Maria Korochkina

How much can children learn about English morphology through book reading?

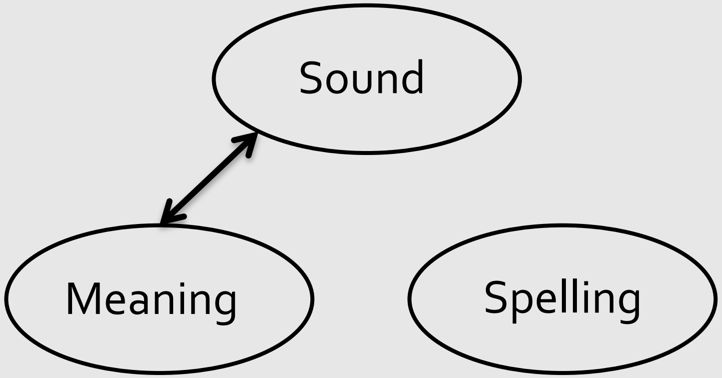
Kathy Rastle
Royal Holloway, University of London

@kathy_rastle
Kathy.Rastle@rhul.ac.uk

SL and Reading

“Learning to read is an exercise in statistical learning”

(Rueckl et al., 2024)



How do the discrete visual symbols of a writing system represent spoken language?

+ Distributional properties of the input

+ Reader’s engagement with the input

- Multiple (conflicting) levels of regularity

- Different degrees of reliability & frequency

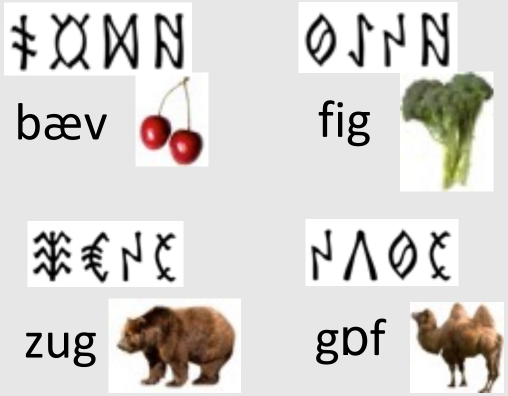
- Very extended time-course (years, decades)

SL in a real reading scenario is hard!

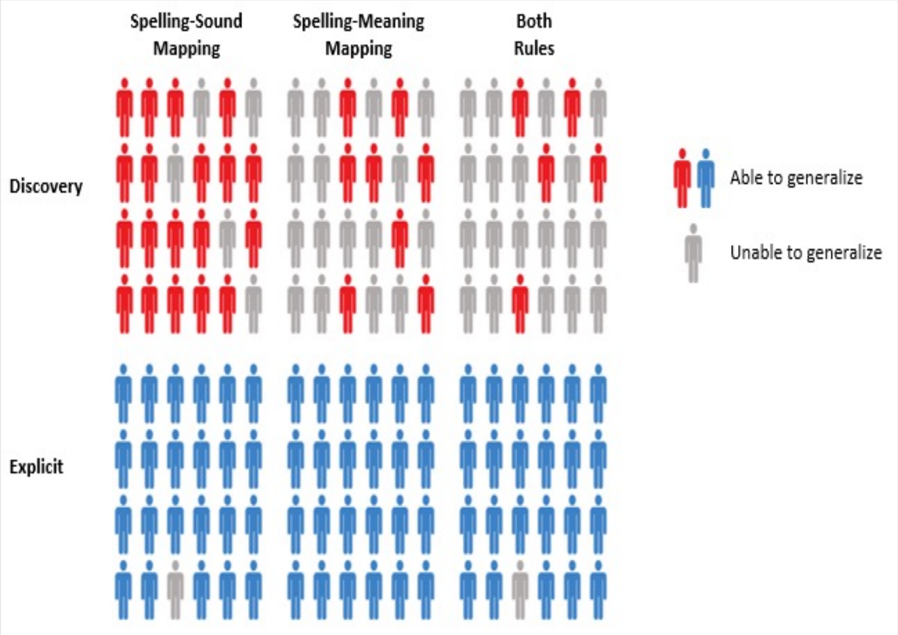
"Discovery learning may be a relatively inefficient way of learning underlying regularities even given years of text experience"

(Rastle et al., 2021)

Adults trained on new words for ~18 hrs. Half had 30 mins instruction on writing system

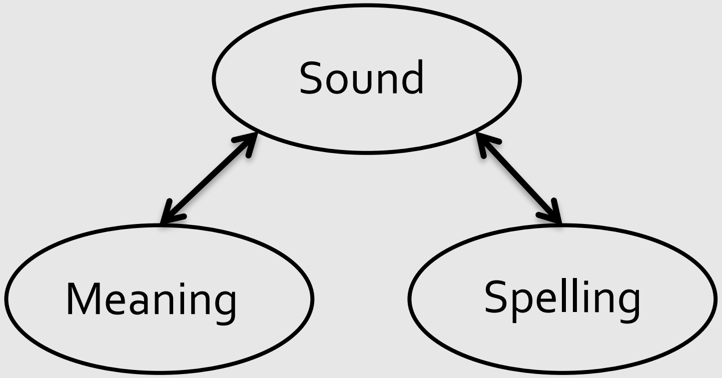


Tested on sounds and meanings of trained words and untrained words.



Poor learning of underlying regularities in the absence of instruction.

OP (orthography-phonology) mapping

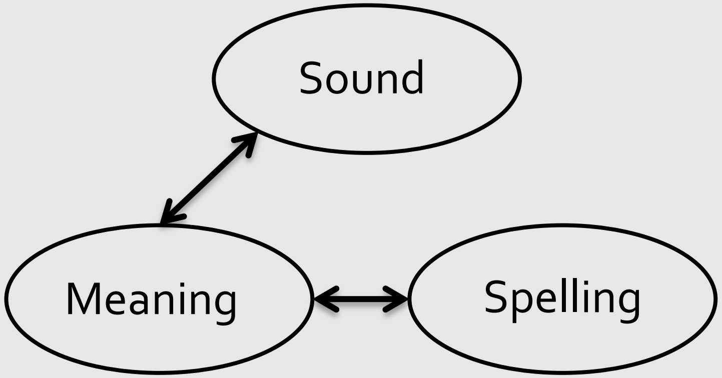


OP (Orthography-Phonology) Mapping

- Highly systematic, even in the least transparent alphabets (English)
- Basic GPCs virtually always instructed (to some extent) in initial years of school.
- SL through text experience builds on this (also to non-instructed regularities e.g. oo, ook); graded according to the salience (*frequency, consistency)* of mapping.
- Knowledge of non-instructed regularities builds very slowly, and may not fully capture highly-systematic regularities even after decades of experience *(Treiman & Kessler, 2019)*

OS (orthography-semantics) mapping

OS (Orthography-Semantics) Mapping



- OS systematicity conveyed via morphology

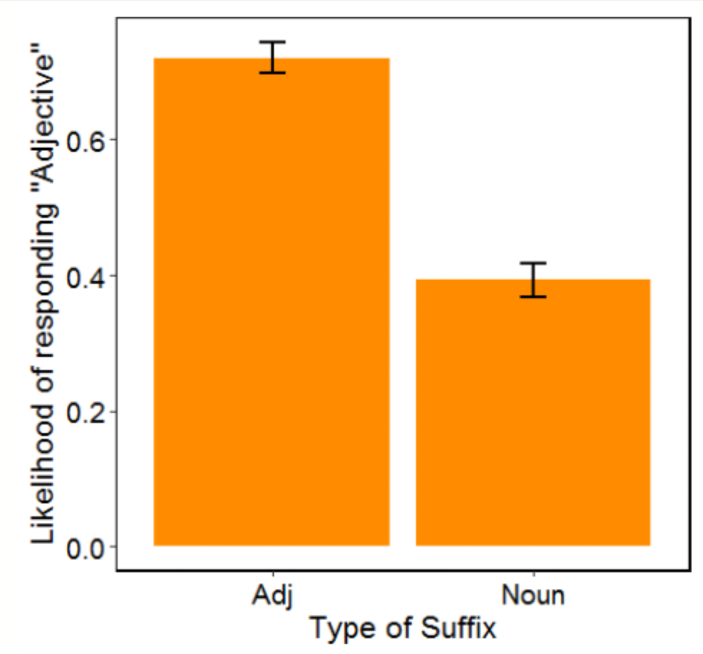
cleaner, cleanly, unclean
teacher, banker, builder

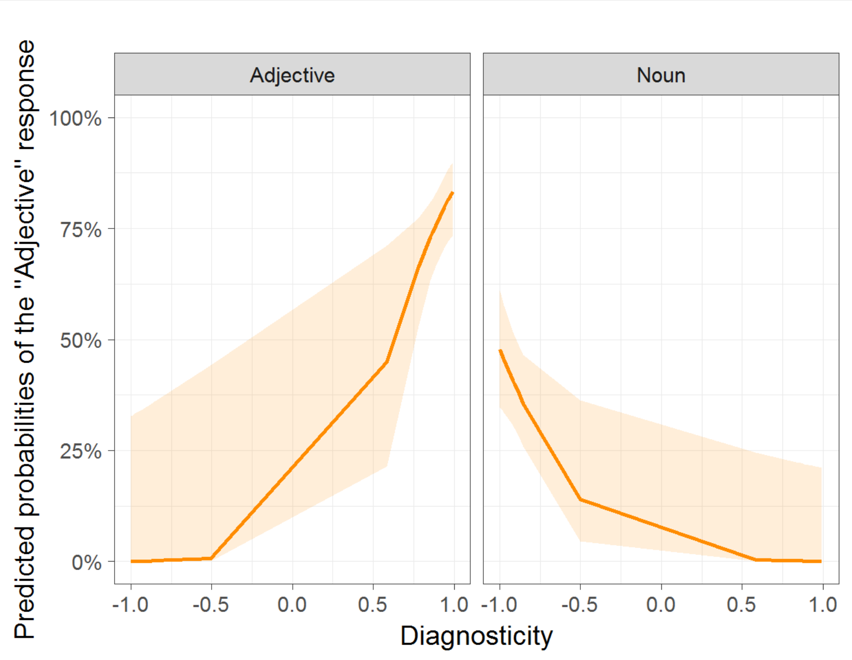
- Systematicity much greater in written English than in spoken English.
- Substantial evidence that undergraduate participants analyse *familiar* and *unfamiliar* (e.g. quickify) words in terms of their morphology in online lexical processing.
- Emerging evidence that this knowledge is *graded* in terms of how reliably the morpheme communicates meaning.

OS (morpheme) sensitivity is graded

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Is it an adjective or noun?
DOMOUS, JIXLET, TERISH, RABNESS …





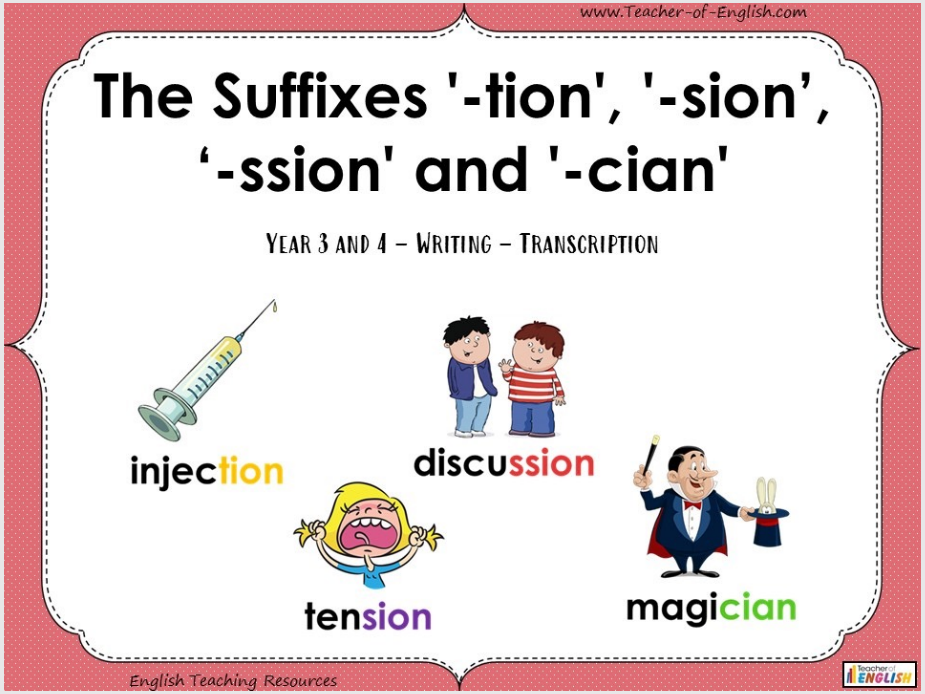
- Explicit knowledge of object / property / act status, linked to strength of cue
- Knowledge superior for adults with higher vocabulary & spelling
- Similar, graded effects in eye-tracking and spelling

Adult knowledge reflects OS distributional structure

Ulicheva, Harvey, Aronoff, & Rastle, 2020, Cognition

OS (morphology) instruction

Teacher knowledge of morphology is patchy; often *no instruction* of derivational morphology or *poor instruction*



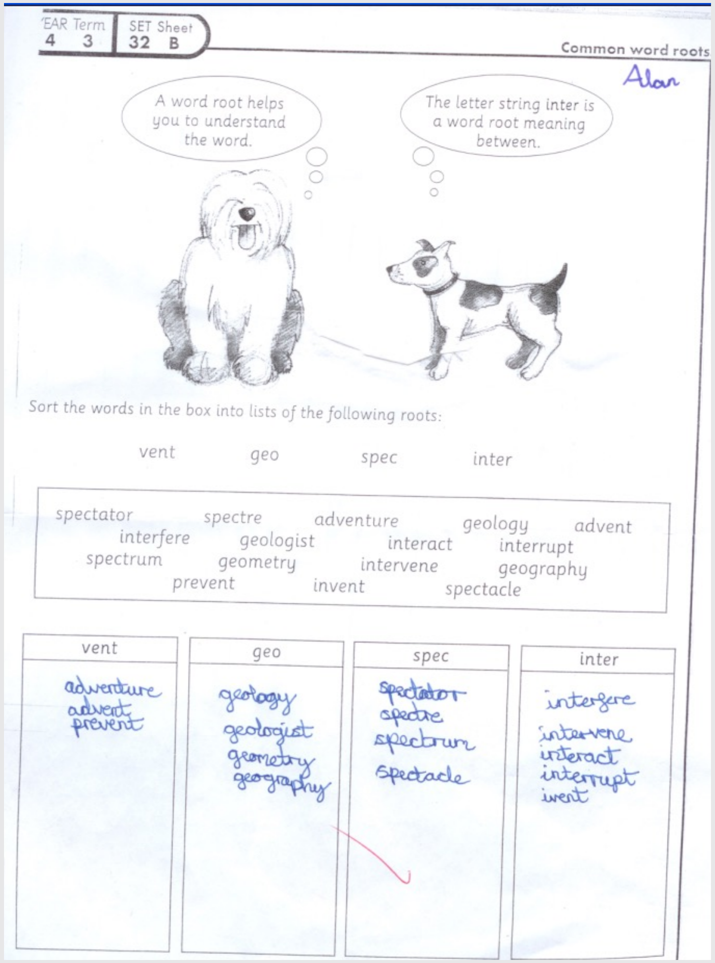
www.Teacher-of-English.com

The Suffixes '-tion', '-sion', '-ssion' and '-cian'

YEAR 3 AND 4 – WRITING – TRANSCRIPTION

English Teaching Resources

Children need to acquire morpheme knowledge via text experience.



YEAR Term 4 3 | SET Sheet 32 B

Common word roots

Alan

A word root helps you to understand the word.
The letter string inter is a word root meaning between.

Sort the words in the box into lists of the following roots:

vent geo spec inter

spectator spectre adventure geology advent
interfere geologist interact interrupt
spectrum geometry intervene geography
prevent invent spectacle

vent	geo	spec	inter
adventure	geology	spectator	intergere
advent	geologist	spectre	intervene
prevent	geometry	spectrum	interact
	geography	spectacle	interrupt
			invent

Learning morphemes

unknown
unfair
unable
untested
unafraid
unconvinced
unaware
unlikely
unpaid
untrue
unselfish
unemployed

peerage
corkage
vicarage
dotage
voltage
package
spillage
breakage
spoilage
parsonage
vassalage
sewerage

proclaim
prodigy
prolapse
prolific
promote
prolong
propel
prorogue
prospect
pronoun
proceed
prohibit

- Must have multiple exemplars (types) *(Tamminen et al., 2015)*
- Must have consistent meaning transformation *(Tamminen et al., 2015)*
- Must be able to identify meaningful parts

What does children's exposure to morphology in text look like?

The CYP-LEX Project

ROYAL HOLLOWAY UNIVERSITY OF LONDON

National reading surveys, publisher data, & book sales statistics

1,200 popular fiction & non-fiction e-books, 400 books per age band

~70 million tokens; 105,694 types

7-9 years

10-12 years

13+ years

Korochkina et al., 2024

Books contain many complex words

Based on words available in MorphoLex *(Sánchez-Gutiérrez et al., 2017)*

	7-9	10-12	13+
Number of unique words	52,851	70,945	90,980
Number of words in MorphoLex	39,149	47,363	54,557
Morph-complex (%)	17,634 **(45%)**	22,564 **(48%)**	27,555 **(51%)**
One or more suffixes (%)	11,559 **(66%)**	14,865 **(66%)**	18,587 **(67%)**
One or more prefixes (%)	4,775 **(27%)**	6,328 **(28%)**	8,105 **(29%)**

- Roughly half of word types are morphologically-complex.
- Increasing percentage as books become more advanced.
- Much greater exposure to suffixed than prefixed words.

But fewer high-frequency complex words

ROYAL HOLLOWAY UNIVERSITY OF LONDON

	7-9	10-12	13+
Number of words in MorphoLex (all)	39,149	47,363	54,557
Morph-complex (%)	17,634 **(45%)**	22,564 **(48%)**	27,555 **(51%)**
Number of words in MorphoLex (10+)	19,769	27,271	35,034
Morph-complex – 10+ occurrences (%)	6,831 **(35%)**	10,540 **(39%)**	14,906 **(43%)**
Number of words in MorphoLex (50+)	9,512	14,047	19,455
Morph-complex – 50+ occurrences (%)	2,636 **(25%)**	4,128 **(29%)**	6,702 **(34%)**

- Readers encounter many morphologically-complex words, but few are repeated frequently.

Unique source of morpheme information

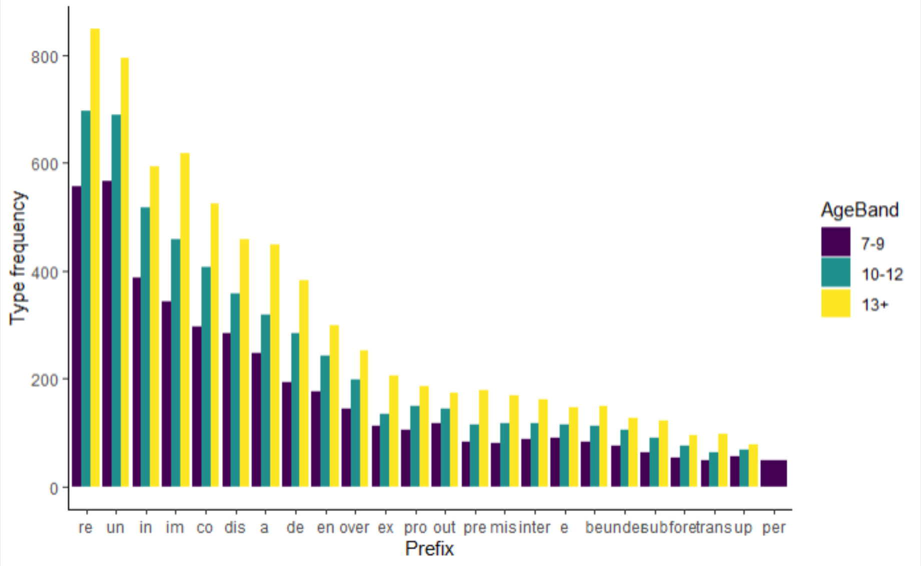
ROYAL HOLLOWAY UNIVERSITY OF LONDON

	7-9	10-12	13+
Number of words in MorphoLex	39,149	47,363	54,557
Number missing from CBBC	8,280	14,050	20,105
Morph-complex (%)	4,924 **(59%)**	8,562 **(61%)**	12,894 **(64%)**
Number missing from SUBTLEX	1,211	2,450	4,602
Morph-complex (%)	888 **(73%)**	1,796 **(73%)**	3,514 **(76%)**

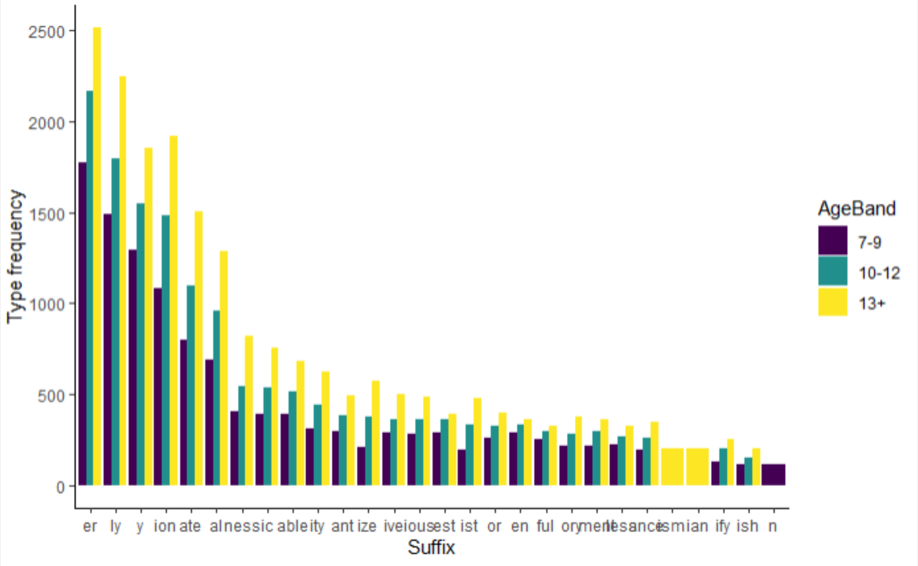
- Most unfamiliar words that children encounter in books are morphologically-complex.
- Books may be an important source of morpheme information.

Only a few affixes very common

Prefixes



Suffixes

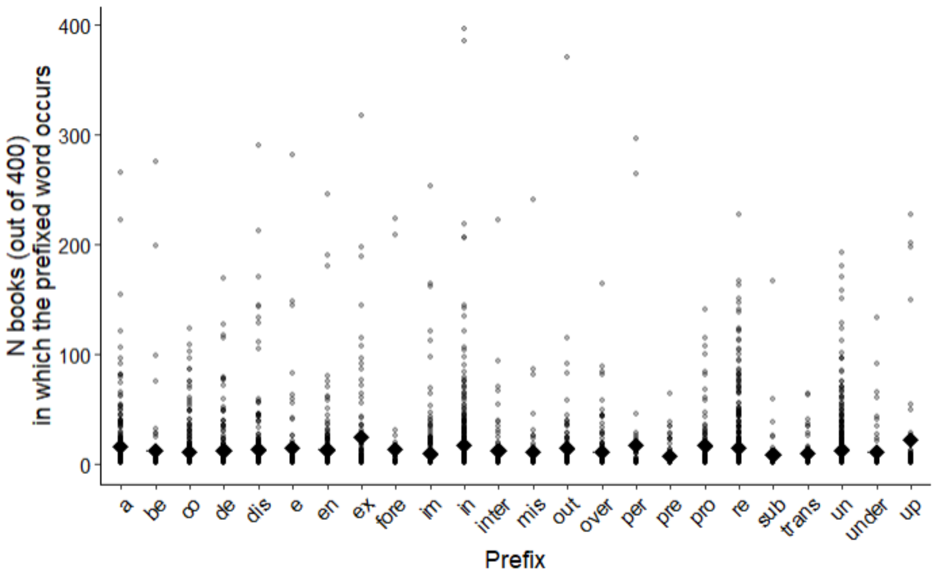


Note differences in Y-axis scale!

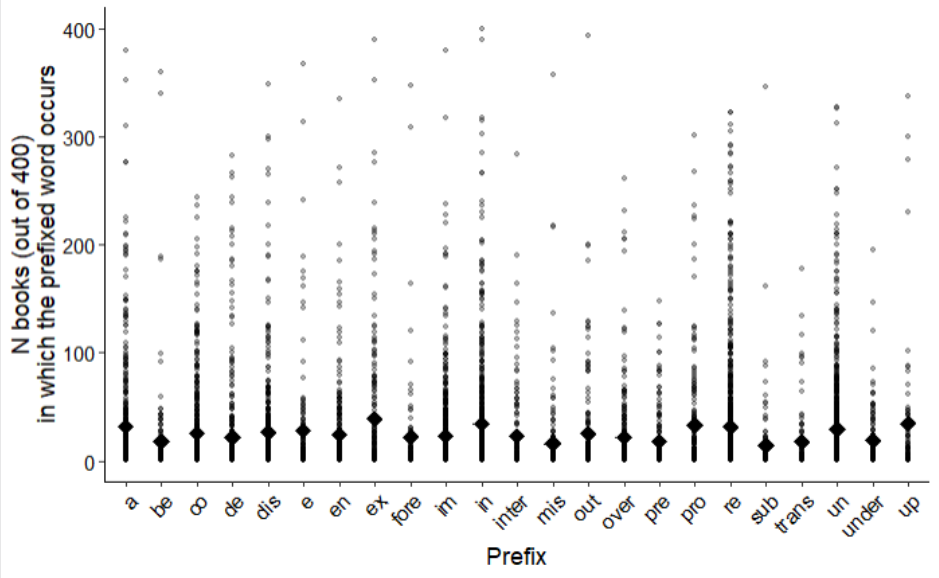
- Prefixes: un-, re-
- Suffixes: -er, -ly, -y, -ion, -ate, -al, -ness, -able, -ic
- Limited exposure to multiple types before 13+ text

Prefixes sparsely represented across books

7-9



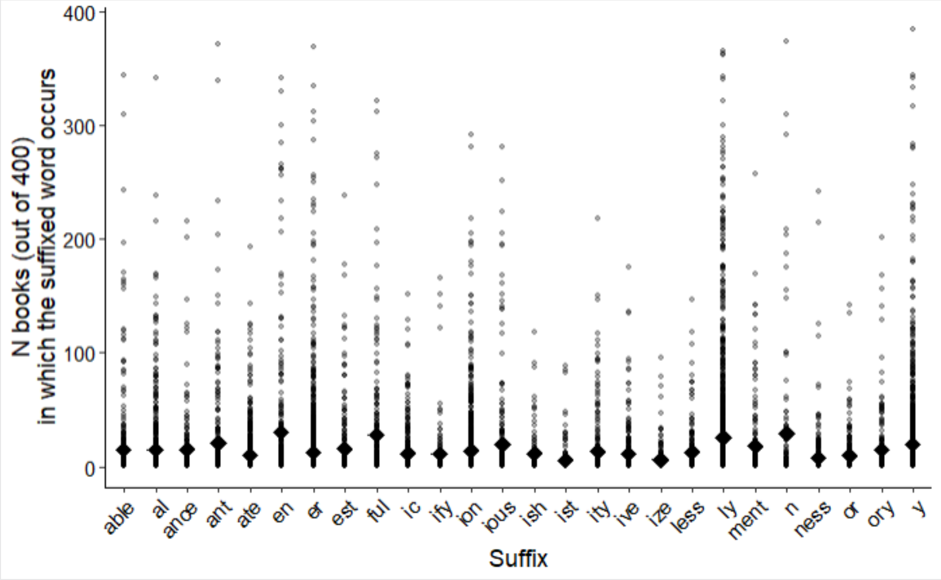
13+



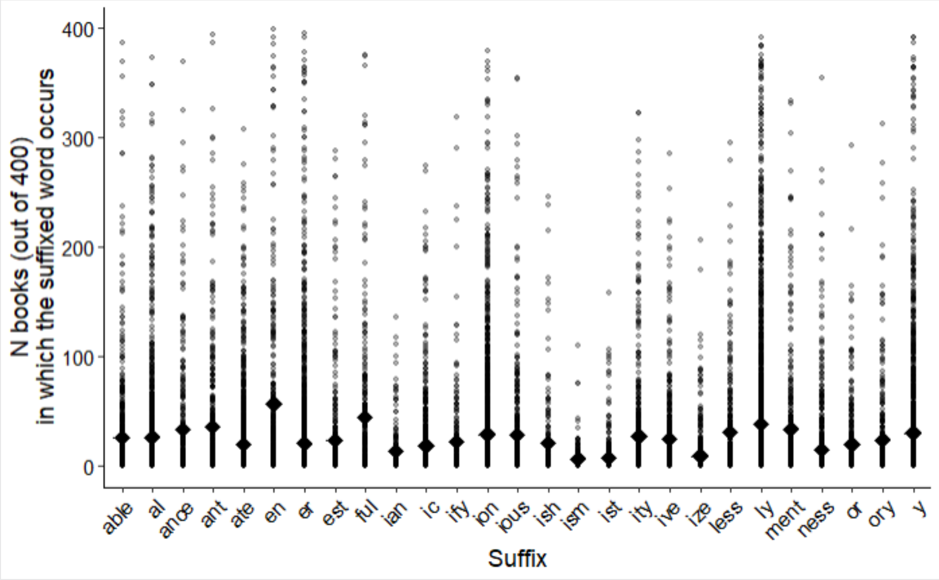
- The average prefixed word does not occur in many books
- re-, un-, and in- has reasonable representation in the 7-9 corpus
- More chance of exposure to different prefixed types in the 13+ corpus

Suffixes sparsely represented across books

7-9



13+



- Suffixes better represented across different books
- The average suffixed word does not occur in many books
- -ly, -y, and –er has reasonable representation in the 7-9 corpus
- More chance of exposure to different suffixed types in the 13+ corpus

How easy is it to “find” the morphemes?

Statistics thus far based on morphology defined etymologically (in the dictionary). How does the picture change when morphemes are defined orthographically?

Built RegEx to detect cases that *appear* to have morphological structure

- Recursive search of legal stem & affix combinations
- Sensitive to common orthographic alterations in morpheme combination

Out of 54,557 words in 13+ corpus available in MorphoLex

	MorphoLex Complex	RegEx (hits)	RegEx (FAs)
Prefixed	8,105	3,811 (47%)	1,510 (3.3%)
Suffixed	18,587	8,801 (47%)	1,735 (4.8%)

- Hits are low because of missing ‘stems’ (e.g. pessimist, exclude) and complex alterations (e.g. sustain -> [sub][tenere])
- False alarms arise because of pseudoaffixation (e.g. corner)

How easy is it to "find" the morphemes?

ROYAL HOLLOWAY UNIVERSITY OF LONDON

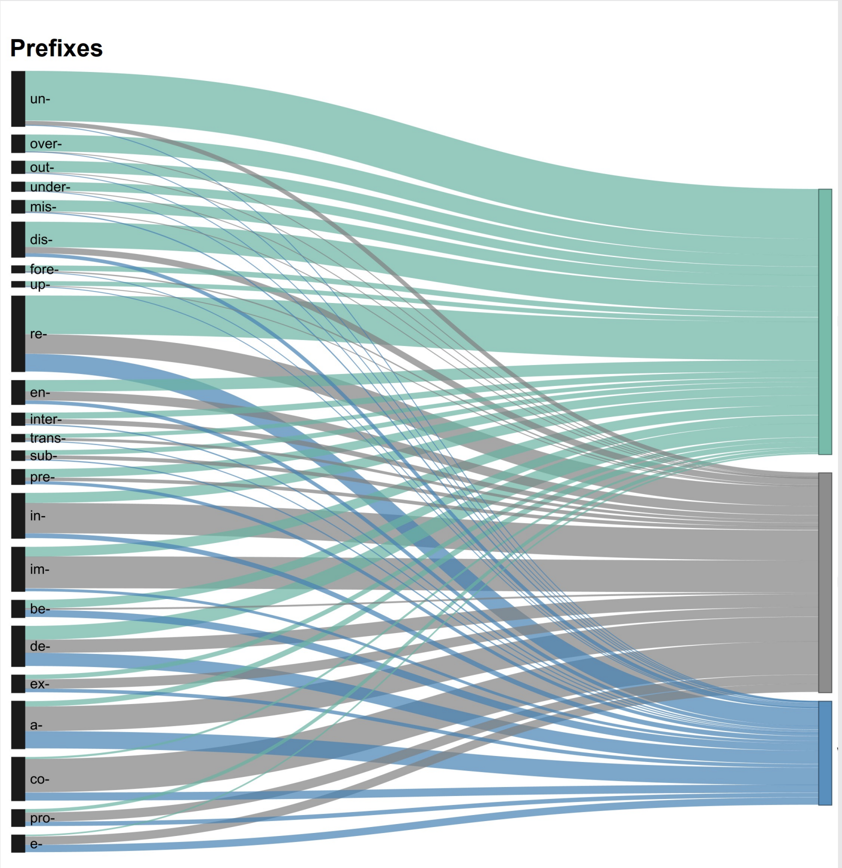
Substantial variation across affixes in how each the morpheme components can be "found" via a simple orthographic algorithm.

Examples from 13+ corpus

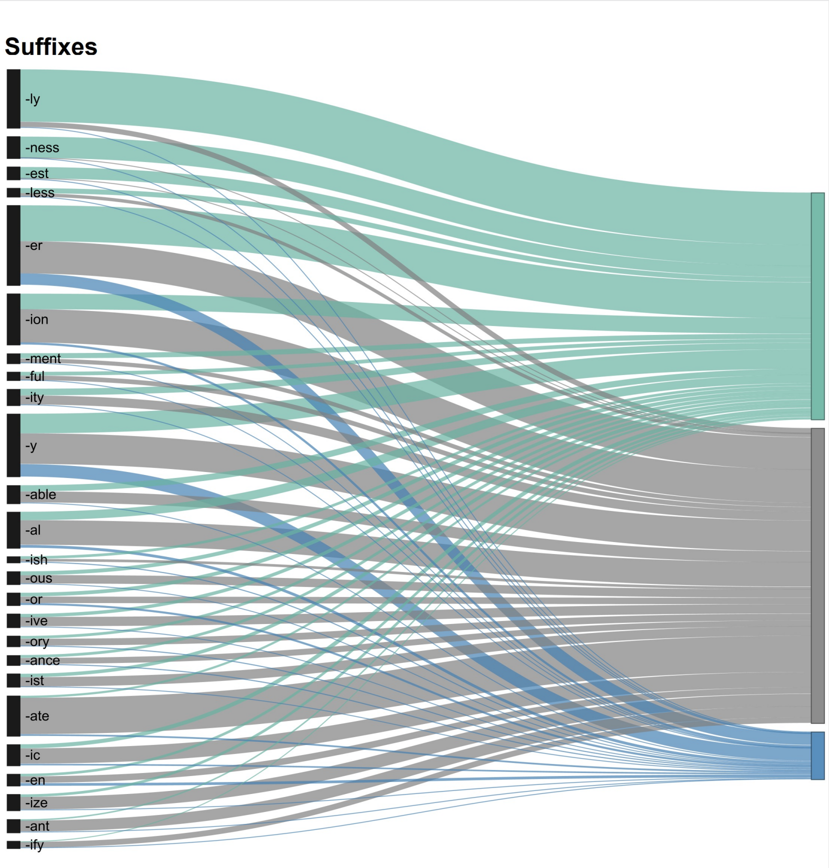
	MorphoLex Complex	RegEx (hits)	RegEx (FAs)
a-	449	80	264
un-	794	729	13
-y	1,850	790	447
-ness	823	812	3

Statistical learning of affixes depends on more than exposure to orthographic chunks; learning may depend on being able to detect a reliable transformation of the stem

How easy is it to “find” the morphemes?



Prefixes



Suffixes

- ● Genuine complex words detected with RegEx
- ● Genuine complex words not detected with RegEx
- ● Simple words identified as complex with RegEx

Conclusions

Morphologically-complex words comprise a large proportion of words in children's books, but morpheme knowledge beyond a handful of affixes will be difficult to acquire from text experience (low frequency, sparse representation, parsing problems, pseudoaffixation).

One consequence may be that children do not show evidence of morpheme knowledge in online reading tasks until late adolescence (age 15-16, >10 years reading experience)

- Morpheme interference effect *(Dawson et al., 2017)*
- Morpheme masked priming *(Dawson et al., 2021)*

Morphemes are graded along several dimensions: type frequency, reliability of communicating meaning, and the ease of detecting morpheme constituents. Important to study these properties and their relationship to learning.

Quick to ascribe distributional knowledge as the result of "statistical learning", but we need to understand why that knowledge seems to be so much more difficult to acquire than in laboratory studies, and why it's so hard to link lab performance to real-world outcomes.

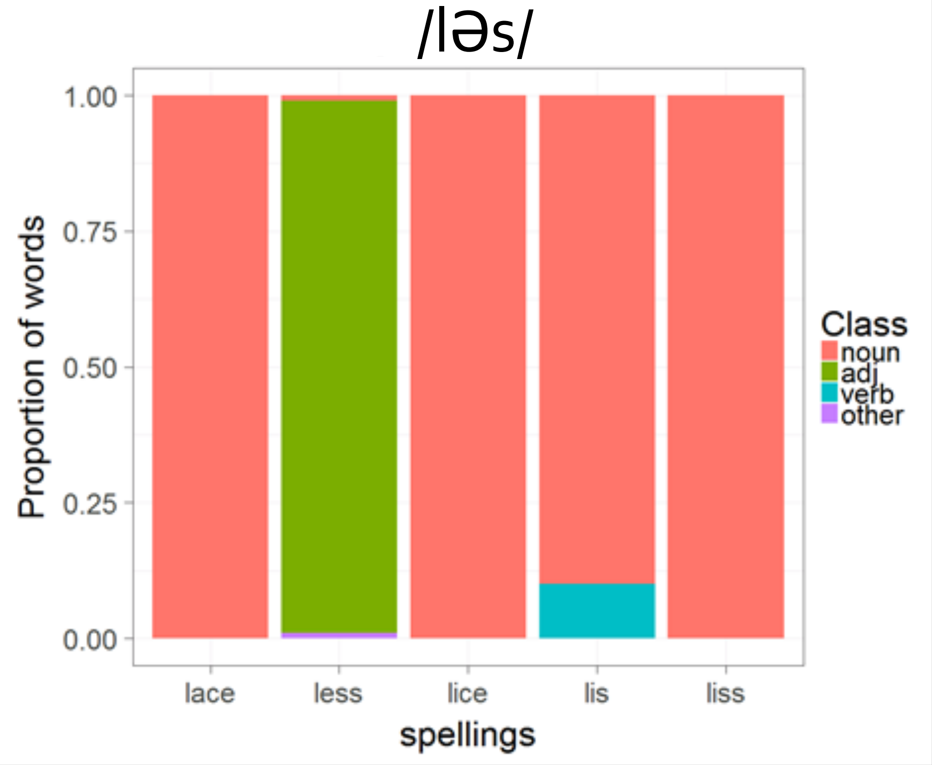
Thank you!

Morphology in English spelling

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Morphology may be highly "visible" in English spelling

- Immediate knowledge of part of speech (*object*, *property*, *act)* status for substantial % of English words.
- OS systematicity trades against OP systematicity; this information is often not available in spoken language
- Systematicity arises across English suffixes, but strength of OS relationship is graded.



Ulicheva, Harvey, Aronoff, & Rastle, 2020, Cognition