

The nature of morphemic regularities in children's literature: What do British children learn about derivational morphology when they read?

Maria Korochkina¹ | Marco Marelli² | Marc Brysbaert³ | Kathy Rastle¹

¹Department of Psychology, Royal Holloway, University of London, UK

²Department of Psychology, University of Milano-Bicocca, Italy

³Department of Experimental Psychology, Ghent University, Belgium

SSSR 2024 Annual conference



Symposium 2E
*Print exposure and its impact on
reading across languages*

11 July 2024



The power of morphology

The power of morphology

- In languages like English, most words consist of several elements that are themselves meaningful — *morphemes* [1, 2]

The power of morphology

- In languages like English, most words consist of several elements that are themselves meaningful — *morphemes* [1, 2]
- Morpheme knowledge is crucial for quick access to the meanings of *known* printed words...

The power of morphology

- In languages like English, most words consist of several elements that are themselves meaningful — *morphemes* [1, 2]
- Morpheme knowledge is crucial for quick access to the meanings of *known* printed words...
 - ▶ ‘un-’ + [stem] → ‘not [stem]’ (e.g., *unclean*, *unfair*)

The power of morphology

- In languages like English, most words consist of several elements that are themselves meaningful — *morphemes* [1, 2]
- Morpheme knowledge is crucial for quick access to the meanings of *known* printed words...
 - ▶ ‘un-’ + [stem] → ‘not [stem]’ (e.g., *unclean*, *unfair*)
 - ▶ [stem] + ‘-ify’ → ‘make [stem]’ (e.g., *simplify*, *purify*)

The power of morphology

- In languages like English, most words consist of several elements that are themselves meaningful — *morphemes* [1, 2]
- Morpheme knowledge is crucial for quick access to the meanings of *known* printed words...
 - ▶ ‘un-’ + [stem] → ‘not [stem]’ (e.g., *unclean*, *unfair*)
 - ▶ [stem] + ‘-ify’ → ‘make [stem]’ (e.g., *simplify*, *purify*)
- ...and for computing the meanings of *unfamiliar* words [3, 4]

The power of morphology

- In languages like English, most words consist of several elements that are themselves meaningful — *morphemes* [1, 2]
- Morpheme knowledge is crucial for quick access to the meanings of *known* printed words...
 - ▶ ‘un-’ + [stem] → ‘not [stem]’ (e.g., *unclean*, *unfair*)
 - ▶ [stem] + ‘-ify’ → ‘make [stem]’ (e.g., *simplify*, *purify*)
- ...and for computing the meanings of *unfamiliar* words [3, 4]
 - ▶ ‘bright’ + ‘-ify’ → *brightify*

The power of morphology

- In languages like English, most words consist of several elements that are themselves meaningful — *morphemes* [1, 2]
- Morpheme knowledge is crucial for quick access to the meanings of *known* printed words...
 - ▶ ‘un-’ + [stem] → ‘not [stem]’ (e.g., *unclean*, *unfair*)
 - ▶ [stem] + ‘-ify’ → ‘make [stem]’ (e.g., *simplify*, *purify*)
- ...and for computing the meanings of *unfamiliar* words [3, 4]
 - ▶ ‘bright’ + ‘-ify’ → *brightify*
- Yet, readers do not show evidence of morpheme knowledge in online reading tasks until mid-to-late adolescence [5, 6]

Morpheme learning

unknown

unfair

unafraid

unlikely

unconvinced

unknown

unfair

unafraid

unlikely

unconvinced

- Must have consistent meaning transformation [7]

unknown

unfair

unafraid

unlikely

unconvinced

- Must have consistent meaning transformation [7]
- Must occur with a high number of distinct stems [7]

unknown

unfair

unafraid

unlikely

unconvinced

- Must have consistent meaning transformation [7]
- Must occur with a high number of distinct stems [7]
- Must be detectable

unknown	subconscious
unfair	subheading
unafraid	submit
unlikely	subject
unconvinced	

- Must have consistent meaning transformation [7]
- Must occur with a high number of distinct stems [7]
- Must be detectable

unknown	subconscious
unfair	subheading
unafraid	submit
unlikely	subject
unconvinced	sustain

- Must have consistent meaning transformation [7]
- Must occur with a high number of distinct stems [7]
- Must be detectable

unknown	subconscious
unfair	subheading
unafraid	submit
unlikely	subject
unconvinced	sustain ('sub-' + 'tenere')

- Must have consistent meaning transformation [7]
- Must occur with a high number of distinct stems [7]
- Must be detectable

unknown	subconscious
unfair	subheading
unafraid	submit
unlikely	subject
unconvinced	sustain ('sub-' + 'tenere')

- Must have consistent meaning transformation [7]
- Must occur with a high number of distinct stems [7]
- Must be detectable
- Little or poor explicit morphological instruction in school

unknown	subconscious
unfair	subheading
unafraid	submit
unlikely	subject
unconvinced	sustain ('sub-' + 'tenere')

- Must have consistent meaning transformation [7]
 - Must occur with a high number of distinct stems [7]
 - Must be detectable
 - Little or poor explicit morphological instruction in school
- Morpheme knowledge must be acquired via **text experience**

unkn <u>own</u>	sub <u>conscious</u>
un <u>fair</u>	sub <u>heading</u>
un <u>afraid</u>	sub <u>mit</u>
un <u>likely</u>	sub <u>ject</u>
un <u>convinced</u>	sus <u>tain</u> ('sub-' + 'tenere')

- Must have consistent meaning transformation [7]
 - Must occur with a high number of distinct stems [7]
 - Must be detectable
 - Little or poor explicit morphological instruction in school
- Morpheme knowledge must be acquired via **text experience**

But what do children experience **in the wild?**

The CYP-LEX database

1,200 books popular with British children & young people

The CYP-LEX database

1,200 books popular with British children & young people

7–9



10–12



13+



The CYP-LEX database

1,200 books popular with British children & young people

7–9



10–12



13+



- 400 books per age band

The CYP-LEX database

1,200 books popular with British children & young people

7–9



10–12



13+



- 400 books per age band
- 70+ mln tokens; 105,694 types

The CYP-LEX database

1,200 books popular with British children & young people

7–9



10–12



13+



- 400 books per age band
- 70+ mln tokens; 105,694 types
- See Korochkina et al., 2024, *QJEP*, for more detail & link [8]

Many complex words in books

Based on MorphoLex (68,624 unique words) [9]

Many complex words in books

Based on MorphoLex (68,624 unique words) [9]

	7–9	10–12	13+
N (%) complex words	17,634 (45%)	22,564 (48%)	27,554 (51%)
N (%) prefixed words	4,775 (27%)	6,328 (28%)	8,105 (29%)
N (%) suffixed words	11,559 (66%)	14,865 (66%)	18,587 (67%)

Many complex words in books

Based on MorphoLex (68,624 unique words) [9]

	7–9	10–12	13+
N (%) complex words	17,634 (45%)	22,564 (48%)	27,554 (51%)
N (%) prefixed words	4,775 (27%)	6,328 (28%)	8,105 (29%)
N (%) suffixed words	11,559 (66%)	14,865 (66%)	18,587 (67%)

- Roughly half of word types are complex

Many complex words in books

Based on MorphoLex (68,624 unique words) [9]

	7–9	10–12	13+
N (%) complex words	17,634 (45%)	22,564 (48%)	27,554 (51%)
N (%) prefixed words	4,775 (27%)	6,328 (28%)	8,105 (29%)
N (%) suffixed words	11,559 (66%)	14,865 (66%)	18,587 (67%)

- Roughly half of word types are complex
- % complex words increases with book target age

Many complex words in books

Based on MorphoLex (68,624 unique words) [9]

	7–9	10–12	13+
N (%) complex words	17,634 (45%)	22,564 (48%)	27,554 (51%)
N (%) prefixed words	4,775 (27%)	6,328 (28%)	8,105 (29%)
N (%) suffixed words	11,559 (66%)	14,865 (66%)	18,587 (67%)

- Roughly half of word types are complex
- % complex words increases with book target age

Many complex words in books

Based on MorphoLex (68,624 unique words) [9]

	7–9	10–12	13+
N (%) complex words	17,634 (45%)	22,564 (48%)	27,554 (51%)
N (%) prefixed words	4,775 (27%)	6,328 (28%)	8,105 (29%)
N (%) suffixed words	11,559 (66%)	14,865 (66%)	18,587 (67%)

- Roughly half of word types are complex
- % complex words increases with book target age

Many complex words in books

Based on MorphoLex (68,624 unique words) [9]

	7–9	10–12	13+
N (%) complex words	17,634 (45%)	22,564 (48%)	27,554 (51%)
N (%) prefixed words	4,775 (27%)	6,328 (28%)	8,105 (29%)
N (%) suffixed words	11,559 (66%)	14,865 (66%)	18,587 (67%)

- Roughly half of word types are complex
- % complex words increases with book target age
- More exposure to suffixed than to prefixed words

Few complex words used repeatedly

Few complex words used repeatedly

		7–9	10–12	13+
Complex	Frequency ≤ 5	50%	42%	35%
	Frequency ≥ 100	8%	11%	15%
Monomorphemic	Frequency ≤ 5	30%	24%	18%
	Frequency ≥ 100	23%	29%	36%

Few complex words used repeatedly

		7–9	10–12	13+
Complex	Frequency ≤ 5	50%	42%	35%
	Frequency ≥ 100	8%	11%	15%
Monomorphemic	Frequency ≤ 5	30%	24%	18%
	Frequency ≥ 100	23%	29%	36%

Few complex words used repeatedly

		7–9	10–12	13+
Complex	Frequency ≤ 5	50%	42%	35%
	Frequency ≥ 100	8%	11%	15%
Monomorphemic	Frequency ≤ 5	30%	24%	18%
	Frequency ≥ 100	23%	29%	36%

Few complex words used repeatedly

		7–9	10–12	13+
Complex	Frequency ≤ 5	50%	42%	35%
	Frequency ≥ 100	8%	11%	15%
Monomorphemic	Frequency ≤ 5	30%	24%	18%
	Frequency ≥ 100	23%	29%	36%

- Morphologically simple words are much more common than complex words

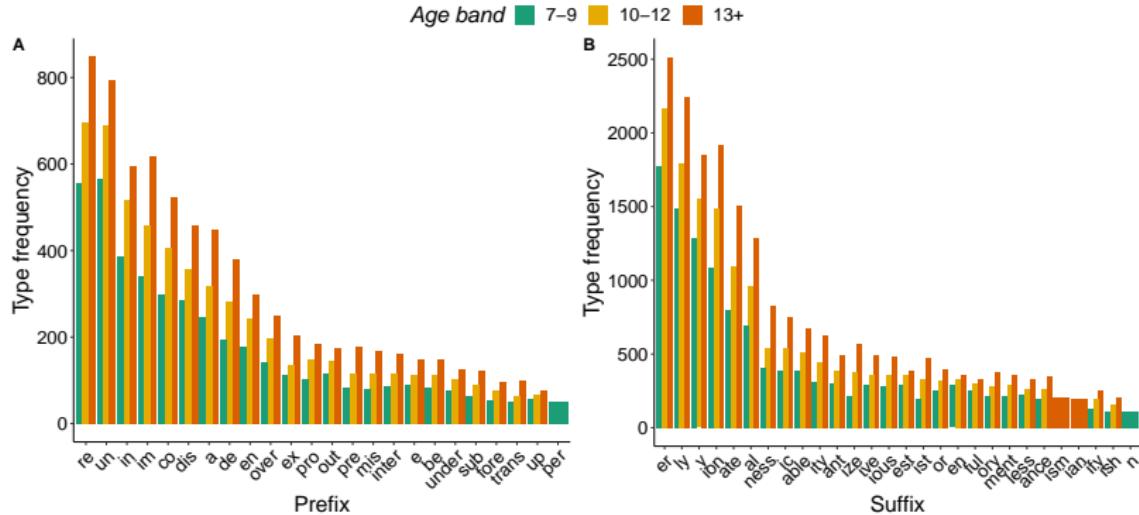
Few complex words used repeatedly

		7–9	10–12	13+
Complex	Frequency ≤ 5	50%	42%	35%
	Frequency ≥ 100	8%	11%	15%
Monomorphemic	Frequency ≤ 5	30%	24%	18%
	Frequency ≥ 100	23%	29%	36%

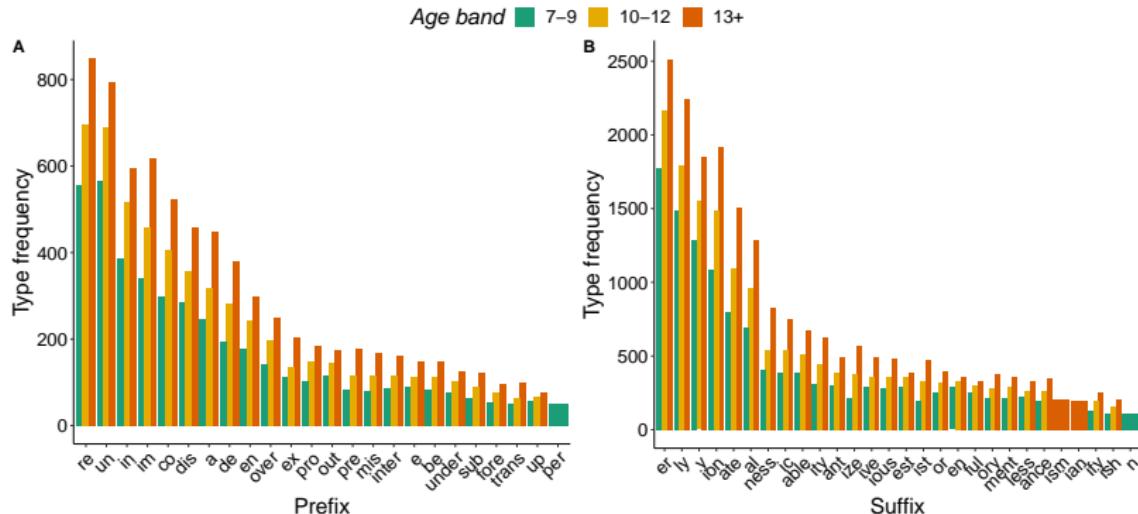
- Morphologically simple words are much more common than complex words
- Readers encounter many morphologically complex words, but few are repeated frequently

Only a few affixes encountered often

Only a few affixes encountered often

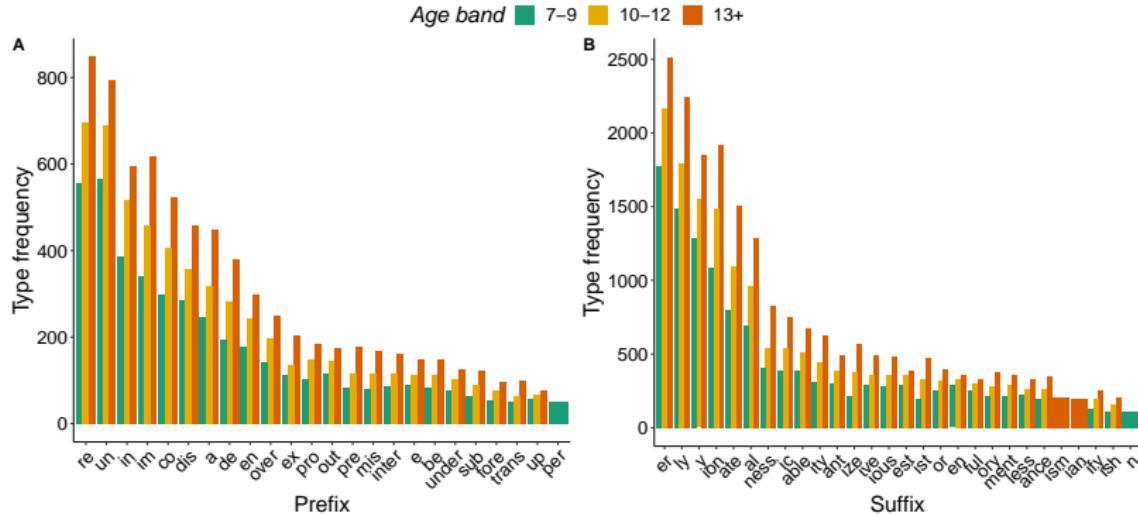


Only a few affixes encountered often



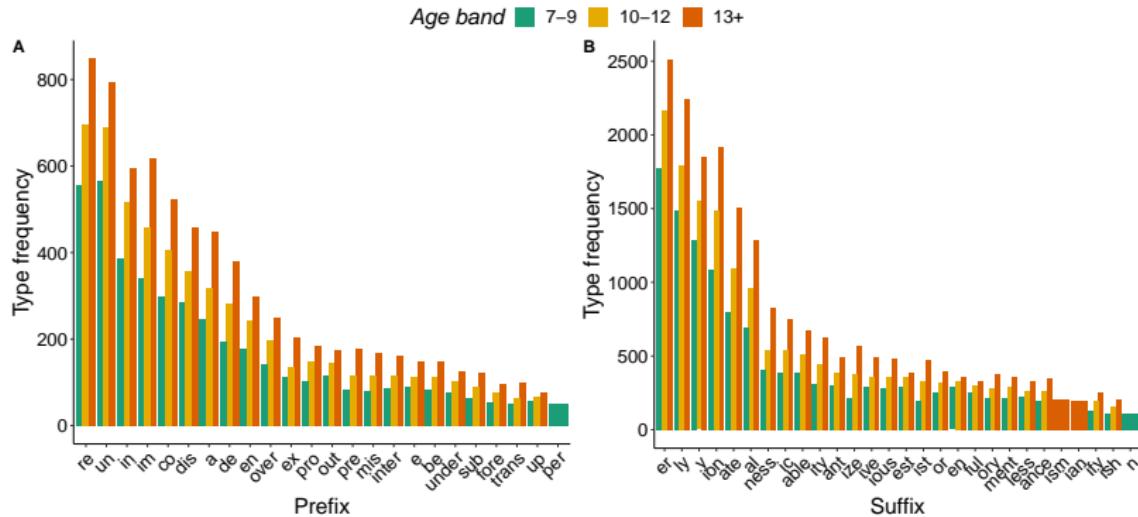
- Prefixes: *re-*, *un-*, *in-*, *co-*

Only a few affixes encountered often



- Prefixes: *re-*, *un-*, *in-*, *co-*
- Suffixes: *-er*, *-ly*, *-y*, *-ion*, *-ate*, *-al*, *-ness*, *-able*, *-ic*

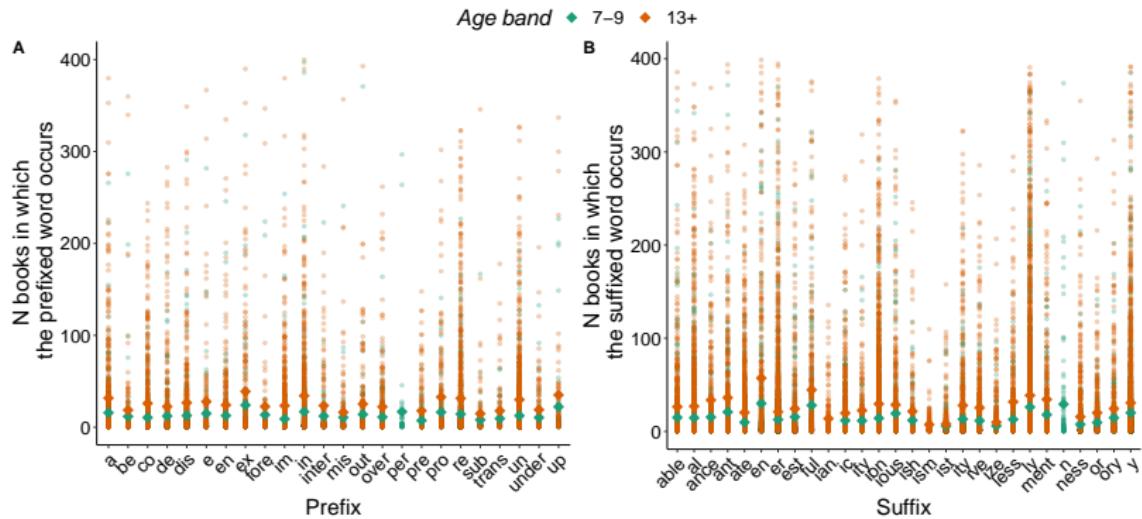
Only a few affixes encountered often



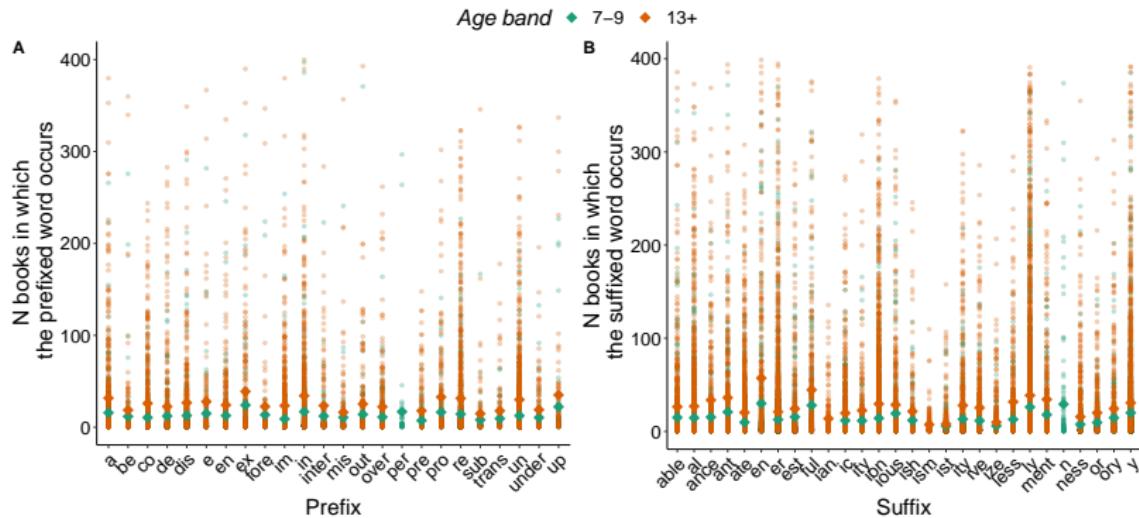
- Prefixes: *re-*, *un-*, *in-*, *co-*
- Suffixes: *-er*, *-ly*, *-y*, *-ion*, *-ate*, *-al*, *-ness*, *-able*, *-ic*
- Limited exposure to multiple types before 13+ texts

Affixes sparsely represented across books

Affixes sparsely represented across books

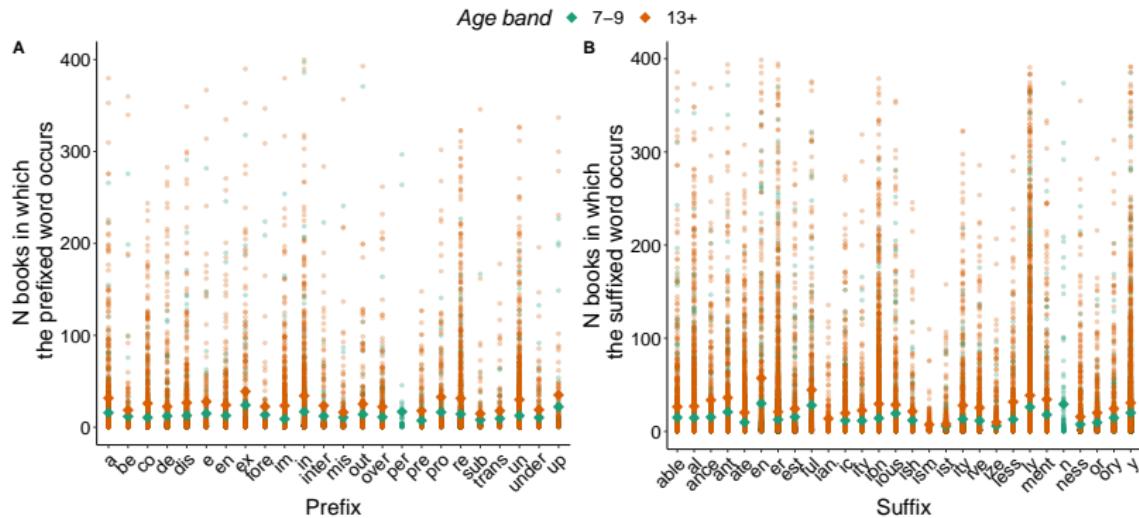


Affixes sparsely represented across books



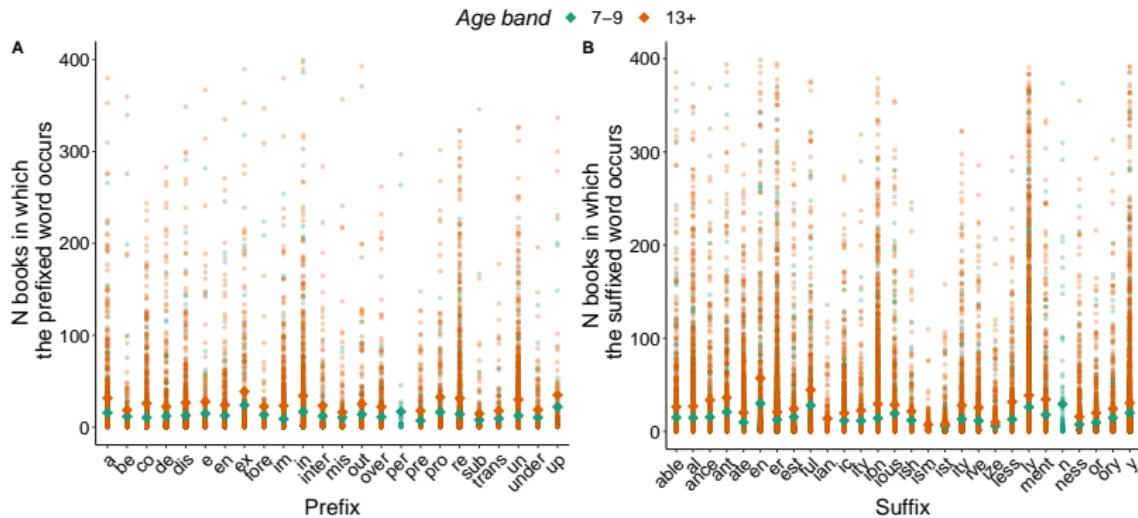
- The average affixed word does not occur in many books

Affixes sparsely represented across books



- The average affixed word does not occur in many books
- Suffixes better represented across books than prefixes

Affixes sparsely represented across books



- The average affixed word does not occur in many books
- Suffixes better represented across books than prefixes
- Few affixes have reasonable representation before 13+ texts

Are morphemes easy to 'find'?

Are morphemes easy to 'find'?

- Complex words are numerous, but infrequent & poorly distributed across books

Are morphemes easy to 'find'?

- Complex words are numerous, but infrequent & poorly distributed across books
- *Likely to see a complex word but unlikely to see it again!*

Are morphemes easy to 'find'?

- Complex words are numerous, but infrequent & poorly distributed across books
 - *Likely to see a complex word but unlikely to see it again!*
- Difficult to learn to recognise complex words by sight

Are morphemes easy to 'find'?

- Complex words are numerous, but infrequent & poorly distributed across books
- *Likely to see a complex word but unlikely to see it again!*
 - Difficult to learn to recognise complex words by sight
 - Being able to break words apart is critical

Are morphemes easy to 'find'?

- Complex words are numerous, but infrequent & poorly distributed across books
- *Likely to see a complex word but unlikely to see it again!*
 - Difficult to learn to recognise complex words by sight
 - Being able to break words apart is critical
- MorphoLex defines morphemes based on etymology

Are morphemes easy to 'find'?

- Complex words are numerous, but infrequent & poorly distributed across books
- *Likely to see a complex word but unlikely to see it again!*
 - Difficult to learn to recognise complex words by sight
 - Being able to break words apart is critical
- MorphoLex defines morphemes based on etymology
 - ▶ *submit*, *subject*, *sustain*,

Are morphemes easy to 'find'?

- Complex words are numerous, but infrequent & poorly distributed across books
- *Likely to see a complex word but unlikely to see it again!*
 - Difficult to learn to recognise complex words by sight
 - Being able to break words apart is critical
- MorphoLex defines morphemes based on etymology
 - *submit, subject, sustain, perfect, include, distribute*

Are morphemes easy to 'find'?

- Complex words are numerous, but infrequent & poorly distributed across books
- *Likely to see a complex word but unlikely to see it again!*
 - Difficult to learn to recognise complex words by sight
 - Being able to break words apart is critical
- MorphoLex defines morphemes based on etymology
 - ▶ *submit, subject, sustain, perfect, include, distribute*
- Yet, children's etymological knowledge is limited

Are morphemes easy to 'find'?

- Complex words are numerous, but infrequent & poorly distributed across books
- *Likely to see a complex word but unlikely to see it again!*
 - Difficult to learn to recognise complex words by sight
 - Being able to break words apart is critical
- MorphoLex defines morphemes based on etymology
 - *submit, subject, sustain, perfect, include, distribute*
- Yet, children's etymological knowledge is limited
 - Exposure to morphology should not be judged on the basis of etymological counts

Are morphemes easy to 'find'?

- Complex words are numerous, but infrequent & poorly distributed across books
- *Likely to see a complex word but unlikely to see it again!*
 - Difficult to learn to recognise complex words by sight
 - Being able to break words apart is critical
- MorphoLex defines morphemes based on etymology
 - *submit, subject, sustain, perfect, include, distribute*
- Yet, children's etymological knowledge is limited
 - Exposure to morphology should not be judged on the basis of etymological counts



Are morphemes easy to detect if defined **orthographically**?

The Regular Expression approach

The Regular Expression approach

A string-searching algorithm to detect words that *appear* complex

The Regular Expression approach

A string-searching algorithm to detect words that *appear* complex

- Recursive search of legal stem–affix combinations

The Regular Expression approach

A string-searching algorithm to detect words that *appear* complex

- Recursive search of legal stem–affix combinations
- Sensitive to common orthographic alterations

The Regular Expression approach

A string-searching algorithm to detect words that *appear* complex

- Recursive search of legal stem–affix combinations
- Sensitive to common orthographic alterations

	<i>N</i> 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%)

The Regular Expression approach

A string-searching algorithm to detect words that *appear* complex

- Recursive search of legal stem–affix combinations
- Sensitive to common orthographic alterations

	N 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 low due to bound stems (*subject*) & complex orthographic alterations (*sustain*)

The Regular Expression approach

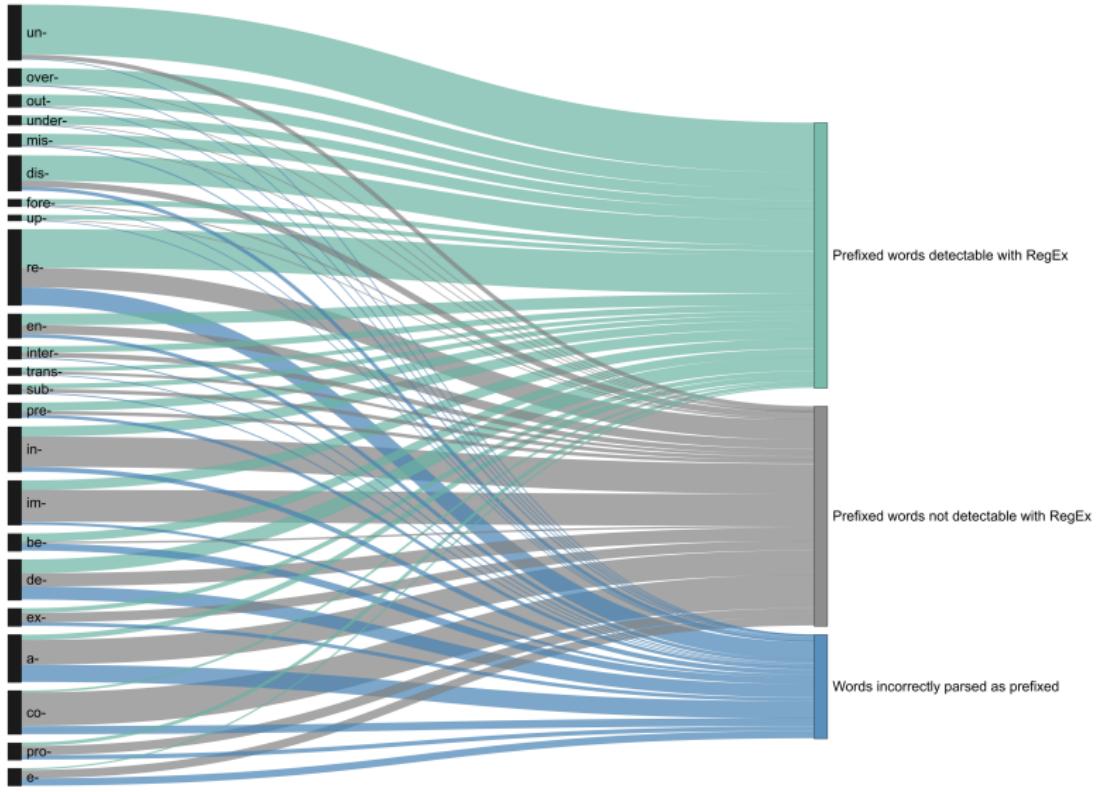
A string-searching algorithm to detect words that *appear* complex

- Recursive search of legal stem–affix combinations
- Sensitive to common orthographic alterations

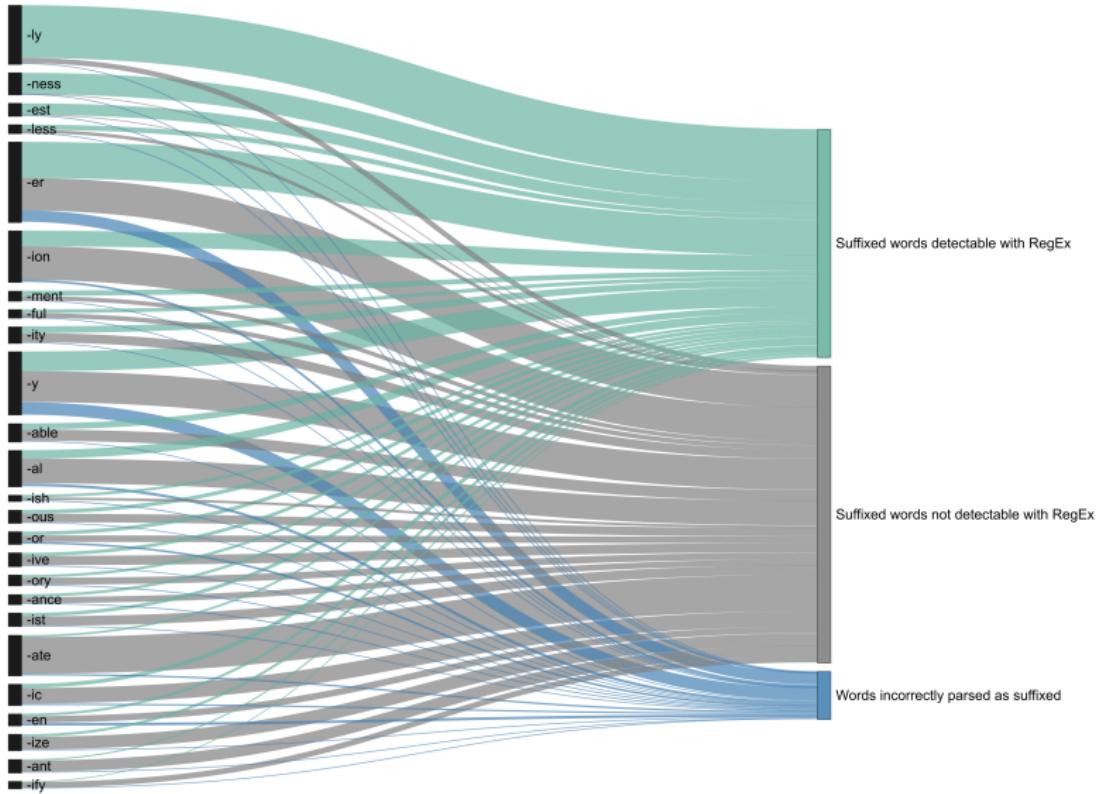
	N MorphoLex complex	RegEx hits	RegEx FAs
Prefixed	8,105	3,811 (47%)	1,510 (3.3%)
Suffixed	18,587	8,801 (47%)	1,735 (4.8%)

- Hits low due to bound stems (*subject*) & complex orthographic alterations (*sustain*)
- False alarms due to pseudoaffixation (*corner*)

Some prefixes are much easier to detect than others



Some suffixes are much easier to detect than others



Conclusions

- Complex words comprise a large proportion of words in children's books

- Complex words comprise a large proportion of words in children's books
- Substantial variation across affixes

- Complex words comprise a large proportion of words in children's books
- Substantial variation across affixes
 - ▶ Frequency, representation across books, detectability

- Complex words comprise a large proportion of words in children's books
- Substantial variation across affixes
 - ▶ Frequency, representation across books, detectability
- *Graded* morpheme information → *graded* morpheme knowledge

- Complex words comprise a large proportion of words in children's books
- Substantial variation across affixes
 - ▶ Frequency, representation across books, detectability
- *Graded* morpheme information → *graded* morpheme knowledge



Beyond a handful of affixes, **morpheme knowledge** will be **difficult to acquire from text**

- Complex words comprise a large proportion of words in children's books
- Substantial variation across affixes
 - ▶ Frequency, representation across books, detectability
- *Graded* morpheme information → *graded* morpheme knowledge



Beyond a handful of affixes, **morpheme knowledge** will be **difficult to acquire from text**

→ Is this why children do not show evidence of morpheme knowledge in online reading tasks until late adolescence?...

- Complex words comprise a large proportion of words in children's books
- Substantial variation across affixes
 - ▶ Frequency, representation across books, detectability
- *Graded* morpheme information → *graded* morpheme knowledge



Beyond a handful of affixes, **morpheme knowledge** will be **difficult to acquire from text**

→ Is this why children do not show evidence of morpheme knowledge in online reading tasks until late adolescence?...

→ A case for systematic morphology instruction?...

Thank you!

Most 'new' words in books are complex

Most 'new' words in books are complex

- SUBTLEX-UK: 9 channels over 3 years; 159,235 types [10]

Most 'new' words in books are complex

- SUBTLEX-UK: 9 channels over 3 years; 159,235 types [10]
- CBBC: aimed at children aged 6–12; 58,691 types

Most 'new' words in books are complex

- SUBTLEX-UK: 9 channels over 3 years; 159,235 types [10]
- CBBC: aimed at children aged 6–12; 58,691 types

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

Most 'new' words in books are complex

- SUBTLEX-UK: 9 channels over 3 years; 159,235 types [10]
- CBBC: aimed at children aged 6–12; 58,691 types

	7–9	10–12	13+
N words in MorphoLex	39,149	47,363	54,557
N words missing from CBBC	8,260	14,050	20,105
N (%) complex words	4,924 (59%)	8,562 (61%)	12,894 (64%)
N words missing from SUBTLEX-UK	1,211	2,450	4,602
N (%) complex words	888 (73%)	1,796 (73%)	3,514 (76%)

Most 'new' words in books are complex

- SUBTLEX-UK: 9 channels over 3 years; 159,235 types [10]
- CBBC: aimed at children aged 6–12; 58,691 types

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

- Most words not encountered on TV are complex

Most 'new' words in books are complex

- SUBTLEX-UK: 9 channels over 3 years; 159,235 types [10]
- CBBC: aimed at children aged 6–12; 58,691 types

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

- Most words not encountered on TV are complex
- Books are an important source of morpheme information

Teacher & publisher knowledge often patchy

Teacher & publisher knowledge often patchy

www.Teacher-of-English.com

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

YEAR 3 AND 4 – WRITING – TRANSCRIPTION

injection

discussion

tension

magician

English Teaching Resources

- [1] P. Angelelli, C. Marinelli, and C. Burani, “The effect of morphology on spelling and reading accuracy: A study on Italian children,” *Frontiers in Psychology*, vol. 5, no. Article 1373, pp. 1–10, 2014. DOI: <https://doi.org/10.3389/fpsyg.2014.01373>.
- [2] R. Baayen, R. Piepenbrock, and H. van Rijn, *The CELEX lexical database, (CD-ROM)*. PA: University of Pennsylvania. Linguistic Data Consortium, 1993.
- [3] K. Rastle and M. Davis, “Morphological decomposition based on the analysis of orthography,” *Language and Cognitive Processes*, vol. 23, no. 7/8, pp. 942–971, 2008. DOI: <https://doi.org/10.1080/01690960802069730>.

- [4] K. Rastle, “The place of morphology in learning to read in English,” *Cortex*, vol. 116, pp. 45–54, 2019. DOI: <https://doi.org/10.1016/j.cortex.2018.02.008>.
- [5] N. Dawson, K. Rastle, and J. Ricketts, “Morphological effects in visual word recognition: Children, adolescents, and adults,” *Journal of Experimental Psychology: Learning Memory and Cognition*, vol. 44, no. 4, pp. 645–654, 2018. DOI: <https://doi.org/10.1037/xlm0000485>.
- [6] N. Dawson, K. Rastle, and J. Ricketts, “Finding the man amongst many: A developmental perspective on mechanisms of morphological decomposition,” *Cognition*, vol. 211, p. 104605, 2021. DOI: <https://doi.org/10.1016/j.cognition.2021.104605>.

- [7] J. Tamminen, M. Davis, and K. Rastle, "From specific examples to general knowledge in language learning," *Cognitive Psychology*, vol. 79, pp. 1–39, 2015. DOI: <https://doi.org/10.1016/j.cogpsych.2015.03.003>.
- [8] M. Korochkina, M. Marelli, M. Brybaert, and K. Rastle, "The Children and Young People's Books Lexicon (CYP-LEX): A large-scale lexical database of books read by children and young people in the United Kingdom," *Quarterly Journal of Experimental Psychology*, 2024. DOI: <https://doi.org/10.1177/17470218241229694>.
- [9] C. H. Sánchez-Gutiérrez, H. Mailhot, S. H. Deacon, and M. A. Wilson, "MorphoLex: A derivational morphological database for 70,000 English words," *Behavior Research Methods*, vol. 50, no. 4, pp. 1568–1580, 2018. DOI: [10.3758/s13428-017-0981-8](https://doi.org/10.3758/s13428-017-0981-8).

- [10] W. J. van Heuven, P. Mandera, E. Keuleers, and M. Brysbaert, “SUBTLEX-UK: A new and improved word frequency database for British English,” *Quarterly Journal of Experimental Psychology*, vol. 67, no. 6, pp. 1176–1190, 2014. DOI: [10.1080/17470218.2013.850521](https://doi.org/10.1080/17470218.2013.850521).