A corpus-based developmental investigation of linguistic complexity in children's writing

Hsiao, Yaling; Dawson, Nicola J.; Banerji, Nilanjana; Nation, Kate

DOI:
10.1016/j.acorp.2024.100084

Applied Corpus Linguistics 4 (2024) 100084

Contents lists available at ScienceDirect

Applied Corpus Linguistics

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

Articles

A corpus-based developmental investigation of linguistic complexity in children's writing☆

Yaling Hsiao [a,*], Nicola J. Dawson [b], Nilanjana Banerji [c], Kate Nation [b]

[a] *University of Birmingham, United Kingdom*
[b] *University of Oxford, United Kingdom*
[c] *Oxford University Press, United Kingdom*

ARTICLE INFO

Keywords:
Lexical development
Grammatical development
Writing development
Corpus analysis
Linguistic complexity
Principal component analysis

ABSTRACT

Writing proficiency is associated with linguistic complexity. We used measures of linguistic complexity to investigate the development of children's narrative writing using a large corpus of short stories (N>100,000) written by children aged 5–13 in the UK. Linguistic complexity was assessed using both lexical ($N = 30$) and syntactic ($N = 14$) measures. Most measures were associated with age, with writing by older children showing greater lexical density, sophistication, and diversity than writing by younger children. Older children also used longer sentences, and longer T-units and clauses, and the density of smaller syntactic units inside larger units was also higher. Principal Component Analysis identified a number of dimensions associated with complexity, with the first two dimensions capturing nearly 50 % of variance. Lexical diversity was mainly represented on the first dimension and syntactic complexity on the second. Across the age range, there was wider variation in syntactic complexity than in lexical diversity, suggesting that syntactic development is subject to more individual differences than the ability to use a diverse set of lexical items. Our findings quantify the nature and content of children's writing through mid-childhood, and we discuss the utility of analysing children's writing using a computational, data-driven approach.

A corpus-based developmental investigation of linguistic complexity in children's writing

A good writer communicates clearly and coherently, and with a tone and register that is appropriate for the communication context. Unlike conversation that happens in the here and now, aided by facial expression, gesture, intonation and a shared context, written language is typically de-contextualised and remote. As such, writing is a form of communication that requires words and sentences to be crafted with precision so that the mind of the writer is recreated for the reader and as a consequence, written language is generally more complex than spoken language (e.g., Biber, 1988; Roland et al., 2007). Notably, this holds for children too: child-directed print is lexically and syntactically more complex than child-directed speech (Dawson et al., 2021; Hsiao et al., 2022; Montag, 2019; Montag et al., 2015; Montag and MacDonald, 2015). These differences start early and are present in books written for pre-school children to listen to in the context of shared reading (for review, see Nation et al., 2022).

When do these linguistic features emerge in children's own writing and how do they build with development? Relatively little research (for reviews, see Crossley, 2020; Durrant et al., 2021) has charted the emergence of linguistic complexity in young children's writing, quantitatively and at-scale, certainly in comparison to the large evidence base that links linguistic complexity and writing proficiency in second language (L2) or English as a Foreign Language (EFL) learning. Our aim was

☆ Author notes: This research was funded by a British Academy Post-Doctoral Fellowship (PF2/180013) and a John Fell COVID Rebuilding Research Momentum Fund grant (0010144 CRRMF) awarded to Yaling Hsiao, a grant from the Nuffield Foundation (EDO/43392) to Kate Nation, and resources made available to Nilanjana Banerji by the department of Children's Dictionaries and Children's Language Data at Oxford University Press. Data and code associated with this paper are available on the Open Science Framework website (https://osf.io/wuzaf). The Oxford Children's Language Corpus is a growing database of writing for and by children developed and maintained by Oxford University Press for the purpose of children's language research. We thank Songjun He for research assistance. Correspondence concerning this article should be addressed to Yaling Hsiao, School of Education, University of Birmingham, Birmingham, B15 2WN, UK. E-mail: y.hsiao@bham.ac.uk.

* Corresponding author.
E-mail address: y.hsiao@bham.ac.uk (Y. Hsiao).

https://doi.org/10.1016/j.acorp.2024.100084
Received 22 August 2022; Received in revised form 10 November 2023; Accepted 4 January 2024
Available online 7 January 2024

to fill this research gap by analysing a large developmental corpus of stories written by 5–13 year-old children in the UK. In the following sections, we describe measures of lexical and syntactic complexity developed in previous studies in L1 and L2. We consider how these have informed our understanding of writing development and we identify the research gaps which we aimed to fill with the current study.

Measures of linguistic complexity

Length of basic units

Zipf's law states that word frequency is highly influenced by word length (Zipf, 1949). That longer words tend to be lower in frequency and later aquired (with r's of −0.40 (Brysbaert and New, 2009) and 0.40 (Brysbaert and Biemiller, 2017), respectively) is a potential consequence of communicative pressure that reflects working memory constraints in interaction with the complexity of information content (Piantadosi, 2014). It is therefore reasonable to expect that children produce longer words in writing with age, yet previous studies have provided mixed evidence. Evidence is also mixed for the hypothesis that better writing is associated with longer words. For example, Deno et al. (1982) found that for 7–12 year-olds, word length was constantly and strongly correlated with grade level. In another study, Olinghouse and Leaird (2009) asked 2nd and 4th grade children to produce two written narratives. One narrative revealed developmental increases in word length and an association with subjective writing quality, but there was no age difference for the other narrative, and only an association with quality for the 2nd grade children.

Sentence length and text length also appear to index writing maturity. These two measures are related: when sentences in a text contain more words, the text is necessarily longer by word count. Older children tend to produce longer sentences, and therefore longer texts. These length measures could be viewed as a proxy of writing "fluency"(Chenoweth and Hayes, 2001; Durrant et al., 2021). In fact, they do not always correlate with age: older children are more likely to have learned to write within a word limit (in standardised assessments for example), and to have been instructed to be concise rather than verbose (Deane and Quinlan, 2010; Myhill, 2008), whereas younger children are more inconsistent in marking sentence boundaries (Golub and Frederick, 1970). Thus, the validity of sentence and text length measures remain unclear.

Measures of lexical complexity

Lexical complexity refers to the breadth and quality of vocabulary use and has been linked to the overall quality of writing (Engber, 1995). Lexical complexity comprises three components: lexical density, lexical sophistication and lexical diversity (Durrant et al., 2021; Lu, 2012; Read, 2000). *Lexical density* is generally operationalised as the ratio of the number of lexical words (i.e. nouns, verbs, adjectives and adverbs derived from adjectives) to the total number of words in a text (Ure, 1971). Not all studies find age differences in lexical density (Berman and Nir, 2010) and greater lexical density does not always correlate with judgements of writing quality (Uccelli et al., 2012). Berman and Nir (2010) reported that children's writing was more lexically dense than their spoken productions, but they found no age differences in written lexical density. Similarly, Uccelli et al. (2012) found no correlation between lexical density and either age or writing quality in their analysis of academic writing in a high school sample. These mixed findings might be associated with small sample sizes, which calls for larger-scale developmental data.

Lexical sophistication refers to the proportion of advanced or difficult words in a text. This tends to be indexed by frequency, with rare words being identified as those beyond a certain frequency rank in a reference corpus (Finn, 1977). In general, lexical sophistication is associated with writing quality. For example, Deno (Deno et al., 1982) found that the use of mature words (i.e. words not in a common word list) by Grade 3 and 6 children was linked with higher writing quality, as assessed by a standardised instrument. This study also observed an increase in lexical sophistication from Grade 3 to Grade 6, consistent with there being a general association between lexical sophistication and age. While this is supported by some studies (e.g., Olinghouse and Leaird, 2009; Olinghouse and Wilson, 2013), the evidence base points to more complex relations. Some studies have failed to find an overall association with age (Crossley et al., 2011; Durrant and Brenchley, 2019). Part of speech and genre also influence variation in lexical sophistication, as does whether the unit of calculation is based on word token or by type (e.g., Durrant and Brenchley, 2019). This complex picture again suggests that further research is needed.

Lexical diversity refers to vocabulary range and breath, and it is generally associated with writing quality and proficiency (Engber, 1995; Lu, 2012; Treffers-Daller et al., 2018; Zenker and Kyle, 2021). It is typically operationalised as the ratio of number of unique word types relative to the total number of word tokens in a text (Templin, 1957). Type-to-token ratio (TTR) is heavily dependent on text length as words are more likely to be repeated in longer texts. That is, as the number of tokens increases linearly, the increase in types steadily decreases. To illustrate, consider two sentences from a piece of writing in the corpus used in this study, written by the same child: "*I wonder if I can have pancake*" and "*I wonder if I can pick my nose with my elbow*". The first sentence has 6 unique word types and 7 word tokens, resulting in a TTR of 0.86. The second sentence has 9 unique word types and 11 word tokens, with a TTR of 0.82 (see also Table 1 for calculation). The lower TTR of the second sentence is due to its longer length (i.e., larger denominator) and the larger number of repetitions of words within the sentence (i.e., smaller nominator). In an attempt to correct for text length, studies have applied transformations (e.g. log transformation) or used a variety of sampling approaches (e.g. mean TTR for every 50-word window) (for reviews, see Lu, 2012; McCarthy and Jarvis, 2010; Zenker and Kyle, 2021). Using the same example sentences, after applying log transformation, both sentences would return a log TTR of 0.92. There is ample discussion in the literature as to the robustness of this type of measure with regard to variation in text length, text genre, and whether the text is produced by L1 or L2 speakers. For example, Zenker and Kyle (2021) applied TTR along with eight other indices of lexical diversity (all developed to address the issue of text length) to a large corpus of L2 argumentative essays. They found that Moving-Average TTR (MATTR) and Measure of Textual Lexical Diversity (MTLD) (see Table 1 and Zenker and Kyle, 2021, for definitions) yielded highly stable values across text length. The authors argued that these metrics should be used to assess writing, in preference to the other lexical diversity measures. This serves as an example of the utility of a large-scale data driven approach to assess L2 writing. In our study, we followed a similar approach to address writing development in L1 writing.

Measures of syntactic complexity

Turning to syntax, complex syntactic structures allow people to express complicated ideas or relationships, and to do so in a more precise and sophisticated manner (Beers and Nagy, 2009). Syntactic complexity is typically quantified using sentence-level length of production units (e.g., T-units, clauses), the amount of subordination or coordination, and the number of particular syntactic structures (Ortega, 2003). There is evidence showing that older children produce longer syntactic units and more varied and complex syntactic structures than younger children (Beers and Nagy, 2009; Crossley et al., 2011; Durrant et al., 2020, 2021; Durrant and Brenchley, 2019; Hunt, 1965; Myhill, 2008), as do second language learners by proficiency level (Lu, 2010; Ortega, 2003).

T-unit is important in the assessment of language proficiency. The definition of a T-unit is a dominant clause and its dependent clauses: it is the "minimally terminable unit" into which sentences can be split (Hunt, 1965). A sentence with a coordinating conjunction structure like

Table 1
The 30 lexical richness measures used in this study. Calculation is illustrated using an example sentence.

#	Code	Measure	Definition	Calculation using the example sentence *I wonder if I can pick my nose with my elbow*
Lexical Density				
1	ld	lexical density	number of content word tokens/number of all tokens	$\frac{4}{11} = 0.36$(content words: *wonder, pick, nose, elbow*)
Lexical Sophistication				
2	ls1	lexical sophistication by token	sophisticated tokens/all tokens	$\frac{2}{11} = 0.18$ (sophisticated tokens: *nose, elbow*)
3	ls2	lexical sophistication by type	sophisticated types/all types	$\frac{2}{9} = 0.22$ (sophisticated types: *nose, elbow*)
4	vs1	verb sophistication 1	sophisticated verb types/all verb tokens	$\frac{0}{2} = 0$
5	vs2	verb sophistication 2	sophisticated verb types/square root of 2 * number of all verb tokens	$\frac{0}{\sqrt{2*2}} = 0$
6	cvs1	verb sophistication 3	square of number of sophisticated verb types/all verb tokens	$\frac{0^2}{2} = 0$
Lexical Diversity				
7	ndw	types	number of different words	9
8	ndwz	types in first 50 words	number of types in the first 50 tokens	9
9	ndwrtz	types in 50-word samples	mean number of types in the 10 samples of 50 random tokens	9
10	ndwesz	types in 50-word sequences	mean number of types in the 10 samples of 50-word sequences	9
11	ttr	type-token ratio	number of all types/number of all tokens	$\frac{9}{11} = 0.82$
12	msttr	mean segmental TTR	splitting the text into 50-word segments, mean TTR of all segments	$\frac{9}{11} = 0.82$
13	cttr	corrected TTR	types/square root of 2 * tokens	$\frac{9}{\sqrt{2*11}} = 1.92$
14	rttr	root TTR	types/square root of tokens	$\frac{9}{\sqrt{11}} = 2.71$
15	logttr	logarithmic TTR	log(types)/log(tokens)	$\frac{\log 9}{\log 11} = 0.92$
16	uber	uber TTR	(Square of log(tokens))/log(tokens/ types)	$\frac{(\log 11)^2}{\log(11/9)} = 12.44$
17	MATTR	moving average TTR	TTRs for a moving window of tokens (e.g. 50 words) from the first to the last token, computing a TTR for each window	$\frac{9}{11} = 0.82$
18	HDD	hypergeometric distribution diversity index	for each word type, the probability of encountering one of its tokens in a random sample of 42 tokens	0, as there is only one sample, with less than 42 tokens
19	MTLD	measure of textual lexical diversity	the average number of words in a row for which a certain TTR is maintained	16.94 (see McCathy & Jarvis, 2010, for operationalisation)
20	MTLD-MA-wrap	moving-average wrapped MTLD	MTLD but instead of calculating partial factors, it wraps to the beginning of the text to complete the last factors	13.0 (see Zenker and Kyle, 2021, for operationalisation)
21	MTLD-bi	moving-average bidirectional MTLD	MTLD in each direction using a moving window	0 (see Zenker and Kyle, 2021, for operationalisation)
22	lv	lexical word variation	content word types/all content word tokens	$\frac{4}{4} = 1$
23	lv1	verb variation 1	verb types/all verb tokens	$\frac{2}{2} = 1$
24	svv1	verb variation 2	verb types/square root of 2 * all verb tokens	$\frac{2}{\sqrt{2*2}} = 1$
25	cvv1	verb variation 3	square of number of verb types/all verb tokens	$\frac{2^2}{2} = 2$
26	vv2	verb variation 4	verb types/all content word tokens	$\frac{2}{4} = 0.5$
27	nv	noun variation	noun types/all content word tokens	$\frac{2}{4} = 0.5$
28	adjv	adjective variation	adjective types/all content word tokens	$\frac{0}{4} = 0$
29	advv	adverb variation	adverb types/all content word tokens	$\frac{0}{4} = 0$
30	modv	modifier variation	adjective + adverb types/all content word tokens	$\frac{0+0}{4} = 0$

"*There was a boy in my class, and he liked playing football*" has two T-units, whereas a sentence clause like "*There was a boy in my class who liked playing football*" has only one T-unit because the relative clause ("who liked playing football") is a dependent clause attached to the dominant main clause ("there was a boy in my class"). Both sentences have two clauses but differ in number of T-units. T-unit length therefore signals the ability to use subordination to combine clauses. Subordination allows for the relationships between elements to be expressed, without the use of coordination (e.g., "and"). It is therefore a way of avoiding long run-on sentences, like those typically produced by younger children (Crosson et al., 2008).

Density measures or measures based on ratios are useful to capture this complexity as they are based on the concept that "complexity is a product of number of component parts within a feature and the number and nature of connections between those parts" (Durrant et al., 2020, p.422). For example, consider the metric T-unit complexity ratio as a measure of the density of clauses in a T-unit. A sentence involving coordinating conjunction like "*There was a boy in my class, and he liked* *playing football*" has a score of 1: the clause count of 2 is divided by the T-unit count of 2. A sentence with a relative clause like "*There was a boy in my class who liked playing football*" has a score of 2: the clause count of 2 is divided by the number of T-units, in this case 1. In this example, the sentence with a relative clause is quantified as more syntactically complex than the coordinated sentence with a conjunction.

Finer-grained analyses of clausal or phrasal features show that adverbial clauses (e.g. "He was happy *if he could play football*") and complex nominals (e.g. "*The involvement of football* in his life makes him happy") predict language development and writing quality, both in L1 and L2 (Durrant et al., 2020; Durrant & Brenchley, 2023; Kyle and Crossley, 2018; Li et al., 2023; Lu, 2010). These indices are also sensitive to genre and communicative purpose. For example, Durrant et al. (2020) detailed analysis of adverbial clauses and Durrant and Brenchley's (2022) analysis of complex noun phrases present a complex developmental picture in which usage varies as children transition with age from mainly fiction writing to more expository or persuasive writing.

Recent advances in Natural Language Processing (NLP) techniques also facilitate the development and application of more sophisticated and accurate assessment of syntactic complexity. Previously, quantifying relative clause usage required laborious manual annotation by skilled linguists. Naturally, this served to limit sample size. With automatic parsing and analysis software readily available, researchers can process much larger language databases and perform standardised analyses, making it easier to compare findings across studies.

The current study

We took a data-driven approach to analysing a large corpus of children's writing (over 100,000 pieces) across a large developmental window (5–13 years of age). As outlined above, lexical and syntactic complexity can each be captured in many different ways, as to be expected given the multidimensional nature of linguistic complexity. This breadth has not been applied to children's first language writing across a range of proficiency levels and within the same study (cf. the larger literature on writing by second language learners, e.g. Ortega, 2003), using automatic analysis tools. Our first aim was to fill this research gap. Rather than focusing on one or two features, we calculated 48 different metrics tapping production unit length, lexical and syntactic complexity and used these to examine developmental change in children's writing through the primary and early secondary school years, as they transition to more academic-like language (Durrant and Brenchley, 2019, 2023; Nippold, 2007).

From this, we took a statistical approach and used Principal Component Analysis (PCA) to identify the underlying relationships between the different measures and components that best explained linguistic complexity in children's writing. We then considered developmental change in these markers of linguistic complexity by comparing the writing of younger and older children. In summary, our overall aim was to describe the nature of linguistic complexity in a large cross-sectional corpus of children's writing and to discover whether this 'macro' approach (in terms of number of measures as well as size of the corpus and its developmental range) has the potential to complement the insights provided by detailed analyses of individual aspects of linguistic complexity in particular age ranges, noting that previous work with smaller samples might generate subtle and unreliable findings (e.g., Durrant and Brenchley, 2019; Myhill, 2008).

Method

The corpus

We used the writing component of the Oxford Children's Language Corpus, held by Oxford University Press. In total, this contains over one million stories written by 5–13-year-old children in the UK. The stories were sourced from BBC Radio 2 500 Words, a national children's writing competition that ran annually for 10 years, 2011–2021. Each year, children were invited to submit entries on any topic – the only constraint was that they must be no more than 500 words. The competition had significant media attention across national TV and radio, and it was also promoted within schools. As such, this resource provided a naturally occurring language sample (Goldstone and Lupyan, 2016) not generated for assessment purposes, nor contaminated by experimenter prompts or cues. This allowed us to analyse children's own free writing, without constraints from time, place, instruction, or topic. This bottom-up approach to analysing language can reveal patterns in how people use language, and from this, help address psychological questions about children's development (Jackson et al., 2021).

We selected all the stories submitted in 2019 ($N = 107{,}273$ stories; approximately 55 million words). We used the Key Stage information available as metadata for each story to approximate developmental stage. Key Stage refers to bandings within the education system of England and Wales, with 5–7-year-olds falling within Key Stage 1, 7–11-year-olds into Key Stage 2 and 11–14-year-olds into Key Stage 3. The majority of entries (59 %) came from children in Key Stage 2; 39 % of entries came from children in Key Stage 3 and only 2 % from the youngest children in Key Stage 1. We note the relative imbalance in sample size between Key Stage 1 and the others. Nevertheless, the total number of stories written by Key Stage 1 children was 3625. This still represents a substantial sample of children at the early stages of learning to write.

The Key Stage information is educationally useful as it relates to the curriculum. Clearly, however, the age bands within a stage are large, and there are likely large differences between, say, 5-year-olds and 7-year-olds within Key Stage 1. To capture development in a more fine-grained way, we repeated all analyses and replaced Key Stage with age in years. The results were similar and for simplicity in reporting and visualising the data, this report focuses on the results based on Key Stage. All analyses by age are openly available on OSF (https://osf.io/wuzaf). Any differences between findings by Key Stage and those by age in years are highlighted in the Results section.

Measures of linguistic complexity

Length of basic units

Under the assumption that older children would produce more words and longer words, and longer sentences, we adopted four measures to capture length of production units: number of letters per word, number of words per sentence, number of words per story and number of sentences per story.

Measures of lexical complexity

The three components of lexical complexity (density, sophistication, diversity) have been captured in numerous ways. Lu (2012) used 24 different measures in a corpus investigation of the quality of L2 speakers' oral narratives. We adopted these in our study, and added six more designed to be less sensitive to text length (Zenker and Kyle, 2021). Table 1 summarises all 30 measures.

Lexical density was defined as the ratio of the number of lexical words (i.e. nouns, verbs, adjectives and adverbs derived from adjectives) to the total number of words in a text (Ure, 1971). This single measure was used to quantify lexical density. *Lexical sophistication,* following Lu's (2012) methodology, used the frequency of rare word types not featured in the top 2000 types in the British National Corpus (BNC) as a proxy (Leech et al., 2001). Note that the BNC is predominately an adult corpus, but it represents a cross-section of British English from a wide range of sources and following Durrant and Brenchley's assumption that "sophistication should be gauged with reference to the sort of discourse towards which children's education aims" (2019, p. 1934), we retained its use here. We also calculated verb sophistication (Harley and King, 1989), defined as the ratio of verb types to the total number of verbs. Overall, we examined children's lexical sophistication using five different measures, as listed in Table 1.

Lexical diversity was captured through type-to-token ratio (TTR), as well as several other methods that attempt to correct for text length, such as those that apply transformations (e.g. log TTR, root TTR, Maas Index) or use a variety of sampling approaches (e.g. MTLD, MSTTR, MATTR, HDD) (for reviews, see Lu, 2012; McCarthy and Jarvis, 2010; Zenker and Kyle, 2021). Our study provided an opportunity to examine these different metrics at the same time in the context of children's narrative writing. We also considered lexical diversity for different parts of speech by calculating separate type-to-token measures for nouns, verbs, adjectives, and adverbs. In total, 24 measures were used to estimate lexical diversity.

Measures of syntactic complexity

We analysed 14 syntactic complexity measures (Table 2), building from Lu (2010) who developed automatic computing software for assessing L2 writing. These included three indices of *length of syntactic production units*, i.e., length of sentences, length of T-units, and length of clauses.

Alongside these holistic measures of production unit length, we also included 11 measures of *syntactic complexity ratio*, listed in Table 2. These provide an indication of clause density within a production unit. Ratios were calculated with the number of specific types of syntactic structure as the nominator and a production unit (e.g., number of T-units) as the denominator. Syntactic structures were concerned with subordination (e.g., complex T-units, dependent clauses), coordination (i.e., coordinate phrases) or other types of fine-grained complex clauses or phrases (e.g., complex nominals).

Table 2
The 14 syntactic complexity measures used in this study. Calculation is illustrated using an example sentence.

#	Code	Measure	Definition	Calculation using the example sentence *I wonder if I can pick my nose with my elbow*
Unit of Production				
1	MLS	mean length per sentence	mean number of words in a sentence	11
2	MLT	mean length per T-unit	mean number of words in a T-unit	11
3	MLC	mean length per clause	mean number of words in a clause	$\frac{2+9}{2}=5.5$(There are two clauses in the sentence: *"I wonder"*, *"if I can pick my nose with my elbow".)*
Complexity Ratio				
4	CS	sentence complexity ratio	mean number of clauses per sentence	$\frac{2}{1}=2$
5	VP.T	verb phrase per T-unit	mean number of verb phrases per T-unit	$\frac{2}{1}=2$(each clause contains a verb phrase)
6	C.T	T-unit complexity ratio	mean number of clauses per T-unit	$\frac{2}{1}=2$
7	DC.C	dependent clause ratio	mean number of dependent clauses per clause	$\frac{0+1}{2}=0.5$(*"with my elbow"* is a dependent clause)
8	DC.T	dependent clause per T-unit	mean number of dependent clauses per T-unit	$\frac{0+1}{1}=1$
9	T.S	T-unit per sentence	mean number of T-unit per sentence	$\frac{1}{1}=1$
10	CT.T	complex T-unit ratio	Mean number of complex T-unit (a T-unit that contains a dependent clause) per T-unit	$\frac{1}{1}=1$
11	CP.T	coordinate phrases per T-unit	mean number of coordinate phrases per T-unit	$\frac{0}{1}=0$
12	CP.C	coordinate phrases per clause	mean number of coordinate phrases per clause	$\frac{0}{2}=0$
13	CN.T	complex nominals per T-unit	mean number of complex nominals per T-unit	$\frac{0}{1}=0$
14	CN.C	complex nominals per clause	mean number of complex nominals per clause	$\frac{0}{2}=0$

Procedure

We pre-processed the corpus by removing stories that were very short or possibly contained mainly nonsense words, i.e., those that contained only one sentence, or less than 30 words, or with average word length of over 10 letters, or average sentence length of over 50 words. For lexical complexity measures specifically, we removed punctuation and converted all characters to lower case to ensure that the same words in different cases were counted as the same type. After pre-processing, the final sample available for analysis comprised 105,065 stories (47.7 million word tokens). Each story was tagged with the child's Key Stage information (Key Stage 1, 2 or 3).

To measure length and compute the various lexical and syntactic complexity measures, we developed a Python script that utilized various natural language processing modules, including the *Natural Language Toolkit* (Loper and Bird, 2002) for tokenization and sentence segmentation*,* the *Lexical Complexity Analyzer for Academic Writing* (Nasseri and Lu, 2020) and the *lexical-diversity package* (Kyle, 2018) for calculating lexical complexity. We used the *L2 Syntactic Complexity Analyzer* (Lu, 2010) for computing syntactic complexity; this uses the Stanford Parser (Klein and Manning, 2003) to generate part-of-speech tags and parse trees, and to extract relevant syntactic units or phrases. The parser has high accuracy performance with an F measure of 89.96 (https://nlp.stanford.edu/software/srparser.html). Note that parser accuracy varies depending on the type of corpus (Gray, 2019). Speech data is particularly challenging for automatic parsers due to colloquial features such as incomplete or ungrammatical sentences, and intervening phrases (Hsiao et al., 2022; Roland et al., 2007). Our corpus comprised written language, albeit being language produced by children, and we expected high accuracy of parser performance. All scripts and resulting values of complexity measures are available at OSF (https://osf.io/wuzaf).

Results

Length of basic units

We calculated the following four count statistics by Key Stage: number of letters per word, number of words per sentence (note this is the same measure as MLS - mean length per sentence - among the syntactic complexity measures), number of words per story and number of sentences per story. We predicted that all these measures would be higher in older children's writing. As shown in Fig. 1, even though the word limit was set at 500 words per story, there was still developmental variation in length (with some pieces going over the 500-word limit). The older children wrote longer stories than younger children, with a large increase of 127 words being seen between Key Stage 1 and 2, compared to 27 words between Key Stage 2 and 3. The older children produced longer words, more sentences per story and more words per sentence and again, differences between Key Stages 1 and 2 were most obvious. Table 3 shows increases across all measures by Key Stage, confirmed by linear regression, although the difference was most evident for story length. Note that a Bonferroni correction of the p values was performed, such that the p value of 0.001 was taken as the threshold

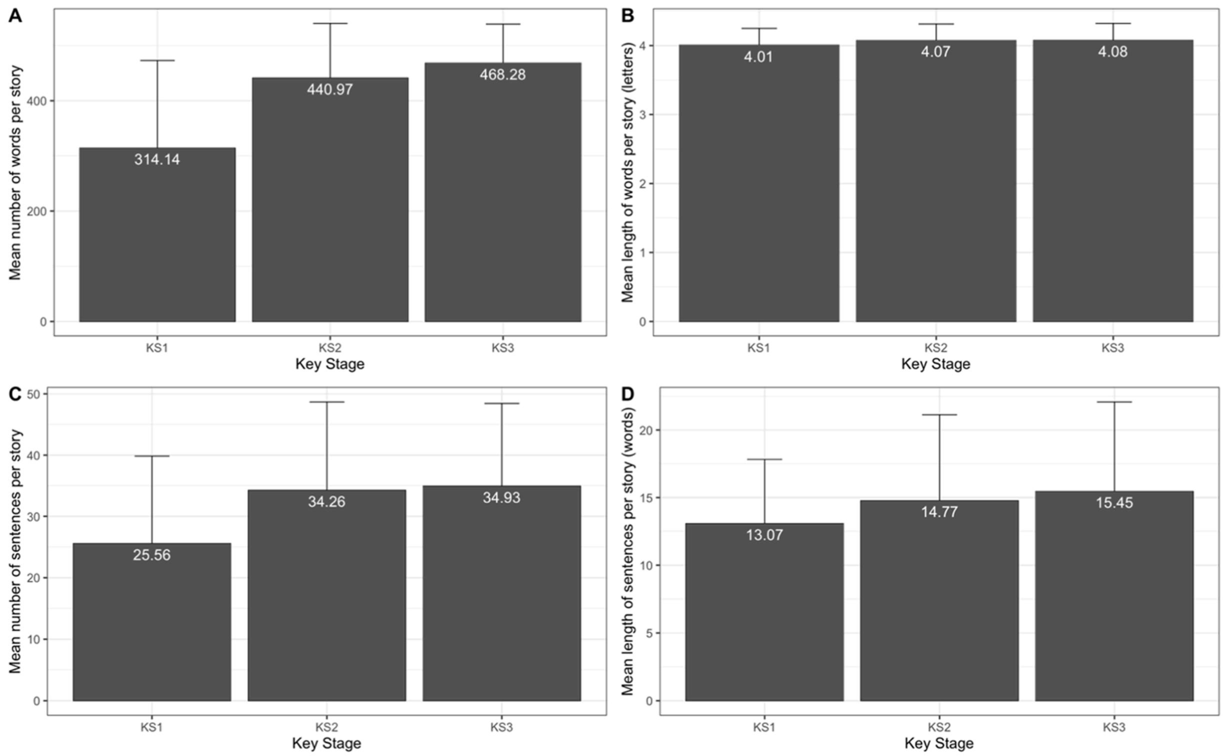

Fig. 1. Number and length of words and sentences in children's writing by Key Stage.

Table 3
Descriptive summary of length of basic units per story in children's writing by Key Stage (KS). Betas indicate the estimated increase (if positive) or decrease (if negative) by unit of the measure as a function of Key Stage. A significant effect of Key Stage (p's< 0.001 after Bonferroni correction) is marked with an asterisk. Note an effect size $\eta^2 = 0.01$ indicates a small effect, 0.06 a medium effect, and 0.14 a large effect.

Measure Mean (SD)	KS1 (N = 3625)	KS2 (N = 62,766)	KS3 (N = 38,674)	*Beta*	η^2
N words	314.14 (158.76)	440.97 (99.19)	468.28 (70.64)	43.03*	0.06
N letters	4.01 (0.24)	4.07 (0.24)	4.08 (0.25)	0.01*	0.0007
N sentences	25.56 (14.29)	34.26 (14.42)	34.93 (13.50)	1.94*	0.005
Sentence length/ words	13.07 (4.75)	14.77 (6.36)	15.45 (6.63)	0.84*	0.005

of significance considering a total of 48 models fitted in the study (0.05/48 = 0.001). Also note that the complexity measures varied in scale. This is due to the differences in how the complexity measures were calculated (e.g., count, percentage, log transformation, ratio) and this in turn is reflected in the means and standard deviations. To address the differences in scale, as well as the potential issue of overpowering given corpus size (large samples could produce statistically significant effects that are negligible) (Egbert et al., 2022), we also report the effect size measure η^2.

Measures of lexical and syntactic complexity

We computed the measures of lexical and syntactic complexity of each story according to each of the described measures in Tables 1 and 2. These two types of complexity are summarised by the child's Key Stage in Tables 4 and 5, respectively. Complexity scores were regressed on Key Stage (and age in years, reported in the supplementary analyses) as a proxy for language proficiency. As indicated by linear regression, Key Stage predicted growth in most measures, with a small number of exceptions: there was a decrease in complexity by Key Stage for three of the lexical diversity measures (verb variation by word type, vv2, $b = -0.003$, SE = 0.0004, $t = -7.09$, $p < .001$, noun variation, nv, $b = -0.007$, SE = 0.001, $t = -6.37$, $p < .0001$, and adjective variation, adjv, $b = -0.006$, SE = 0.0005, $t = -13.98$, $p < .001$), and two of the syntactic complexity measures (coordinate phrase per T-unit, CP.T, $b = -0.020$, SE = 0.001, $t = -18.6$, $p < .001$; coordinate phrase per clause, CP.C, $b = -0.015$, SE = 0.0005, $t = -29.89$, $p < .001$). When using age in years as an alternative for Key Stage, we found the same effects except two lexical complexity measures were not significant: lexical sophistication by token (ls1, $b = 0.0005$, SE = 0.0004, $t = 1.52$, $p = .13$) and verb variation 4 (vv2, $b = 0.000002$, SE = 0.0002, $t = 0.14$, $p = .89$). All syntactic complexity measures were significant and in the same direction as Key Stage, when age in years was the predictor, except for C.T (mean number of clauses per T unit), which produced a borderline p value at 0.001 (C.T, $b = 0.006$, SE = 0.002, $t = 3.26$, $p = .001$). In terms of effect size, the effect of Key Stage was stronger in lexical diversity measures than others (e.g., ndw, cttr, rttr show medium to large effects, also in the analysis using age as a predictor). Many of the variables, including the ones mentioned above with developmental decrease and insignificant growth, showed small to minimal effects. The effect sizes in syntactic complexity measures were mostly very small except for a few measures like mean length per clause (MLC) and complex nominals per clause (CN.C).

Relationships among measures of lexical and syntactic complexity

To reduce the number of dimensions and identify how the variables cluster together, 44 lexical and syntactic complexity variables were entered into a Principal Component Analysis (PCA). This type of analysis can represent variation in the entire dataset by generating new variables (i.e., principal components or dimensions) which are orthogonal to each other. The analysis was conducted in R using the FactoMineR package (Lê et al., 2008) and the results were visualised using the factoextra

Table 4
Mean (SD) score for each lexical richness measure by Key Stage. Reference numbers correspond to descriptors in Table 1. Betas indicate the estimated increase (if positive) or decrease (if negative) in units of the measure as a function of Key Stage. A significant effect of Key Stage (*p*'s< 0.001 after Bonferroni correction) is marked with an asterisk. Note an effect size $\eta^2 = 0.01$ indicates a small effect, 0.06 a medium effect, and 0.14 a large effect.

Measure Mean (SD)	KS1 (*N* = 3625)	KS2 (*N* = 62,766)	KS3 (*N* = 38,674)	*Beta*	η^2
Lexical Density					
1. ld	0.04 (0.02)	0.05 (0.02)	0.05 (0.02)	0.002*	0.004
Lexical Sophistication					
2. ls1	0.31 (0.25)	0.35 (0.18)	0.35 (0.16)	0.006*	0.0003
3. ls2	0.44 (0.06)	0.45 (0.05)	0.45 (0.06)	0.003*	0.0007
4. vs1	0.13 (0.28)	0.25 (0.34)	0.31 (0.37)	0.07*	0.01
5. vs2	0.17 (0.39)	0.37 (0.61)	0.49 (0.71)	0.13*	0.01
6. cvs1	0.13 (0.26)	0.27 (0.34)	0.33 (0.37)	0.07*	0.01
Lexical Diversity					
7. ndw	140.07 (61.71)	206.14 (45.69)	225.37 (36.13)	26.63*	0.10
8. ndwz	35.76 (3.84)	38.41 (3.28)	39.10 (3.11)	0.996*	0.03
9. ndwerz	37.35 (3.09)	39.89 (1.92)	40.40 (1.65)	0.83*	0.05
10. ndwesz	35.89 (3.14)	38.45 (2.17)	39.19 (1.85)	1.03*	0.06
11. ttr	0.48 (0.09)	0.47 (0.06)	0.48 (0.05)	0.008*	0.006
12. msttr	0.72 (0.06)	0.77 (0.04)	0.78 (0.03)	0.02*	0.07
13. cttr	5.53 (1.17)	6.92 (0.95)	7.36 (0.82)	0.59*	0.11
14. rttr	7.82 (1.66)	9.78 (1.34)	10.40 (1.16)	0.83*	0.11
15. logttr	0.87 (0.02)	0.88 (0.02)	0.88 (0.02)	0.006*	0.03
16. uber	18.45 (3.00)	21.51 (3.07)	22.78 (3.07)	1.56*	0.07
17. MATTR	0.70 (0.06)	0.75 (0.04)	0.76 (0.03)	0.02*	0.07
18. HDD	0.75 (0.09)	0.80 (0.04)	0.81 (0.03)	0.16*	0.05
19. MTLD	45.00 (13.29)	59.54 (16.18)	65.70 (16.71)	7.49*	0.06
20. MTLD_wrap	44.94 (13.21)	60.08 (16.27)	66.36 (16.82)	7.68*	0.06
21. MTLD_bi	42.30 (13.69)	58.04 (16.23)	64.46 (16.56)	7.89*	0.06
22. lv	0.70 (0.23)	0.71 (0.16)	0.72 (0.14)	0.01*	0.001
23. vv1	0.44 (0.42)	0.59 (0.40)	0.64 (0.40)	0.06*	0.007
24. svv1	0.65 (0.75)	1.09 (1.00)	1.23 (1.08)	0.18*	0.009
25. cvv1	0.43 (0.38)	0.62 (0.41)	0.66 (0.42)	0.07*	0.008
26. vv2	0.08 (0.11)	0.07 (0.07)	0.07 (0.06)	-0.003*	0.0005
27. nv	0.70 (0.31)	0.72 (0.20)	0.71 (0.19)	-0.007*	0.0004
28. adjv	0.10 (0.14)	0.09 (0.07)	0.08 (0.06)	-0.006*	0.002
29. advv	0.20 (0.18)	0.25 (0.13)	0.27 (0.13)	0.03*	0.01
30. modv	0.30 (0.21)	0.34 (0.14)	0.35 (0.13)	0.02*	0.007

Table 5
Mean (SD) score for each syntactic complexity measure by Key Stage. Reference numbers correspond to descriptors in Table 2. Betas indicate the increase (if positive) or decrease (if negative) in units of the measure as a function of Key Stage. All betas showed a significant effect of Key Stage (*p*'s< 0.001 after Bonferroni correction). Note an effect size $\eta^2 = 0.01$ indicates a small effect, 0.06 a medium effect, and 0.14 a large effect.

Measure	KS1 (*N* = 3625)	KS2 (*N* = 62,766)	KS3 (*N* = 38,674)	*Beta*	η^2
Length of Production Unit					
1. MLS	13.38 (4.81)	15.36 (6.70)	15.90 (6.90)	0.77*	0.004
2. MLT	11.05 (3.86)	12.75 (4.73)	13.04 (4.64)	0.51*	0.003
3. MLC	7.14 (1.19)	7.26 (1.13)	7.52 (1.21)	0.24*	0.01
Complexity Ratio					
4. C.S	1.88 (0.65)	2.13 (0.93)	2.13 (0.94)	0.04*	0.0006
5. VP.T	1.84 (0.65)	2.14 (0.80)	2.17 (0.76)	0.07*	0.002
6. C.T	1.56 (0.57)	1.77 (0.65)	1.74 (0.61)	0.02*	0.0002
7. DC.C	0.27 (0.12)	0.33 (0.11)	0.33 (0.11)	0.01*	0.003
8. DC.T	0.46 (0.45)	0.63 (0.48)	0.62 (0.45)	0.02*	0.0006
9. T.S	1.21 (0.22)	1.19 (0.19)	1.20 (0.21)	0.008*	0.0004
10. CT.T	0.34 (0.17)	0.40 (0.16)	0.41 (0.16)	0.01*	0.002
11. CP.T	0.32 (0.21)	0.31 (0.19)	0.29 (0.18)	-0.02*	0.003
12. CP.C	0.21 (0.12)	0.18 (0.09)	0.16 (0.08)	-0.02*	0.008
13. CN.T	0.88 (0.61)	1.16 (0.62)	1.20 (0.59)	0.08*	0.005
14. CN.C	0.55 (0.20)	0.64 (0.18)	0.68 (0.20)	0.05*	0.02

package (Kassambara and Mundt, 2020). The scree plot (Fig. 2) shows the amount of variance explained by the top 10 components/dimensions, each with eigenvalues over 1. Together, these ten dimensions explained 83 % of variance. The scree plot shows a strong decrease and then a plateauing in amount of variance accounted for after the 3rd dimension. With the first two dimensions accounting for nearly 50 % of the variance, we therefore focus our discussion on the first two dimensions below, with some description on the 3rd dimension.

We correlated the 44 linguistic complexity variables with the first three dimensions, using coordinates of the variables on the principal components. Fig. 3 displays these correlations with blue squares indicating a positive relationship, red squares a negative relationship and the depth of colour indicating the strength of the relationship. Dimension 1 accounted for 29 % of the variance and was correlated most with lexical complexity, in particular those measures of lexical diversity devised to combat sensitivity to text length (i.e., uber, MTLD-wrap, MATTR, MTLD-bi, MTLD, cttr, rttr, msttr, $r = 0.86{\sim}.90$). The number of unique word types also showed a strong positive correlation with Dimension 1 (i.e., ndwesz, ndwerz, ndw, $r = 0.72{\sim}.82$). In contrast, the simple type-token ratio (i.e., ttr), correlated less well with this component, $r = 0.57$.

The second dimension explained 18 % of the variance and was associated with measures of syntactic complexity, particularly those measures that used T-unit as the production unit, for example, mean length of T-unit (MLT, $r = 0.92$), and complexity ratios based on T-units (C.T, VP.T, DC.T, C.T, $r = 0.85{\sim}.91$). Other measures that used sentence as the base unit were also highly correlated with Dimension 2 (i.e., MLS, C.S, with $r = 0.87$ and 0.80, respectively).

The third dimension explained around 10 % of the variance. This

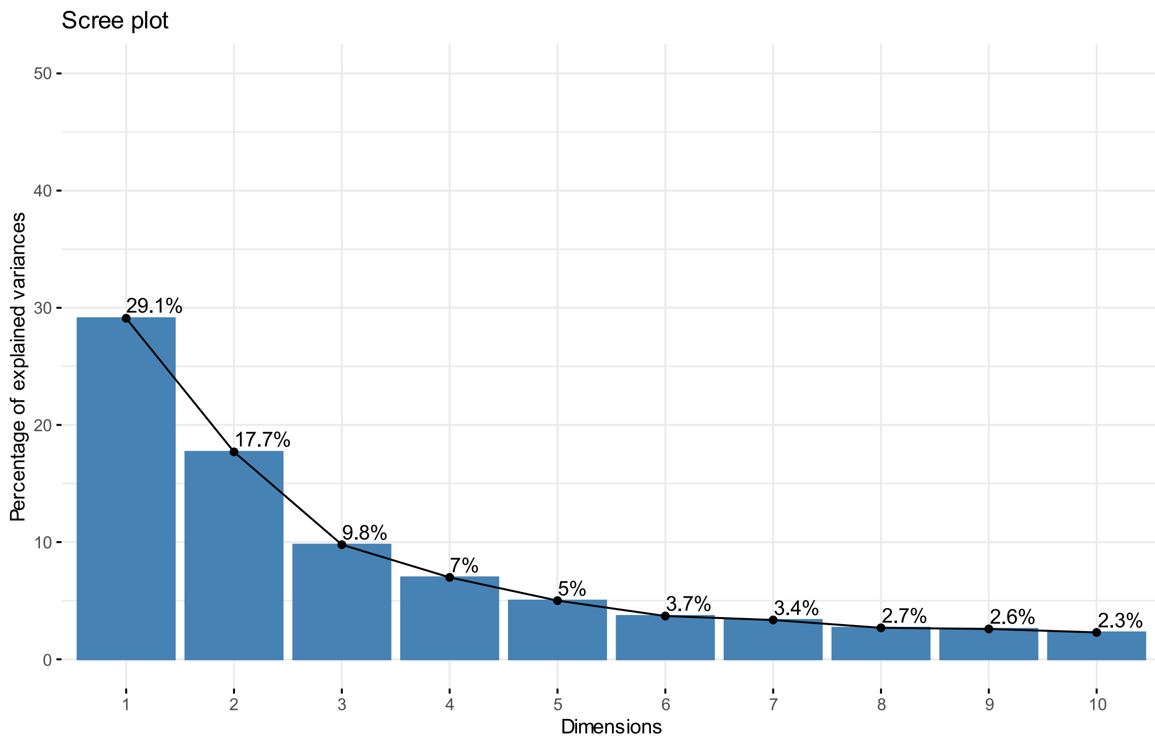

Fig. 2. Scree plot of the top 10 dimensions of the PCA on lexical and syntactic complexity measures of children's writing.

correlated positively with lexical sophistication, and with lexical diversity and sophistication measures relating to verbs (i.e., cvv1, cvs1, svv1, vs2, vv2, vs1, vv1, $r = 0.67\sim.79$).

We then examined the first two principal components in detail, using $\cos^2$ (or squared cosine or squared coordinates; $\cos^2$ is equal to the square of the values presented in Fig. 3). A high $\cos^2$ indicates a good representation of the variable on the principal component, and in turn, it shows the importance of a principal component for a given variable. For any given variable, the sum of the $\cos^2$ across all the principal components is equal to one. Fig. 4 visualises the $\cos^2$ of each individual measure of complexity on two dimensions, corresponding to the first two components. If a variable is perfectly represented by the two dimensions, the arrow will fall on the circumference of the circle. The longer the arrow (i.e., the closer the arrow to the circumference of the circle, as opposed to the centre of the circle), the higher the quality of that variable's representation. The same information is also conveyed by colour, with warmer or redder colours indicating higher quality. As can be seen from Fig. 4, the first dimension separates lexical complexity from syntactic complexity. The higher-quality variables on this dimension are those that represent lexical diversity, particularly those that reduce dependence on text length (e.g. uber, MATTR, MTLD_wrap, MTLD_bi, MTLD), followed by number of unique word types (e.g., ndwesz, ndwerz, ndw). The second dimension reflects linguistic complexity overall, with syntactic complexity being well represented in this dimension, especially measures with T-unit as the base unit (e.g. MLT, C.T, VP.T, CN.T, DC.T) and those involving sentence as the unit (e.g. MLS, C.S). Patterns around these two dimensions align with the patterns of correlations reported in Fig. 3. Fig. 4 also shows that almost all lexical complexity measures showed varying degrees of representation in both dimensions, whereas most of the syntactic complexity measures had positive representation in the second dimension and negative representation for the first (with the exception of certain clause-based measures, i.e., MLC, CN.C).

As a first step towards identifying measures that might usefully assess children's writing quality, we examined the amount of variance, or the percentage of contribution, along the two dimensions captured by the top 10 variables. The percentage was calculated by dividing the $\cos^2$ value of a complexity measure by the sum of $\cos^2$ of the dimension. As shown in Fig. 5, the first dimension was explained by lexical diversity measures that adjusted for text length. The second dimension was captured most by syntactic complexity, particularly those measures that used T-unit as the base unit. It is also notable that the percentage of variance described by the top variables was higher for the second dimension than the first dimension.

Finally, to assess the quality of representation by developmental stage, we investigated how individual data points (each representing an individual child's story) clustered on the first two dimensions as a function of the child's Key Stage. Fig. 6 shows that green data points (representing stories written by Key Stage 1 children) cluster at the lower left quadrant, suggesting that younger children's writing was poorly represented by both dimensions. Younger children's narrative writing was less lexically diverse and syntactically complex than older children's writing. In contrast, the purple dots representing stories written by Key Stage 3 children were visible at the upper right quadrant, indicating higher quality of representation along both dimensions. Overall, there was more variation in the distribution of data points in the upper left quadrant, which represents the dimension accounted for by syntactic complexity, compared to other quadrants. This suggests the presence of higher degree of individual differences in syntactic complexity compared to lexical complexity, across Key Stages. Some children may have more advanced skills in constructing complex sentences than other children of similar age, whereas most children seem to reach the milestone of producing increasingly diverse vocabulary at similar pace. In other words, the ability to construct complex sentences may be less uniform across children of different ages compared to being able to use a diverse set of lexical items.

Table 6 shows the excerpts from stories written by children in Key Stage 2 to illustrate the differences in levels of lexical diversity and syntactic complexity, even among children in the same Key stage. The stories were selected based on the two metrics each contributed the most to the first two dimensions: uber and MLT. The story represented highly in both lexical diversity and syntactic complexity contains more varied vocabulary and more dependent or subordinate clauses. On the other hand, the story low in the two metrics is composed of repeated and

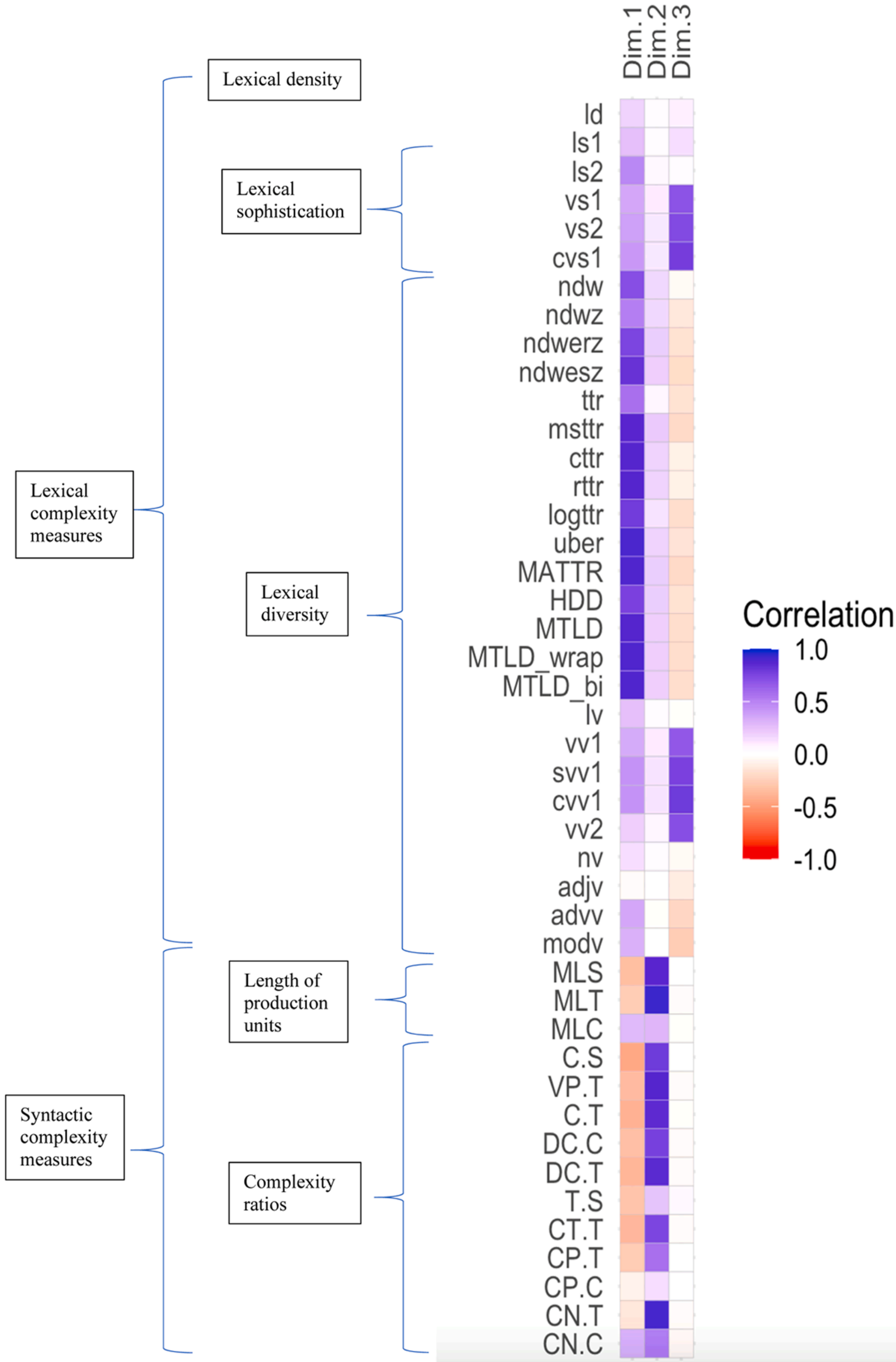

Fig. 3. Pattern of correlations between each variable and the three strongest dimensions produced by the PCA.

sometimes misspelled words, as well as short sentences without embedded clauses. It can also be noticed that stories low in syntactic complexity contain interactive dialogues between characters. Speech is usually shorter and less lexically and syntactically rich than written narratives (Dawson et al., 2021; Montag et al., 2015; Hsiao et al., 2022). Therefore, stories that include more dialogues could result in lower syntactic complexity scores compared to those that contain mostly description of events.

Discussion

Our aim was to quantify the nature and content of children's writing through mid-childhood by analysing a large cross-sectional sample of stories written by 5–13-year-olds. Previous investigations of children's writing development in first language have tended to be small in scale and either looked in detail at a restricted set of linguistic features, or they have focused within a more restricted age range. In contrast, our approach was to consider a range of language features across a broad age range and within a very large sample. This computational approach was informed by studies of second language learning (e.g., Lu, 2010, 2012).

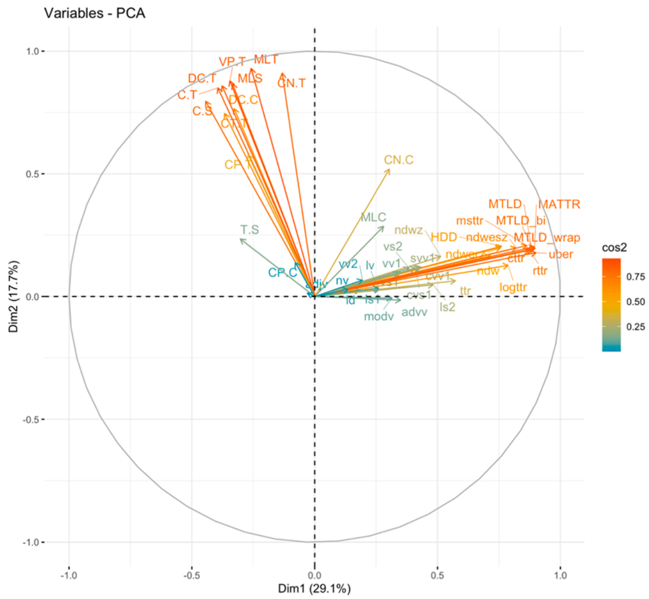

Fig. 4. Two-dimensional visualisation of the quality of representation of each complexity measure on the first two principal components, using $\cos^2$.

There are some fundamental differences between the current study and most L2 studies. Not only are there inherent differences between first language and second language acquisition, but our developmental window also comprised 9 years spanning mid-childhood through a cross-sectional dataset, whereas studies of L2 tend to chart longitudinal changes in proficiency over a short window, often associated with a short-term intensive language training across a semester in college-age students. Given these differences, we prioritise the discussion of our findings in the context of first language writing development, with a secondary focus on the insights provided by L2 research. Our study also brings an opportunity to consider how our quantitative at-scale findings complement those derived from different methodological approaches, and to set out directions for future work.

We begin by summarising our key findings on the emergence of lexical and syntactic complexity in children's writing. Growth was seen across a large set of lexical and syntactic features. If we accept that these features mark writing quality, their growth provides evidence that writing becomes increasingly complex and higher in quality with age. While this might not appear surprising, some of these features have not always been associated with changes in age or proficiency in earlier work. Take for example word length. Olinghous and Leaird (2009) found a difference between Grade 2 and Grade 4 children's narrative writing (around 100 children in each group) in word length, but only in an experimental task and not in a standardized writing assessment. They also found a correlation between word length and subjective writing quality in Grade 2 but not Grade 4 children, and only for the standardized test. We found clear evidence for a developmental effect of length: younger children produced words that were on average shorter than those produced by older children. This was a small effect, however, and this may explain the mixed findings in the literature on the basis that age differences only become significant when the sample size is large. Further work is needed to better understand the underlying variables (and their interactions) that are responsible for increases in word length as children's writing develops. As noted by Durrant et al. (2021), word length is conceptually complex. Longer words tend to be lower in frequency, more abstract and are more likely to be morphologically complex than shorter words. These observations indicate that increases in word length by age reflect increases in lexical sophistication.

The previous literature also provides mixed findings on the development of lexical density in children's writing (Berman and Nir, 2010; Uccelli et al., 2012) and in L2 writing (Engber, 1995; Lu, 2012). We observed significant growth in lexical density by Key Stage, perhaps reflecting both the size of our sample and its broader age range, relative to previous work. Once again, it is important to note that the effect size is small, and to ask what lexical density is measuring. Lexical density is typically considered a marker of lexical richness, but it might also reflect syntactic competence (Durrant et al., 2021). In our PCA, however, lexical density clustered more closely with other measures of lexical richness. It did not feature on the principal component most associated with syntactic complexity, unlike some of the other measures of lexical complexity.

Lexical sophistication was measured using five measures designed to capture the proportion of advanced words, with advanced defined in terms of frequency. Three of these measures focused on verbs and these all showed significant growth with Key Stage, as did overall word type; the younger children obtained lower scores for lexical sophistication by token, but this was a small effect. This insensitivity was confirmed by the regression model using age in years as the predictor. This finding corroborates previous studies that failed to find a consistent age effect (Durrant and Brenchley, 2019; Olinghouse and Leaird, 2009). Furthermore, Durrant and Brenchley (2019)'s analyses demonstrated that lower frequency words were used by children when writing literary text compared to non-literary text. They also observed a decrease of mean frequency by age in verbs and adjectives (i.e., indicating increased lexical sophistication) but an increase of frequency by age for nouns (i.e., decreased lexical sophistication). These effects were only present when analysing by word token, not by word type. Durrant and Brenchley suggested that when children are engaged in literary writing, their stories are characterised by fairy tale and fantastical themes, and that this encourages the use of nouns that are rare in adult writing. This is likely to be the case in our study too. Moreover, younger children in Durrant and Brenchley's study tended to repeat these nouns in their writing. On the other hand, older children produced more nouns considered to be academic words, and without as many repetitions as do their younger counterparts. Taken together, these findings suggest that while lexical sophistication is associated with growth in children's writing, frequency itself might not be the critical variable, but instead it is a by-product of broader linguistic influences and perhaps not really separable from lexical diversity (Durrant and Brenchley, 2019; Durrant and Durrant, 2022). In support of this, our measures of lexical sophistication clustered together and loaded on the first principal component, together with the lexical diversity measures. They were also strongly represented on the third dimension, this time alongside lexical diversity measures that tapped verb usage.

Lexical diversity can be captured in various ways and reflecting this, we included 24 different measures. Generally, these clustered together, and stories written by older children showed greater levels of lexical diversity than those written by younger children. These findings mirror those already reported in the literature (Berman and Verhoeven, 2002; Durrant and Brenchley, 2019; Malvern et al., 2004; Wagner et al., 2011). Most of the measures, especially those designed to adjust for text length rather than simple type-token ratio, showed medium-to-large effect sizes. Three of the 24 variables showed a decline with Key Stage (vv2: verb variation by word type, nv: noun variation and adjv: adjective variation; note the non-significance found for vv2 when using age as a predictor). The decline does not necessarily reflect a decrease in the repertoire of verbs, nouns and adjectives as children grow, but could potentially be a consequence of older children writing longer texts – a typical issue of type-token ratio calculations with text length. It would also be due to the levelling off of the growth in the vocabulary in these parts of speech. Lu (2012) found no effects of proficiency on lexical variation in verbs (vv2), nouns (nv), and adjectives (adjv) in L2 oral narratives, suggesting we need to take caution when interpreting the negative association with age in our results, as evidenced by the small effect sizes. It could be also the consequence of word usage difference between younger and older children: nouns related to fairy tales and

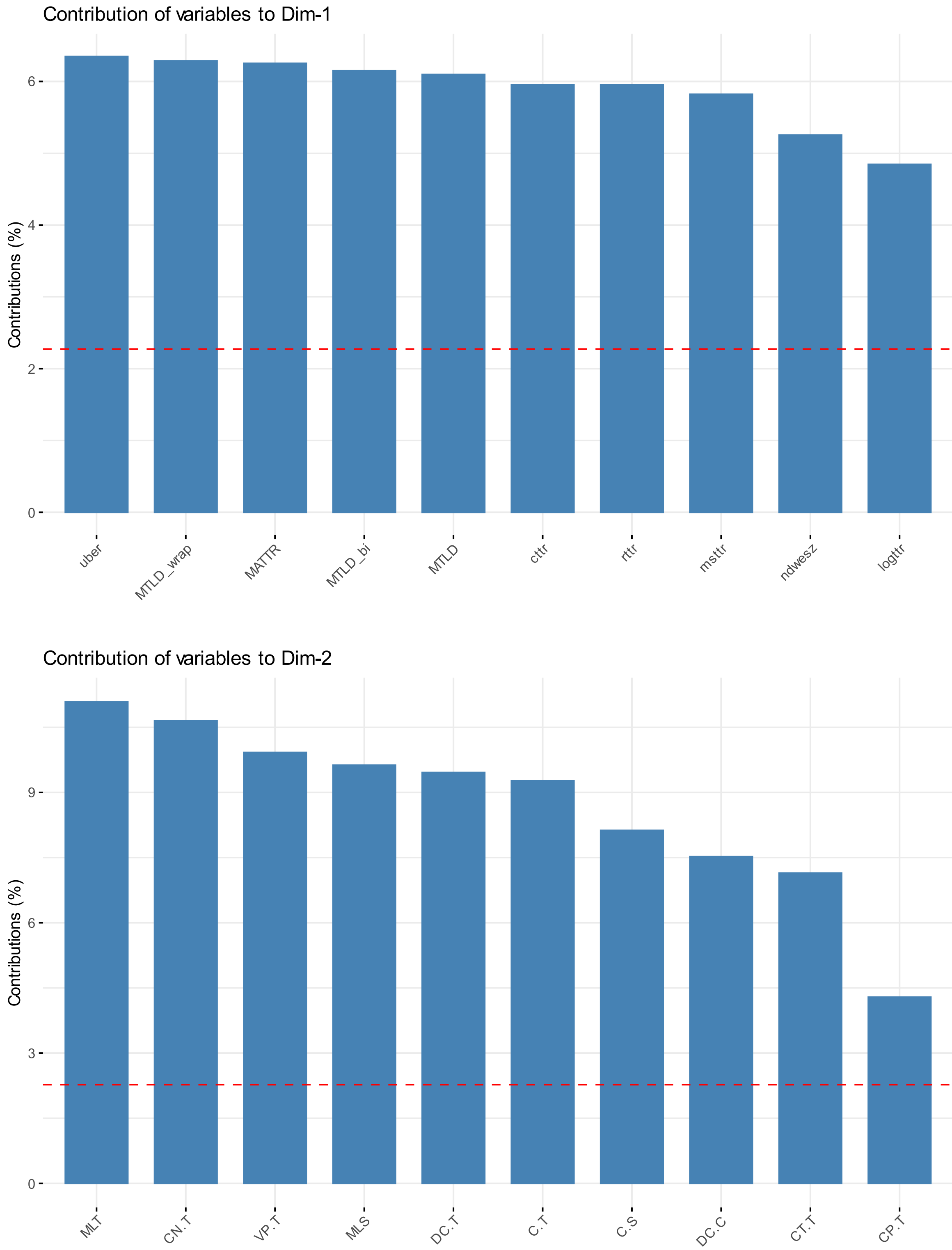

Fig. 5. Percentage of contribution by the top 10 measures of complexity in the first dimension (top panel) and the second dimension (lower panel). The red dotted line shows expected average contribution, at 2.2 % (100 % divided by the total number of features measured, $N = 44$).

fantasies dominate young children's writing, which later cease to be represented as children grow – a point mentioned earlier related to lexical sophistication (Durrant and Brenchley, 2019; Durrant and Durrant, 2022). Regarding the PCA results, the lexical diversity measures were strongly represented on the first dimension, especially those calculated with methods designed to reduce the confounding effect of text length (Malvern et al., 2004). Our findings support previous L2 research in showing that certain lexical complexity measures are resistant to length variation (e.g. MATTR, MTLD, MTLD_wrap), and that therefore these should be preferred when measuring lexical diversity (Zenker and Kyle, 2021). We found these same measures had higher quality representation in and made a higher contribution to the first principal component. As noted above, diversity measures tapping variation in verb usage were also represented on the third component, along

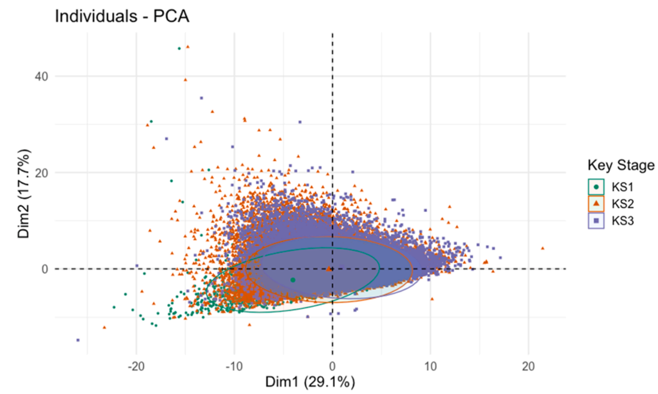

Fig. 6. The quality of representation of each child's story along dimensions 1 and 2 as a function of child's Key Stage. The circles indicate concentration ellipses for each Key Stage.

Table 6
Excerpts from stories written by children from Key Stage 2, as examples to demonstrate the differing levels of representation in Dimension 1 and Dimension 2. Statistics of uber and MLT of each story, not the excerpt, are presented in the parentheses. Square brackets mark the boundaries of dependent clauses. Underlines indicate repetitions of words.

Level of representation		Dimension 2 (Mean MLT of KS2 stories: 12.75)	
		High	Low
Dimension 1 (mean uber of KS2 stories: 21.51)	**High**	Their mother [who worked for the local hospital as a nurse] was very protective over her children and had always tried to hide them from any wardens [who attempted to seek out the children to be evacuated]. But [when father came back from the war injured with a serious fracture in his leg] they decided to escape to the countryside. (uber: 24.94, MLT: 18.59)	There was a place [that could fix me]. The man there was called Ben and he was very friendly. He chatted to Lucy for a while. About ten minutes later Lucy came out with an extremely sulky face. "I can't believe [they can't fix you]! They are useless!" She looked very miserable. What is she doing?? This is awful! (uber: 25.47, MLT: 7.76)
	Low	[As we were walking through the forest] the parents decided to make us split up me and Chloe went together on one quad. We spent ages trying to find out [were the ponies were]. We were starting to get a little bit worried [until we heard some music]. We followed the sound and it was getting louder and louder [until we found a cave [where the sound was coming from]]. (uber: 19.96, MLT: 18.34)	Ding dong. "Who will that be? Its a parcle. Its from Kevin the prince. Wow there shoes. Lets put them on." "I got the ring and Im reddy to go. Hears the ring." "Wow" The ring turned the prince into a…rabbit! "Oh dear this ring is not a real ring" She shouted. "What is it then"? "Its growing" The ring was actually a gummy bare. (uber: 19.32 MLT: 7.32)

with measures of lexical sophistication.

Regarding complexity at the sentence level, older children produced more sentences than younger children, and they used longer sentences too. Longer sentences are traditionally associated with language growth as they allow complex ideas to be communicated (Bear, 1939; Golub and Frederick, 1970; Hunt, 1965; Myhill, 2008). We also found increases in the production of T-units and clauses as a function of Key Stage, replicating previous work (Golub and Frederick, 1970; Hunt, 1965; Peltz, 1973; Rubin and Piché, 1979; Wagner et al., 2011). Syntactic complexity was mainly represented by the second dimension in the PCA., Despite small effect sizes, developmental growth was seen across most measures except for those tapping the use of coordination phrases (CP.T and CP.C). Other studies have found a non-positive correlation between age and coordination usage (Golub and Frederick, 1970; Hunt, 1965; Peltz, 1973), perhaps because it emerges relatively early as a sentence combining operation, but is then replaced by other types of complex grammar later on. As suggested by others (Bardovi-Harlig, 1992; Norris and Ortega, 2009; Wolfe-Quintero et al., 1998), syntactic complexity develops over stages from independent, uncoordinated utterances (stage 0), to utterances then linked by coordination (stage 1), then by subordination (stage 2), and finally by a noun phrase (stage 3), as illustrated by Kiuken and Vedder (2019):

Stage 0: *I have a son. He is 12 years old.* (2 T-units, average length 4.5 words)
Stage 1: *I have a son and he is 12 years old.* (2 T-units, average length 4.5 words)
Stage 2: *I have a son who is 12 years old.* (1 T-units, average length 9 words)
Stage 3: *My 12-year-old son.* (1 T-units, average length 3 words)

Verhoeven et al. (Verhoeven et al., 2002) found that children's writing used more coordination whereas adults used more subordination. They also found effects of genre, with narrative writing containing more coordination and expository writing containing more subordination. Subordination is associated with increases in T-unit length, and this is generally considered an indication of syntactic maturation. However, using complex nominals (like the stage 3 example above), serves to reduce T-unit length. It is therefore important to consider complexity ratios (e.g., clause per T-unit, dependent clauses per T-unit, complex nominals per T-unit). Our results showed that these complexity ratios (i.e., C.T, DC.T, CN.T) were strongly represented on the second principal component. This confirms both the utility of T-unit as a marker of children's writing development (Gaies, 1980; Hunt, 1965), and density ratios as a marker of syntactic sophistication (Hunt, 1965; Lei and A., 2012; Rubin and Piché, 1979; Verhoeven et al., 2002; Wagner et al., 2011). However, recent research has questioned the sensitivity of T-units and related complexity ratio measures in capturing the types and functions of diffrent syntactic structures (Biber et al., 2020; Norris and Ortega, 2009). For example, it is not entirely clear how changes in T-unit length reflect usage of different types of dependent clauses (e.g., relative clauses, like the Stage 2 example above) and dependent phrases (e.g., complex nominals, like the Stage 3 example above), even though both could contain long and complex embeddings. Researchers (e.g., Biber et al., 2020) have advocated against these omnibus measures and instead for measures that capture the use of specific structural types and syntactic features in the grammatical system (e.g., within finite dependent clauses, distinguish among adverbial clauses, complement clauses and noun modifier clauses).

It is interesting to consider whether the two major dimensions indicated by our PCA demonstrate that lexical vs. syntactic complexity account for distinct portions of variance in children's writing. Lexical knowledge and syntactic development are closely associated in infancy (Bates and Goodman, 1997; Devescovi et al., 2005; Moyle et al., 2007). Children learn a word by knowing not only its meaning and how to pronounce it, but also where to place it in a sentence and how it changes in form depending on other constraints. We also saw an interrelationship between lexical and syntactic complexity in our analyses, as shown in Figs. 3 and 4. Although the first dimension was primarily lexical and the second dimension primarily syntactic, all types of complexity measures were positively correlated with Dimension 2. However, there is some evidence to suggest different trajectories across syntactic and lexical development (Huttenlocher et al., 2010), with syntactic development generally being slower (Fisher et al., 1994; Gleitman, 1990). Consistent with this, in our analysis, children across all Key Stages exhibited wider variation in syntactic complexity. This might reflect that syntactic knowledge matures or is expressed later in time compared to lexical knowledge. This could also potentially explain why syntactic complexity was associated with smaller effect sizes compared to lexical diversity, with the former being more subtle and requiring larger samples to detect. However, the variation in syntactic complexity could also result from the narrative style that individual children chose to adopt. As shown in Table 6, texts containing more dialogues scored lower in syntactic complexity as speech is often short and less grammatically complex (Hsiao et al., 2022; Montag, 2019).

Our study has several strengths. The stories were not written for this investigation, nor was their content prompted by pictures or other experimental instructions. Our naturally occurring dataset is massive, and the stories were written by children across a wide age range. These features permit a comprehensive overview of linguistic complexity as it emerges in children's writing across the mid-childhood period. It is also important to recognise the limitations of our approach. Writing proficiency and writing quality are clearly multidimensional. Notwithstanding the 44 measures of lexical and syntactic complexity used in our study, other factors are also important (Crossley et al., 2014) including domain knowledge (Kellogg, 2006) and the ability to write towards different discourses, as appropriate for register and usage (McCormick et al., 1992). Cohesion and coherence are also important. Children use cohesive devices, such as connectives (e.g., *because, so*) and referential pronouns to create links and flow in the writing (McNamara et al., 2014); these are not necessarily captured by automated measures of complexity. Although we used measures that have been validated against human judgement in previous studies (e.g., Lu, 2010, 2012), not all features of writing quality can be easily quantified using a computational, data-driven approach. Consider, for example, the organisation of ideas, adaptation to different audience, and the coherence and rhetoric of the writing (Ferris, 1994). Valid and reliable assessment of these aspects of writing quality requires judgement by raters who are trained and committed to careful evaluation and hand-coding, or advanced machine learning algorithms that capture the underlying topical or discoursal structure. Future work could look to complement our 'bottom-up' approach with qualitative assessments by humans or artificial agents that build from global and 'top-down' criteria. We should also note that our analyses are restricted to the narrative domain and therefore our findings might not be generalisable to expository writing or argumentative essays.

A different type of limitation is that our data are cross-sectional. We examined writing samples from over 100,000 5–13-year-olds. This provides a perspective on how aspects of writing change with age, but without additional data about the children (e.g., language background, home literacy environment, cognitive ability and literacy skill), it is impossible to know which factors beyond age are associated with individual and developmental differences in writing ability. Longitudinal data are particularly important in helping to trace complex relationships and how they unfold over time. Unfortunately, however, there is a dearth of large-scale studies that examine the development of writing in the elementary school years in a longitudinal context (Loban, 1976). This should be a priority for future work. In the meantime, however, our study provides an attempt to understand children's writing development from the perspective of lexical and syntactic complexity, using a large-scale sample of narrative writing produced by young children.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.acorp.2024.100084.

References

Bardovi-Harlig, K., 1992. A Second Look at T-Unit Analysis: reconsidering the Sentence. TESOL Quarterly 26 (2), 390–395. https://doi.org/10.2307/3587016.

Bates, E., Goodman, J.C., 1997. On the inseparability of grammar and the lexicon: evidence from acquisition, aphasia, and real-time processing. Lang. Cogn. Process 12 (5–6), 507–584. https://doi.org/10.1080/016909697386628.

Bear, M.V., 1939. Children's growth in the use of written language. Elementary English Rev. 16 (8), 312–319.

Beers, S.F., Nagy, W.E., 2009. Syntactic complexity as a predictor of adolescent writing quality: which measures? Which genre? Read Writ. 22 (2), 185–200. https://doi.org/10.1007/s11145-007-9107-5.

Berman, R.A., Nir, B., 2010. The lexicon in writing–speech-differentiation. Written Lang. Literacy 13 (2), 183–205. https://doi.org/10.1075/wll.13.2.01ber.

Berman, R.A., Verhoeven, L., 2002. Cross-linguistic perspectives on the development of text-production abilities. Written Lang. Literacy 5 (1), 1–43. https://doi.org/10.1075/wll.5.1.02ber.

Biber, D., 1988. Variation Across Speech and Writing. Cambridge University Press. https://doi.org/10.1017/CBO9780511621024.

Biber, D., Gray, B., Staples, S., Egbert, J., 2020. Investigating grammatical complexity in L2 English writing research: linguistic description versus predictive measurement. J. Eng. Acad. Purposes 46. https://doi.org/10.1016/j.jeap.2020.100869.

Brysbaert, M., Biemiller, A., 2017. Test-based age-of-acquisition norms for 44 thousand English word meanings. Behav. Res. Methods 49 (4), 1520–1523. https://doi.org/10.3758/s13428-016-0811-4.

Brysbaert, M., New, B., 2009. Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behav. Res. Methods 41 (4), 977–990. https://doi.org/10.3758/BRM.41.4.977.

Chenoweth, N.A., Hayes, J.R., 2001. Fluency in writing. Written Commun. 18, 80–98. https://doi.org/10.1177/0741088301018001004.

Crossley, S.A., 2020. Linguistic features in writing quality and development: an overview. J. Writ. Res. 11 (3), 415–443. https://doi.org/10.17239/JOWR-2020.11.03.01.

Crossley, S.A., Roscoe, R., McNamara, D.S., 2014. What is successful writing? An investigation into the multiple ways writers can write successful essays. Written Commun. 31 (2), 184–214. https://doi.org/10.1177/0741088314526354.

Crossley, S.A., Weston, J.L., McLain Sullivan, S.T., McNamara, D.S., 2011. The Development of Writing Proficiency as a Function of Grade Level: a Linguistic Analysis. Written Commun. 28 (3), 282–311. https://doi.org/10.1177/0741088311410188.

Crosson, A.C., Lesaux, N.K., Martiniello, M., 2008. Factors that influence comprehension of connectives among language minority children from Spanish-speaking backgrounds. Appl. Psycholinguist 29 (4), 603–625. https://doi.org/10.1017/S0142716408080260.

Dawson, N., Hsiao, Y., Tan, A.W.M., Banerji, N., Nation, K., 2021. Features of lexical richness in children's books: comparisons with child-directed speech. Lang. Develop. Res. 2021, 1–45. https://doi.org/10.34842/5we1-yk94.

Deane, P., Quinlan, T., 2010. What automated analyses of corpora can tell us about students' writing skills. J. Writ. Res. 2 (2), 151–177. https://doi.org/10.17239/jowr-2010.02.02.4.

Deno, S., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., Jenkins, J., 1982b. The Use of Standard Tasks to Measure Achievement in Reading, Spelling, and Written Expression: A Normative and Developmental Study.

Devescovi, A., Caselli, M.C., Marchione, D., Pasqualetti, P., Reilly, J., Bates, E., 2005. A crosslinguistic study of the relationship between grammar and lexical development. J. Child Lang. 32 (4), 759–786. https://doi.org/10.1017/S0305000905007105. IssueCambridge University Press.

Durrant, P., Brenchley, M., 2019. Development of vocabulary sophistication across genres in English children's writing. Read Writ. 32 (8), 1927–1953. https://doi.org/10.1007/s11145-018-9932-8.

Durrant, P., Brenchley, M., 2022. Development of noun phrase complexity across genres in children's writing. Appl. Linguist. https://doi.org/10.1093/applin/amac032.

Durrant, P., Brenchley, M., McCallum, L., 2021. Understanding Development and Proficiency in Writing: Quantitative Corpus Linguistic Approaches. Cambridge University Press. https://doi.org/10.1017/9781108770101.

Durrant, P., Clarkson, R., Brenchley, M., 2020. Syntactic development across genres in children's writing:the case of adverbial clauses. J. Writ. Res. 12 (2) https://doi.org/10.17239/jowr-2020.12.02.04. Issue.

Durrant, P., Durrant, A., 2022. Appropriateness as an aspect of lexical richness: what do quantitative measures tell us about children's writing? Assessing Writing 51. https://doi.org/10.1016/j.asw.2021.100596.

Egbert, J., Biber, D., Gray, B., 2022. Distribution considerations. Designing and Evaluating Language Corpora. Cambridge University Press, pp. 122–155. https://doi.org/10.1017/9781316584880.005.

Engber, C.A., 1995. The relationship of lexical proficiency to the quality of ESL compositions. J. Second Language Writing 4 (2), 139–155. https://doi.org/10.1016/1060-3743(95)90004-7.

Ferris, D.R., 1994. Rhetorical Strategies in Student Persuasive Writing: differences between Native and Non-Native English Speakers. Res. Teach. English 28 (1), 45–65.
Finn, P.J., 1977. Computer-aided description of mature word choices in writing. In: Cooper, C.R., ee Odell, L. (Eds.), Evaluating writing: Describing, measuring, Judging. National Council of Teachers of English, pp. 69–89.
Fisher, C., Hall, D.G., Rakowitz, S., Gleitman, L., 1994. When it is better to receive than to give: syntactic and conceptual constraints on vocabulary growth. Lingua 92, 333–375. https://doi.org/10.1016/0024-3841(94)90346-8.
Gaies, S.J., 1980. T-unit analysis in second language research: applications, problems and limitations. TESOL Quarterly 14 (1), 53–60. https://doi.org/10.2307/3586808.
Gleitman, L., 1990. The structural sources of verb meanings. Lang Acquis 1 (1), 3–55. https://doi.org/10.1207/s15327817la0101_2.
Goldstone, R.L., Lupyan, G., 2016. Discovering psychological principles by mining naturally occurring data sets. Top Cogn. Sci. 8 (3), 548–568. https://doi.org/10.1111/tops.12212.
Golub, L., Frederick, W., 1970. An analysis of children's writing under different stimulus conditions. Res. Teach. Eng. 4 (2), 168–180.
Harley, B., King, M.Lou, 1989. Verb Lexis in the Written Compositions of Young L2 Learners. Stud. Second Lang. Acquis 11 (4), 415–439.
Hsiao, Y., Dawson, N.J., Banerji, N., Nation, K., 2022. The nature and frequency of relative clauses in the language children hear and the language children read: a developmental cross-corpus analysis of English complex grammar. J. Child Lang. 1–26. https://doi.org/10.1017/S0305000921000957.
Hunt, K.W. (1965). Grammatical structures written at three grade levels. National Council of Teachers of English, Research Report No. 3, 1–176. https://eric.ed.gov/?id=ED113735.
Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., Hedges, L.V., 2010. Sources of variability in children's language growth. Cogn. Psychol. 61 (4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002.
Jackson, J.C., Watts, J., List, J.-M., Puryear, C., Drabble, R., Lindquist, K.A., 2021. From text to thought: how analyzing language can advance psychological science. Perspect. Psychol. Sci. 17 (3), 805–826. https://doi.org/10.1177/17456916211004899.
Kassambara, A., Mundt, F., 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses.
Kellogg, R.T., 2006. Professional writing expertise. In: Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R. (Eds.), The Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press, pp. 389–402. https://doi.org/10.1017/CBO9780511816796.022.
Klein, D., Manning, C.D., 2003. Accurate Unlexicalized Parsing, pp. 423–430. https://doi.org/10.3115/1075096.1075150.
Kuiken, F., Vedder, I., 2019. Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish. Int. J. Appl. Linguist. (United Kingdom) 29 (2), 192–210. https://doi.org/10.1111/ijal.12256.
Kyle, K., 2018. *Lexical-Diversity Python Package* (0.1.1).
Kyle, K., Crossley, S.A., 2018. Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. Modern Lang J 102 (2), 333–349. http://www.jstor.org/stable/44981064.
Lê, S., Josse, J., Husson, F., 2008. {FactoMineR}: a package for multivariate analysis. J. Stat. Softw. 25 (1), 1–18. https://doi.org/10.18637/jss.v025.i01.
Leech, G., Rayson, P., Wilson, A., 2001. Word Frequencies in Written and Spoken English, 1st ed. Taylor & Francis.
Lei, S., A, N.M, 2012. Narrative writing in children and adolescents: examining the literate Lexicon. Lang. Speech Hear. Serv. Sch. 43 (1), 2–13. https://doi.org/10.1044/0161-1461(2011/10-0099.
Li, Y.ke, Lin, S., Liu, Y., Lu, X., 2023. The predictive powers of fine-grained syntactic complexity indices for letter writing proficiency and their relationship to pragmatic appropriateness. Assessing Writing 56. https://doi.org/10.1016/j.asw.2023.100707.
Loban, W., 1976. Language development: kindergarten through grade twelve. NCTE committee on research report No. 18. National Council Teachers English.
Loper, E., Bird, S., 2002. *Nltk. July 2002* 63–70. https://doi.org/10.3115/1118108.1118117.
Lu, X., 2010. Automatic analysis of syntactic complexity in second language writing. Int. J. Corpus Linguistics 15 (4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu.
Lu, X., 2012. The relationship of lexical richness to the quality of ESL learners' oral narratives. Modern Language J. 96 (2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232.x.
Malvern, D., Richards, B., Chipere, N., Durán, P., 2004. Lexical Diversity and Language Development. Lexical Diversity and Language Development. https://doi.org/10.1057/9780230511804.
McCarthy, P.M., Jarvis, S., 2010. MTLD, vocd-D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment. Behav. Res. Methods 42 (2), 381–392. https://doi.org/10.3758/BRM.42.2.381.
McCormick, C.B., Busching, B.A., Potter, E.F., 1992. Children's knowledge about writing: the development and use of narrative criteria. In: Pressley, M., Harris, K.R., Gutherie, J.T. (Eds.), Promoting Academic Competence and Literacy in School. Academic Press, pp. 331–336.
McNamara, D.S., Graesser, A.C., McCarthy, P.M., Cai, Z., 2014. Automated Evaluation of Text and Discourse with Coh-Metrix. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664.
Montag, J.L., 2019. Differences in sentence complexity in the text of children's picture books and child-directed speech. First Lang. 39 (5), 527–546. https://doi.org/10.1177/0142723719849996.
Montag, J.L., Jones, M.N., Smith, L.B., 2015. The Words Children Hear : picture Books and the Statistics for Language Learning. Psychol. Sci. 26 (9), 1489–1496. https://doi.org/10.1177/0956797615594361.
Montag, J.L., MacDonald, M.C., 2015. Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. J. Exp. Psychol. Gen. 144 (2), 447.
Moyle, M.J., Weismer, S.E., Evans, J.L., Lindstrom, M.J., 2007. Longitudinal relationships between lexical and grammatical development in typical and late-talking children. J. Speech Lang. Hear. Res. 50 (2), 508–528. https://doi.org/10.1044/1092-4388(2007/035.
Myhill, D., 2008. Towards a linguistic model of sentence development in writing. Lang. Educ. 22 (5), 271–288. https://doi.org/10.1080/09500780802152655.
Nasseri, M., Lu, X., 2020. Updated LCA-AW For Python 3, Lexical Complexity Analyzer for Academic Writing, Version 2.2. https://doi.org/10.5281/zenodo.4147000.
Nation, K., Dawson, N., Hsiao, Y., 2022. 'Book language' and its implications for children's language, literacy, and development. Curr. Dir. Psychol. Sci.
Nippold, M.A., 2007. Later Language Development: School-age Children, Adolescents, and Young Adults.
Norris, J.M., Ortega, L., 2009. Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. Appl. Linguist. 30 (4), 555–578. https://doi.org/10.1093/applin/amp044.
Olinghouse, N.G., Leaird, J.T., 2009. The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. Read. Writ. 22 (5), 545–565. https://doi.org/10.1007/s11145-008-9124-z.
Olinghouse, N.G., Wilson, J., 2013. The relationship between vocabulary and writing quality in three genres. Read. Writ. 26 (1), 45–65. https://doi.org/10.1007/s11145-012-9392-5.
Ortega, L., 2003. Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing. Appl. Linguist. 24 (4), 492–518. https://doi.org/10.1093/applin/24.4.492.
Peltz, F.K., 1973. The effect upon comprehension of repatterning based on students' writing patterns. Read. Res. Q. 9 (4), 603–621. https://doi.org/10.2307/747005.
Piantadosi, S.T., 2014. Zipf's word frequency law in natural language: a critical review and future directions. Psychon. Bull. Rev. 21 (5), 1112–1130. https://doi.org/10.3758/s13423-014-0585-6.
Read, J., 2000. Assessing vocabulary. Cambridge Language Assessment. Cambridge University Press. https://doi.org/10.1017/CBO9780511732942.
Roland, D., Dick, F., Elman, J.L., 2007. Frequency of basic english grammatical structures: a corpus analysis. J. Mem. Lang. 57 (3), 348–379. https://doi.org/10.1016/j.jml.2007.03.002.
Rubin, D.L., Piché, G.L., 1979. Development in syntactic and strategic aspects of audience adaptation skills in written persuasive communication. Res. Teach. Eng. 13 (4), 293–316.
Templin, M.C., 1957. Certain language skills in children; their development and interrelationships. Certain Language Skills in children; Their Development and Interrelationships. University of Minnesota Press.
Treffers-Daller, J., Parslow, P., Williams, S., 2018. Back to basics: how measures of lexical diversity can help discriminate between CEFR levels. Appl. Linguist. 39 (3), 302–327. https://doi.org/10.1093/applin/amw009.
Uccelli, P., Dobbs, C.L., Scott, J., 2012. Mastering academic language: organization and stance in the persuasive writing of high school students. Written Commun. 30 (1), 36–62. https://doi.org/10.1177/0741088312469013.
Ure, J. (1971). Lexical density and register differentiation. In G. Perren & J.L.M. Trim (Eds.), *Applications of Linguistics* (pp. 443–452).
Verhoeven, L., Aparici, M., Cahana-Amitay, D., van Hell, J.G., Kriz, S., Viguié-Simon, A., 2002. Clause packaging in writing and speech. Written Lang. Literacy 5 (2), 135–161. https://doi.org/10.1075/wll.5.2.02ver.
Wagner, R.K., Puranik, C.S., Foorman, B., Foster, E., Wilson, L.G., Tschinkel, E., Kantor, P.T., 2011. Modeling the development of written language. Read. Writ. 24 (2), 203–220. https://doi.org/10.1007/s11145-010-9266-7.
Wolfe-Quintero, K., Inagaki, S., Kim, H.-Y., 1998. Second language development in writing : measures of fluency, accuracy, & complexity. University of Hawai'i, Second Language Teaching & Curriculum Center.
Zenker, F., Kyle, K., 2021. Investigating minimum text lengths for lexical diversity indices. Assessing Writing 47. https://doi.org/10.1016/j.asw.2020.100505.
Zipf, G.K., 1949. Human behavior and the principle of least effort. Human Behavior and the Principle of Least Effort. Addison-Wesley Press.