

Dr Maria Korochkina

Modelling affix learning from reading: Insights from compositional distributional semantics

[TeaP](#) @ Tübingen
17 March 2026

Pre-requisites for affix learning

- **High type frequency** – affix must occur within a high number of distinct words

*Count all instances where a complex-looking word is **historically formed through derivation***

<u>d</u> eactivate	<u>d</u> emand
<u>d</u> ecode	<u>d</u> eceive
<u>d</u> ecompose	<u>d</u> epend
<u>d</u> efrost	<u>d</u> elete
<u>d</u> ehydrate	<u>d</u> eliver
<u>d</u> evalue	<u>d</u> etail
<u>d</u> ethrone	<u>d</u> efeat

Pre-requisites for affix learning

- **High type frequency** – affix must occur within a high number of distinct words

Count all instances where a complex-looking word is **historically formed through derivation**

deactivate
decode
decompose
defrost
dehydrate
devalue
dethrone

demand
deceive
depend
delete
deliver
detail
defeate



Pre-requisites for affix learning

- **High type frequency** – affix must occur within a high number of distinct words

Count all instances where a complex-looking word is **historically formed through derivation**

deactivate
decode
decompose
defrost
dehydrate
devalue
dethrone

demand
deceive
depend
delete
deliver
detail
defeat



Pre-requisites for affix learning

- **High type frequency** – affix must occur within a high number of distinct words

Count all instances where a complex-looking word is **historically formed through derivation**

deactivate
decode
decompose
defrost
dehydrate
devalue
dethrone

demand
deceive
depend
delete
deliver
detail
defeat



Pre-requisites for affix learning

- **High type frequency** – affix must occur within a high number of distinct words

Count all instances where a complex-looking word is **historically formed through derivation**

deactivate
decode
decompose
defrost
dehydrate
devalue
dethrone

demand
deceive
depend
dele
deliver
detail
defeat



Pre-requisites for affix learning

- **High type frequency** – affix must occur within a high number of distinct words

Count all instances where a complex-looking word is **historically formed through derivation**

deactivate
decode
decompose
defrost
dehydrate
devalue
dethrone



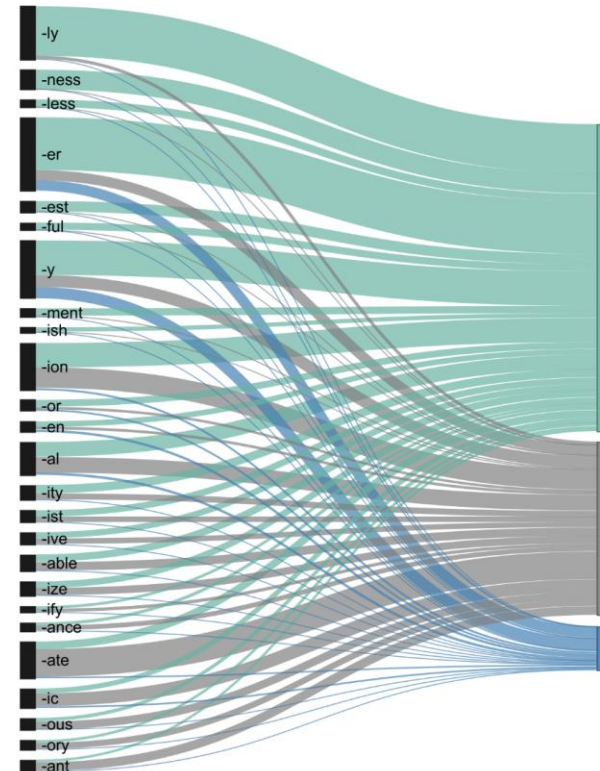
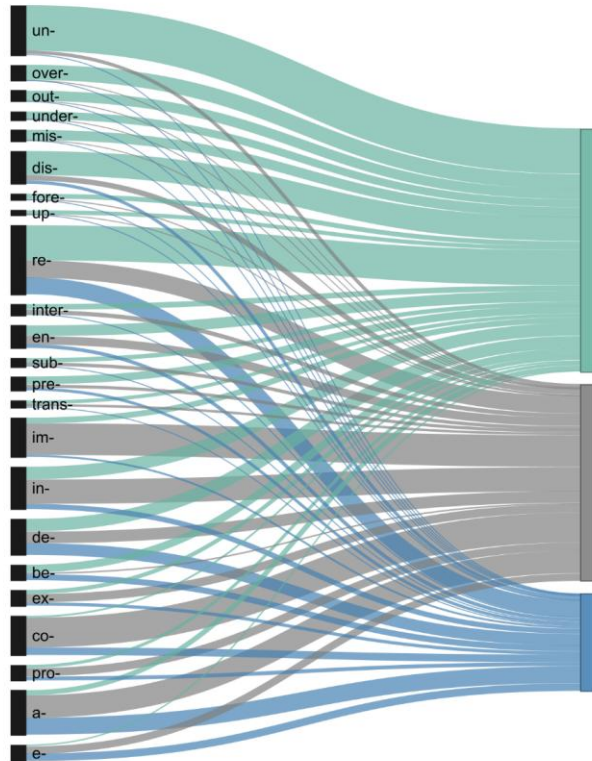
demand
deceive
depend
dele
deliver
detail
defeat



Count only those instances where affixes are **identifiable without specialised knowledge**

Affixes vary in detectability and false-alarm rates

Data from
a 70-mln-word
corpus of children's
books (CYP-LEX)



Pre-requisites for affix learning

- **High type frequency** – affix must occur within a high number of distinct words
 - This number changes if we take readers' perspective into account
- Affixes must induce a **consistent meaning transformation**

teacherr
bankerr

Pre-requisites for affix learning

- **High type frequency** – affix must occur within a high number of distinct words
 - This number changes if we take readers' perspective into account
- Affixes must induce a **consistent meaning transformation**

teacherr
bankerr
dresserr

Pre-requisites for affix learning

- **High type frequency** – affix must occur within a high number of distinct words
 - This number changes if we take readers' perspective into account
- Affixes must induce a **consistent meaning transformation**

teacherr
bankerr
dresserr
cornerr

Pre-requisites for affix learning

- **High type frequency** – affix must occur within a high number of distinct words
 - This number changes if we take readers' perspective into account
- Affixes must induce a **consistent meaning transformation**
 - Affixes communicate meaning in a quasiregular way



What can readers learn about affix meanings through reading?



Compositional distributional semantics

CAOSS: Compounding as Abstract Operation in Semantic Space

undo

CAOSS: Compounding as Abstract Operation in Semantic Space



un

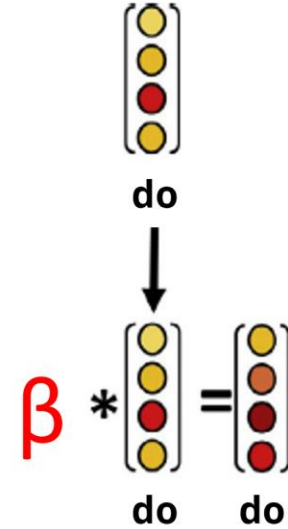
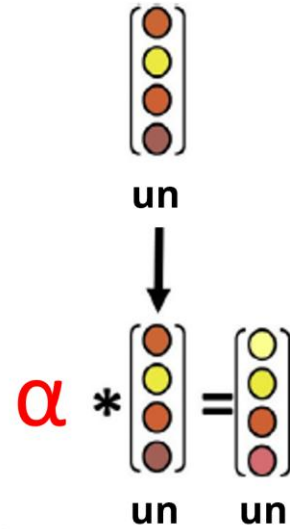
undo



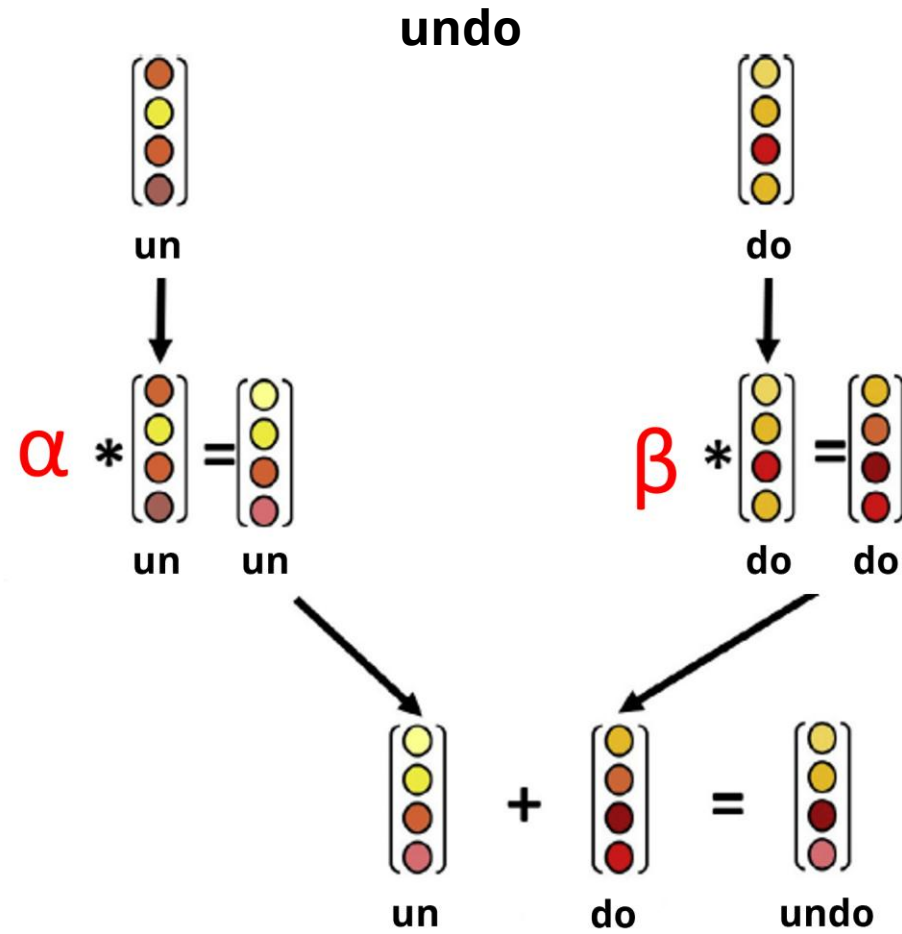
do

CAOSS: Compounding as Abstract Operation in Semantic Space

undo



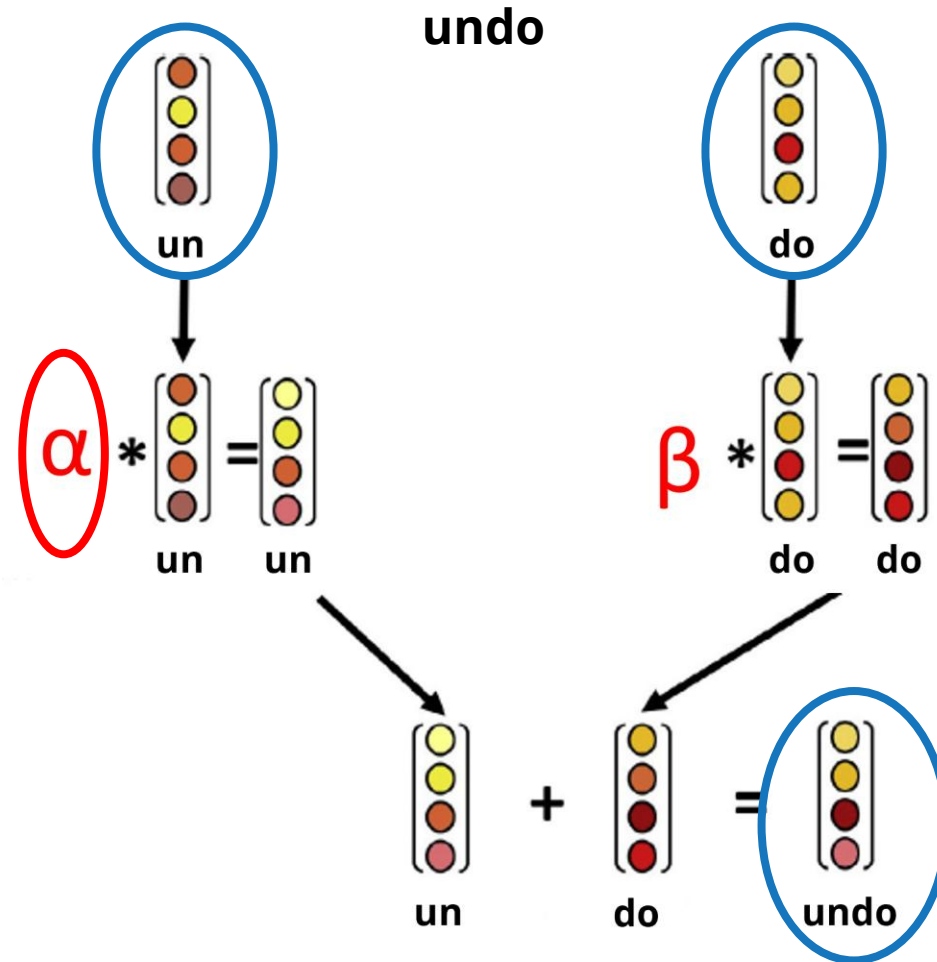
CAOSS: Compounding as Abstract Operation in Semantic Space



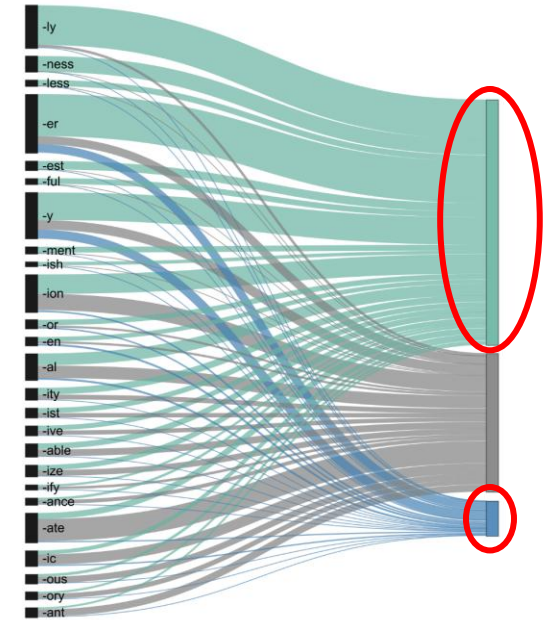
Our modelling approach

Lexicalised representations
for *words* & *affixed words*
from *subs2vec*
(van Paridon & Thompson, 2021)

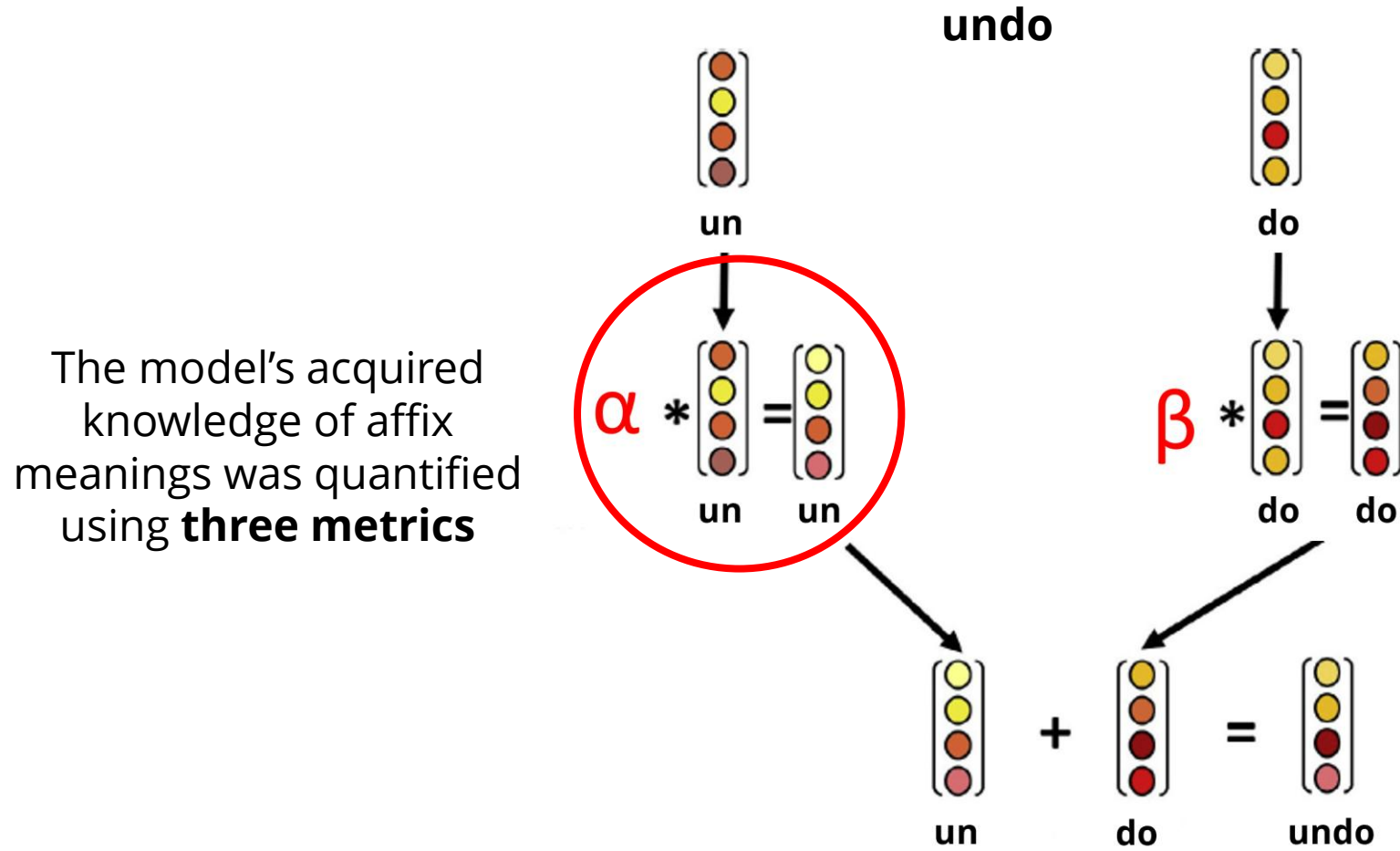
Affix representations
Average of vector representations of all
words with this affix
(Westbury & Hollis, 2019)



Training set



Our modelling approach



Affix diffuseness
Degree of diffusion and uncertainty in the affix meaning

Affix richness
Richness and complexity of affix meaning

Affix-word coherence
Similarity between the meaning of the affix and the meanings of words that contain it

The morpheme interference paradigm

woodness

word not a word

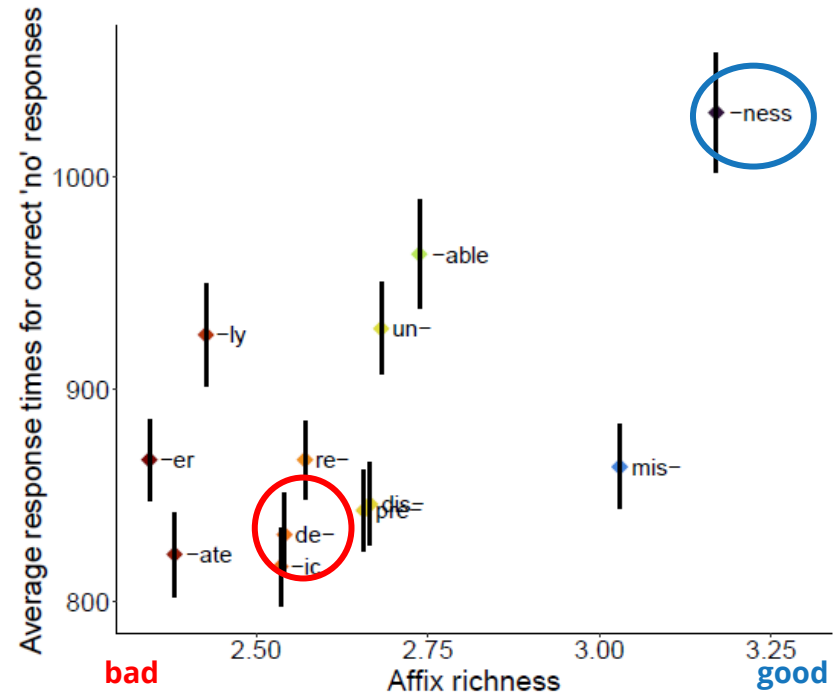
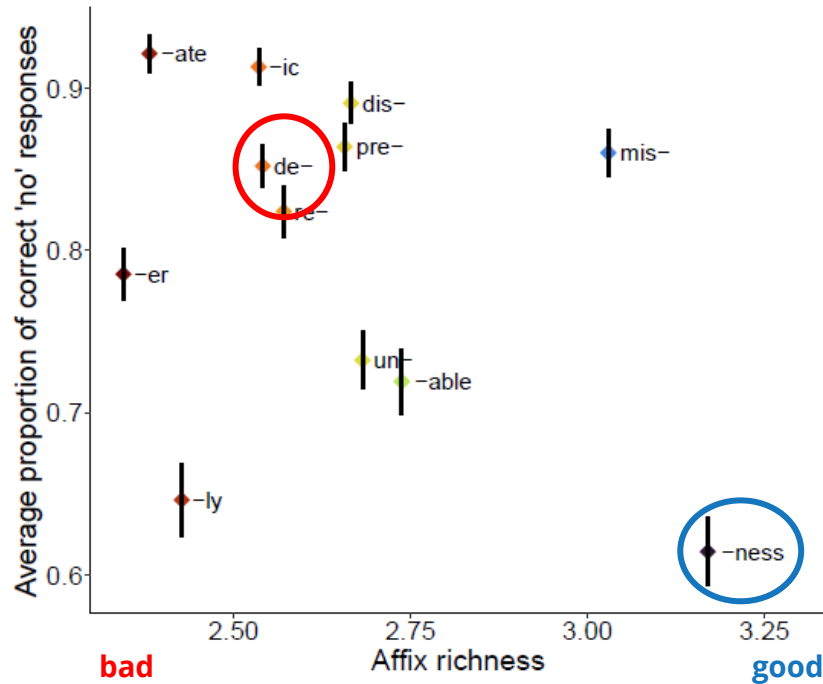
woodnels

word not a word

Readers are **less accurate and slower**
at rejecting morphologically structured nonwords (*woodness*)

- 6 prefixes: *un-*, *mis-*, *dis-*, *pre-*, *de-*, *re-*
- 6 suffixes: *-ness*, *-ly*, *-able*, *-er*, *-ic*, *-ate*

Nonwords with “better” affixes are harder to reject



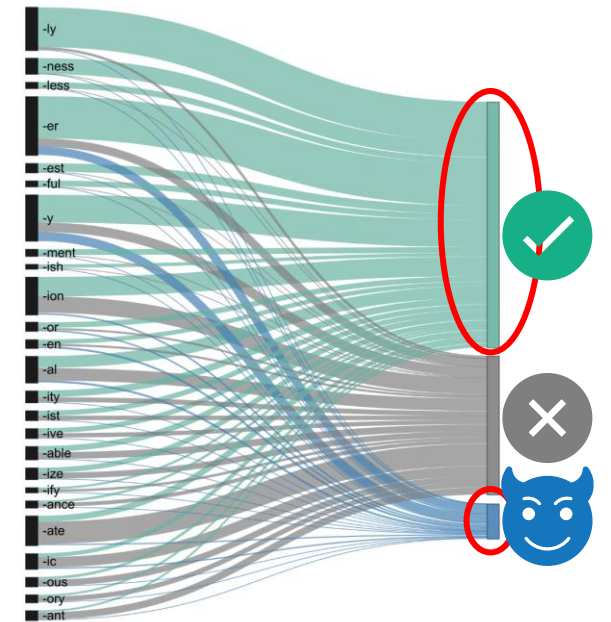
- Same result for the other two metrics (*affix-word coherence*, *affix diffuseness*)
- CAOSS captures systematic patterns in readers' lexical processing

Conclusions

- **Principled & scalable** account of **morpheme learning in the wild**
- **Innovation** in computational **modelling** of affix semantics
 - **First** use of CAOSS with **“noisy”** input
 - **First** attempt to model **prefix semantics**

→ Natural text has **enough structure to support learning** of some core aspects of affix semantics...

→ ...and it shapes both our knowledge of individual affixes and the perceived plausibility of morphemic combinations containing them



Article | [Open access](#) | Published: 05 May 2025

Morphology in children's books, and what it means for learning

[Maria Korochkina](#)  & [Kathleen Rastle](#)


npj Science of Learning **10**, Article number: 22 (2025) | [Cite this article](#)

4985 Accesses | 23 Altmetric | [Metrics](#)

Psychonomic Bulletin & Review (2025) 33:3
<https://doi.org/10.3758/s13423-025-02830-2>

BRIEF REPORT

Morpheme knowledge is shaped by information available through orthography

Maria Korochkina¹  · Holly Cooper¹  · Marc Brysbaert²  · Kathleen Rastle¹ 

Journal of Memory and Language 148 (2026) 104746



ELSEVIER

Contents lists available at [ScienceDirect](#)

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml



Morphemes in the wild: Modelling affix learning from the noisy landscape of natural text

Maria Korochkina ^a , Marco Marelli ^b , Kathleen Rastle ^a 

^a Department of Psychology, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, United Kingdom

^b Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, Milan, 20126, Italy



Holly Cooper



Marco Marelli



Marc Brysbaert



Kathy Rastle

Thank you!

m.korochkina@aston.ac.uk

<https://mariakna.github.io/>