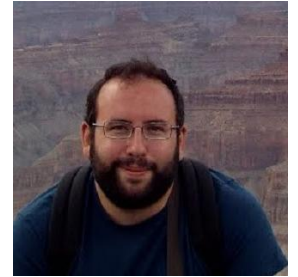




Kathy Rastle



Marc Brysbaert



Marco Marelli

Maria Korochkina & Holly Cooper

The graded nature of morphological information in text, and what it means for morpheme learning

FRiLL, 12 December 2024

The power of morphology

- Most English words are built by **recombining stems and affixes**
 - cleaner, cleanly, unclean
 - teachere, bankere, buildere
- Morpheme knowledge enables rapid access to the meanings of **familiar** words
- It is also crucial for computing the meanings of **unfamiliar** words
 - bright + -ify → brightify
- Limited time for explicit teaching of morphology, so morpheme knowledge acquired primarily through **text experience**

Many complex words in children's books

7-9 years



10-12 years



13+ years



CYP-LEX: The Children and Young People's Books Lexicon

1,200 popular books
400 books per age band
Over 70 mln words
Over 100,000 distinct words

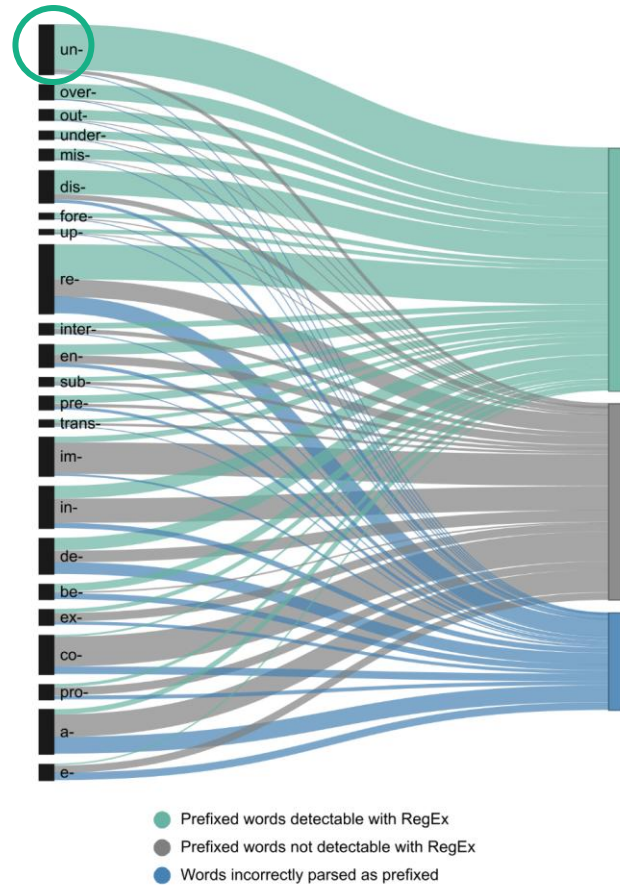
- Roughly **half of all distinct words** are complex
- **Few** complex words are **used repeatedly** or in many books
- Children are **likely to see** a complex word but **unlikely to see** this word **again**

Pre-requisites for morpheme learning

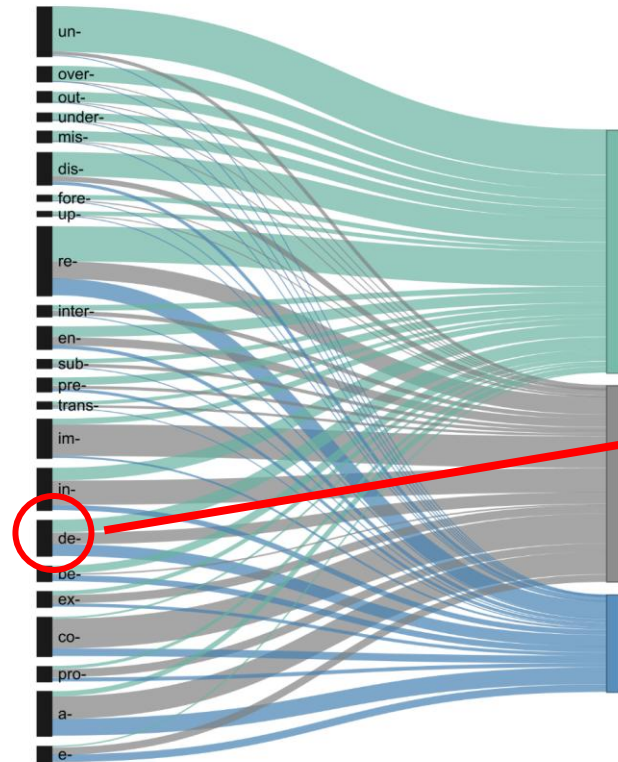
<u>u</u> nknown	<u>d</u> eactivate
<u>u</u> nfair	<u>d</u> ecode
<u>u</u> nafraid	<u>d</u> ecompose
<u>u</u> nlikely	<u>d</u> emand
<u>u</u> nconvinced	<u>d</u> eceive
<u>u</u> nsure	<u>d</u> epend
<u>u</u> nwell	<u>d</u> eliver (de + -liberare)

- Must have **consistent meaning** transformation
- Must occur with a **high number of distinct stems**
- Must be **detectable**

Few affixes are easy to detect



Few affixes are easy to detect



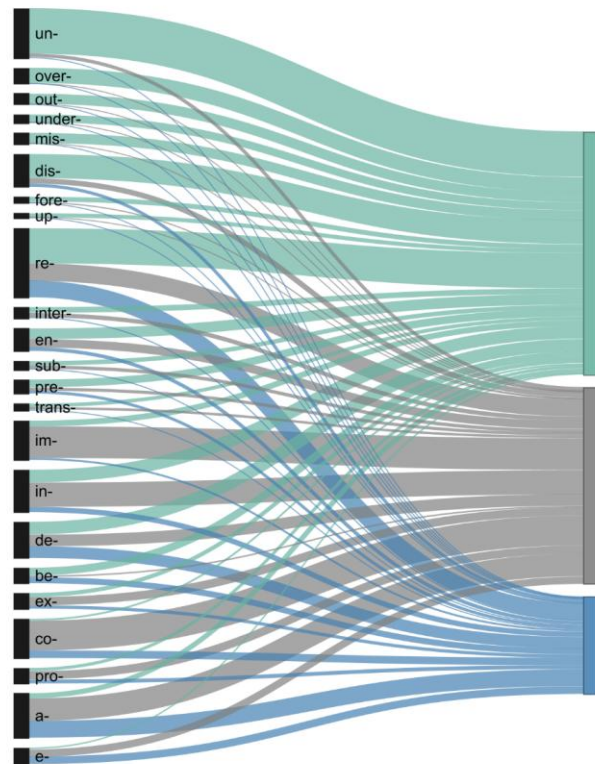
1/3 detectable
deactivate, decode, decompose

1/3 undetectable
demand, deceive, depend

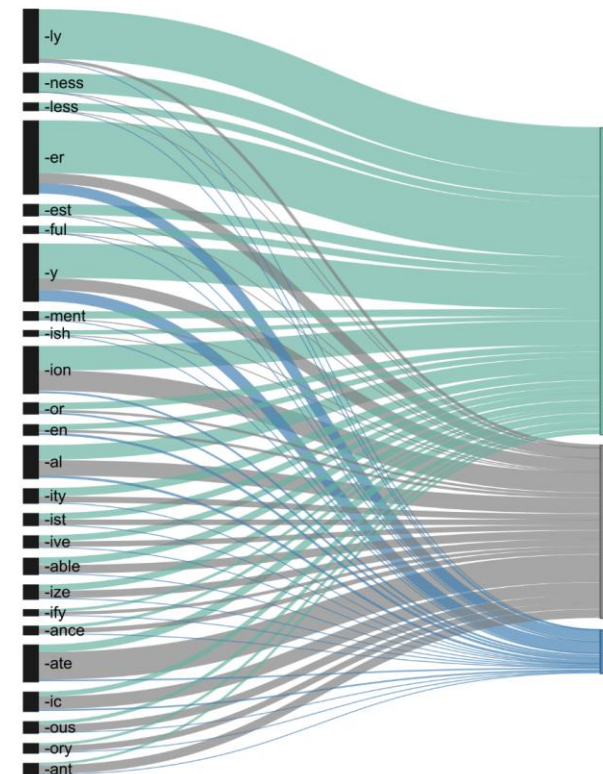
1/3 false alarms
deliver, detail, defeat

- Prefixed words detectable with RegEx
- Prefixed words not detectable with RegEx
- Words incorrectly parsed as prefixed

Few affixes are easy to detect

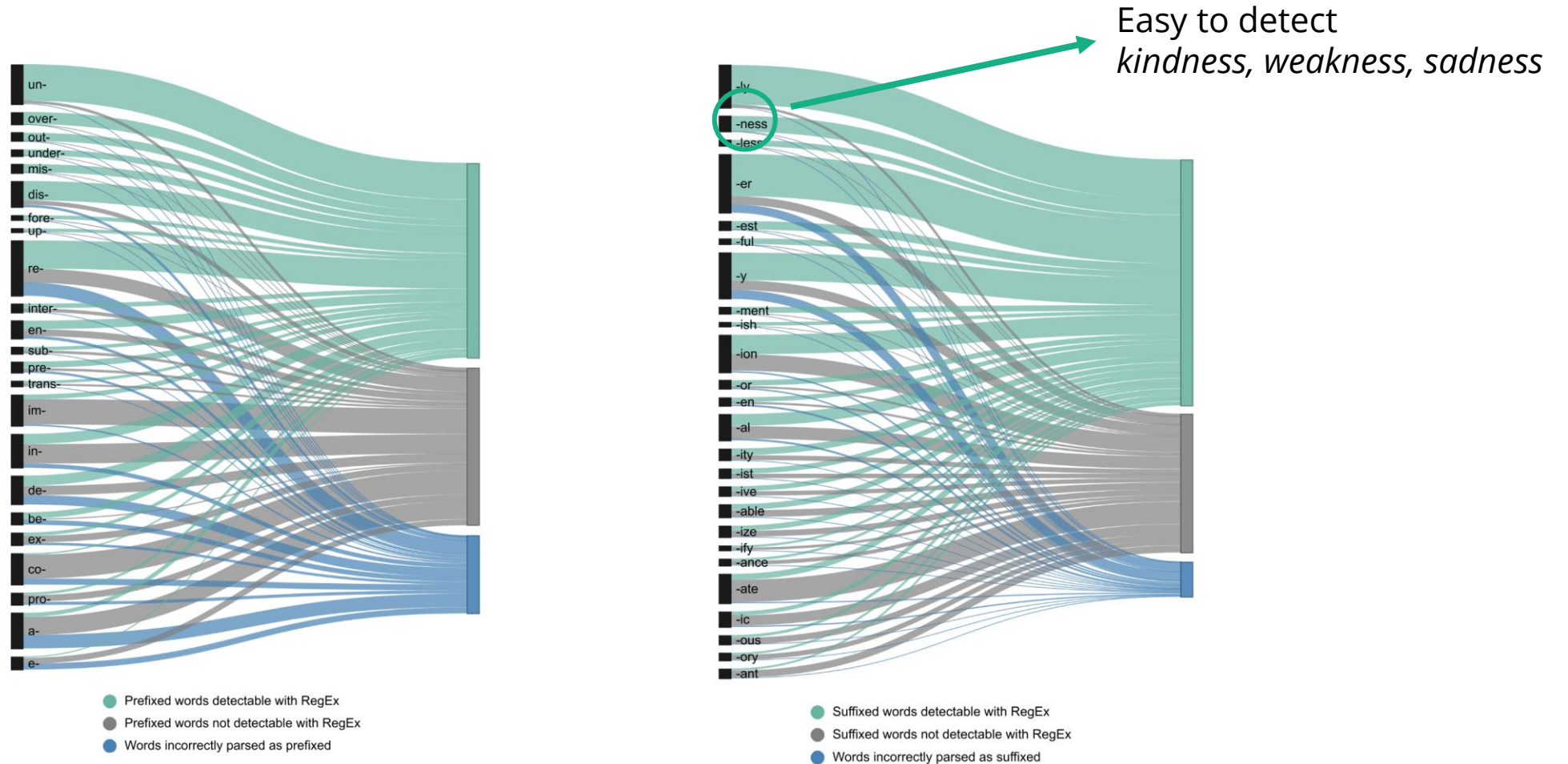


- Prefixed words detectable with RegEx
- Prefixed words not detectable with RegEx
- Words incorrectly parsed as prefixed

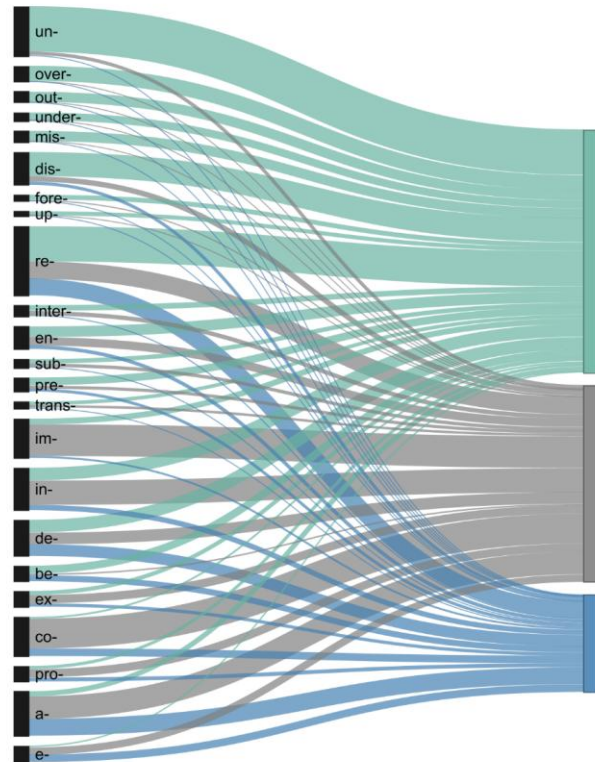


- Suffixed words detectable with RegEx
- Suffixed words not detectable with RegEx
- Words incorrectly parsed as suffixed

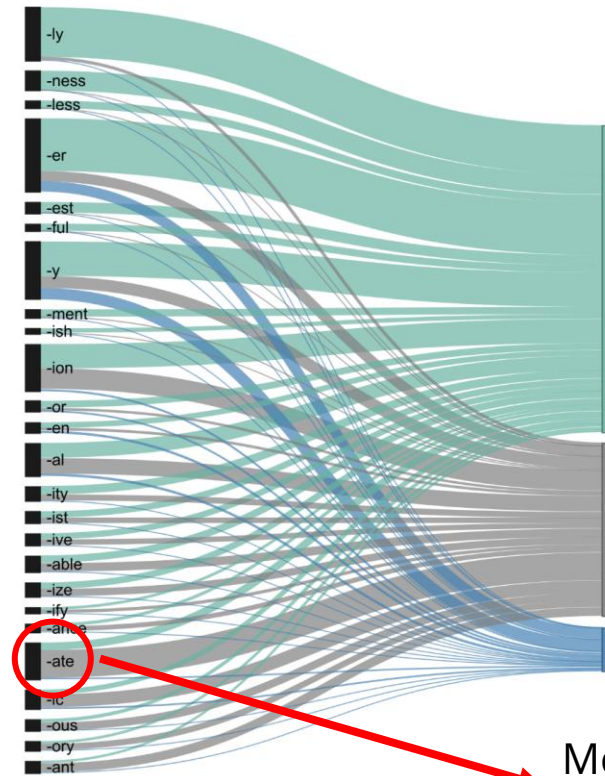
Few affixes are easy to detect



Few affixes are easy to detect



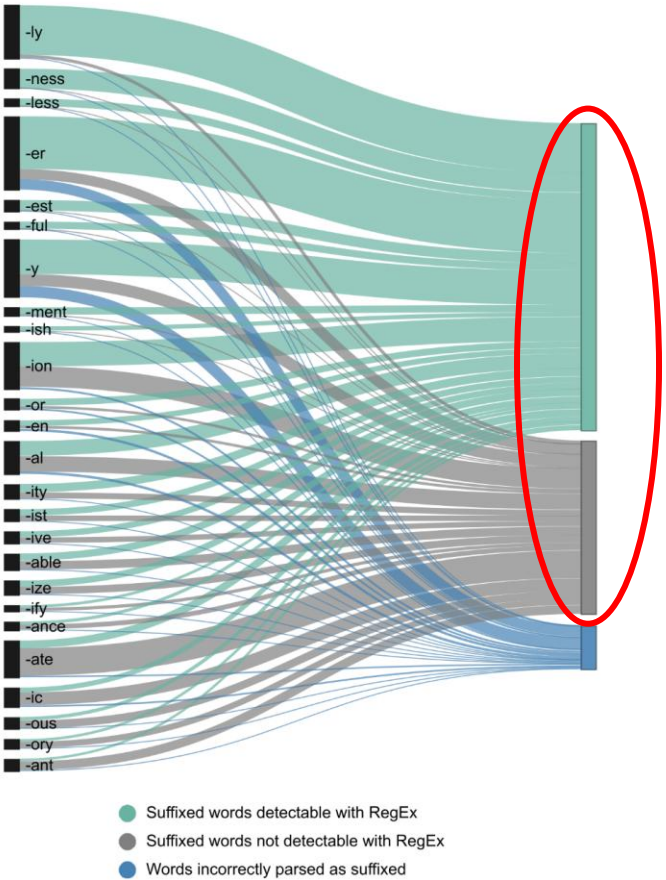
- Prefixed words detectable with RegEx
- Prefixed words not detectable with RegEx
- Words incorrectly parsed as prefixed



- Suffixed words detectable with RegEx
- Suffixed words not detectable with RegEx
- Words incorrectly parsed as suffixed

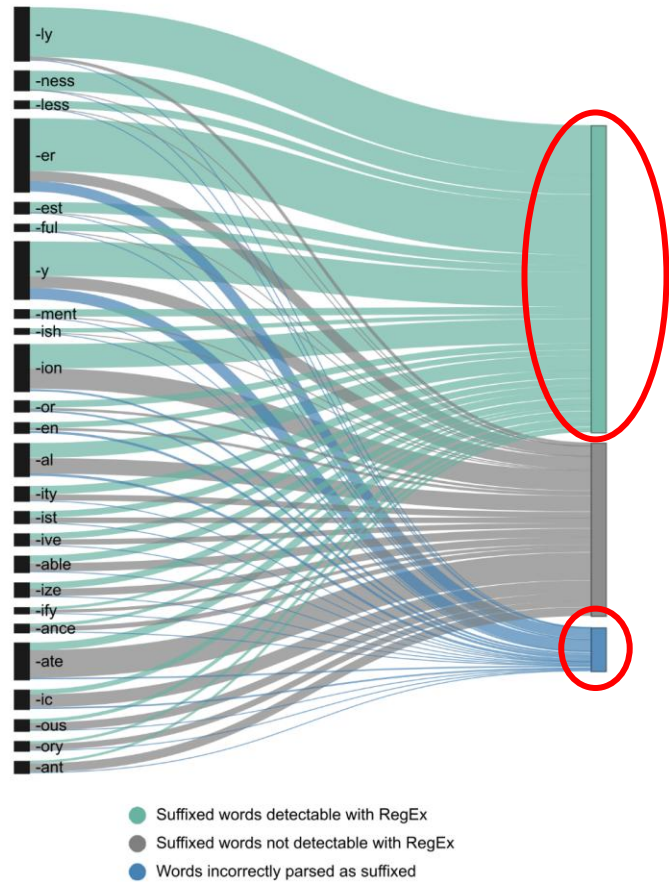
Mostly undetectable
appreciate, generate, integrate

Theories of morpheme learning



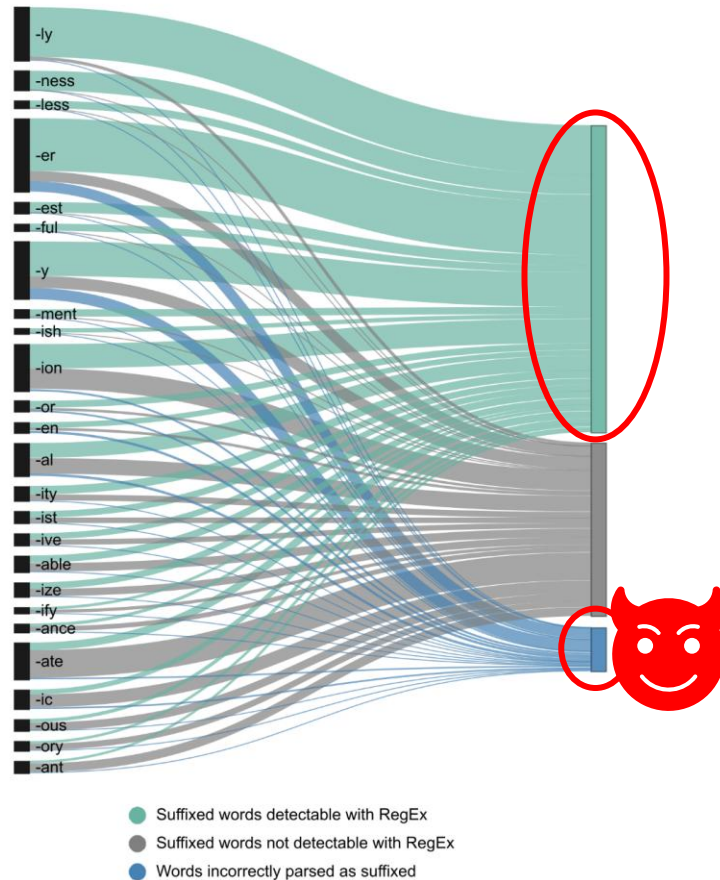
1. Morpheme learning based on **dictionary counts**

Theories of morpheme learning



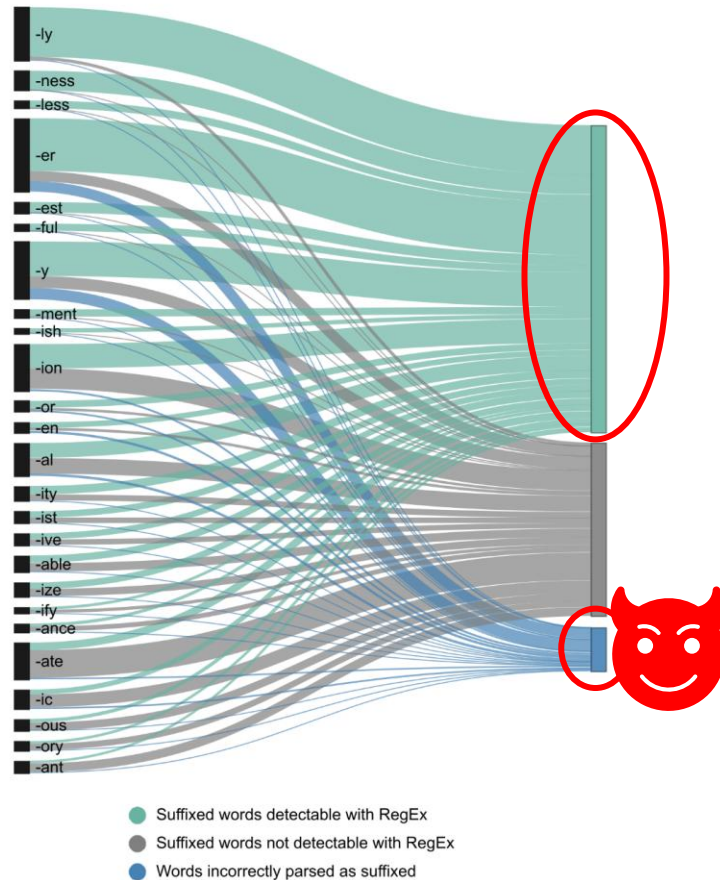
1. Morpheme learning based on **dictionary counts**
2. Morpheme learning based on **what's detectable**

Theories of morpheme learning



1. Morpheme learning based on **dictionary counts**
2. Morpheme learning based on **what's detectable**
3. Morpheme learning based on **what's detectable** but there is a **penalty for false alarms**

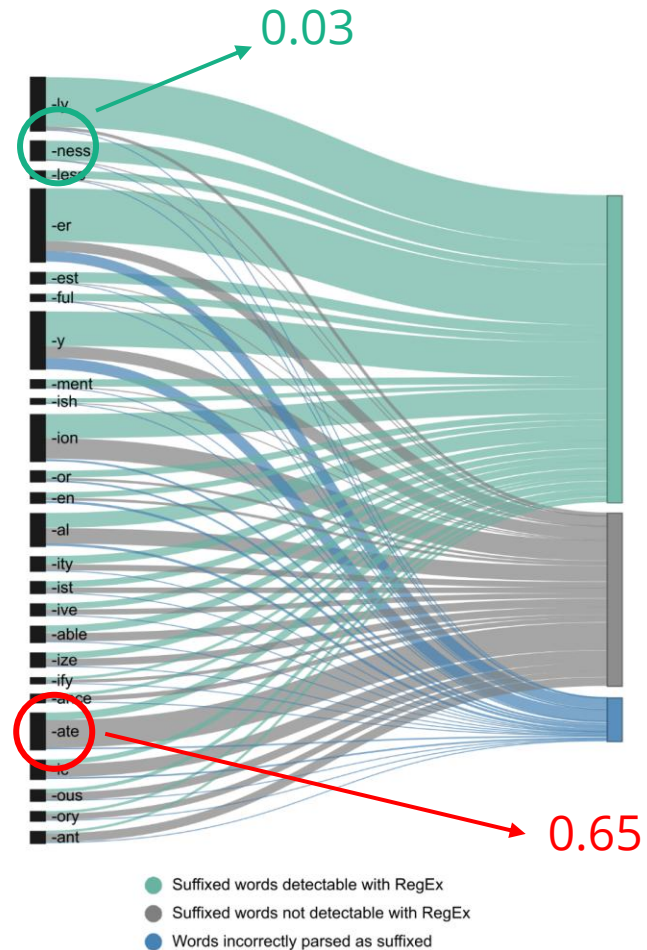
Quantifying the penalty



Shannon entropy

- Quantifies the **uncertainty** associated with identifying whether a word is **genuinely complex** or a **false alarm**

Quantifying the penalty



Shannon entropy

- Quantifies the **uncertainty** associated with identifying whether a word is **genuinely complex** or a **false alarm**
- **Low** entropy → little uncertainty
- **High** entropy → a lot of uncertainty

Theories in action

Which theory best explains human behaviour?

The morpheme interference effect

guilt**ness**

word not a word

guilt**nels**

word not a word

- Morphologically-structured nonwords are more difficult, and take longer, to reject
- Skilled readers segment complex-looking words into morphemes

Stimuli

- 6 prefixes
 - *un-, mis-, dis-, pre-, de-, re-*
- 6 suffixes
 - *-ness, -ly, -able, -er, -ic, -ate*
- Morphologically structured nonwords
 - *unguilt, guiltness*
- Nonwords with no morphological structure
 - *ubguilt, guiltnels*
- Each participant saw...
 - Each affix with 10 stems (120 morphologically structured nonwords)
 - Yoked controls (120 nonwords with no morphological structure)
 - 120 morphologically complex + 120 morphologically simple words

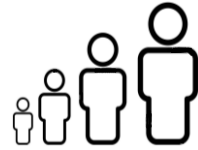
Stimuli

- 6 prefixes
 - *un-, mis-, dis-, pre-, de-, re-*
- 6 suffixes
 - *-ness, -ly, -able, -er, -ic, -ate*
- Morphologically structured nonwords
 - *unguilt, guiltness*
- Nonwords with no morphological structure
 - *ubguilt, guiltnels*
- Each participant saw **480 letter strings**
 - Each affix with 10 stems (120 morphologically structured nonwords)
 - Yoked controls (120 nonwords with no morphological structure)
 - 120 morphologically complex + 120 morphologically simple words

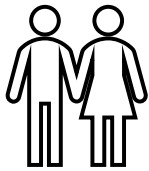
Participants



120 participants



18 – 40 years old

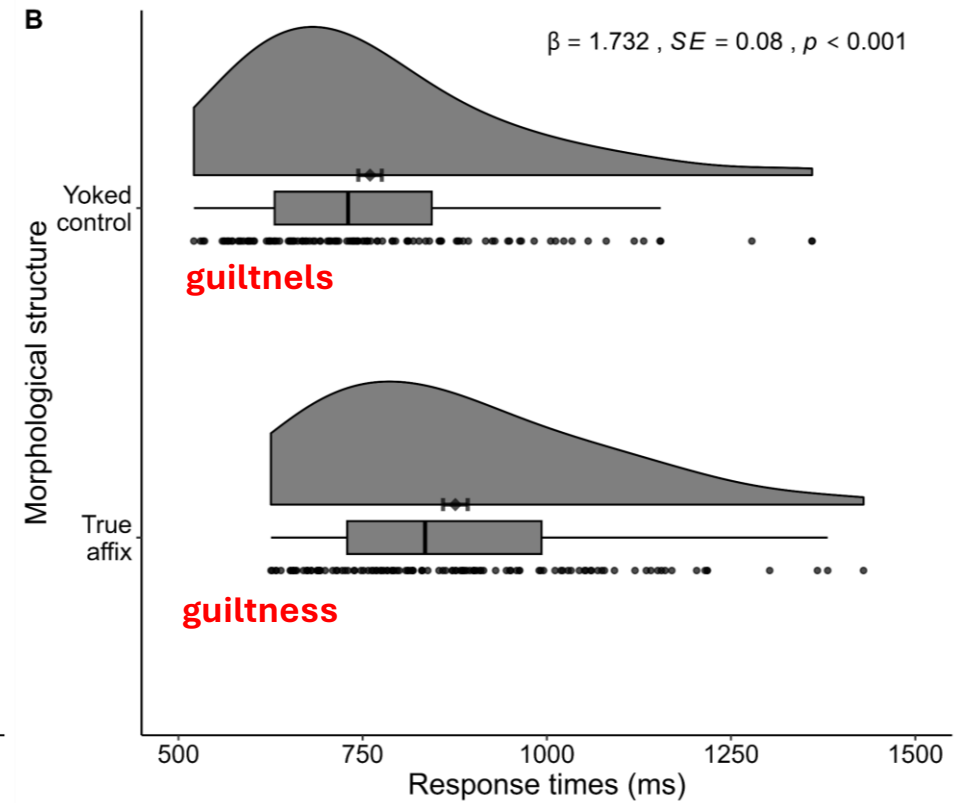
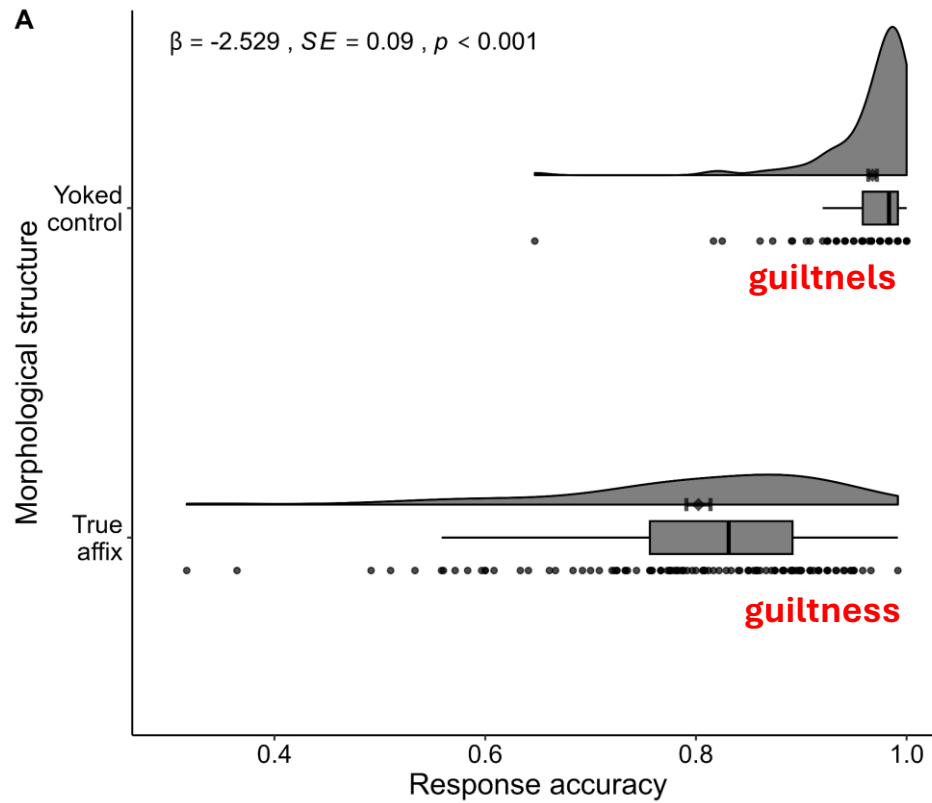


63 female
56 male
1 non-binary

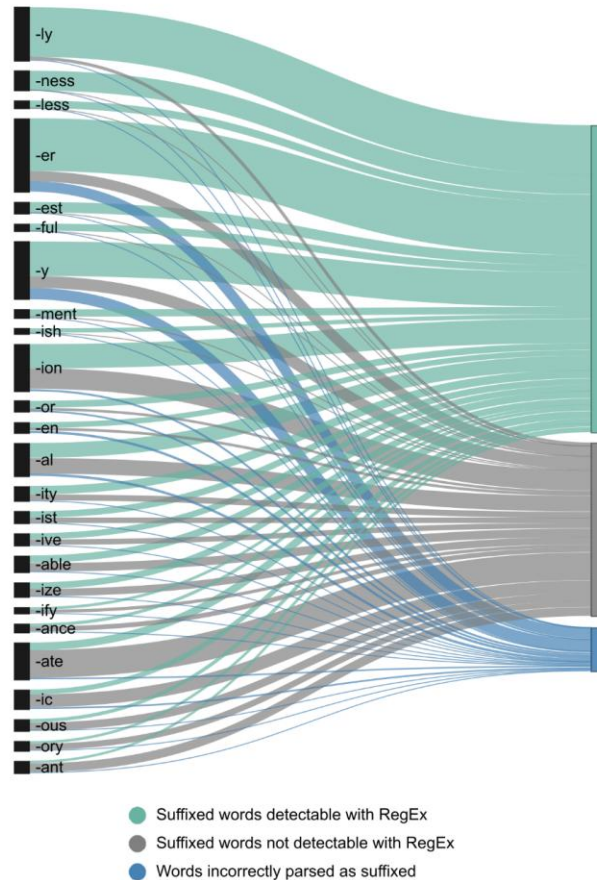


UK based
English as a first language
No language disorders

Morphological structure matters

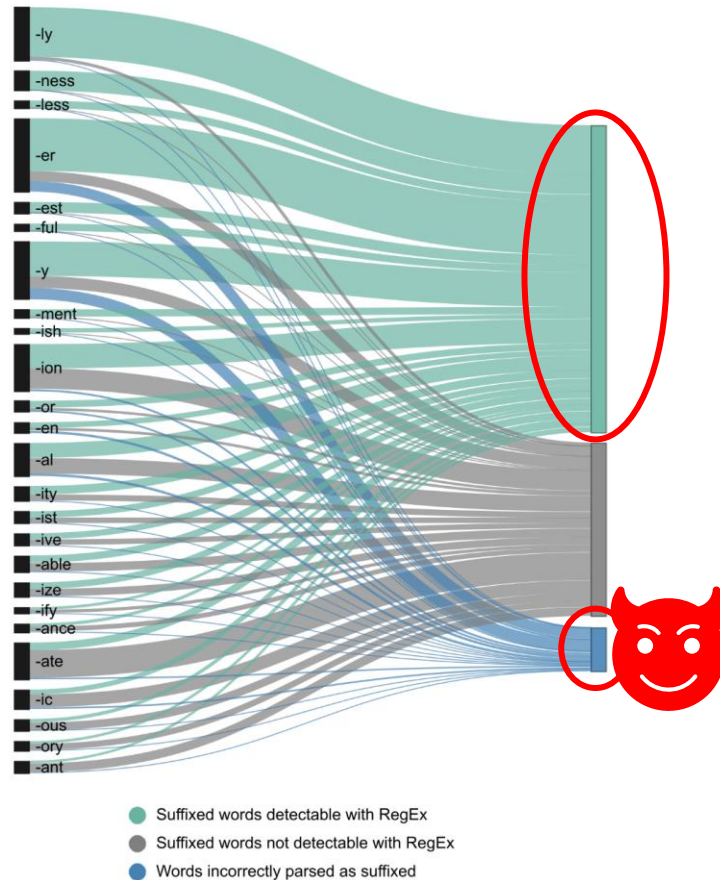


Which theory?



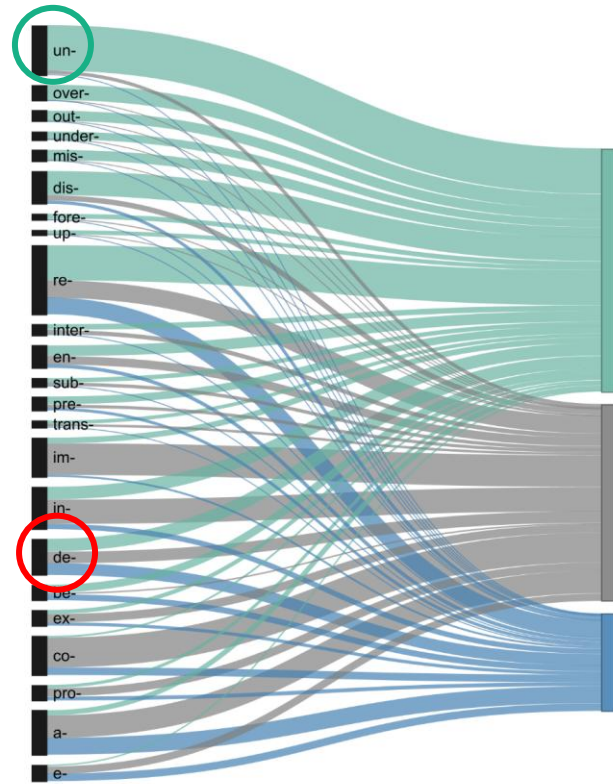
1. Morpheme learning based on **dictionary counts**
2. Morpheme learning based on **what's detectable**
3. Morpheme learning based on **what's detectable** but there is a **penalty for false alarms**

Theory 3 explains data best

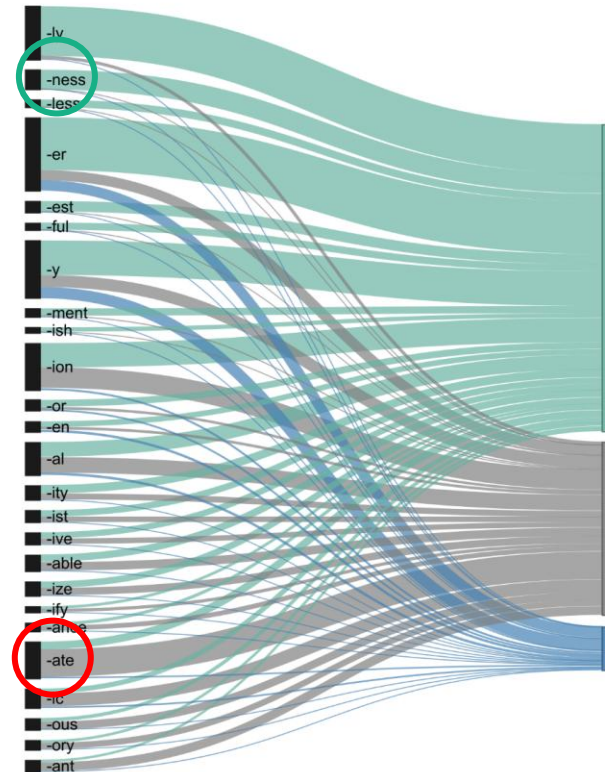


1. Morpheme learning based on **dictionary counts**
2. Morpheme learning based on **what's detectable**
3. Morpheme learning based on **what's detectable** but there is a **penalty for false alarms**

Exploring specific affixes

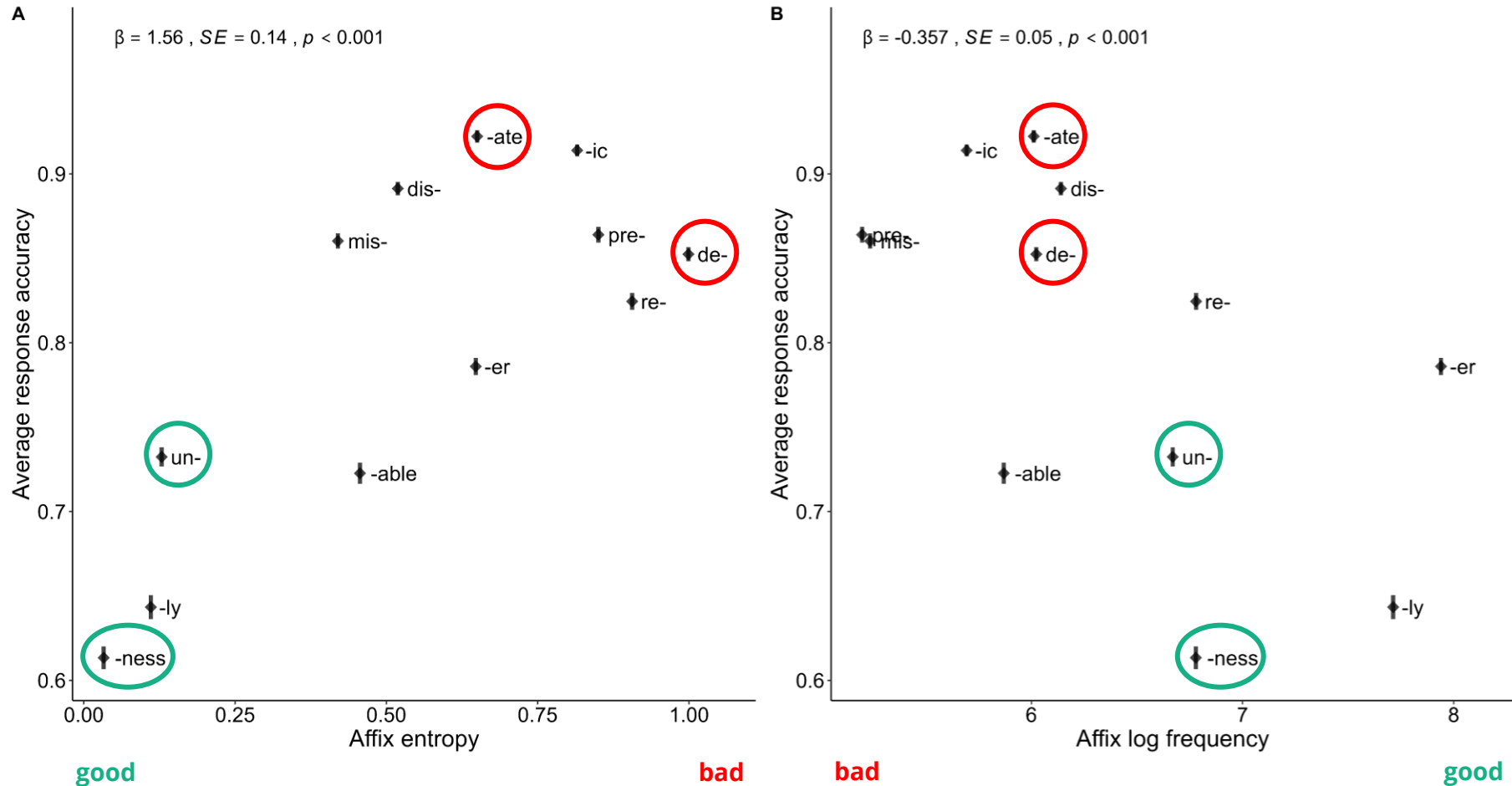


- Prefixed words detectable with RegEx
- Prefixed words not detectable with RegEx
- Words incorrectly parsed as prefixed

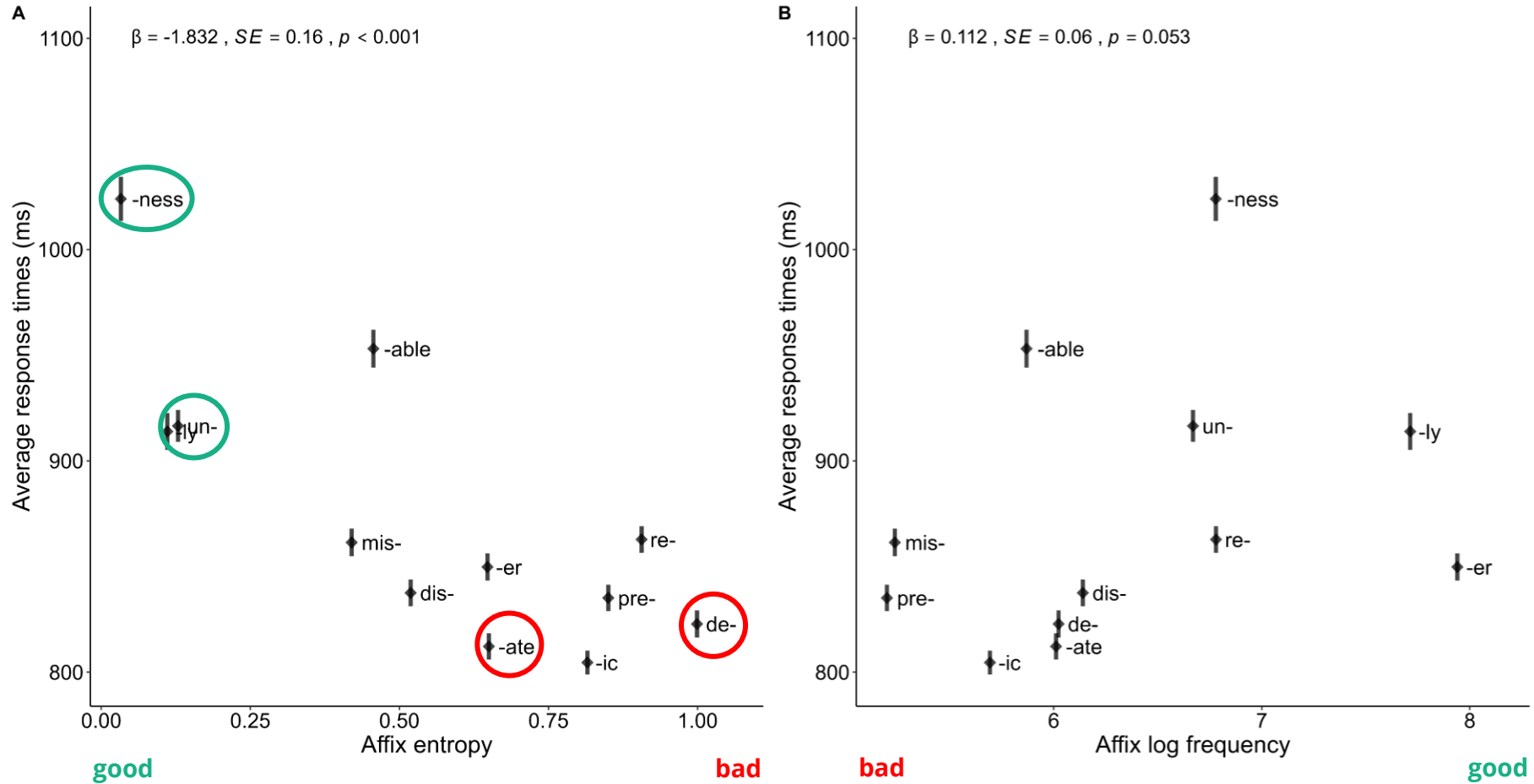


- Suffixed words detectable with RegEx
- Suffixed words not detectable with RegEx
- Words incorrectly parsed as suffixed

Nonwords with “good” affixes are hard to reject...



... and these rejections take time



Conclusions

Quantified morpheme **experience** in print

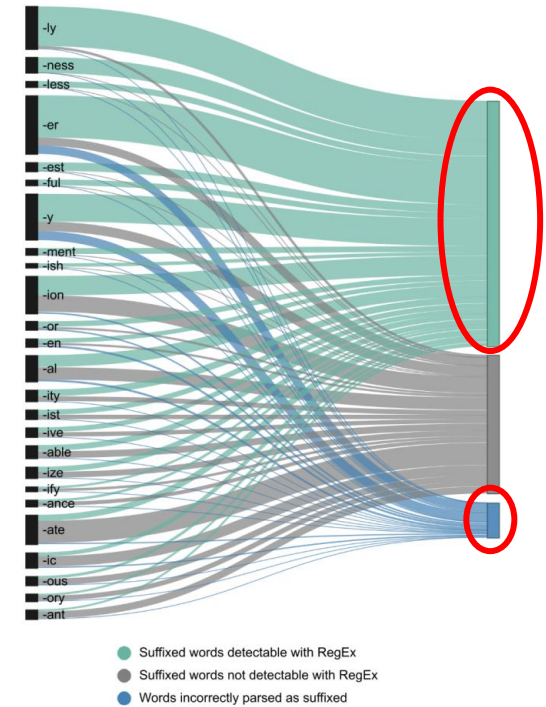


Developed a theory of morpheme learning



Tested this theory against human data

- Theories of learning **must reflect** real-world **experience**
- Learning is driven by what's **detectable**...



Conclusions

Quantified morpheme **experience** in print

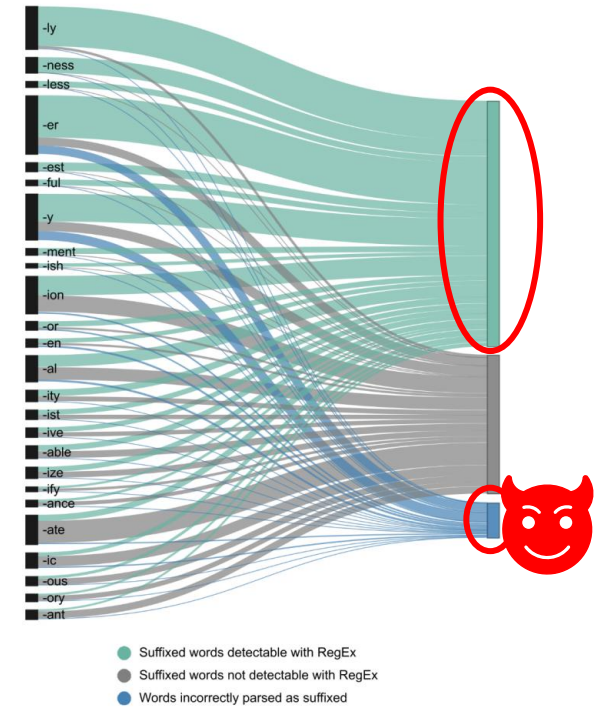


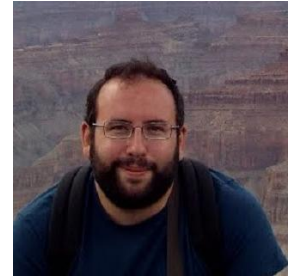
Developed a theory of morpheme learning



Tested this theory against human data

- Theories of learning **must reflect** real-world **experience**
- Learning is driven by what's **detectable...** and **false alarms harm learning**
- Graded experience → graded knowledge





Thank you!

Maria.Korochkina@rhul.ac.uk

Holly.Cooper@rhul.ac.uk