

What do we know about the words in teenagers' books?

Professor Kathy Rastle

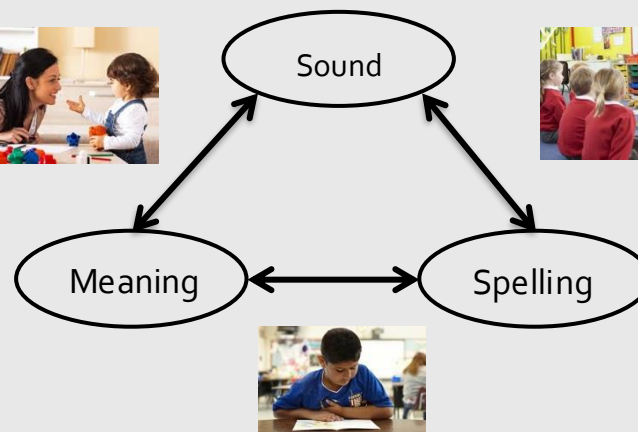


The journey to skilled reading



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

Oral language
foundations



Phonics
instruction



Text experience



What challenges does text experience pose and what opportunities does it bring?

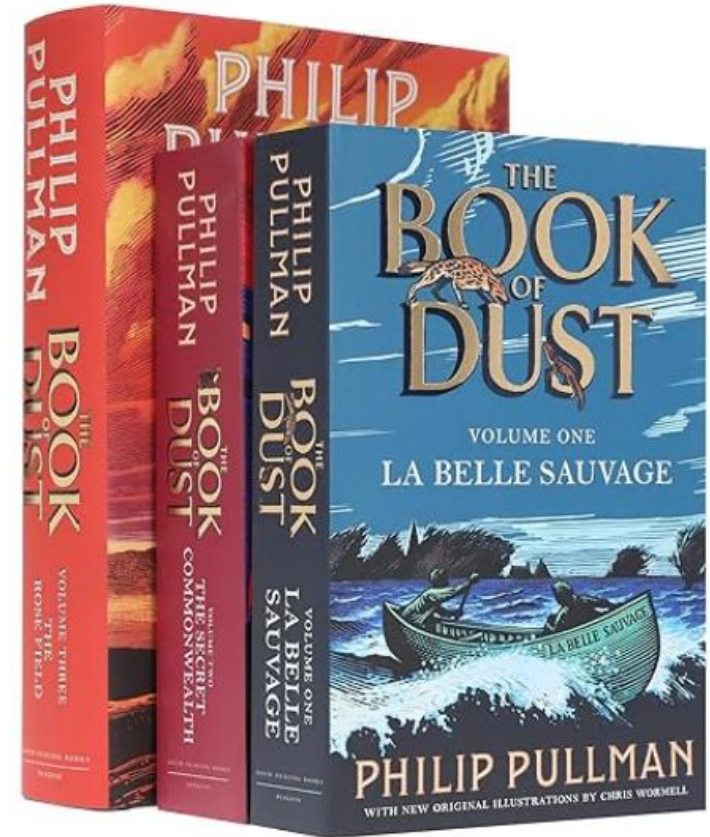
The richness of text



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

But he enjoyed the life of the inn. He especially enjoyed the conversations he overheard, whether they concerned the venal rascality of the River Board, the helpless idiocy of the government, or more philosophical matters such as whether or not the stars were the same age as the Earth.

Sometimes Malcolm became so interested in the latter sort of conversation that he'd rest his armful of empty glasses on the table and join in, but only after having listened intently. He was known to many of the scholars and other visitors, and generously tipped, but becoming rich was never an aim of his; he took tips to be the generosity of providence, and came to think of himself as lucky, which did him no harm in later life. If he'd been the sort of boy who acquired a nickname he would no doubt have been known as 'Professor', but he wasn't that sort of boy. He was liked when noticed, but not noticed much, and that did him no harm either.



Suitable for 12+ years old

Rastle, 2026

More sophisticated than spoken language



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

But he enjoyed the life of the inn. He especially enjoyed the conversations he overheard, whether they concerned the venal rascality of the River Board, the helpless idiocy of the government, or more philosophical matters such as whether or not the stars were the same age as the Earth.

Sometimes Malcolm became so interested in the latter sort of conversation that he'd rest his armful of empty glasses on the table and join in, but only after having listened intently. He was known to many of the scholars and other visitors, and generously tipped, but becoming rich was never an aim of his; he took tips to be the generosity of providence, and came to think of himself as lucky, which did him no harm in later life. If he'd been the sort of boy who acquired a nickname he would no doubt have been known as 'Professor', but he wasn't that sort of boy. He was liked when noticed, but not noticed much, and that did him no harm either.

Bloody school tomorrow!

Going back tonight ... or tomorrow morning ?

Tomorrow morning .

What early?

I'm gonna have to get up at four. Get home by seven .

Have you hurt your eyes ?

Yeah. Rubbing it ... Got eyelash in my eye ... I get a funny twitch in it . The nerve goes , phew {laugh} It's horrible when it {laughing} A bit dodgy !

Yeah. Was the club busy last night ?

Er , yeah it was actually packed , it was really packed. It was a good night actually.

Bergen Corpus of London
Teenage Language (1993)

Rastle, 2026

CYP-LEX corpus



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

National reading surveys, publisher data, book sales statistics from Amazon, BookTrust, Goodreads, LoveReading4Kids, etc.



1,200 popular books, 400 books per age band

7-9 years



10-12 years



13+ years



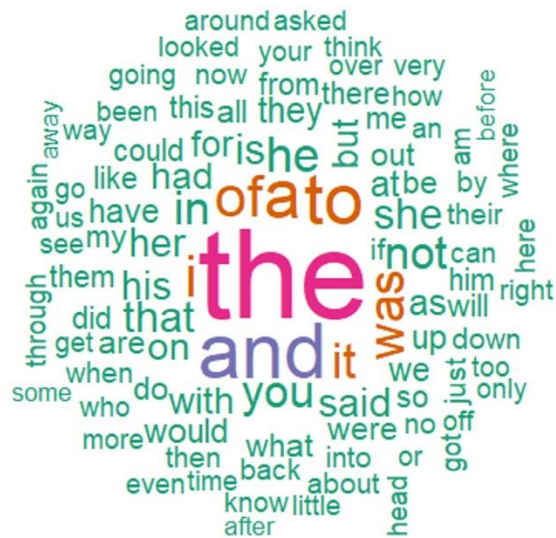
Over 70 million words in the corpus
Over **100,000 distinct words**

100 words make up 54% of the books...

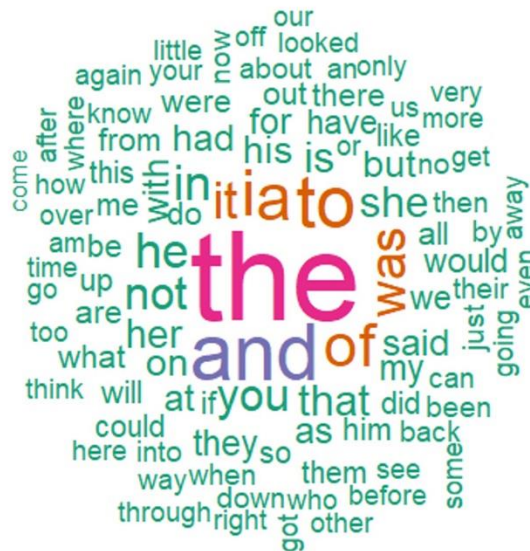


**ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON**

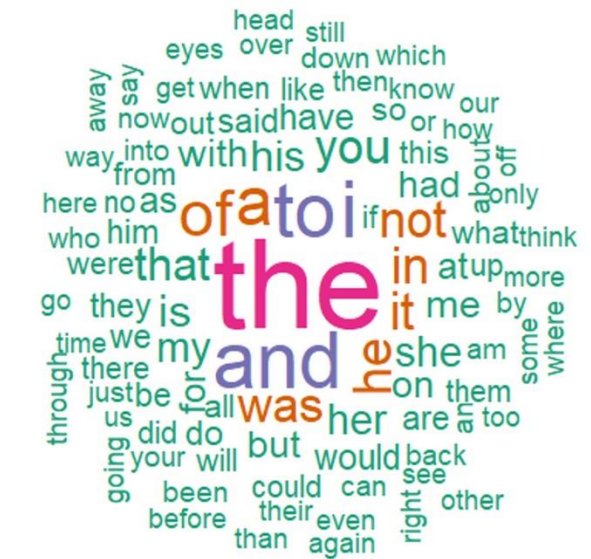
A 7-9 age band



B 10-12 age band



C 13+ age band



Yet, these words aren't very useful for understanding

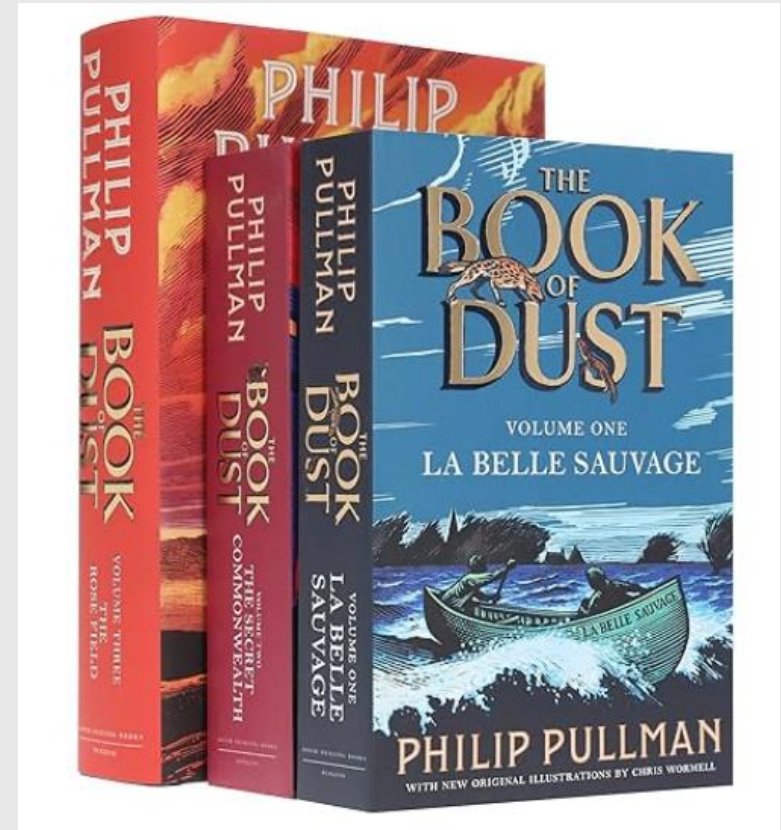


ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

But he ... the ... of the He the ...
he ..., whether they ... the of the ...
..., the of the ..., or more such
as ... whether or not the ... were the
as the

... .. so ... in the of ... that he'd ...
his ... of on the ... and ... in, but only
after He was ... to many of the ...
and, and, but was ... an ...
of his; he... .. to be the ... of ... and came
to ... of himself as ... which did him no ...
in If he's been the ... of ... who ... a
... he would no ... have been ... as ..., but
he wasn't that ... of He was ... when
... but not, and that did him no ...

....



- Readers will quickly learn to recognise common words by sight
- But recognising every second word is not sufficient!

Most words are not used repeatedly



	% words used more than 100 times (across 400 books)	% words used less than 50 times (across 400 books)
7-9 age band	12%	81%
10-12 age band	14%	79%
13+ age band	16%	77%

- Most words are **rare**
- **Not enough exposure** to learn them by sight



It is crucial that readers acquire **strong analytical skills** to unpick them

Many words are unfamiliar



Percentage of CYP-LEX words **not** encountered on TV

	CBeebies (<i>up to 6 yrs</i>) + CBBC (<i>6-12 yrs</i>)	9 BBC channels
7-9 age band	28%	
10-12 age band	40%	
13+ age band		21%

- Readers encounter many words in books that are **not in their spoken vocabulary**



Readers **will be needing considerable support** to understand the words they encounter in books

Books vary greatly in the words they use



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

- In each age band, 30% of distinct words appear in **1 book out of 400**
- Most of these words are used **once** in that book
- **Low similarity** in vocabulary **across the individual books**



Each book contains many words that are not encountered in any other book



- **Each book is a new challenge**
- To encounter as many different words as possible as often as possible, teenagers **must read widely**

How do GCSE English Lit books compare?



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

32 prose books from the AQA and EdExcel specifications for English Literature GCSE



How do these books compare to the popular books?

Highly dense vocabulary in the GCSE books



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON



- GCSE books are **half as long as** popular books, but contain a **similar number of distinct words**
- GCSE books are much **less homogenous** in the words they use



- **More** vocabulary through **less** text in the GCSE books
- May be **harder to understand the text as a whole**

Many unfamiliar words in the GCSE books

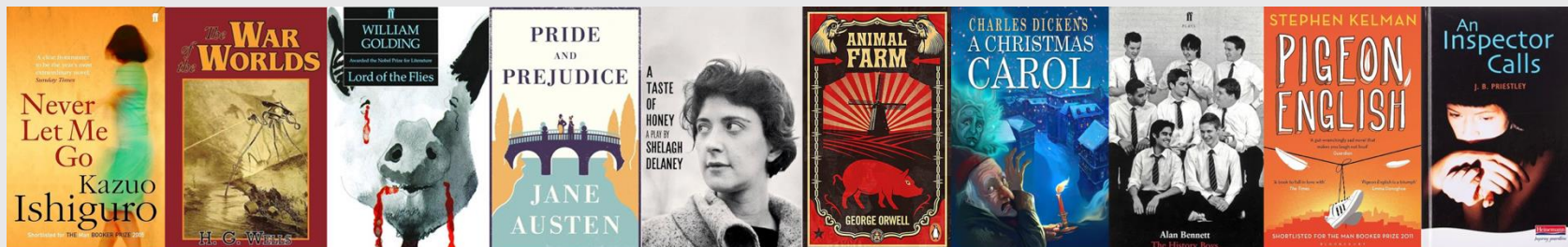


ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

- Only **33%** of distinct words in GCSE books **occur regularly** in 13+ samples
- **27%** of distinct words in GCSE books **don't occur at all** in 13+ samples
- **3,000** distinct words **never used on 9 BBC TV channels over 3 years**
 - *poulterer, bonneted, dowerless, bedight, sepulchre, catechize*
 - *brusquely, docilely, imploringly, beatifically, superciliously*



- GCSE texts **will stretch** even those who read widely
- Weaker readers **may not be able to engage** with the GCSE texts at all



Many unfamiliar words in GCSE books are new roots



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

- Children **often** encounter unfamiliar words in books
- In popular books, these “new” words are usually **morphologically complex**
→ Meanings **can be derived** from the words’ constituents: [mourn] + [-ful] + [-ly]
- In GCSE books, many additional unfamiliar words but most are **new roots**
→ Meanings **cannot be derived** from smaller units
 - *aspidistra, crimplene, beseech, coccidia, gambol*



- Pupils **must rely on context or instruction** to understand these words
- Deriving meaning from context requires **advanced language and reading skills**

Conclusions



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

- 💡 Books offer a wonderful **opportunity** to build vocabulary
- 💡 Yet, book vocabulary is very **challenging** – unfamiliar words, rare words, sparsely represented across different books
- 💡 The challenge ramps up further in GCSE English Literature texts – density, unfamiliar vocabulary, prevalence of **new roots**
- 💡 Young people need **strong foundational reading skills** to access popular books and to **read widely** to build vocabulary
- 💡 Young people with good foundational reading skills will be able to read, understand what they are reading, and derive pleasure from books, **leading to a virtuous circle**

Further reading



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

Quarterly Journal of Experimental Psychology
OnlineFirst

© Experimental Psychology Society 2024, Article Reuse Guidelines
<https://doi.org/10.1177/17470218241229694>

Sage Journals

Original Article

Open Access



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

Maria Korochkina ¹, Marco Marelli ², Marc Brysbaert ³, and Kathleen Rastle ¹



Maria Korochkina & Kathy Rastle · Mar 12 · 5 min read

rastlelab.com/blog

What Words do Children Encounter When They Read for Pleasure?

The English Association Journal
for Teachers of English

THE USE OF english

Edited by Kayte Haselgrove and Chris Green

12 Maria Korochkina and
Kathleen Rastle

The vocabulary barrier in the
General Certificate of Secondary
Education (GCSE) in English
Literature

tes
magazine

Teaching & Learning ▾ Scotland Leadership ▾ Newsletters Jobs and more ▾

What makes GCSE English lit so hard for students?

New research has analysed GCSE literature texts against popular fiction – and found three key reasons why so many young people struggle with the qualification

4th August 2024, 8:00am

Maria Korochkina and Kathleen Rastle



Rastle, 2026

Thank you!



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON