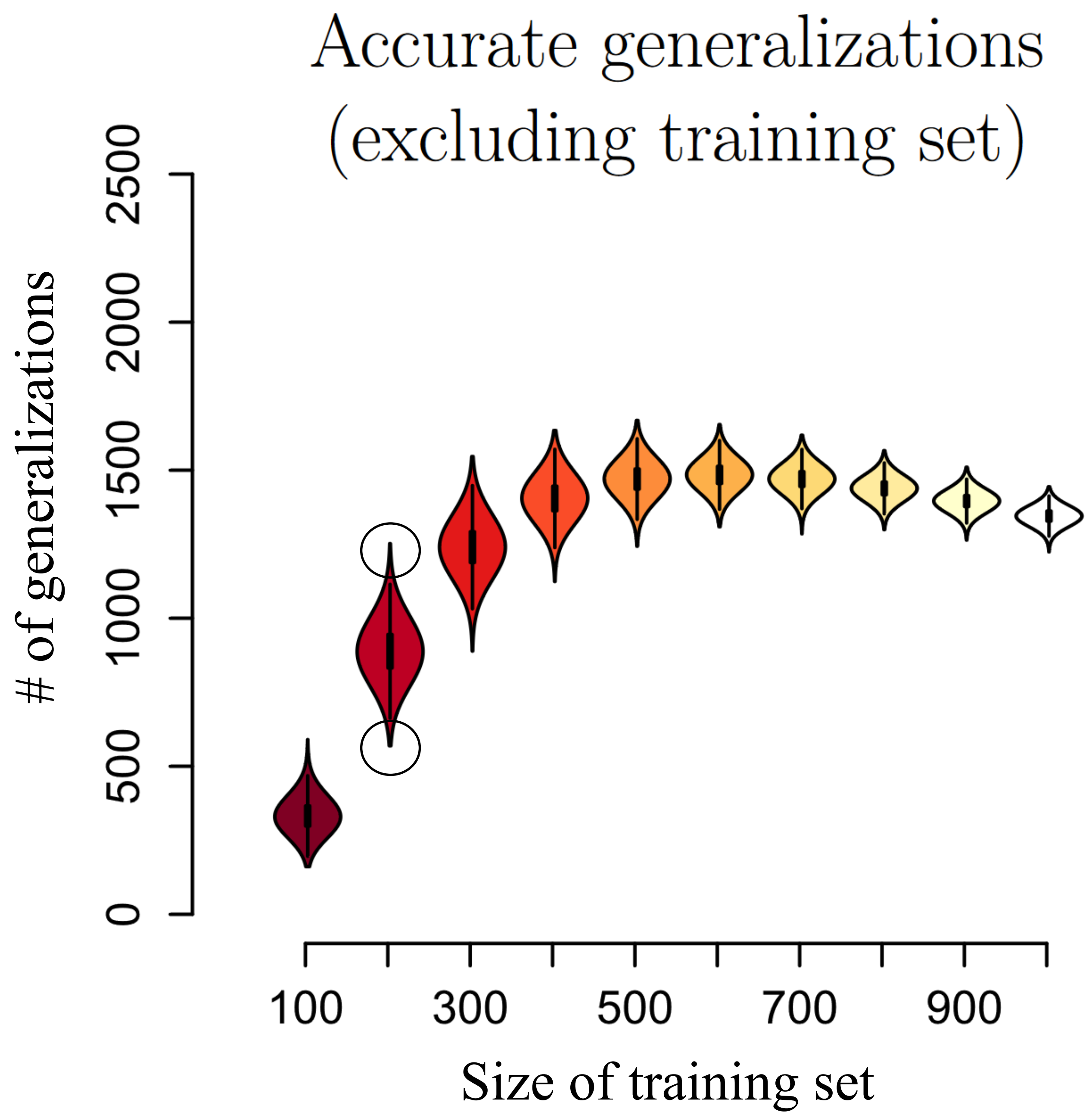
- All training sets were learned nearly perfectly
- Data focus on generalization accuracy
- Analysis:

Across models:

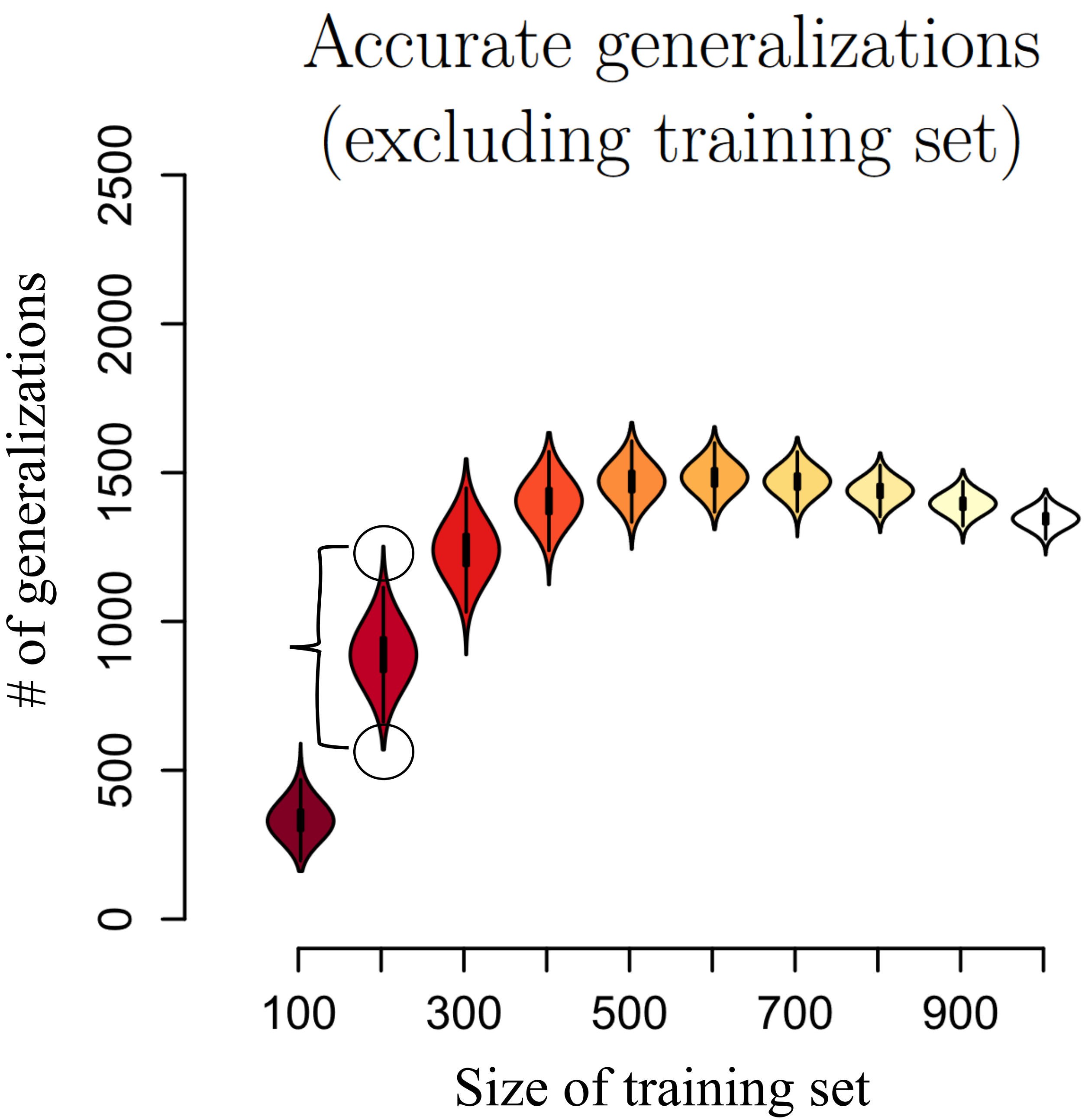
$$\frac{\#\ words\ learned\ via\ generalization}{total\ \#\ of\ words}$$



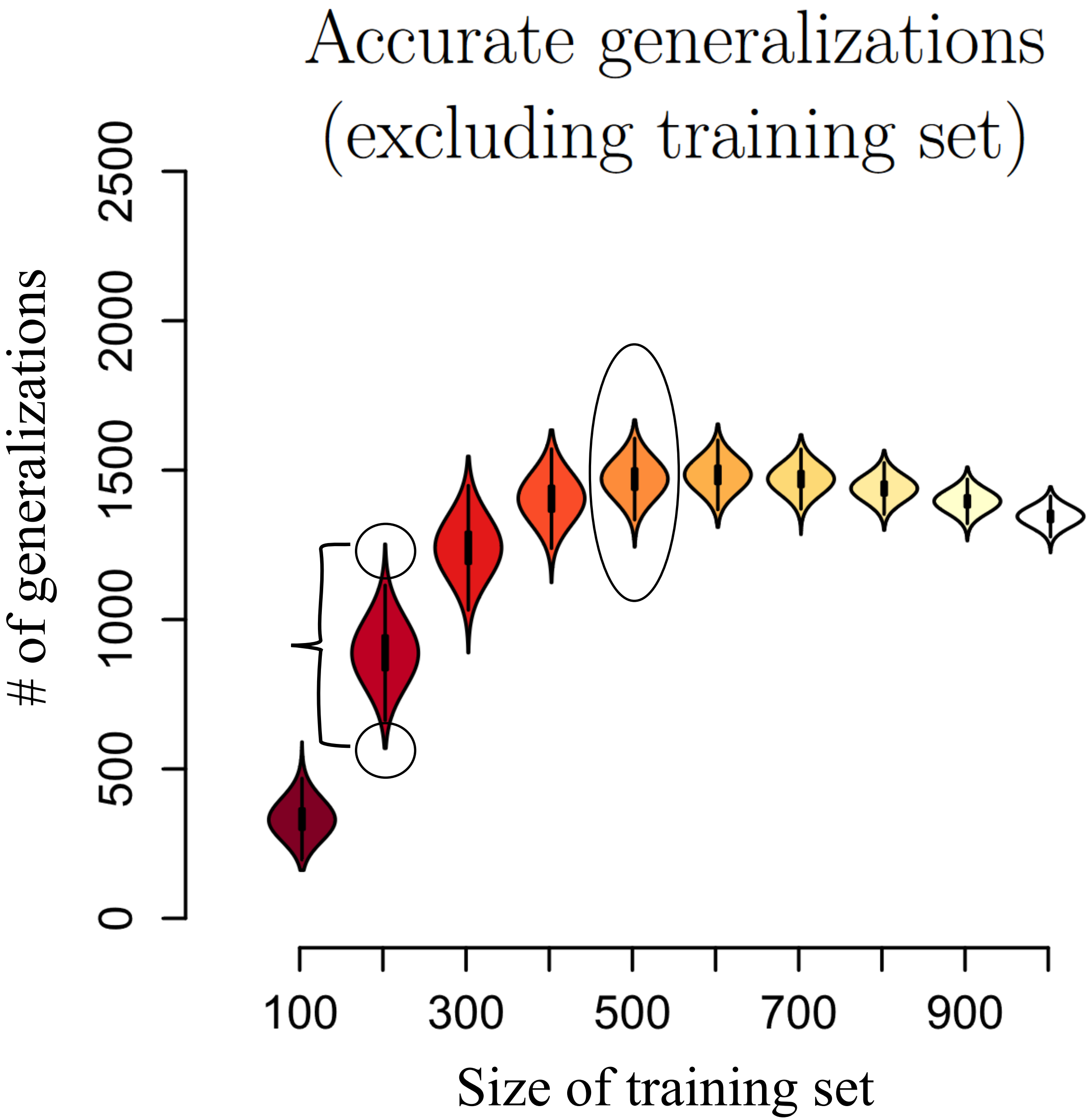
- Best models: high generalization (good)



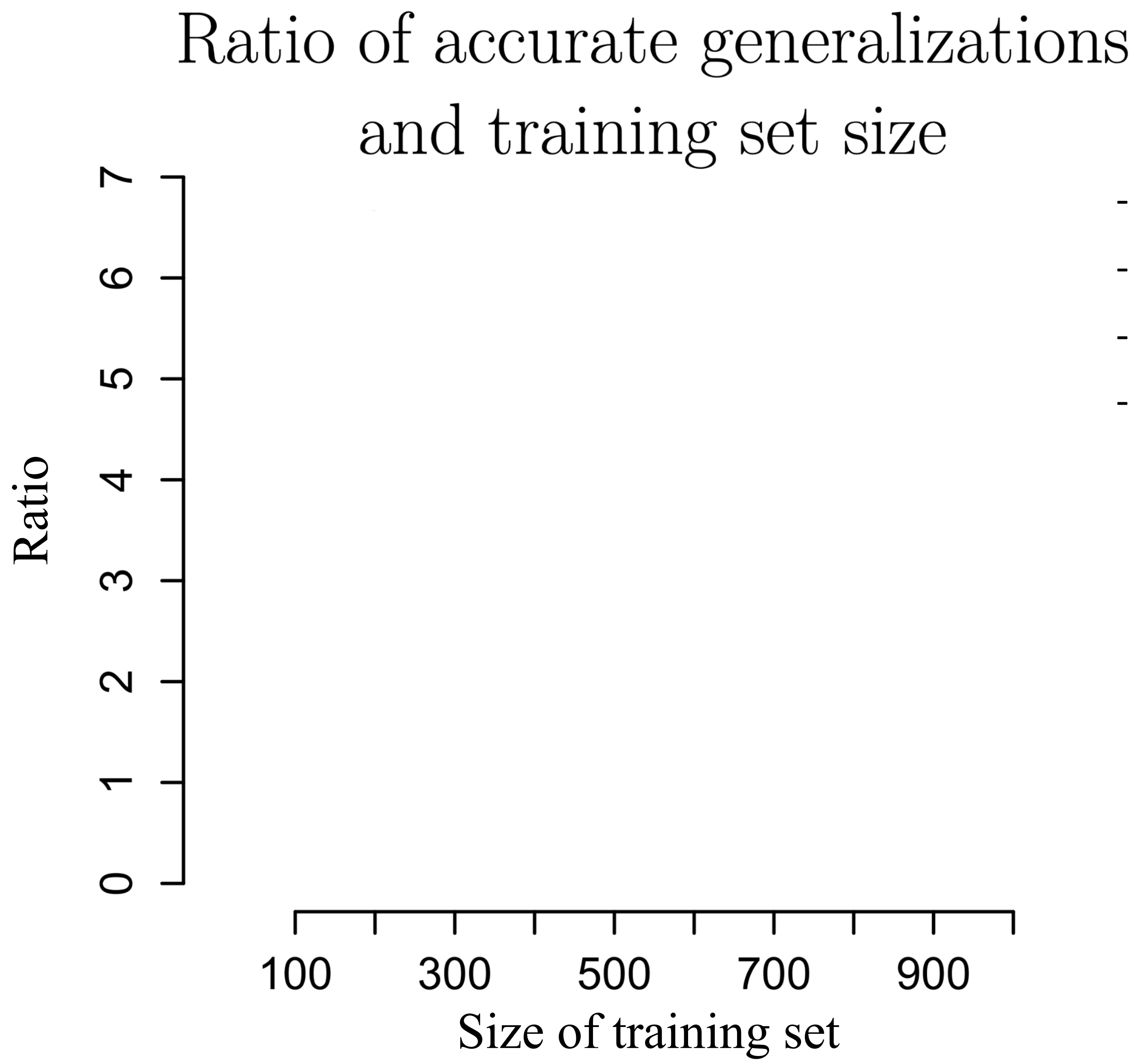
- Best models: high generalization (good)
- Worst models: low generalization (bad)



- Best models: high generalization (good)
- Worst models: low generalization (bad)
- Variation within training set size



- Best models: high generalization (good)
- Worst models: low generalization (bad)
- Variation within training set size
- Get generalization you need by 500



- Peak ratio within 200
- Most bang for your buck
- (in terms of generalization)
- Have limited time? *this is where you search*

What Characteristics of Words in a Training Set are Predictive of Generalization Accuracy

Model-level Variable	η_p^2	ΔR^2
Orthographic length		
Orthographic neighbors		
Phonological neighbors		
Consistency		

Consistent words are important (with a mix of factors)

Why? …*Generalization*

Contrast with old idea: get the *inconsistent* words

What about the order in which words are learned?

- Optimize the order in which words are taught
- Same type of computational model
- Take the best 200 words from the previous experiments
- Give the model a finite set of training trials (10K)
- Optimize: the probability of sampling each word
- That's basically what frequency captures
- Test: the words that aren't directly taught (at the end of training)
- Question: can an optimized set beat frequency-weighted sampling?

Sen, Cox, Cooper Borkenhagen, Seidenberg, & Zhu (2020) *ArXiv*

Trial 1: $Word_1$, $Word_{2,}$ $Word_3$ … $Word_{3000}$ $Word_x$ sampled $\longrightarrow$ Train model

T2: $Word_1$, $Word_{2,}$ $Word_3$ … $Word_{3000}$ $Word_x$ sampled $\longrightarrow$ Train model

T3: $Word_1$, $Word_{2,}$ $Word_3$ … $Word_{3000}$ $Word_x$ sampled $\longrightarrow$ Train model

…

T10K: $Word_1$, $Word_{2,}$ $Word_3$ … $Word_{3000}$ $Word_x$ sampled $\longrightarrow$ Train model

Trial 1: *Word*$_1$, *Word*$_2$, *Word*$_3$ … *Word*$_{3000}$ *Word*$_x$ sampled → Train model

T2: *Word*$_1$, *Word*$_2$, *Word*$_3$ … *Word*$_{3000}$ *Word*$_x$ sampled → Train model

T3: *Word*$_1$, *Word*$_2$, *Word*$_3$ … *Word*$_{3000}$ *Word*$_x$ sampled → Train model

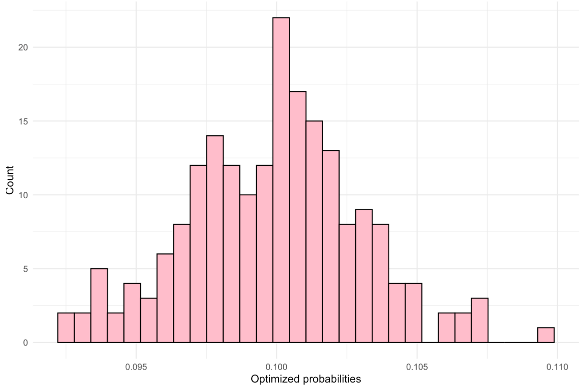
…

T10K: *Word*$_1$, *Word*$_2$, *Word*$_3$ … *Word*$_{3000}$ *Word*$_x$ sampled → Train model

Optimized probabilities - in two flavors

(Just a fancy way of finding a good word “frequency”)

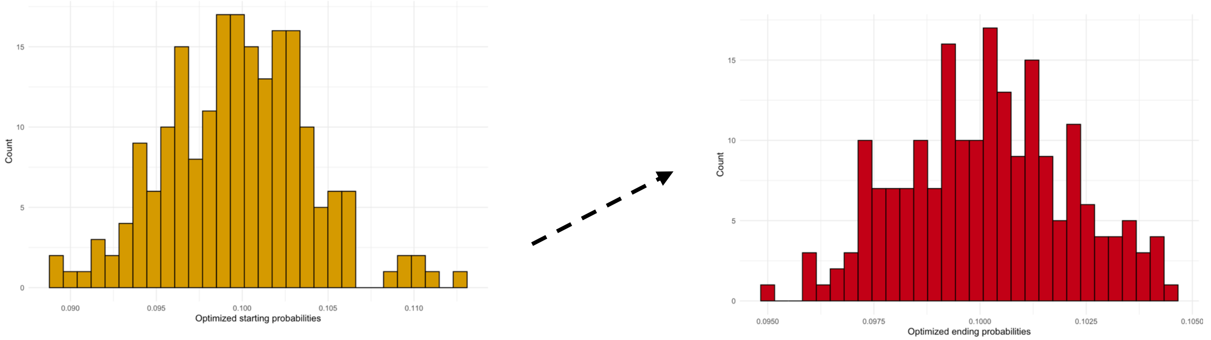
Condition 1: A single set of frequencies



= one set of probabilities throughout learning

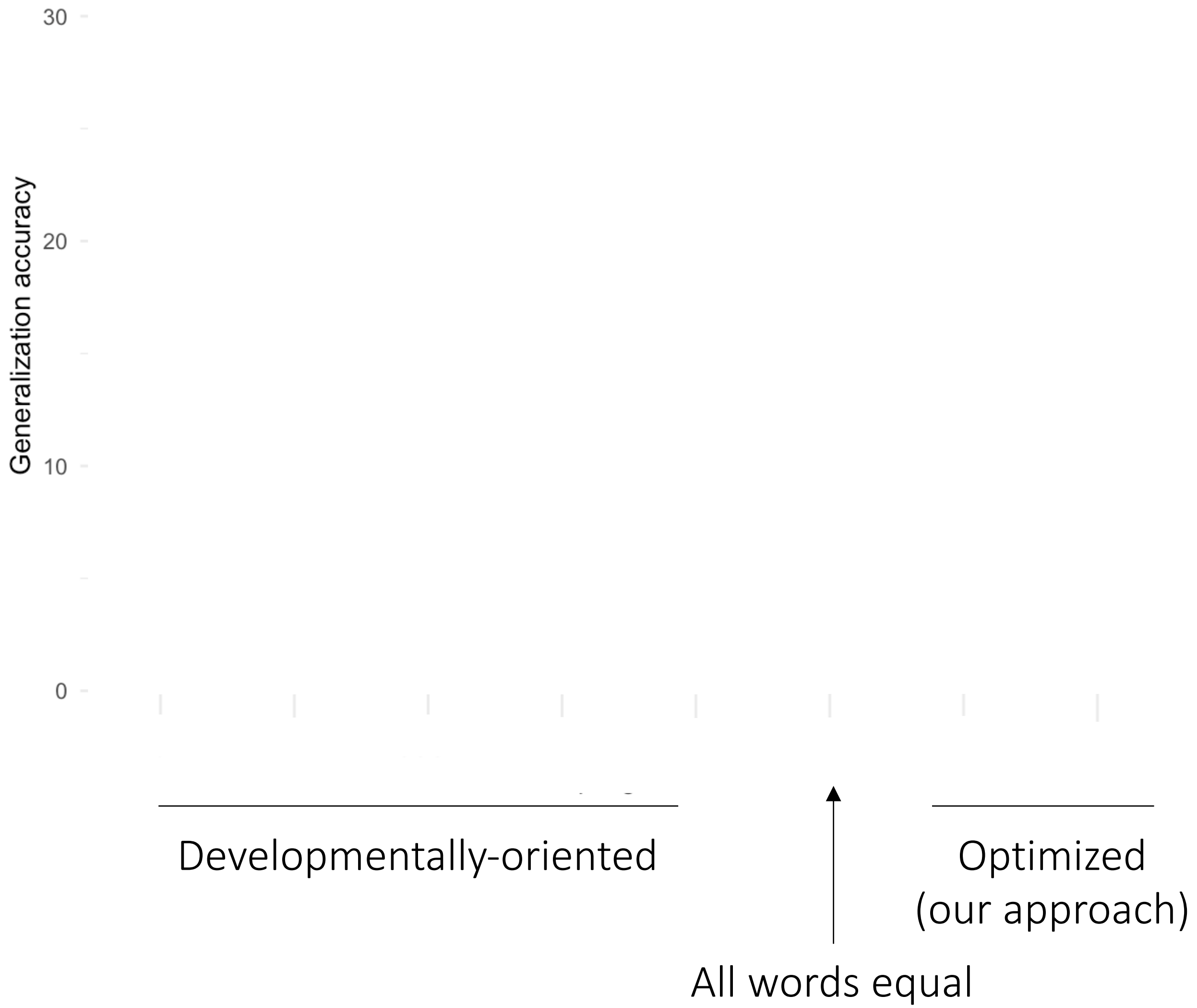
“+”

Condition 2: The frequencies changes start to finish



= different probabilities at start + end of learning

“++”



30
20
10
0
Generalization accuracy
Developmentally-oriented
All words equal
Optimized
(our approach)

Sequential optimization

Optimized sampling…

- Better than various frequency-weighted approaches
- Better than randomly picking (uniform)
- The order in which we teach words can benefit learning

A new approach to special words

- Start with a theoretical model of learning
- Implement it in computational form
- Teach it words, find what is good for learning (the special words part)
- Result: a tractable and scalable alternative (to orthodox ED theories)
- Complements experimentation with kids in lab, classroom

You end up with something different than the old idea / educational orthodoxy

The takeaway

- Variable practices in the classroom (outcomes?)
- This results from application of outdated ideas; bad translation
- Psychology has theories too! (what we do)
- Computational theories can contribute
- ID words that benefit learning, generalization
- Help us specify what to teach, when to teach it
- More granular solutions/ research on the horizon
- The basic theory + method ingredients are there to help

The future

- Beyond simple words
- Involves more complex + exciting computational solutions, resources
- Modeling variability in language background - knowledge – skill
- Implementation in classroom
- Application of more advanced quantitative methods

100 Sight Words

the	he	at	but	there
of	was	be	not	use
and	for	this	what	an
a	on	have	all	each
to	are	from	were	which
in	as	or	we	she
is	with	one	when	do
you	his	had	your	how
that	they	by	can	their
it	I	words	said	if

will	some	two	my	find
up	her	more	than	long
other	would	write	first	down
about	make	go	water	day
out	like	see	been	did
many	him	number	call	get
then	into	no	who	come
them	time	way	am	made
these	has	could	its	may
so	look	people	now	part

smells	strike	fox	drake	field
smart	seemed	trick	close	dodge
lot	chip	which	grave	dear
hoot	shelves	splashed	beds	smirk
course	crush	crack	mow	sharp
phrase	tells	grey	flies	spain
got	trace	guts	your	need
kid	cake	wool	eyed	ripped
ping	earn	stepped	ropes	kind
plan	check	jot	from	faced
rocked	rake	tuck	path	strength
woo	stair	weep	blast	plop
would	swan	stops	wow	shines
start	those	fair	doze	suds
plum	babe	break	sty	lets
want	screamed	dog	ton	toed
curls	boss	saw	star	paint
did	sound	waist	bar	rests
fan	tube	grip	them	build
rack	cole	fields	chairs	head

Thank you

Collaborators
Christopher R. Cox (LSU)
Ayon Sen (Meta)
Mark S. Seidenberg (UW Madison)
Jerry Zhu (UW Madison)
Lauren Schilling (UW Madison)

Language & Cognitive Neuroscience Lab
Production · Comprehension · Reading · Dyslexia · Behavior · Brain · Development

Funding
Work supported by the Vilas Trust at the UW-Madison and the Institute of Education Sciences, US Department of Education through Award #R305B150003 to UW-Madison. The opinions expressed here are those of the authors and do not represent views of the US Department of Education.

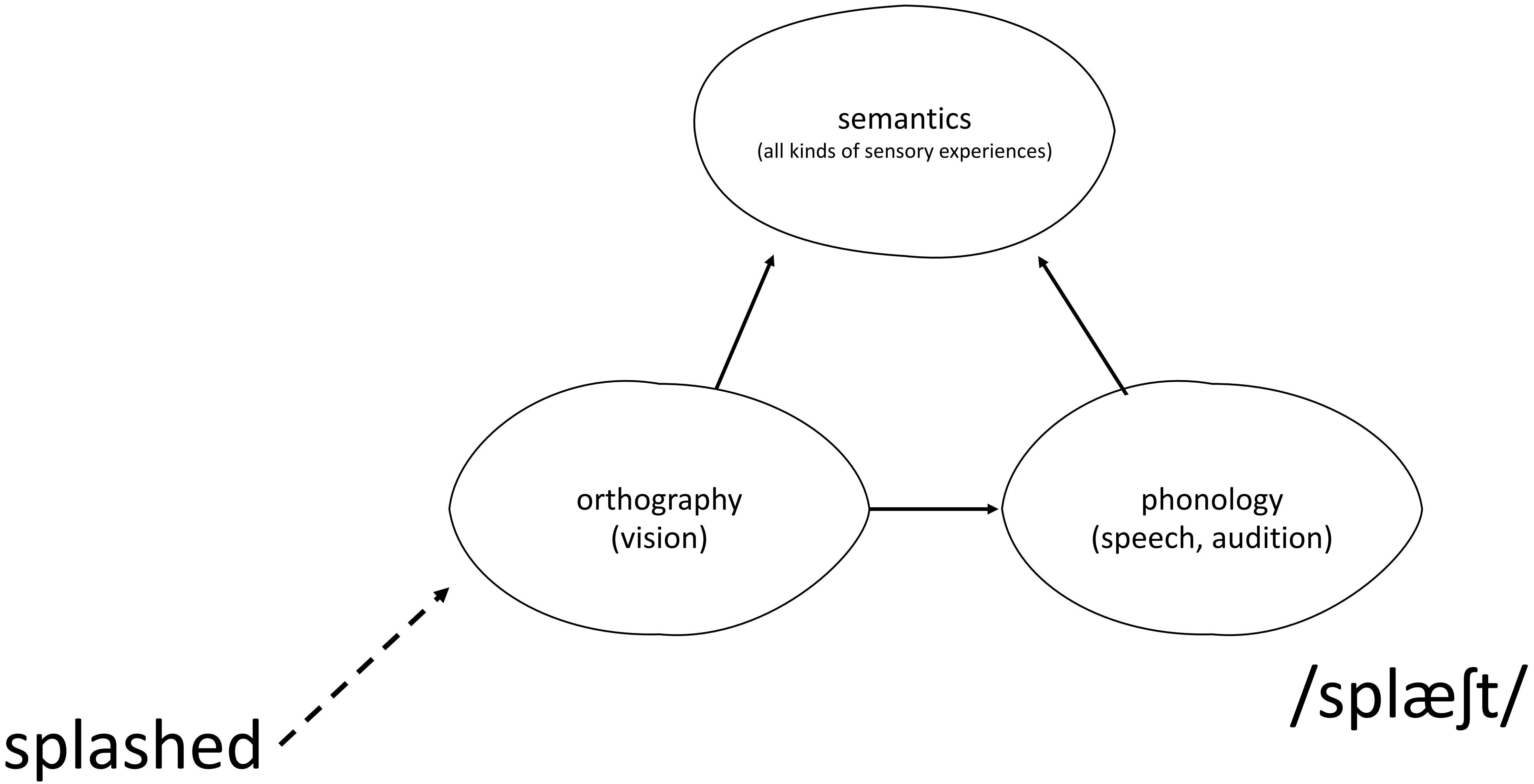
ITP Interdisciplinary Training Program in Education Sciences

ies INSTITUTE OF EDUCATION SCIENCES

WISCONSIN UNIVERSITY OF WISCONSIN–MADISON

Extra stuff

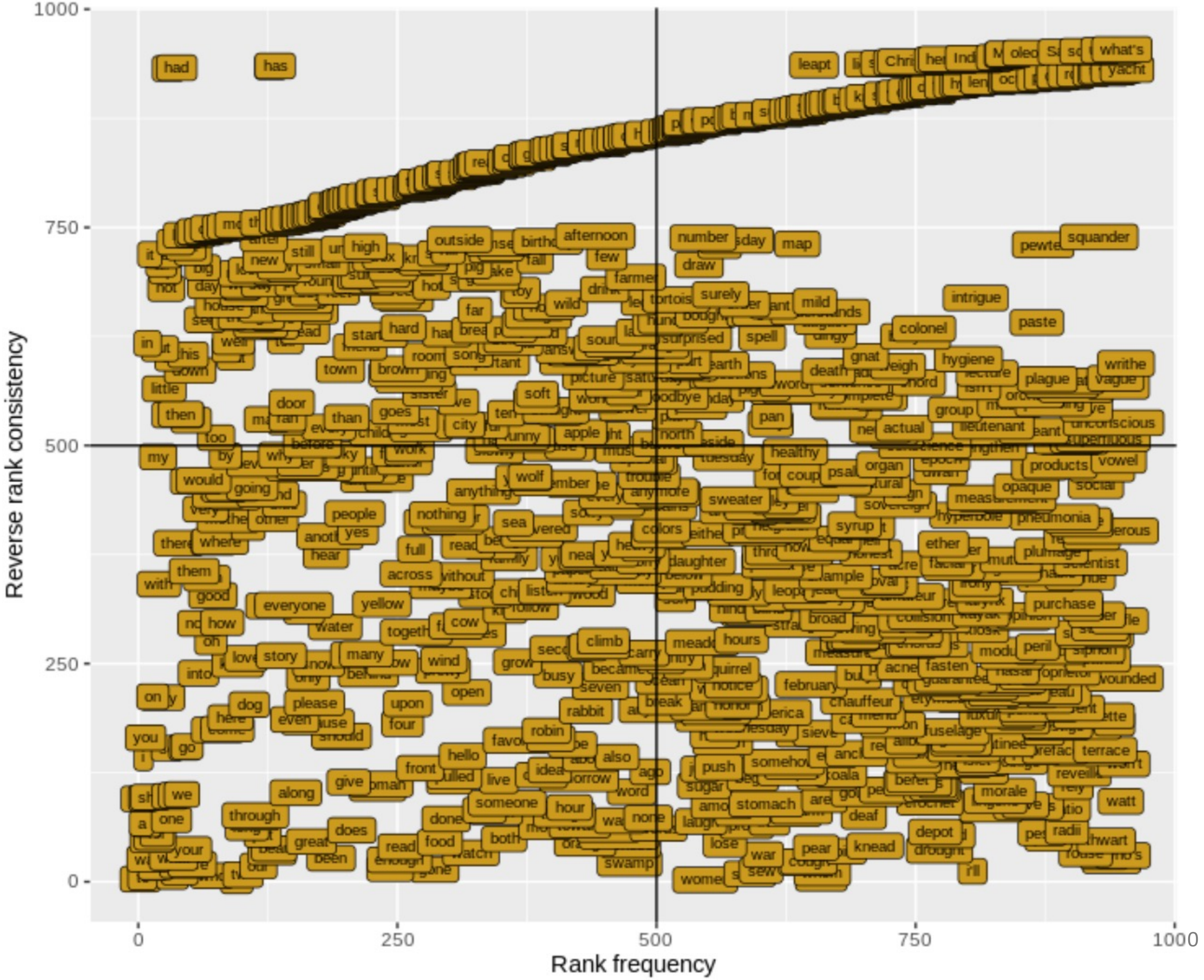
< water, movement, action, past >



All words across 5 programs

High frequency +
Low consistency

Low frequency +
Low consistency



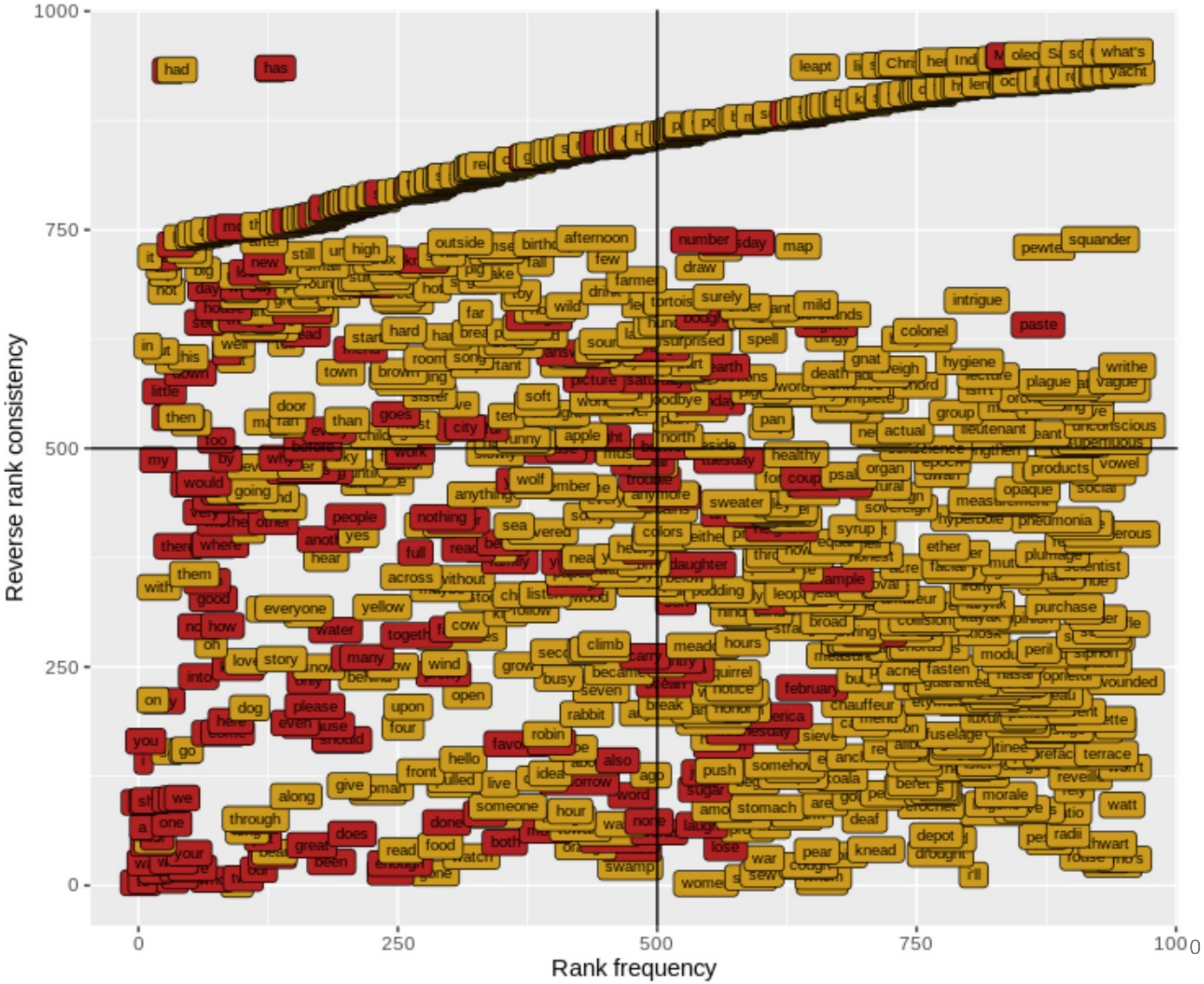
High frequency +
High consistency

Low frequency +
High consistency

Fundations words (in red)

High frequency +
Low consistency

Low frequency +
Low consistency



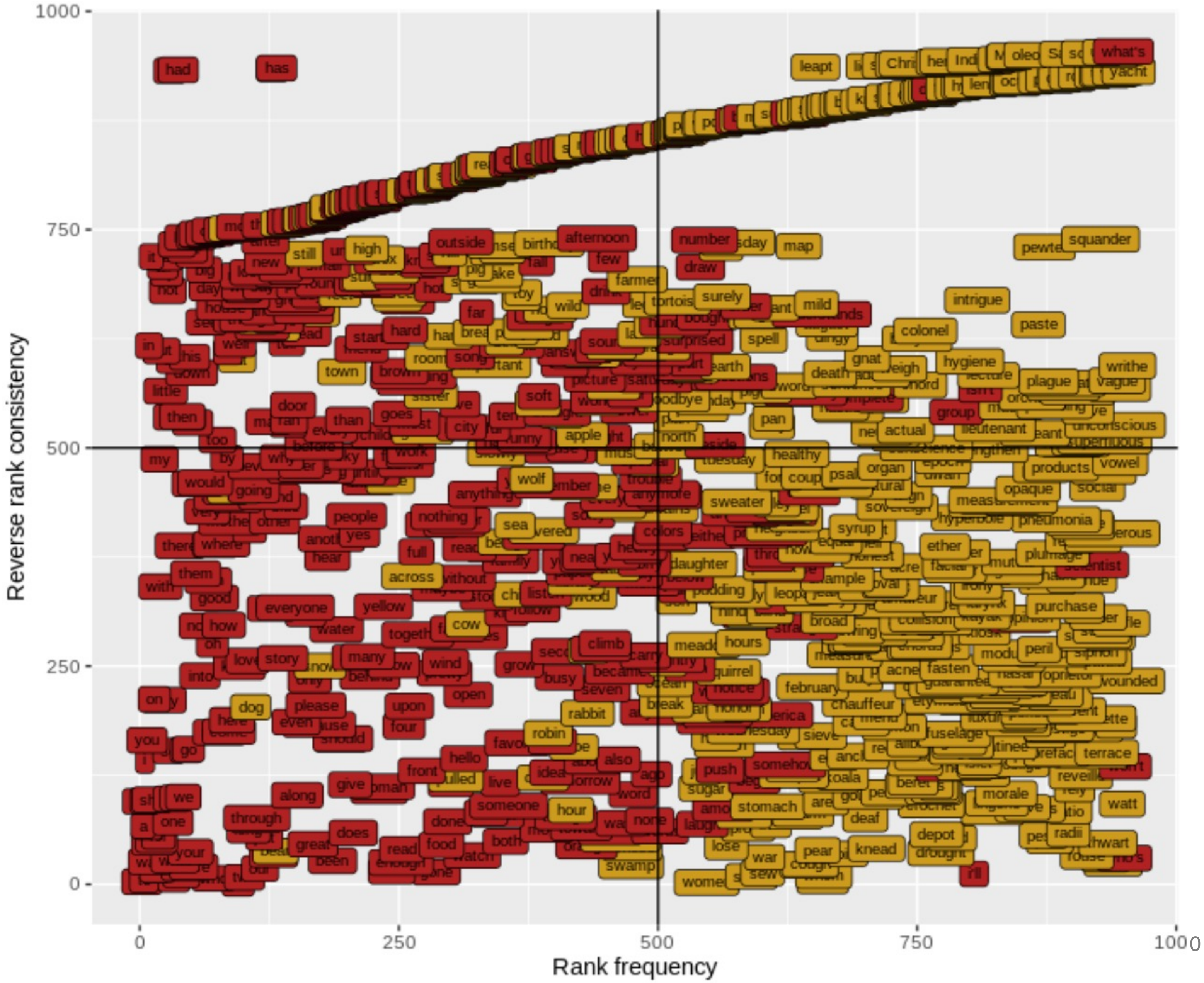
High frequency +
High consistency

Low frequency +
High consistency

Wonders words (in red)

High frequency +
Low consistency

Low frequency +
Low consistency



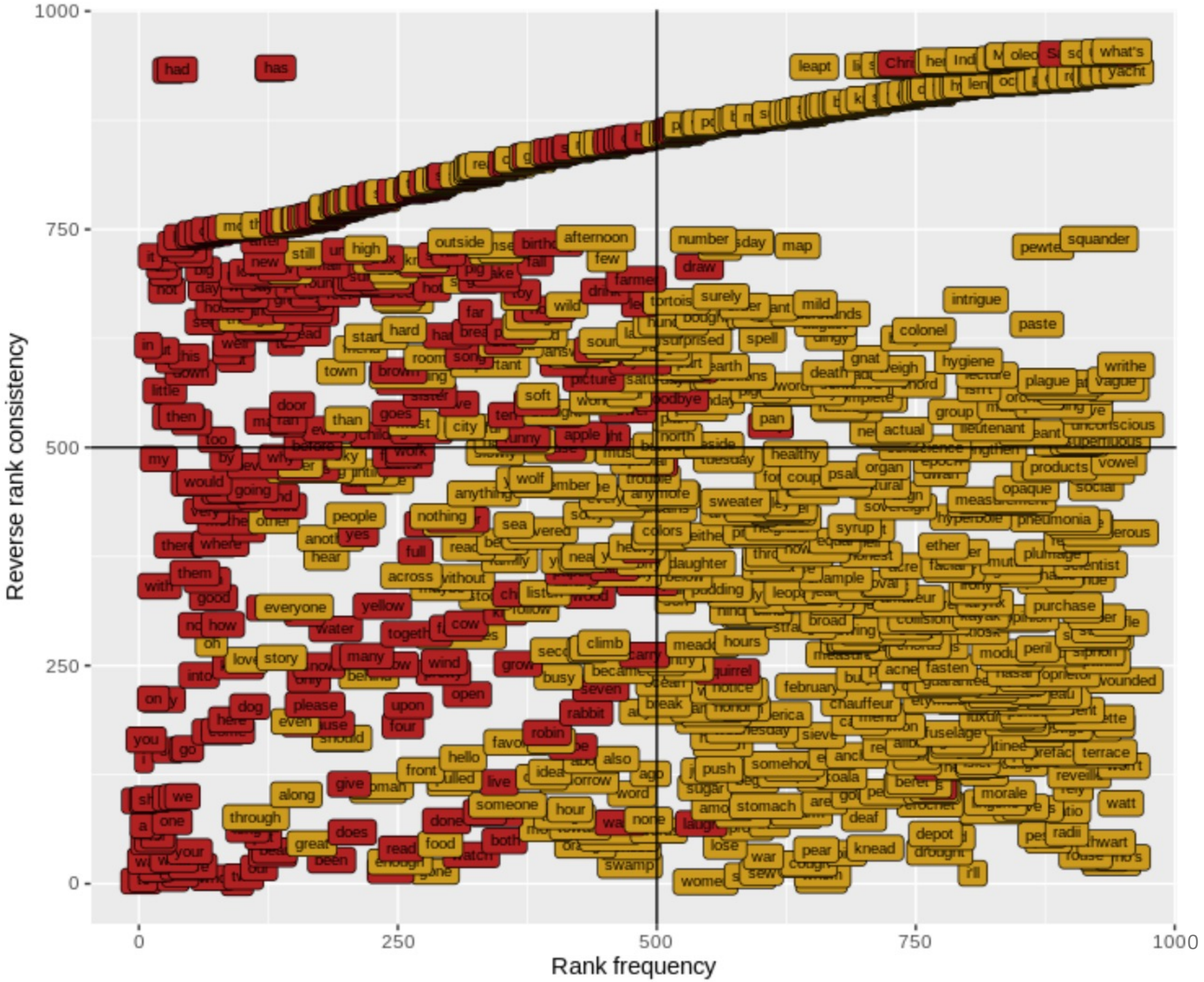
High frequency +
High consistency

Low frequency +
High consistency

Dolch words (in red)

High frequency +
Low consistency

Low frequency +
Low consistency



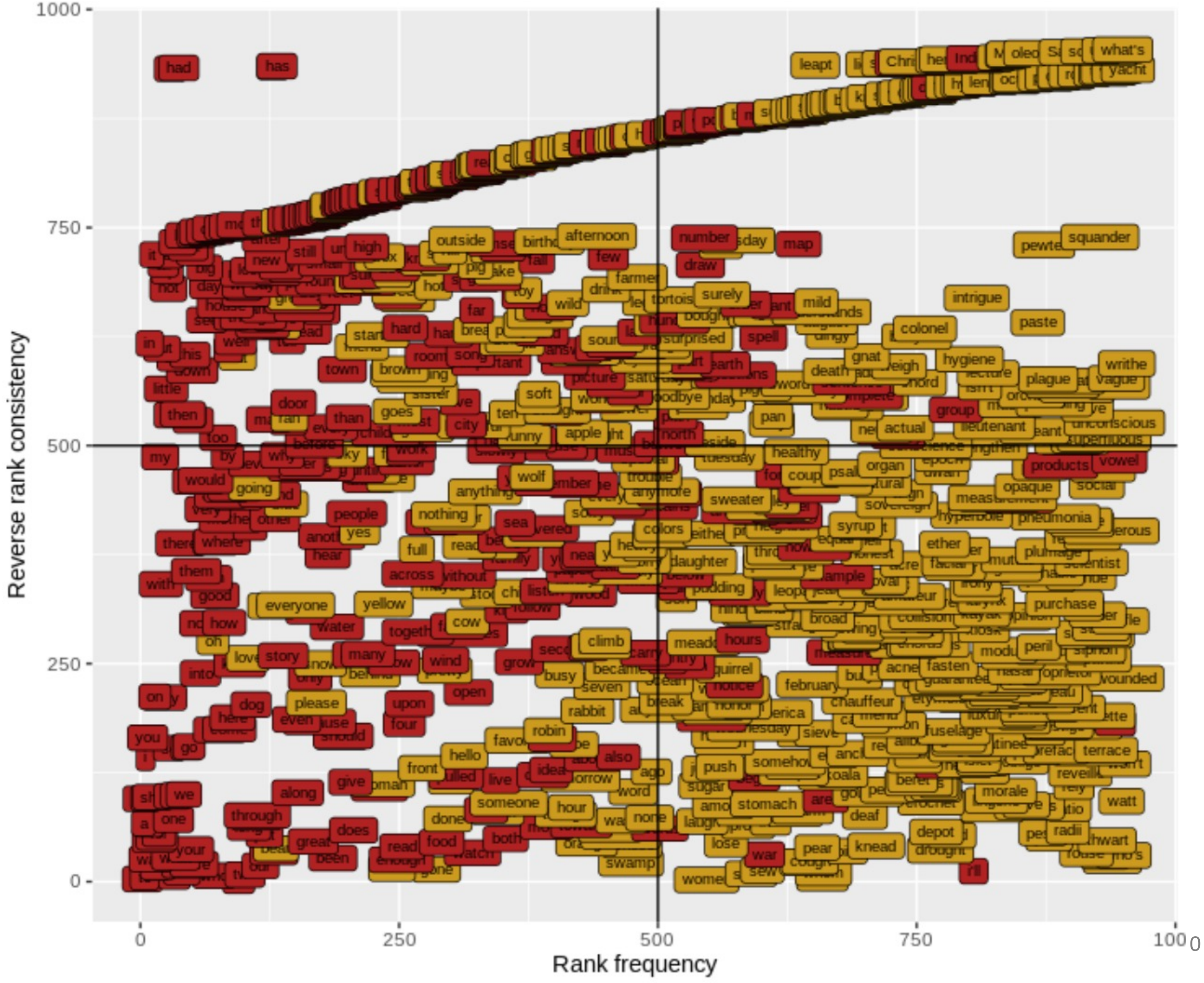
High frequency +
High consistency

Low frequency +
High consistency

Fry words (in red)

High frequency +
Low consistency

Low frequency +
Low consistency



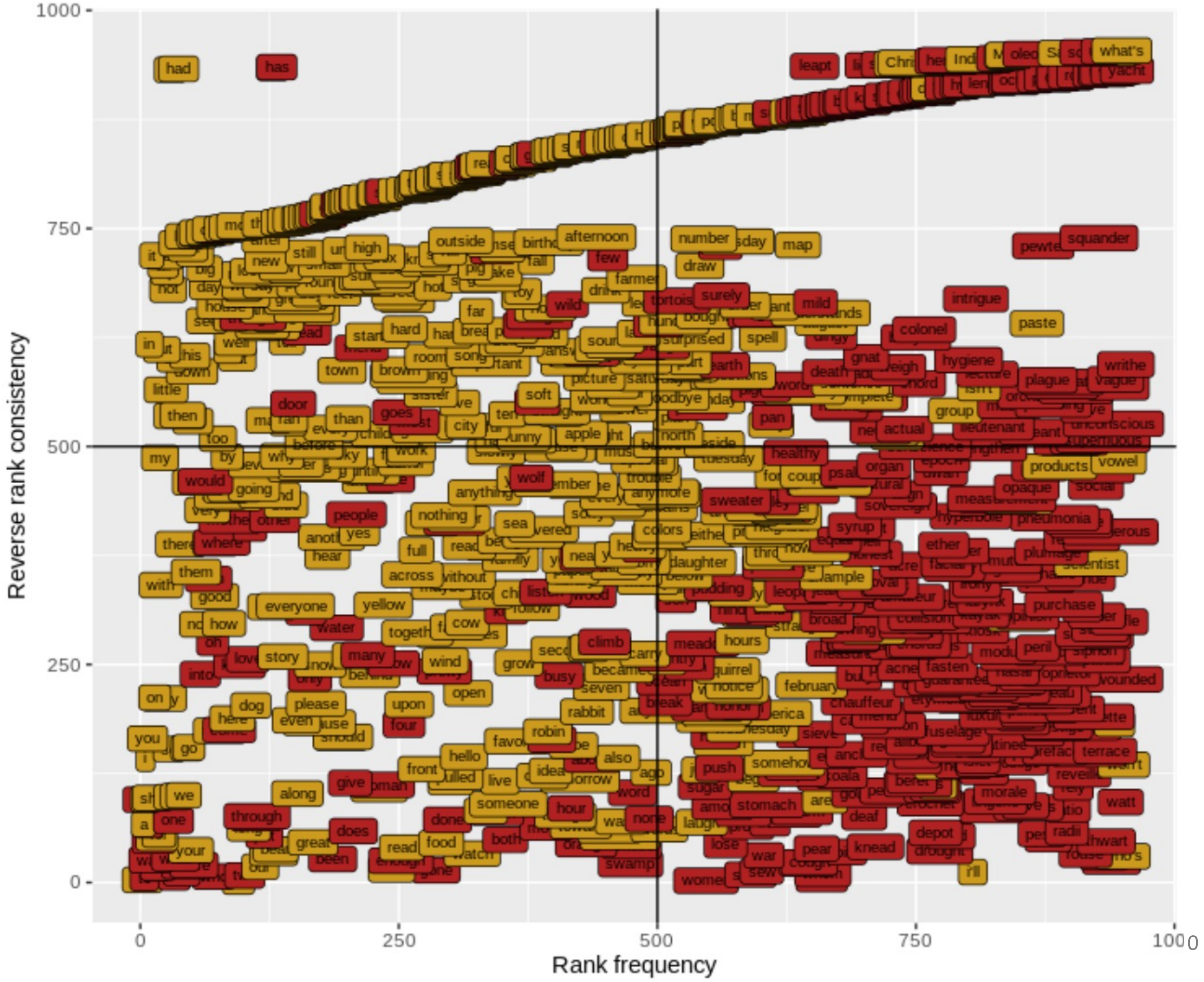
High frequency +
High consistency

Low frequency +
High consistency

Kilpatrick words (in red)

High frequency +
Low consistency

Low frequency +
Low consistency

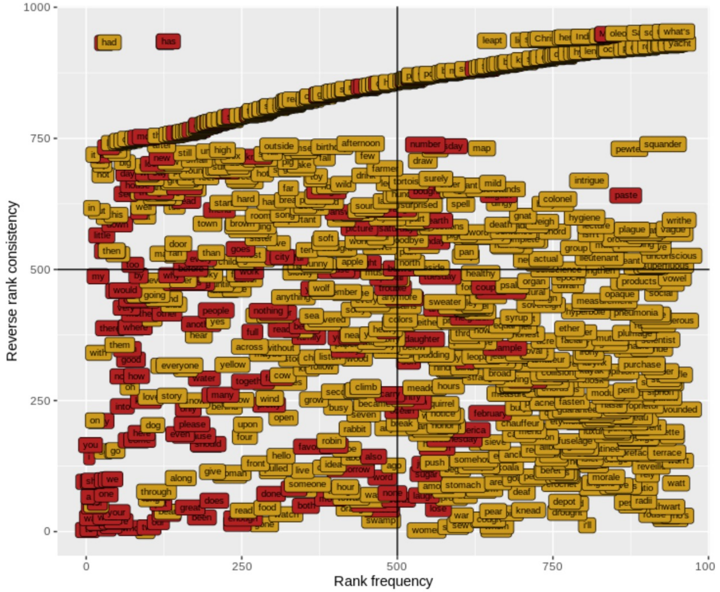


High frequency +
High consistency

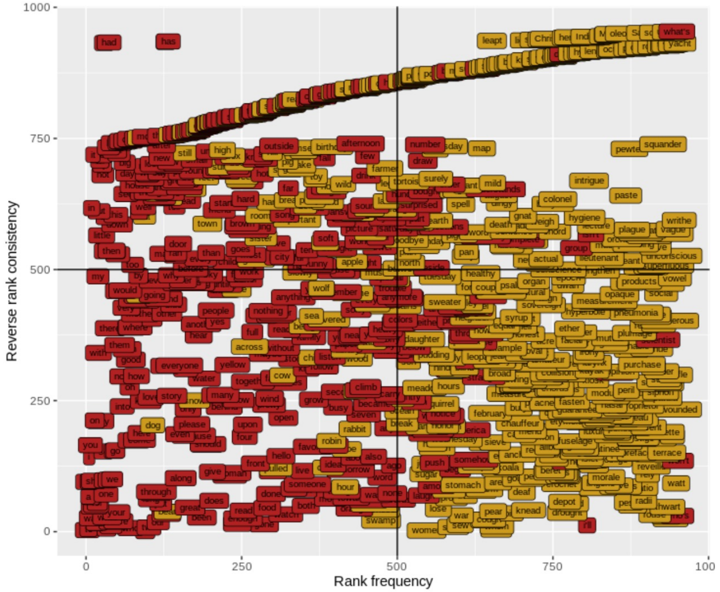
Low frequency +
High consistency

HF+LC	LF+LC
HF+HC	LF+HC

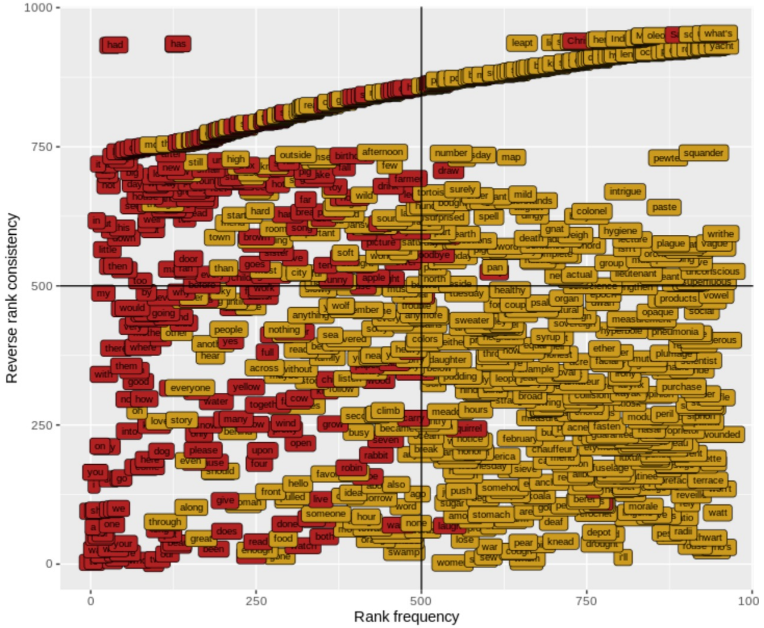
Fundations



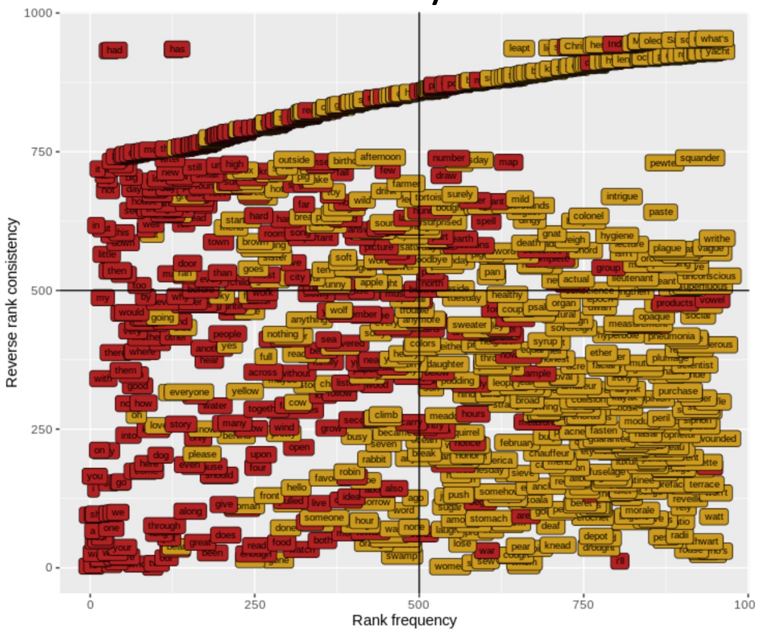
Wonders



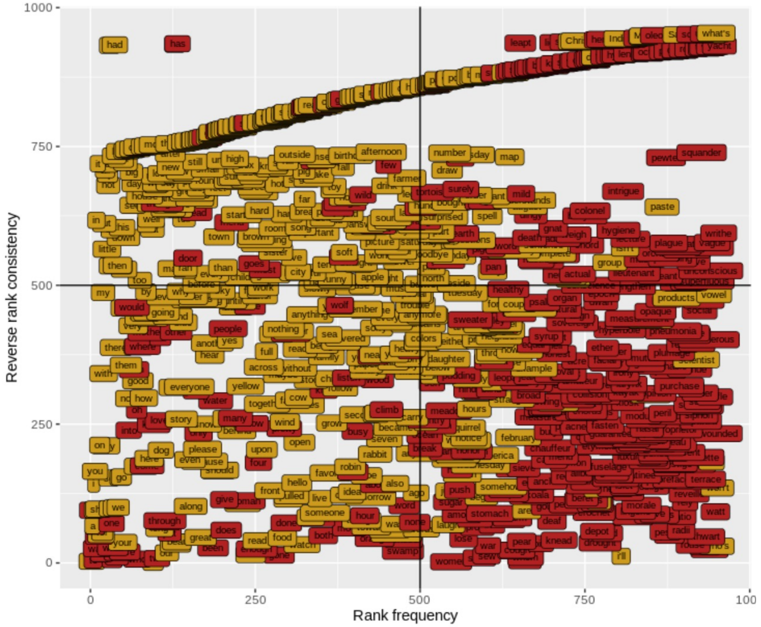
Dolch



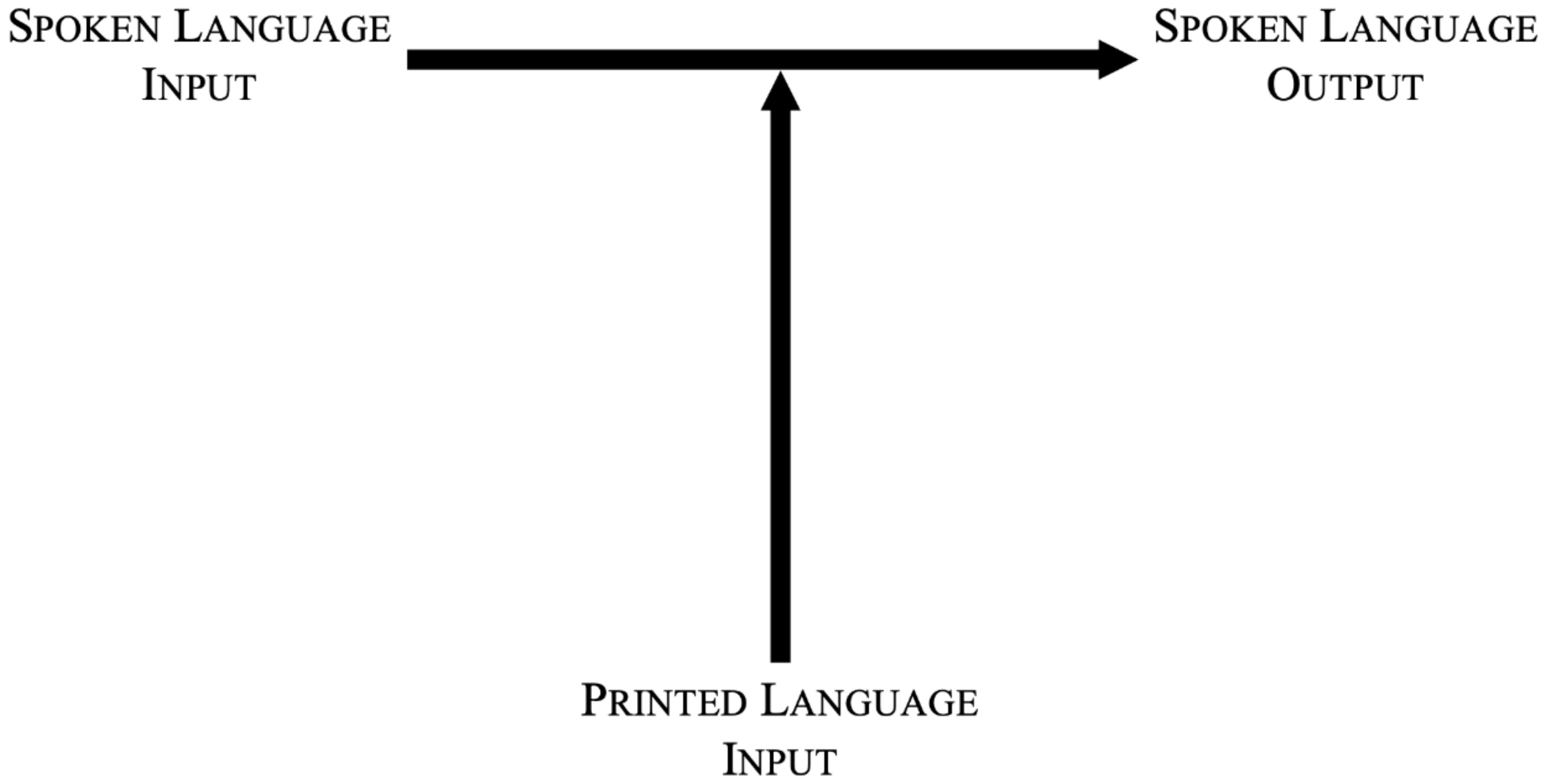
Fry



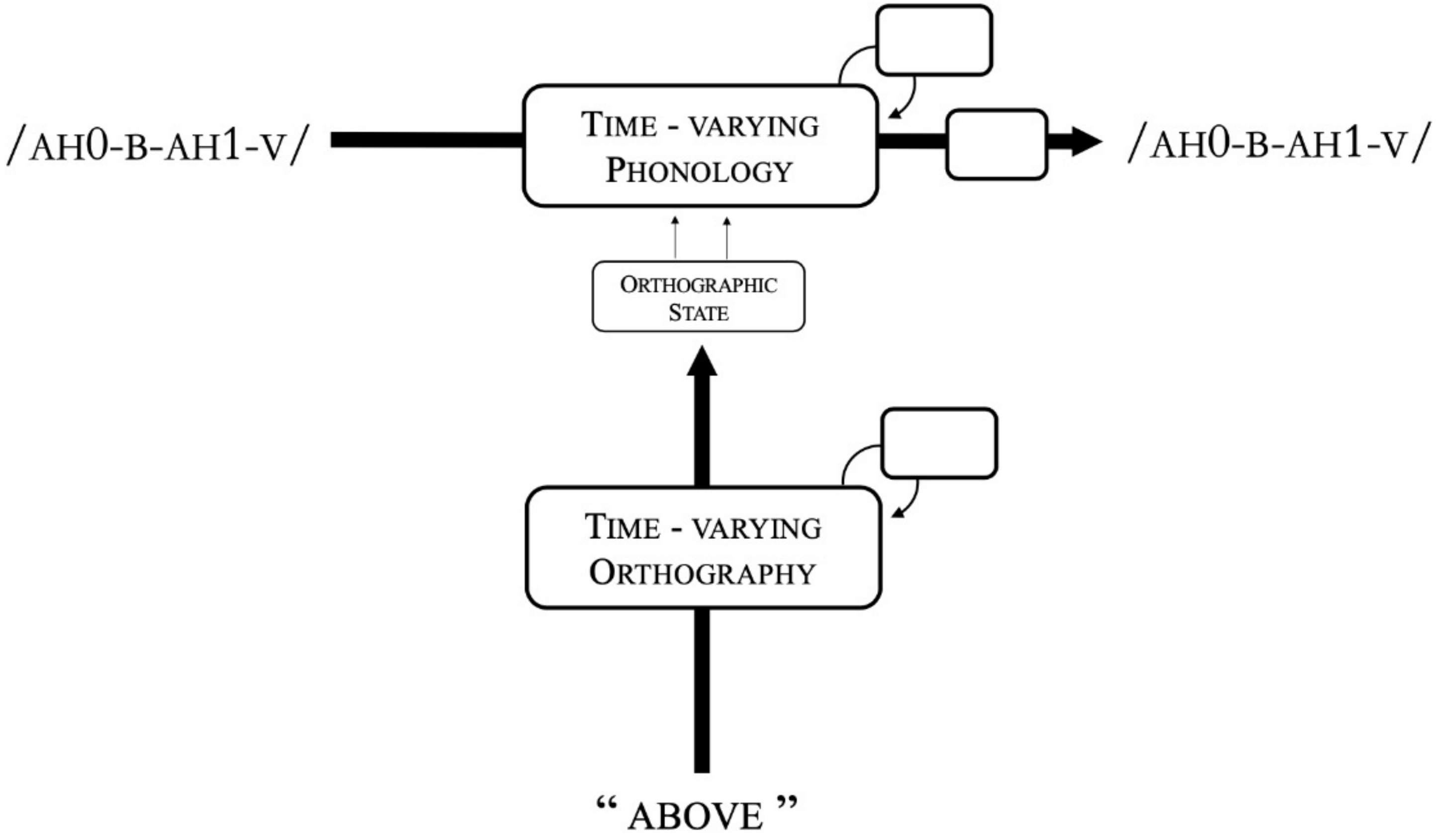
Kilpatrick

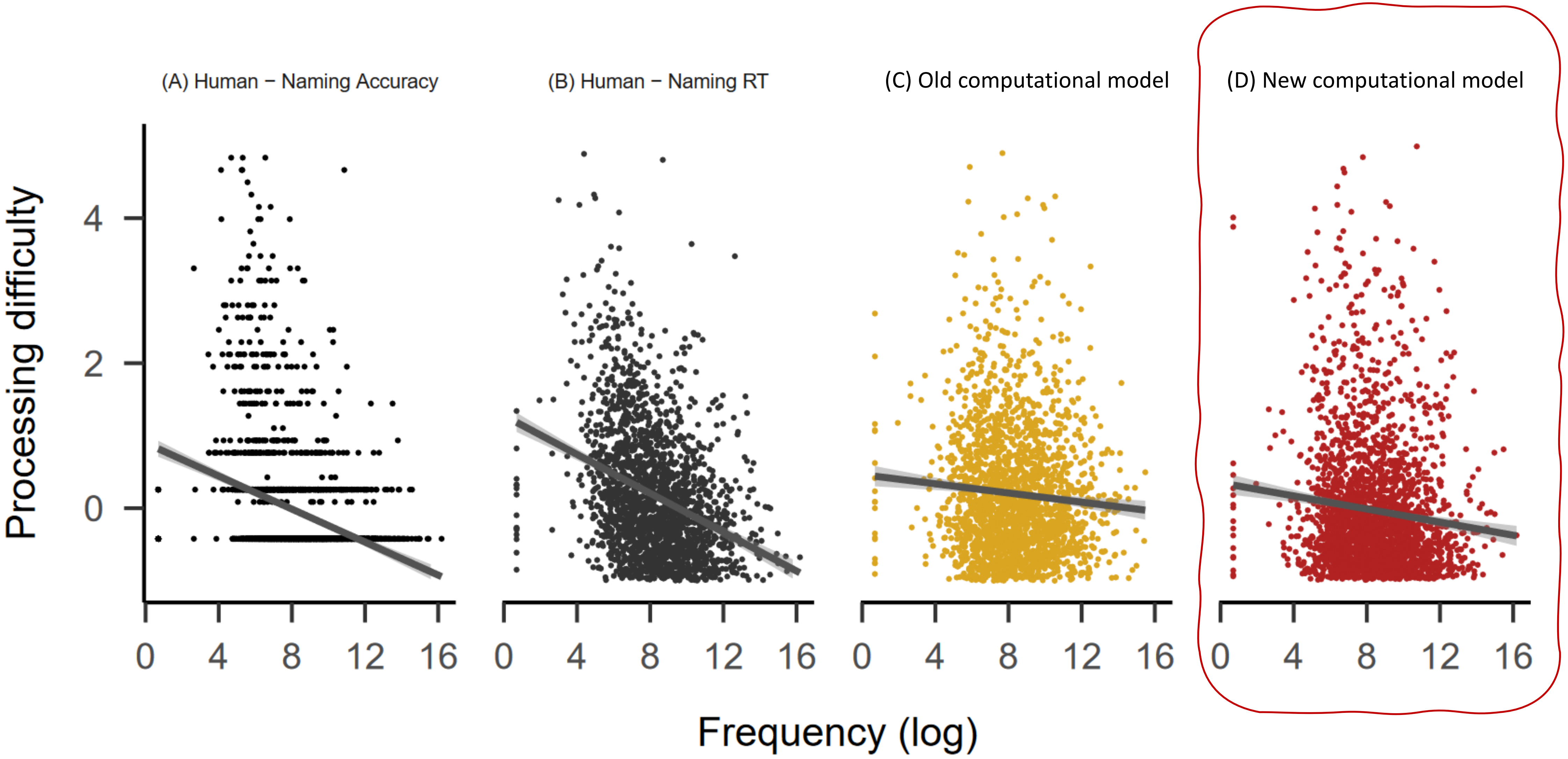


Monosyllabic words aren't good enough!



We can (should!) use longer words too





(A) Human − Naming Accuracy
(B) Human − Naming RT
(C) Old computational model
(D) New computational model
Processing difficulty
4
2
0
0 4 8 12 16
0 4 8 12 16
0 4 8 12 16
0 4 8 12 16
Frequency (log)

Word frequency

- Measured several ways
- Tell you different things, depending on data source
  - CHILDES: parental speech to children (MacWhinney, 2000)
  - TASA: books for kids in academic contexts (Zeno et al., 1995)
  - COCA: adult-directed texts, fiction (Davies, 2010)
  - WCBC: our corpus—books read to young children (Lewis et al., 2021)

Comparing to language corpora

- To facilitate comparison, we used each of four corpora
- The sample of words in a corpus = reference point (norm)
- Measured frequency of each word (# of occurrences)
- Used the rank-ordered, standardized frequency (- mean/ SD)
- Then subset the words from each instructional resource
- This allows comparison across sources
- and relative to normative sample (each corpus)

Descriptive Statistics of Rank Frequency by Source across Corpora

Source	WCBC	TASA	COCA	CHILDES
Dolch	-1.63 (0.12)	-1.68 (0.08)	-1.73 (0.01)	-1.71 (0.02)
Fry	-1.53 (0.36)	-1.67 (0.08)	-1.73 (0.01)	-1.68 (0.07)
Fundations	-1.51 (0.37)	-1.64 (0.15)	-1.73 (0.01)	-1.69 (0.07)
Kilpatrick	-0.74 (1.00)	-1.16 (0.75)	-1.68 (0.07)	-1.41 (0.52)
Wonders	-1.59 (0.21)	-1.68 (0.08)	-1.73 (0.01)	-1.70 (0.06)

Note. Means and standard deviations (in parentheses) are shown over standardized rank frequency values for all words in a given instructional source. Frequencies are standardized based on all words in a given corpus before subsetting and calculating the mean and spread by instructional source. As a result, the mean rank frequency shown represents how far from the mean of all words in a given corpus (zero) the distribution within an instructional source is.

Descriptive Statistics of Other Word Variables for each Source across WCBC

Source	AoA	Consistency	Letters	Syllables	Imageability
Dolch	-0.37 (0.69)	-0.18 (1.35)	-1.07 (0.57)	-0.80 (0.48)	-0.47 (1.28)
Fry	-0.35 (0.69)	-0.20 (1.31)	-0.89 (0.67)	-0.66 (0.64)	-0.71 (1.13)
Fundations	-0.57 (0.63)	-0.94 (1.37)	-0.75 (0.85)	-0.43 (0.86)	-1.12 (1.01)
Kilpatrick	-0.27 (0.78)	-0.85 (1.42)	-0.64 (0.63)	-0.54 (0.68)	-0.49 (1.11)
Wonders	-0.45 (0.65)	-0.60 (1.38)	-0.89 (0.73)	-0.61 (0.68)	-0.96 (0.98)

Note. Means and standard deviations (in parentheses) are shown over Z transformed rank frequency values for all words in a given instructional source. Standardization happens relative to the Wisconsin Children's Book Corpus. Therefore the mean value for a source for a given variable indicates the average distance from the overall mean within WCBC.

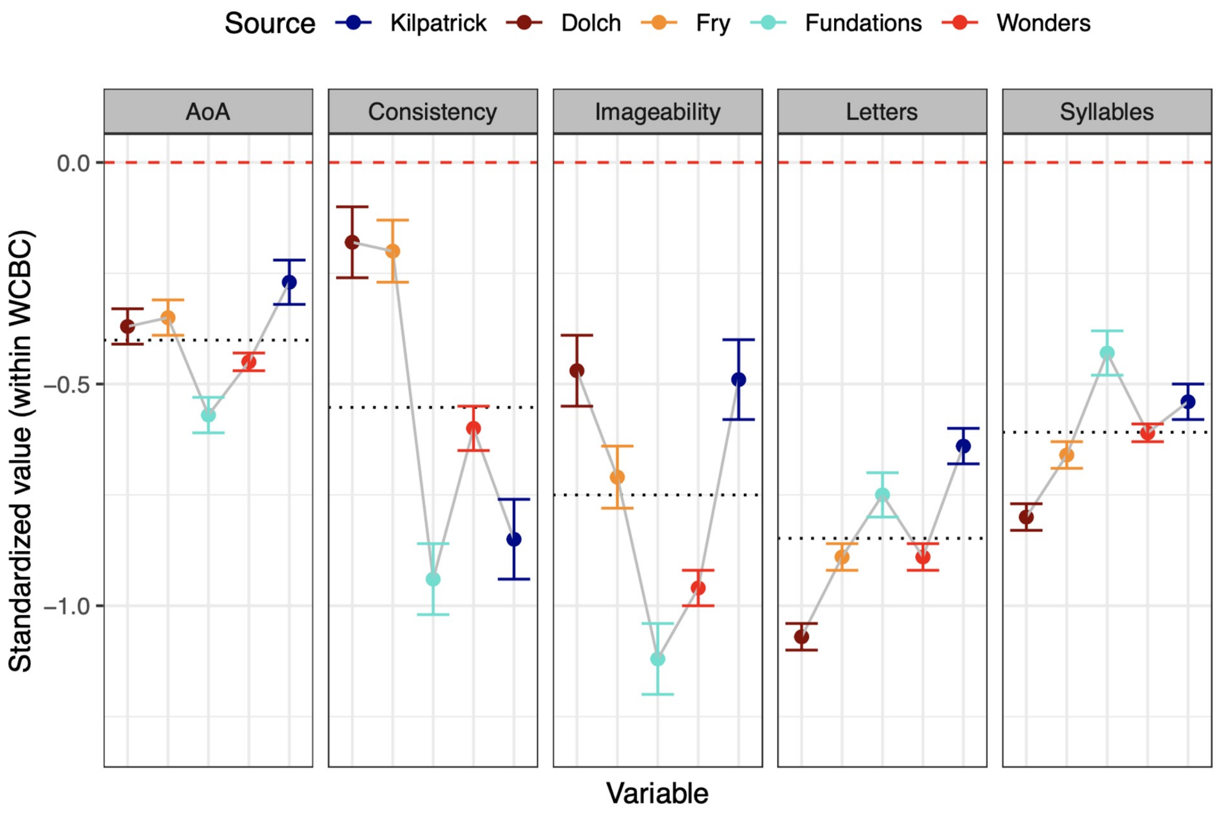


Figure 2. Means and standard errors for a range of word variables within sources are shown. Means are calculated over standardized measures within WCBC, therefore the position of the point for a source for a variable can be interpreted as the average distance away from the mean of all words within WCBC (i.e., 0 shown with a red dashed line) for that variable in terms of standard deviation units. The mean on a given variable across all sources is shown as a grey dotted line just for convenience. Estimates were derived using t.test() in base R.

Model Estimates for Statistical Tests of Source Means against Means from Normative Sample

Source	AoA	Consistency	Imageability	Letters	Syllables
Dolch	**-0.37** (0.04)	**-0.18** (0.08)	**-0.47** (0.08)	**-1.07** (0.03)	**-0.80** (0.03)
Fry	**-0.35** (0.04)	**-0.20** (0.07)	**-0.71** (0.07)	**-0.89** (0.03)	**-0.66** (0.03)
Fundations	**-0.57** (0.04)	**-0.94** (0.08)	**-1.12** (0.08)	**-0.75** (0.05)	**-0.43** (0.05)
Kilpatrick	**-0.27** (0.05)	**-0.85** (0.09)	**-0.49** (0.09)	**-0.64** (0.04)	**-0.54** (0.04)
Wonders	**-0.45** (0.02)	**-0.60** (0.05)	**-0.96** (0.04)	**-0.89** (0.03)	**-0.61** (0.02)

Note. Estimates obtained from using t.test() from base R, with standard errors in parentheses. Bolded parameter estimates are statistically significant. For a full accounting of parameter estimates see appendix. Values were Z transformed on the distribution from the WCBC prior to subsetting and testing by source.

Parameter Estimates for Source Means on Age-of-Acquisition Relative to Normative Sample from WCBC

Source	*b*	*t*	*df*	*SE*	95% CI	*p*
Dolch	**-0.37**	-9.33	307	0.04	[-0.45, -0.29]	$< .001$
Fry	**-0.35**	-9.75	374	0.04	[-0.42, -0.28]	$< .001$
Fundations	**-0.57**	-15.85	304	0.04	[-0.64, -0.50]	$< .001$
Kilpatrick	**-0.27**	-5.17	228	0.05	[-0.37, -0.16]	$< .001$
Wonders	**-0.45**	-19.22	770	0.02	[-0.49, -0.40]	$< .001$

Note. Bolded coefficients are statistically significant and can be interpreted as Cohen's *d* based on the statistical test performed.

Consistency

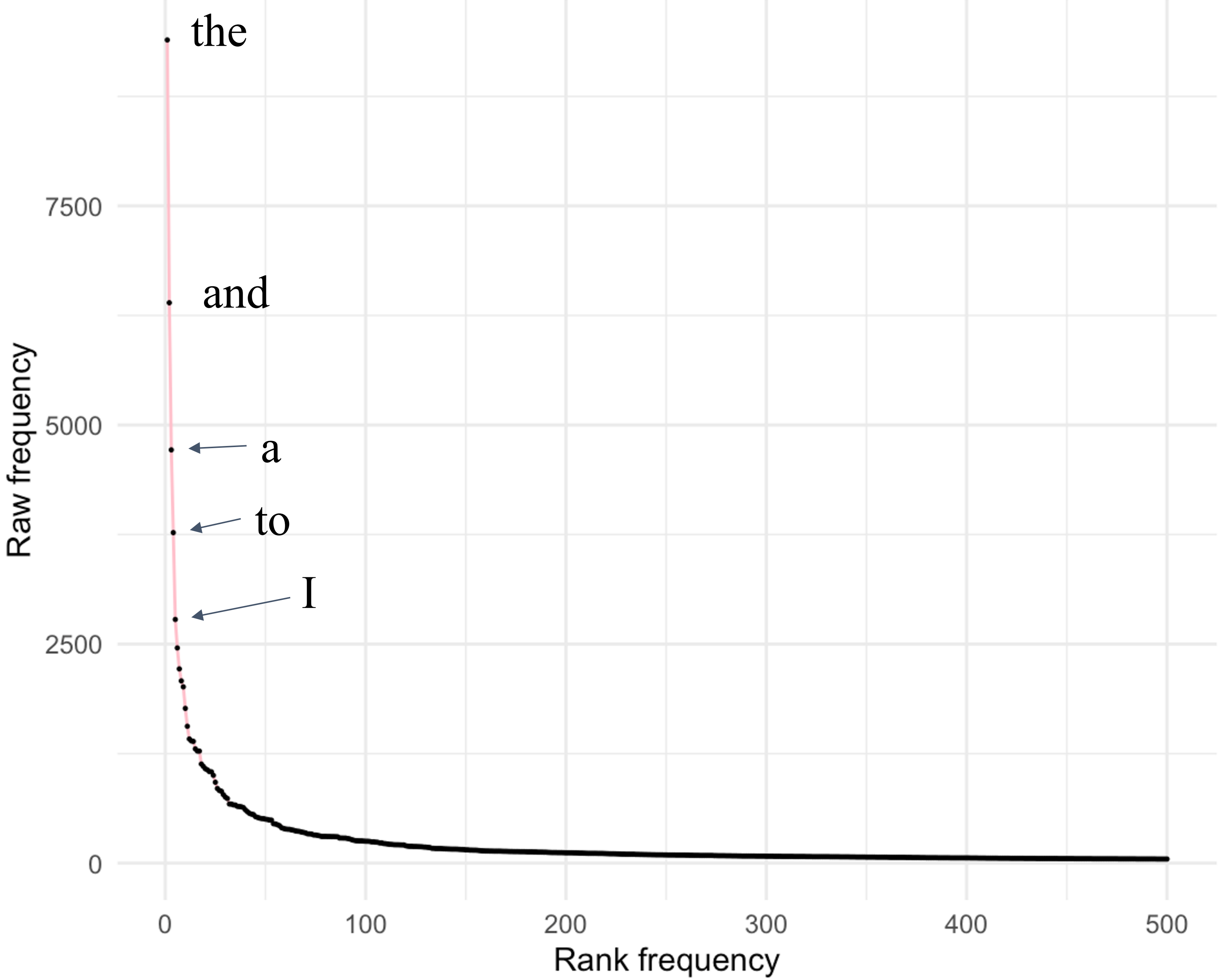
Source	*b*	*t*	*df*	*SE*	95% CI	*p*
Dolch	**-0.18**	-2.38	306	0.08	[-0.34, -0.03]	$< .05$
Fry	**-0.20**	-2.95	384	0.07	[-0.33, -0.07]	$< .01$
Fundations	**-0.94**	-11.98	305	0.08	[-1.09, -0.78]	$< .001$
Kilpatrick	**-0.85**	-9.03	228	0.09	[-1.03, -0.66]	$< .001$
Wonders	**-0.60**	-12.10	783	0.05	[-0.70, -0.50]	$< .001$

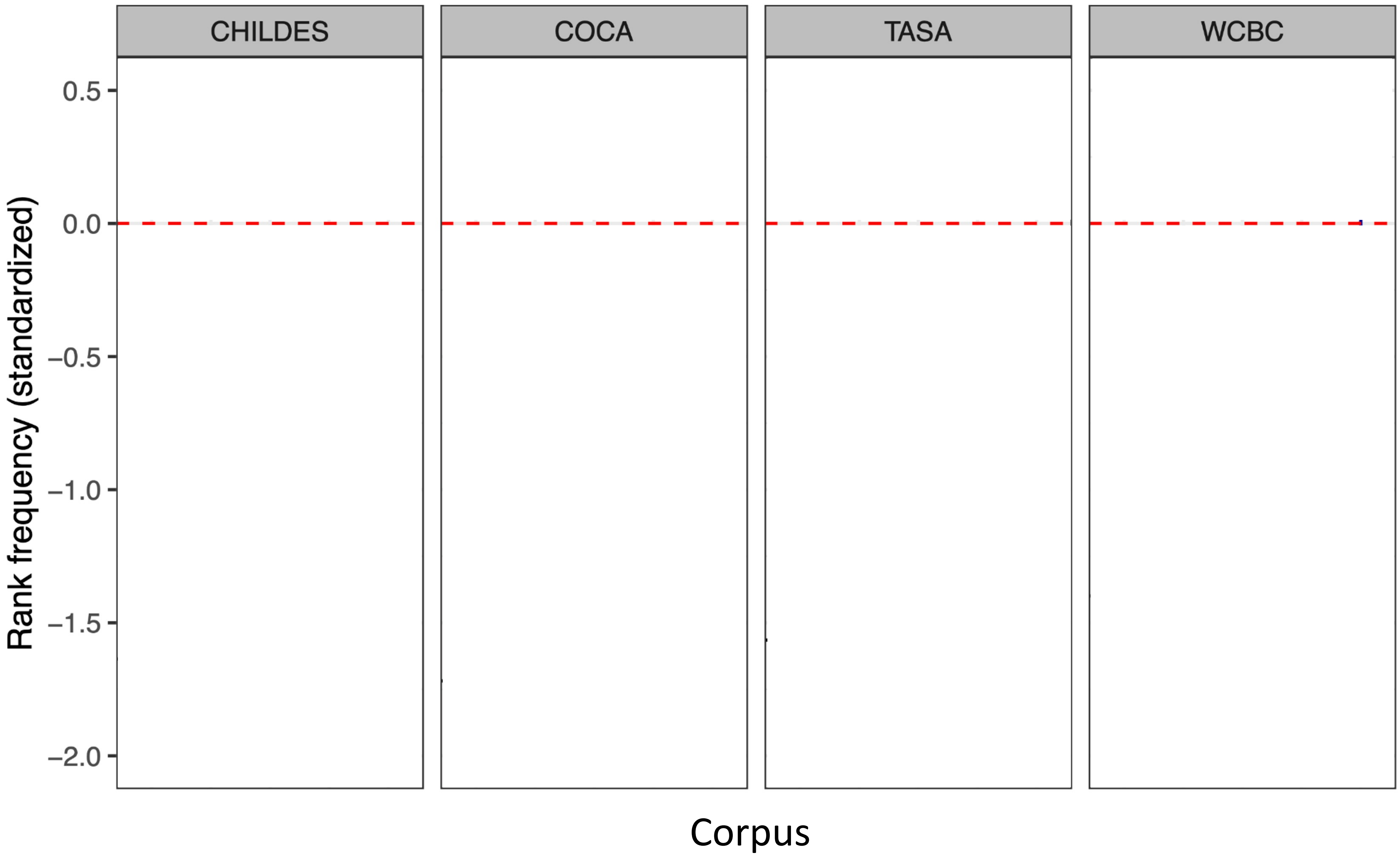
Imageability

Source	b	t	df	SE	95% CI	p
Dolch	**-0.47**	-5.68	240	0.08	[-0.63, -0.31]	$< .001$
Fry	**-0.71**	-10.08	257	0.07	[-0.85, -0.57]	$< .001$
Fundations	**-1.12**	-14.55	171	0.08	[-1.28, -0.97]	$< .001$
Kilpatrick	**-0.49**	-5.26	140	0.09	[-0.68, -0.31]	$< .001$
Wonders	**-0.96**	-22.35	518	0.04	[-1.04, -0.87]	$< .001$

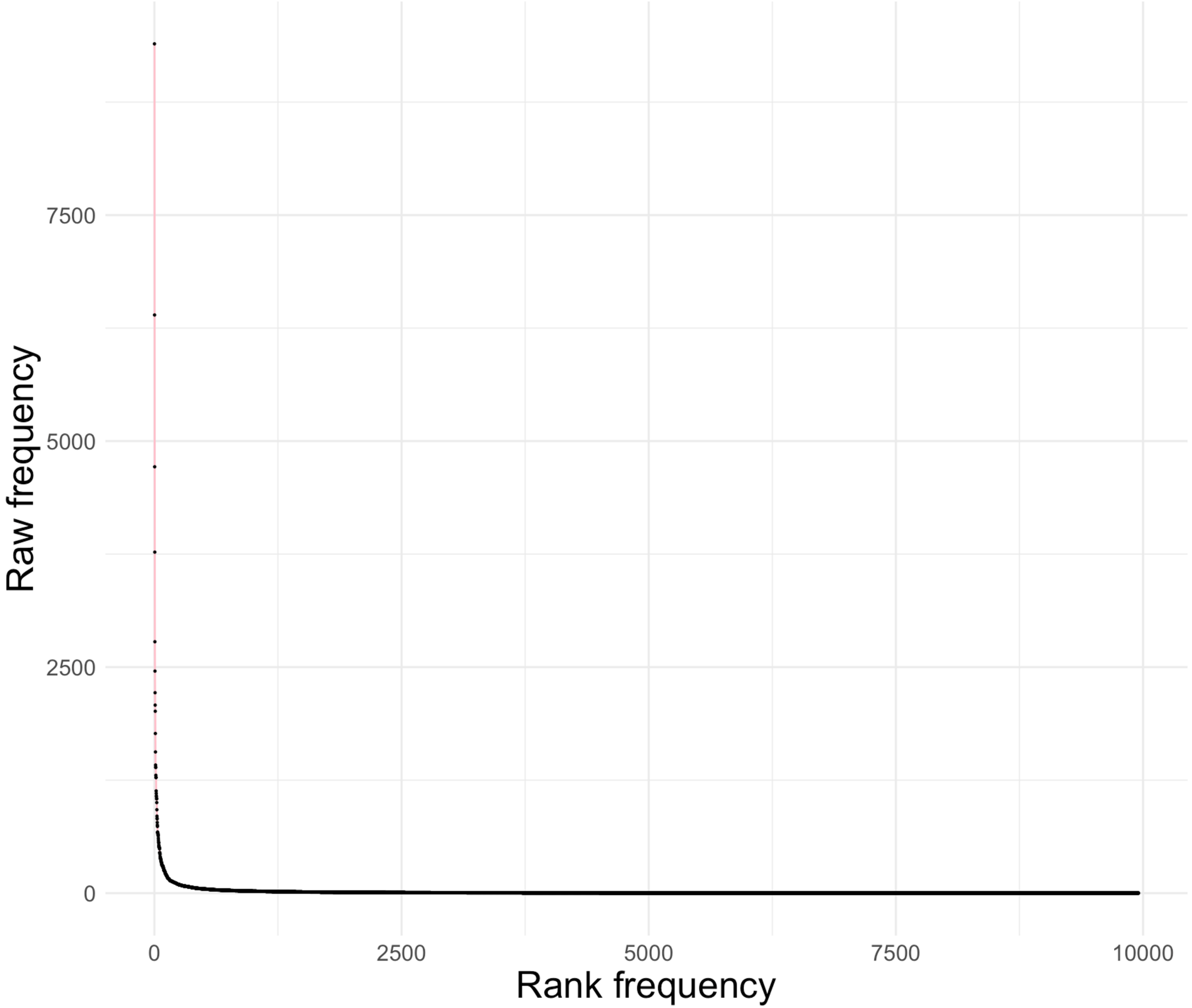
Number of Syllables per Word

Source	b	t	df	SE	95% CI	p
Dolch	**-0.80**	-29.22	310	0.03	[-0.85, -0.75]	$< .001$
Fry	**-0.66**	-20.05	387	0.03	[-0.72, -0.59]	$< .001$
Fundations	**-0.43**	-8.91	309	0.05	[-0.53, -0.34]	$< .001$
Kilpatrick	**-0.54**	-12.13	232	0.04	[-0.63, -0.45]	$< .001$
Wonders	**-0.61**	-25.27	790	0.02	[-0.66, -0.56]	$< .001$

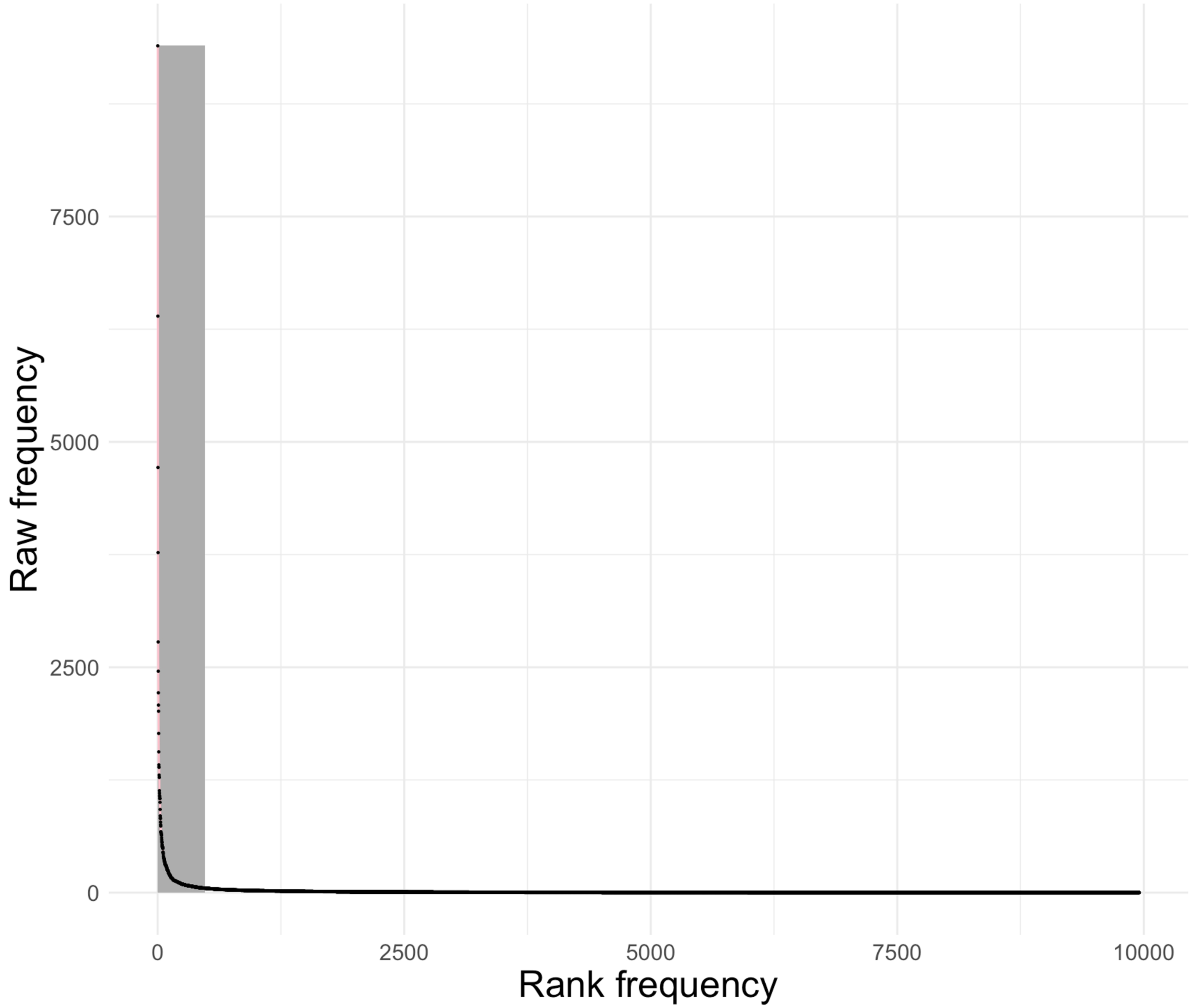

the
and
a
to
I
7500
5000
2500
0
Raw frequency
0
100
200
300
400
500
Rank frequency

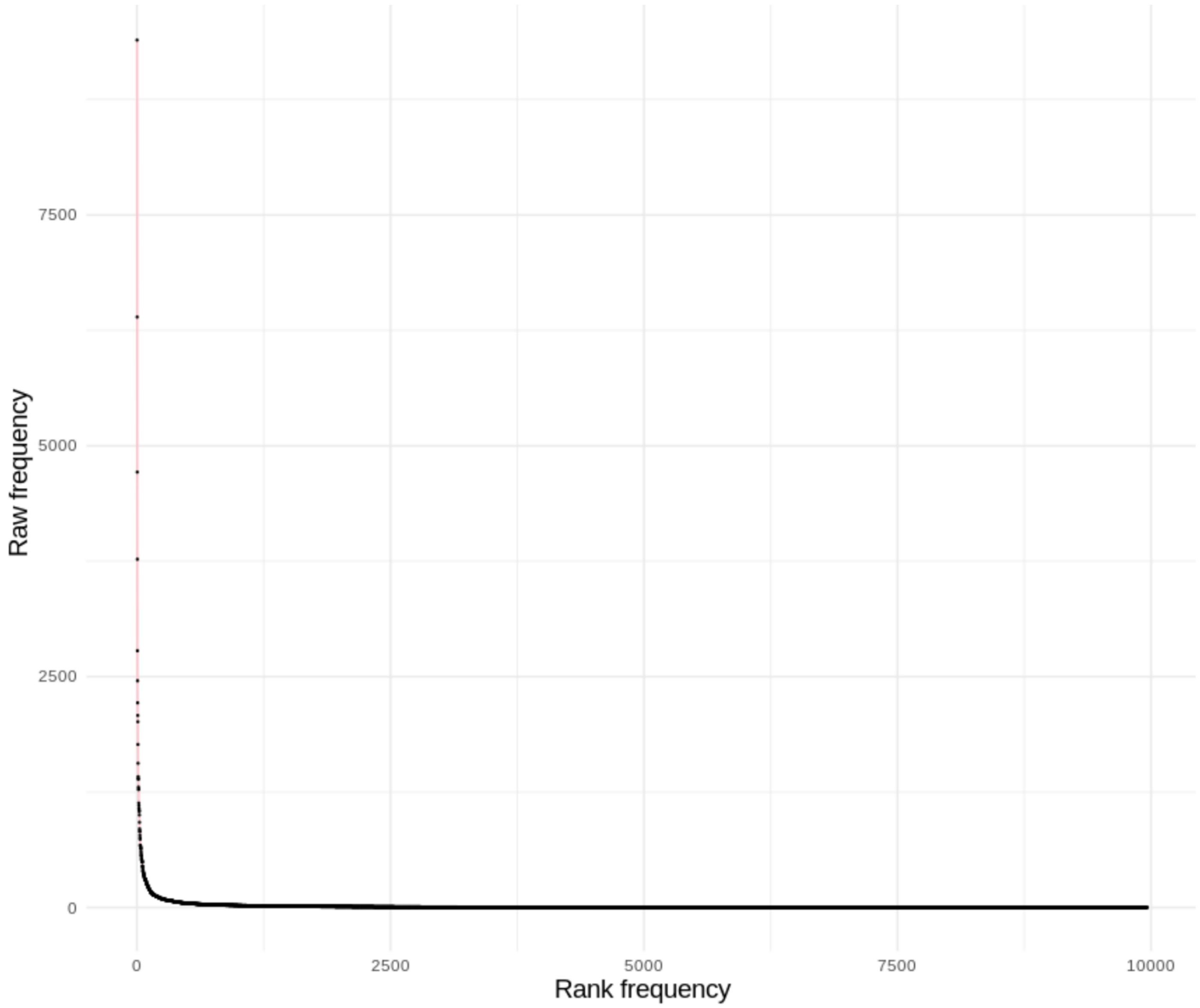

CHILDES
COCA
TASA
WCBC
0.5
0.0
−0.5
−1.0
−1.5
−2.0
Rank frequency (standardized)
Corpus

Source	b	t	df	SE	95% CI	p

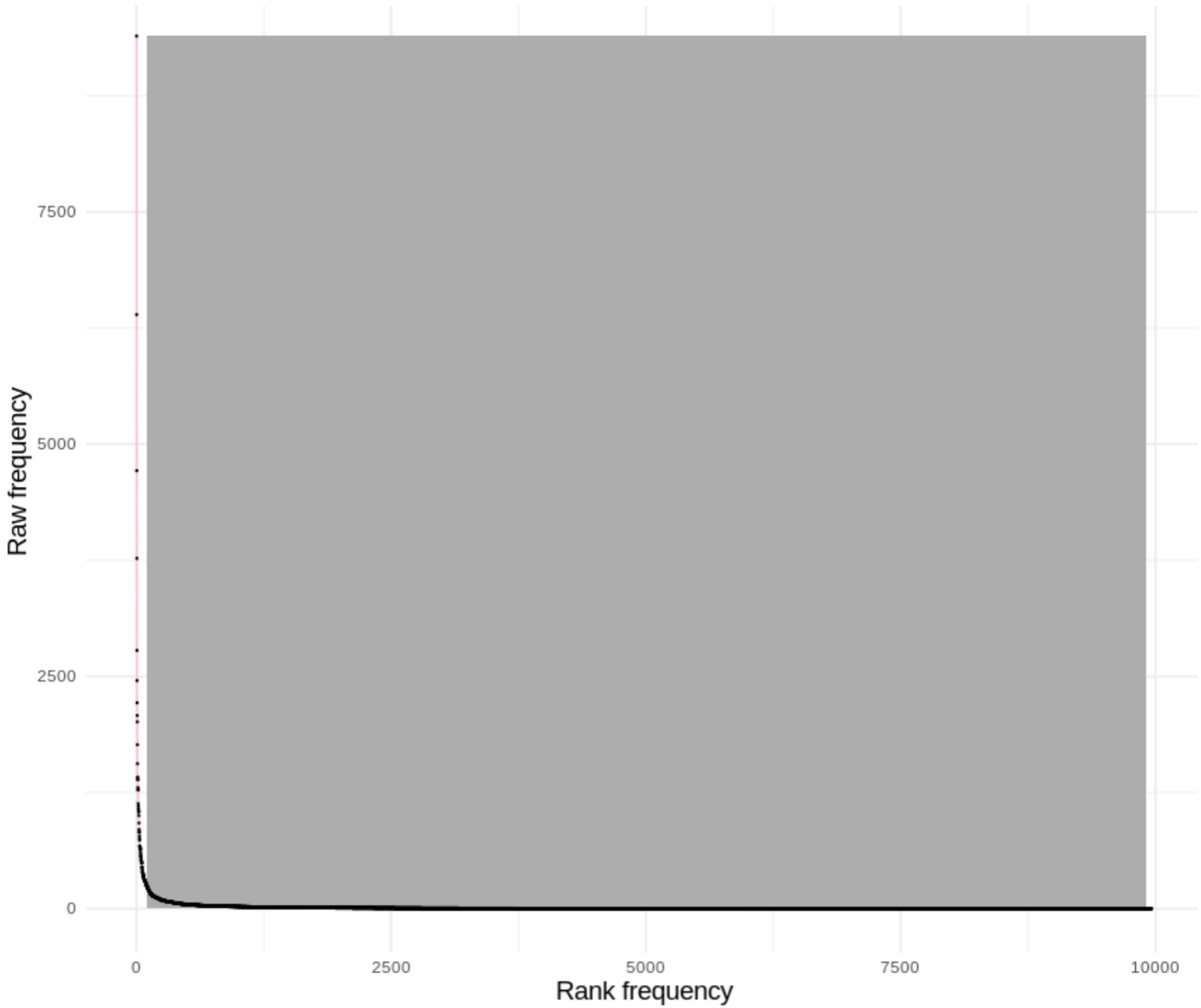


Raw frequency
7500
5000
2500
0
Rank frequency
0
2500
5000
7500
10000

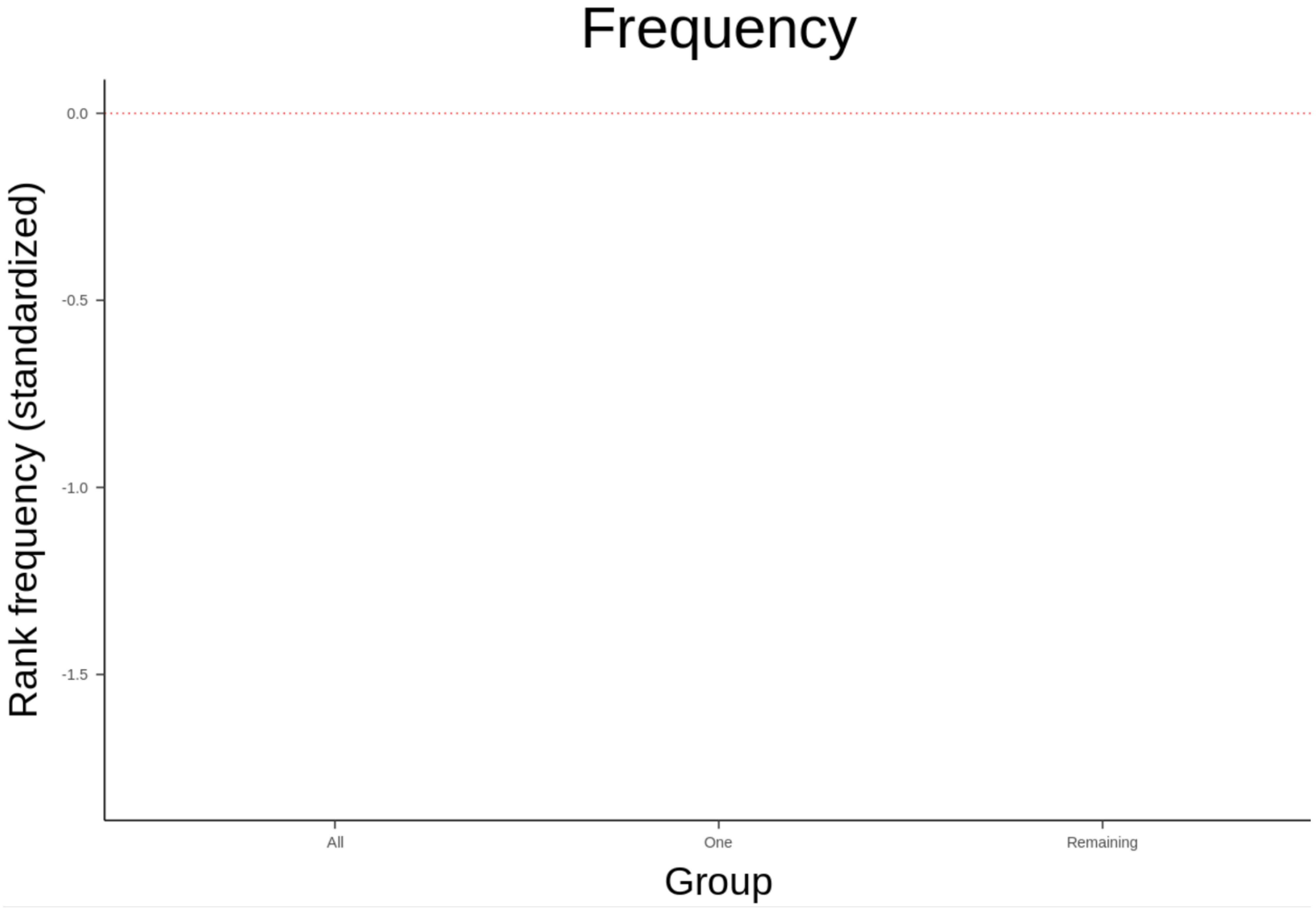

7500
5000
2500
0
Raw frequency
0
2500
5000
7500
10000
Rank frequency



Raw frequency
7500
5000
2500
0
0
2500
5000
7500
10000
Rank frequency



Raw frequency
Rank frequency
0
2500
5000
7500
10000


Frequency
Rank frequency (standardized)
0.0
-0.5
-1.0
-1.5
All
One
Remaining
Group

These are **theories**

- They give teachers a model for their learner
- And what should be learned
- That is, they are educational theories

learn, earth

Variable	Spearman's ρ
Orthographic length	**0.16**
Phonological length	**0.23**
Orthographic neighbors	-0.1
Phonological neighbors	-0.08
Phonological density	**0.15**
Morphology	**-0.14**
Oncleus entropy	**0.14**
Vowel entropy	**0.22**
Rime entropy	0
Age of acquisition	0.11
Child text frequency	-0.12
Adult text frequency	-0.12

Results

- All training sets were learned nearly perfectly
- Data focus on generalization accuracy
- Two types of analyses:

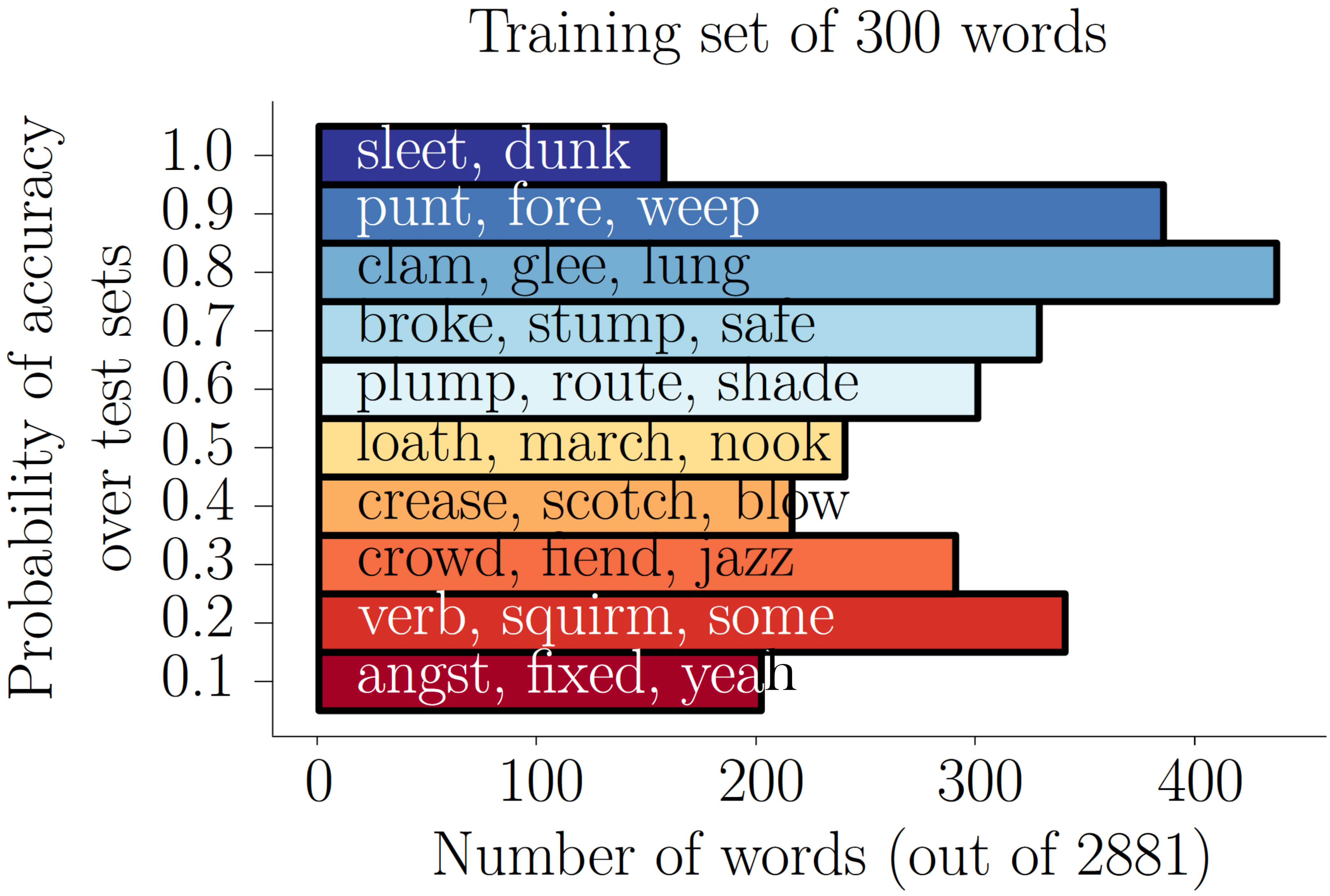
Across models:

$$\frac{\#\ words\ learned\ via\ generalization}{total\ \#\ of\ words}$$

By item:

How often it was correct on generalization (not trained)

$$\frac{\#\ of\ correct\ generalizations}{total\ \#\ of\ models\ where\ item\ not\ taught}$$



Training set of 300 words
Probability of accuracy over test sets
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
sleet, dunk
punt, fore, weep
clam, glee, lung
broke, stump, safe
plump, route, shade
loath, march, nook
crease, scotch, blow
crowd, fiend, jazz
verb, squirm, some
angst, fixed, yeah
0
100
200
300
400
Number of words (out of 2881)

High Frequency "Trick" Words

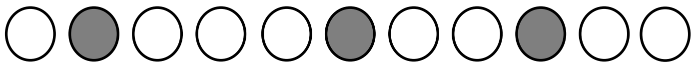
High frequency words are the words that appear most often in print. They are the very common everyday words of the English language (such as "they" and "what"). Some of these words are phonetically regular, but many are irregular, non-phonetic words which do not follow the "system" of the language. These high frequency words, whether phonetic or irregular, are used so commonly in English they need to be recognized and spelled quickly and easily even if their phonemic patterns have not yet been taught. Thus, these words are presented to be memorized.

In Fundations, the high frequency words are called trick words, and these words are not tapped out. The trick words were selected from common high frequency word lists such as Fry, and the American Heritage Word Frequency Book.

Representations are across features

- They are distributed
- A /b/ sound:
  - Uses the lips (is *bilabial*)
  - Engages the vocal chords (is *voiced*)
  - Involves a complete stoppage (is *plosive*)

/b/ : 0 1 0 0 0 1 0 0 1 0 0



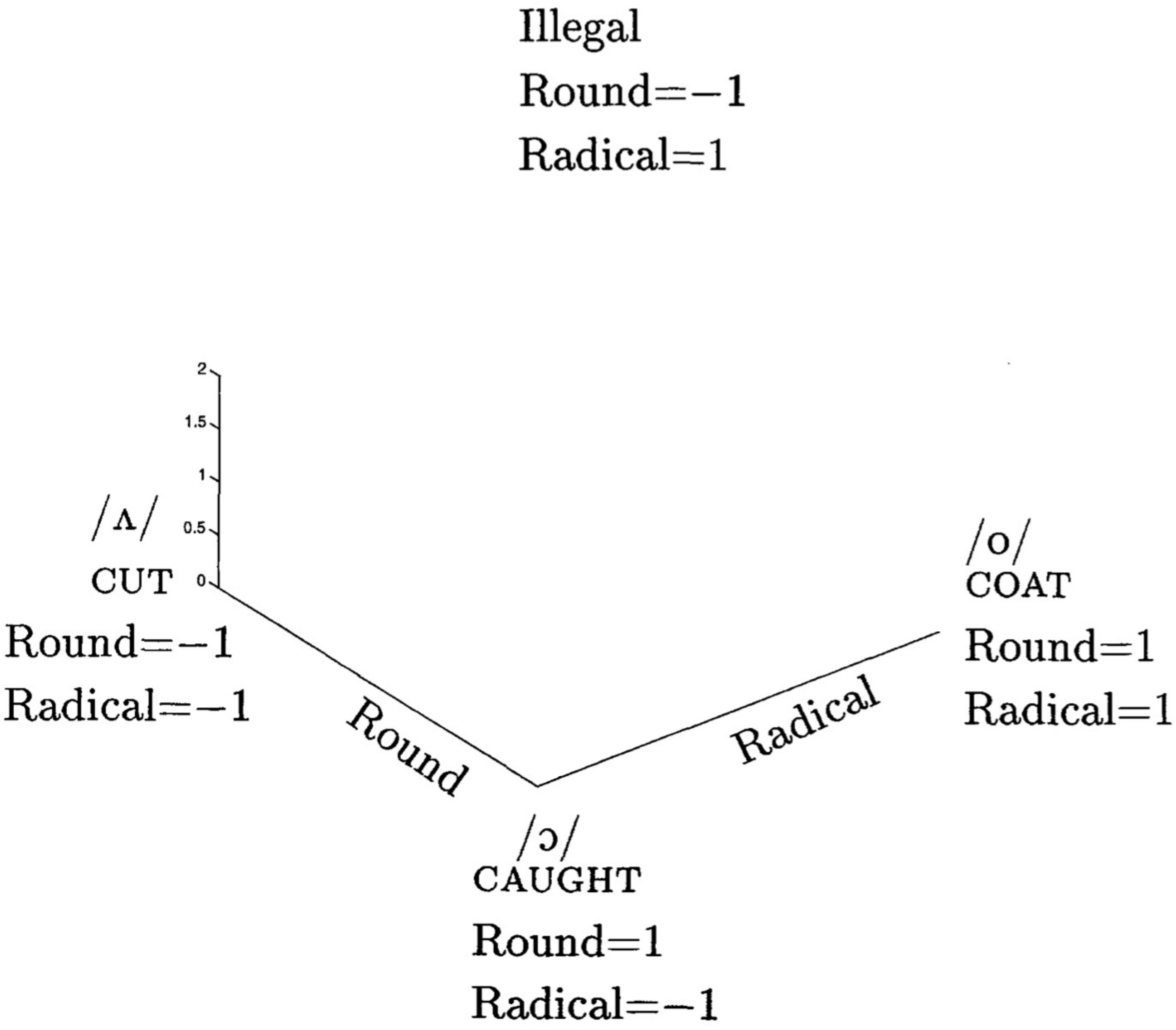


Figure 7. Attractor basins plotted spatially (see Figure 6). A point landing on this space would "roll" into one of the three legal attractor states. The z axis measures distance a point travels in the dynamic space.

Illegal
Round=−1
Radical=1

Palatal=−1

/o/
COAT
Round=1
Radical=1

/ʌ/
CUT
Round=−1
Radical=−1

/ɔ/
CAUGHT
Round=1
Radical=−1

Figure 6. Phonological attractor, depicting three legal phonemes derived from combinations of two features: round and radical.

Trial 1: $Word_1$, $Word_{2,}$ $Word_3$ … $Word_{3000}$ $Word_x$ sampled → Train model

T2: $Word_1$, $Word_{2,}$ $Word_3$ … $Word_{3000}$ $Word_x$ sampled → Train model

T3: $Word_1$, $Word_{2,}$ $Word_3$ … $Word_{3000}$ $Word_x$ sampled → Train model

…

T10K: $Word_1$, $Word_{2,}$ $Word_3$ … $Word_{3000}$ $Word_x$ sampled → Train model

Test via sampling after training + optimization processes

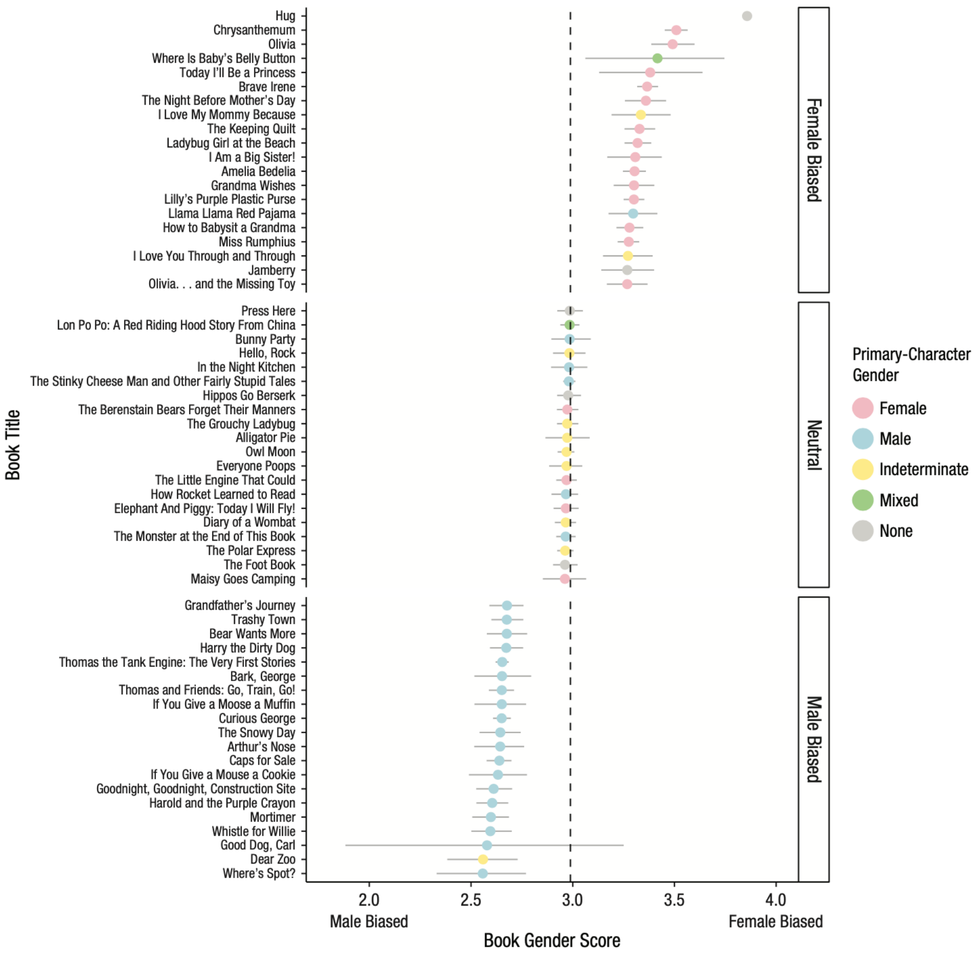


Fig. 1. Mean gender bias of a subset of books from Study 1b as a function of the gender of the book's primary character. The subset contains the 20 books with the highest feminine-bias scores, the 20 with the highest masculine-bias scores, and 20 from the neutral range. Bias scores were calculated from the mean gender ratings of words in each book (tokens). The dashed line indicates the overall mean across books. Error bars show bootstrapped 95% confidence intervals.

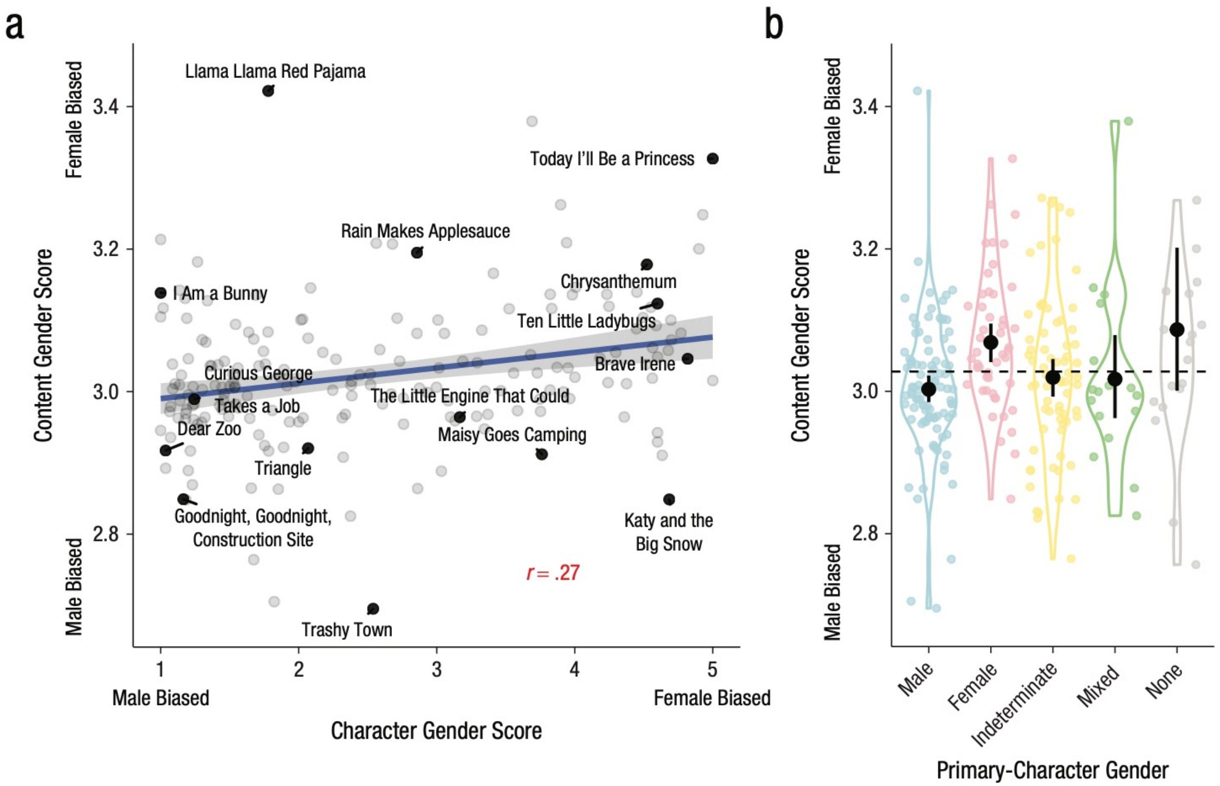


Fig. 2. Gender content bias as a function of character gender (Study 1b). The scatterplot (a) shows the relation between mean content gender score for each book and mean character gender score. The solid line shows the best-fitting linear regression, and the error band shows standard errors. The violin plots (b) shows the distribution of content gender score across books as a function of the gender of the book's primary character. Colored points show individual books (one point is excluded for visibility). The dashed line shows the grand mean content gender score. Black points show means, and error bars show bootstrapped 95% confidence intervals. The width of each violin indicates the density of the data.

Fig. 3. Proportion of books with main characters in each gender category (male, female, mixed, and indeterminate) as a function of publication decade (Study 1b). Bar width corresponds to the number of books in the Wisconsin Children's Book Corpus published in that decade.

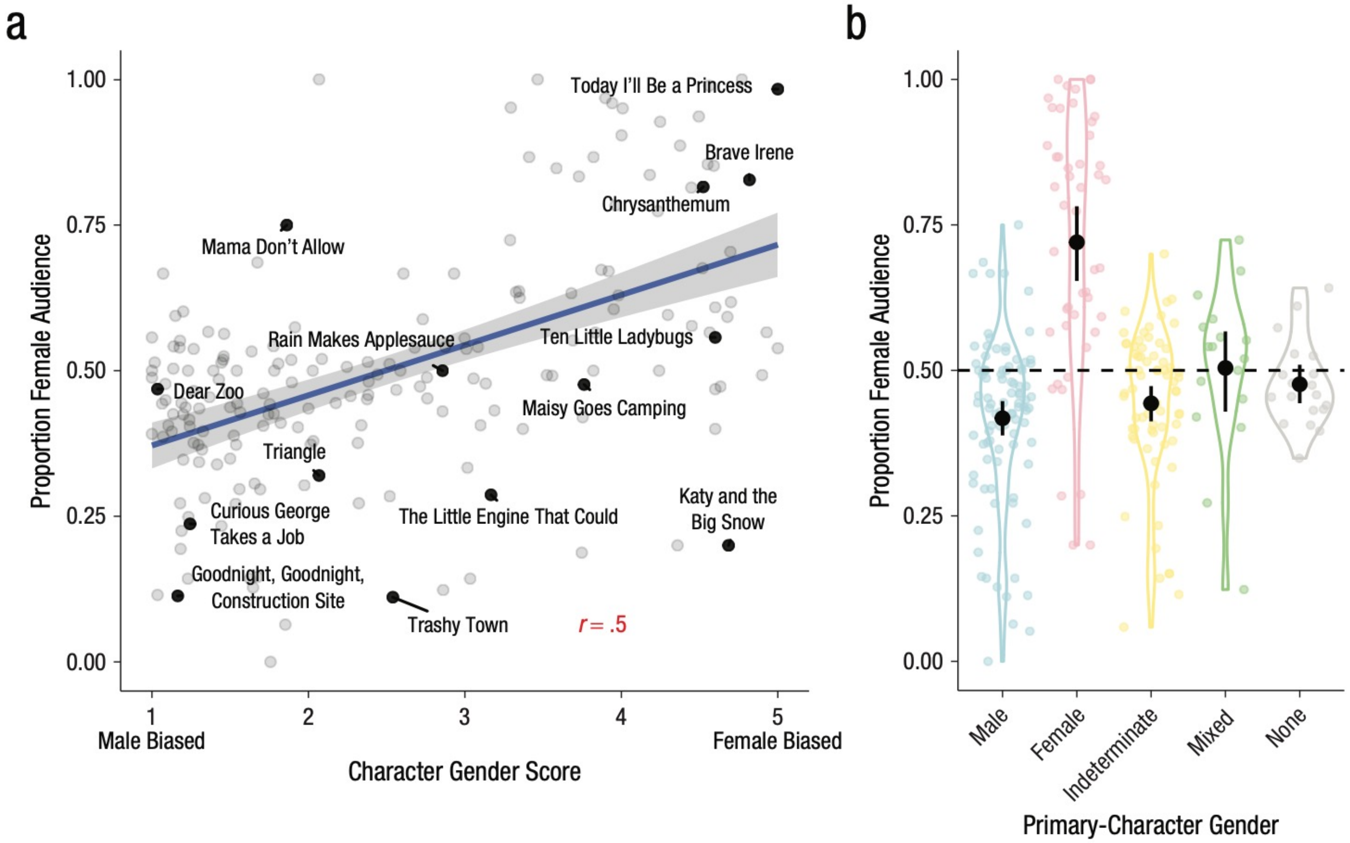


Fig. 6. Proportion of female audience as a function of character gender (Study 3). The scatterplot (a) shows the relation between the female audience for each book and the book's mean character gender score. The solid line shows the best-fitting linear regression, and the error band shows standard errors. The violin plots (b) show the distribution of female audience members across books as a function of primary-character gender. Colored points show individual books. The dashed line corresponds to an audience that is half female. Black points show means, and error bars show bootstrapped 95% confidence intervals. The width of each violin indicates the density of the data.

Grounding early reading instruction in a theory of learning: Deciding what to teach when, and why.

There is a widely-held view that early reading instruction matters, and that the properties of words taught are an important factor in instruction. This implies that there might be common ground in decision-making about words selected for instruction, to the extent that we have established theories of learning employed in developing educational theories along these lines. I will present data that show that many common instructional approaches to teaching word reading vary widely, despite similarities in how they conceive of the process of learning to read words. Additionally, I'll report on a range of research that shows how a theory of learning and its computational implementation can support decisions about what we teach early readers, when, and why.