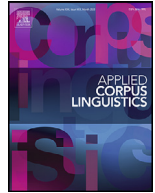




ELSEVIER

Contents lists available at ScienceDirect

## Applied Corpus Linguistics

journal homepage: [www.elsevier.com/locate/acorp](http://www.elsevier.com/locate/acorp)

## Studying children's writing development with a corpus

Philip Durrant

University of Exeter, St Luke's Campus, Exeter EX1 2LU, UK

## ARTICLE INFO

## Keywords:

Children's writing  
School writing  
Learner corpus  
Grammar  
Vocabulary  
Collocation

## ABSTRACT

One of Randi Reppen's major contributions has been her pioneering corpus research into school children's writing. In this paper, I will discuss how such research can contribute to both theory and educational practice. I will then look at two sets of unresolved methodological issues in this area: the issue of defining appropriate linguistic and textual categories, and the issue of drawing valid developmental inferences.

The issue of categories arises because corpus analysis depends on abstracting away from specific instances of language use in specific texts to make claims about the use of linguistic categories (e.g., noun phrases, low-frequency vocabulary) in textual categories (e.g., stories, science reports). Such abstraction enables researchers to draw out patterns of language variation that are difficult to spot by other means. But it also raises the problem of how to define categories that are reliably operationalizable, that capture consistent developmental patterns, and that are theoretically and educationally informative.

The issue of drawing valid inferences stems from the fact that corpus data record the products of complex, contextually contingent writing processes, involving the interaction of many variables. Capturing the combined outcomes of these complex processes promotes ecological validity. However, it also creates challenges for researchers who want to draw conclusions about specific aspects of the writing process, such as writers' knowledge of vocabulary or grammar, or their emerging awareness of audience.

This paper will discuss these issues in detail, illustrating their impact and suggesting ways forward for educationally informative corpus research.

## 1. Introduction

A central theme in Randi Reppen's work has been her commitment to informing teaching through corpus analysis of school children's writing (Reppen, 1994, 2001, 2002, 2007, 2009). Like Randi, I believe that research of this sort has a great deal to offer. For many children, "writing conventions will remain a mystery unless teachers are able to bring these forms and patterns of language use to conscious awareness" (Reppen, 2002, p. 321), and for this to happen, teachers themselves need to know what those forms and patterns are. Because intuitions about such matters are notoriously imprecise and inaccurate, this raises the need for systematic, empirically based descriptions of the sorts of texts that children are asked to write (Rose and Martin, 2012).

Of course, not all child writing can be taken as a model, and not all child writing is equally successful. This raises the question of what successful writing looks like. Though teachers may have a tacit sense of what makes for *good* writing, they are often unable to formulate this in any but the vaguest of terms (Lines, 2014; Myhill, 1999). This is problematic because, unless a concept of quality writing can be brought out of the heads of teachers and given an external formulation, it can't be made available to learners (Sadler, 1989). It is also problematic because, unless it is explicitly formulated, the concept of quality is protected from

critique. If we are clear about what we are valuing in student writing, we can have a discussion about whether we are valuing the right things, and about whether our practices unfairly advantage some children over others. Since at least the 1960s (e.g., Hillocks 1964, Percival 1966, Potter 1967, Sampson 1964), researchers have been analysing children's texts to identify sets of linguistic features that distinguish those awarded higher or lower grades by teachers. This endeavour has both yielded a rich and complex picture of the construct of quality itself and highlighted the ways that this construct varies across contexts (Durrant, Brenchley, & McCallum, 2021).

A further way in which corpus analysis of children's texts can inform teaching is by tracing how writing develops as children progress through school. As noted above, explicit descriptions of the types of texts that children are asked to write are an important tool in designing a course of study. However, since this writing changes through the course of their education, teachers need descriptions that are specific to each educational level. Further, by tracing changes in children's writing over time, we can evaluate the impact of educational programmes on writing and, when analyses are repeated across multiple contexts, identify general developmental patterns that can inform syllabus planning.

While these goals could be (and have been) pursued with many different methods, a distinctively corpus linguistic approach - that is,

E-mail address: [p.l.durrant@exeter.ac.uk](mailto:p.l.durrant@exeter.ac.uk)

<https://doi.org/10.1016/j.acorp.2022.100026>

Received 27 July 2022; Accepted 4 August 2022

Available online xxx

2666-7991/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

one that incorporates computational methods to enable analysis of large numbers of systematically sampled texts - has some powerful advantages (Durrant et al., 2021). Firstly, by working with many texts, it enables us to spot subtle patterns that might not be visible in a smaller sample and to make generalizable claims about variation across contexts, text types, levels of quality, and stages of education. Second, instructing a computer to identify a linguistic feature requires us to articulate clearly what that feature means and how it can be spotted in a text. Corpus research thus promotes transparency and reliability by pushing researchers to explicitly define and operationalize the features they are analyzing. Finally, it places a helpful distance between the analyst and the texts. Dealing with texts as data in a corpus, rather than solely as objects to be engaged with on an individual basis, “offers us a new perspective on the familiar” (Hunston, 2002, p. 3) and can help analysts notice patterns that might not fit their usual expectations or presuppositions about how language works.

These considerations make a powerful case for including corpus linguistic methods as a key tool in our endeavour to understand children’s writing. As with all research however, the strengths of these methods come with corresponding limitations and raise significant methodological dilemmas. In this paper, I will focus on two of these: the use of linguistic and textual categories, and the ways in which quantitative findings from corpus data can be validly interpreted. I argue that both issues point to the need for ongoing and cyclical interpretive engagement with a corpus and the contexts in which it was collected. Our analytical categories need to be held as provisional and kept under constant review as we progress through our analyses. The meanings of quantitative patterns likewise need to be constantly interrogated through both quantitative and qualitative post-hoc analyses and researchers need to be alert to the possibility that patterns may reflect aspects of the writing context that they had not originally anticipated.

## 2. The problem of defining categories

Corpus linguistics achieves much of its analytical power by abstracting away from specific instances of language in specific texts to generate counts of linguistic categories across large numbers of texts. Thus, rather than observing that *Hilary Mantel begins her novel The Mirror and the Light with the sentence “Once the queen’s head is severed, he walks away”*, a corpus linguist is more likely to say that *fiction texts often use fronted adverbial clauses and cataphoric reference*.

Abstraction of this sort relies on categories. Specifically, it relies on categories of linguistic features (such as *fronted adverbial clause* and *cataphoric reference*) and on categories of texts (such as *fiction*). In each case, categories can be defined at different levels of abstraction. The book opening quoted above, for example, could be described, with increasing degrees of abstraction, as including *a fronted adverbial of time clause*, *a fronted adverbial clause*, *an adverbial clause*, *a subordinate clause*, or just *a clause*. In each case, the category becomes broader, encompassing a wider range of forms. Each level of abstraction has the potential to yield new generalizations, but with a concomitant loss of detailed description. Similarly, the text might be categorized as *one of Hilary Mantel’s trilogy of historical novels about Thomas Cromwell*, *an historical novel*, *a novel*, *a work of fiction*, *a piece of writing*, or *a sample of the English language*. Again, greater abstraction may enable greater generality of claims. But it also leads to a loss of detail and risks conflating important patterns. The following sections will discuss these two types of categorization in turn.

### 2.1. Linguistic categories

Our choice of linguistic categories typically needs to respond to four, potentially competing, demands. First, categories must be identifiable with a high level of accuracy in a corpus. At one level, this is tricky because language can be fundamentally ambiguous between linguistic categories. The word *writing* in the phrase *writing is hard*, for example, could be counted either as a noun or as a verb (Durrant et al., 2021;

Huddleston and Pullum, 2002). Moreover, when dealing with large corpora<sup>1</sup>, we typically face the additional issue of needing to identify categories by (semi-)automated means. Although tools now exist that aim to analyse a wide range of features automatically, their accuracy remains a matter of debate and ongoing research (e.g., Huang et al. 2018, Newman and Cox 2020). Crucially, overall accuracy rates reported for such tools do not necessarily imply high accuracy for the particular features that a researcher wants to study in the particular set of texts with which they are working (Gray, 2019). A further concern is that such tools may limit the researcher to studying only those features that their creators have seen necessary (or found computationally convenient) to include. This inevitably leaves out much that may be of interest, both because some types of features (e.g., those that require distinguishing between the senses of polysemic words) are difficult to identify computationally and because each tool relies on a particular choice of analytical framework. Thus, a great deal of research is currently based on the grammatical model embodied in Stanford CoreNLP’s *Universal Dependencies* framework (Schuster and Manning, 2016), but few, if any, capture the categories of Systemic Functional Linguistics, which previous research has shown to give important insights into the development of children’s writing (e.g., Christie 2012, Christie and Derewianka 2008, Coffin 2006, Rose and Martin 2012).

Second, categories must identify regular developmental patterns. Development in the use of some linguistic categories may be so idiosyncratic to individual children or to specific contexts that they yield no generalizable patterns. Categories of this sort would not allow for meaningful corpus analysis. An example of a frequently analysed feature that may be of this type is *lexical density* (the proportion of words in a text that are lexical - i.e., adjectives, adverbs, nouns, or lexical verbs), which does not appear to show any consistent L1 developmental patterns either across time (Berman and Nir, 2010; Golub and Frederick, 1979) or measures of quality (Golub and Frederick, 1979; Uccelli, Dobbs, and Scott, 2013).

Third, categories must not misleadingly conflate patterns that may be visible at a finer-grained level of analysis. Linguistic categories gain their power by adjusting our analytical lens such that the details of individual cases go out of focus, allowing a broader picture to emerge. This move is, in some ways, analogous to stepping back from a photomosaic, so that we lose sight of the pictures that form its individual pixels but gain a view of a new larger-scale picture. In a corpus analysis, the new picture is what emerges when we find a statistically reliable pattern that holds across large numbers of cases.

The analogy is imperfect in that, in corpus analysis, interesting pictures can emerge at several different levels of abstraction. The challenge facing the analyst is to identify the level(s) that offer the most faithful picture(s) of writing development. Most corpus research into child writing has chosen to focus on the big picture, relying on a small number of very abstract linguistic categories (e.g., *words per clause*; *clauses per t-unit*) that yield some highly generalizable developmental patterns (Durrant et al., 2021). However, these broad patterns are likely to conflate important trends that would be observed on a finer-grained analysis, thus presenting a distorted picture of development. Biber et al. (2020) show, for example, that measures such as *words per t-unit* conflate clausal and phrasal complexity and that these two categories have very different developmental profiles. Analyses that do not distinguish such features may not only miss important patterns; they may produce actively misleading results as distinct patterns are averaged out into a central tendency that fails to represent anything meaningful.

An example of this averaging-out effect can be seen in Fig. 1, which shows the frequency of adverbial clauses in stories written by children

<sup>1</sup> Children’s writing corpora frequently comprise several hundreds, or even thousands, of texts – as shown in the summary table of relevant studies at <https://phildurrant.net/quantitative-measures-of-written-language-use/>. Last accessed 8 August 2022.

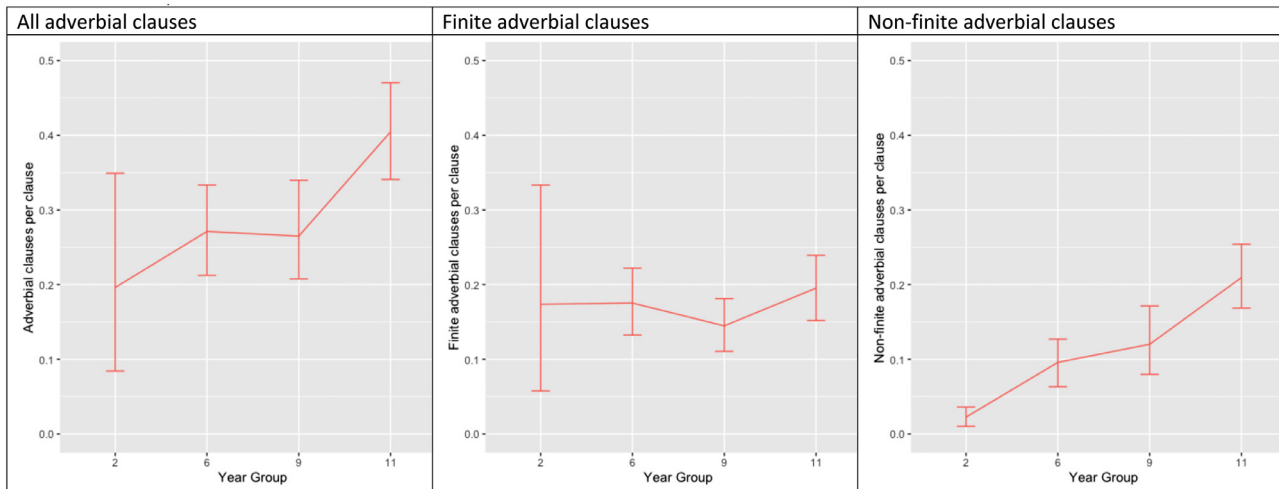


Fig. 1. Adverbial clauses per clause in children's stories.

at schools in England (Durrant, Brenchley, & Clarkson, 2020). When all cases of adverbial clauses are treated as a single group (represented in the far-left chart), the number of adverbial clauses per main clause is found to remain more-or-less constant for children in Years 2, 6, and 9. The slight differences between the means for these year groups stay well within the bounds of their respective 95% confidence intervals (represented by the vertical bars). There is then a marked increase at Year 11. Such a pattern might lead researchers to conclude that use of adverbial clauses is a feature of mature writing, mostly developing in the later stages of secondary school. However, it is important to remember that the general category of adverbial clause includes two rather different subtypes:

- finite adverbial clauses, e.g.:
  - *The party was full of laughter and people socialising as classical music faded into the background.*
  - *there must have been a power cut because there were no lights.*
- non-finite adverbial clauses, e.g.:
  - *Next, Batman punched the Joker, causing him to stumble.*
  - *Built from the very trees that guard it, it grew on the same soil.*

As the second and third charts in Fig. 1 show, these have distinct developmental profiles. Finite clauses are relatively common from the start of primary school and do not show any increase in frequency throughout the years of schooling. Non-finite clauses, in contrast, are very infrequent in the youngest children's writing, but their use increases in a

more-or-less linear fashion throughout the school years. The shape seen for adverbial clauses as a whole conflates these to show a pattern that fails to represent the development of either type, so giving a misleading picture of syntactic development.

The fourth, and perhaps overriding, demand to be considered when defining linguistic categories is that they need to be educationally meaningful. There is little value in finding that a particular linguistic category is reliably correlated with writing development if that category is unable to inform teaching. Again, the generic categories exemplified by *words per t-unit* provide a good example. Although reliably correlated with development, their ability to inform a programme of teaching is limited. It is not clear what a teacher could do with the information that texts with more words per t-unit are typically awarded higher grades. The information that texts with more complex noun phrases are typically scored higher is more useful, however, because *noun phrase* is a category that could be built into a syllabus and worked on in class. The information that texts that use more relative clauses and postmodifying prepositional phrases are scored higher is better still, since it pinpoints the types of noun phrase complexity that are most pertinent.

2.2. Textual categories

An example of the sort of issue raised by classification of texts can be seen in Fig. 2, which represents the use of academic verb-object collocations in children's non-literary writing (Durrant & Brenchley, 2021).

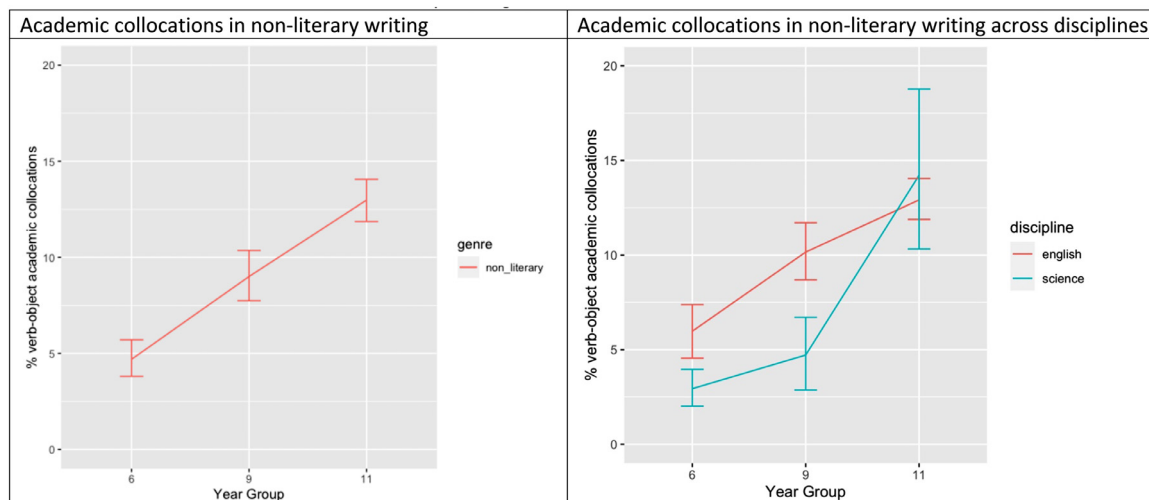


Fig. 2. Academic collocations in children's non-literary writing.

**argument(ative)**; argumentative expository; classification; compare/contrast; creative; description/descriptive; essay; explanatory; explorative/exploratory; **expository**; expository letter; expressive; factual; imaginative; imaginative descriptive; informative/informational; literary response; **narrative**; narrative descriptive; opinionative discursive; **persuasive/persuasion**; procedure; reporting; story; technical explanatory

terms used in 10 or more studies are highlighted in bold

**Fig. 3.** Genre categories from corpus studies of child writing.

Academic collocations are defined here as word pairings that occur in a corpus of mature academic writing (in this case, the BAWE<sup>2</sup> corpus) with a frequency greater one per million words and with a mutual information score of greater than five. These are pairings like *play-role*, *solve-problem*, and *provide-evidence*. As the left-hand chart shows, when all non-literary writing is taken as a single category, we see a clear linear increase in the use of such collocations across year groups. However, as the right-hand chart shows, when texts are divided into those written for English classes and those written for Science classes, a more complex pattern emerges. The Science writing makes significantly less use of academic collocations than English writing in Years 6 and 9, and there is little evidence of an increase in use between the two groups. However, from Year 9 to Year 11, we see a dramatic increase, such that Year 11 Science writing slightly overtakes that of English writing. This both shows a clear disciplinary difference in younger children's writing and suggests that an important shift of some kind takes place in Science writing in the second half of secondary school. Both these points are missed when analysis focuses on the broader textual category without considering discipline. This has substantive implications for our understandings of children's writing. The granularity of our analyses determines the contextual variables that are brought into play in our explanations of writing development. Optimizing this granularity is therefore central to identifying the primary communicative and genre-based contexts that drive such development (Mark Brenchley, personal communication, 25 February 2022).

Four main types of text category are commonly used in corpus research on children's writing: the educational stages at which texts were written; the grades assigned to texts by markers; the subject area (academic discipline) within which texts were written; and text genre (Durrant et al., 2021).

Each of these comes with its own complications. Time could be defined in literal chronological terms, focusing on a child's age in years and months, or in educational terms, focusing on a child's stage in their school career. In both cases, decisions need to be made about the appropriate breadth and nature of each period. Children might be divided, for example, into primary vs. secondary school, year groups, terms or semesters of study, or even more finely, weeks in a course.

Categories based on the grades assigned to texts introduce complex issues surrounding who does the grading, under what conditions, and according to what criteria. In studies focusing on authentic writing contexts, for example, it is common for very different grading systems and practices to be found in different schools, making analysis of grades in a corpus compiled across contexts highly problematic. In the context of England, for example, the 2014 National Curriculum (Department for Education, 2014) saw a deliberate shift away from a set of national attainment levels that tracked children's progress throughout their school careers to a policy of each school being expected to develop its own assessment practices. In the wake of this, the corpus of school writing that I collected with my colleagues between 2015 and 2018 (Durrant et al., 2021) included a wide array of different grading scales, including

<sup>2</sup> BAWE was developed at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800). More details can be found at the corpus website: [www.coventry.ac.uk/bawe/](http://www.coventry.ac.uk/bawe/).

descriptive three-point scales (e.g. *below/at/above age-related expectations*; *commenced/developing/secure*); systems of letters (e.g. *C to A\**); numbers (e.g. percentages); letters plus numbers (e.g. *A0 to B6*); and numbers plus letters (e.g. *3A–8C*).

Disciplines also are defined in different ways in different contexts. A glance at the literature on disciplinarity (e.g., Becher and Trowler 2001, Trowler et al. 2012) or at the range of disciplinary distinctions used in Applied Linguistic research (reviewed in Durrant, 2017) demonstrates that categories are far from fixed and can be defined at many levels of specificity. From the perspective of child writing, we can add the complication that school disciplines can take different forms at different stages of a child's education, making any disciplinary comparison across year groups challenging.

Arguably the trickiest set of issues, however, surrounds the category of genre. A genre is a group of texts that share a common purpose or set of purposes and that, in line with those purposes, share features of schematic structure, content, and style (Christie and Derewianka, 2008; Hyland, 2008; Rose and Martin, 2012; Swales, 1990). The difficulties of applying this idea to categorise actual texts are hinted at by the diversity of practice in the research literature. In our review of 104 corpus studies of child writing development published between 1945 and 2015, my colleagues and I (Durrant et al., 2021) found a wide range of terms used to categorise texts into genres (see Fig. 3). In some cases, different researchers apparently used different terms to refer to the same genre: *narratives* and *stories*, for example, appear to have the same thing in mind, as do *persuasive/persuasion* and *argument(ative)*. Perhaps more worryingly, some terms are used in different ways by different researchers. The common term *expository/exposition* is a central example of this. For Cox et al. (1991), an exposition "expresses factual information in order to inform or explain". For Rose and Martin (2012, p. 56), in contrast, it "argu[es] for a point of view". For Hall-Mills and Apel (2015), exposition is a broader term that encompasses both of these, involving "conveying facts or describing procedures, sharing basic information, relating cause-effect relationships, or arguing a point of view".

This diversity of definitions is reflected in the tasks used to elicit expositions in various studies. Tasks labelled as expository include those that require relating information:

- *the child was asked to assume the role of a ... nonfiction book scientist-writer and compose ... an information report on a familiar topic (Fang, 2002);*

Providing an explanation:

- *Pretend you are a super hero and you are being interviewed on the news. Tell everyone what special powers you would have. Also, explain what you would do with them to help the world. (Hall-Mills and Apel, 2015);*

setting out and persuading someone of a viewpoint:

- *Your class has just received some money for a field trip. In several paragraphs write an essay to your teacher explaining where you think your class should go, and why? (Koutsoftas and Gray, 2012).*
- *Imagine that there is a large undeveloped space in your school yard. Every student in the school has been asked for ideas about what to put there. A committee of teachers and parents will choose the best suggestion. (Roessingh et al. 2015).*

and stating and explaining a personal preference:

- *I like (person, place, or thing) \_\_\_ because \_\_\_\_\_. (Yates et al., 1995).*

For some studies, exposition is used so broadly that its defining feature is the negative one of not being a story:

- *Participants...were asked to give their ideas and thoughts on the topic of “problems between people” and were explicitly instructed not to tell or write a story but to discuss the issue (Berman and Nir-Sagiv, 2007).*

The key lesson of these examples is that the literature does not demonstrate a single coherent set of agreed genre terms. Researchers both use different terms to refer to similar types of texts and use the same term to refer to different types. This emphasizes the importance of transparency in defining the genre categories used in research; we cannot simply take for granted that all researchers will understand genre terms in the same way. Unfortunately, such transparency is often missing in published research.

A further difficulty with defining genres is that the types of writing children are asked to do change considerably as they progress through school. Indeed, a shift in the range and types of genres with which children engage is a key aspect of what counts as writing development (Beers and Nagy, 2011; Coffin, 2006). Thus, Christie and Derewianka (2008), for example, show how children’s writing about history progresses from simple chronological recounts – initially often personal and descriptive, later more generalized and including elements of cause and effect – to non-chronological genres such as descriptions of historical periods or sites (e.g. *Egyptian houses; Ancient Rome*), before moving on to assignments that involve interpreting and explaining events, arguing for a stance on an historical issue, or testing and evaluating the reliability of sources.

This evolution in genres becomes especially problematic when analysts want to trace changes in language across phases of schooling. As Fig. 2 showed, dividing texts into appropriate categories is important so that diverging patterns in different types of text are not fudged into a misleading ‘pattern’. But such analyses also require us to have categories that remain constant across time. Without this, no coherent comparison across year groups would be possible. This implies the need for genre categories that are not only sufficiently detailed to avoid misleading conflation of distinct patterns but also sufficiently coarse-grained to classify texts in a consistent way across year groups.

### 2.3. Interim conclusion: defining categories

Defining linguistic and textual categories is a crucial step in designing research into children’s writing. One key concern is practicality. We need linguistic and textual categories that can be reliably implemented at scale in a potentially large corpus. A second concern is usability. Which linguistic and textual categories are most informative will depend, in part, on which categories are in-play in the context we are trying to inform. An analysis that uses grammatical terms or generic labels that differ from the metalanguage of a particular educational context, for example, may struggle to say anything of interest to teachers in that context. Finally, we have the concern of achieving maximum fidelity to the data. As the examples above have demonstrated, our choice of categories can have a large effect on the nature of the results that emerge. While grouping texts or linguistic features into categories enables us to highlight differences between categories, it also makes our analysis blind to any differences within the categories. When we organise data into categories, we are therefore making informed guesses about where variation is most likely to occur; we expect variation between categories but not within them.

Good decision making therefore depends on the nature of linguistic variation in the studied context. While previous research can help inform such decisions, our systematic review suggests that the literature is often too sparse to give any real confidence (Durrant et al., 2021). We therefore face a chicken-and-egg problem of the sort described by Biber (1993) with regard to setting an appropriate sample size. That is, we can’t be sure what categories will work until we have spent some time exploring data from the corpus itself. This implies the need for a

flexible, cyclical process, in which analysis is carried out at different levels of specificity and categories are left open to revision in response to what we learn about patterns of variation from our initial analyses. As we will see below, the need for a flexible and cyclical approach to analysis is also pressed by our second set of problems: those surrounding interpretation.

### 3. The problem of interpretation

The second set of issues I will discuss in this paper concerns how we should interpret patterns derived from corpus data. Put another way, this is the question of what inferences can be drawn from our data. Inferences are important in corpus research because we rarely wish to make claims purely about the countable contents of the particular texts we are studying. A conclusion like *the set of writing comprising texts 2\_6c, 2\_8c, 2\_9b, [...] included, on average, fewer complex noun phrases than the set of writing comprising texts 13\_756d, 13\_757b, 13\_758b [...] would be pretty uninteresting*. The promise of corpus methods lies in what they tell us about broader constructs that our data are theorized to represent. We might want to learn, for example, about the contrasting communication strategies found in different genres of child writing; about changes in writing abilities, or particular aspects of linguistic proficiency, that occur over the course of a programme of study; about linguistic factors that predict raters’ grading behaviour, and so on.

Drawing an inference from the contents of our corpora to a broader theoretical construct requires us to make clear links from countable text features (e.g., number of distinct word types or mean words per noun phrase) to the theoretical constructs that are our primary object of interest (e.g., writers’ vocabulary repertoire or developing syntactic proficiency). This can be highly problematic. One set of issues, discussed at length elsewhere (Bulté and Housen, 2012, 2014; Norris and Ortega, 2009), concerns the ways in which linguistic constructs (such as *fluency* or *complexity*) should be operationalised in terms of countable objective features such as type-token ratios or frequencies of subordination.

Beyond this, however, is a broader set of issues concerning how differences in linguistic constructs should themselves be understood. If, for example, the syntactic complexity or lexical diversity of a child’s writing is higher in a text written at the end of a programme of study than in one written at the beginning, should we thereby infer that some property of that child’s mental linguistic system has changed? The chief difficulty in drawing such conclusions stems from the fact that written texts are the products of complex processes that incorporate many different types of influence. As various models of writing (Grabe and Kaplan, 1996; Hayes, 2012; McNamara, 1996; Weigle, 2002) and of language production more broadly (Bachman and Palmer, 1996; Canale and Swain, 1980) highlight, the language that ends up in any given text is influenced by a wide range of variables, including the topic, the genre, the intended audience, the available resources (e.g., time, reference works, computers, peer or teacher support), various aspects of the writer’s language proficiency, topic knowledge, motivation and attitude towards the task, and so on. Part of the strength of corpus research is that, unlike less naturalistic methods, it can capture the outcomes of these complex processes as they play out in real communicative situations. While this promotes the ecological validity of our work, however, it means that claims about individual constructs – such as lexical or syntactic proficiency – can be problematic. The linguistic features that appear in a text are not pure reflections of their writers’ language proficiencies, but rather the outcomes of countless interacting variables.

Considerations of this sort are commonplace in the language testing literature (e.g., Bachman and Palmer 1996. Fulcher 2017, McNamara 1996, Weigle 2002, Weir 2005). A typical response in that context is to attempt to control writing tasks such that the influence of variables we are not attempting to measure is neutralized (e.g., by ensuring all texts are written on the same topic and under the same contextual

<p><b>TEXT A</b> Should the Spangebob liv ande de si? <b>We have been discussting wheetheer</b> Spongbob should live under the sea. <b>Many people believe that</b> Spongbok should live ander the see, <b>bicase firstly</b> he soaks up water. <b>Secenly</b> hi wash the rocs in the see. <b>On the adh meni pepo beliv</b> he shut not liv in the see <b>bicoss</b> hi is not e fihs. <b>In conclusion there are many great reasons for and against</b> for spong bov should liv andr the see <b>bicos deris</b> hi hous. <b>Wat do uia finc?</b></p>
<p><b>TEXT B</b> Should Sbungab lifundur The see? <b>We have been discussing whether</b> Sbranchbab should lif under the sea. <b>Many people believe that</b> Sbonchbab should live undur the sea <b>because fersly</b> Sbonbob can sacup the sea and <b>secle</b> Sbanbab can clen the racs. <b>on the uver hand many dat blef</b> that Ssongbab should not live under the sea <b>because</b> he isent a fish and he lifs in a haus and lifs bsid the sea. <b>in conclusion there are many great reasons for and against</b> wefer Ssongbob should lif under The sea. but I blef that Ssongbob should lif under the sea because thas wer hes home is.</p>
<p><b>TEXT C</b> Should Jhingerbred man be eaten? <b>We have been discasing wether</b> Jhingrbred man should be eaten or not. <b>Many people belive that</b> the Jinjrbred man should be eaten <b>because firstly</b> he is quit mean and he runs away from the cow and the silly old horse. <b>Secondly</b> the cow and the silly old horse are hungry. <b>On the ather hand people disagry</b> they think Jinjer bread man shouldn't be eaten. <b>In conclusion there are many great reasons</b> why Gingrbread man shouldn't be eaten because he woudnt fill the animals up. But I belive he shouldn't be eaten because he is tiny. I think he should be eaten because he myte feell angry and chase the animals <b>what do you think?</b></p>
<p><b>TEXT D</b> <u>Should</u> Sleeping Beauty <u>stay</u> asleep? <b>We have been discussing wheather</b> Sleeping Beauty <u>should stay</u> asleep. <b>Many people belive that</b> Sleeping Beauty <u>should stay</u> <b>because fistly</b>, she has been sleeping for one hundred years. <b>Secondly</b>, she wants to be lazy. <b>Lastly</b> she doesn't want to work! <b>On the other hand some people dissagre</b> that's because she has been sleeping nearly forever. <b>In conclusion, there are many great reasons for and against</b> for sleeping Beauty to <u>stay</u> asleep but I think that Sleeping Beauty should have made the corect choice She has to wacke up because she is a princess. <b>What do you think?</b></p>

Fig. 4. *Balanced argument* texts from the Growth in Grammar corpus.

conditions), and much corpus research into child writing development has followed that example (e.g., [Beers and Nagy 2011](#), [Berman and Nir 2010](#), [Deane and Quinlan 2010](#), [Nippold et al., 2005](#); [Olinghouse and Wilson 2013](#)). The disadvantage of this approach, however, is that it means giving up on some of the key strengths of corpus research. Elicited texts are, by definition, not authentic (though they may be designed to emulate features of authentic texts). Both the types of texts that are written and the conditions under which they are produced (in terms of, for example, available resources, writer motivation, links to class content) may differ from that of children's usual educational world. By eliciting texts, we therefore lose much of the ecological validity that we were hoping to capture. On a more pragmatic note, because eliciting texts involves asking children to do something they would not otherwise be doing, so placing an additional burden on their usual schoolwork, it may not be possible to get texts in the large numbers that are typical of corpus research.

One example of the sorts of issues that arise for researchers who stick with the more ecologically valid *corpus as found texts* approach, can be seen in [Fig. 4](#). These texts were written by Year 2 children from a single school. They are representative examples from a group of 34 similar texts, written by 21 different children. In each text, the child follows the same formulaic scheme to offer a motion for debate, present arguments for and against that motion, and draw a conclusion. The texts cover a total of 15 different topics, though (as [Table 1](#) shows) some recur much more frequently than others.

This is a clear example of the linguistic content of children's texts being heavily influenced by the resources that have been made available to them, this time in the form of teacher input and class materials. These

**Table 1**  
Topics for the *balanced argument* task.

Topic	Occurrences
Should Spongebob live under the sea?	8
Should gingerbread man be eaten?	5
Should school be open five days a week?	4
Should Rapunzel stay in the tower?	3
Should Toys R Us close?	3
Should Jack steal the golden goose?	2
Should children walk to school?	1
Should Elsa have powers?	1
Should Sleeping Beauty stay asleep?	1
Should Spongebob go to work?	1
Should Tiana kiss the frog?	1
Should we do homework?	1
Should we eat cookies?	1
Should we eat healthy food?	1
Should we go to bed at 7.30?	1

children have obviously been given a skeleton argument structure, built around some useful formulaic expressions and cohesion markers, and asked to fill out the structure with appropriate content. The repetition of ideas between some texts (compare A and B in [Fig. 4](#)) suggests that some of the content is based on notes from a class discussion, whereas we can guess that topics that appear only once (e.g. D in [Fig. 4](#)) may be based on the writers' own ideas. The fact that individual children wrote multiple texts of this sort suggests that the structure was revisited on different occasions to explore different topics.

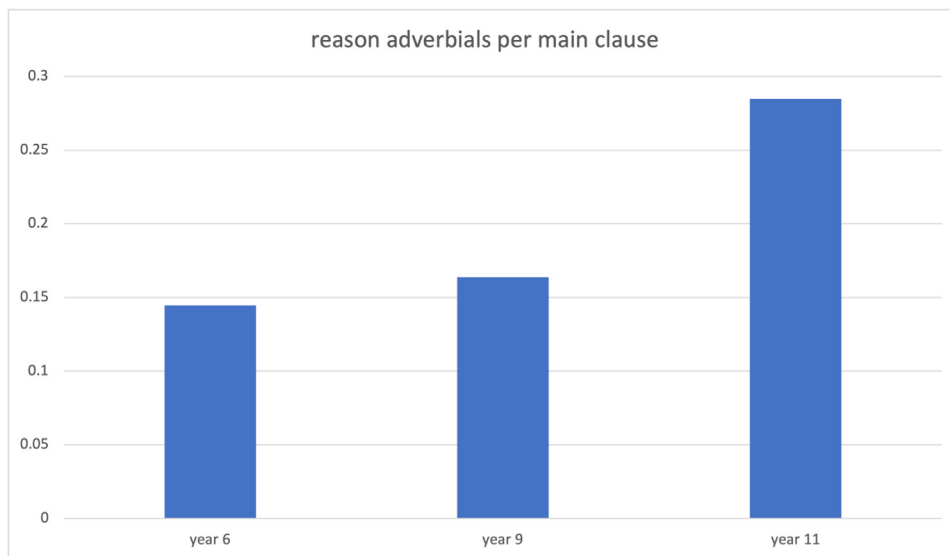


Fig. 5. Adverbial clauses of reason in children's non-literary writing.

The structures that are repeated across these texts are likely to lead to certain feature counts being strikingly high in this part of the corpus. In particular, there will be high frequencies for:

- Adverbial clauses: e.g., the many clauses starting *whether* or *because*.
- Complement clauses in object position: e.g., the many clauses following *believe* or *think*.
- The cohesion markers *firstly*, *secondly*, *thirdly*, and *on the other hand*.
- The auxiliary verb *should*.

The high levels of repetition between texts, and the shortness of each text, would result in counts for such features being high in terms of occurrences per word or occurrences per clause, and, except in a very large corpus, these would be likely to influence overall counts for Year 2 texts in general. The set of text-organising adverbials *firstly*, *secondly*, *thirdly*, for example occur in total 14.5 times per 10,000 words across the 278 Year 2 non-literary English texts in this corpus. This compares with only 1.4 occurrences per 10,000 words in the 368 corresponding Year 11 texts. Inspection of the cases identified in the Year 2 corpus reveals that all cases come from the task illustrated in Fig. 4.

So how should such differences be interpreted? Clearly, we cannot draw a straight line from these data to conclusions about differences between Year 2 and Year 11 students' proficiency with linking devices. Rather, these frequency data highlight an interesting feature of the class in which a particular group of students are writing. That is, it is telling us something about the pedagogical context of a particular Year 2 English class.

One response to a case like this could be simply to conclude that the subcorpus of 278 Year 2 texts needs to be expanded such that the influence of individual contexts is washed out. If the corpus were increased in size by a factor of ten, the cited frequency of text-organising adverbials would fall in line with that seen in Year 11 writing, assuming no further classes of a similar sort happened to be included. In general, this line of reasoning might run, we need to make sure that our corpora are sufficiently large that any quirks introduced by individual classrooms, teachers, topics, or any other influences that we do not want to focus on, are randomly distributed and therefore averaged out.

This approach runs into immediate practical difficulties, however. Collecting corpora of school writing requires the extensive goodwill of schools, busy teachers, their students, and those students' parents/guardians. It also requires painstaking transcription and anonymisation of mostly handwritten texts that are not always easy to decipher. These considerations mean that of creating a corpus of child writing large enough to iron out 'unwanted' influences may require resources that are beyond the reach of most projects.

More fundamentally, it could be argued that local patterns of the sort illustrated in Fig. 4 can provide important insights into writing development and that we should value the insights they provide, rather than attempting to wash them out. Writing development is, after all, a contextualized process and understanding the impact of individual classrooms, individual teachers, individual children's stylistic preferences, etc. should be an important part of what we study. In this regard, mixed-effects models (Gries, 2015; Winter, 2019) have much to offer corpus linguists in terms of identifying where such random variables impact on language use.

A second example of the variety of variables that influence school corpus data can be found in the use of adverbial clauses in the same corpus, as reported in Durrant et al. (2020). Amongst other things, that study investigated a striking pattern in the use of clauses indicating reasons. That is, clauses like:

- I believe that festival should not be banned **because it bring excitement and enjoyment to people from all areas.**
- It checks to see if there is a similar pattern in your work.

As Fig. 5 illustrates, our initial quantitative analysis suggested a sharp increase in the use of such clauses from Years 9 to 11. This raised the question of what such a substantial increase should be ascribed to. Unlike the previous example, it was not the product of an individual group of students. Reason clauses were found in writing from six of the nine schools that contributed Year 11 texts, with writing from four schools evidencing 21 or more examples each. Greater insight can be gained if we alter the way that texts are categorized. Specifically (and focusing now just on the gap between Year 9 and 11), if texts are divided between those written for English classes and those written for Science classes (as in Fig. 6), we see the frequency of reason clauses in Year 11 English classes remains in line with that of writing in both disciplines at Year 9 while their frequency in Year 11 Science increases dramatically. Whereas frequencies in previous year groups and in English writing at Year 11 hovers at around one reason adverbial for every five main clauses, in Year 11 Science, this ratio rises to one in two. We can therefore conclude that the increase in reason clauses at Year 11 is due to something particular to their Science writing.

So, what is driving this increase? The key appears to lie in the types of question that are set. Year 9 Science writing tasks are predominantly descriptive. Students are usually asked either to report on an experiment they have done to answer a particular question, e.g.

- Which metal is the softest?
- How does the temperature affect the number of times a ball bounces?

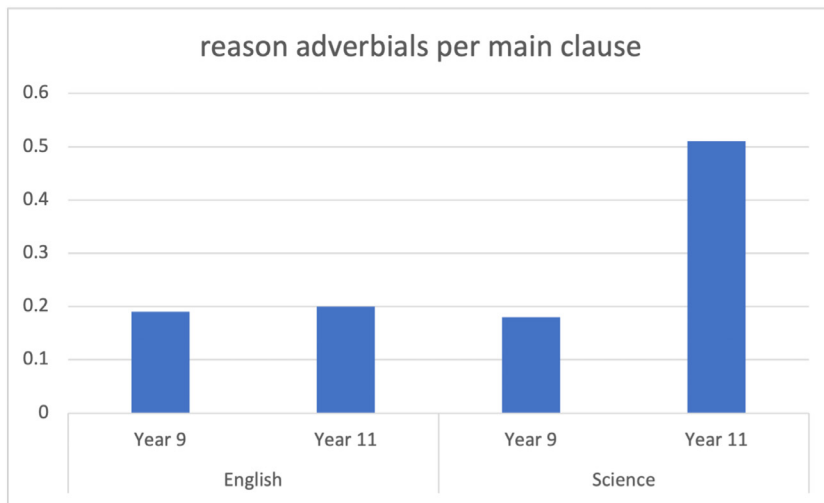


Fig. 6. Adverbial clauses of reason across years and disciplines.

or to describe a given phenomenon, e.g.

- Describe how radio waves and gamma rays differ in their properties and uses.
- Describe the harmful effects of infrared and ultraviolet waves, relating them to the frequencies of the waves.

Year 11 tasks, in contrast, almost invariably combine such descriptions with an explicit request for students to provide a reason for something, e.g.:

- After his operation Tim will regularly need to take a drug like aspirin or warfarin. His doctor shows him this graph. Warfarin reduces the rate at which fibrin is made. Aspirin stops platelets gathering together in blood vessels. Explain why aspirin and warfarin have the effects shown in the graph and how the graph convinces Tim that he should take one of the drugs.
- Look at the graphs. They show how the biomass of phytoplankton and zooplankton changes over a year in two places, the Arctic and the North Atlantic oceans. Describe the similarities and differences between Graph A and Graph B and suggest reasons for these differences.

It appears to be this shift in task type that is primarily responsible for the increase in reason adverbial clauses at this age group. Science teacher trainers inform me that this shift is driven by the need to prepare students for the tasks they are set in the national GCSE exams that children sit at the end of Year 11. As with the example of cohesion markers above therefore, a striking quantitative difference in the use of a particular linguistic features tells us primarily about something that is happening in the pedagogical context.

The general lesson of examples such as these is that researchers who work with authentic (rather than controlled/elicited) corpora of child writing need to maintain a broad view of the inferences to be drawn from their data. I argued in Section 2.3 that it may not be possible to determine in advance the most adequate linguistic and textual categories for analysis, and that a process of constructive cyclicity may be needed to work towards such categories. In much the same way, it may not be possible to determine in advance what inferences should be drawn from our quantitative analyses. The case of adverbial reason clauses illustrates how an initial quantitative analysis can draw our attention to a striking phenomenon that requires more detailed post-hoc analyses and engagement with individual (groups of) texts and their conditions of production before conclusions can be drawn.

#### 4. Looking ahead

Corpus methods have much to offer our understanding of children's writing. However, for this research to reach its full potential, it is im-

portant to reflect on their limitations. I have tried in this paper to draw out two sets of issues: categorization and valid interpretation of data. In both cases, I have argued for an analytical approach that involves sustained engagement with corpora, an engagement that continually interrogates the nature of our analytical categories and the meanings of our emerging findings. We need to ask how our results might have looked if we had categorised linguistic features or texts in a different way, or at a different level of resolution. We also need to ask what features of the writing context might best explain any quantitative patterns. This will often require us to form new hypotheses, to re-analyse data in ways that we had not expected, and to look in detail at the contents and contexts of individual texts and groups of texts. We may also find that we end up drawing conclusions about aspects of writing that we had not anticipated studying.

Fig. 7 attempts to summarize what this approach involves. Four analytical activities are key. Because these do not occur in a fixed order, and each can both build on and be informed by each of the others, I

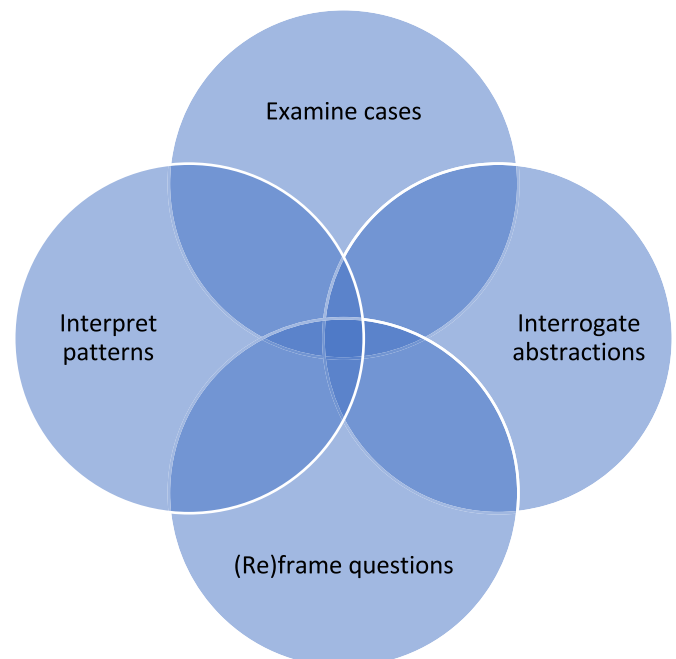


Fig. 7. Analytical activities.



**Table 2**  
Example register scores.

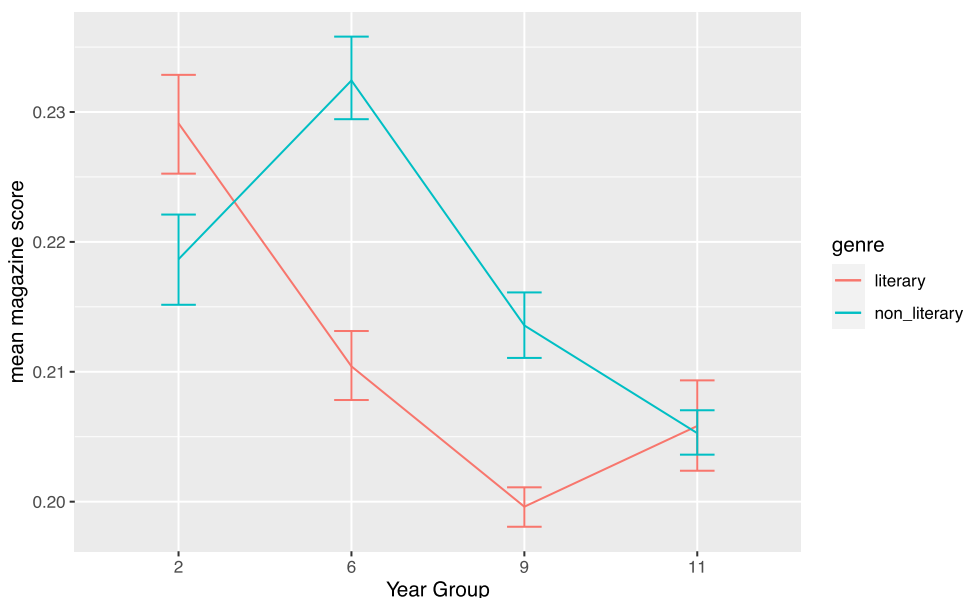
Lemma	POS	Register score			
		Academic	Fiction	Magazine	News
the	determiner	.28	.24	.24	.24
analysis	noun	.83	.02	.08	.07
happy	adjective	.07	.41	.28	.24

have represented them as overlapping fields, rather than a directional process.

The activities can be glossed as follows:

- *Interrogating abstractions* is the process of reflecting on the linguistic and textual categories used in an analysis. It involves asking questions like *how might my categories be more finely specified to reveal hidden patterns?* And *how might my categories be combined to reveal a bigger picture?*
- *Interpreting patterns* is the process of reflecting on what patterns in our data mean. It involves asking questions like *what aspects of the writing situation might have influenced these findings?* And *how might these patterns usefully inform pedagogy?*
- *Examining cases* is the process of studying examples of language use to inform the interpretation of patterns.
- *(Re)framing questions* is the process of determining what we want to learn from the data. I have called the activity *(re)framing*, rather than simply *framing* to emphasize that new questions are likely to emerge as the analysis proceeds.

To understand how these work in practice, and how they can mutually inform each other, I will finish by describing a brief case study that extends an analysis originally reported in Durrant & Durrant (2022). That study considered how the register of children's vocabulary changes as they progress through their schooling in England using a measure which assigns words a register profile comprising four scores, as illustrated in Table 2. The scores represent how strongly a word is associated with each of four registers – academic, fiction, magazine, and news – and is calculated based on that word's frequency across registers in a reference corpus (in this case, the Corpus of Contemporary American – COCA – Davies, 2008). The four scores always sum to 1.0, so a score of .25 indicates no bias towards a register. Thus, *the* is relatively neutral, whereas *analysis* and *happy* show strong biases towards academic writing and fiction, respectively.



**Fig. 8.** Magazine vocabulary across year groups and genres in children's writing.

To trace register variation across year groups, each word in each learner text is assigned a register profile. Mean scores are then calculated for each register for each text. Thus, each text had a set of four register scores showing, on average, the extent to which its vocabulary is biased towards each register. Our previous research had shown how children's developing writing is characterized by an increase in academic vocabulary and a decrease in fiction-like vocabulary, especially when they are writing in non-literary genres (such as essays, historical recounts, and scientific reports) (Durrant & Brenchley, 2019). In Durrant & Durrant (2022), we expanded this analysis to include the magazine and news registers. This was intended to give a broader picture of how children's vocabulary use orients towards other key written registers to which they are likely to be exposed. In this analysis, the magazine register showed the intriguing pattern illustrated in Fig. 8.

**Interpreting this pattern**, we might say that both literary and non-literary writing tended to avoid magazine-like vocabulary overall, since all scores are below .25. We might also say that writing tends to become less magazine-like as children mature, though this trend is not linear. In general, therefore, magazine vocabulary looks like a dispreferred feature, which children use less as they mature. Given this conclusion, however, two features of Fig. 8 are puzzling and might push us to **frame** some further **questions**. Specifically, we might ask why scores increase in non-literary writing at Year 6 and why they are higher in non-literary than in literary writing at both Years 6 and 9.

To address these questions, we can **examine cases** by looking at the words which underlie the high scores in Year 6 non-literary writing. Words contribute to an overall text score as a function of two things: the strength of their bias towards the magazine register, and their frequency in the text. The words which contribute most to the score will be those which are both strongly magazine-like and highly frequent. To quantify this, a *magazine weight* was calculated for each word by multiplying its magazine register score by its frequency of use within a year group and genre. For each year group, the top 25 items accounted for a large majority of the total magazine weight for the subcorpus (Year 2: 92%; Year 6: 90%; Year 9: 85%). Examining these words can therefore give a strong idea of the reasons for the overall patterns.

Fig. 9 shows the top 25 words by weight in the Year 6 non-literary subcorpus. Two things stand out immediately: all of these words appear to be more closely associated with Science than with other subject areas, and a majority are nouns. These considerations might lead us to **interrogate** our **abstractions** in two ways. First, the initial analysis conflates texts from three subject areas: English, Humanities, and Science. The words in Fig. 9 suggest that we might benefit from unpacking these.

yeast (noun); sugar (noun); microbe (noun); bacteria (noun); meteor (noun); bulb (noun); cup (noun); flour (noun); amp (noun); mixture (noun); layer (noun); brightness (noun); meteorite (noun); stir (verb); tablespoon (noun); tortoise (noun); antibacterial (adjective); croc(noun); penguin (noun); colour (noun); epiphyte (noun); gel (noun); battery (noun); yogurt (noun); heat (noun)

Fig. 9. Strongest magazine-like words in Year 6 non-literary writing.

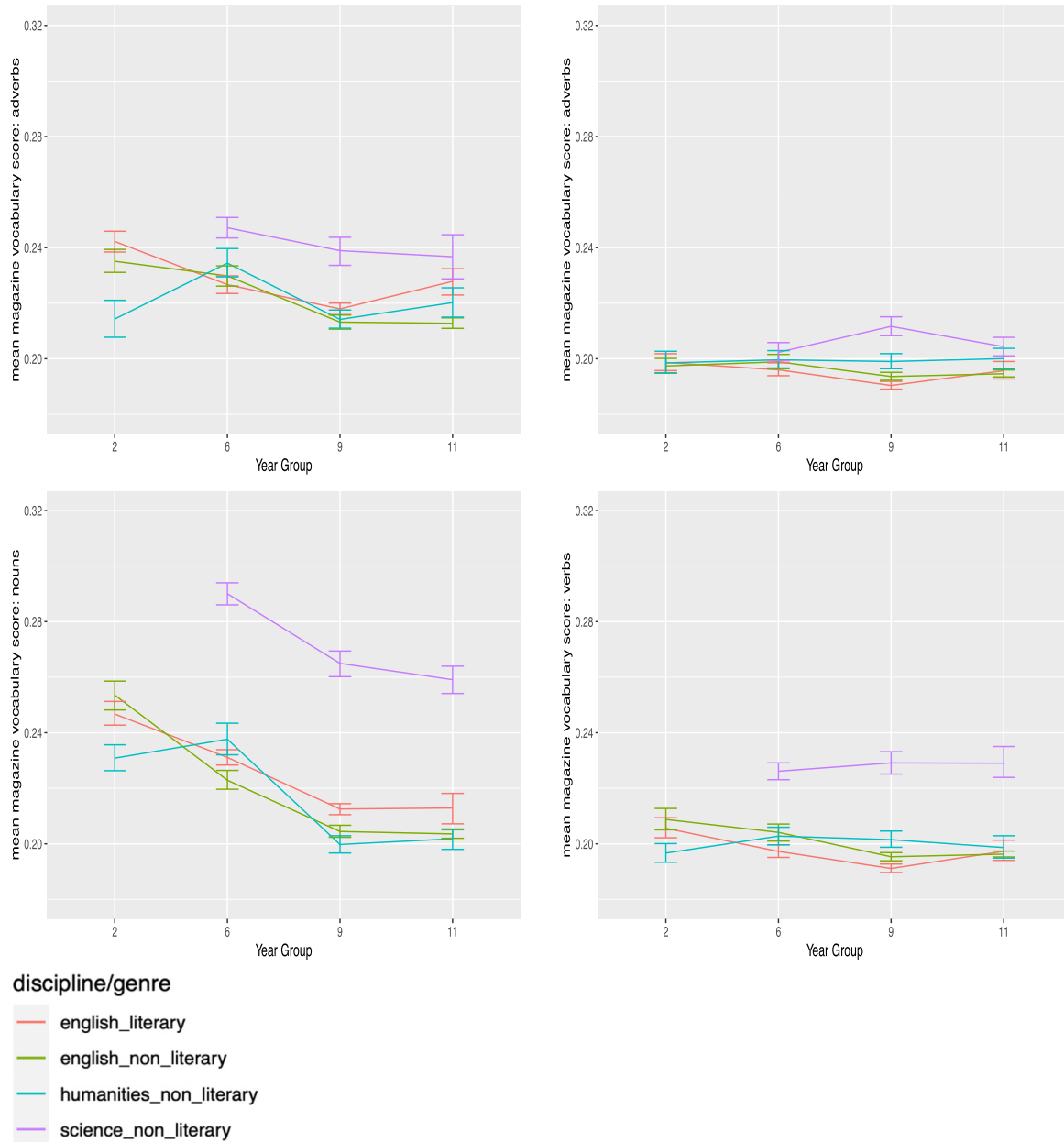


Fig. 10. Magazine vocabulary in different parts of speech.

Second, the analysis does not distinguish between words with different parts of speech. Further insights might therefore be gained if these are separated.

Fig. 10 repeats the quantitative analysis from Fig. 8 but with writing in different subject areas and words with different parts of speech distinguished. These confirm the trends we suspected from our inspection of Fig. 9. That is, use of magazine-like vocabulary is a feature of Science writing, rather than of non-literary writing in general, and this

bias is strongest for nouns. While Science writing does show a tendency towards magazine words across all parts of speech, the difference from other areas is by far the greatest in nouns, and this part of speech alone has mean scores above the 'neutral' score of .25. It also appears to be nouns which drive the decrease in scores from Years 6 to 11. Unpacking the data in this way also highlights the fact that there is no Year 2 Science writing in the corpus. This is likely to explain the increase in scores seen from Years 2 to 6 in Fig. 8.

**Table 3**  
Most heavily weighted magazine words in science writing.

Field of use	Year 6	Year 9	Year 11
Food & nutrition	yeast; sugar; microbe; flour; mixture; tablespoon; yogurt	acid; microwave; diet; fruit; nutrient; vitamin	enzyme; acid; cheese; carrot; rennet; casein; carbonate; calcium; texture; protein; layer; whey; nutrient; sugar
Experiments	yeast; sugar; bulb; cup; amp; mixture; brightness; tablespoon; croc; battery; teaspoon	colour; acid; magnesium; aluminium; heat; squash; zinc; bulb; calcium; battery; powder colour	enzyme; carrot; rennet; colour; heat; texture; protein; thermostat; whey; salt
Animals, their properties, behaviours and environment	meteor; layer; meteorite; tortoise; penguin; color; heat		
Disease and illness	bacteria; microbe; chickenpox	cancer; acid; microwave; heat; retina; clot; fruit; nutrient; vitamin	acid; clot; heartburn
Plants	epiphyte	colour; stem; pea; seedling; tomato	
Electrical/electronic	bulb; amp; brightness; croc; battery	bulb; battery	nutrient

Armed with this new analysis, we can **reframe** our **questions**, asking why magazine-like nouns are so prominent in Science writing, and why this prominence decreases across year groups.

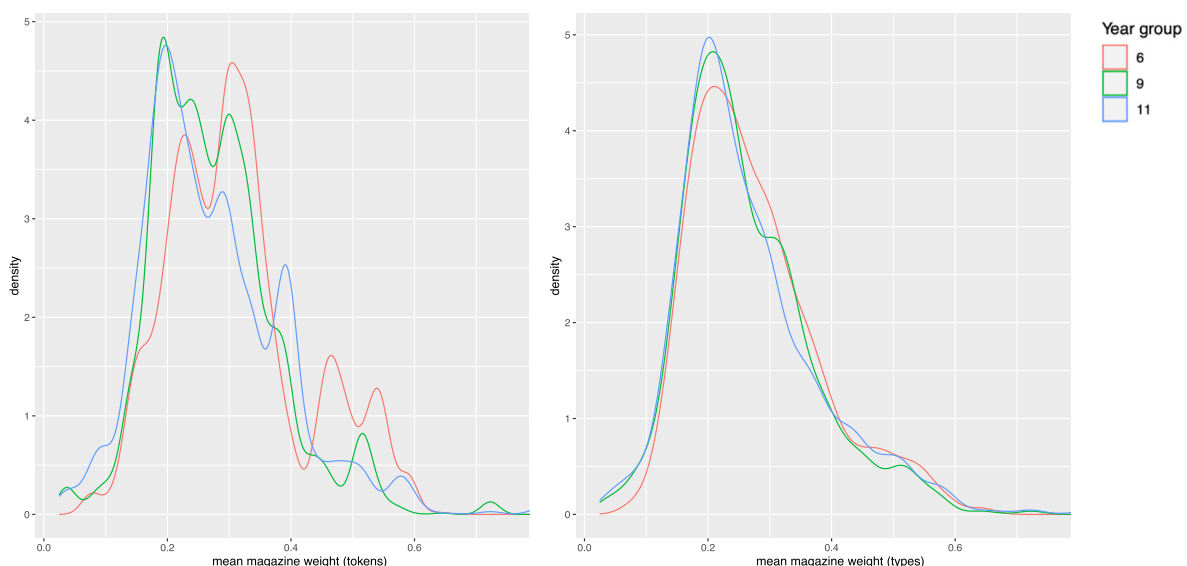
To address the first question, we can again **examine cases** by looking at nouns with the highest magazine weights in Science writing. Each word was examined in context within the learner corpus and classified according to its overall field of use. As [Table 3](#) summarizes, these words were organised into six main categories. In order of frequency, these were: food and nutrition; experiments; animals (including their environment); disease and illness; plants; and electrical/electronic. These appear to be key topics in school science writing. Importantly, they overlap strongly with the topics which are prominent in the magazine section of the COCA corpus. A separate analysis ([Durrant & Durrant, 2022](#)) of the fine-grained classifications of COCA texts in which magazine-like words that are prominent in child writing occurred showed that the majority were found in *Health and Home* magazines, and were used in the context of food (e.g., *flour, fruit, heat*), health and nutrition (e.g., *anti-bacterial; cancer; diet*), and gardening (e.g., *bulb; epiphyte; stem*). It seems that the high magazine scores seen in children's Science writing primarily reflects this overlap in topics.

Our second question was why the prominence of magazine-like words decreased across year groups. This cannot be accounted for in terms of topics: while some (e.g., *animals*) decrease across the year groups, others (e.g., *food and nutrition*) increase. We can, however, get fresh insight by again **interrogating** our (linguistic) **abstractions**. Our initial analysis has been based on word *tokens*. We could move to a higher level of abstraction by combining repeated uses of each word

to focus on word *types*. [Fig. 11](#) shows density plots which compare the magazine weights of Science texts based on tokens (left-hand box) and types (right-hand box). These demonstrate how, while the younger children's texts frequently score very highly on the magazine measure when we look at word tokens, the three groups are almost exactly equal when we look at word types. **Interpreting** these **patterns**, we can say that the repertoire of words used by the younger children is no more magazine-like than that used by older children. Their higher scores must, rather, be due to extensive repetition of high-scoring words.

We can get an understanding of this tendency by again **examining cases**. [Fig. 12](#) illustrates how four words with high magazine scores (*yeast, sugar, microbe, and bacteria*) are used and re-used in Year 6 Science texts. This highlights how writers at this level maintain cohesion through frequent repetition of nouns that are central to their topic. Since, as we have seen, these topic-related nouns are the principal drivers of the bias towards magazine words in Science writing, this results in higher overall scores for texts in this year group.

This analysis started with a graph ([Fig. 8](#)) that purported to tell us about the register appropriateness of children's vocabulary. It appeared to show that magazine-like words were generally not used in children's writing and that their use decreased across year groups. This could be taken to suggest that learning to write with a less magazine-like vocabulary is an aspect of mature school writing. However, closer inspection of the data, involving examination of cases, reframing of questions, interpretation of patterns, and interrogation of textual and linguistic abstractions, has ultimately led us to rather different conclusions: that school science writing has a strong topical overlap with *Health and Home* style



**Fig. 11.** Mean magazine score for each Science text, across year groups.

To observe and record the differences between the growth of **yeast** cells with and without **sugar**. 2 plastic cups flour **yeast** solution & **sugar** 2 cylinders bowl or tray with hot water. First I labelled both cups one saying **yeast** and water and the other saying **yeast**, water and **sugar**. Then I added four dessert of flour in each cup. Next I added a solution of **yeast** to cup A until it had the consistency of a thick, milkshake. After that I added enough **yeast** water and **sugar** to the solution. Then I poured solution A and B into two different cylinders up to 30ml. Next record and observe the solutions every five minutes. Also put the cylinders in warm water. Looking at our results **yeast** and **sugar** rose the most because **yeast** needs to eat so the **yeast** needs the **sugar** to make it rise. As the water gets cooler and the **yeast** has eaten all the **sugar** it will start to go down. **Yeast** is a fast eating bacteria.

**Microbes** can be good or they can be bad. **Microbes** have been round for 3.8 BN years. If there was no **microbes** the world would be over!!!! **Microbes** are all around you. They are even inside you. They are everywhere!!!!!!!!!!!!!!!!!!!!!! In 1 litre of sea water there is 1BN **microbes** and in 1 tea spoon of soil there is 100,000 **microbes**. There is three classes of **microbes**. The first one is **bacteria**. The second one is virus. The third one is fungi! Wash your hands before cooking, before eating. You need to put on deodorant. You better clean your teeth. The **microbes** in the world are mostly bad. Fungi, **bacteria** and virus are mostly all bad, but some **bacteria** and virus is good. So you better look out. There is some good **microbes** in the world. Penicillin and yeast is good for you but it is **microbes**. This shows the **microbes** can be good or bad. Did you know that in one tea spoon of soil there is 100,000 **microbes**!!! Did you know that in one litre of sea water there is 1BN **microbes**!!! Every living creature on earth has **bacteria** on the body. We are a good habitat for **bacteria** because we are warm and our skin is moist: our body releases sweat, salt and chemicals that **bacteria** feast on. **Bacteria** also feasts on leftover sugar that is why **bacteria** lives in your mouth. How much **bacteria** is there on me? **Bacteria** looks like this.

magazines (especially their interests in food, health, and gardening), and that some younger children maintain text cohesion by frequent repetition of topic-related nouns. Although this was not the type of conclusion we were initially looking for, it appears to be more faithful to our data than the original interpretation, and has taught us something new about the changing nature of children's school writing.

From one perspective, this process has been rather inconvenient. Instead of specifying in advance that we want to learn something about linguistic development, then devising and carrying out an analysis that will lead to clear and confident conclusions about that construct, we are forced to acknowledge that our planned analyses might not turn out to give us all the information we need and that we cannot even be sure in advance what constructs we will learn most about.

It is also possible to put a more positive spin on things, however. Prolonged engagement with corpus data from multiple perspectives, where analysis can develop organically and in ways not initially envisaged by the researcher as they work towards a coherent and satisfying interpretation, can be seen as a less positivistic way of working with a corpus, and one that is likely to yield more valid interpretations of data. It also highlights the ways that authentic (as opposed to elicited) learner corpora can provide insights into a rich range of educationally important elements of school writing: genres, disciplines, topics, classroom practices, and so on. This may provide a picture that is both more educationally useful and more faithful to reality. Learning to write is a highly contextualized process in which children engage with language to express themselves in relation to specific topics, within specific genres and disciplines, in specific, dynamic, classroom settings. Attempting to describe such development in terms of a few pre-determined categories is likely both to be educationally unenlightening and to provide a distorted picture of development. In this context, I would argue that there is much to be gained from embracing an exploratory and holistic approach to the analysis of school corpora.

## Funding

This work was supported by a grant from the [Economic and Social Research Council \(ES/M00967X/1\)](#).

## 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.

## Acknowledgments

The author would like to thank Mark Brenchley and Hildegunn Dirdal, who commented on an earlier draft of this paper, as well as the editors and anonymous reviewers, whose helpful suggestions have strengthened the final product.

## References

- Bachman, L., Palmer, A., 1996. *Language Testing in Practice*. Oxford University Press, Oxford.
- Becher, T., Trowler, P., 2001. *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines*. SRHE and Open University Press, Buckingham.
- Beers, S., Nagy, W.E., 2011. Writing development in four genres from grades three to seven: syntactic complexity and genre differentiation. *Read. Writ.* 24, 183–202.
- Beers, S.F., Nagy, W.E., 2011. Writing development in four genres from grades three to seven: syntactic complexity and genre differentiation. *Read. Writ.* 24 (2), 183–202.
- Berman, R.A., Nir, B., 2010. The lexicon in writing-speech-differentiation. *Writ. Lang. Lit.* 13 (2), 183–205.
- Berman, R.A., Nir-Sagiv, B., 2007. Comparing narrative and expository text construction across adolescence: a developmental paradox. *Discourse Process.* 43 (2), 79–120.
- Biber, D., 1993. Representativeness in corpus design. *Lit. Linguist. Comput.* 8 (4), 243–257.
- Biber, D., Gray, B., Staples, S., Egbert, J., 2020. Investigating grammatical complexity in L2 English writing research: linguistic description versus predictive measurement. *J. Engl. Acad. Purp.* 46, 1475–1585.

Fig. 12. Repetition of magazine words in Year 6 Science writing.

- Bulté, B., Housen, A., 2012. Defining and operationalising L2 complexity. In: Housen, A., Kuiken, K., Vedder, I. (Eds.), *Dimensions of L2 Performance and Proficiency. Complexity, Accuracy and Fluency in SLA*. John Benjamins, Amsterdam, pp. 21–46.
- Bulté, B., Housen, A., 2014. Conceptualizing and measuring short-term changes in L2 writing complexity. *J. Second Lang. Writ.* 26, 42–65. doi:10.1016/j.jslw.2014.09.005.
- Canale, M., Swain, M., 1980. Theoretical bases of communicative approaches to second-language teaching and testing. *Appl. Linguist.* 1, 1–47.
- Christie, F., 2012. A language theory for educational practice. *Lang. Learn.* 62, 1–31.
- Christie, F., Derewianka, B., 2008. *School Discourse: Learning to Write Across the Years of Schooling*. Continuum, London.
- Coffin, C., 2006. Learning the language of school history: the role of linguistics in mapping the writing demands of the secondary school curriculum. *J. Curric. Stud.* 38 (4), 413–429.
- Cox, B.E., Shanahan, T., Tinzmann, E., 1991. Children's knowledge of organization, cohesion, and voice in written exposition. *Res. Teach. Engl.* 25 (2), 179–218.
- Davies, M. (2008). *The Corpus of Contemporary American: 450 million words, 1990-present*. Retrieved from <http://corpus.byu.edu/coca/> on 1 May 2018.
- Deane, P., Quinlan, T., 2010. What automated analyses of corpora can tell us about students' writing skills. *J. Writ. Res.* 2 (2), 151–177.
- Department for Education. (2014). *The National Curriculum in England. Framework document*. Retrieved from <https://www.gov.uk/government/publications/national-curriculum-in-england-framework-for-key-stages-1-to-4> on 22 April 2018.
- Durrant, P., 2017. Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories. *Applied Linguistics* 38 (2), 165–193.
- Durrant, P., Brenchley, M., 2019. Development of vocabulary sophistication across genres in English children's writing. *Reading and Writing* 32 (8), 1927–1953.
- Durrant, P., Brenchley, M., 2021. The development of academic collocations in children's writing. In: Szudarski, P., Barclay, S. (Eds.), *Vocabulary Theory, Patterning and Teaching*. Multilingual Matters, Bristol, pp. 99–120.
- Durrant, P., Brenchley, M., Clarkson, R., 2020. Syntactic development across genres in children's writing: The case of adverbial clauses. *Journal of Writing Research* 12 (2), 419–452.
- Durrant, P., Brenchley, M., McCallum, L., 2021. *Understanding development and proficiency in writing: quantitative corpus linguistic approaches*. Cambridge University Press, Cambridge.
- 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.
- Fang, Z., 2002. The construction of literate understanding in a literature-based classroom. *Journal of Research in Reading* 25 (1), 109–126.
- Fulcher, G., 2017. Scoring performance tests. In: Fulcher, G., Davidson, F. (Eds.), *The Routledge Handbook of Language Testing*. Routledge, pp. 378–392.
- Golub, L.S., Frederick, W.C., 1979. *Linguistic Structures in the Discourse of Fourth and Sixth Graders*. Center for Cognitive Learning, The University of Wisconsin.
- Grabe, W., Kaplan, R.B., 1996. *Theory and Practice of Writing: An Applied Linguistic Perspective*. Longman, New York.
- Gray, B., 2019. Tagging and counting linguistic features for multi-dimensional analysis. In: Berber Sardinha, T., Pinto, M.V. (Eds.), *Multi-Dimensional Analysis: Research Methods and Current Issues*. Bloomsbury Academic, London, pp. 43–66.
- Gries, S.T., 2015. The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora* 10 (1), 95–125.
- Hall-Mills, S., Apel, K., 2015. Linguistic feature development across grades and genre in elementary writing. *Lang. Speech Hear. Serv. Sch.* 36, 242–255.
- Hayes, J.R., 2012. Modelling and remodeling writing. *Writ. Commun.* 29 (3), 369–388.
- Hillocks, G., 1964. An analysis of some syntactic patterns in ninth grade themes. *J. Educ. Res.* 57 (8), 417–420.
- Huang, Y., Murakami, A., Alexopoulou, T., Korhonen, A., 2018. Dependency parsing of learner English. *Int. J. Corpus Linguist.* 23 (1), 28–54.
- Huddleston, R., Pullum, G.K., 2002. *Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge.
- Hunston, S., 2002. *Corpora in Applied Linguistics*. Cambridge University Press, Cambridge.
- Hyland, K., 2008. Genre and academic writing in the disciplines. *Lang. Teach.* 41 (4), 543–562.
- Koutsoftas, A.D., Gray, S., 2012. Comparison of narrative and expository writing in students with and without language-learning disabilities. *Lang. Speech Hear. Serv. Sch.* 43, 395–409.
- Lines, H., 2014. *It's a Matter of Individual Taste, I Guess: Secondary School English Teachers' and Students' Conceptualisations of Quality in Writing*. University of Exeter, Exeter (PhD).
- McNamara, T.F., 1996. *Measuring Second Language Performance*. Longman, London.
- Myhill, D., 1999. Writing matters: linguistic characteristics of writing in GCSE English examinations. *Engl. Educ.* 33, 3.
- Newman, J., Cox, C., 2020. Corpus annotation. In: Paquot, M., Gries, S.T. (Eds.), *A Practical Handbook of Corpus Linguistics*. Springer Nature, pp. 25–48.
- Nippold, M.A., Ward-Loneragan, J.M., Fanning, J.L., 2005. Persuasive writing in children, adolescents, and adults: a study of syntactic, semantic, and pragmatic development. *Lang. Speech Hear. Serv. Sch.* 36 (2), 125–138.
- 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.
- Olinghouse, N.G., Wilson, J., 2013. The relationship between vocabulary and writing quality in three genres. *Read. Writ. Interdiscip. J.* 26, 45–65.
- Percival, E., 1966. The dimensions of ability in English composition. *Educ. Rev.* 18 (3), 205–212.
- Potter, R.R., 1967. Sentence structure and prose quality: an exploratory study. *Res. Teach. Engl.* 1 (1), 17–28.
- Reppen, R., 1994. *Variation in Elementary Student Language: A Multi-Dimensional Perspective*. Northern Arizona University (Doctor of Philosophy in Applied Linguistics).
- Reppen, R., 2001. Register variation in student and adult speech and writing. In: Conrad, S., Biber, D. (Eds.), *Variation in English: Multi-Dimensional Studies* (Eds.). Taylor & Francis.
- Reppen, R., 2002. A genre-based approach to content writing instruction. In: Richards, J.C., Renandya, W.A. (Eds.), *Methodology in Language Teaching: An Anthology of Current Practice*. Cambridge University Press, pp. 321–327.
- Reppen, R., 2007. First language and second language writing development of elementary students: two perspectives. In: Kawaguchi, Y., Takagaki, T., Tomimori, N., Tsuruga, Y. (Eds.), *Corpus-Based Perspectives in Linguistics*. John Benjamins Publishing Company, pp. 147–167.
- Reppen, R., 2009. Exploring L1 and L2 writing development through collocations: a corpus-based look. In: Barfield, A., Gyllstad, H. (Eds.), *Researching Collocations in Another Language*. Palgrave Macmillan, pp. 49–59.
- Roessingh, H., Elgie, S., Kover, P., 2015. Using lexical profiling tools to investigate children's written vocabulary in grade 3: an exploratory study. *Lang. Assess. Q.* 12 (1), 67–86.
- Rose, D., Martin, J.R., 2012. *Learning to Write/Reading to Learn: Genre, Knowledge and Pedagogy in the Sydney School*. Equinox, Sheffield.
- Sadler, D.R., 1989. Formative assessment and the design of instructional systems. *Instr. Sci.* 18 (2), 119–144.
- Sampson, O.C., 1964. A linguistic study of the written compositions of ten-year old children. *Lang. Speech.* 7 (3), 176–182.
- Schuster, S., Manning, C.D., 2016. Enhanced English universal dependencies: an improved representation for natural language understanding tasks. In: *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pp. 2371–2378.
- Swales, J., 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press, Cambridge.
- Trowler, P., Saunders, M., Bamber, V., 2012. *Tribes and Territories in the 21st Century*. Routledge, London.
- Uccelli, P., Dobbs, C.L., Scott, J., 2013. Mastering Academic Language: Organization and Stance in the Persuasive Writing of High School Students. *Written Communication* 30 (1), 36–62.
- Weigle, S.C., 2002. *Assessing Writing*. Cambridge University Press, Cambridge.
- Weir, C., 2005. *Language Testing and Validation: An Evidence-Based Approach*. Palgrave Macmillan, Basingstoke.
- Winter, B., 2019. *Statistics for Linguists: An Introduction Using R*. Routledge, London.
- Yates, C., Berninger, V., Abbott, R., 1995. Specific writing disabilities in intellectually gifted children. *Talents and Gifts* 19 (2), 131–155.