

Dr Maria Korochkina

The words children meet in books: Evidence from a 70-mln word corpus

Come & Learn Breakfast session @ Renaissance UK
22 April 2026

By the end of this session, you will...

- Understand the **general properties of vocabulary** offered by books **popular** with children and young people
- Learn about the importance of **knowledge about the internal structure of words** for skilled reading and how children acquire this knowledge through reading (or not!)
- Recognise **how GCSE English Lit texts differ** from popular books and the **additional barriers** they may present
- Reflect on the **classroom implications** of the properties of texts that children choose, or are required to, read

CYP-LEX

The Children and Young People's Books Lexicon

1,200 *popular* books, 400 books per age band

7-9 years



10-12 years



13-16 years



The **lightbulb** moments

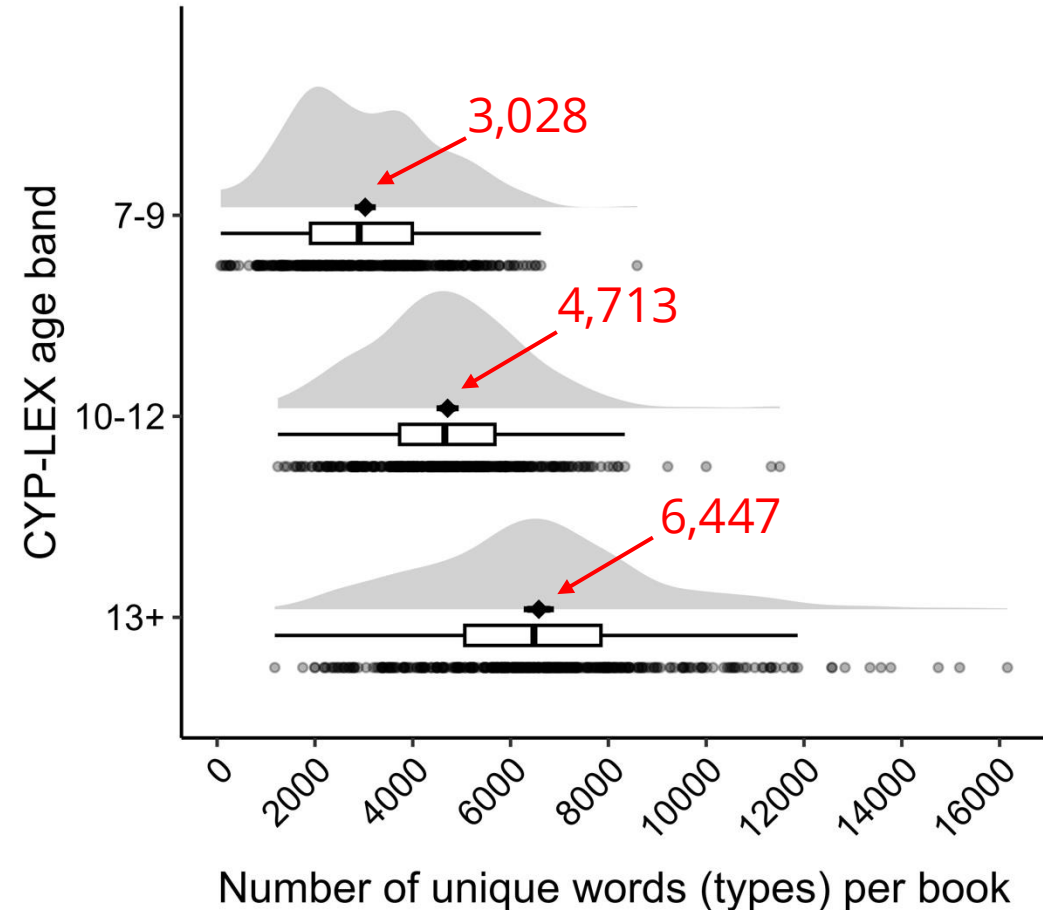


Books contain **vast** numbers of words

- Over **70 million words** and **over 100,000 distinct words** in 1,200 books
- **50,000+** distinct words in the 7-9 age band alone

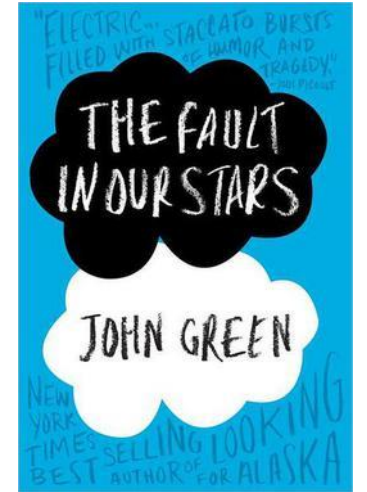


- **Memorising words by rote is not an option**
- **Phonics is key**



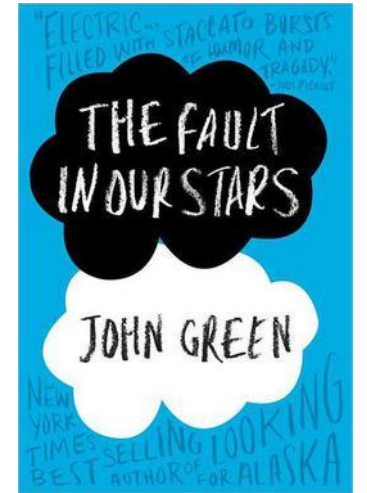
Few words are used repeatedly

“Late in the winter of my seventeenth year, my mother decided I was depressed, presumably because I rarely left the house, spent quite a lot of time in bed, read the same book over and over, ate infrequently, and devoted quite a bit of my abundant free time to thinking about death”.



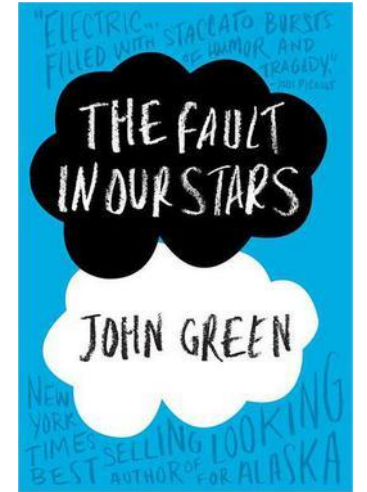
Few words are used repeatedly

" in the of my ,
my I was ,
because I the ,
a lot of in , the
same over and over, ,
and a bit of my
to about ".



Few words are used repeatedly

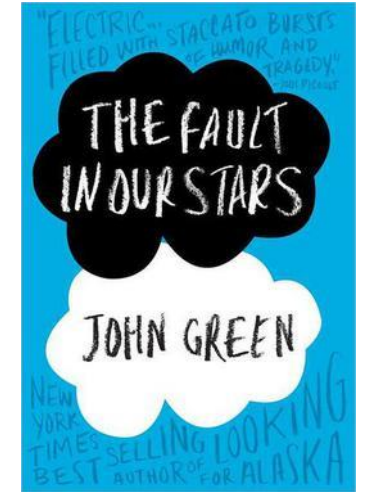
" in the of my ,
my I was ,
because I the ,
a lot of in , the
same over and over, ,
and a bit of my
to about ".



- Recognising every second word **does not take you very far**

Few words are used repeatedly

“Late in the winter of my seventeenth year, my mother decided I was depressed, presumably because I rarely left the house, spent quite a lot of time in bed, read the same book over and over, ate infrequently, and devoted quite a bit of my abundant free time to thinking about death”.



- Recognising every second word **does not take you very far**
- **Gaining repeated exposure** needed for learning **is not trivial**

Many **unfamiliar** words in books

- **40%** of distinct words don't appear on **children's TV**



- **In 13+ books, 21%** of distinct words do not appear even in TV programmes for adults



Books offer a **unique opportunity** to build vocabulary...

... but they also present a **huge challenge** from the earliest years of independent reading!

Books vary greatly in the words they use

- 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**
- Books for younger children are **less similar to one another** than those for older kids



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



- **Each book is a challenge**
- To maximise learning, it's important to **read widely**

A challenge, or a blessing?

 Access to **language beyond everyday use**

 Wonderful **opportunity to build vocabulary**

...However, the other side of the coin is that...

 **Strong phonics skills** are required to access even popular books

 Building vocabulary & reading proficiency requires **reading widely**

Half of distinct words have **complex structure**



Sound + less + ly

Half of distinct words have **complex structure**



- Sound + less + ly
- Memory + ise
- Travel + er
- Numb + ly
- Guilt + y + ly
- Sorrow + ful + ly

Understanding **morphemes**

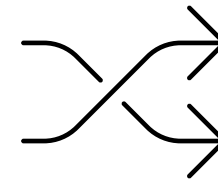
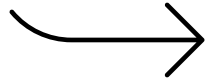
- Most English words are built by **recombining morphemes**

soundlessly, tirelessly, endlessly
tasty, tasteless, tasteful

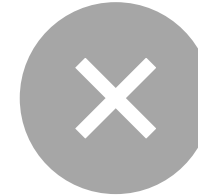
- Morpheme knowledge enables **rapid access to the meanings of familiar** words
- It is also **crucial for computing the meanings of unfamiliar** words

brightify = bright + -ify

Assuming the perspective of a reader



useful

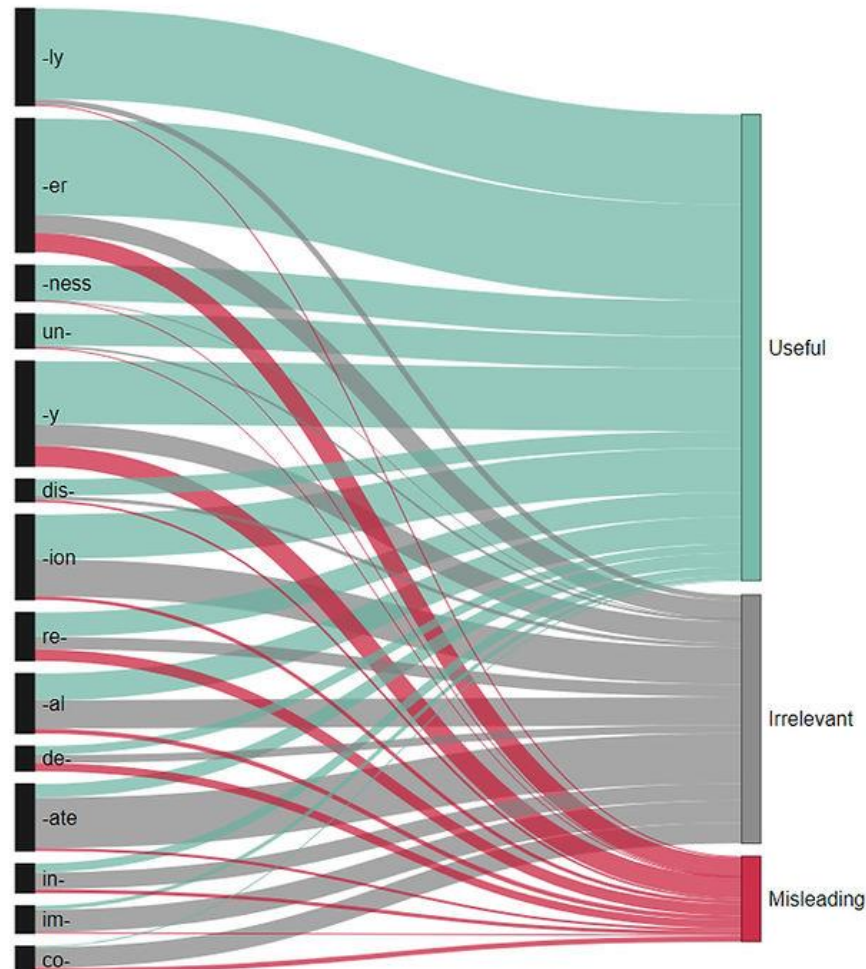


irrelevant



misleading

Words **do not contribute equally** to learning



Kindly, openly, nicely, teacher, banker, driver, darkness, sadness, unlock, unafraid, difficulty, sunny, disobey, disagree, reassure, decode, relation, celebration



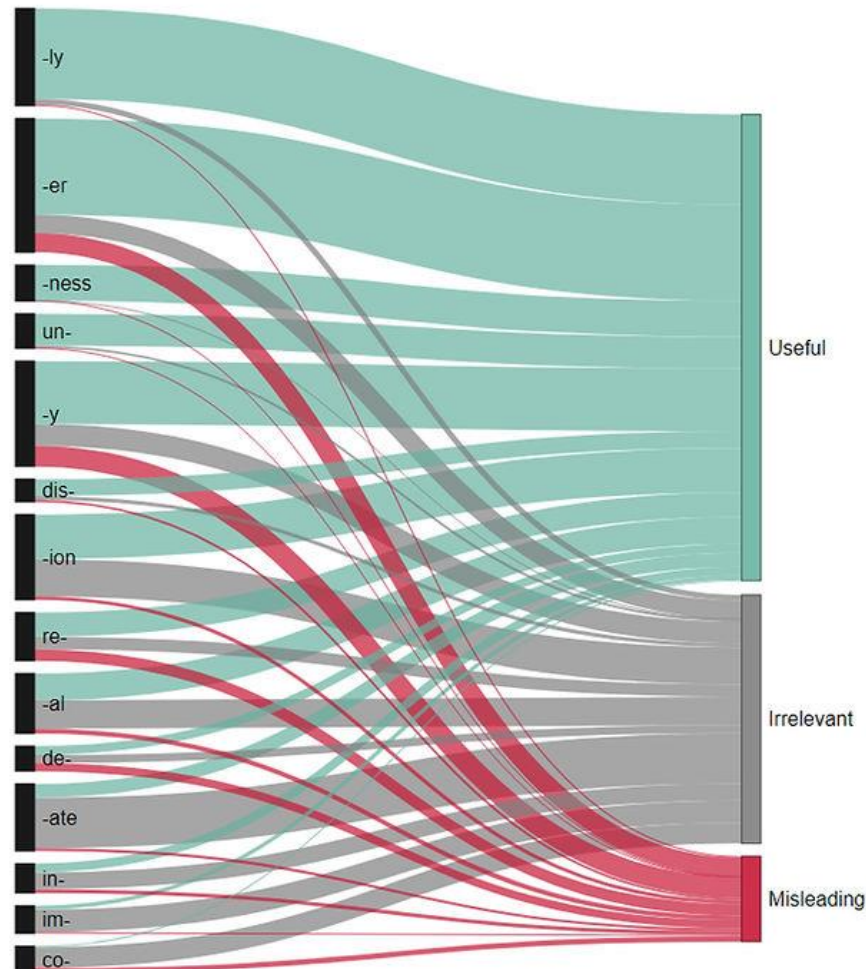
Recognise, resume, reflect, demand, deduce, involve, insist, sorcerer, inquisitive, generate, appreciate, compensate, corrupt, contain, journal, abysmal



Brother, army, brandy, forty, deliver, detail, record, retired, comedian



Some words are **useful** for learning morphemes



Kindly, openly, nicely, teacher, banker, driver, darkness, sadness, unlock, unafraid, difficulty, sunny, disobey, disagree, reassure, decode, relation, celebration



kind + **ly**
dark + **ness**

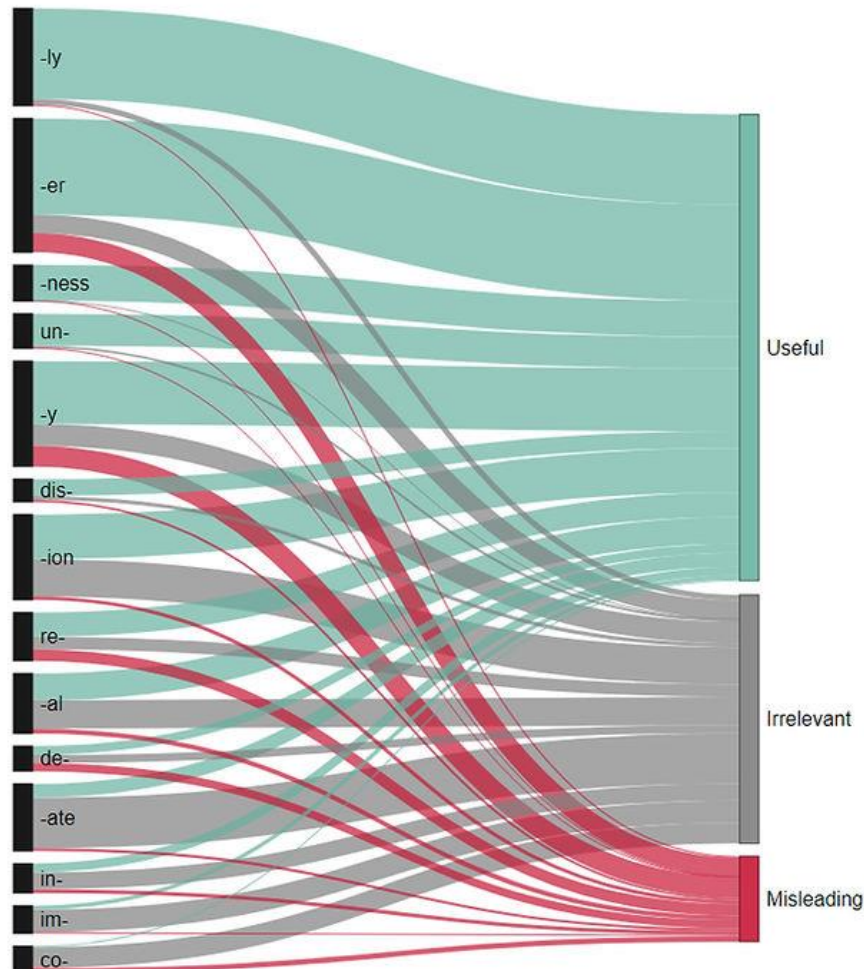
Recognise, resume, reflect, demand, deduce, involve, insist, sorcerer, inquisitive, generate, appreciate, compensate, corrupt, contain, journal, abysmal



Brother, army, brandy, forty, deliver, detail, record, retired, comedian



In many words, morphemes **can't be identified**



Kindly, openly, nicely, teacher, banker, driver, darkness, sadness, unlock, unafraid, difficulty, sunny, disobey, disagree, reassure, decode, relation, celebration



Recognise, resume, reflect, demand, deduce, involve, insist, sorcerer, inquisitive, generate, appreciate, compensate, corrupt, contain, journal, abysmal

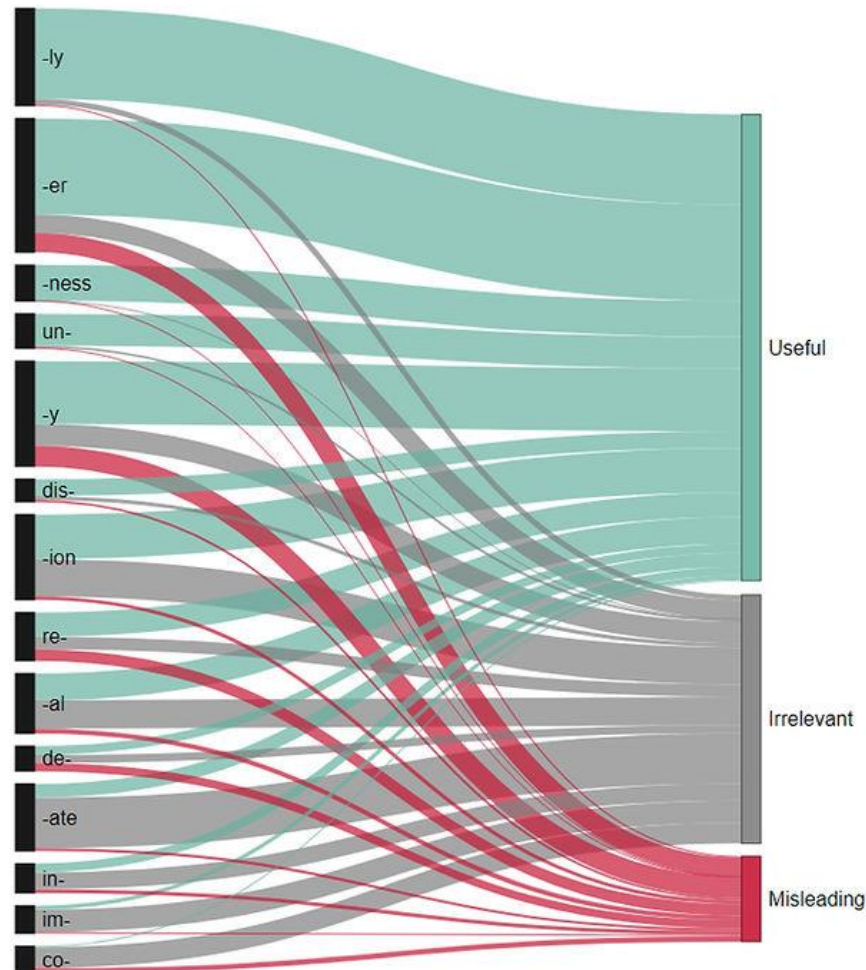


Brother, army, brandy, forty, deliver, detail, record, retired, comedian



resume
demand

Some words only **appear** complex



Kindly, openly, nicely, teacher, banker, driver, darkness, sadness, unlock, unafraid, difficulty, sunny, disobey, disagree, reassure, decode, relation, celebration



Recognise, resume, reflect, demand, deduce, involve, insist, sorcerer, inquisitive, generate, appreciate, compensate, corrupt, contain, journal, abysmal

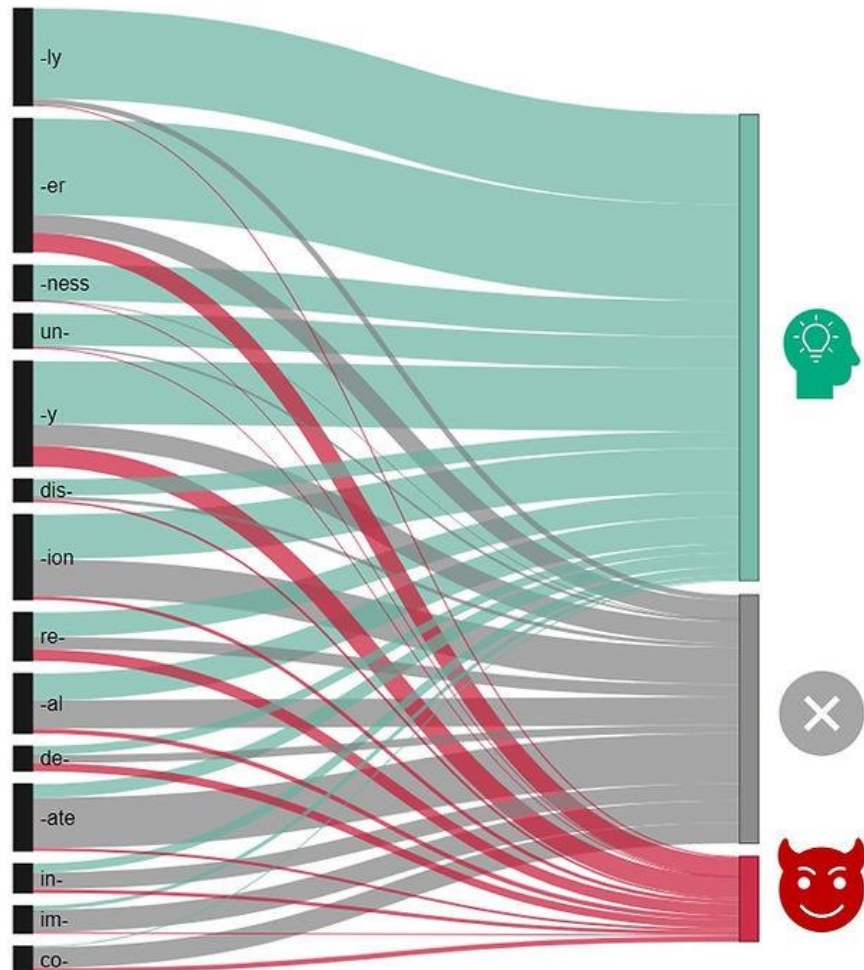


Brother, army, brandy, forty, deliver, detail, record, retired, comedian



forty
deliver

Learning morphemes is **challenging**



Reading alone is unlikely to
be enough to learn the
meanings of many individual
morphemes

Key insights so far...

- 💡 Access to **language beyond everyday use**
- 💡 **Wonderful opportunity** to build vocabulary, **but challenging** from the get-go
- 💡 Building vocabulary & reading proficiency requires **reading widely**
- 💡 **Strong phonics & morphology skills** are a must
- 💡 **Plenty** of opportunities to experience **morphology**
- 💡 **Skills** to access complex words **can't be gained from reading alone**
- 💡 **Pupils will require support** to make the most of what reading can offer



There is a **virtuous cycle** between **reading skills** and **reading motivation**



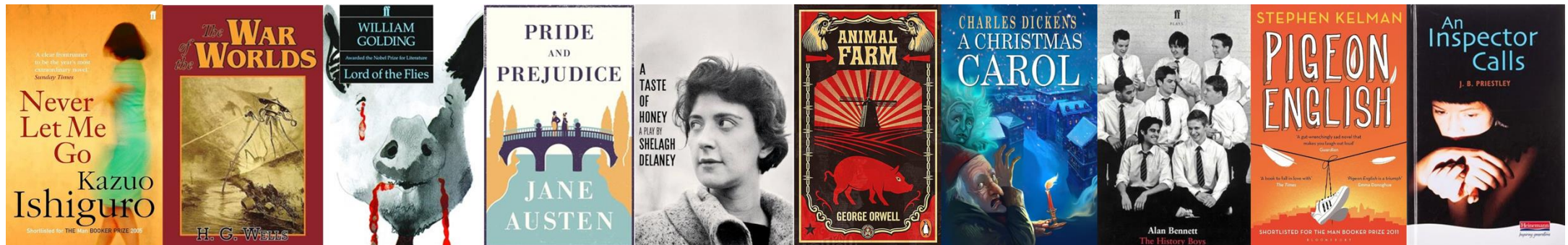
A closer look at the 13+ books

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



GCSE English Literature: Overview

- Curriculum set by DfE
- Qualifications delivered by exam boards (e.g., AQA, EdExcel, OCR)
- Teachers may choose from recommended texts, provided they include...
 - at least one play by Shakespeare
 - at least one 19th century novel
 - fiction or drama from the British Isles from 1914 onwards
 - a selection of post-1789 poetry (must include Romantic poetry and consist of at least 300 lines & 15 poems)



The 19th century novels

Hitherto it had touched him on the intellectual side alone; but now his imagination also was engaged, or rather enslaved; and as he lay and tossed in the gross darkness of the night and the curtained room, Mr. Enfield's tale went by before his mind in a scroll of lighted pictures. | He would be aware of the great field of lamps of a nocturnal city; then of the figure of a man walking swiftly; then of a child running from the doctor's; and then these met, and that human Juggernaut trod the child down and passed on regardless of her screams. | Or else he would see a room in a rich house, where his friend lay asleep, dreaming and smiling at his dreams; and then the door of that room would be opened, the curtains of the bed plucked apart, the sleeper recalled, and lo! there would stand by his side a figure to whom power was given, and even at that dead hour he must rise and do its bidding.

011 Robert Louis Stevenson: *The Strange Case of Dr. Jekyll and Mr. Hyde*

Read the following extract from Chapter 2 (Search for Mr Hyde) of *The Strange Case of Dr. Jekyll and Mr. Hyde* and then answer the question that follows.

In this extract, Utterson is alone thinking about Mr Hyde.

That was the amount of information that the lawyer carried back with him to the great, dark bed on which he tossed to and fro, until the small hours of the morning began to grow large. It was a night of little ease to his toiling mind, toiling in mere darkness and besieged by questions.

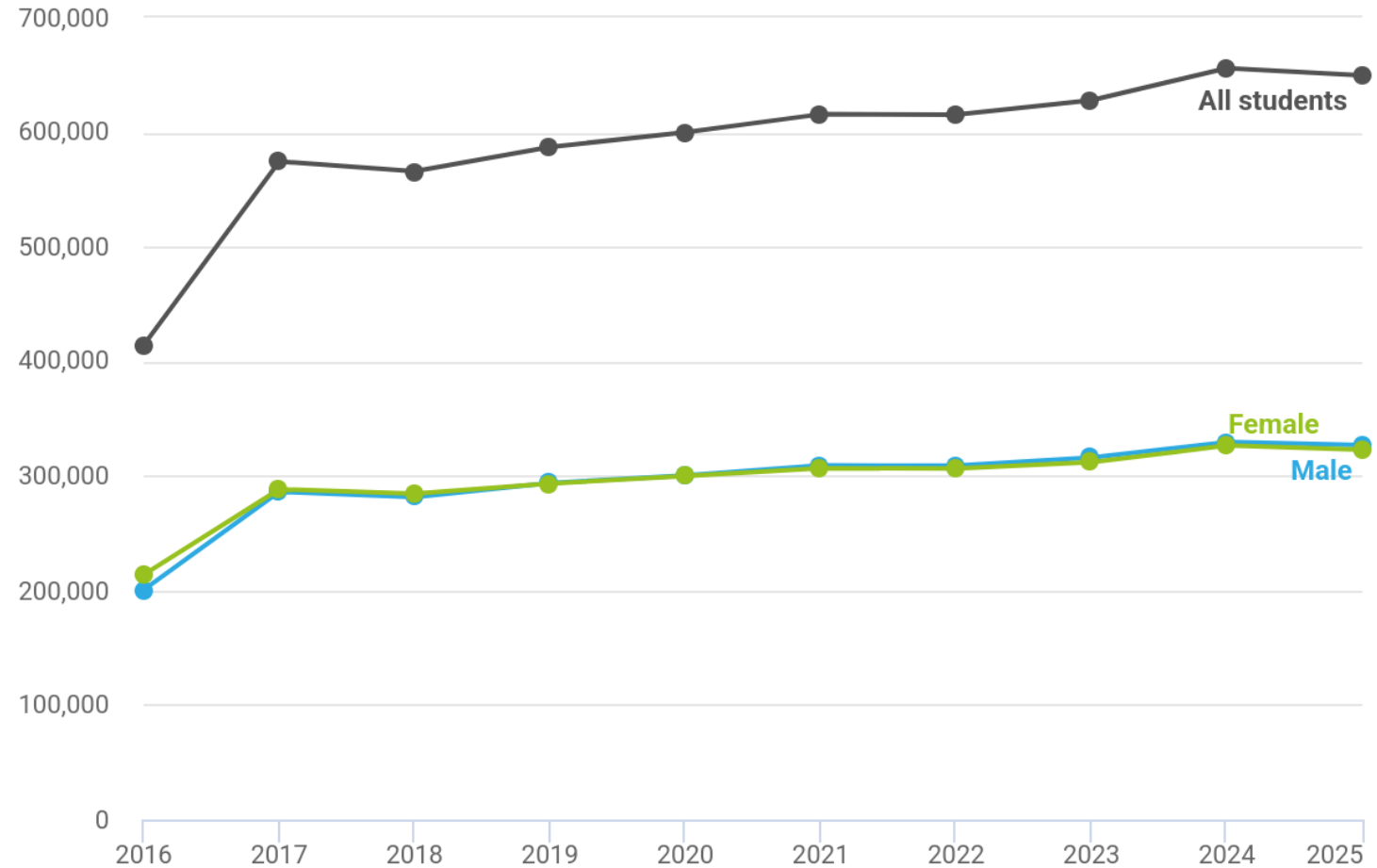
5 Six o'clock struck on the bells of the church that was so conveniently near to Mr. Utterson's dwelling, and still he was digging at the problem. Hitherto it had touched him on the intellectual side alone, but now his imagination also was engaged, or rather enslaved, and as he lay and tossed in the gross darkness of the night and the curtained room, Mr. Enfield's tale went by before his mind in a scroll of lighted pictures. He would be aware of the great field of lamps of a nocturnal city; then of the figure of a man walking swiftly; then of a child running from the doctor's; and then these met, and that human Juggernaut trod the child down and passed on regardless of her screams. Or else he would see a room in a rich house, where his friend lay asleep, dreaming and smiling at his dreams, and then the door of that room would be opened, the curtains of the bed plucked apart, the sleeper recalled, and lo! there would stand by his side a figure to whom power was given, and even at that dead hour he must rise and do its bidding. The hours in these two phases haunted the lawyer all night, and if at any time he dozed over, it was but to see it glide more stealthily through sleeping houses, or move the more swiftly and still the more swiftly, even to dizzy heights through wider labyrinths of lamp-lighted city, and at every street corner crush a child and leave her screaming. And still the figure had no face by which he might know it, even in his dreams, it had no face, or one that baffled him and melted before his eyes, and thus it was that there sprang up and grew apace in the lawyer's mind a singular story, almost an anecdote, contrary to belief the features of the real Mr. Hyde. If he could but once set eyes on him, he thought the mystery would lighten and perhaps roll altogether away, as was the habit of mysterious things when well examined. He might see a reason for his friend's strange preference or bondage (call it which you please) and even for the startling clauses of the will. And at least it would be a face worth seeing; the face of a man who was without bowels of mercy; a face which had but to show itself to raze up, in the mind of the unrepentable Enfield, a spot of enduring hatred.

30 From that time forward, Mr. Utterson began to haunt the door in the by-street of shops. In the morning before office hours, at noon when business was plenty and time scarce, at night under the face of the fogged city moon, by all lights and at all hours of solitude or concourse, the lawyer was to be found on his chosen spot.

How does Stevenson create the atmosphere of fear and danger in this extract and in the novel as a whole?

GCSE entries in English literature, 2016-2025

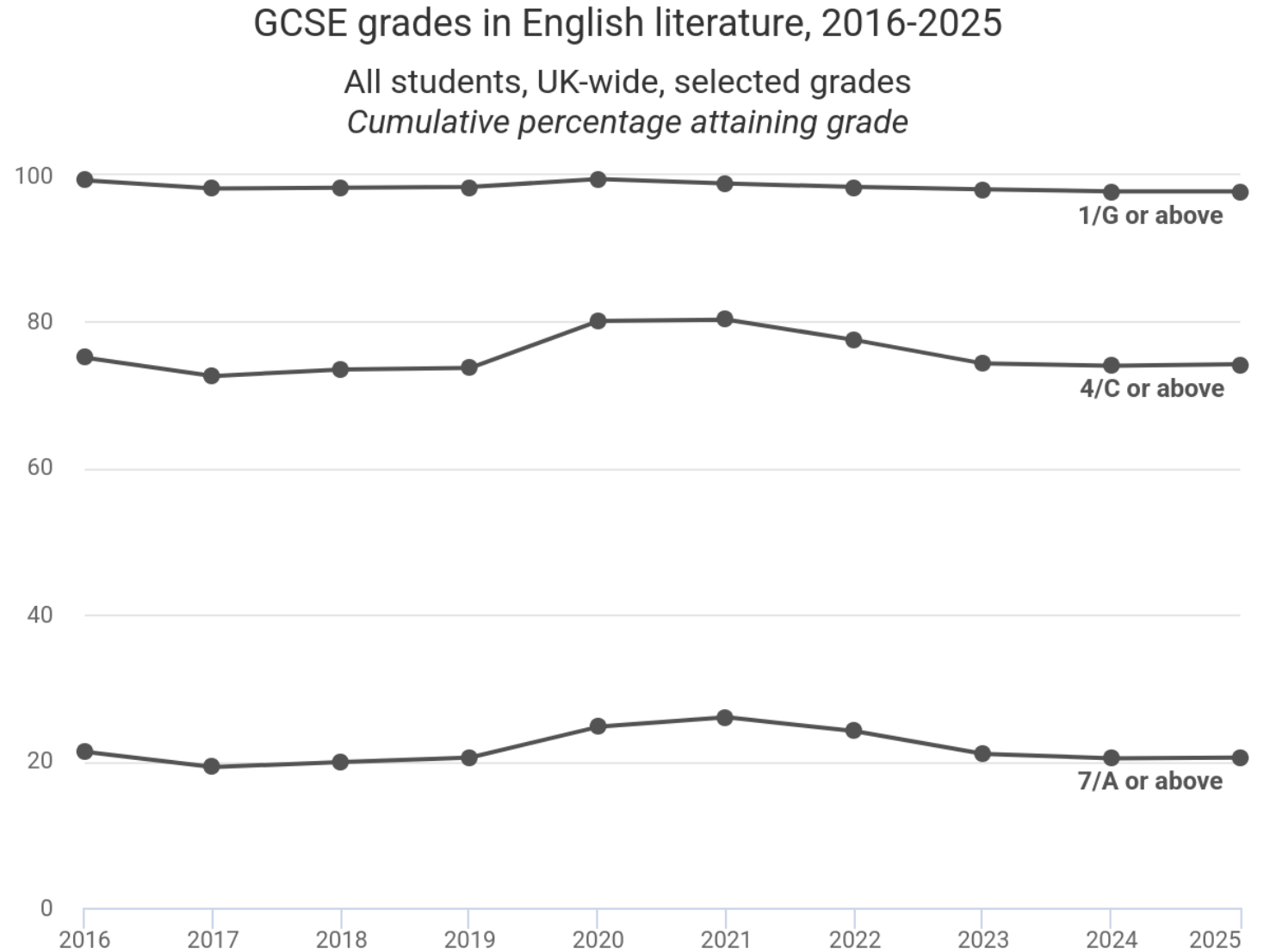
All students, UK-wide
Number of entries



**Over 500,000
entries per year
since 2017**

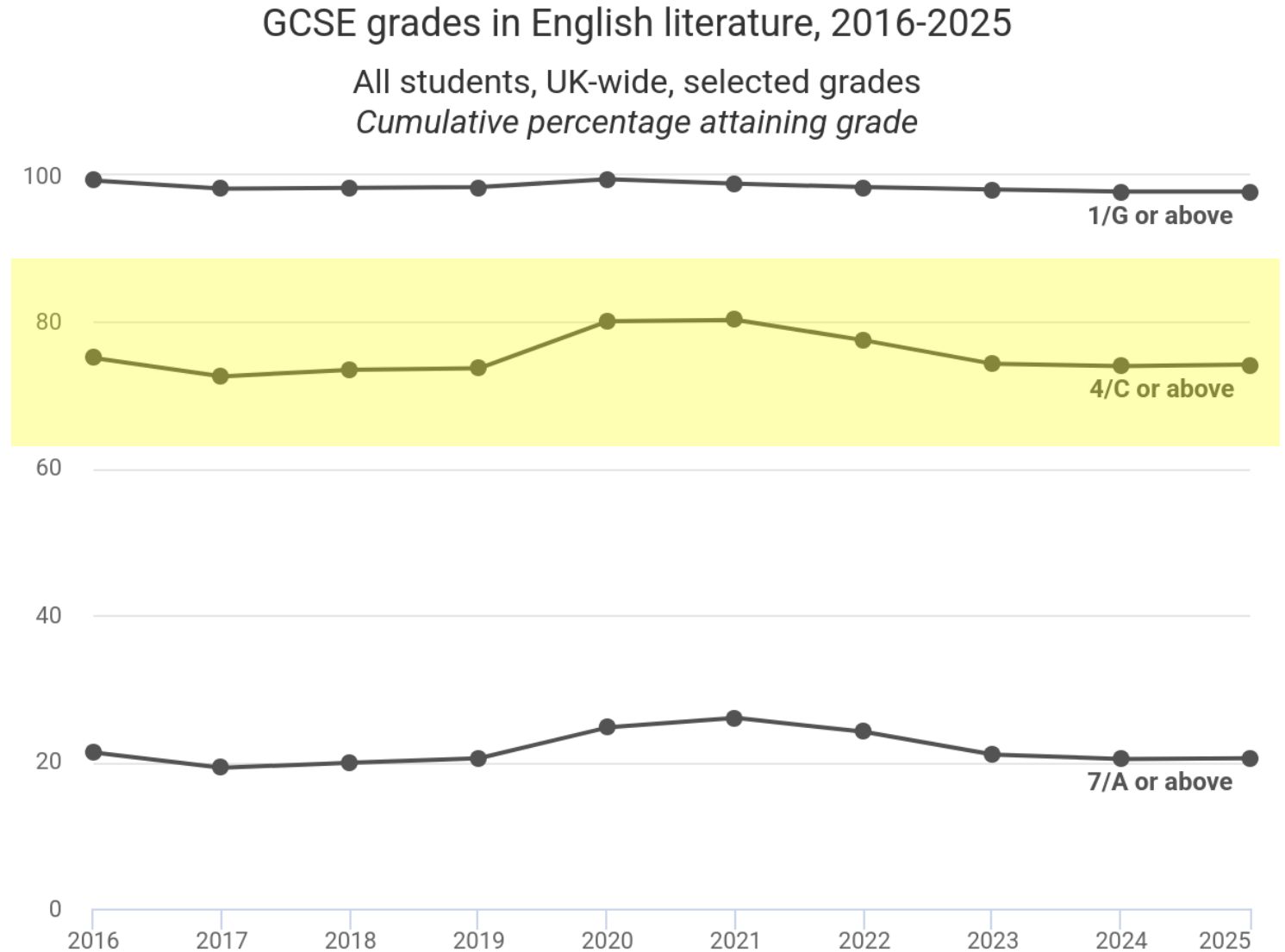
Source: FFT Education Datalab analysis of JCQ data
Project funded by the Nuffield Foundation

Pass rates **higher**
than in other
subjects



Source: FFT Education Datalab analysis of JCQ data
Project funded by the Nuffield Foundation

And yet, **25%** of candidates **do not achieve a pass grade**



Source: FFT Education Datalab analysis of JCQ data
Project funded by the Nuffield Foundation

How do GCSE texts compare to popular books?

...And what makes them so challenging?

Highly dense vocabulary in the GCSE books



- 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



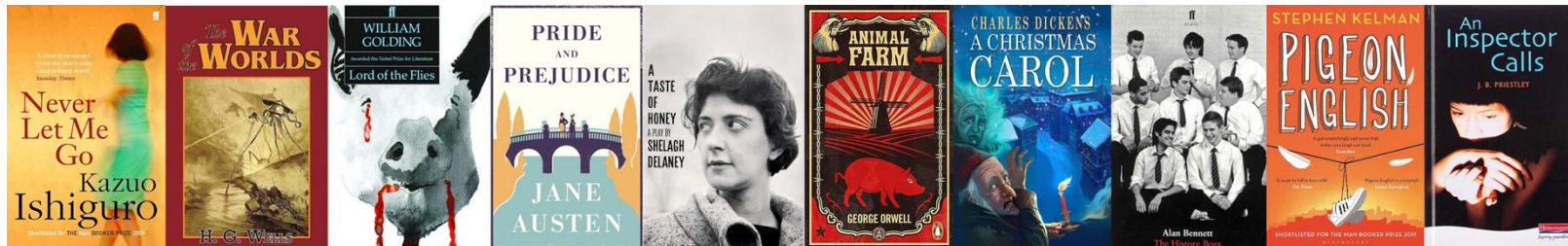
- **More vocabulary** through **less text** in the GCSE books
- Texts **harder to understand** as a whole
- More bang for your buck **but at what cost?**

Many unfamiliar words in the GCSE books

- Only **33%** of the distinct words **occur regularly** in popular books
- The remaining **67%** are **used sparsely**
- **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 these texts at all



Many of the new GCSE words are **new roots**

- In popular books, many of the unfamiliar words contain **multiple morphemes**

→ Meanings **can be derived** from the words' constituents: mourn + ful + ly

- In GCSE books, most “new” words are **new roots**

→ Meanings **cannot be derived** from the words' structure

- *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**

Is GCSE English Lit **fit for purpose**?

GCSE English Literature should enable students to. . .

- read in depth, critically, and evaluatively
- develop the habit of reading widely and often
- appreciate the depth and power of the English literary heritage

Yet...

- Less than 30% of children read daily in their free time
- Many will lack the skills required to access the GCSE course



Curriculum review **should consider** the huge **variation** in children's language and reading skills...

Policy changes underway

- **2024, The English Association [report](#)**
 - GCSE English Lang & English Lit are “not fit for purpose”
 - Reform needed to “make them more diverse and inclusive in content and scope”
- **July 2024, Curriculum & Assessment [Review](#)**
 - “promote a wider range of literary voices while retaining core texts”
 - “continued study of Shakespeare and 19th century literature”, BUT “more diverse array of authors from the British Isles”
 - “without adding to the mandatory volume of content”
 - No mention of variation in pupils’ reading and language skills
- **AQA [response](#)**
 - Consultation in 2026, new curriculum in spring 2027, first teaching in 2028

Reflection points...

- 💡 Reading is **powerful but challenging**. What implications does this have for our role in pupil's reading?
- 💡 Reading is **important** for learning about morphemes, but it's **not enough**. Time to rethink morphology instruction?
- 💡 There is a **wide variation** in language and reading skills that pupils bring into the classroom. How can teaching & assessment respond?



<https://doi.org/10.1038/s41539-025-00313-6>

Original Article



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

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

Check for updates

Maria Korochkina & Kathleen Rastle

<https://cyp-lex.rastlelab.com/>



Maria Korochkina & Kathy Rastle
Mar 12, 2024 · 5 min read

What Words do Children Encounter When They Read for Pleasure?

The ability to read opens up worlds. Reading enables children to progress into post-primary education and provides the basis for lifelong...

<https://www.rastlelab.com/blog>



Maria Korochkina & Kathy Rastle
May 19 · 5 min read

What can children learn about morphology from reading for fun?

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

<https://www.tes.com/magazine>

THE VOCABULARY BARRIER IN THE GENERAL CERTIFICATE OF SECONDARY EDUCATION (GCSE) IN ENGLISH LITERATURE

Maria Korochkina^{*} and Kathleen Rastle

The English Association Journal
for Teachers of English

THE USE OF english

The power of reading: books, vocabulary and learning

<https://www.tooledupeducation.com/>

Further reading



Marco Marelli



Marc Brysbaert



Kathy Rastle

Thank you!

m.korochkina@aston.ac.uk

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



Economic
and Social
Research Council

