Lexical Richness and Syntactic Complexity in Children's Story Writing

Yaling Hsiao, Nicola Dawson, Nilanjana Banerji, and Kate Nation

1. Introduction

The ability to produce sophisticated words and complex sentences are hallmarks of language and literacy development. Unlike conversation that happens in the here and now, aided by facial expression, gesture, intonation and a shared context, written language is 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). These differences start early and are present in books written for pre-school children to listen to in the context of shared reading (e.g., Dawson et al., 2021). When do these linguistic features emerge in children's own writing and how do they build with development? The aim of the current study is to chart the emergence of linguistic complexity in young children's narrative story writing, quantitively and at-scale.

1.1. Linguistic Complexity

We calculated a range of metrics tapping both lexical and syntactic complexity, and examined development change through middle childhood. Our study thus provides a comprehensive examination of lexical and syntactic development in writing throughout primary and early secondary school years. We begin by reviewing how lexical and syntactic complexity have been defined and measured before describing how we used these metrics to capture children's writing development.

1.1.1. Lexical complexity

This refers to the breadth and quality of vocabulary use and has been linked to the holistic quality of writing (Engber, 1995). Lexical complexity comprises three components: lexical density (the ratio of the number of lexical words to the total number of words), lexical sophistication (proportion of advanced or difficult words in a text) and lexical diversity (the ratio of number of unique word types to the total number of word tokens in a text) (Durrant et al., 2021; Lu, 2012; Read,

*Yaling Hsiao, University of Oxford, yaling.hsiao@psy.ox.ac.uk.

 Proceedings of the 46th annual Boston University Conference on Language Development, ed. Ying Gong and Felix Kpogo, 325-337. Somerville, MA: Cascadilla Press.

2000). A range of measures capture these lexical features. Lu (2012) used 24 different measures in a corpus investigation of second language writing quality. We adopted these 24 measures in our study, and added six more designed to be less sensitive to text length (McCarthy & Jarvis, 2010). Table 1 summarises all 30 measures.

Table 1. The 30 lexical richness measures used in this study

#	Code	Measure	Definition
Lexical Density			
1	ld	Lexical density	number of content word tokens/number of all tokens
Lexical Sophistication			
2	ls1	lexical sophistication by token	sophisticated tokens/all tokens
3	ls2	lexical sophistication by type	sophisticated types/all types
4	vs1	Verb sophistication 1	sophisticated verb types/all verb tokens
5	vs2	Verb sophistication 2	sophisticated verb types/square root of 2 x number of all verb tokens
6	cvs1	Verb sophistication 3	square of number of sophisticated verb types/all verb tokens
Lexical Diversity			
7	ndw	types	number of different words
8	ndwz	types in first 50 words	number of different words in the first 50 words
9	ndwrtz	types 50 word samples	mean number of different words in the 10 samples of 50 words
10	ndwesz	types 50 word sequences	mean number of different words in the 10 samples of 50 word sequences
11	ttr	type-token ratio	number of all types/number of all tokens
12	msttr	mean segmental TTR	splitting the text into 50 word segments, mean TTR of all segments
13	cttr	corrected TTR	types/square root of 2 x tokens
14	rttr	root TTR	types/square root of tokens
15	logttr	bilogagarithmic TTR	log(types)/log(tokens)
16	uber	uber ttr	Log(square of tokens)/Log(tokens/types)
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
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
19	MTLD	measure of textual lexical diversity	the average number of words in a row for which a certain TTR is maintained

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
21	MTLD-bi	moving-average bidirectional MTLD	MTLD in each direction using a moving window
22	lv	lexical word variation	content word types/all content word tokens
23	lv1	verb variation 1	verb types/all verb tokens
24	svv1	verb variation 2	verb types/square root of 2 x all verb tokens
25	cvv1	verb variation 3	square of number of verb types/all verb tokens
26	vv2	verb variation 4	verb types/all content word tokens
27	nv	noun variation	noun types/all content word tokens
28	adjv	adjective variation	adjective types/all content word tokens
29	advv	adverb variation	adverb types/all content word tokens
30	modv	modifier variation	adjective + adverb types/all content word tokens

1.1.2. Syntactic complexity

Complex syntactic structures allow people to express complicated ideas or relationships, and to do so in a more concise and sophisticated manner (Beers & Nagy, 2009). Not surprisingly then, greater syntactic complexity is associated with more advanced writing (Myhill, 2008). Syntactic complexity is typically quantified using the length of production units (e.g., length of sentences, T-units, clauses), the amount of subordination or coordination, and the number of particular syntactic structures (Ortega, 2003). As language develops, production unit length increases and older children use complex syntactic structures more often (e.g., Beers & Nagy, 2009; Durrant et al., 2020; Hunt, 1965). We analysed 14 syntactic complexity measures (Table 2), building from Lu (2010) who developed automatic computing software for L2 assessment. These included three indices of *length of production units*, i.e., length of sentences, length of T-units, and length of clauses, and measures of *syntactic complexity ratio*, which 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.

Table 2. The 14 syntactic complexity measures used in this study

#	Code	Measure	Definition
Unit of Production			
1	MLS	mean length per sentence	mean number of words in a sentence
2	MLT	mean length per T-unit	mean number of words in a T-unit
3	MLC	mean length per clause	mean number of words in a clause
Complexity Ratio			
4	CS	sentence complexity ratio	mean number of clauses per sentence

5	VP.T	verb phrase per T-unit	mean number of verb phrases per T-unit
6	C.T	T-unit complexity ratio	mean number of clauses per T-unit
7	DC.C	dependent clause ratio	mean number of dependent clauses per clause
8	DC.T	dependent clause per T-unit	mean number of dependent clauses per T-unit
9	T.S	sentence coordination ratio	mean number of coordinate phrases per sentence
10	CT.T	complex T-unit ratio	mean number of complex T-unit per T-unit
11	CP.T	coordinate phrases per T-unit	mean number of coordinate phrases per T-unit
12	CP.C	coordinate phrases per clause	mean number of coordinate phrases per clause
13	CN.T	complex nominals per T-unit	mean number of complex nominals per T-unit
14	CN.C	complex nominals per clause	mean number of complex nominals per clause

Our overall aim was to describe in detail the writing component of the Oxford Children's Corpus, held by Oxford University Press. In total, this contains over a million stories written by 5–13-year-old children in the UK. As Tables 1 and 2 make clear that lexical and syntactic complexity can each be captured in 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 ages and within the same study. Thus, our first goal was to do this using 44 different measures of linguistic complexity. From this, we took a statistical approach and used Principal Component Analysis to identify relationships between the different measures. We then considered developmental change in these markers of linguistic complexity by comparing the writing of younger and older children.

2. Method

2.1. The corpus

The Oxford Children's Corpus contains all stories submitted as part of the BBC Radio 2's 500 Words children's writing competition, an annual competition that ran between 2011 and 2020. We selected all those submitted in 2019, N= 107,273 (approximately 55-million word tokens), the year with the largest number of submission to date. Children aged 5-13 years were invited to submit a story on any theme or topic, so long as the word count was no greater than 500 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.

2.2. Procedure

Pre-processing involved removing punctuation and converting all characters to lower case to ensure that the same words in different cases were counted as the same type. We also removed 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 over 10 letters, or average sentence length of over 50 words. 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 & Bird, 2002) for tokenization and sentence segmentation, the *Lexical Complexity Analyzer for Academic Writing* (Nasseri & Lu, 2020) and the *lexical-diversity package* (Kyle, 2018) for calculating lexical richness. We used the *L2 Syntactic Complexity Analyzer* (Lu, 2010) for computing syntactic complexity; this uses the Stanford Parser (Klein & Manning, 2003) to generate part-of-speech tags and parse trees, and to extract relevant syntactic units or phrases.

3. Results

3.1. Lexical and syntactic complexity

We computed the lexical and syntactic complexity of each story according to each of the 44 metrics described in Tables 1 and 2. Scores were regressed on Key Stage 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 < .0001).

3.2. Relationships between individual measures of linguistic complexity

To reduce the number of dimensions and identify how the variables cluster together, all 44 variables were entered into a Principle Component Analysis (PCA). This type of analysis generates new variables, or principal components or dimensions, which are orthogonal to each other to represent the variation present in the original dataset. The analysis was conducted in R using the FactoMineR package (Lê et al., 2008) and the results were visualised using the factoextra

package (Kassambara & Mundt, 2020). The scree plot in Figure 1 shows the amount of variance explained by the top ten components/dimensions, each with eigenvalues over 1. Together, these ten dimensions explained 83% of variance. The scree plot shows a significant decrease and then a plateau in 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.

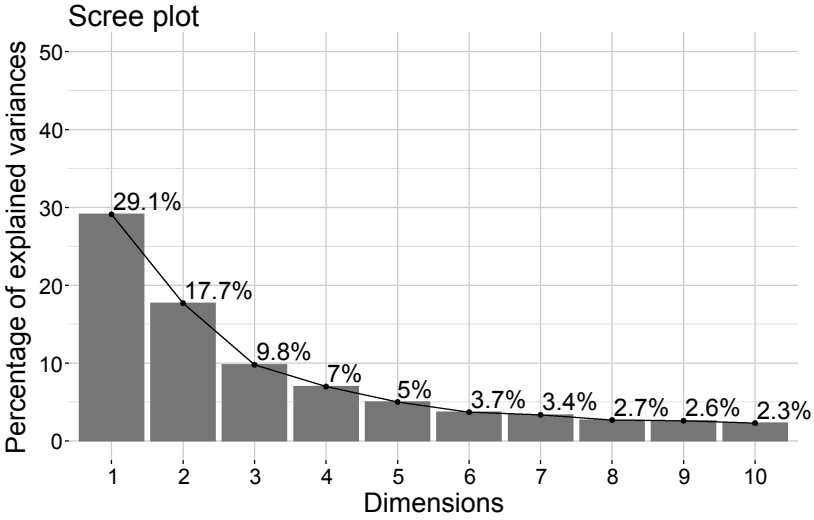

Figure 1. Scree plot of the Principle Component Analysis on lexical and syntactic complexity measures of children's writing

We examined the first two principal components in detail, using cos2 (or squared cosine or squared coordinates; cos2 is equal to the squared values). A high cos2 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 cos2 across all the principal components is equal to one. Figure 2 visualises the cos2 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 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 darker colours indicating higher quality. As can be seen from Figure 2, the first dimension separates lexical complexity from syntactic complexity. The high-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). Figure 2 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).

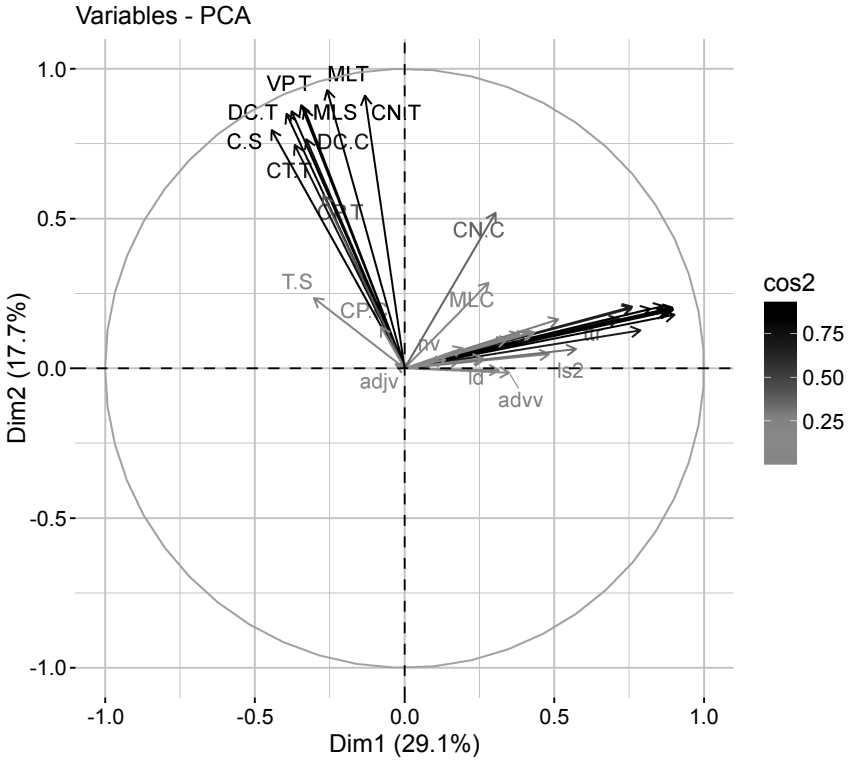

Figure 2. Two-dimensional visualisation of the quality of representation of each complexity measure on the first two principal components, using cos2

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 cos2 value of a complexity measure by the sum of cos2 of the dimension. As shown in Figure 3, the first dimension was explained by lexical diversity measures 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 in the second dimension were higher than those described by the first dimension.

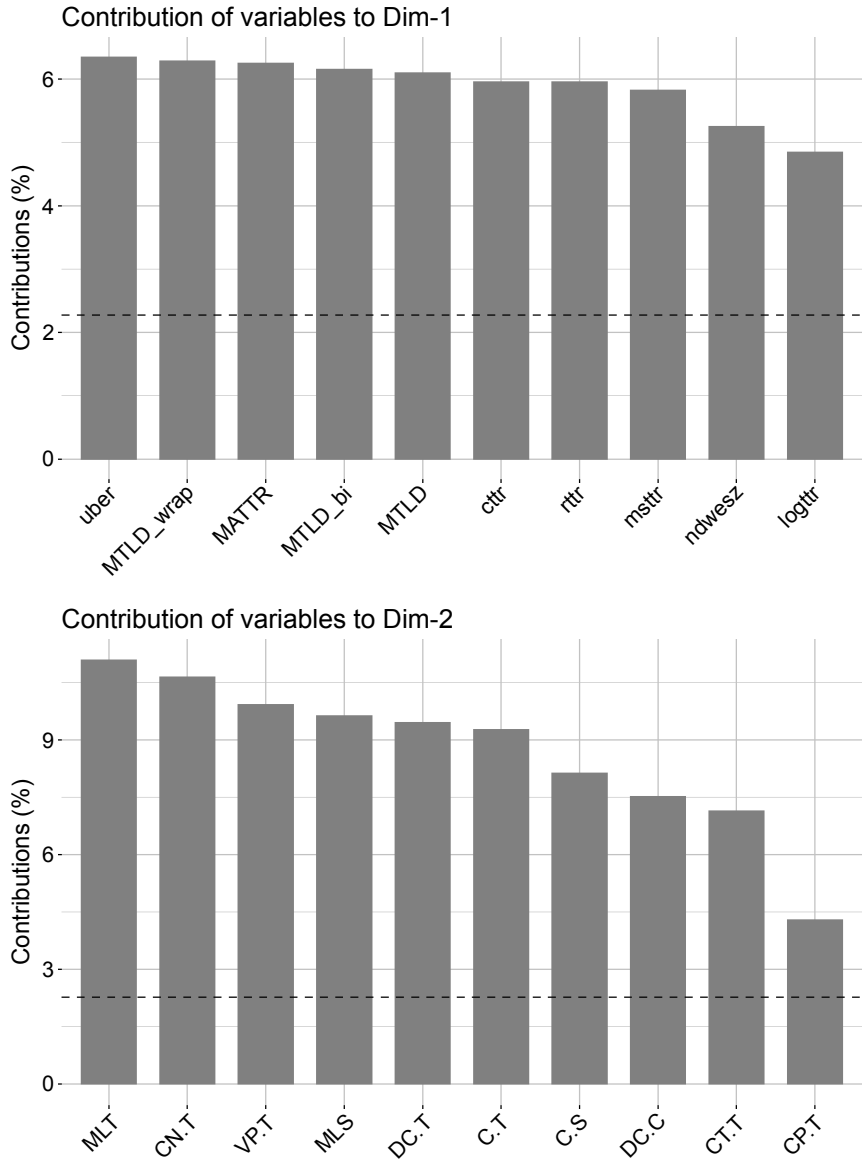

Figure 3. Percentage of contribution by the top 10 measures of complexity in the first dimension (top panel) and the second component (lower panel). The dark dotted line indicates expected average contribution, at 2.2% (100% divided by the total number of features measured, N= 44).

Finally, to assess the quality of representation by age, we investigated how individual data points (each representing a text) clustered on the first two dimensions as a function of the author's Key Stage. Figure 4 shows visible light data points representing stories written by Key Stage 1 children at the lower left quadrant, suggesting that younger children's writing was poorly represented by both of the dimensions. On the contrary, the dark 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 compared to other quadrants. This suggests that syntactic complexity varied among individual children, across Key Stage. 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.

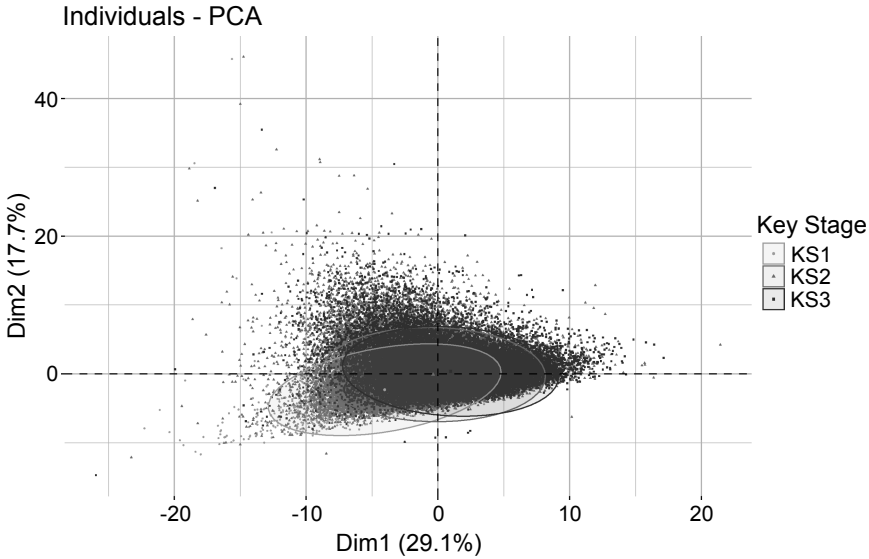

Figure 4. The quality of representation of each story along the two dimensions as a function of author Key Stage. The circles indicate concentration ellipses for each Key Stage.

4. 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. The stories were not written for this investigation, nor was their content prompted by pictures or other experimental instructions. The only constraint was that each story should be no longer than 500 words. While other studies on first language writing development have looked in detail at a more restricted set of linguistic features, or focused within a more restricted age range

or in a smaller sample, our approach was to consider a range of language features across a broad age range and within a very large sample.

We begin by examining the relationship between age and linguistic complexity. Growth was seen across a large set of lexical and syntactic features. If we accept that these features mark writing quality, they provide evidence that writing becomes increasingly complex and higher in quality with age. We included 30 different measures of lexical diversity. Generally, these clustered together, and stories written by older children showed greater levels of lexical diversity than those written by younger children, mirroring findings in the literature (Berman & Verhoeven, 2002; Durrant & Brenchley, 2019; Malvern et al., 2004; Wagner et al., 2011). Three of the 30 variables showed no age effect (verb variation by word type) or a decline with age (noun variation and adjective variation). Lexical diversity measures were strongly represented on the first component, especially those calculated with methods designed to reduce the confounding effect of text length (Malvern et al., 2004). They were also represented on the second component that primarily captured syntactic complexity.

Regarding complexity at the sentence level, we observed that sentences grew in number and by length with age, even in our sample with a word limit. Longer sentences generally allow building of complex ideas and use of sentence structures, and have traditionally been treated as indications for language growth (Bear, 1939; Golub & Frederick, 1970; Hunt, 1965; Myhill, 2008). Similar positive correlation with age was seen in the length of two other production units: T-units and clauses, replicating the findings in the literature of children's writing (Golub & Frederick, 1970; Hunt, 1965; Peltz, 1973; Rubin & Piché, 1979; Wagner et al., 2011). Syntactic complexity was mainly represented by the second component. Growth was seen by developmental stage in most measures except for CP.T and CP.C, which involved the operation of using coordination phrases. Coordination has been found to exhibit non-positive correlation with age (Golub & Frederick, 1970; Hunt, 1965; Peltz, 1973), perhaps because it emerges relatively early as a sentence combining operation, which is replaced by other complex grammar later on.

Our Principal Component Analysis showed two major dimensions that accounted for possibly distinct portions of variance in children's writing: lexical vs. syntactic complexity. Although this contrasts to the close relationship between lexical and syntactic development seen in early language acquisition in the spoken domain (Bates & Goodman, 1997; Devescovi et al., 2005; Moyle et al., 2007), some evidence suggests that syntactic development may follow a different trajectory from lexical development. Studies showed that differences in the syntactic structures caregivers use affect children's language growth, suggesting a causal flow (Huttenlocher et al., 2010), whereas lexical growth had a bi-directional influence between caregiver and child speech. Lexical acquisition, particularly nouns, requires a simpler word-to-world mapping, compared to the more complicated sentence-to-world mapping for argument structures (Fisher et al., 1994; Gleitman, 1990). We indeed noticed that syntactic complexity developed in distinct patterns from lexical complexity, in that children across all

Key Stages exhibited wider variation in syntactic complexity and could imply later maturation or expression of syntactic knowledge compared to lexical development.

The current study serves as an important 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 children in primary and early secondary school. The Principal Component Analysis suggests lexical complexity and syntactic complexity accounted for major proportions of variances and that certain measures have better representation in these dimensions. The lexical and syntactic complexity appeared to have different developmental profiles in children's writing development.

References

Bates, Elizabeth, & Goodman, Judith C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes*, *12*(5–6), 507–584. https://doi.org/10.1080/016909697386628

Bear, Mata V. (1939). Children's Growth In The Use Of Written Language. *The Elementary English Review*, *16*(8), 312–319. http://www.jstor.org/stable/41383155

Beers, Scott F., & Nagy, William E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing*, *22*(2), 185–200. https://doi.org/10.1007/s11145-007-9107-5

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

Biber, Douglas. (1988). *Variation across Speech and Writing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

Dawson, Nicola, Hsiao, Yaling, Tan, Alvin Wei Ming, Banerji, Nilanjana, & Nation, Kate. (2021). Features of lexical richness in children's books: Comparisons with child-directed speech. *Language Development Research*, *2021*, 1–45. https://doi.org/10.34842/5we1-yk94

Devescovi, Antonella, Caselli, Maria Cristina, Marchione, Daniela, Pasqualetti, Patrizio, Reilly, Judy, & Bates, Elizabeth. (2005). A crosslinguistic study of the relationship between grammar and lexical development. In *Journal of Child Language* (Vol. 32, Issue 4, pp. 759–786). Cambridge University Press. https://doi.org/10.1017/S0305000905007105

Durrant, Philip, & Brenchley, Mark. (2019). Development of vocabulary sophistication across genres in English children's writing. *Reading and Writing*, *32*(8), 1927–1953. https://doi.org/10.1007/s11145-018-9932-8

Durrant, Philip, Brenchley, Mark, & McCallum, Lee. (2021). *Understanding Development and Proficiency in Writing: Quantitative Corpus Linguistic Approaches*. Cambridge University Press. https://doi.org/DOI: 10.1017/9781108770101

Durrant, Philip, Clarkson, Rebecca, & Brenchley, Mark. (2020). Syntactic Development across Genres in Children's Writing: The Case of Adverbial Clauses. In *Journal of Writing Research* (Vol. 12, Issue 2). https://doi.org/10.17239/jowr-2020.12.02.04

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

Fisher, Cynthia, Hall, D. Geoffre., Rakowitz, Susan, & Gleitman, Lila. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, *92*, 333–375. https://doi.org/https://doi.org/10.1016/0024-3841(94)90346-8

Gleitman, Lila. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, *1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Golub, Lester, & Frederick, Wayne. (1970). An Analysis of Children's Writing under Different Stimulus Conditions. *Research in the Teaching of English*, *4*(2), 168–180.

Hunt, Kellogg 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, Janellen, Waterfall, Heidi, Vasilyeva, Marina, Vevea, Jack, & Hedges, Larry V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Kassambara, Alboukadel, & Mundt, Fabian. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. https://cran.r-project.org/package=factoextra

Klein, Dan, & Manning, Christopher D. (2003). *Accurate unlexicalized parsing*. 423–430. https://doi.org/10.3115/1075096.1075150

Kyle, Kristopher. (2018). *lexical-diversity Python package* (0.1.1). https://github.com/kristopherkyle/lexical_diversity

Lê, Sébastien, Josse, Julie, & Husson, François. (2008). {FactoMineR}: A Package for Multivariate Analysis. *Journal of Statistical Software*, *25*(1), 1–18. https://doi.org/10.18637/jss.v025.i01

Loper, Edward, & Bird, Steven. (2002). *Nltk*. *July 2002*, 63–70. https://doi.org/10.3115/1118108.1118117

Lu, Xiaofei. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

Lu, Xiaofei. (2012). The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives. *The Modern Language Journal*, *96*(2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232.x

Malvern, David, Richards, Brian, Chipere, Ngoni, & Durán, Pilar. (2004). Lexical Diversity and Language Development. In *Lexical Diversity and Language Development*. https://doi.org/10.1057/9780230511804

McCarthy, Philip M., & Jarvis, Scoot. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

Moyle, Maura Jones, Weismer, Susan Ellis, Evans, Julia L., & Lindstrom, Mary J. (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research*, *50*(2), 508–528. https://doi.org/10.1044/1092-4388(2007/035)

Myhill, Debra. (2008). Towards a linguistic model of sentence development in writing. *Language and Education*, *22*(5), 271–288. https://doi.org/10.1080/09500780802152655

Nasseri, Maryam, & Lu, Xiaofei. (2020). *Updated LCA-AW for Python 3, Lexical Complexity Analyzer for Academic Writing, version 2.2*. https://doi.org/10.5281/zenodo.4147000

Ortega, Lourdes. (2003). Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Synthesis of College-level L2 Writing. *Applied Linguistics*, *24*(4), 492–518. https://doi.org/10.1093/applin/24.4.492

Peltz, Fillmore Kenneth. (1973). The Effect upon Comprehension of Repatterning Based on Students' Writing Patterns. *Reading Research Quarterly*, *9*(4), 603–621. https://doi.org/10.2307/747005

Read, John. (2000). Assessing Vocabulary. In *Cambridge Language Assessment*. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511732942

Roland, Douglas, Dick, Frederic, & Elman, Jeffrey L. (2007). Frequency of Basic English Grammatical Structures: A Corpus Analysis. *Journal of Memory and Language*, *57*(3), 348–379. https://doi.org/10.1016/j.jml.2007.03.002

Rubin, Donald L., & Piché, Gene L. (1979). Development in Syntactic and Strategic Aspects of Audience Adaptation Skills in Written Persuasive Communication. *Research in the Teaching of English*, *13*(4), 293–316. http://www.jstor.org/stable/40170773

Wagner, Richard K., Puranik, Cynthia S., Foorman, Barbara, Foster, Elizabeth, Wilson, Laura Gehron, Tschinkel, Erika, & Kantor, Patricia Thatcher. (2011). Modeling the development of written language. *Reading and Writing*, *24*(2), 203–220. https://doi.org/10.1007/s11145-010-9266-7

Proceedings of the 46th annual Boston University Conference on Language Development

edited by Ying Gong
and Felix Kpogo

Cascadilla Press Somerville, MA 2022

ISSN 1080-692X
ISBN 978-1-57473-077-7 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com